Optimal False Data Injection Attack against Automatic Generation Control in Power Grids

Rui Tan[1] Hoang Hai Nguyen[2] Eddy. Y. S. Foo[3] Xinshu Dong[4]

David K. Y. Yau[1,5] Zbigniew Kalbarczyk[2] Ravishankar K. Iyer[2] Hoay Beng Gooi[3]

[1] School of Computer Science & Engineering, Nanyang Technological University

[2] Coordinated Science Lab, University of Illinois at Urbana-Champaign

[3] School of Electrical & Electronic Engineering, Nanyang Technological University

[4] Advanced Digital Sciences Center, Illinois at Singapore

[5] Singapore University of Technology and Design

Attacks against Power Grids

- **59%** attacks on critical infrastructures target grids [DHS'13]
  - *Night Dragon*: grid operation **data exfiltration** [McAfee'11]
  - *Dragonfly*: **Trojan horses** in grid control systems [Symantec'14]

Frequency Control

- Maintain freq. at 50 or 60 Hz when loads change

Frequency-insensitive load

Frequency-sensitive load

$$P_L + \boxed{D \cdot \omega} = P_G$$

Generation

[Kundur 94]

D: load-damping constant

ω: system frequency

- Widespread and costly impact of failure
  - System frequency is global
  - −0.5 Hz: load shedding (regional blackout)
  - ±2.0 Hz: permanent equipment damage

Automatic Generation Control (AGC)

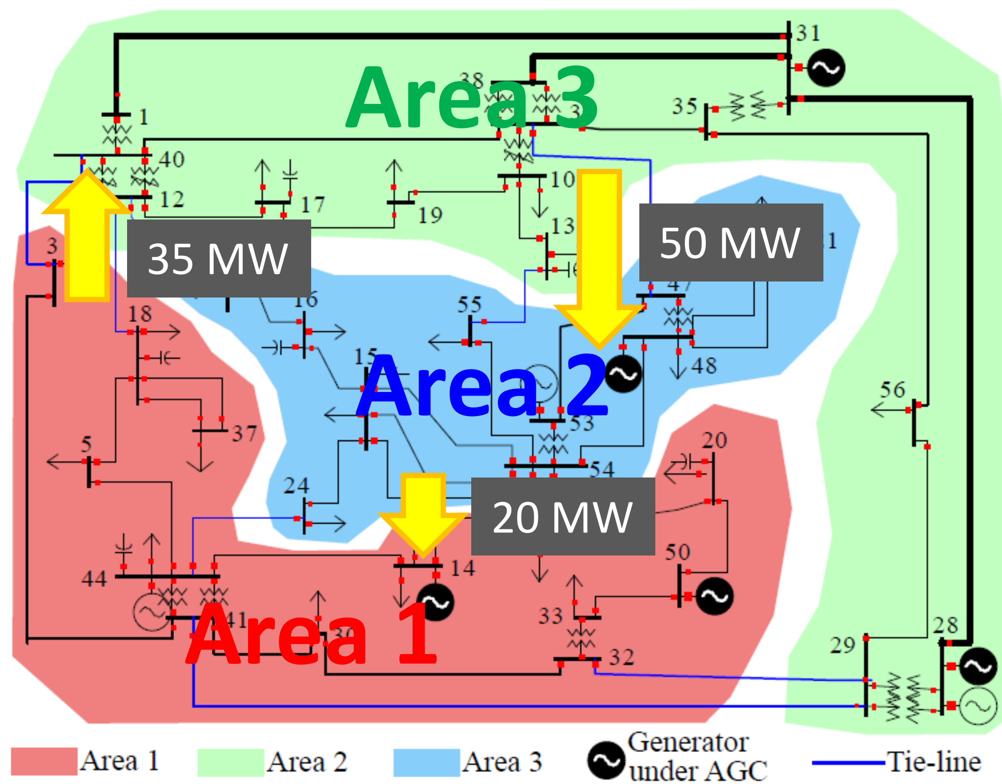
- More than frequency control
  - Regulates power exchanges btw *control areas*
  - **Input:** freq. deviation, area power export deviations
  - **Output:** Area Control Error (ACE)

