Truth Will Out: Departure-Based Process-Level Detection of Stealthy Attacks on Control Systems

Wissam Aoudi
Chalmers University

Mikel Iturbe
Mondragon University

Magnus Almgren
Chalmers University

CHALMERS
UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG

Industrial Control Systems

Industrial Control Systems (ICS)

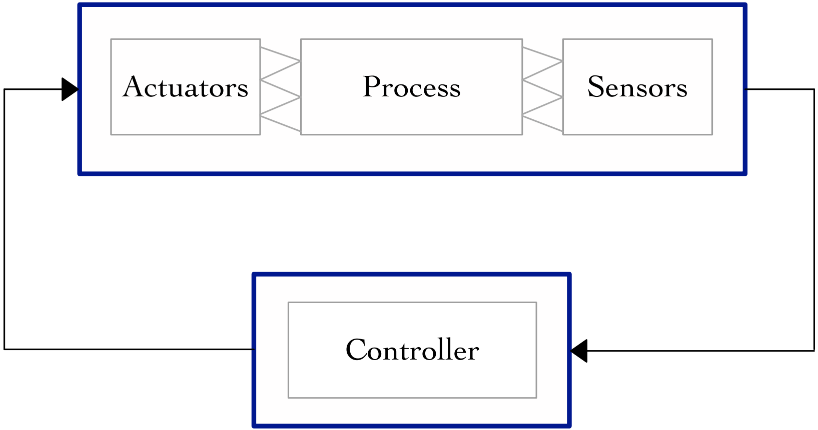
- control industrial processes;
- typically operate on critical infrastructures.

The Problem

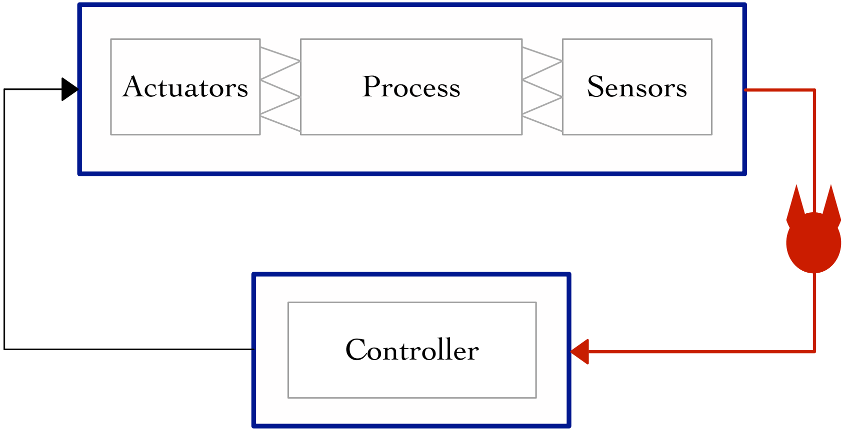
- Attacks on ICS are increasing.
- Successful attacks on ICS
  - can be highly rewarding for attackers;
  - may have devastating consequences on society at large.
- Classical IT-based security is not sufficient.

Process-Level Attack Detection

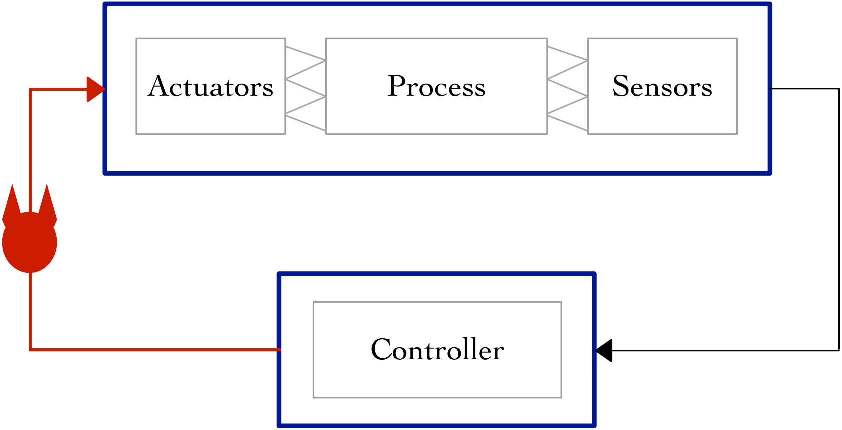
Why?	Because ICS combine both IT and OT technologies.
What?	Check if **physical process** deviates from **the norm**.
How?	By monitoring **process output** in real time.


Actuators
Process
Sensors
Controller

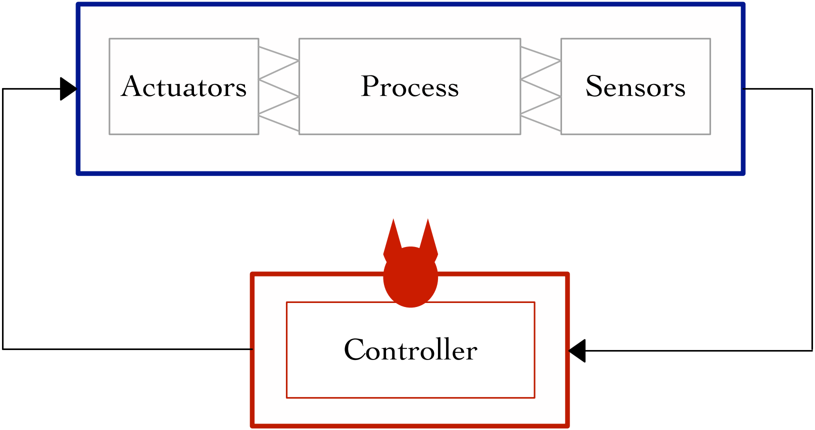
Control Loop and Attacker Model



Control Loop and Attacker Model



Control Loop and Attacker Model



CHALMERS

ICS-Specific Features

- Controllers (e.g., PLCs) operate in a **cyclic** manner.
- Signals repeat $\Rightarrow$ level of **determinism** is relatively high.
- Normal behavior can be **learned** or **modeled**.

ICS-Specific Features

- Controllers (e.g., PLCs) operate in a **cyclic** manner.

Regularity of ICS behavior enables data-driven approaches.

- Normal behavior can be **learned** or **modeled**.

Existing Methodology[1]

Build a model of the physical process

↓

Use the model to **predict** future system behavior

↓

Monitor residuals: Is |**observed** − **predicted**| too large?

[1] Urbina, David I., et al. "Limiting the impact of stealthy attacks on industrial control systems." Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2016.

Existing Methodology[1]

Build a model of the physical process

↓

Use the model to predict future system behavior

Solving a more general problem as an intermediate step!

↓

Monitor residuals: Is $|\textbf{observed} - \textbf{predicted}|$ too large?

