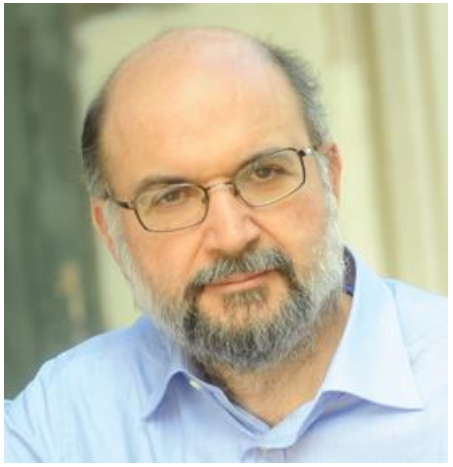
PRINCETON
UNIVERSITY

REACT to Cyber-Physical Attacks on the Power Grid

Saleh Soltan
Department of Electrical Engineering
Princeton University

Collaborators



Mihalis Yannakakis
Department of Computer Science
Columbia University

Gil Zussman
Department of Electrical Engineering
Columbia University

Saleh Soltan, Mihalis Yannakakis, Gil Zussman, "REACT to Cyber Attacks on Power Grids," to appear on IEEE Transactions on Network Science and Engineering, 2018.

Infrastructure Networks

- Almost all infrastructure networks are monitored and controlled by Supervisory Control And Data Acquisition Systems (SCADA)
- The physical components of these networks along with their control network form a *cyber-physical system*
- Due to their direct control of the infrastructure networks, SCADA systems have been the main targets of cyber attacks (e.g., Stuxnet virus)