A 37-bus grid

$$ACE = \alpha \bullet \Delta\omega + \beta \bullet \Delta p_E$$

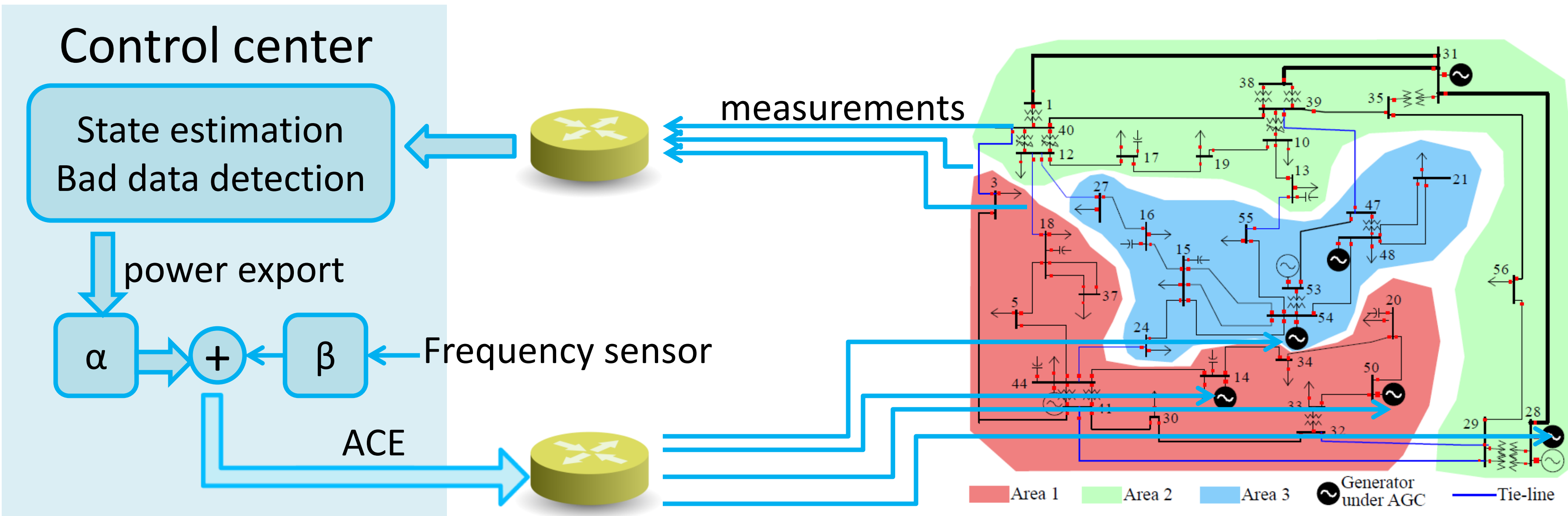
Frequency deviation

Power export deviation



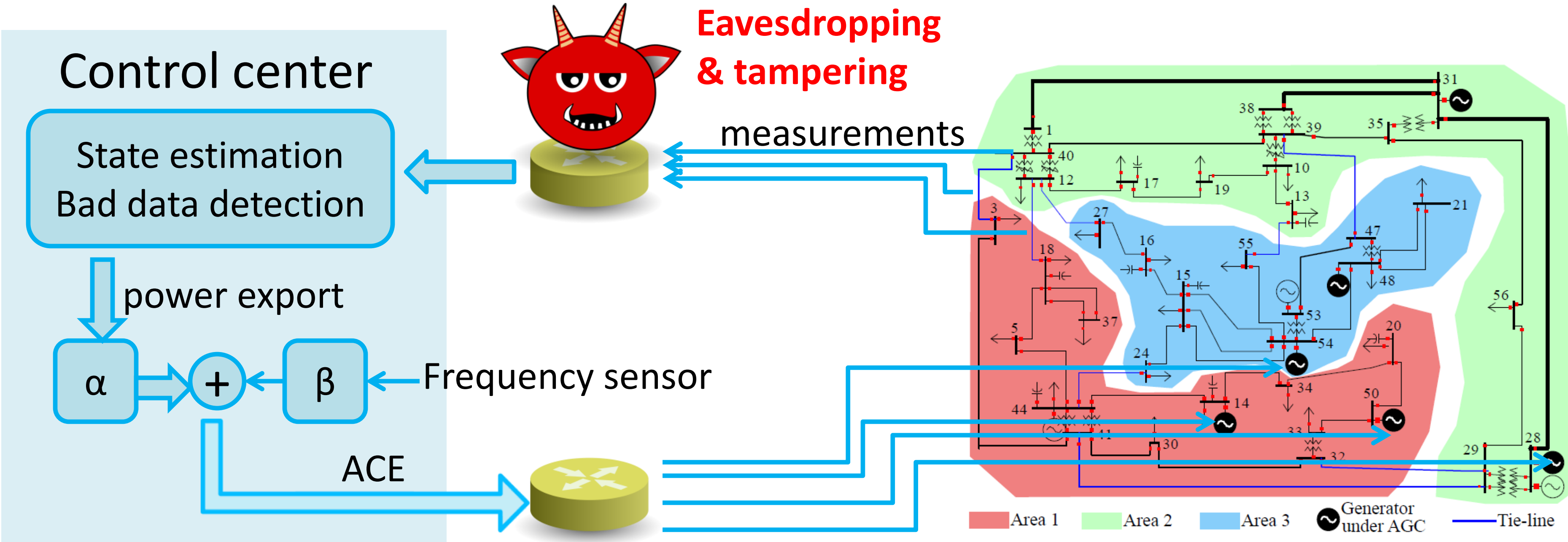
    - Further dispatched to generators
  - **Cycle:** 2 to 4 secs

Implementation & Threat



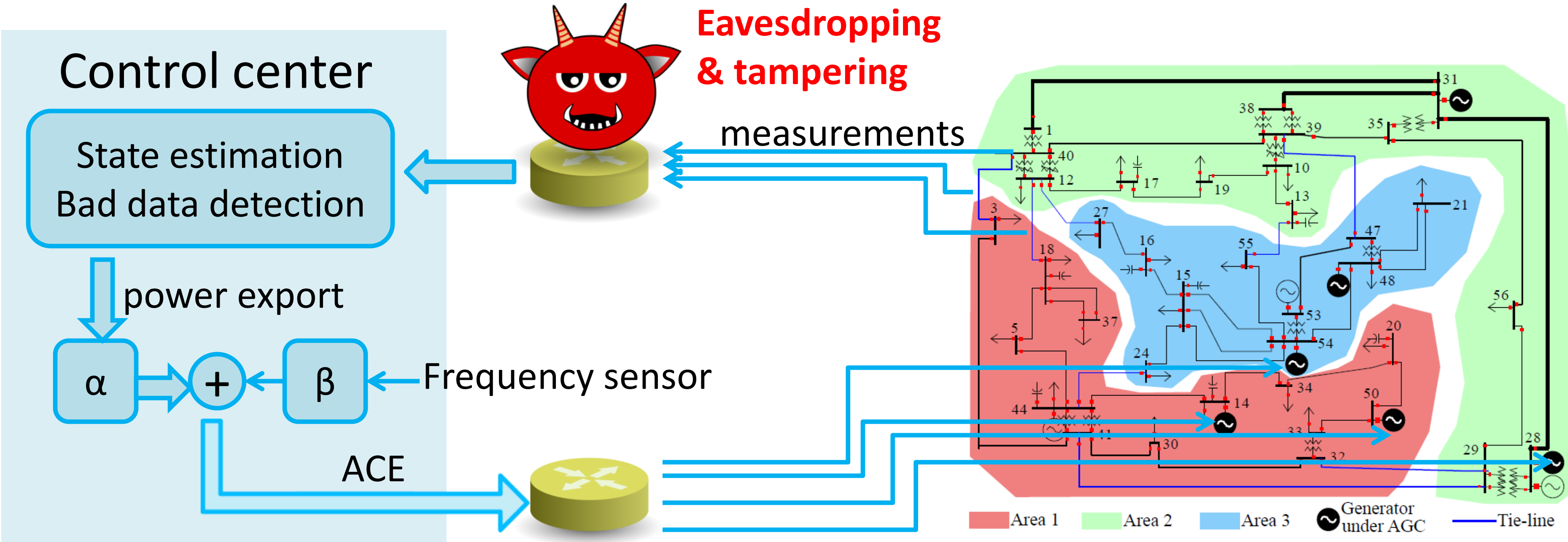
- Networked control system
  – Distributed, networked sensors
  – Control center
  – Generators

Implementation & Threat



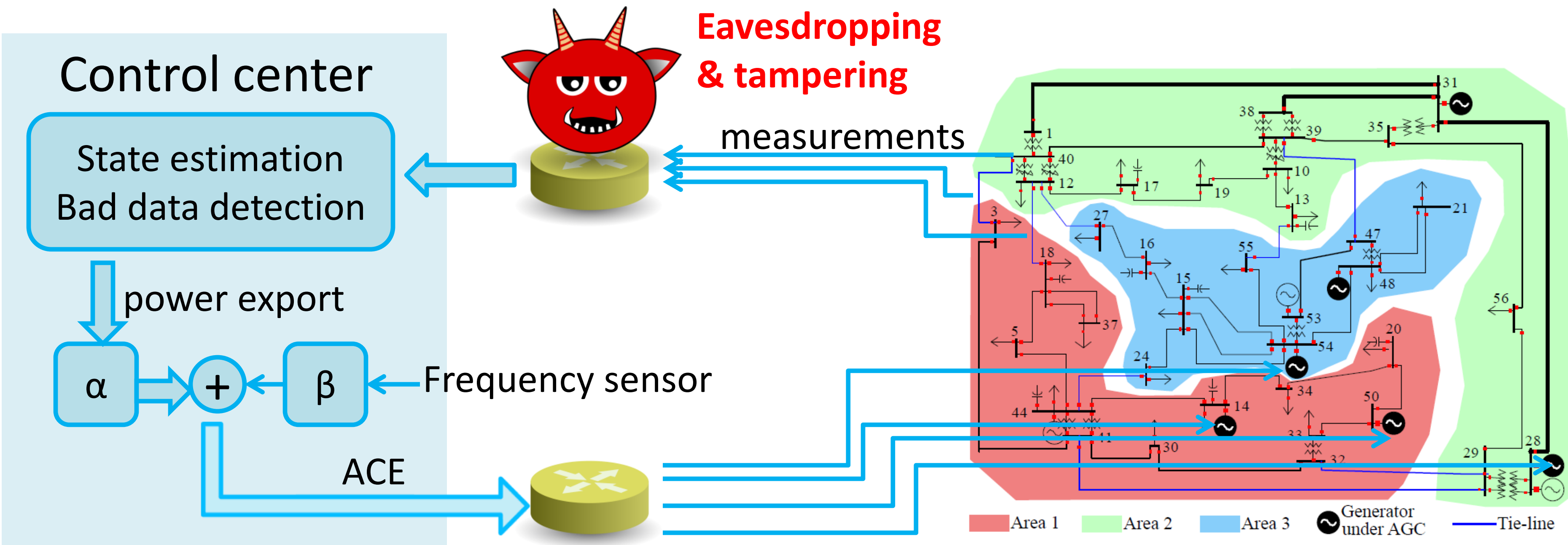
- Networked control system
  - Distributed, networked sensors
    - Logically isolated links (e.g., VPN) in existing networks
  - Control center
  - Generators

