An Integrated Generative Adversarial Network for Identification and Mitigation of Cyber-Attacks in Wide-Area Control

Jishnudeep Kar  
PhD Student  
North Carolina State University  
(jkar@ncsu.edu)

Aranya Chakrabortty  
Professor  
North Carolina State University  
(achakra2@ncsu.edu)
About DoS/FDI attacks

About DoS
• The communicating servers are jammed with malicious request
• Server becomes unable to respond to legitimate requests.

About FDI
• The data shared over the network is corrupted by adding bias.
• This could cause controllers and operators to actuate wrong control actions which may cause closed-loop instability.
About DoS/FDI attacks

- Detect and localize the attack.
- Operate for a varying range of operating conditions.
- Estimate missing or attacked states using GAN to sustain closed-loop stability.

Fig. 2: Cyber-physical architecture for WAC

\[
\Delta x = A \Delta x + Bu
\]
\[
u = -K \Delta x
\]
\[
J = \int_0^\infty \left( \Delta x^T Q \Delta x + u^T Ru \right) dt, \quad Q \geq 0, \quad R > 0
\]

\[
\Delta u^a_{G_i}(t) = - \left( \sum_{G_j \in S_i} K_{i,j} (x_{G_j}(t) - x_{o,G_j}) + \sum_{G_k \in A_i} K_{i,j} (x^a_{G_k}(t) - x_{o,G_k}) \right)
\]
Why is resiliency needed?

Small-signal power grid model
\[ \Delta \dot{x} = A \Delta x + Bu \]

Design a **damping** control input
\[ u = -K \Delta x \]

Minimizing LQR cost
\[ J = \int_{0}^{\infty} (\Delta x^T Q \Delta x + u^T Ru) dt, \quad Q \geq 0, \quad R > 0 \]
Generative Adversarial Network
Our proposed GAN

**Transmission Network**

- G1, G2, G3, G4: Generators
- RCM: Resilient Control Module
- Δu: Change in control signal

**GAN**

- Discriminator (D)
- Generator (G)

**Decentralized KF**

\[ x = [x_{G1}, x_{G2}, x_{G3}, x_{G4}] \]

**Training**

- Input to discriminator is a concatenation of real data and generator output.
- Input to LSTM encoder is attacked state vector from moving data window.

**Discrimination**

- Input to discriminator is data from moving window for attack detection and identification.
- Output of LSTM decoder is imputed healthy states to form augmented state vector.

**Generator (G)**

\[ P_{a}^{G_j}(t) = \left( D(x_{t}^{G_j}) + D(x_{t-1}^{G_j}) + \ldots + D(x_{t-d+1}^{G_j}) \right) / d. \]

**Moving average – WHY?**

- GANs cannot be trained to 100% accuracy.
- Instances of anomalous scores.
- Averaging removes anomaly.
Simulation

Fig. 4: IEEE 68-bus system

- Attacked links are shown by red lines.
- Training data consists of 5000 operating points.

- Communication delays are:
  - Intra-area = 30ms, Inter-area = 60ms.
  - Deviation = +/- 10%
Detection and identification (FDI)

It is seen that during a FDI attack, the average discriminator $P_a$ shows a sudden drop.

The threshold can be estimated based on the best score obtained during training phase.

Generator wise reconstruction error is computed between received and predicted states.

$$L_{r}^{G_j}(t) = \left\| \left( \mathcal{X}_t - \mathbf{G}(\mathcal{X}_t) \right) \odot \Omega_{G_j} \right\|$$
Closed-loop results

Fig. 6a: FDI attack causes closed-loop instability

Fig. 6b: Resiliency to FDI using GAN resiliency

Fig. 7a: DoS causes closed-loop instability

Fig. 7b: Resiliency to DoS using GAN
Conclusions and references

- Neural network-based methods benefit in not requiring the actual model to ensure resiliency during a cyber-attack.
- Can be implemented in a decentralized manner ensuring model privacy.
- Proposed GAN based method work effectively to both localize and mitigate both FDI and DoS cyber-attacks.

- **Future Work**: Large changes in operating points, non-linear controller, IBRs

---

**References**