New York State ISO Control Room

Attacks and Failures in Power Systems

Physical Attacks

Power Grid
Physical Infrastructure

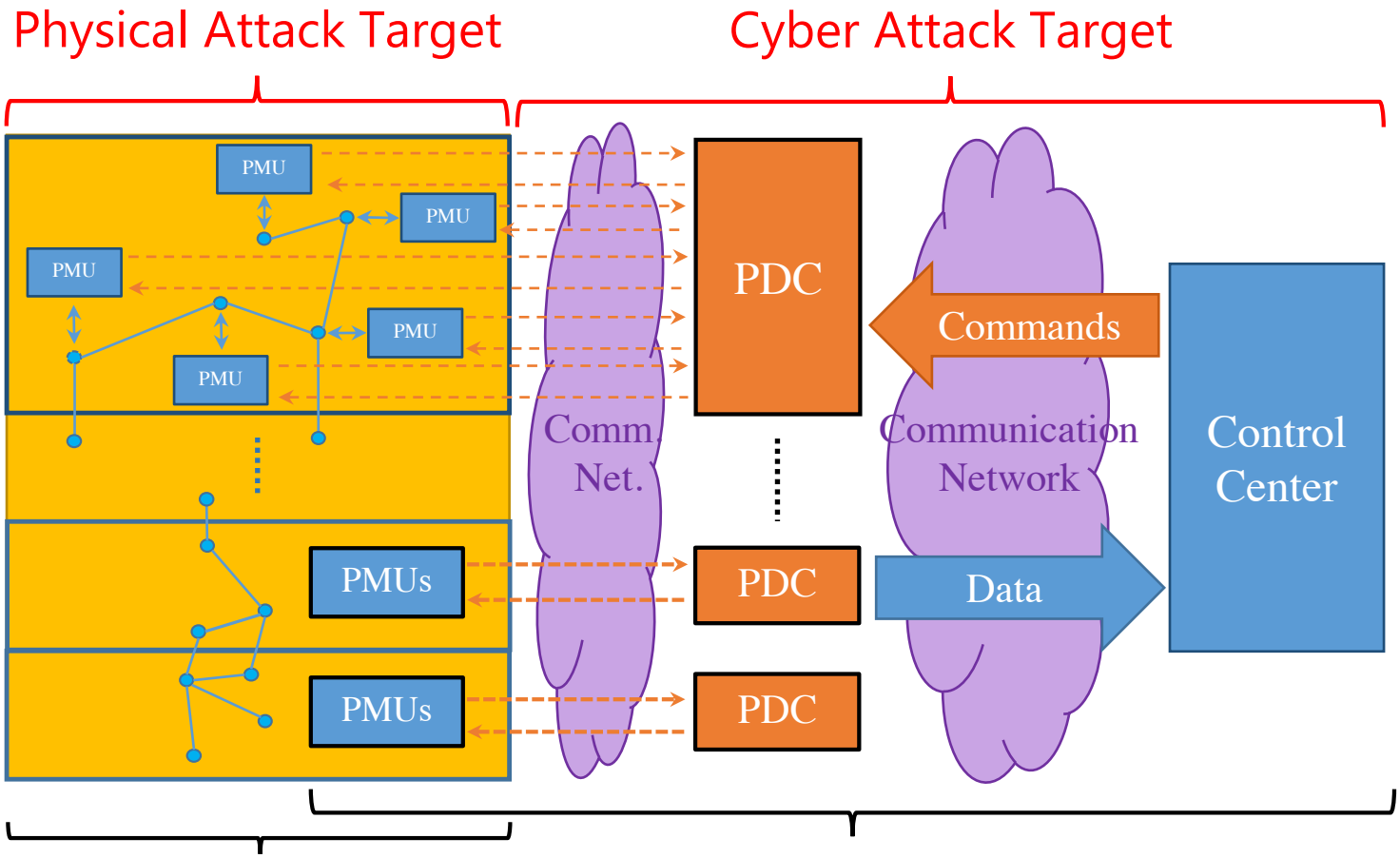
Commands

Data

Cyber Attacks

Supervisory Control and Data Acquisition (SCADA) system

Components of Power Grid SCADA



Power Grid
Physical Infrastructure

Supervisory Control and Data Acquisition (SCADA) system

PMU: Phasor Measurement Unit
PDC: Phasor Data Concentrators

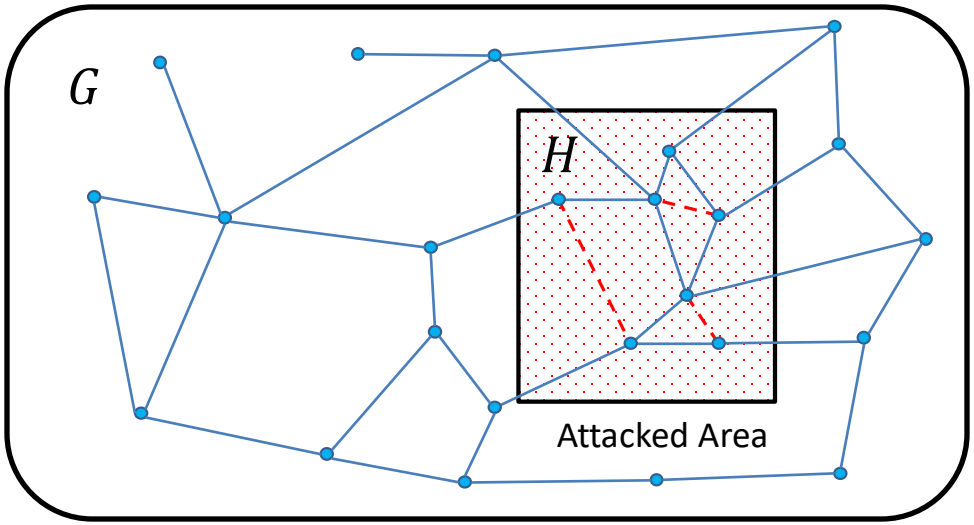
IED: Intelligent Electronic Devices
DC: Data Concentrators

Cyber Attack on the Ukrainian Grid

- Unplugged 225,000 people from the Ukrainian electricity grid in December 2015
  - Steal credentials for accessing the SCADA system, *before June 2015*
  - Explore of SCADA system and attack planning, *June-Dec. 2015*
  - Remotely operate circuit breakers, *day of attack*
  - Phone jamming attacks keeps operators unaware, *day of attack*
- "An attacker can simply replay, modify, and spoof the traffic to SCADA devices"

Attack Model

- An adversary attacks the grid by
  - Manipulating the measurements (cyber)
    - **Block** the measurements
    - **Falsify** the measurements (false data injection)
  - Disconnecting lines within the attacked area (physical)
- **Goal:** Efficiently detect the attacked area and the disconnected lines to avoid further failures



Saleh Soltan, Mihalis Yannakakis, Gil Zussman, "Joint Cyber and Physical Attacks on Power Grids: Graph Theoretical Approaches for Information Recovery," in Proc. ACM SIGMETRICS'15, June 2015.

AC Power Flows

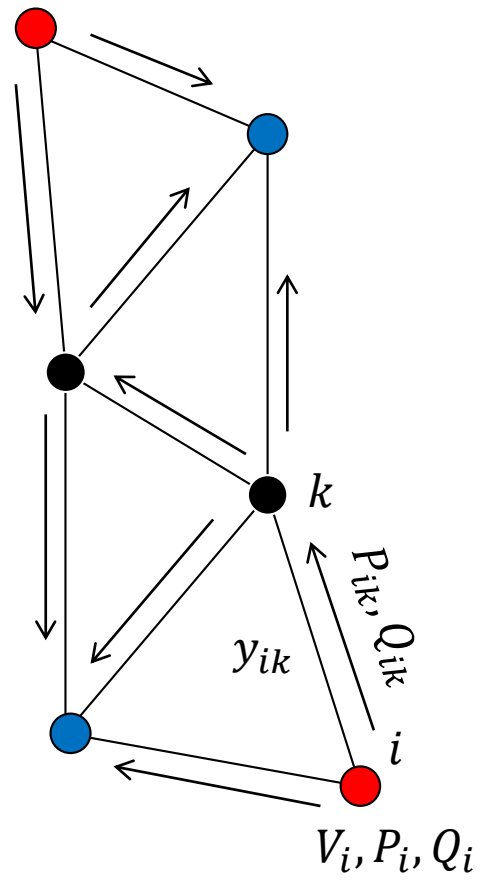
- Present the grid by a connected graph $G = (N, E)$
- In the phasor domain
- $V_i = |V_i| e^{\mathbf{i}\theta_i}$
  $|V_i|$ is the Voltage magnitude
  θ_i is the phase angle
- Transmission line (i, k) is characterized by series admittance $y_{ik} = g_{ik} + \mathbf{i} b_{ik}$
- The active and reactive power flows:

$$P_{ik} = |V_i|^2 g_{ik} - |V_i||V_k| g_{ik} \cos\theta_{ik} - |V_i||V_k| b_{ik} \sin\theta_{ik}$$

$$Q_{ik} = -|V_i|^2 b_{ik} + |V_i||V_k| b_{ik} \cos\theta_{ik} - |V_i||V_k| g_{ik} \sin\theta_{ik}$$

and $\theta_{ik} = \theta_i - \theta_k$

- Active and reactive power at node i:
  $P_i = \sum P_{ik}$, $Q_i = \sum Q_{ik}$
- Given a subset of P, Q, V values, compute the rest → nonlinear and not unique



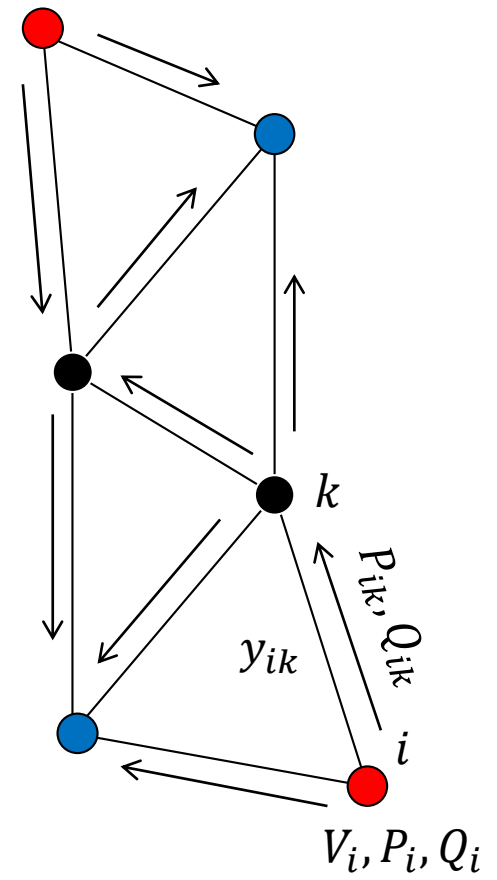
Power Flows - DC Approximation

- In the stable state of the system
  - $|V_i| \approx 1\ p.u.$ for all i
  - $\left|\frac{g_{ik}}{b_{ik}}\right| \ll 1$ for all lines $\Rightarrow y_{ik} \approx \boldsymbol{i} b_{ik}$
  - $\theta_{ik} \ll 1 \Rightarrow \cos(\theta_{ik}) \approx 1$ and $\sin(\theta_{ik}) \approx \theta_{ik}$