Implementation & Threat



- Networked control system
  - Distributed, networked sensors
    - Logically isolated links (e.g., VPN) in existing networks
  - Control center
    - Well protected
  - Generators

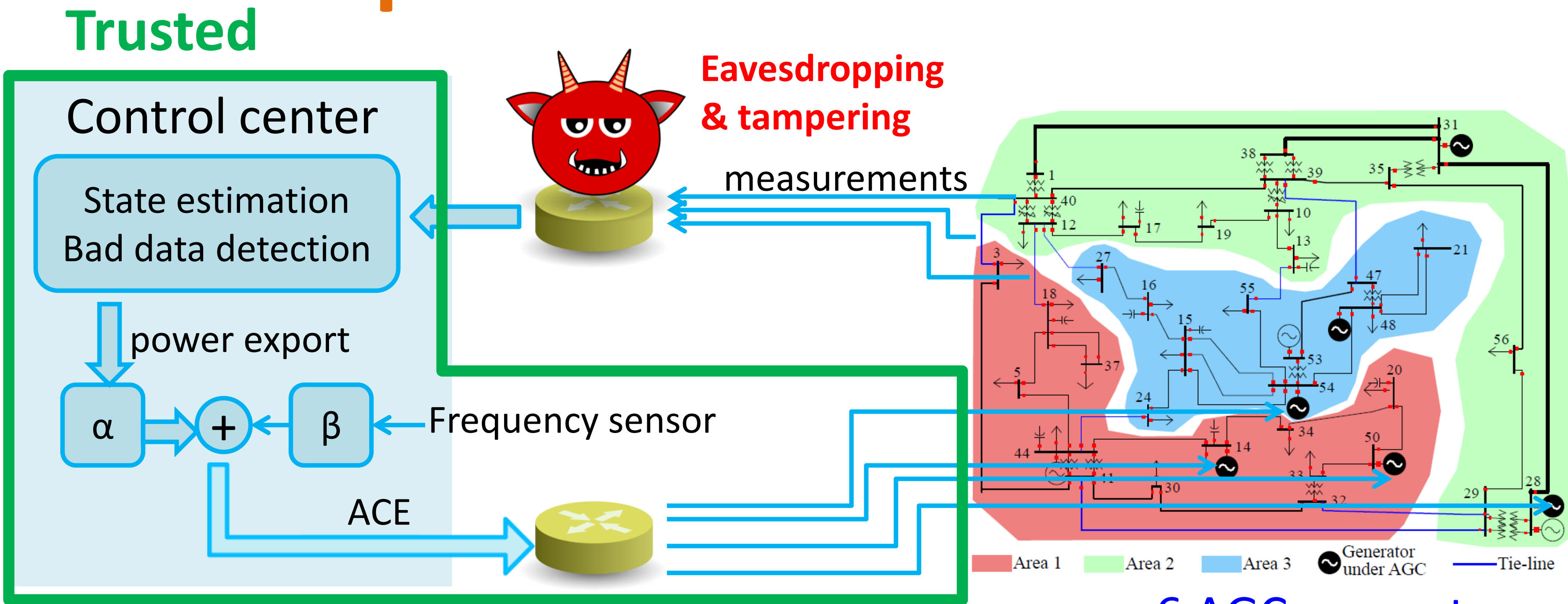
Implementation & Threat



6 AGC generators vs. 81 sensors

- Networked control system
  - Distributed, networked sensors
    - Logically isolated links (e.g., VPN) in existing networks
  - Control center
    - Well protected
  - Generators
    - Limited #, physically isolated links

Implementation & Threat



6 AGC generators vs. 81 sensors

- Networked control system
  - Distributed, networked sensors
    - Logically isolated links (e.g., VPN) in existing networks
  - Control center
    - Well protected
  - Generators
    - Limited #, physically isolated links

Outline

- Motivation & Background
- **Attack Model**
- Optimal Attack
- Simulations & Testbed Experiments

False Data Injection (FDI)

- Corrupt measurements

$$\mathbf{z}' = \mathbf{z} + \mathbf{a}$$

Attack vector

Measurement vector

  – Can read **z** and corrupt a subset of **z** elements
  – Stealthy to fault/attack detectors
    - Bypass bad data detection [Liu CCS'09]

$$\mathbf{a} = \mathbf{H}\mathbf{c}$$

Arbitrary vector

Measurement matrix

    - Bypass data quality checks

$$-\mathbf{a}_{\max} \prec \mathbf{a} \prec \mathbf{a}_{\max}$$

Research Question

- **Optimal attack** via worst-case analysis
  - A sequence of FDIs to deviate frequency to unsafe level **in shortest time**
    Protection assessment given attack response delay
  - **How to compute?**
  - **Achievable in practice?**

- Existing work on AGC security
  - Simulations based on predefined attack templates
    scaling, ramps, surges, random noises, time delays
    [Bose 2004 2005] [Sridhar 2010 2014]
  - Reachability analysis [Esfahani 2010]

Outline

- Motivation & Background
- Attack Model
- **Optimal Attack**
  - **How to compute?**
  - Achievable?
- Simulations & Testbed Experiments

Attack Impact Model

In Laplace domain: $$\Delta \omega = \mathbf{\Phi} \cdot \Delta \mathbf{p} + \mathbf{\Phi \Lambda T} \cdot \mathbf{a}$$

Freq. deviation

Vector of load changes of areas

Attack vector

- **T**: constant integer matrix
  - From grid topology
  - Can be obtained by attacker (social engineering)
- **Φ**, **Λ**
  - Transfer functions of turbines, generators, transmission lines, etc
  - Closed-forms unknown

Optimal Attack

- Time-to-emergency (TTE): remaining time before

$$\Delta\omega \notin (\Delta\omega_{\min}, \Delta\omega_{\max})$$

  - $\Delta\omega_{min}$ = −0.5 Hz: load shedding (regional blackout)

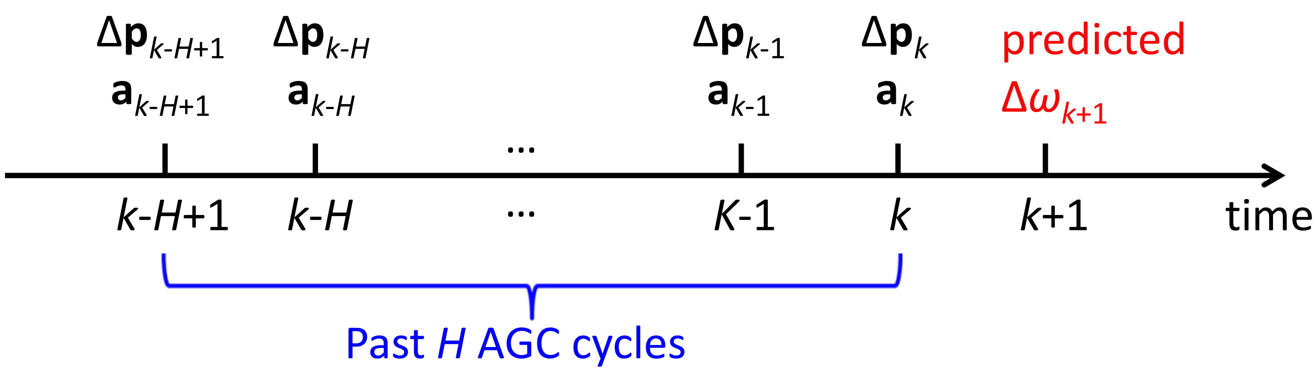
- Compute a series of ***a*** to minimize TTE subject to
  - Write access
  - Stealthiness

