Robust Resilient Signal Reconstruction under Adversarial Attacks

Yu Zheng
Olugbenga (Moses) Anubi*
Lalit Mestha
Hema Achanta

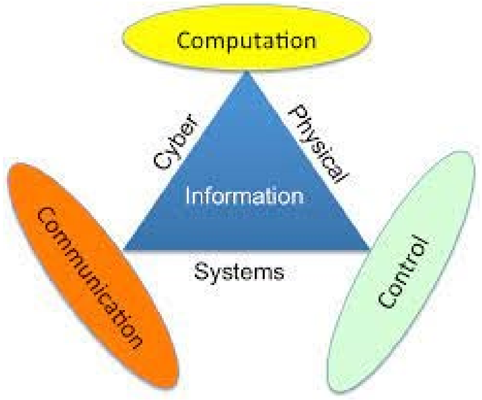
RAS lab
Physics-Data-Driven

FAMU-FSU Engineering

FLORIDA STATE UNIVERSITY
CENTER FOR ADVANCED POWER SYSTEMS

Motivation and Preliminary

Cyber-Physical Systems



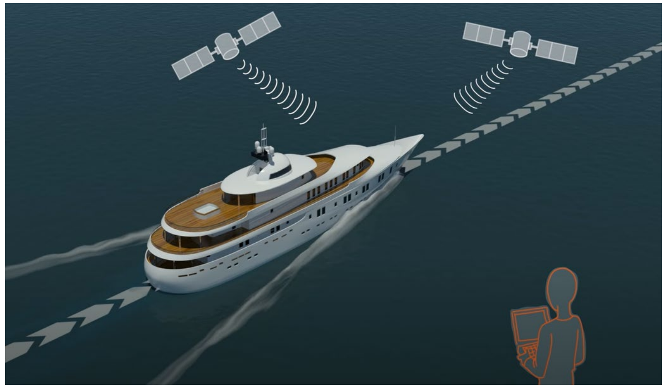
Credit: https://dev.to/ruthvikraja_mv/cyber-physical-system-security-vulnerabilities-4bak

Cyber Threats

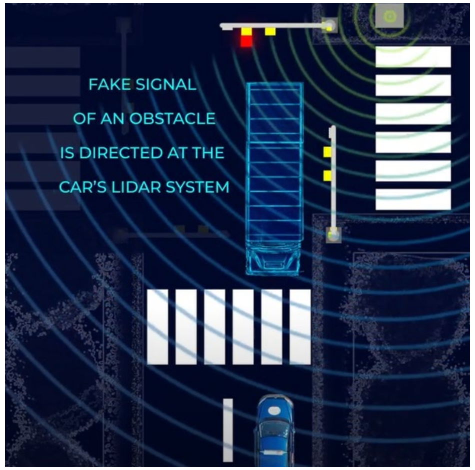
- 2000: SCADA system of **Maroochy Shire Sewage** in Australia was compromised
- 2003: A cyberattack penetrated a computer network at the **Davis-Besse nuclear power plant**
- 2010: **Stuxnet worm** attacked Iran's Natanz nuclear fuel-enrichment facility
- 2015: Blackenergy **Power generation and distribution** (Ukaine)
- 2016: **IronGate**: Designed to target Industrial Control systems (ICS)
- 2020: A malicious cyber-attack attempted to raise the chlorine level in **Israel's water supply** to a dangerous proportion
- 2021: Cyberattacks on the **Colonial Pipeline**
- 2022: Ghost Attacks on **Tesla**
- 2023: **Lidar/Radar** Spoofing attack

Sensor Attacks

False Data Injection Attack
Sensor Spoofing Attack
Poisoning Attack
Deceptive Attack



GPS Spoofing

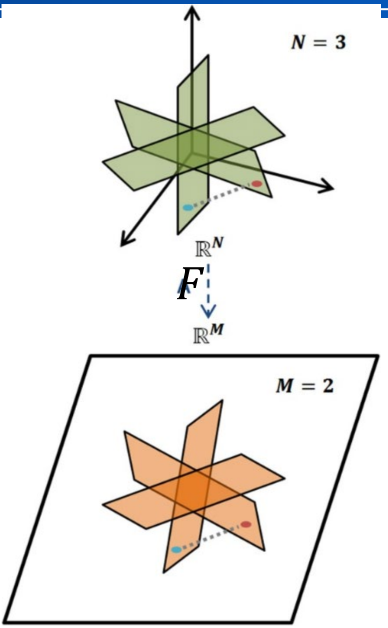


Lidar/Radar Spoofing

Camera Spoofing

Motivation and Preliminary

Figure Credits: Petrantonakis, Panagiotis C., and Panayiota Poirazi. "A compressed sensing perspective of hippocampal function." Frontiers in systems neuroscience 8 (2014): 141.