- The power flow equations reduce to

$$f_{ik} := P_{ik} = -b_{ik}(\theta_i - \theta_k)$$
$$\sum_k P_{ik} = P_i$$

- The DC power flows only considers active powers



DC Power Flows (Matrix Form)

- The DC power flow can be written in matrix form:

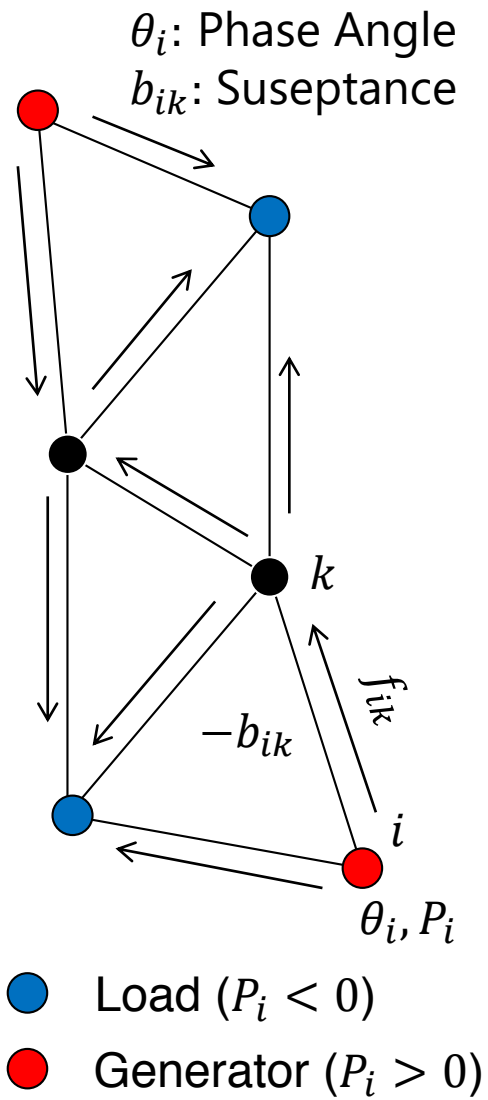
$$YD^T\,\vec{\theta} = \vec{f}$$
$$A\vec{\theta} = \vec{p}$$

$D \in \{-1,0,1\}^{n\times m}$: the **incidence matrix** of the grid:

$$d_{ij} = \begin{cases} 0, & \text{if } e_j \text{ is not incident to node } i, \\ 1, & \text{if } e_j \text{ is coming out of node } i, \\ -1, & \text{if } e_j \text{ is going into of node } i, \end{cases}$$

$Y \in \mathbb{R}^{m\times m}$: the diagonal matrix of suseptance values,

and $A = DYD^T$: the **admittance matrix** of the grid



Assumptions and Objective

- Assume that the phase angles $\vec{\theta}$ are measured directly at all the nodes
- Correct phase angles after the attack: $\vec{\theta}' = \begin{bmatrix} \vec{\theta}'_H \\ \vec{\theta}'_{\bar{H}} \end{bmatrix}$
- Measured phase angles after the attack: $\vec{\theta}^* = \begin{bmatrix} \vec{\theta}^*_H \\ \vec{\theta}^*_{\bar{H}} \end{bmatrix}$
  - $\vec{\theta}^*_{\bar{H}} = \vec{\theta}'_{\bar{H}}$

Objective: Use the measurements after the attack ($\vec{\theta}^*$) and the information before attack ($A, \vec{\theta}$) to:

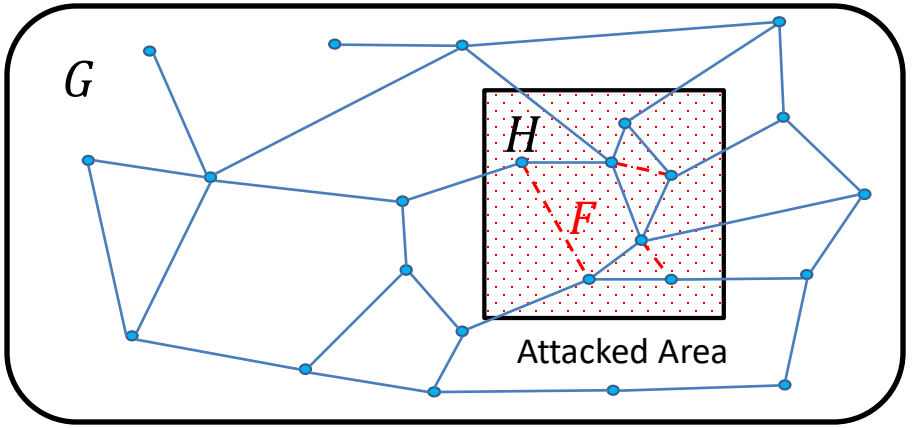
- *Detect the attack area (H)*
- *Detect the disconnected lines (F)*

H : an induced subgraph of G that represents the attacked area

$\bar{H}$: $G \backslash H$

F : Set of failed lines

O' : The value of O after an attack



False Data Injection

- Assume two types of *data attacks*:
  - ***Data distortion:*** the attacker adds large noise to the measurements coming from the attacked area:

$$\vec{\theta}_H^* = \vec{\theta}_H' + \vec{z}$$

  - ***Data replay:*** the attacker replays measurements from previous hours/days instead of the actual measurements coming from the attacked area:

$$\vec{\theta}_H^* = \vec{\theta}_H''$$

    in which $A\vec{\theta}'' = \vec{p}''$ and $\vec{p}_H'' = \vec{p}_H$.