[1] Urbina, David I., et al. "Limiting the impact of stealthy attacks on industrial control systems." Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2016.

PASAD

1. solves an easier problem;
2. requires limited knowledge of system dynamics;
3. is capable of detecting subtle changes in system behavior.

CHALMERS

PASAD

1. solves an easier problem:

Learns normal behavior from historical data

↓

Measures to what extent **present** readings **conform** with the estimated dynamics.

PASAD

1. solves an easier problem:

Learns normal behavior from historical data

No need to predict the future!

Measures to what extent **present** readings **conform** with the estimated dynamics.

PASAD

2. requires limited knowledge of system dynamics:

- It is entirely data-driven.
- Uses only **raw** sensor readings.
- It is model-free.

CHALMERS

PASAD

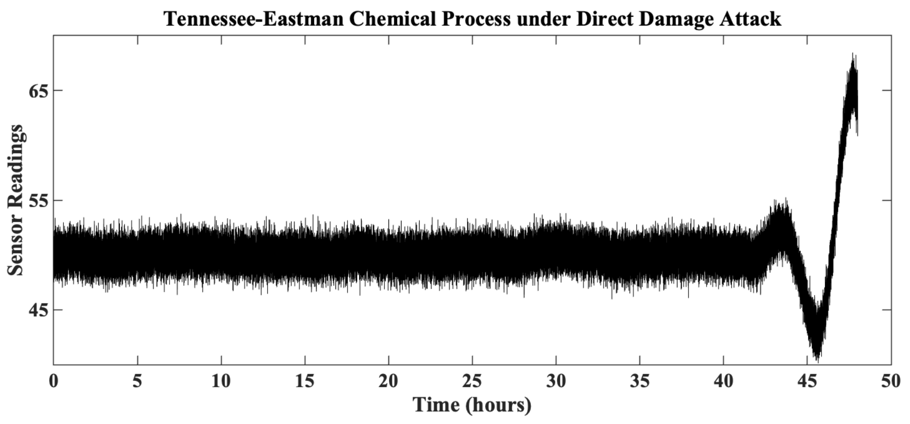
2. requires limited knowledge of system dynamics:

- It is entirely d
- Uses only **rav**
- It is model-free.

PASAD is specification-agnostic.
Applicable to various systems.

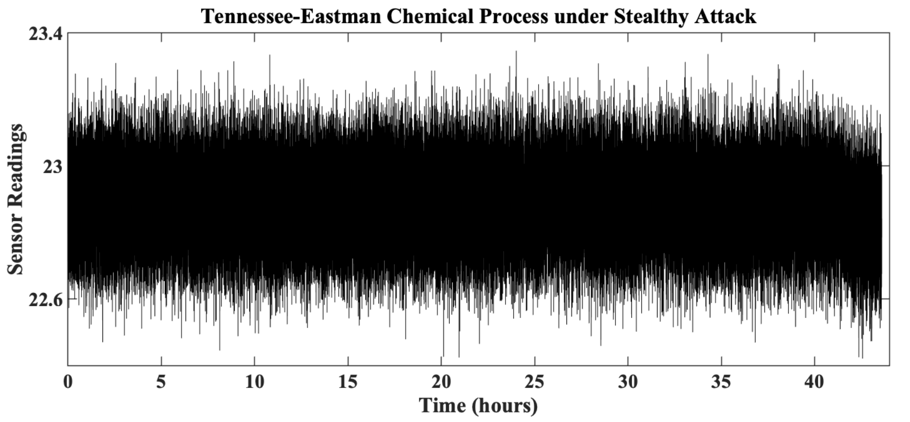
PASAD

3. is capable of detecting subtle changes in system behavior:



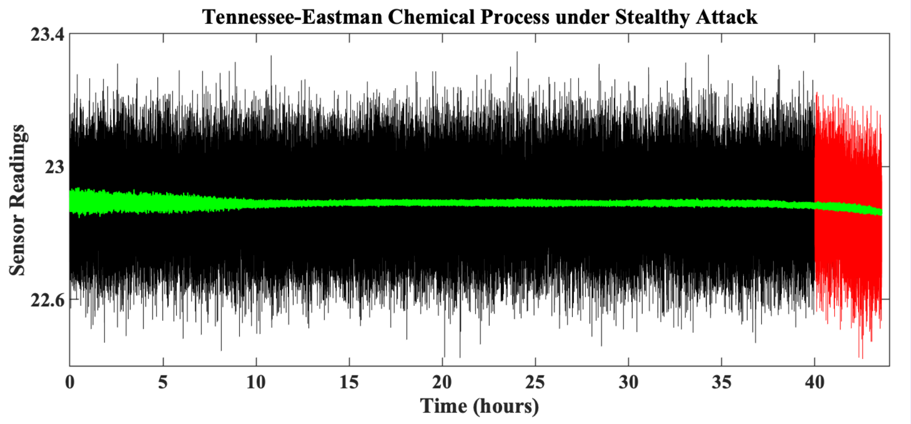
PASAD

3. is capable of detecting subtle changes in system behavior:



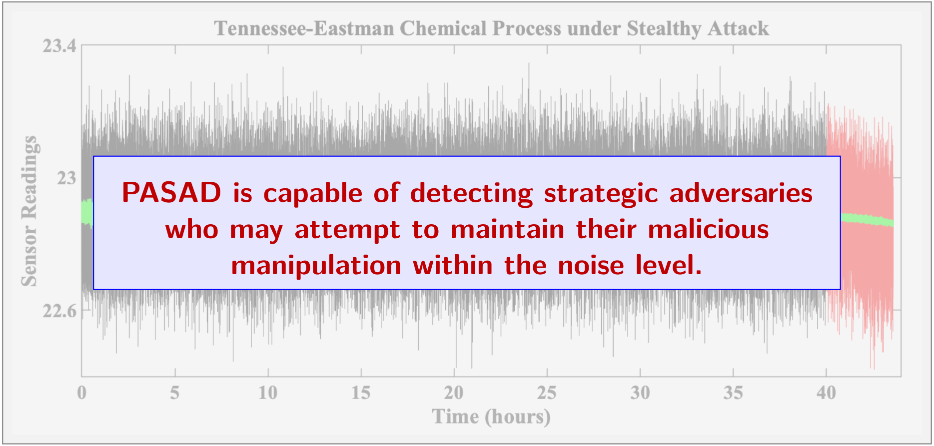
PASAD

3. is capable of detecting subtle changes in system behavior:



PASAD

3. is capable of detecting subtle changes in system behavior:



PASAD is capable of detecting strategic adversaries who may attempt to maintain their malicious manipulation within the noise level.

PASAD — Process-Aware Stealth-Attack Detection

Rationale: Detect attacks on ICS by monitoring sensor measurements for unusual behavior.