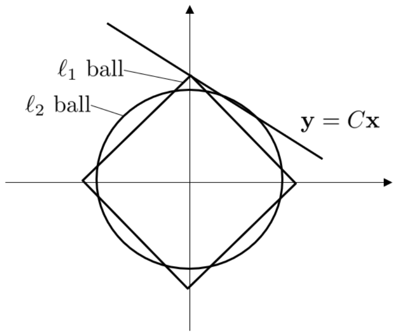


Modeling Adversarial Attacks

$$y = Cx + e + v \qquad y, e, v \in \mathbb{R}^m, x \in \mathbb{R}^n$$

Assumptions:

1. Redundancy: $m > n$
2. Bounded Noise: $||v|| \leq \epsilon$
3. Sparse Corruption: $supp(e) \leq k < m \quad [supp(e) = \{i | e_i \neq 0\}]$
4. Attack-Noise Orthogonality: $e^{\top} v = 0$



Resiliency Properties

$$Given\ a\ coding\ matrix\ F\ (FC = 0): \qquad y' = Fy = Fe$$

$$\boxed{\boldsymbol{Minimize\ \|e\|_0 \quad subject\ to: y' = Fe}}$$

1. Uniqueness condition:

Any k-sparse e can be recovered if

$$\mathcal{N}(F) \cap \Sigma_k = \emptyset \quad [\Sigma_k = \{x \in \mathbb{R}^m, |supp(x)| < k]$$

(2k-observability)

2. NP-hard problem

$$\boxed{\boldsymbol{Minimize\ \|e\|_1 \quad subject\ to: y' = Fe}}$$

1. Equivalence condition:

Restricted Isometry Property (RIP)

$$(1 - \delta_k)\|h\|^2 \leq \|F(\mathcal{T})h\|^2 \leq (1 + \delta_k)\|h\|^2$$

$$for\ any\ h \in \mathbb{R}^{|\mathcal{T}|}\ and\ all\ \mathcal{T}\ with\ |\mathcal{T}| \leq k$$

2. Uniqueness condition:

$$\delta_k + \delta_{2k} + \delta_{3k} < 1$$

RAS lab Physics-Data-Driven

FAMU-FSU Engineering

Reconstruct with Exact Support Prior

$$\begin{aligned} Minimize: &\quad \|e\|_0 + \|v\|_2 \\ Subject\ to: &\quad y = Cx + e + v \\ &\quad e^\top v = 0 \end{aligned}$$

$$\mathcal{T} = supp(e) \Rightarrow$$

$$Minimize: \quad \|y_{\mathcal{T}} - C_{\mathcal{T}}x\|_2$$

Theorem 1 (Least Square Reconstruction). *Given the linear model*

$$\mathbf{y} = C\mathbf{x} + \boldsymbol{\nu}, \tag{9}$$

where $\mathbf{y} \in \mathbb{R}^m$ is a vector of measurements, $\mathbf{x} \in \mathbb{R}^n, n \leq m$ is a vector of internal states (or features), $C \in \mathbb{R}^{m \times n}$, and $\boldsymbol{\nu}$ is the model error with the associated error bound $\|\boldsymbol{\nu}\| \leq \varepsilon$ for a known constant $\varepsilon > 0$.

Consider any partial measurement $\mathbf{y}_1 \in \mathbb{R}^{m_1}, m_1 > n$ satisfying

$$\mathbf{y}_1 = C_1\mathbf{x}^* + \boldsymbol{\nu}_1, \tag{10}$$

where $C_1 \in \mathbb{R}^{m_1 \times n}$ is a matrix of the corresponding rows of C and $\boldsymbol{\nu}_1$ is the associated model error, the vector $\mathbf{x}^* \in \mathbb{R}^n$ is the unknown actual internal state associated with the complete measurement set as in (9).

The least-square estimator

$$\hat{\mathbf{x}} = \arg\min \left\{ \frac{1}{2} \|\mathbf{y}_1 - C_1\mathbf{x}\|^2 \right\}, \tag{11}$$

of $\mathbf{x}^*$, satisfies the error bound

$$\|\hat{\mathbf{x}} - \mathbf{x}^*\| \leq \frac{2}{\sigma_1}\varepsilon, \tag{12}$$

Remark 1 (Rank-deficiency and RIP): Necessarily $|\mathcal{T}^c| \geq n$, otherwise the reconstruction error $\|\hat{\mathbf{x}} - \mathbf{x}^*\|$ is unbounded. Consequently, one can conclude that: $\|\hat{\mathbf{x}} - \mathbf{x}^*\| \leq \frac{2}{\delta_n}\varepsilon$, where δ_n is the n-restricted isometry constant of $C^\top$.