- Measurements remain locally consistent after a *replay attack*

Outline

- Hardness
- Attacked Area Approximation
  - Data distortion
  - Data replay
  - ATtacked Area Containment (ATAC) module
- Line Failures Detection
- REcurrent Attack Containment and deTection (REACT) Algorithm
- Numerical Results

Hardness

Lemma. Given A, $\vec{\theta}$ and $\vec{\theta}'$, it is strongly NP-hard to determine if there exists a set of line failures F such that:

$$A^{(F)}\vec{\theta}' = A\vec{\theta}$$

- Reduction from 3-partition problem

Lemma. Given A, $\vec{\theta}$, H and $\vec{\theta}'_{\bar{H}}$, it is strongly NP-hard to determine if there exists a set of line failures F in H and a vector $\vec{\theta}'_H$ such that

$$A^{(F)}\begin{bmatrix}\vec{\theta}'_H \\ \vec{\theta}'_{\bar{H}}\end{bmatrix} = A\vec{\theta}$$

Lemma. Given A, $\vec{\theta}$ and $\vec{\theta}^*$, it is strongly NP-hard to determine if there exists a subgraph H_0 with $|V_{H_0}| \leq |V|/2$, set of line failures F in H_0, and a vector $\vec{\theta}'_{H_0}$ such that

$$A^{(F)}\begin{bmatrix}\vec{\theta}'_{H_0} \\ \vec{\theta}^*_{\overline{H_0}}\end{bmatrix} = A\vec{\theta}$$

Attacked Area Approximation

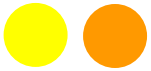
Data Distortion Attack

- For any $i \in int(\bar{H}), A_i\vec{\theta}^* = p_i$
- For any $i \in V \backslash int(\bar{H}), A_i\vec{\theta}^* \neq p_i$

$$int(\bar{H}) = V \backslash \text{supp}(A\vec{\theta}^* - \vec{p})$$

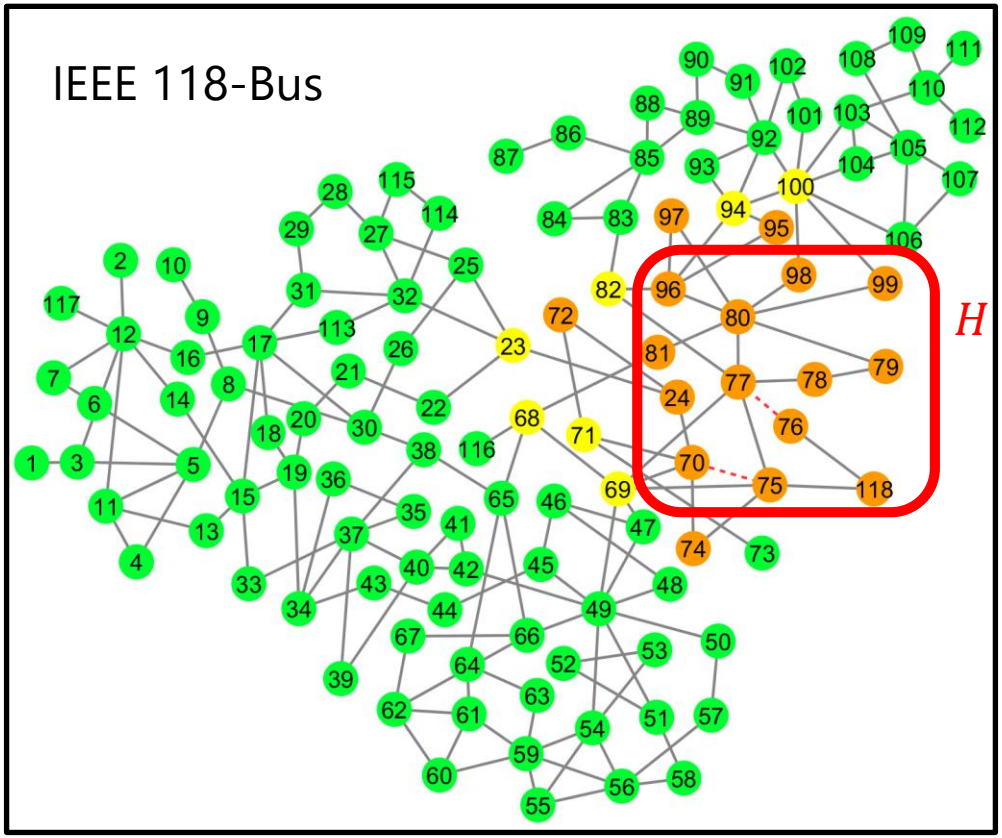
- $S_0 \coloneqq G[\text{supp}(A\vec{\theta}^* - \vec{p})]$
- $V_H \subseteq int(S_0)$

$int(\bar{H})$→ 

S_0 → 

$int(S_0)$→ 

$int(S)$ ≔ nodes in S such that their neighbors are also in S



Data Replay Attack

- Detecting the attacked area is more challenging
- For any $i \in int(\bar{H}) \cup int(H), A_i\vec{\theta}^* = p_i$
- For any $i \in \partial(\bar{H}) \cup \partial(H), A_i\vec{\theta}^* \neq p_i$

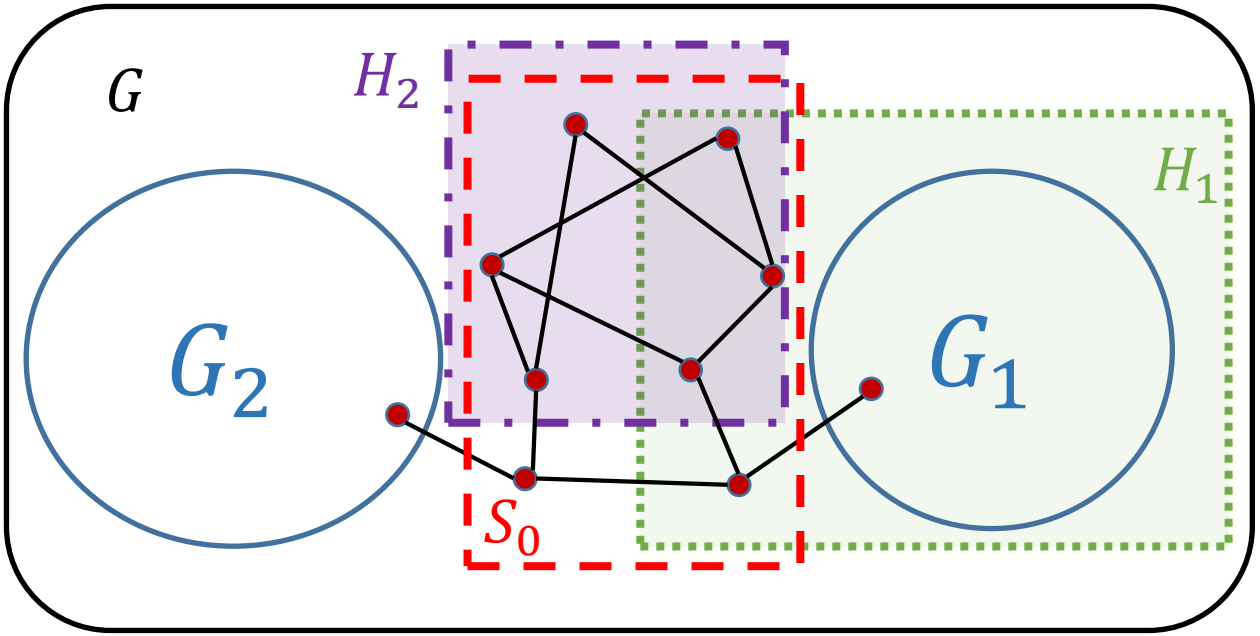
$$\text{supp}\left(A\vec{\theta}^* - \vec{p}\right) = \partial(H) \cup \partial(\bar{H})$$