PASAD works in two phases: *Offline learning* and *online detection*.

PASAD — Process-Aware Stealth-Attack Detection

Rationale: Detect attacks on ICS by monitoring sensor measurements for unusual behavior.

PASAD works in two phases: *Offline learning* and *online detection*.
Learning Phase: Create a mathematical representation of the *norm*

- Extract noise-reduced signal information from noisy time series of sensor readings.
- Construct *Signal Subspace* and project training vectors.
- Compute centroid of the cluster formed by training vectors.

PASAD — Process-Aware Stealth-Attack Detection

Rationale: Detect attacks on ICS by monitoring sensor measurements for unusual behavior.

PASAD works in two phases: *Offline learning* and *online detection*.
Learning Phase: Create a mathematical representation of the *norm*

- Extract noise-reduced signal information from noisy time series of sensor readings.
- Construct *Signal Subspace* and project training vectors.
- Compute centroid of the cluster formed by training vectors.

Detection Phase: Track distance from the centroid

- Project most recent measurement vector onto the subspace.
- Compute a *departure score*: distance from the centroid.
- Raise an alarm if a certain threshold is crossed.

Input: $\mathcal{T} = x_1, x_2, \cdots, x_N, x_{N+1}, \cdots$
Output: Alarm upon departure from normal behavior.

Learning Phase

Step 1: (Embedding)

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & \dots & x_{N-L+1} \\ x_2 & x_3 & \dots & x_{N-L} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \dots & x_N \end{bmatrix}$$

Input: $\mathcal{T} = x_1, x_2, \cdots, x_N, x_{N+1}, \cdots$
Output: Alarm upon departure from normal behavior.

Learning Phase

Step 2: (Singular Value Decomposition)

- Compute svd($\mathbf{X}$) to obtain the L eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_L$ of $\mathbf{X}\mathbf{X}^T$.
- Select $r < L$ leading eigenvectors.

The Two Phases of PASAD

Input: $\mathcal{T} = x_1, x_2, \cdots, x_N, x_{N+1}, \cdots$
Output: Alarm upon departure from normal behavior.

Learning Phase

Step 3: (Projection onto the Signal Subspace)

- Let $\mathbf{U} = [\mathbf{u}_1 : \mathbf{u}_2 : \cdots : \mathbf{u}_r]$ and $\mathcal{L}^r = \text{range}(\mathbf{U})$.
- Compute centroid as $\tilde{\mathbf{c}} = \mathbf{P}\mathbf{c}$, where $\mathbf{P} = \mathbf{U}\mathbf{U}^T$ is a projection matrix and $\mathbf{c}$ is the sample mean of training vectors.

The Two Phases of PASAD

Input: $\mathcal{T} = x_1, x_2, \cdots, x_N, x_{N+1}, \cdots$
Output: Alarm upon departure from normal behavior.

Detection Phase

Step 4: (Distance Tracking)

For every test vector $\mathbf{x}_j$ $(j > N - L + 1)$

- Compute the *departure score* as $D_j = ||\tilde{\mathbf{c}} - \mathbf{P}\mathbf{x}_j||^2$.
- Generate an alarm whenever $D_j \geq \theta$ for some threshold θ.

Input: $\mathcal{T} = x_1, x_2, \cdots, x_N, x_{N+1}, \cdots$
Output: Alarm upon departure from normal behavior.

Detection Phase

Step 4: (Distance Tracking)

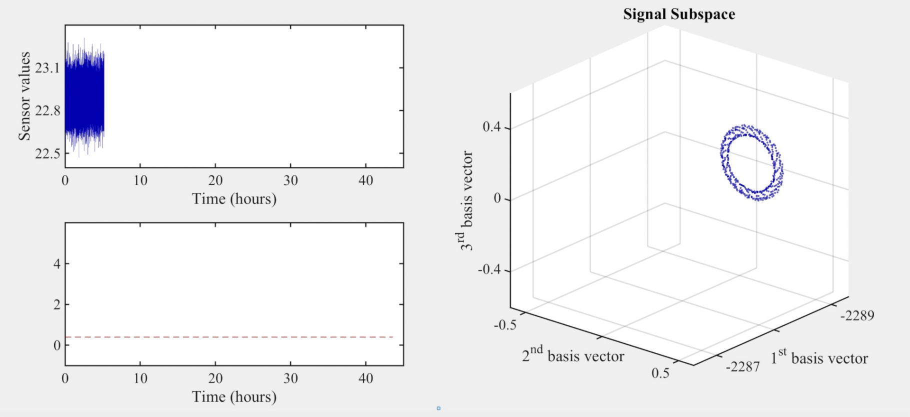
For every test vector $\mathbf{x}_j$ $(j > N - L + 1)$

- Compute the *departure score* as $D_j = ||\tilde{\mathbf{c}} - \mathbf{P}\mathbf{x}_j||^2$.
- Generate an alarm whenever $D_j \geq \theta$ for some threshold θ.

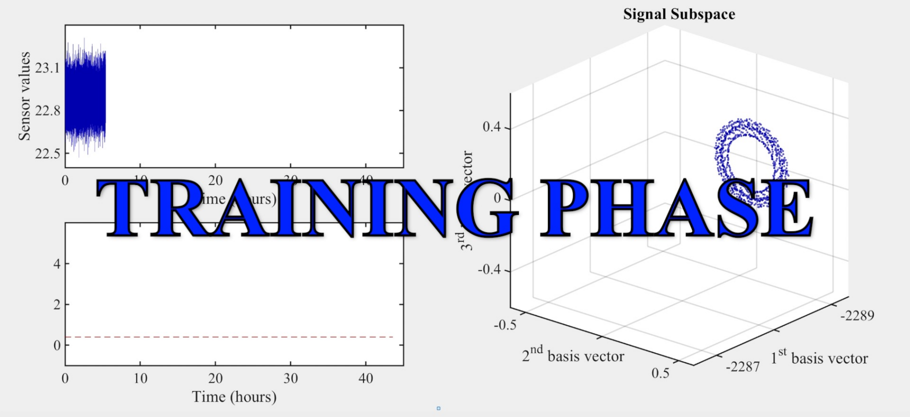
We show mathematically that

- the departure score can be computed more efficiently as $D_j = ||\tilde{\mathbf{c}} - \mathbf{U}^T\mathbf{x}_j||^2$ using *implicit* projection onto the signal subspace **(isometry trick)**, and
- that $\mathcal{L}^r$ is **isomorphic** to $\mathbb{R}^r$, which allows for visualizing the process behavior in the signal subspace.

