Constrained Concealment Attacks against Reconstruction-based Anomaly Detectors in Industrial Control Systems

Alessandro Erba, Riccardo Taormina, Stefano Galelli, Marcello Pogliani, Michele Carminati, Stefano Zanero, Nils Ole Tippenhauer

alessandro.erba@cispa.saarland

ACSAC 2020
December 10th 2020

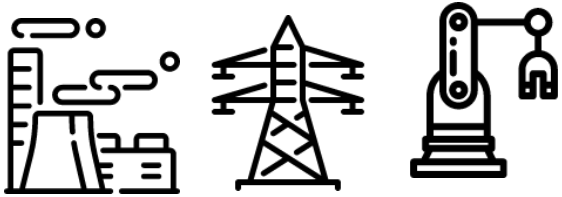
CISPA
HELMHOLTZ CENTER FOR INFORMATION SECURITY

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

POLITECNICO
MILANO 1863

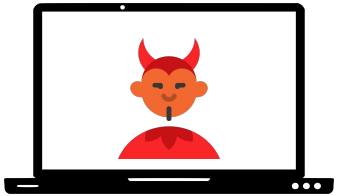
TU Delft

Introduction

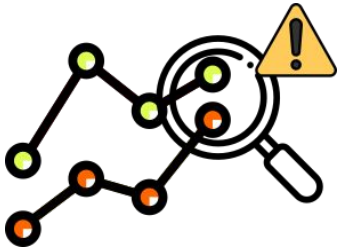


Critical infrastructures are nowadays interconnected

Security features are hard to implement



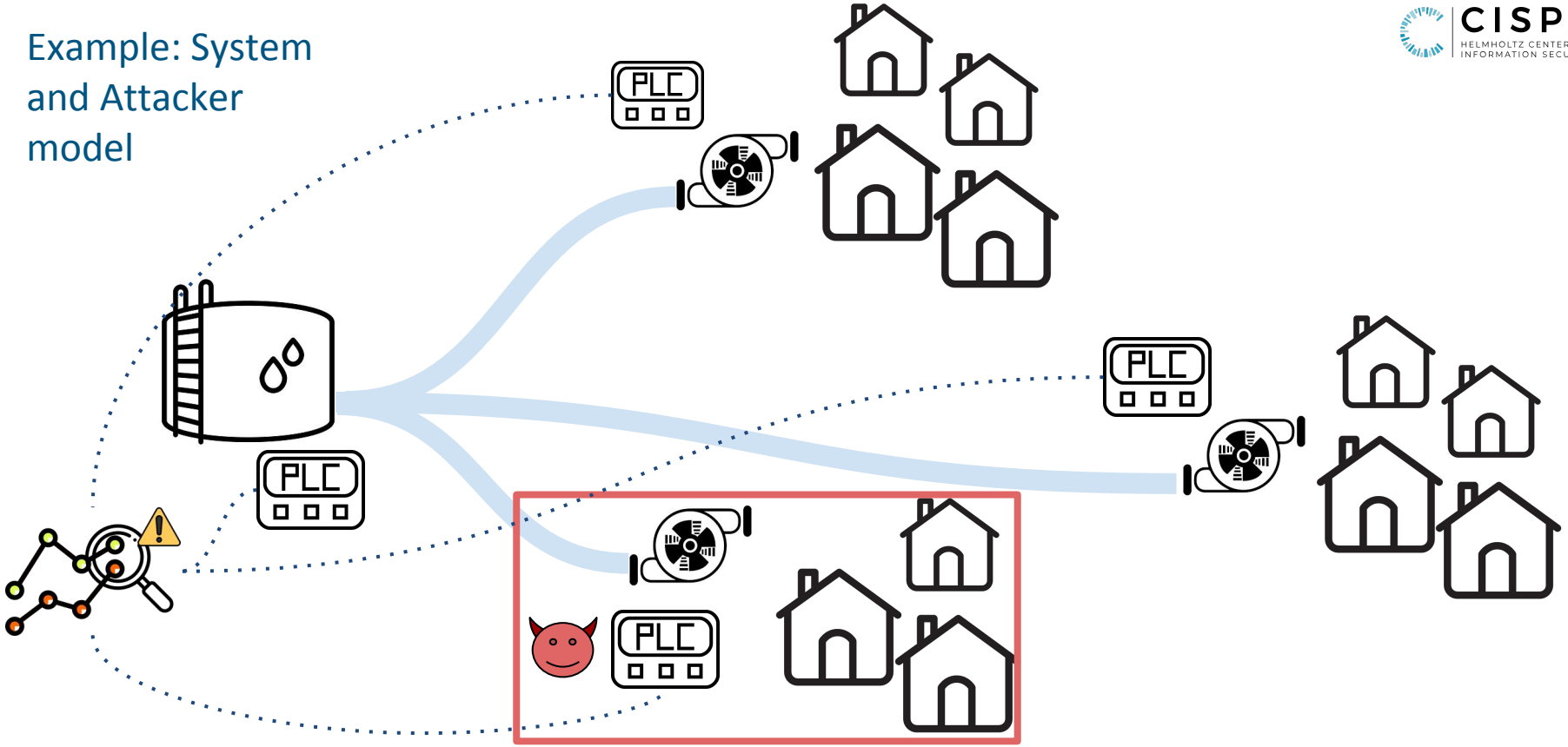
Eavesdrop
Malicious commands
Sensor spoofing