- $S_0 \coloneqq \text{G}[\text{supp}\left(A\vec{\theta}^* - \vec{p}\right)]$ does <u>not</u> contain the attacked area in this case

Data replay attack

Data distortion attack

$int(S) \coloneqq$ nodes in S that their neighbors are also in S
$\partial(S) \coloneqq$ nodes in S that have neighbors also in $\bar{S}$



ATtacked Area Containment (ATAC)

- Provide multiple areas that may contain the attacked area

$int(\bar{H})$

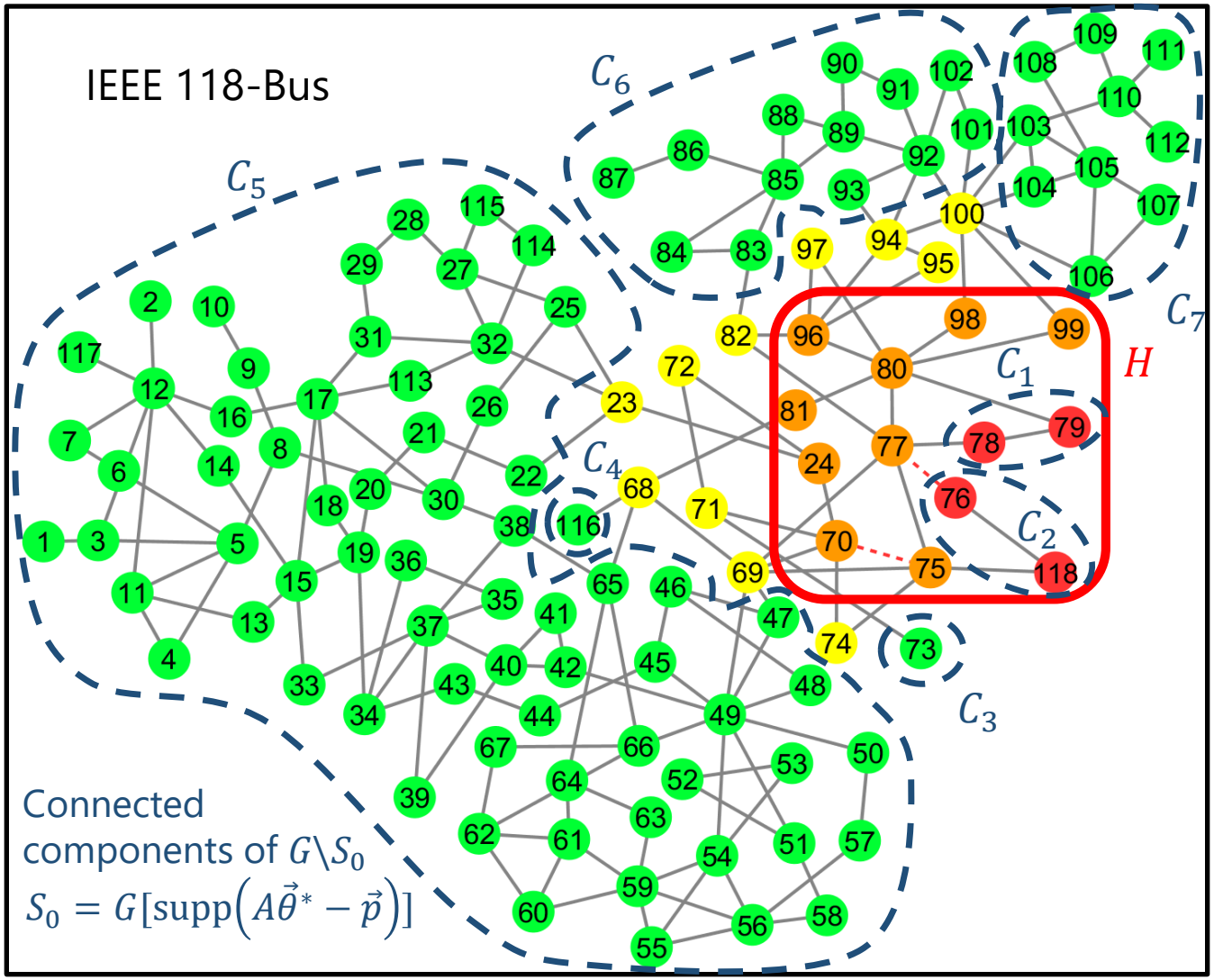
$int(H)$

$\partial(H)$

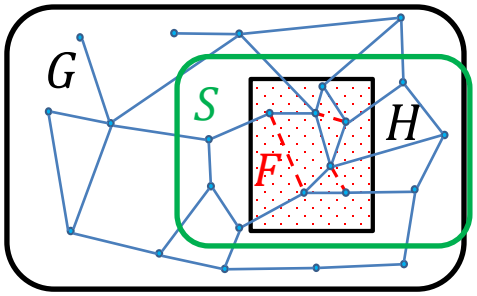
$\partial(\bar{H})$

$G_1 := C_1 \cup C_2$
$G_2 := C_4 \cup C_5$
$G_3 := C_3$
$G_4 := C_6 \cup C_7$

At least one of the $S_0, S_i := G \backslash G_i$ contain the attacked area



Line Failures Detection



$\bar{S}$: $G \backslash S$

O' : The value of O after an attack

O^* : The modified value of O after an attack

- Assume $S_0, S_1, \ldots, S_t$ are the subgraphs from ATAC
- Assume that S contains H → $\vec{\theta}^*_{\bar{S}} = \vec{\theta}'_{\bar{S}}$
- Brute force search algorithm

$$\min_{F,\vec{y}} ||A_{G|\bar{S}}\vec{\theta}^*_{\bar{S}} + A^{(F)}_{G|S}\vec{y} - \vec{p}||_2$$