- $\boxed{\Delta\omega = \mathbf{\Phi} \cdot \Delta\mathbf{p} + \mathbf{\Phi\Lambda T} \cdot \mathbf{a}}$ complex differential eqns
  - Exhaustive search with prohibitive complexity

Regression

Laplace domain: $\Delta\omega = \mathbf{\Phi}\cdot\Delta\mathbf{p} + \mathbf{\Phi\Lambda T}\cdot\mathbf{a}$

Time domain: $\Delta\omega_{k+1} = \sum_{i=0}^{H-1} \mathbf{u}_i \cdot \Delta\mathbf{p}_{k-i} + \mathbf{v}_i \cdot \mathbf{a}_{k-i}$



Regression

Laplace domain: $\Delta \omega = \mathbf{\Phi} \cdot \Delta \mathbf{p} + \mathbf{\Phi \Lambda T} \cdot \mathbf{a}$

Time domain: $\Delta \omega_{k+1} = \sum_{i=0}^{H-1} \mathbf{u}_i \cdot \Delta \mathbf{p}_{k-i} + \mathbf{v}_i \cdot \mathbf{a}_{k-i}$

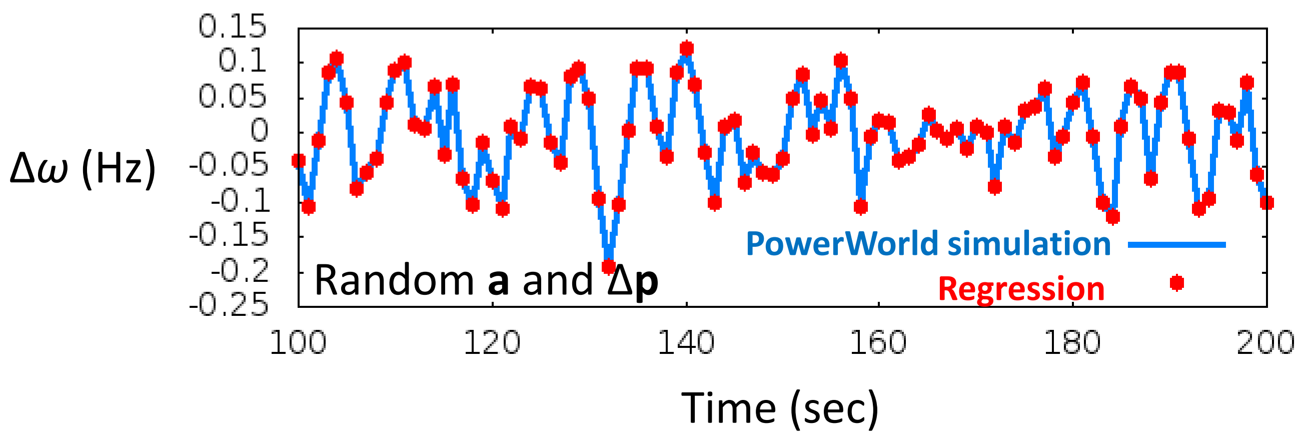
- Coefficients **u** and **v**
  - Trained using data generated by Laplace-domain model

Regression

Laplace domain: $\Delta\omega = \mathbf{\Phi} \cdot \Delta\mathbf{p} + \mathbf{\Phi\Lambda T} \cdot \mathbf{a}$

Time domain: $\Delta\omega_{k+1} = \sum_{i=0}^{H-1} \mathbf{u}_i \cdot \Delta\mathbf{p}_{k-i} + \mathbf{v}_i \cdot \mathbf{a}_{k-i}$

- Coefficients **u** and **v**
  - Trained using data generated by Laplace-domain model



Optimal Attack Algorithm

At prediction horizon *h*:

Eavesdropped past load changes

$$\Delta\omega_{k+h} = \begin{bmatrix} \mathbf{u}_{H-1} \\ \vdots \\ \mathbf{u}_{h} \\ \mathbf{u}_{h-1} \\ \vdots \\ \mathbf{u}_{0} \end{bmatrix}^{\mathrm{T}} \cdot \begin{bmatrix} \Delta\mathbf{p}_{k-H+h+1} \\ \vdots \\ \Delta\mathbf{p}_{k} \\ \Delta\hat{\mathbf{p}}_{k+1} \\ \vdots \\ \Delta\hat{\mathbf{p}}_{k+h} \end{bmatrix} + \begin{bmatrix} \mathbf{v}_{H-1} \\ \vdots \\ \mathbf{v}_{h} \\ \mathbf{v}_{h-1} \\ \vdots \\ \mathbf{v}_{0} \end{bmatrix}^{\mathrm{T}} \cdot \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{a}_{k+1} \\ \vdots \\ \mathbf{a}_{k+h} \end{bmatrix}$$

Attack vectors to be scheduled

Predicted future load changes

- Increasing *h* from 1, minimize and maximize $\Delta\omega_{k+h}$ until

$$\Delta\omega_{k+h} \notin (\Delta\omega_{\min}, \Delta\omega_{\max})$$

  - Optimal, modulo approx err of regression
  - Linear programming

Outline

- Motivation & Background
- Attack Model
- **Optimal Attack**
  - How to compute?
  - **Achievable?**
    How to learn attack impact model?
    What prior information needed?
- Simulations & Testbed Experiments

A Baseline Approach

$$\Delta \omega_{k+1} = \sum_{i=0}^{H-1} \mathbf{u}_i \cdot \Delta \mathbf{p}_{k-i} + \mathbf{v}_i \cdot \mathbf{a}_{k-i}$$

- Inject small attack vectors to collect training data to learn the coefficients
- Less stealthy

Passive Monitoring

$$\Delta\omega = \mathbf{\Phi} \cdot \Delta\mathbf{p} + \mathbf{\Phi\Lambda T} \cdot \mathbf{a}$$

- Learn **Φ**, **Λ** from eavesdropped measurements

Passive Monitoring

$$\Delta \omega = \boxed{\mathbf{\Phi}} \cdot \Delta \mathbf{p} + \mathbf{\Phi \Lambda T} \cdot \mathbf{a}$$

- Learn **Φ**, **Λ** from eavesdropped measurements
  - **Φ**: from $\Delta\omega$ and $\Delta\mathbf{p}$

$$\Delta \omega = \mathbf{\Phi} \cdot \Delta \mathbf{p}$$

Passive Monitoring

$$\Delta \omega = \mathbf{\Phi} \cdot \Delta \mathbf{p} + \mathbf{\Phi}\boxed{\mathbf{\Lambda}\mathbf{T}} \cdot \mathbf{a}$$