Anomaly detection based on sensor readings is promising to detect anomalies caused by attackers

Reconstruction-based detectors are promising to identify anomalies in the system

Example: System and Attacker model



In this work, we propose and evaluate constrained concealment attacks against reconstruction-based anomaly detectors

Research challenges

- **R1:** Due to the distributed nature of the system, the attacker is constrained to manipulate only a subset of features.
- **R2:** Adversarial examples must meet the temporal and spatial correlations expected from the observed physical processes.
- **R3:** Craft adversarial examples to optimize the Mean Squared Error loss instead of optimizing the cross-entropy loss.
- **R4:** Real-time constraints of ICS environment require novel techniques to compute adversarial examples

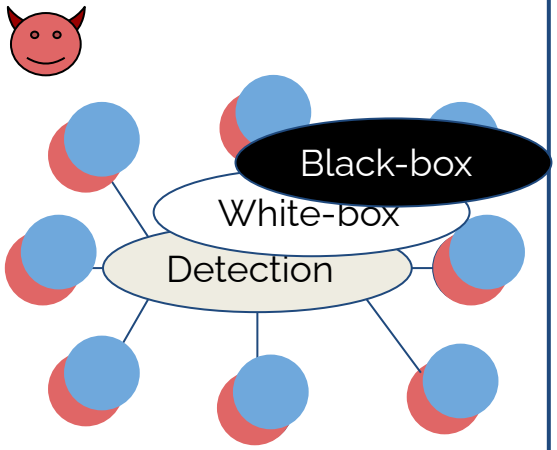
Attacker Model

Attacker can read the feature

Attacker can spoof the feature

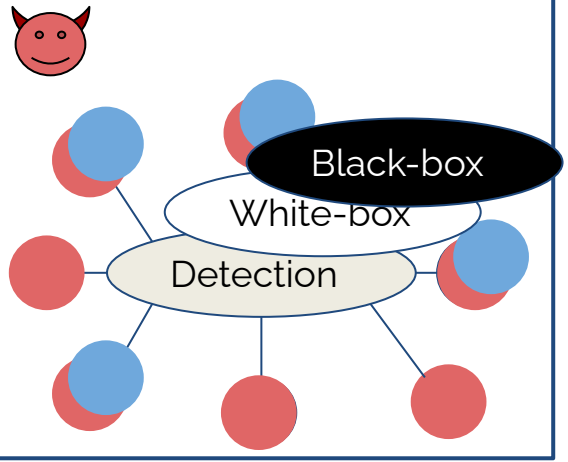
Unconstrained Attacker

The attacker can read and spoof all the sensor readings.