- Not efficient → specially that we don't know if S contains the attacked area or not
- Solution $\vec{x}$ and $\vec{y}$ to the following linear program can detect the phase angles and line failures

Fewest number of line failures →

$$\min \| \vec{x} \|_1 \ \ s.t.$$

$$A_{S|G}\left(\vec{\theta} - \vec{\theta}'\right) = D_S\vec{x} \rightarrow A_{S|S}\left(\vec{\theta}_S - \vec{y}\right) + A_{S|\bar{S}}(\vec{\theta}_{\bar{S}} - \vec{\theta}^*_{\bar{S}}) = D_S\vec{x}$$

$$A_{\bar{S}|G}\left(\vec{\theta} - \vec{\theta}'\right) = 0 \rightarrow A_{\bar{S}|S}\left(\vec{\theta}_S - \vec{y}\right) + A_{\bar{S}|\bar{S}}(\vec{\theta}_{\bar{S}} - \vec{\theta}^*_{\bar{S}}) = 0$$

$(*)$

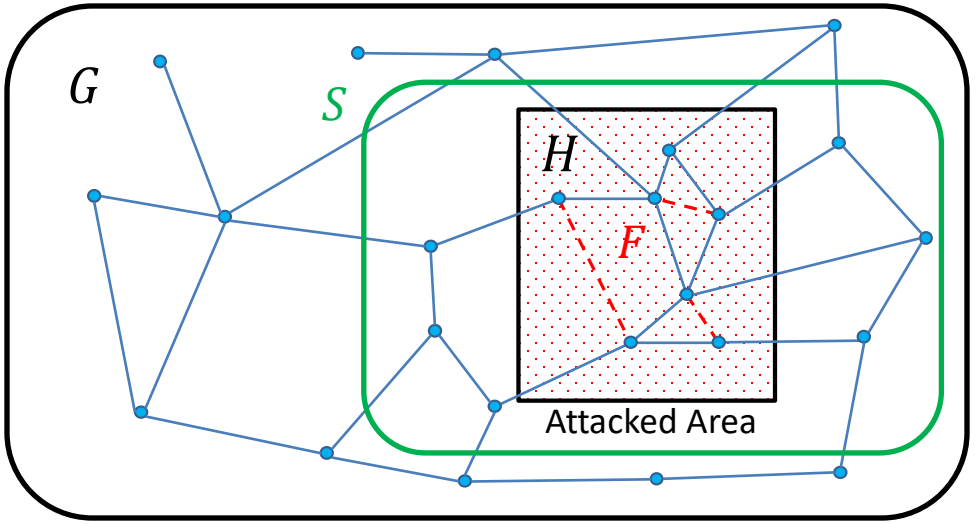
under some conditions, $\text{supp}(\vec{x}) = F$ and $\vec{y} = \vec{\theta}'_S$.

Saleh Soltan, Mihalis Yannakakis, Gil Zussman, "Joint Cyber and Physical Attacks on Power Grids: Graph Theoretical Approaches for Information Recovery," in Proc. ACM SIGMETRICS'15, June 2015.

Conditions and Limitations

External Conditions	Internal Conditions	Attack Constraints
Matching	Acyclic	None
Matching	Planar	Less than half of the edges in each cycle are failed
Partial Matching	Acyclic	Less than half of the edges connected to an internal node are failed
Partial Matching	Planar	Two above conditions

- Since at the time of a data replay attack, S might be much larger than H, in most of the cases S may not have the above conditions



Use Random Weights

- For a good diagonal matrix of random weights W, the solution to the following LP detects the line failures

$$\begin{aligned} &\min \| W\vec{x} \|_1 \ \ s.t. \\ &A_{S|S}\left(\vec{\theta}_S - \vec{y}\right) + A_{S|\bar{S}}(\vec{\theta}_{\bar{S}} - \vec{\theta}^*_{\bar{S}}) = D_S\vec{x} \\ &A_{\bar{S}|S}\left(\vec{\theta}_S - \vec{y}\right) + A_{\bar{S}|\bar{S}}(\vec{\theta}_{\bar{S}} - \vec{\theta}^*_{\bar{S}}) = 0 \end{aligned} \qquad (**)$$

- Confidence of the solution

$$c(F, \vec{y}) := \left(1 - ||A_{G|\bar{S}}\vec{\theta}^*_{\bar{S}} + A^{(F)}_{G|S}\vec{y} - \vec{p}||_2/||\vec{p}||_2\right) \times 100$$

- Generate random weights, solve $(**)$
  - check if for $F = \text{supp}(\vec{x})$ and $\vec{y}$, $||A_{G|\bar{S}}\vec{\theta}^*_{\bar{S}} + A^{(F)}_{G|S}\vec{y} - \vec{p}||_2$ is small enough
  - if not, regenerate W and solve $(**)$
- One can proves that in some cases, a good W can be obtained in *expected polynomial time* → details in the paper