Corollary 1 (Constrained Least Square Reconstruction). *Let $\mathcal{X} \subset \mathbb{R}^n$ be a set characterized by $\|\mathbf{x}_1 - \mathbf{x}_2\| \leq \delta$ for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$ and some $\delta > 0$. Consider the constrained least-square estimator:*

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x} \in \mathcal{X}} \left\{ \frac{1}{2} \|\mathbf{y}_1 - C_1\mathbf{x}\|^2 \right\}. \tag{17}$$

If $\mathbf{x}^* \in \mathcal{X}$, then the reconstruction error can be bounded as:

$$\|\hat{\mathbf{x}} - \mathbf{x}^*\| \leq 2\min\left\{ \frac{\delta}{2}, \frac{\varepsilon}{\delta_n} \right\}. \tag{18}$$

Reconstruction with Inexact Support Prior

Unknown Attack Support

$$\mathcal{T} = supp(e)$$

$$\mathbf{q}_i = \begin{cases} 1 & \text{if } i \in \mathcal{T}^c \\ 0 & \text{otherwise} \end{cases}$$

Estimated Attack Support

$$\hat{\mathcal{T}}$$

Agreement Model:

$$\mathbf{q}_i = \epsilon_i \hat{\mathbf{q}}_i + (1 - \epsilon_i)(1 - \hat{\mathbf{q}}_i)$$

$\epsilon_i \sim \mathcal{B}(1, \mathbf{p}_i)$, with known $\mathbf{p}_i \in R_+$
given by $\mathbf{p}_i = E[\epsilon_i] = \Pr\{\epsilon_i = 1\}$

$$PPV(\hat{\mathrm{x}}, \mathrm{x}) = \frac{TP}{TP + FP} \Rightarrow PPV = \frac{1}{|\hat{\mathcal{T}}^c|} \sum_{i \in \hat{\mathcal{T}}^c} \epsilon_i .$$

Proposition 1. *The underlying AADL outperforms a random flip of a fair coin if and only if*

$$\sum_{i=1}^{m} \mathbf{p}_i > m p_A \qquad (21)$$

where $p_A \in (0, 1)$ is the expected percentage of attacked nodes. Furthermore, if the maximum percentage of attacked nodes is P_A, then (21) is only a sufficient condition.

Definition 2 (Pruning, Pruning algorithm, PPV_η). *A pruning algorithm is a procedure returning a subset support prior $\hat{\mathcal{T}}_\eta^c \subset \{1, \cdots, m\}$ of $\hat{\mathcal{T}}^c$ satisfying*