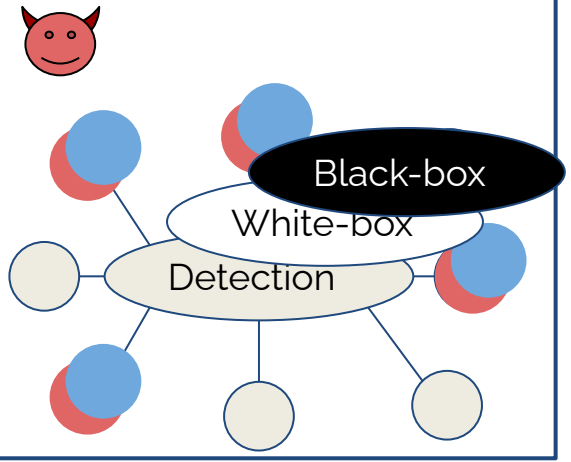
Partially Constrained Attacker

The attacker can read all the sensor readings and spoof only a subset of them.



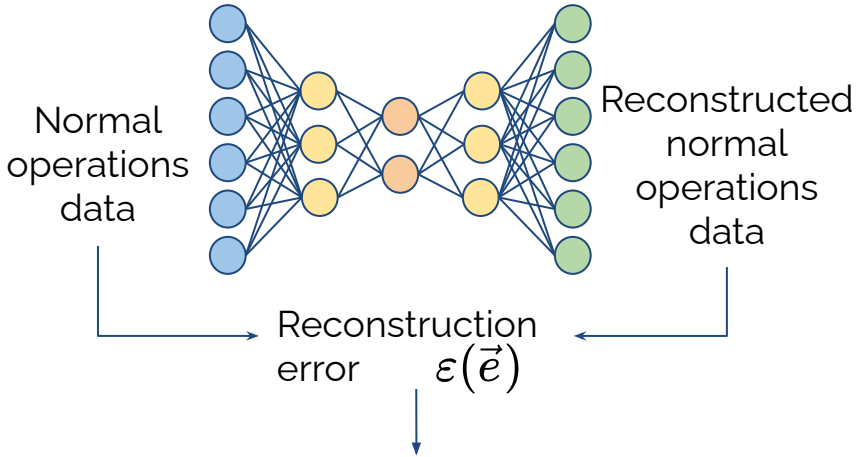
Fully Constrained Attacker

The attacker can read and spoof a subset of sensor readings.

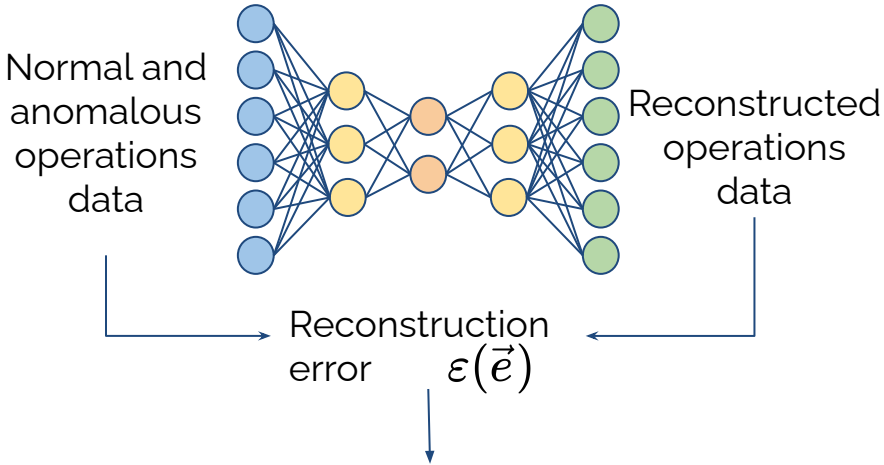


Reconstruction-based anomaly detectors

Training



Testing



$$y(\vec{x}) = \begin{cases} \text{`under attack' if } \varepsilon(\vec{e}) > \theta \\ \text{`safe' otherwise} \end{cases}$$

Attacker Goal Formalization

Given: $\vec{x}$ (Feature vector of sensor readings)

$y(\vec{x})$ (Classification function)

such that: $y(\vec{x}) =$ `under attack'