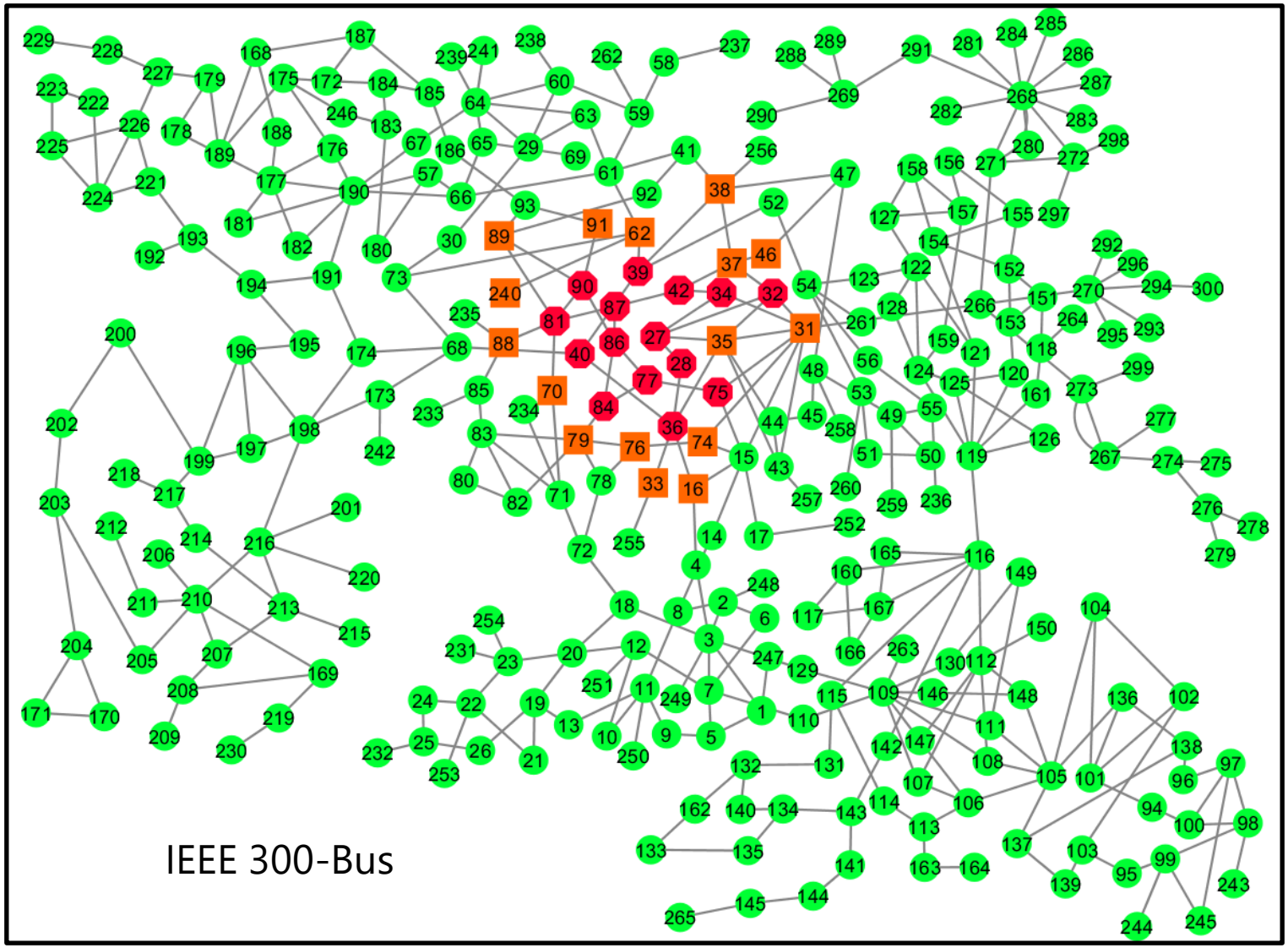
REACT Algorithm

- REcurrent Attack Containment and deTection (REACT)
  1. Obtain $S_0, S_1, \ldots, S_t$ using the ATAC module
  2. For each $i = 1$ to t, compute $S = G[int(S_i)]$
  3. If $(**)$ is not feasible go to the next i
  4. While $c(F, \vec{y}) < 99.9$ and $counter < T$
  5. Generate a random weight matrix W
  6. Solve $(**)$
  7. Return a solution with the highest confidence

$$\min \| W\vec{x} \|_1 \quad s.t.$$
$$A_{S|S}\left(\vec{\theta}_S - \vec{y}\right) + A_{S|\bar{S}}(\vec{\theta}_{\bar{S}} - \vec{\theta}^*_{\bar{S}}) = D_S\vec{x}$$
$$A_{\bar{S}|S}\left(\vec{\theta}_S - \vec{y}\right) + A_{\bar{S}|\bar{S}}(\vec{\theta}_{\bar{S}} - \vec{\theta}^*_{\bar{S}}) = 0 \qquad (**)$$