$$\hat{\mathcal{T}}_\eta^c \subseteq \hat{\mathcal{T}}^c . \qquad (22)$$

And the corresponding precision of the pruned support prior can be calculated as

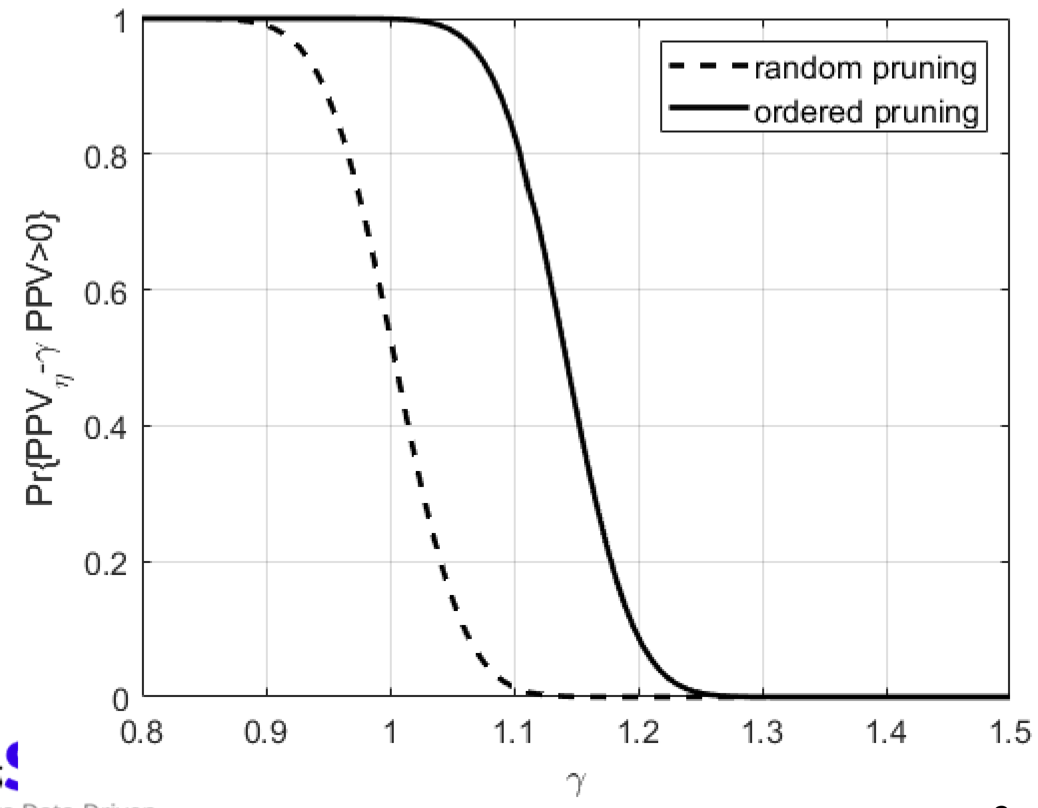
$$PPV_\eta = \frac{\sum_{i \in \hat{\mathcal{T}}_\eta^c} \epsilon_i}{|\hat{\mathcal{T}}_\eta^c|} . \qquad (23)$$

Reconstruction with Inexact Support Prior

Probability – Possibility Tradeoff behind Pruning

Proposition 2. *Given $\gamma_0 > 0$, then*

$$Pr\{PPV_\eta - \gamma_0 PPV > 0\} > 0 \text{ if and only if } \gamma_0 |\hat{\mathcal{T}}_\eta^c| < |\hat{\mathcal{T}}^c|.$$



Theorem V.2. *Given an estimated attack support $\hat{\mathcal{T}} \subseteq \{1, 2, \cdots, Tm\}$ with the uncertainty characteristic described in (30). Let $\hat{\mathcal{T}}_\eta$ be a pruned support estimate satisfying $\hat{\mathcal{T}}_\eta^c \subseteq \hat{\mathcal{T}}^c$, then*

$$Pr\{PPV_\eta - \gamma PPV \geq 0\} \geq \sum_{j=1}^{|\hat{\mathcal{T}}_\eta^c|+1} \left(\mathbf{r}_\eta(j) \sum_{i=1}^{\Phi_{j-1}+1} \tilde{\mathbf{r}}(i) \right), \tag{35}$$

where,

$$\mathbf{r}_\eta = \left(\prod_{i \in \hat{\mathcal{T}}_\eta^c} \mathbf{p}_i \right) \begin{bmatrix} -\mathbf{s}_{\eta,1} \\ 1 \end{bmatrix} * \begin{bmatrix} -\mathbf{s}_{\eta,2} \\ 1 \end{bmatrix} * \cdots * \begin{bmatrix} -\mathbf{s}_{\eta,|\hat{\mathcal{T}}_\eta^c|} \\ 1 \end{bmatrix},$$

$$\tilde{\mathbf{r}} = \left(\prod_{i \in \hat{\mathcal{T}}^c \setminus \hat{\mathcal{T}}_\eta^c} \mathbf{p}_i \right) \begin{bmatrix} -\tilde{\mathbf{s}}_1 \\ 1 \end{bmatrix} * \begin{bmatrix} -\tilde{\mathbf{s}}_2 \\ 1 \end{bmatrix} * \cdots * \begin{bmatrix} -\tilde{\mathbf{s}}_{|\hat{\mathcal{T}}^c \setminus \hat{\mathcal{T}}_\eta^c|} \\ 1 \end{bmatrix},$$

and $\Phi_k = \min\left\{ \left\lceil \frac{|\hat{\mathcal{T}}^c|}{\gamma |\hat{\mathcal{T}}_\eta^c|} - 1 \right\rceil k, |\hat{\mathcal{T}}^c| - |\hat{\mathcal{T}}_\eta^c| \right\}$, $\mathbf{s}_{\eta,i} = -\frac{1-\mathbf{p}_{\hat{\mathcal{T}}_\eta^c,i}}{\mathbf{p}_{\hat{\mathcal{T}}_\eta^c,i}}$, $\tilde{\mathbf{s}}_i = -\frac{1-\mathbf{p}_{\hat{\mathcal{T}}^c \setminus \hat{\mathcal{T}}_\eta^c,i}}{\mathbf{p}_{\hat{\mathcal{T}}^c \setminus \hat{\mathcal{T}}_\eta^c,i}}$.

