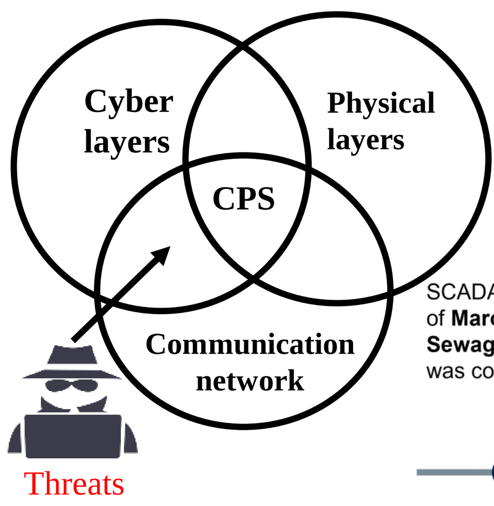
Moving-horizon False Data Injection Attack Design against Cyber-Physical Systems

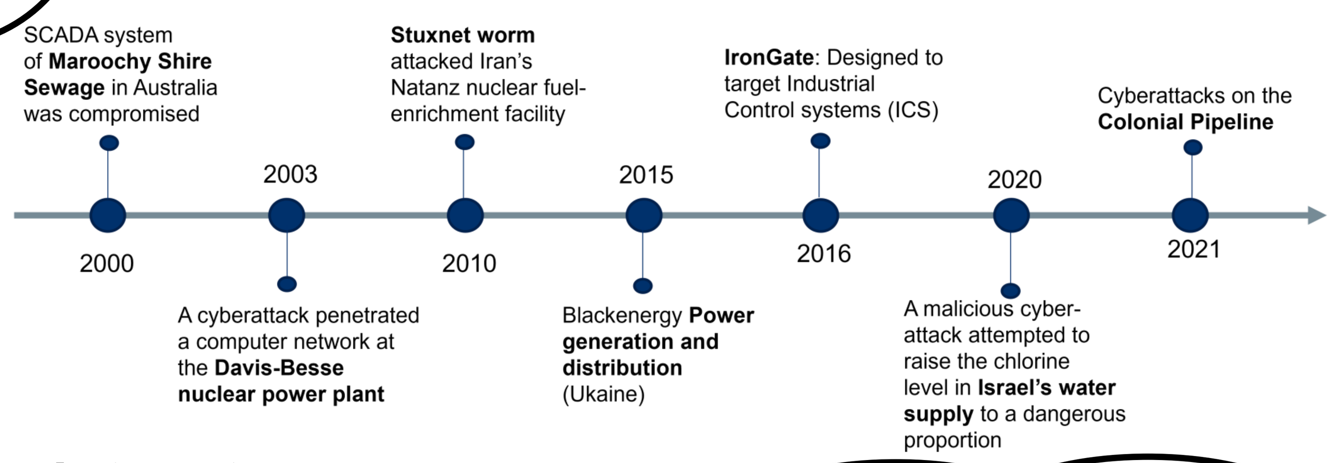
Yu Zheng, Sridhar Babu Mudhangulla, Olugbenga Moses Anubi

RAS lab Physics-Data-Driven | FAMU-FSU Engineering

Introduction and Motivation

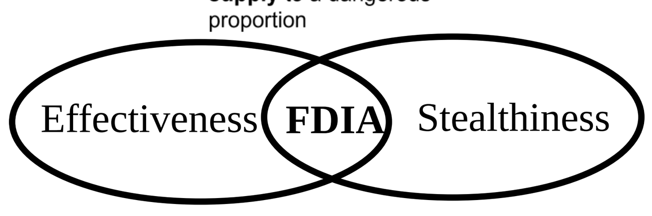


Closed-loop interaction between cyber world and Physical world ensure the chance of cyber attacks to cause catastrophe in physical world



False data injection attacks (FDIAs)

- Maximizing effectiveness (closeness to intended degradation)
- Maintaining stealthiness (potential to bypass BDD)