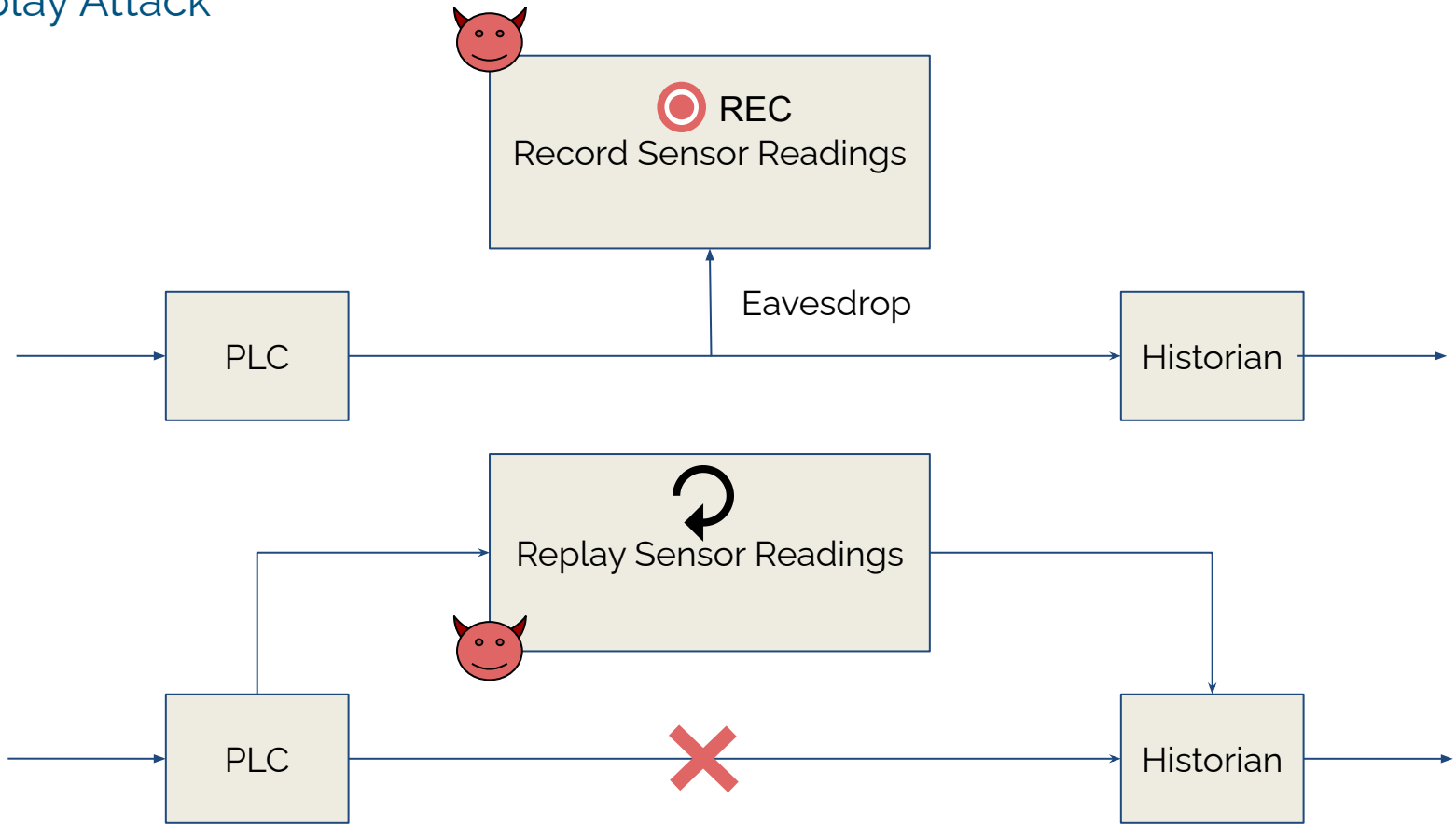
minimize $$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{x_i} - (x_i + \delta_i))^2$$