Physics-Data-Driven

FAMU-FSU Engineering

Reconstruction with Inexact Support Prior

A Robust Pruning Algorithm

Algorithm 1 A Robust Pruning Algorithm

i. Obtain the maximum quantity l_η of safe channels that are localized by $\hat{\mathcal{T}}^c$ correctly with a probability of at least $\eta \in (0,1)$:

Aggressive parameter

$$l_\eta = \max\left\{ |\mathcal{I}| \,\middle|\, \prod_{i\in\mathcal{I}} \mathbf{p}_i \geq \eta,\ \ \mathcal{I} \in \hat{\mathcal{T}}^c \right\}. \qquad (24)$$

ii. Use the current localization prior $\hat{\mathbf{q}}$ and the AADL's historical performance $\mathbf{p}$ to extract the l_η safest nodes as follows.

$$\hat{\mathcal{T}}_\eta^c = \{\mathsf{argsort} \downarrow (\mathbf{p} \odot \hat{\mathbf{q}})\}_1^{l_\eta}. \qquad (25)$$

Consider historical performance

where, $\{\cdot\}_1^{l_\eta}$ is an index extraction from the first elements to l_η elements.

Consider current conclusion

Theorem 2. *Suppose there exists an AADL generating estimated support prior $\hat{\mathcal{T}}^c$ satisfying (19), through the* Algorithm V, *the precision of the pruned support prior $\hat{\mathcal{T}}_\eta^c$ satisfies*

$$Pr\{PPV_\eta = 1\} \geq \eta.$$

Theorem 3 (Least Square Reconstruction with Prior Pruning). *Consider the linear measurement model given in* (6). *Suppose there exists an AADL that gives an estimate, $\hat{\mathcal{T}}$, of $\mathsf{supp}(\mathbf{e})$ with uncertainty described by* (19). *Given a parameter $\eta \in (0,1]$ with corresponding quantity l_η given by* (24), *let $\hat{\mathcal{T}}_\eta$ be a new support with the indicator $\hat{\mathbf{q}}_\eta$ defined by* (25). *If $l_\eta - |\mathrm{supp}(\mathbf{e})| \geq n$, then, with a probability of at least η, the least-square estimator* (26) *satisfies the error bound*

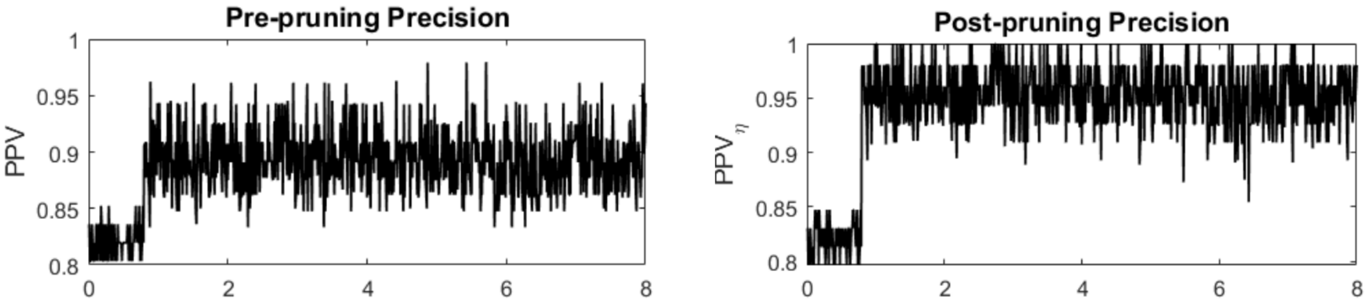
$$\|\hat{\mathbf{x}}_\eta - \mathbf{x}^*\| \leq 2\min\left\{\frac{\delta}{2}, \frac{\varepsilon}{1-\delta_n}\right\}, \qquad (27)$$