Attack Generation Problem in Literature:

Full Model Knowledge	Least Square Estimator (LSE), residual – based BDD Kalman filter with $\chi 2$ detector	Inefficient and less pragmatic FDIAs
Reduced Model Knowledge	Limited access to sensors Incomplete knowledge of system dynamics Incomplete knowledge of implemented state estimators	Inefficient but more pragmatic FDIAs
Data-driven Approaches	Learn system model from runtime data Generative network	Inefficient but most pragmatic FDIAs

★ How does the attack history affect the feasibility of the FDIA? (Recursive Feasibility)
★ Design focus of MHE - how to guarantee feasibility, Over the next window?

MH - FDIA

Model Development

Model: Discrete LTI model to approx. plant model

Closed-from dynamical model

$$\mathbf{x}_{i+1} = A\mathbf{x}_i$$
$$\mathbf{y}_i = C\mathbf{x}_i + \mathbf{v}_i,$$
$$where, A = A' + B'K,$$
$\mathbf{v}_I \in \mathbb{R}^{Tm}$ – noise vector.

Measurement model on the window I of length T

$$\mathbf{y}_I = H\mathbf{x}_i + \mathbf{v}_I$$

$$H = \begin{bmatrix} CA^{1-T} \\ CA^{2-T} \\ \vdots \\ C \end{bmatrix} = [U_1 \quad U_2] \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^\top$$

MHE:

Def: Operator $\mathcal{D}: \mathbb{R}^{Tm} \to \mathbb{R}^n$

Returns: estimate of state vector (T-horizon observation)

Stability: $\|\mathcal{D}(\mathbf{y}_I) - \mathbf{x}_i\|_2 \le \tau_0 \|\mathbf{x}_i\|_2 + \varepsilon_0$, $where\ \tau_0, \varepsilon_0 < \infty$

$\boldsymbol{\ell_2}$ ***MHE***: $\mathcal{D}_2(\mathbf{y}_I) \triangleq \arg\min_{\mathbf{x}} \| \mathbf{y}_I - H\mathbf{x} \|_2 = H^\dagger \mathbf{y}_I$

BDD:

Obj: Monitor state estimate and detect malicious inputs

Designed: Based on residual, $\|\mathbf{y}_I - H\mathcal{D}(\mathbf{y}_I)\|_2$

Def : $\mathrm{BDD}(\mathbf{y}_I) = \begin{cases} 1 & if\ \|\mathbf{y}_I - H\mathcal{D}(\mathbf{y}_I)\|_2 > \delta \\ 0 & otherwise \end{cases}$

Problem Formulation

Quantifying Effectiveness:

Estimation error, $\|\mathcal{D}(\mathbf{y}_I) - \mathcal{D}(\mathbf{y}_I + \mathbf{e}_I)\|_2$

Quantifying Stealthiness:

Estimation residual, $\|\mathbf{y}_I + \mathbf{e}_I - H\mathcal{D}(\mathbf{y}_I + \mathbf{e}_I)\|_2$

Definition (Successful FDIA):

Given the estimator-detector pair $(\mathcal{D}(\mathbf{y}_I), \mathrm{BDD}(\mathbf{y}_I))$, the attack vector $\mathbf{e}_I \in \Sigma_k$ is said to be (α, ϵ)-successful if:

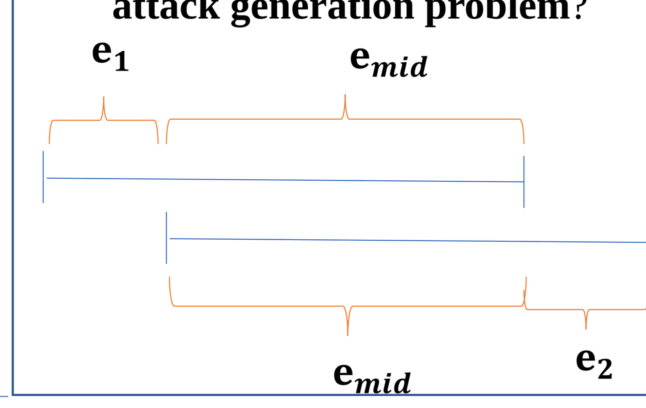
$$\|\mathcal{D}(\mathbf{y}_I) - \mathcal{D}(\mathbf{y}_I + \mathbf{e}_I)\|_2 \ge \alpha, \|\mathbf{y}_I + \mathbf{e}_I - H\mathcal{D}(\mathbf{y}_I + \mathbf{e}_I)\|_2 \le \epsilon$$