- Learn **Φ**, **Λ** from eavesdropped measurements
  - **Φ**: from $\Delta\omega$ and $\Delta\mathbf{p}$

$$\Delta \omega = \mathbf{\Phi} \cdot \Delta \mathbf{p}$$

  - **Λ**: no training data

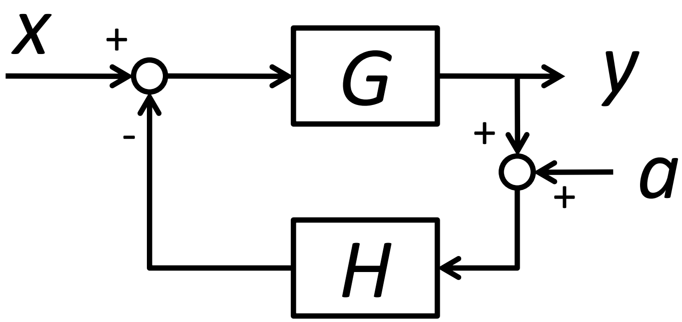
Passive Monitoring

$$\Delta\omega = \mathbf{\Phi}\cdot\Delta\mathbf{p} + \mathbf{\Phi}\boxed{\mathbf{\Lambda T}}\cdot\mathbf{a}$$

- Learn **Φ**, **Λ** from eavesdropped measurements
  - **Φ**: from Δω and Δ**p**

$$\Delta\omega = \mathbf{\Phi}\cdot\Delta\mathbf{p}$$

  - **Λ**: no training data



$$y = \frac{G}{1+GH}x - \frac{GH}{1+GH}a$$

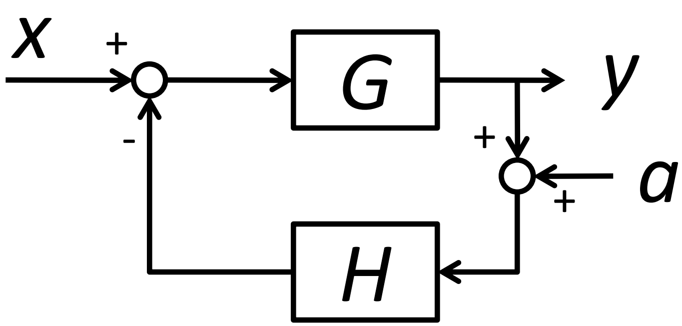
Passive Monitoring

$$\Delta\omega = \mathbf{\Phi}\cdot\Delta\mathbf{p} + \mathbf{\Phi}\boxed{\mathbf{\Lambda T}}\cdot\mathbf{a}$$

- Learn **Φ**, **Λ** from eavesdropped measurements
  - **Φ**: from Δω and Δ**p**

$$\Delta\omega = \mathbf{\Phi}\cdot\Delta\mathbf{p}$$

  - **Λ**: no training data



$$y = \boxed{\frac{G}{1+GH}}x - \frac{GH}{1+GH}a$$

Estimated as a whole,
but not individual G and H

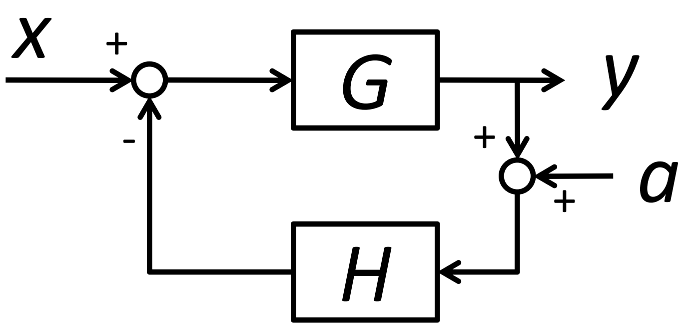
Passive Monitoring

$$\Delta\omega = \mathbf{\Phi}\cdot\Delta\mathbf{p} + \mathbf{\Phi}\boxed{\mathbf{\Lambda T}}\cdot\mathbf{a}$$

- Learn **Φ**, **Λ** from eavesdropped measurements
  - **Φ**: from Δω and Δ**p**

$$\Delta\omega = \mathbf{\Phi}\cdot\Delta\mathbf{p}$$

  - **Λ**: no training data



$$y = \boxed{\frac{G}{1+GH}}x - \boxed{\frac{GH}{1+GH}}a$$

Estimated as a whole,
but not individual G and H

?

Passive Monitoring

$$\Delta \omega = \mathbf{\Phi} \cdot \Delta \mathbf{p} + \mathbf{\Phi} \boxed{\mathbf{\Lambda T}} \cdot \mathbf{a}$$

- Learn **Φ**, **Λ** from eavesdropped measurements
  - **Φ**: from $\Delta\omega$ and $\Delta\mathbf{p}$

$$\Delta \omega = \mathbf{\Phi} \cdot \Delta \mathbf{p}$$

  - **Λ**: from $\Delta\omega$, $\mathbf{z}$ in normal state and 4 system constants:
    - ACE weight parameters α and β: public
    - Load damping constant and total inertia of generators

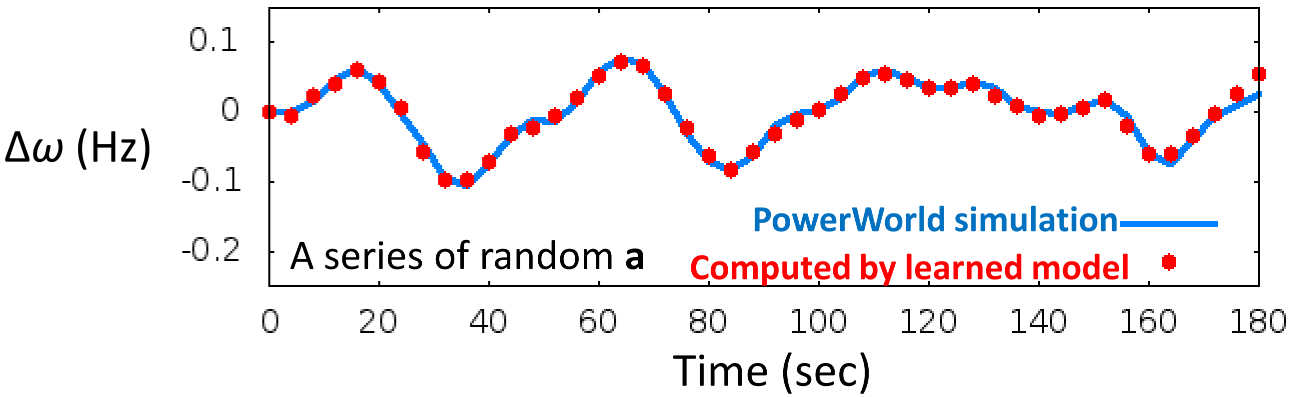
Passive Monitoring

$$\Delta\omega = \boldsymbol{\Phi}\cdot\Delta\mathbf{p} + \boldsymbol{\Phi}\boxed{\boldsymbol{\Lambda}\mathbf{T}}\cdot\mathbf{a}$$

- Learn **Φ**, **Λ** from eavesdropped measurements
  - **Φ**: from $\Delta\omega$ and $\Delta\mathbf{p}$

$$\Delta\omega = \boldsymbol{\Phi}\cdot\Delta\mathbf{p}$$

  - **Λ**: from $\Delta\omega$, **z** in normal state and 4 system constants:
    - ACE weight parameters α and β: public
    - Load damping constant and total inertia of generators