RAS lab
Physics-Data-Driven

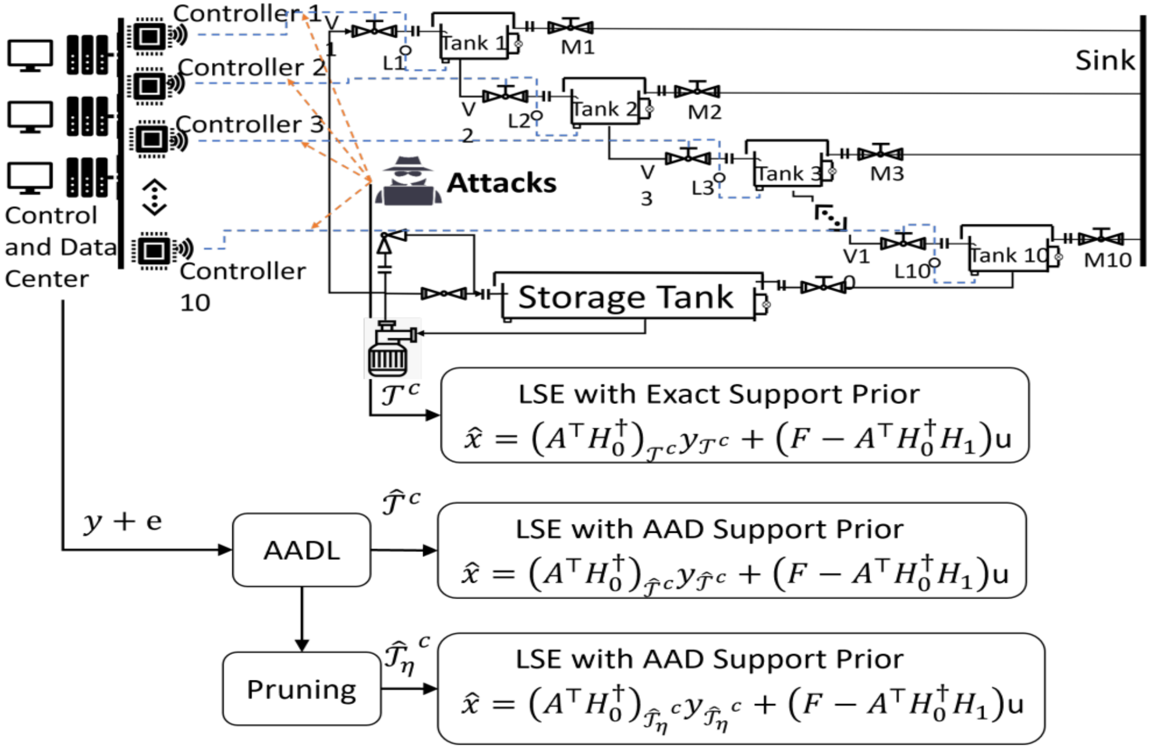
FAMU-FSU Engineering

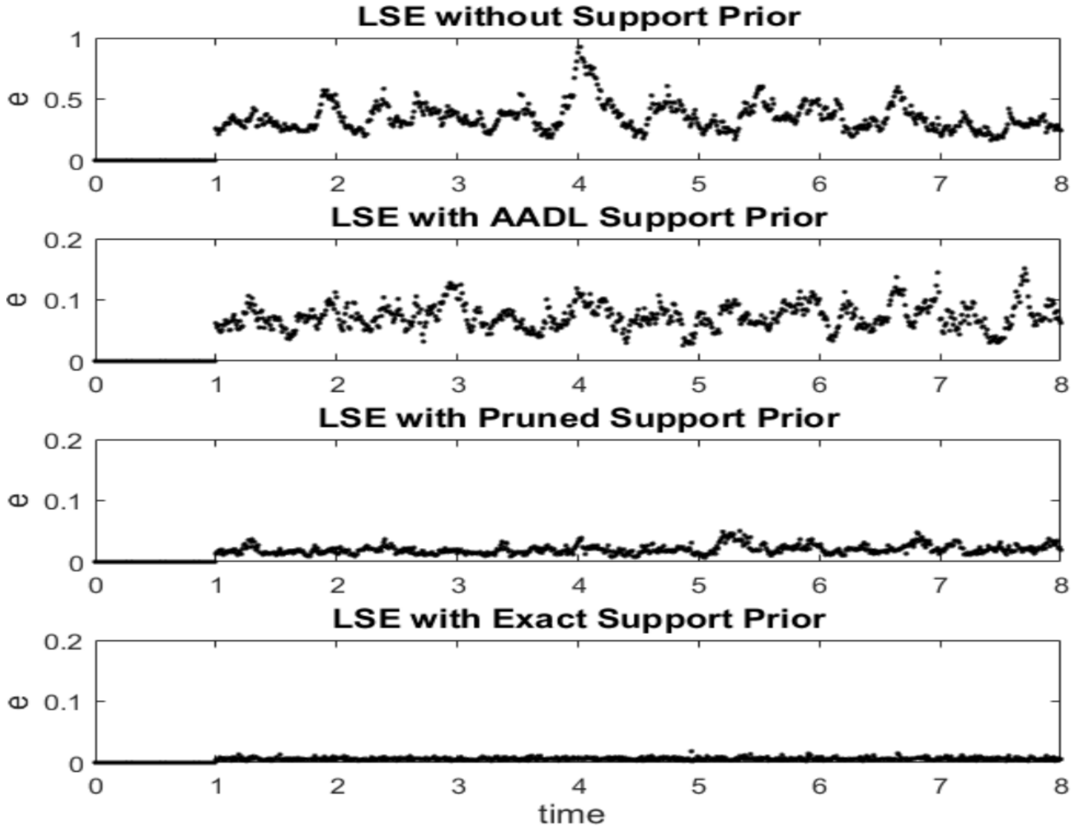
Simulation

- Pruning-based resilient estimation



Water Tank System



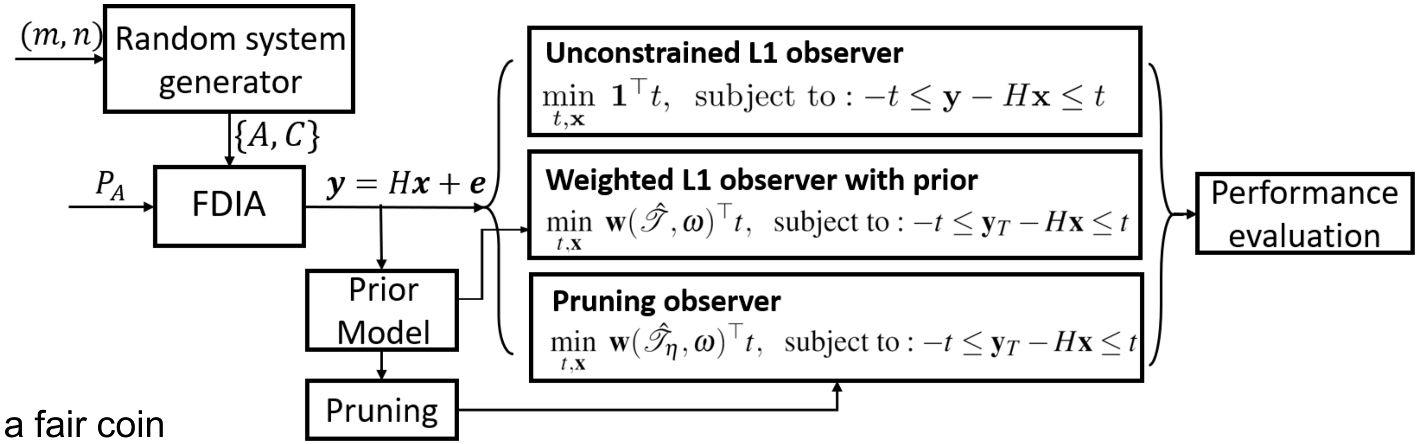


RAS lab
Physics-Data-Driven

FSU Engineering

Simulation

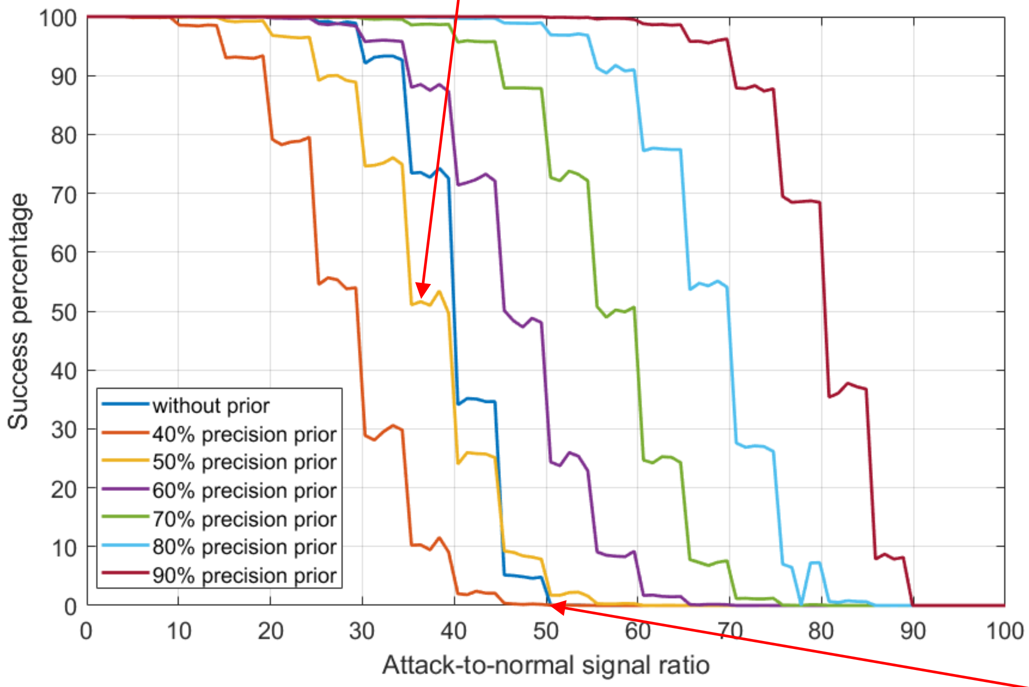
- Pruning-based resilient estimation