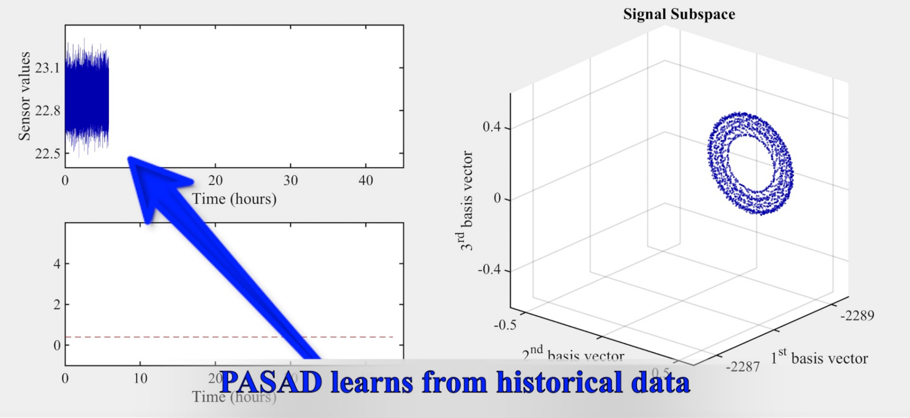
Validation I — Visualizing the Departure



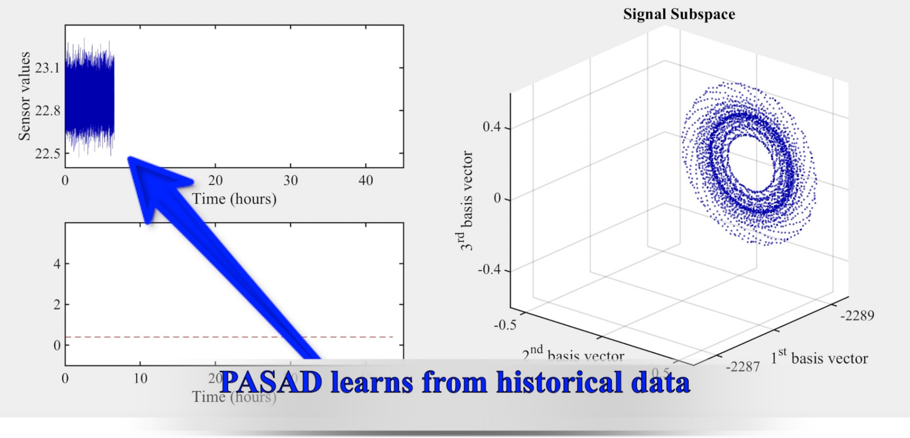
Validation I — Visualizing the Departure



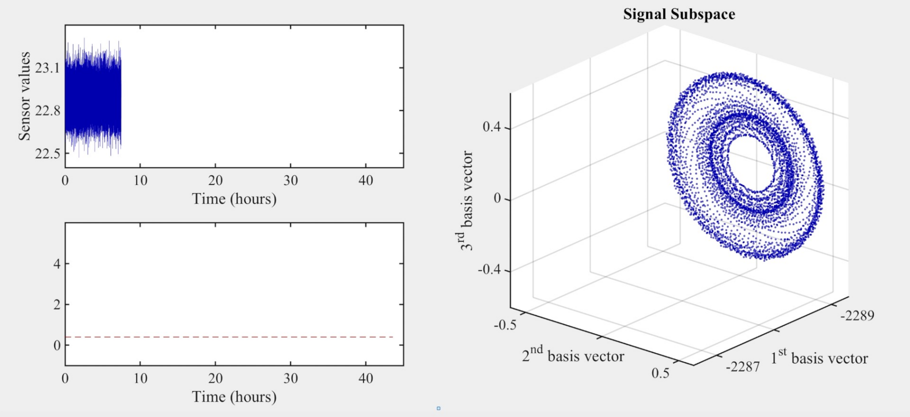
Validation I — Visualizing the Departure

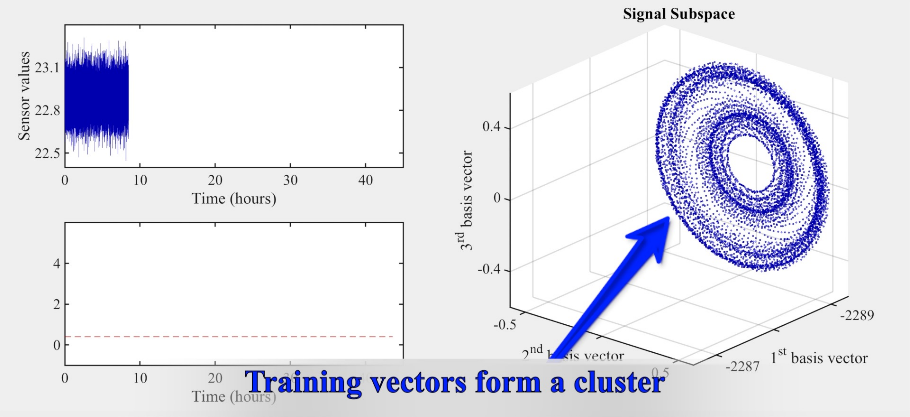


Validation I — Visualizing the Departure



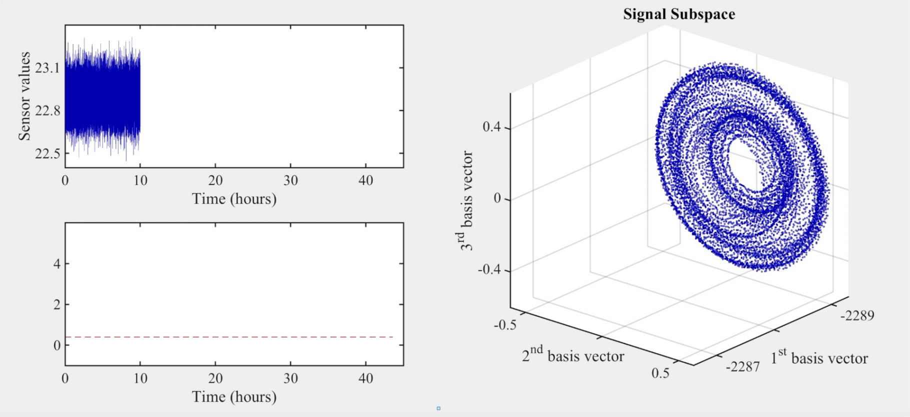
Validation I — Visualizing the Departure