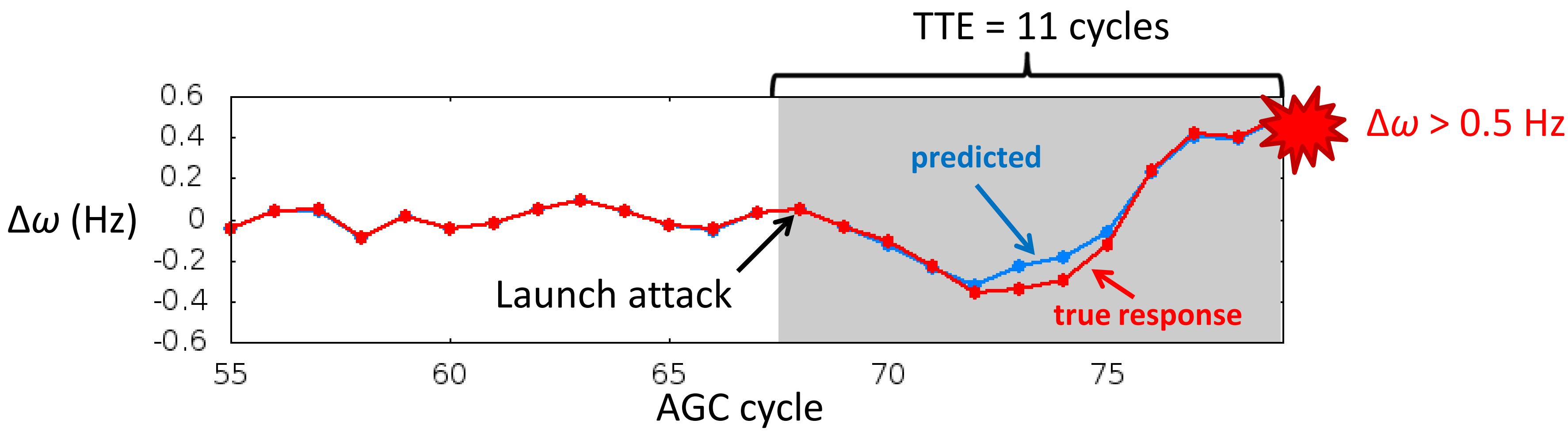


Outline

- Motivation and Background
- Attack Model
- Optimal Attack
- **Simulations & Testbed Experiments**

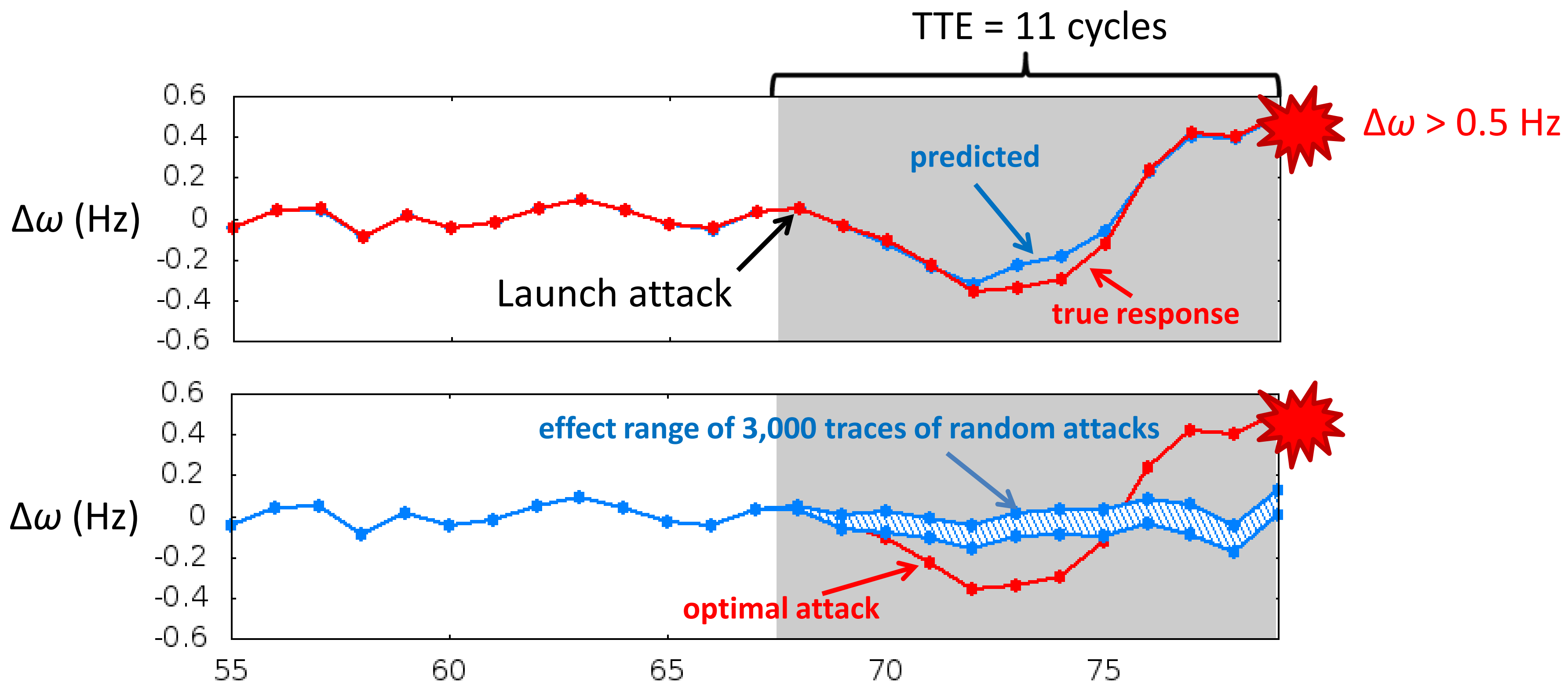
PowerWorld Simulations

37-bus 3-area grid, 81 sensors (all compromised), frequency safety range: (-0.5, +0.5) Hz