s.t. $\vec{\delta} \in$ constraint space

real-time constraints imposed by CPS

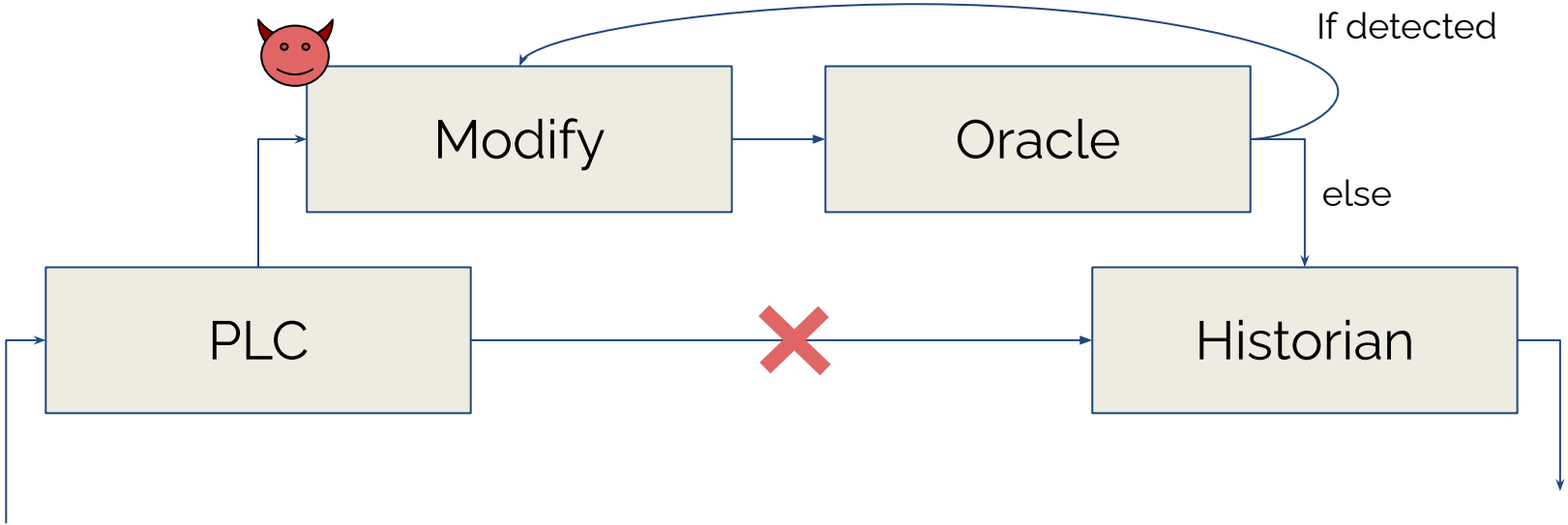
$y(\vec{x} + \vec{\delta}) =$ `safe'