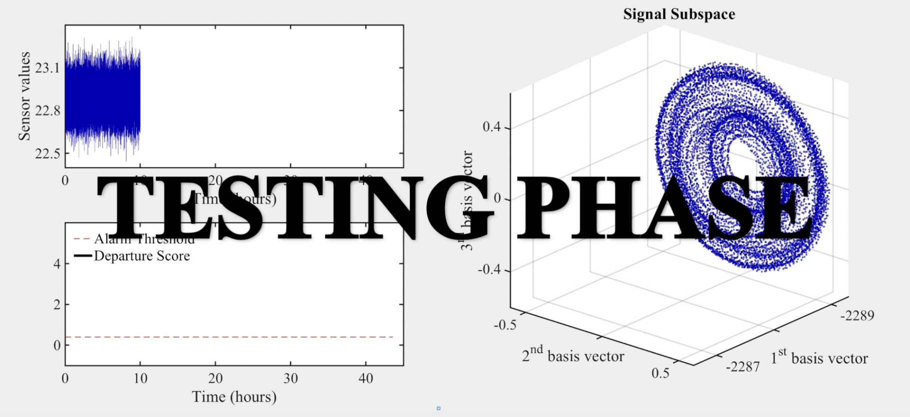


Sensor values
23.1
22.8
22.5
0
10
20
30
40
Time (hours)
4
2
0
Time (hours)
Signal Subspace
3rd basis vector
0.4
0
-0.4
-0.5
2nd basis vector
0.5
-2287
-2289
1st basis vector
Training vectors form a cluster

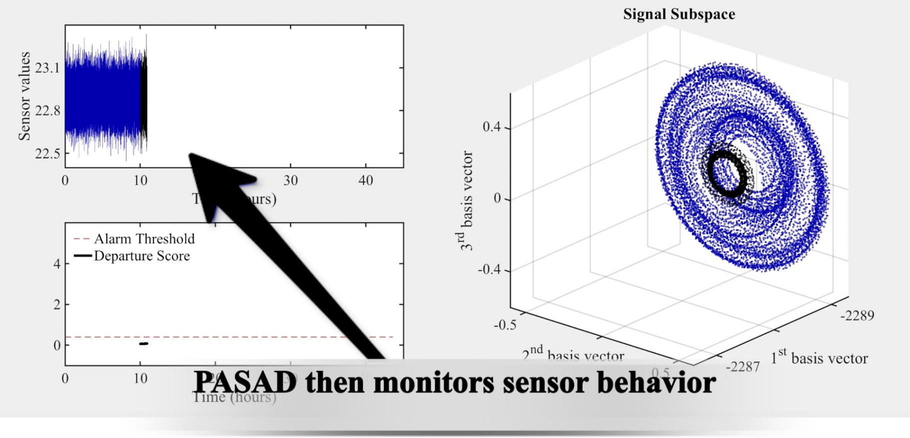
Validation I — Visualizing the Departure



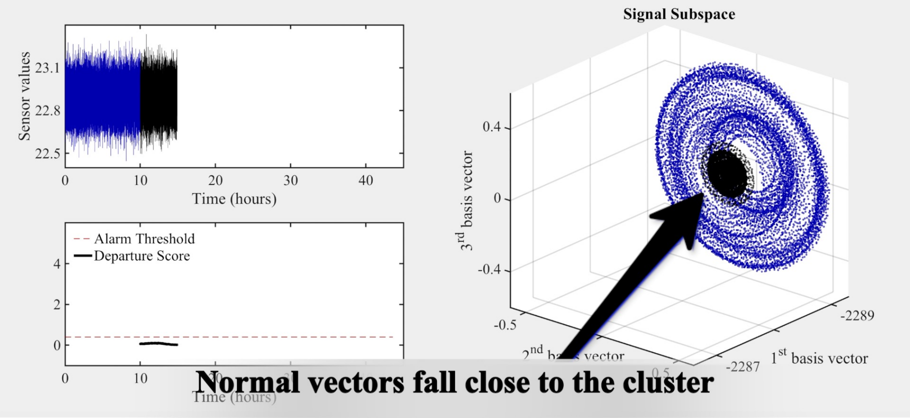
Validation I — Visualizing the Departure



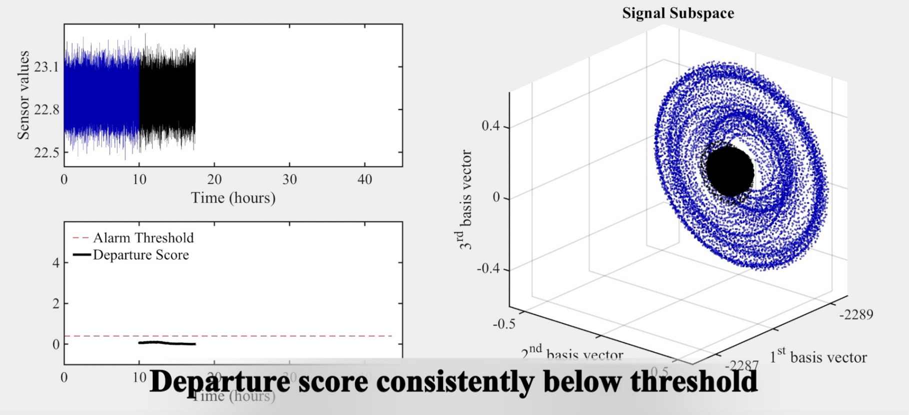
Validation I — Visualizing the Departure



Validation I — Visualizing the Departure

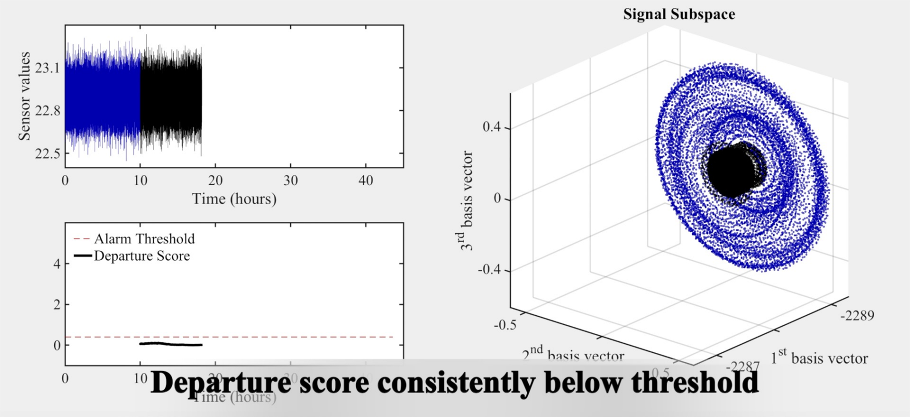


Validation I — Visualizing the Departure



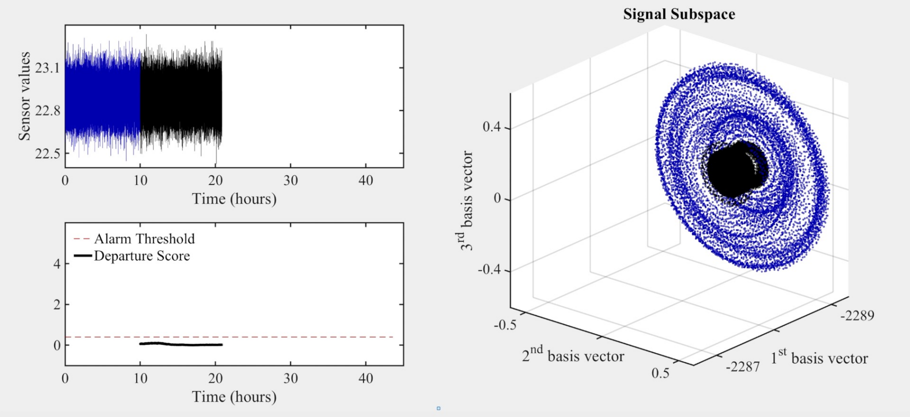
Departure score consistently below threshold