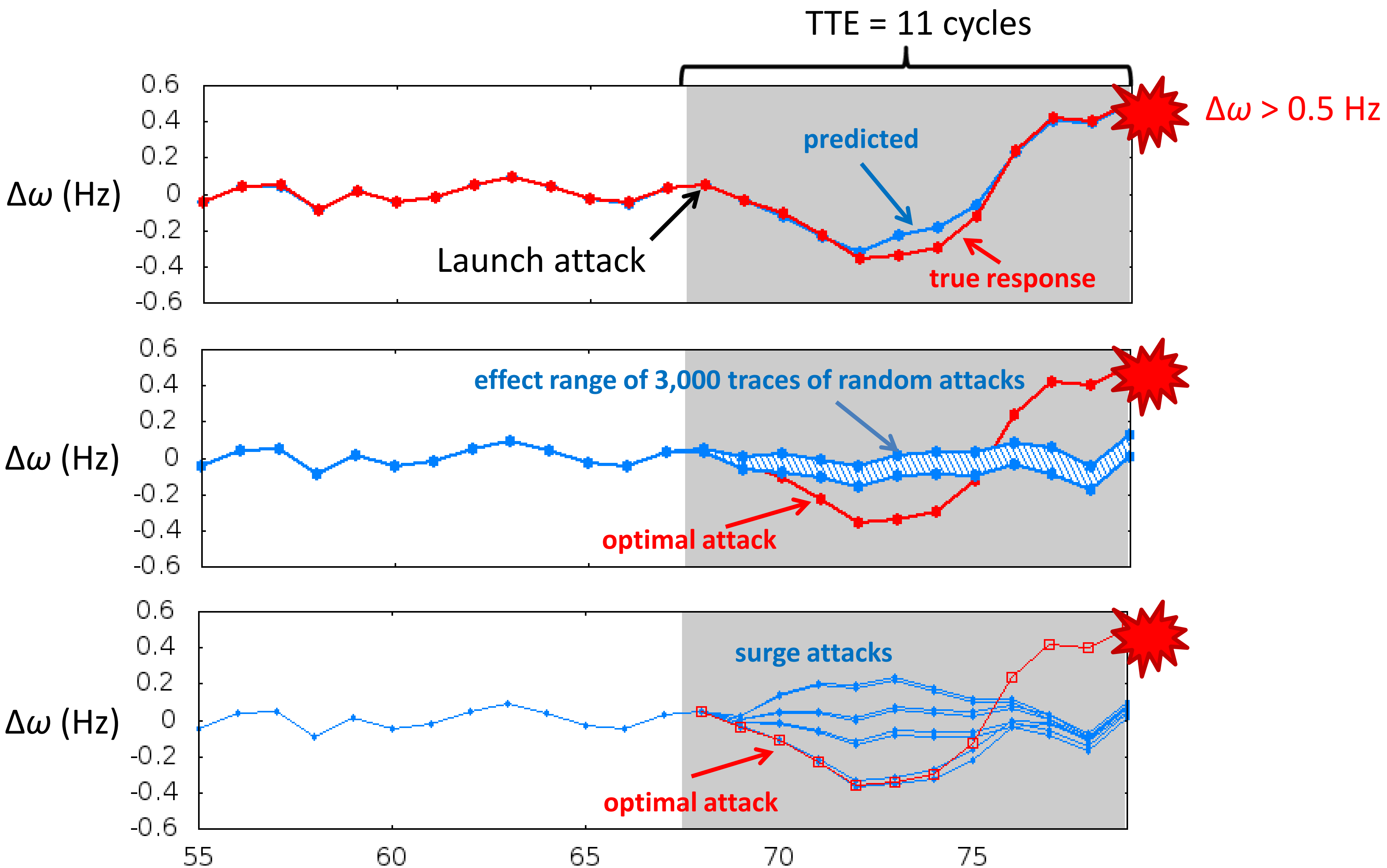
PowerWorld Simulations

37-bus 3-area grid, 81 sensors (all compromised), frequency safety range: (-0.5, +0.5) Hz



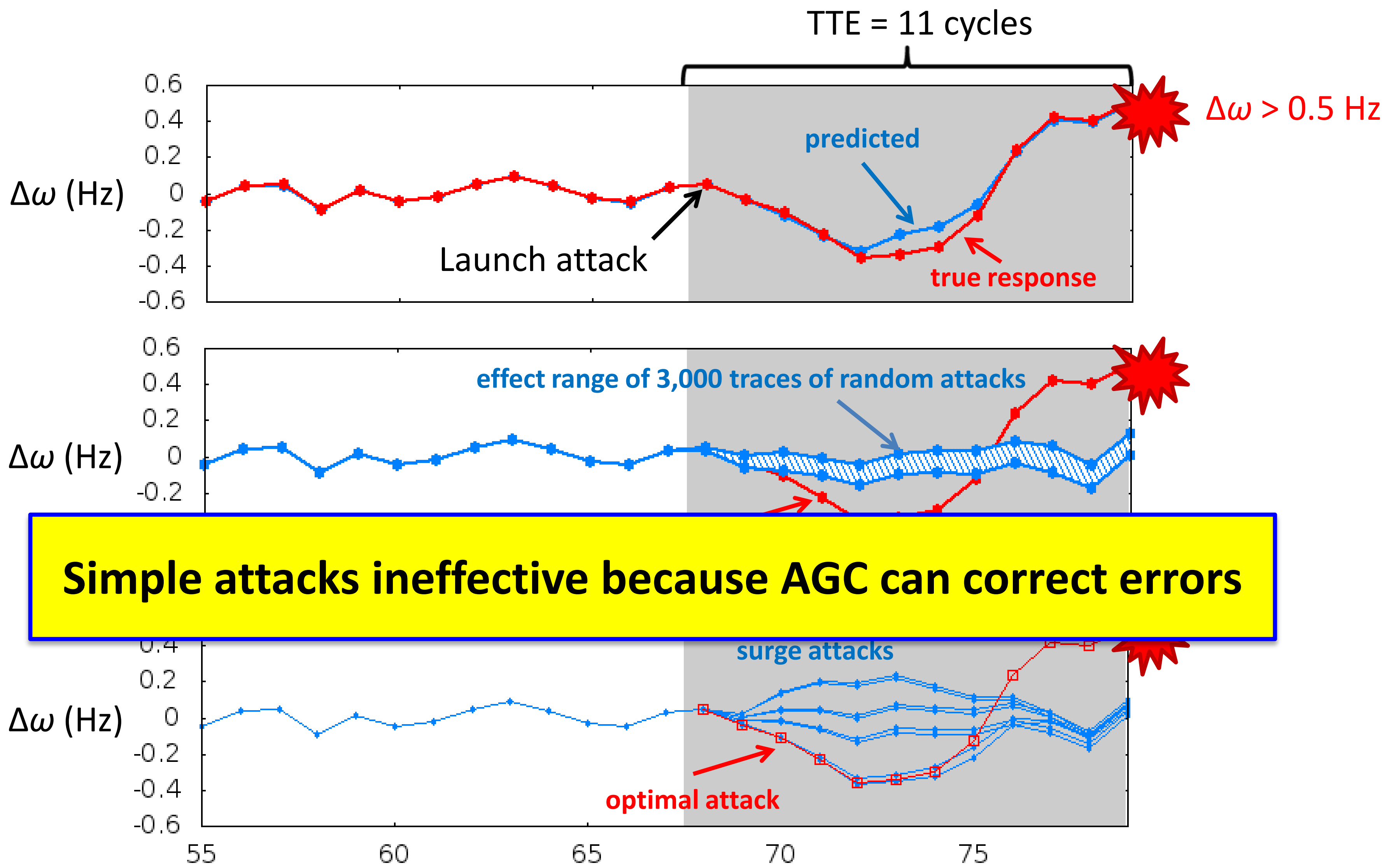
PowerWorld Simulations

37-bus 3-area grid, 81 sensors (all compromised), frequency safety range: (-0.5, +0.5) Hz