Given an attack history, $\mathbf{e}_{I^-} = [\mathbf{e}_{i-T+1}^\top, \mathbf{e}_{i-T+2}^\top, \cdots, \mathbf{e}_{i-1}^\top]^\top$

Satisfies — If $\mathbf{e}_I = [\mathbf{e}_{I^-}^\top, \mathbf{e}_i^\top]^\top \Rightarrow \mathbf{e}_i^\top\ is (\alpha, \epsilon)$–successful

$\boldsymbol{\ell_2}$ ***MHE*** $\Rightarrow \| H^\dagger \mathbf{e}_I \|_2 \ge \alpha, \| (I - HH^\dagger)(\mathbf{y}_I + \mathbf{e}_I) \|_2 \le \epsilon$

What is **recursive feasibility** in **attack generation problem**?



Moving horizon FDIA

Static Successful FDIA

Given, $\boldsymbol{w}_1 \in \mathbb{R}^n$, $\boldsymbol{w}_2 \in \mathbb{R}^{mT-n}$

A T sequence vector of attack:

$$\mathbf{e}_I = U_1 \Sigma \boldsymbol{w}_1 + U_2 \Sigma \boldsymbol{w}_2$$

$is\ (\| \boldsymbol{w}_1 \|_2, \| \boldsymbol{w}_2 \|_2)$ – successful

+

Recursive Feasibility

$$\begin{bmatrix} \mathbf{e}_{I^-} \\ \mathbf{e}_i \end{bmatrix} = \mathbf{e}_I = U_1 \Sigma \boldsymbol{w}_1 + U_2 \Sigma \boldsymbol{w}_2$$

Given attack history $\mathbf{e}_{I^-}$, Decide $\boldsymbol{e}_i$

Feasibility: $\| N_2 \mathbf{v} + \mathbf{w}_2^- \|_2 \le \tilde{\epsilon}$

Effectiveness: $\alpha(\boldsymbol{v}) = \| N_1 \mathbf{v} + \mathbf{w}_1^- \|_2$

Asides:

$$\mathbf{w}_1^- = \Sigma^{-1} U_1^\top \begin{bmatrix} \mathbf{e}_{I^-} \\ 0 \end{bmatrix}, \mathbf{w}_2^- = U_2^\top \begin{bmatrix} \mathbf{e}_{I^-} \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{e}_i \end{bmatrix} = U_1 \Sigma N_1 \boldsymbol{v} + U_2 \Sigma N_2 \boldsymbol{v}$$

$\begin{bmatrix} N_1 \\ N_2 \end{bmatrix}$ *is any matrix in the null space of* $[U_1 \Sigma \ \ U_2]_{\mathcal{T}^c}$, $\mathcal{T}$ *is the support of* e_i

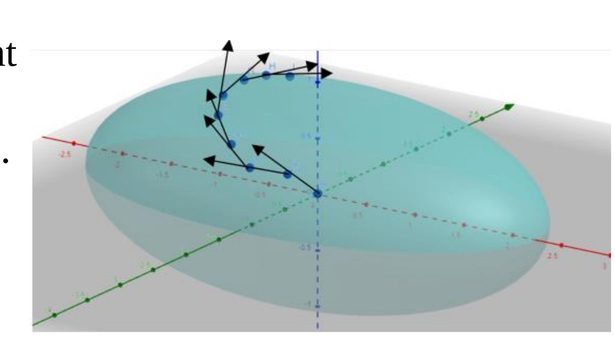
=

Successful MH-FDIA Algorithm

(Projected Gradient Ascent)