Validation I — Visualizing the Departure

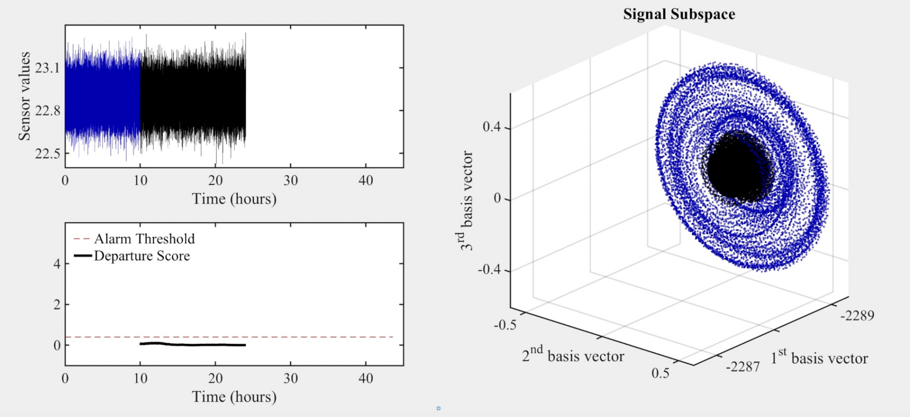


Departure score consistently below threshold

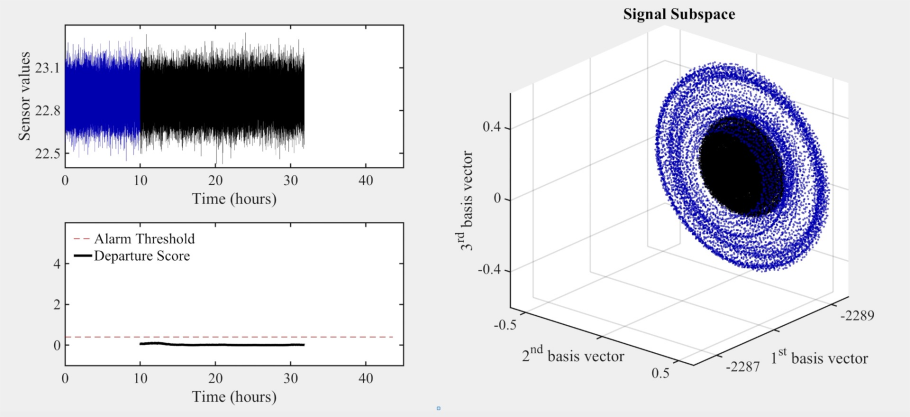
Validation I — Visualizing the Departure



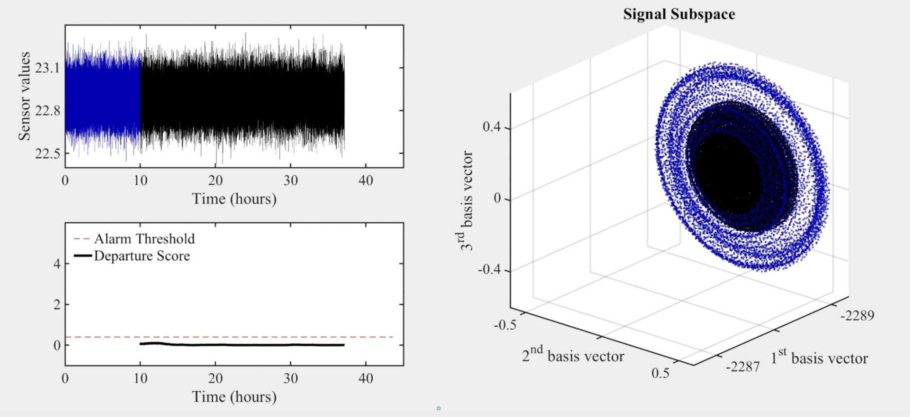
Validation I — Visualizing the Departure



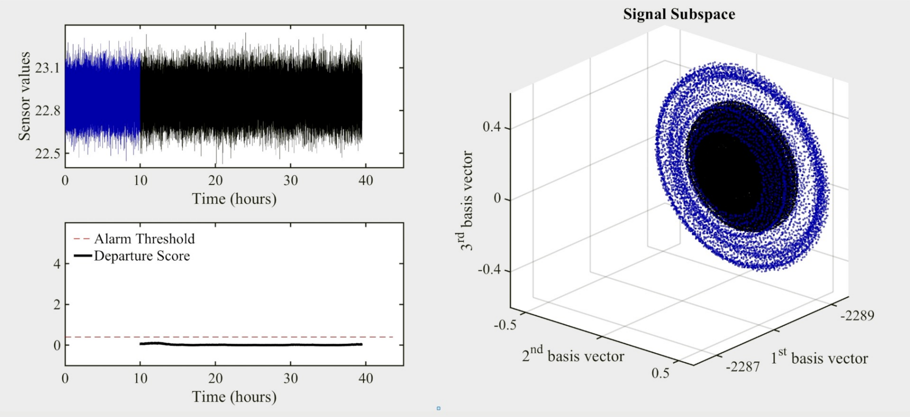
Validation I — Visualizing the Departure