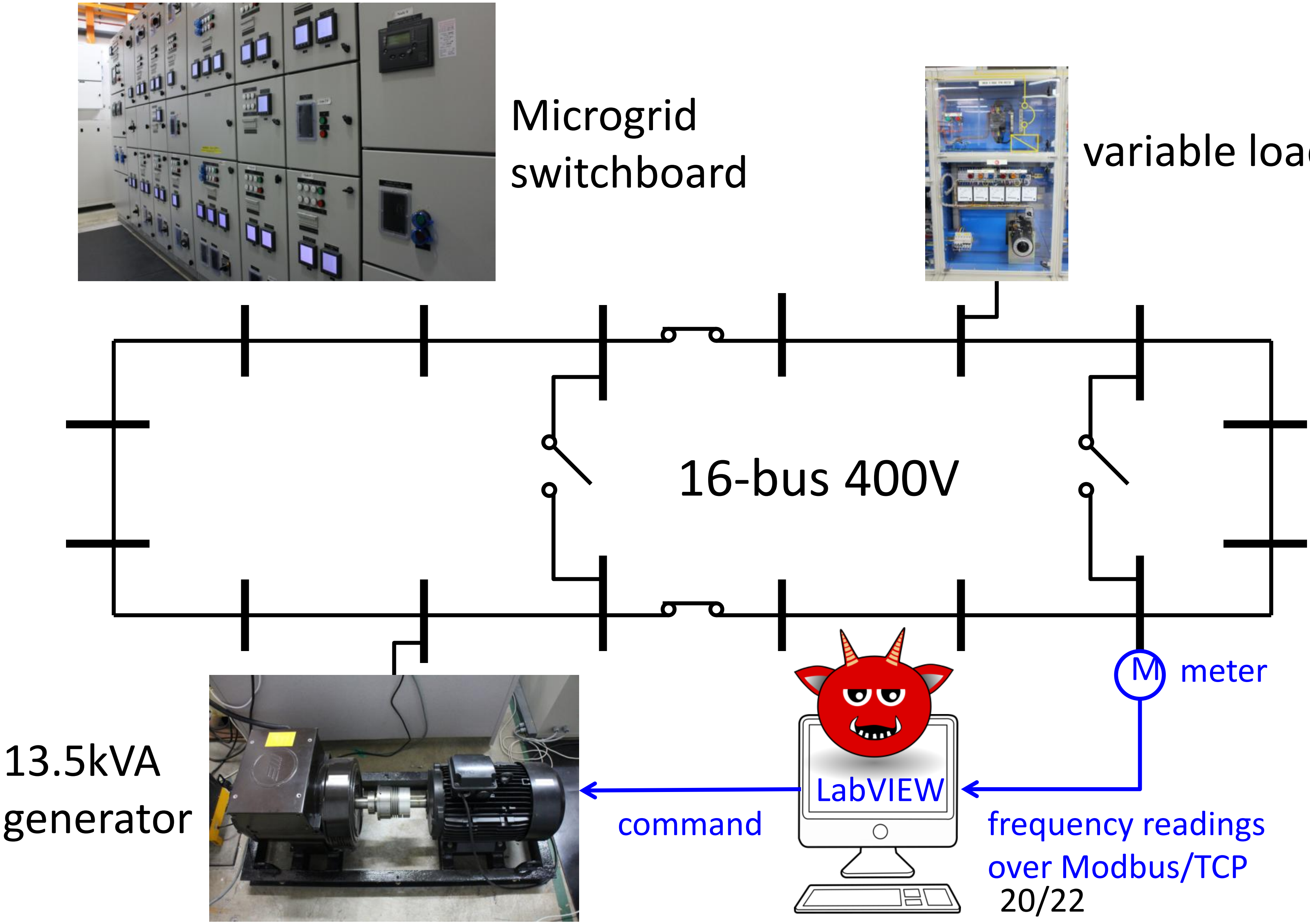
PowerWorld Simulations

37-bus 3-area grid, 81 sensors (all compromised), frequency safety range: (-0.5, +0.5) Hz



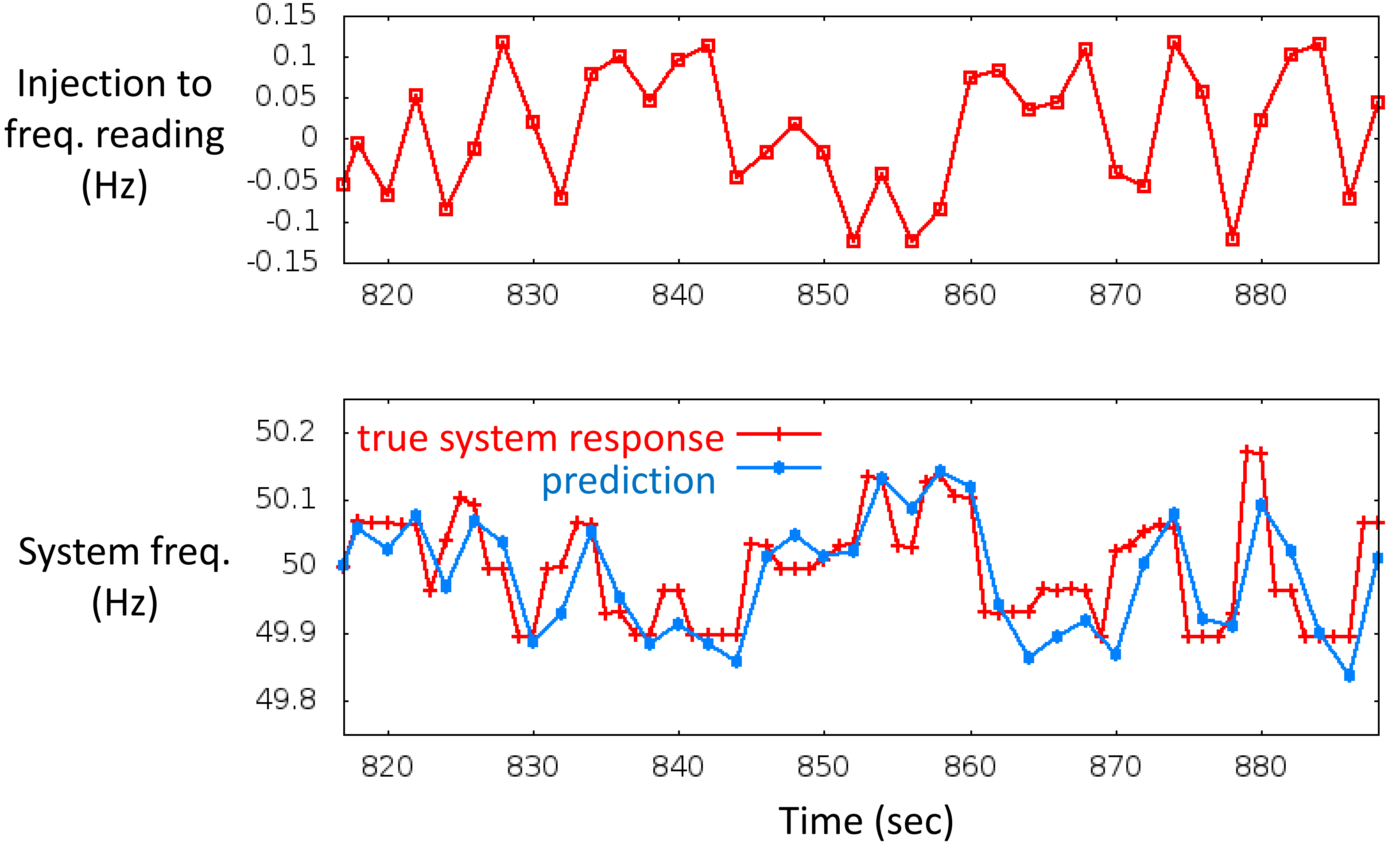
Simple attacks ineffective because AGC can correct errors

Microgrid Testbed



Accuracy of Learned Attack Model

A random bounded injection sequence to avoid damage to testbed



From learned model, 30 seconds to achieve 0.5 Hz deviation

Conclusion

- **FDI attack again AGC**
  - Attack impact model
  - Learned using data in normal state
  - Minimize time-to-emergency

- **Evaluation**
  - PowerWorld simulations
  - Experiments on a real power system

- **Ongoing work**
  - Attack detection, identification (which measurements compromised?), mitigation