Replay Attack

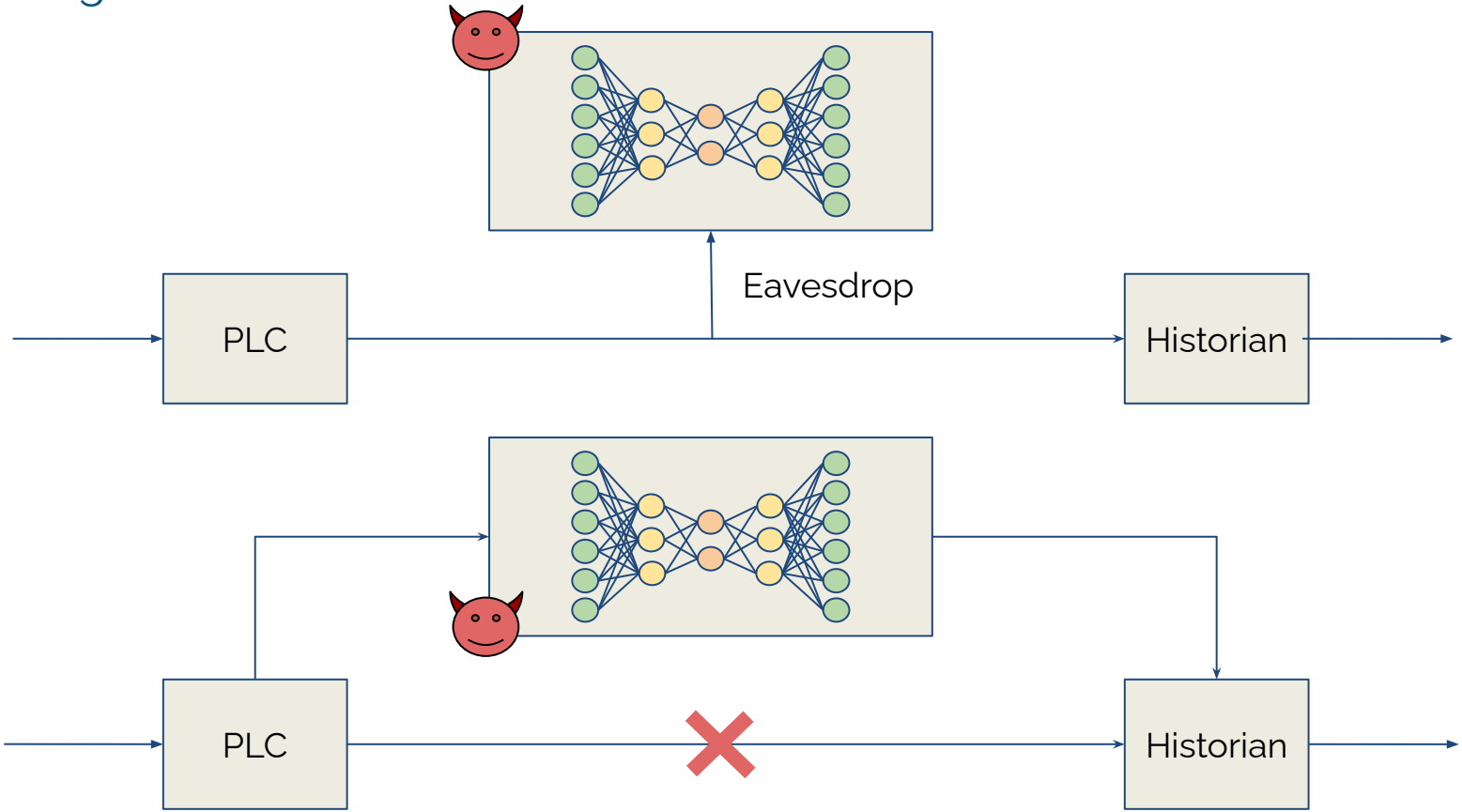


Iterative Attack

The White-box attacker can start an iterative attack to decrease the reconstruction error.



Learning Based attack



Evaluation

We consider two Datasets:

1. Batadal Dataset, water distribution network simulated with epanetCPA. 43 features
2. WADI Dataset, a real-world water distribution testbed. 82 features

Each dataset is divided in two data chunks:

A. Training data → no anomalies
B. Test data → with anomalies

We train 3 Reconstruction based anomaly detectors proposed in prior work:

1. Fully connected autoencoder
2. LSTM autoencoder
3. CNN autoencoder

Attacks are evaluated in terms of reduction of Recall score of the anomaly detector:

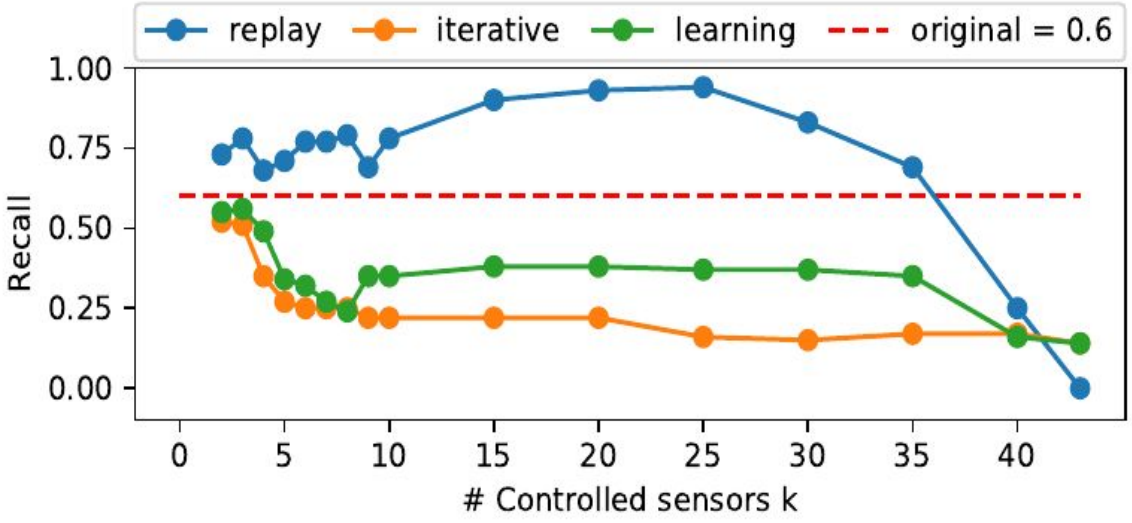
$$Recall = \frac{TP}{TP+FN}$$

Unconstrained Concealment Attack

	Detection Recall			
Dataset	**Original Detection**	**Replay**	**Iterative**	**Learning Based**
Batadal	0.6	0	0.14	0.14
WADI	0.68	0.07	0.07	0.31

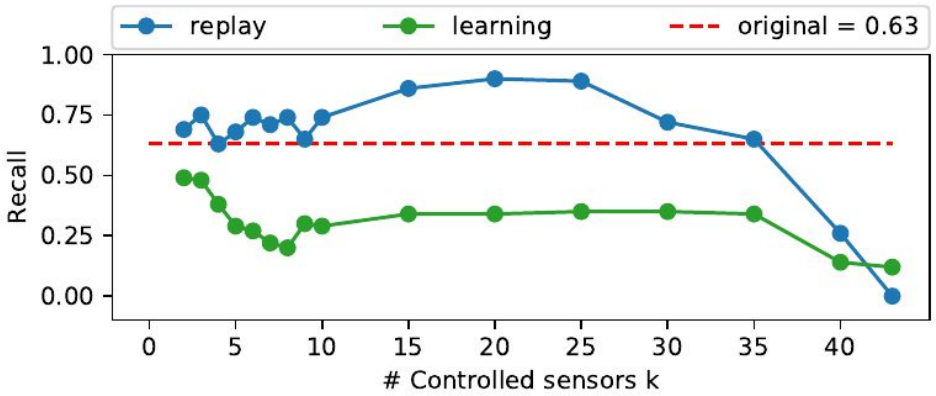
Computational time in seconds					
	Replay	**Iterative**		**Learning Based**	
Data		mean	std	mean	std
Batadal	-	2.28	2.46	0.002	0.005
WADI	-	0.6	0.41	0.005	0.002

Partially Constrained Concealment attacks

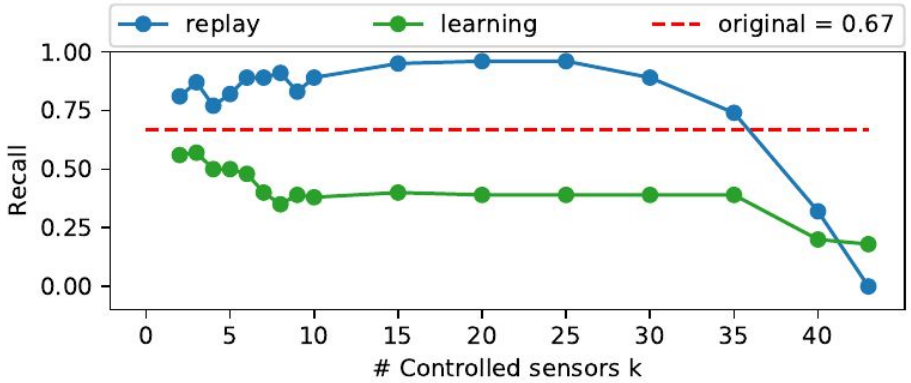


BATADAL

Transferability



LSTM



CNN

CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

Conclusions

- We are the first to formalize a constrained attacker model for real-world ICS.
- We presented the first real-time concealment attacks on reconstruction-based anomaly detectors for ICS
- We show that the attacks present four unique challenges, and propose two attacks to tackle them
- We demonstrated that our attacks are feasible in general, and outperform replay attacks when the attacker is constrained to control of less the 95% of the features
- We demonstrate our attacks in Real-time in a water distribution testbed

Please refer to our paper for the complete results of our evaluation
The code and data are available at: https://github.com/scy-phy/ICS-Evasion-Attacks

Thank you!

alessandro.erba@cispa.saarland