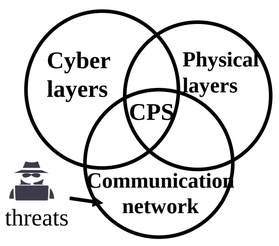
Resilient Cyber-physical System against False Data Injection Attacks

Concurrent Learning and Resilient Estimation

Yu Zheng

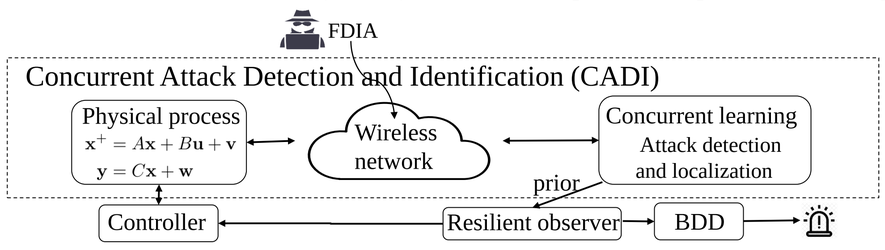
Introduction and Motivation



Example cyber threats in the history:
1) Suspected cyber intruder took control of the power system control center in western Ukraine in December 2015;
2) In 2015, 25 cyber attacks were disclosed in several water systems [1];
3) In 2020, a malicious cyber-attack attempted to raise the chlorine level in Israel's water supply to dangerous proportion ……

False data injection attacks (FDIAs) have been shown to be capable of triggering erroneous state estimation without detected by bad data detector (BDD), thereby resulting in malicious operations in physical processes.
Traditional attack-resilient observer designs, such as L1 decoder, event-trigger luenberger observer…, require **half of measurements to be clean**.

By taking the advantage of CPS, we propose a concurrent learning structure, named as **concurrent attack detection and identification (CADI)**, which could provide improved resiliency on observer design.



However, the following **challenges** should be solved:
1) How to design attack detection and localization algorithm?
2) How to improve the accuracy of attack detection and localization algorithm?
3) What is the symmetrical way to use the uncertain prior information in resilient observer design?

Attack Generation

Definition (successful FDIA):

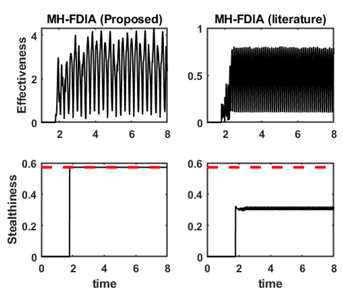
An attack vector $\boldsymbol{e}$ is said to be (α, ϵ)-successful against a decoder-detector pair $\left(D(\boldsymbol{y}), D_\epsilon(\boldsymbol{y})\right)$ if

$$||D(\boldsymbol{y}) - D(\boldsymbol{y}+\boldsymbol{e})|| \geq \alpha, \qquad ||\boldsymbol{y}+\boldsymbol{e} - HD(\boldsymbol{y}+\boldsymbol{e})|| \leq \epsilon$$

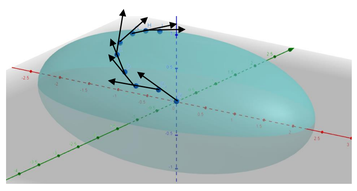
Effectiveness **Stealthiness**

Model-driven approach:

Moving-horizon FDIA design against L2 observer
A **gradient-ascent searching algorithm** is proposed
$h = ||D(\boldsymbol{y}) - D(\boldsymbol{y}+\boldsymbol{e})||_2$ (objective function)
$S = \{\boldsymbol{e}: ||\boldsymbol{y}+\boldsymbol{e} - HD(\boldsymbol{y}+\boldsymbol{e})||_2 \leq \epsilon\}$ (ellipsoid constrain)



The algorithm searches on the gradient ascent direction of h inside the ellipsoid S. When approach the boundary, it will stay on the boundary.

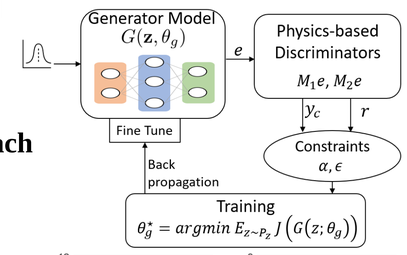


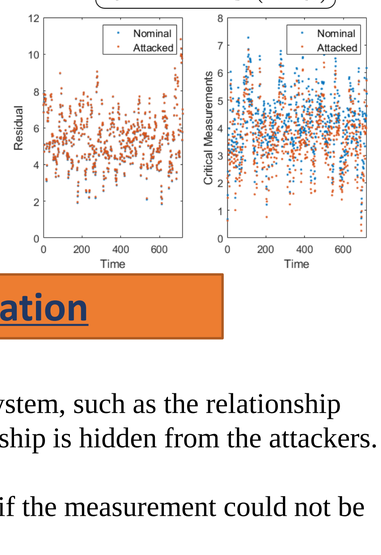
Data-driven approach:

A physics-based Generative adversary network (GAN) approach
For a linear system, M_1, M_2 are transfer matrices from the attack signal to the detection residual (stealthiness) and critical measurements (effectiveness).
Indicator-like loss function:

$$J(e) = ReLU\left(\alpha - \left||M_1\boldsymbol{e}|\right|_2\right) + ReLU\left(\left||M_2\boldsymbol{e}|\right|_2 - \epsilon\right)$$

Advantage: This attack generator framework does not require training attack examples, and M_1, M_2 are not required to be exact relationship.





Attack Detection and Localization

Gaussian process regression (GPR) approach [2]

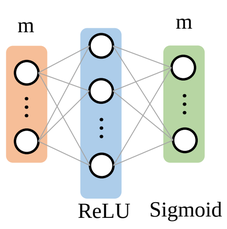
This approach is to train an GP model to learn a hidden feature of system, such as the relationship between a hidden state and each measurement. This kind of relationship is hidden from the attackers.

$$GP_i(x) = y_i, i = 1,2,\dots,m$$

Then the trained GPs are detection model of attacked measurement if the measurement could not be explained by the trained GP model with high likelihood.

step1: capture posterior distribution $GP_i(x) \rightarrow \{\mu_i, \Sigma_i\}, i = 1,2,\dots,m$
step2: calculate Z-score $z_i = \frac{y_i - \mu_i}{\Sigma_i}$
step3: calculate probability $p_i = 1 - p_X(|x| \leq |z_i|) = \int_{|x_i|} \frac{e^{x^2/2}}{\sqrt{2\pi}}$
step4: output localization result $q_i(support\ prior) = 0(attacked)\ if\ p_i \leq 0.5$

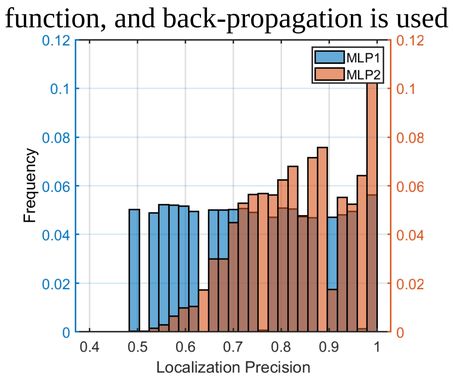
Multi-layer perceptron (MLP) approach



This approach is to train a narrow neural network-based classifier to identify attacks on each measurement node.
The dimension of input and output layers are the same as the dimension of measurements.
Loss function is cross-entropy loss function, and back-propagation is used to train the network.

The following result shows how the localization accuracy of MLP-detector is improved by adding the above GAN-based attack generator in the training process:

MLP1: trained by random attack generator
MLP2: trained by random attack generator and GAN-based generator



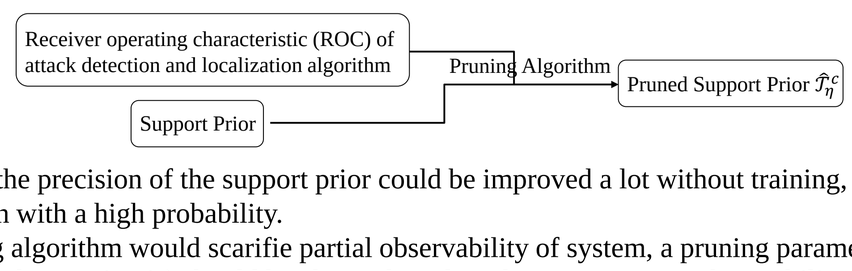
Resilient Estimation

Pruning Algorithm + Weighted L1 Estimator

Support prior: the estimate of real attack location produced by attack detection and localization algorithm.
The challenge to use support prior in dynamical estimation is: (1) the inherent uncertainty of the probabilistic prior information, (2) the long training time.
The uncertainty of prior could be modeled by **Bernoulli distribution.**

Pruning Algorithm [3]

A statistics-based pruning operation of binary classification results.



After pruning, the precision of the support prior could be improved a lot without training, even achieve 100% precision with a high probability.
Notice, pruning algorithm would scarifie partial observability of system, a pruning parameter (control how aggressive the pruning is) should be chosen based on the measurement observability.