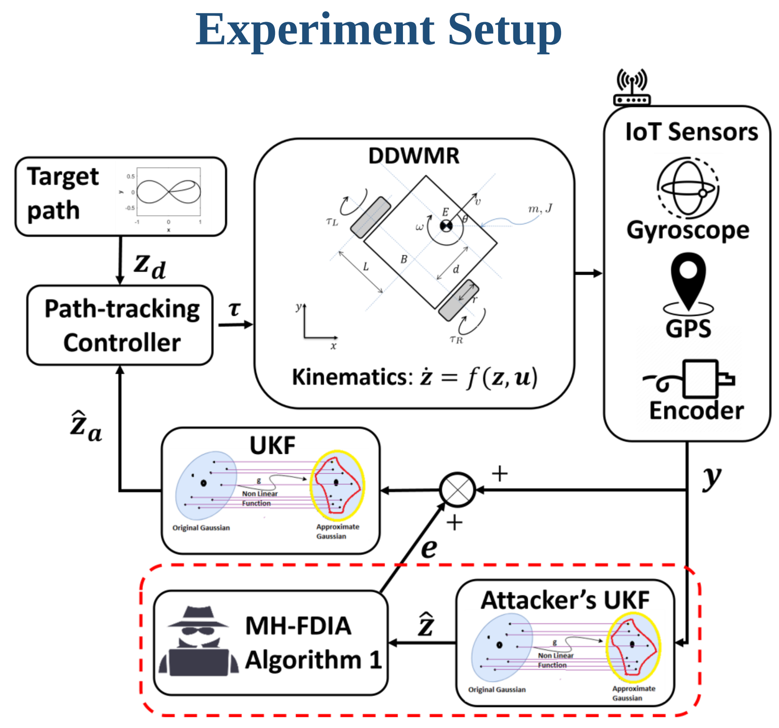
The algorithm searches on the gradient ascent direction of $\alpha(\mathbf{v})$ inside the ellipsoid $S = \{\mathrm{v} | \| N_2 \mathbf{v} + \mathbf{w}_2^- \|_2 \le \tilde{\epsilon}\}$.

When approach the boundary, it will stay on the boundary.