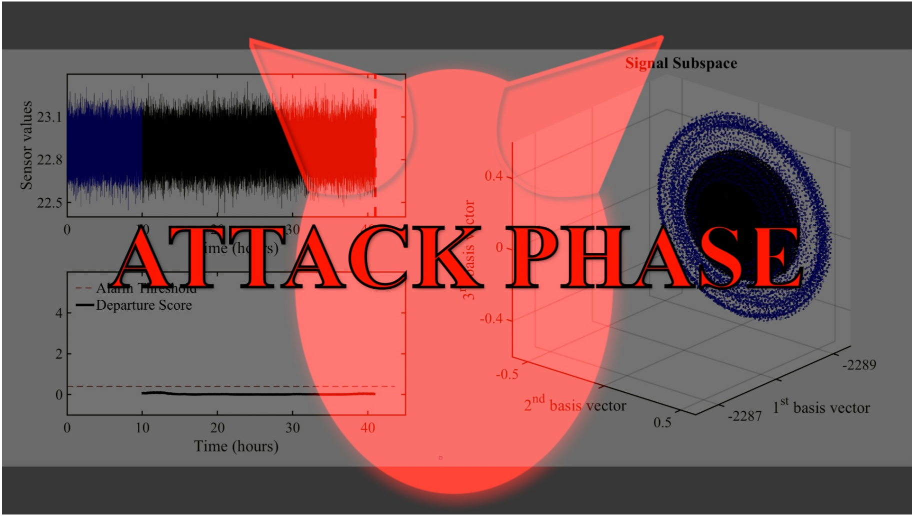
Validation I — Visualizing the Departure



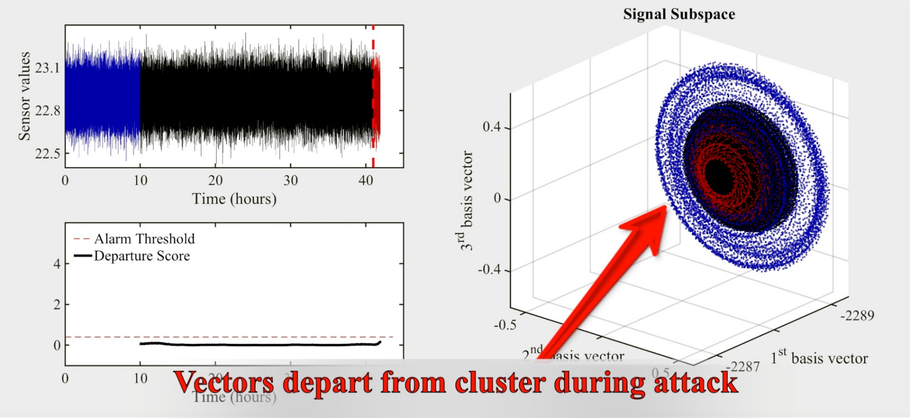
Validation I — Visualizing the Departure



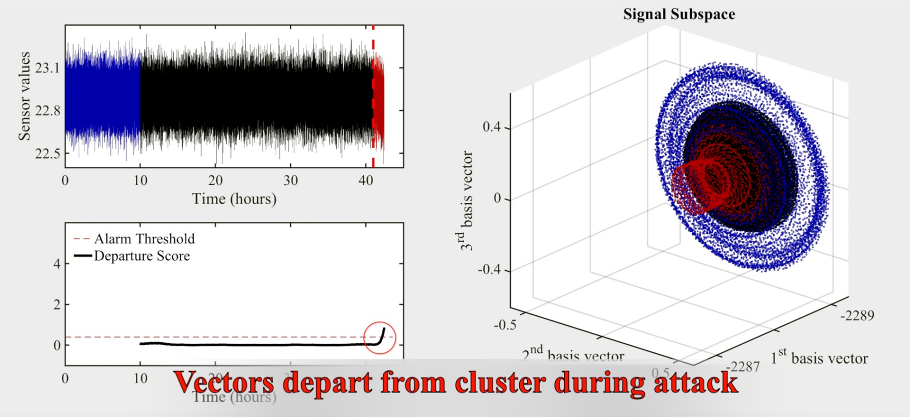
Validation I — Visualizing the Departure



Validation I — Visualizing the Departure

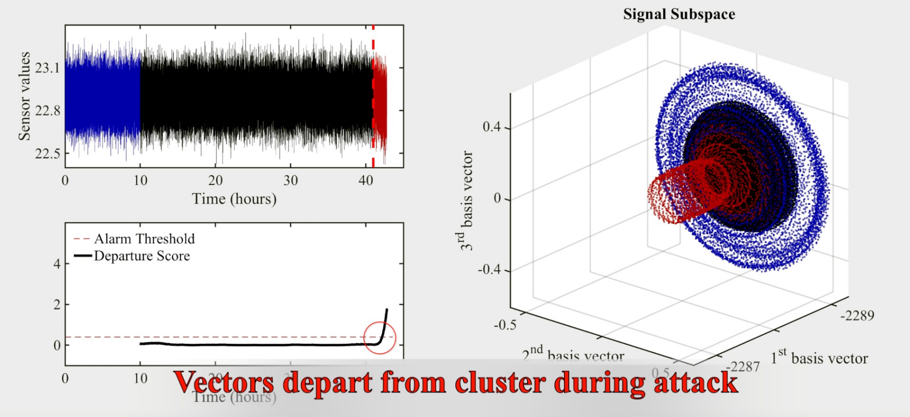


Validation I — Visualizing the Departure

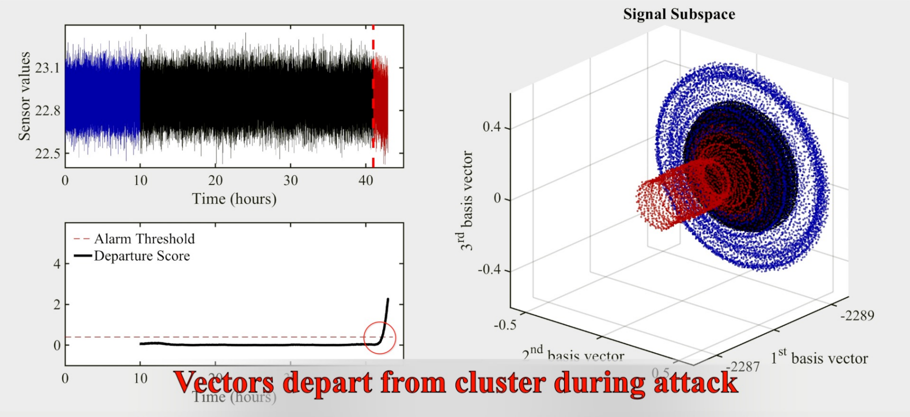


Vectors depart from cluster during attack

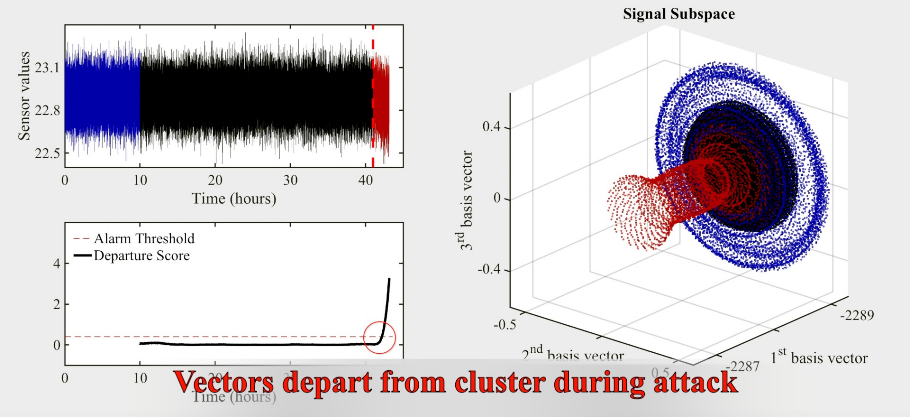
Validation I — Visualizing the Departure