Weighted L1 observer design with prior pruning [3]

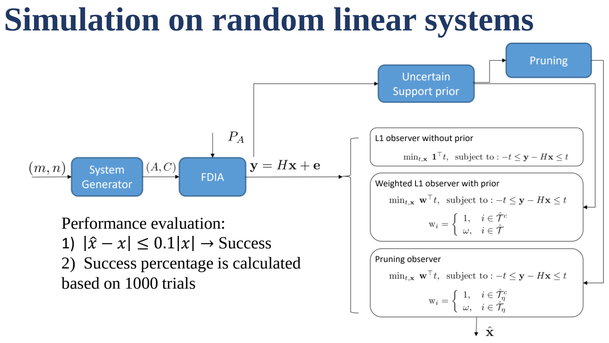
A moving-horizon observer scheme is given below, where the weights are given based on the pruned support prior.
If there is attack based on the pruned support prior, we should set a small weight to that measurement, and how small that value is depends on how much we trust the performance of the support prior.

$$\text{Minimize} \sum_{p=i-T+1}^{i} \|\mathbf{y}_p - C\mathbf{z}_p\|_{1,w}$$

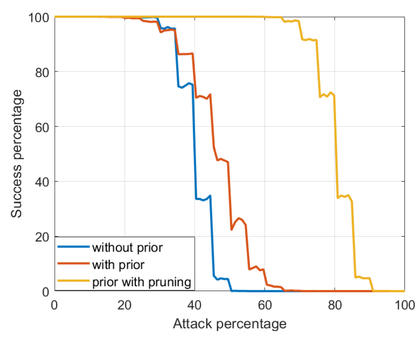
$$\text{Subject to } \mathbf{z}_{p+1} - A\mathbf{z}_p = 0, \quad p = i-T+1, \cdots, i-1$$

$$w = \begin{cases} 1, & j \in \hat{\mathcal{T}}_\eta^c \\ \omega, & j \in \hat{\mathcal{T}}_\eta \end{cases}$$

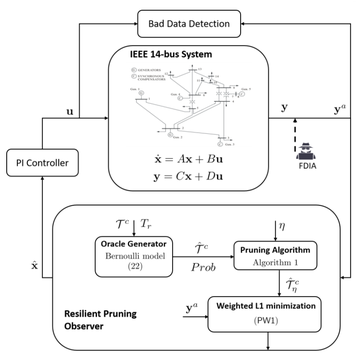
Simulation on random linear systems

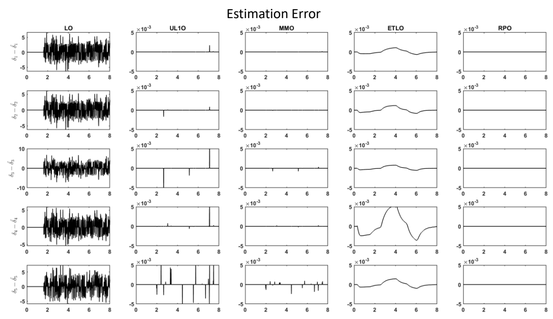


Performance evaluation:
1) $|\hat{x} - x| \leq 0.1|x| \rightarrow$ Success
2) Success percentage is calculated based on 1000 trials



Simulation on IEEE 14-bus system





LO: Luenburger Observer ETLO: Event triggered Luenbueger Observer
UL1O: Unconstrained L1 Observer RRO: proposed Pruning Observer
MMO: Multi-model Observer

Conclusion and Future Work

This poster presents a complete framework of resilient cyber-physical system design including attack detection and localization algorithm with automated attack generator, pruning algorithm and resilient estimator.

1) The proposed attack generators provide abundant adversary training datatset. Compared to literature, the proposed methods are more realistic and implementable, and are potential to generate more different types of attacks.

2) Two different ideas of attack detection and localization are given, GPR approach is to learn the features of system, while the MLP approach is to learn the features of attacks.

3) Pruning method is a way to improve the precision of results of localization algorithm without training. The proposed symmetrical design provides quantified guarantee for resilient estimation error.

4) The weight L1 observer with prior pruning is an improved resilient observer with capability of correctly estimating state in worse adversary environment.

Reference

[1]. R. M. Clark, S. Panguluri, T. D. Nelson, and R. P. Wyman, "Protectingdrinking water utilities from cyberthreats,"Journal of the AmericanWater Works Association, vol. 109, no. INL/JOU-16-39302, 2017

[2]. Y. Zheng and O. Anubi, "Resilient Observer Design for Cyber-Physical Systems with Data-Driven Measurement Pruning", Security and Resilience in Cyber-Physical Systems, edited by M. Abbaszadeh and A.Zemouche, Springer, 2021 [to appear]

[3]. Y. Zheng, OM Anubi, "Attack-Resilient Weighted L1 Observer with Prior Pruning", 2021 American Control Conference.

Contacts

Yu Zheng, Ph.D. Candidate
IEEE Student Member
Florida State University
Email: yz19b@fsu.edu
Github: https://github.com/ZYblend

Advisor:
Olugbenga Moses Anubi
IEEE Senior Member
Florida State University
Email: oanubi@fsu.edu
Website: https://web1.eng.famu.fsu.edu/~anubi/