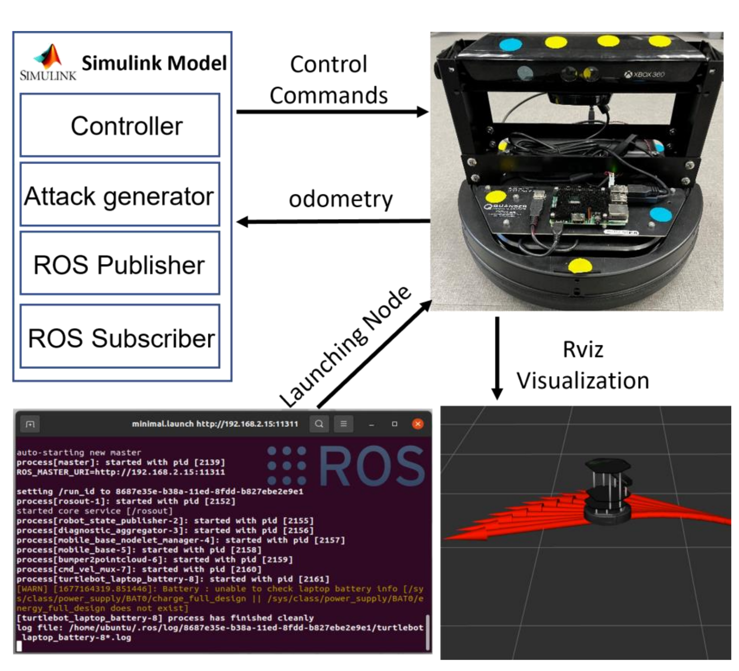


Experiment

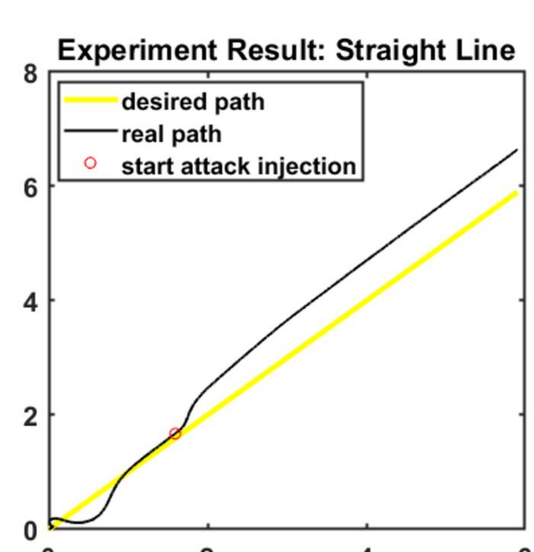
Experiment Setup

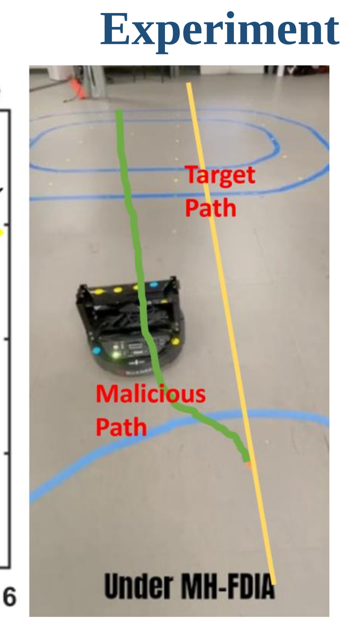


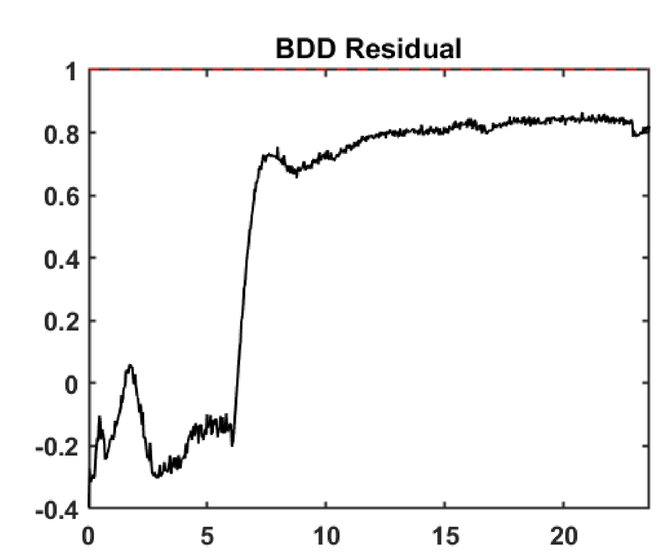
Experiment Platform



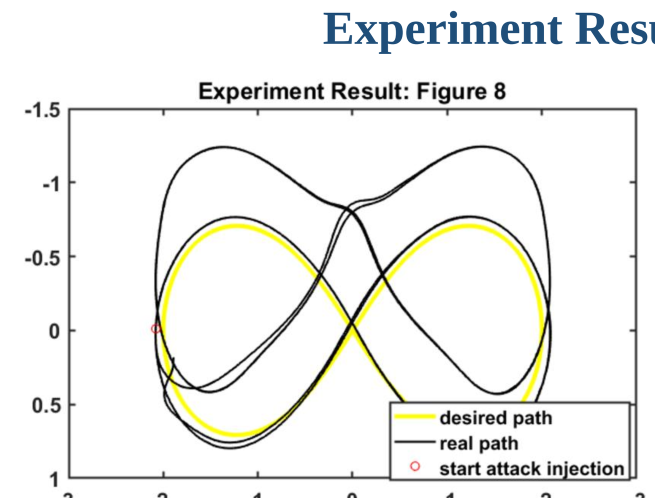
Experiment Result: Line

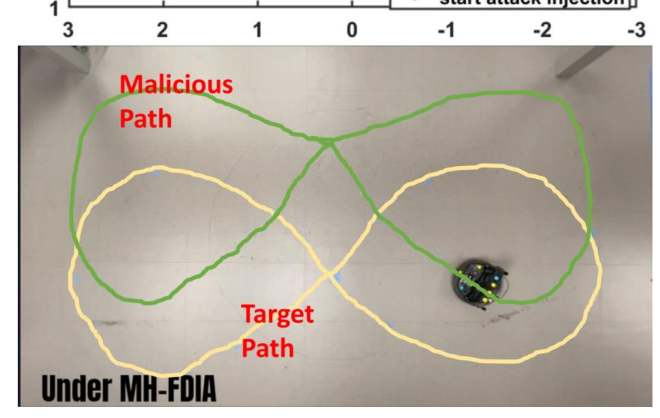


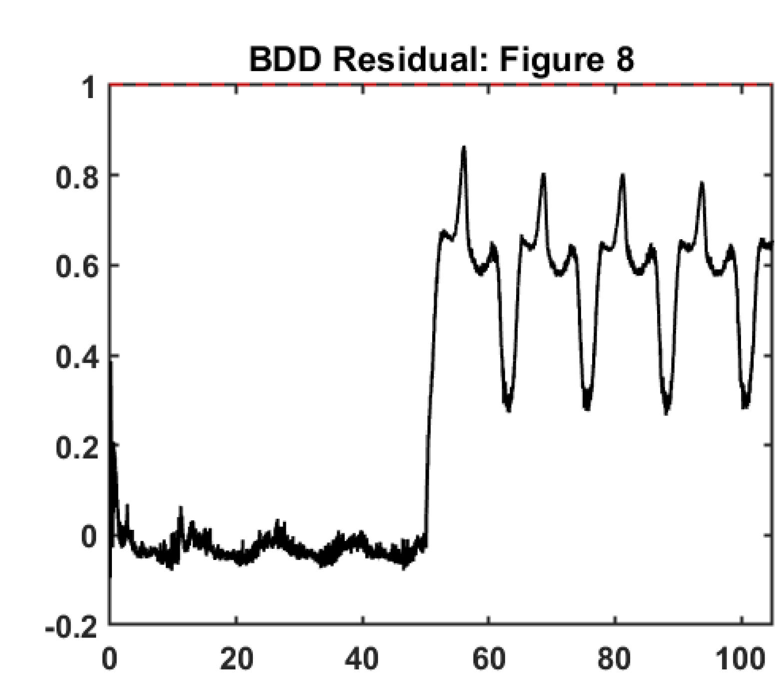




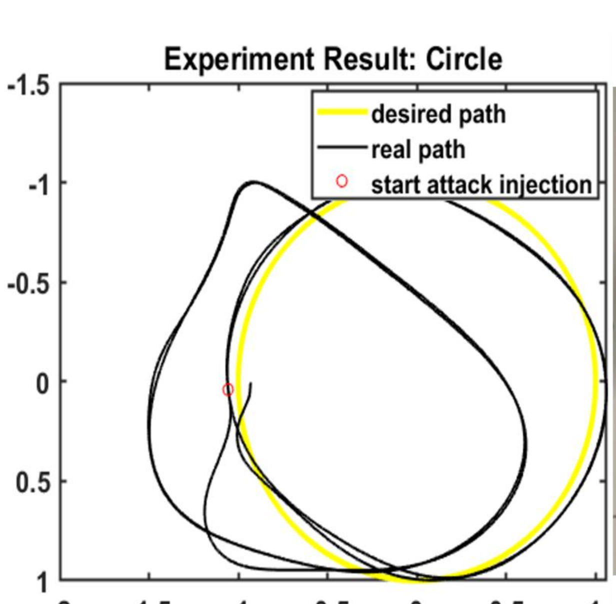
Experiment Result: Figure-8

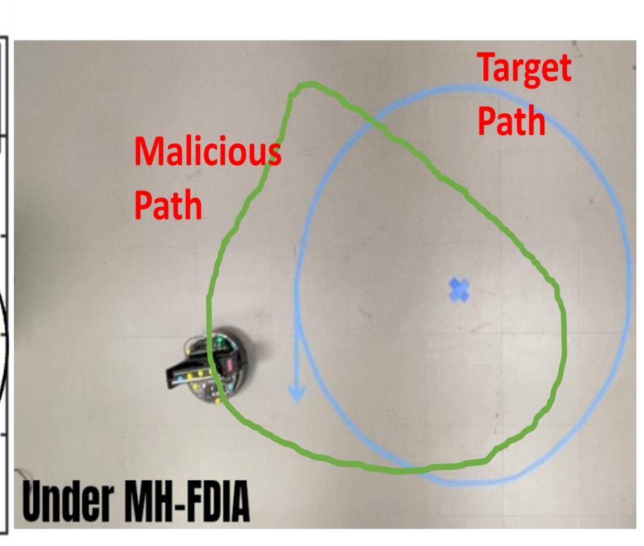


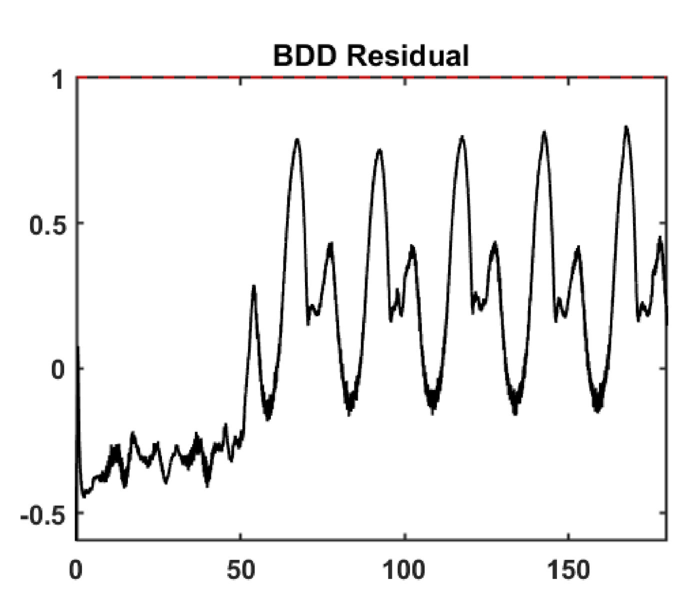




Experiment Result: Circle







Conclusion and Future work

This poster presents a complete framework of MH-FDIA design including attack effectiveness improvement algorithm.

1) Based on a formal definition of successful FDIA, the MH-FDIA design is given against $\boldsymbol{\ell_2}$ MHE and BDD, and shown to be (α; ϵ)-successful.
2) An adaptive algorithm is proposed to search for the most successful FDIAs while preserving recursive feasibility

Future work:

Given a pre-defined malicious trajectory, how to design successful MH-FDIA.

Reference

Paper:

Y. Zheng, S. Mudhangulla and O. Anubi, "Moving-horizon False Data Injection Attack Design against Cyber-Physical Systems ", Control Engineering Practice, 2023, [under review]

Experiment video Link:
(36) Safe Autonomy - YouTube

Contacts

Yu Zheng, IEEE Student Member, Florida State University, Email: yzheng6@fsu.edu

Sridhar Babu Mudhangulla, Florida State University, Email: sm19ch@fsu.edu

Olugbenga Moses Anubi, IEEE Senior Member, Florida State University, Email: oanubi@fsu.edu