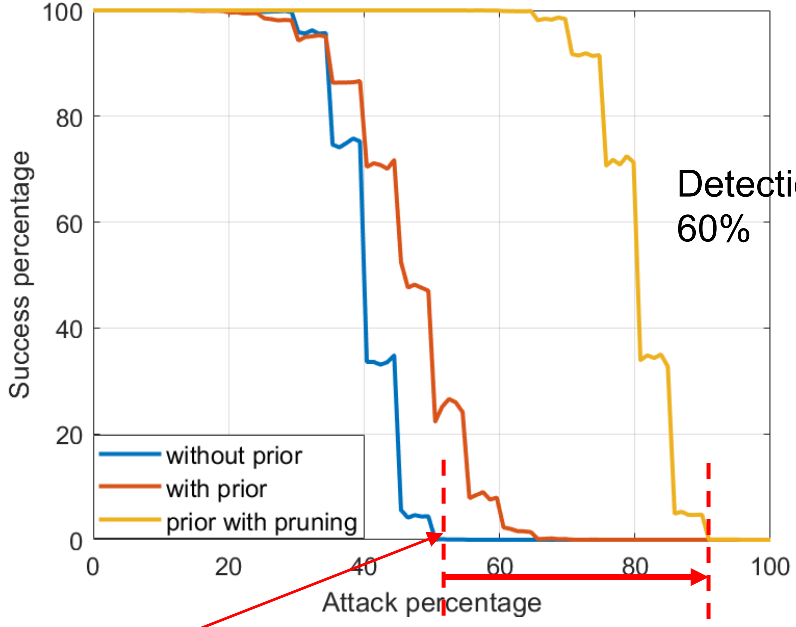


Numerical Simulation (Monte-Carlos)

Require Prior precision > 50%

Better than random flip of a fair coin





Detection precision = 60%

50% Limitation in the literature

Our Contribution

RAS lab Physics-Data-Driven

FAMU-FSU Engineering

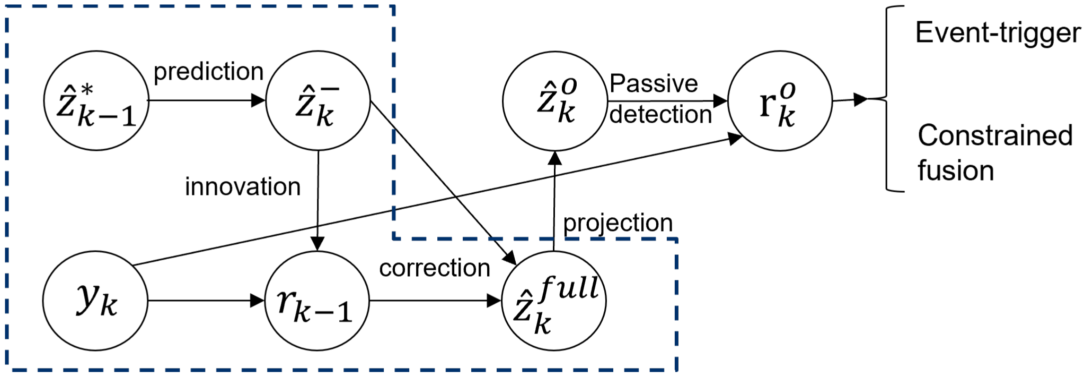
Summary and Future Work

Summary

1. Resilient Signal Reconstruction with Data-driven Prior
2. Model False Positive Uncertainty, Pruning Algorithm

Future Work

1. Instead of attack isolation, a better way to incorporate data-driven prior in 2-norm observers



2. For resilient 1-norm observers, a Lyapunov analysis framework is expected

THANK YOU

Yu Zheng, yzheng6@fsu.edu
Olugbenga Moses Anubi, oanubi@fsu.edu