Validation I — Visualizing the Departure

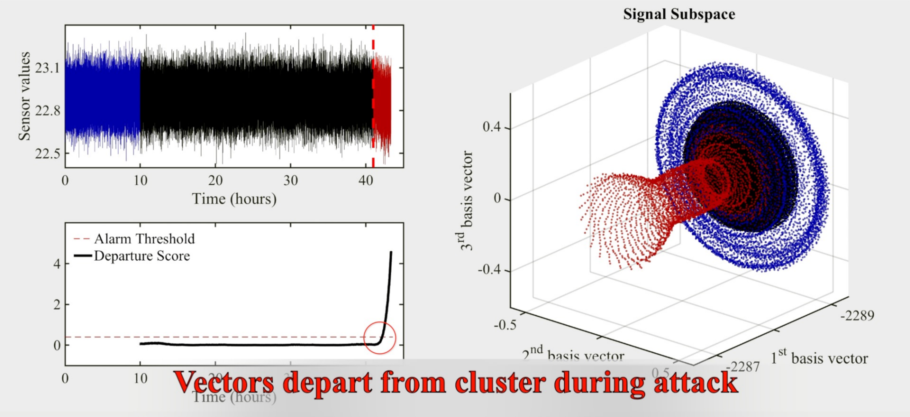


Validation I — Visualizing the Departure



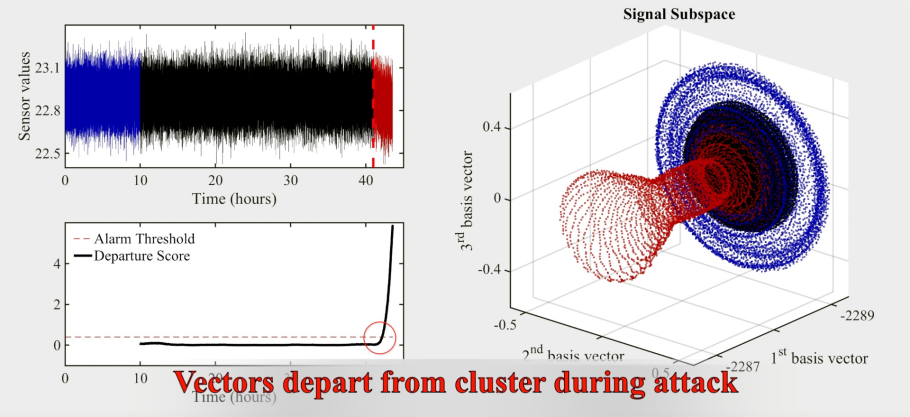
Vectors depart from cluster during attack

Validation I — Visualizing the Departure



Vectors depart from cluster during attack

Validation I — Visualizing the Departure

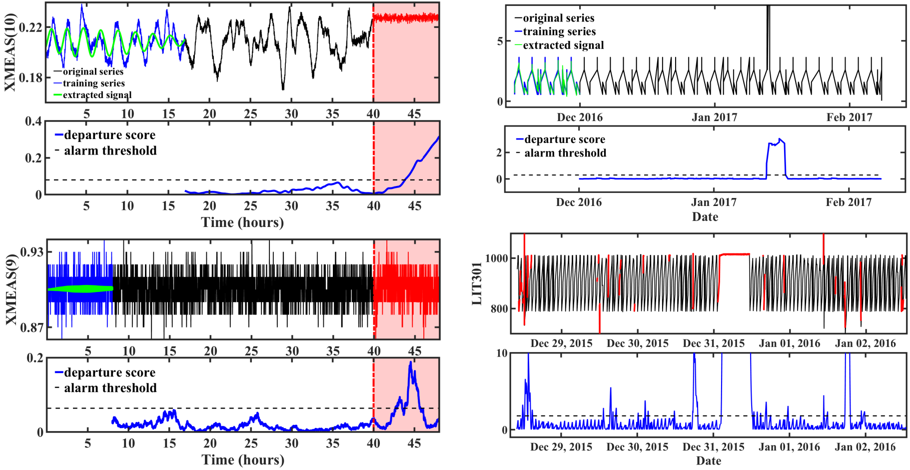


Validation II — Evaluation on Various Systems

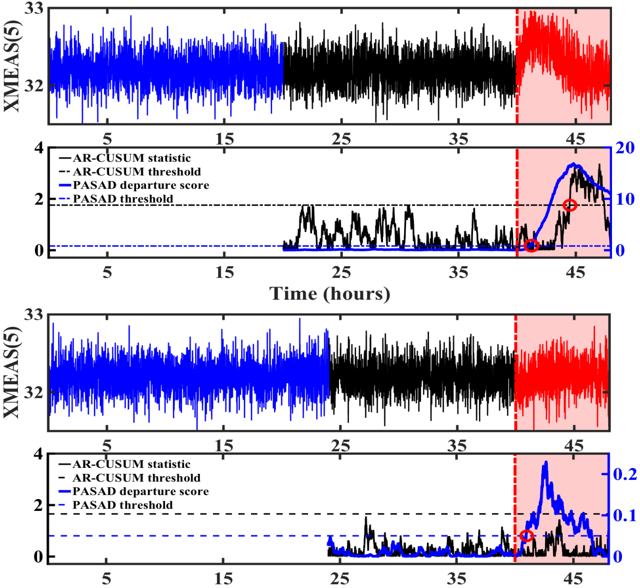
PASAD was tested on

- Tennessee-Eastman Process: a simulation model of a chemical plant.
- SWaT dataset: data from the SWaT water treatment testbed.
- Real data: from a water distribution plant in Gothenburg.

Validation II — Evaluation on Various Systems



Validation III — Comparison with Auto-Regression



Validation IV — Deploying a Prototype in a Real Environment

- A full-fledged PASAD prototype was deployed in a real control system (paper mill north of Gothenburg).
- System operation was monitored for 75 days.
- Stable performance: no technical issues encountered.

Summary

- Attacks on ICS are worryingly increasing.
- Process-level attack detection proves a viable approach in this domain.
- Existing methods solve more general problems.
- PASAD is a model-free detection method that
  - has sound theoretical basis,
  - is specification-agnostic,
  - efficient and lightweight, and
  - noise-tolerant.

Questions?

CHALMERS