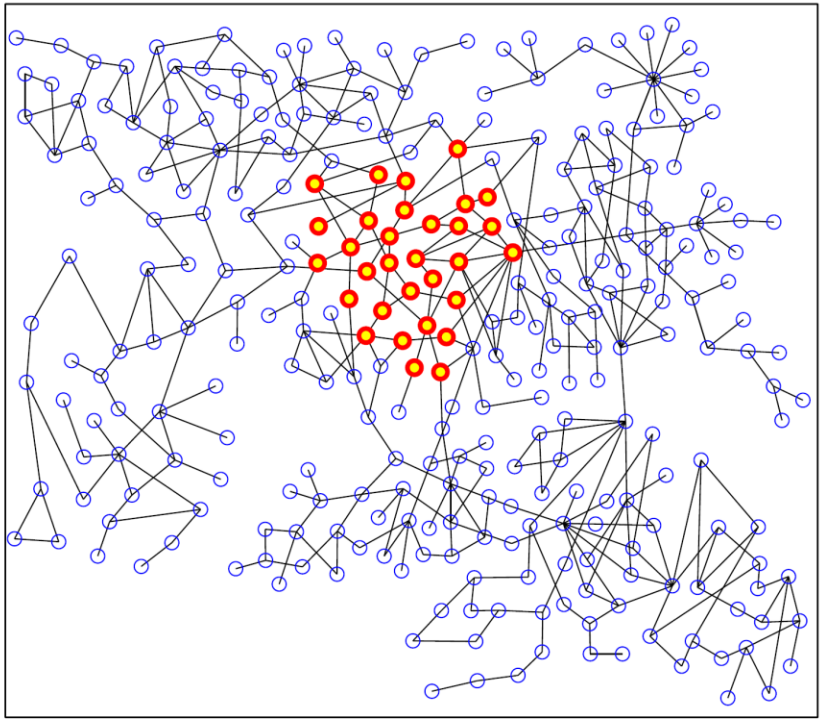
Numerical Results

- Two attacked areas: one with 31 nodes and the other with 15 nodes

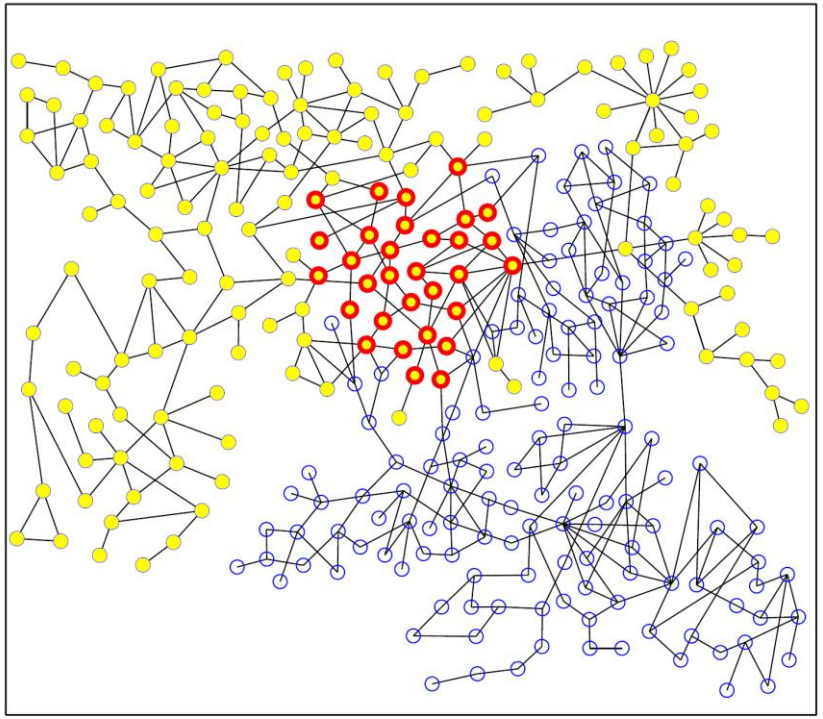


Data Distortion vs. Data Replay

- Difficulty in detecting the attacked area after a data replay attack



(a) Data Distortion Attack

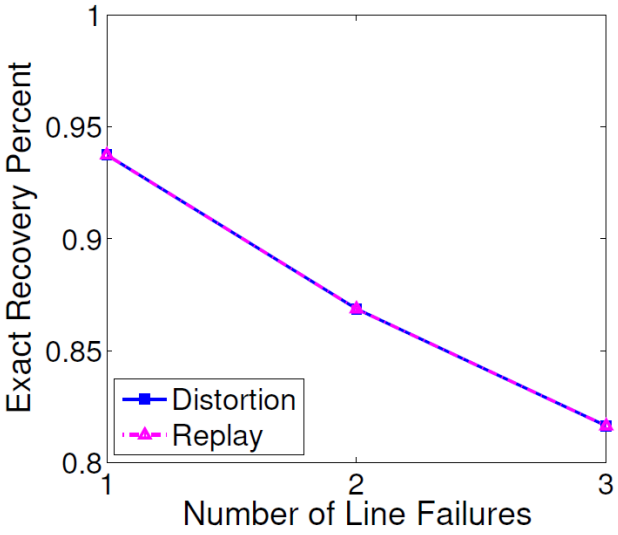


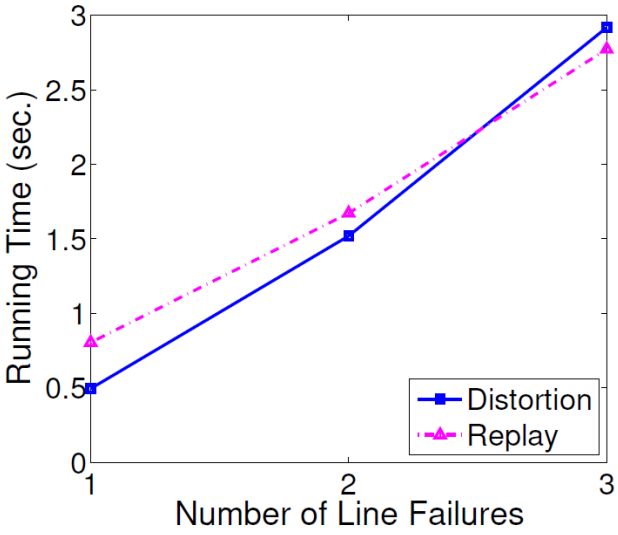
(b) Data Replay Attack

Data Distortion vs. Data Replay

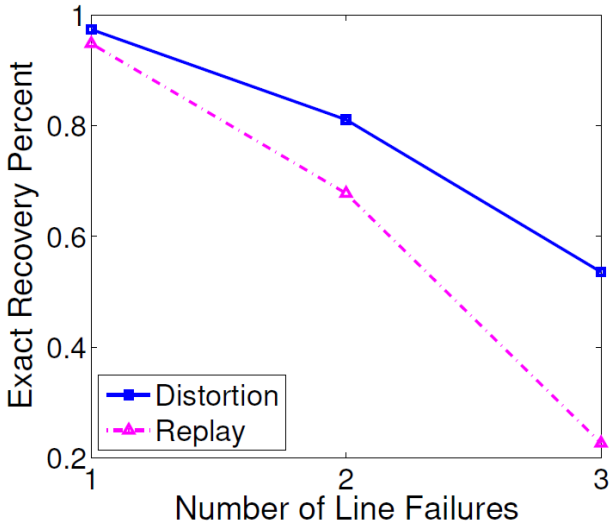
- $T = 20$

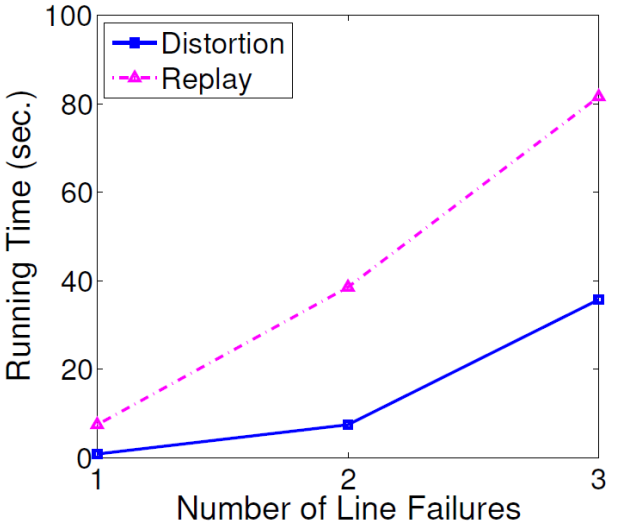
Smaller Attacked Area





Larger Attacked Area





Conclusions

- Modeled cyber-physical attacks on the power grid
- Studied hardness
- Showed that in general replay attacks (or more sophisticated data attacks) are harder to deal with
- Provided a stochastic REACT algorithm to detect the attacked area and line failures → trade-off between accuracy and running time

- Extension to the AC power flow model
- Extension to the noisy scenarios

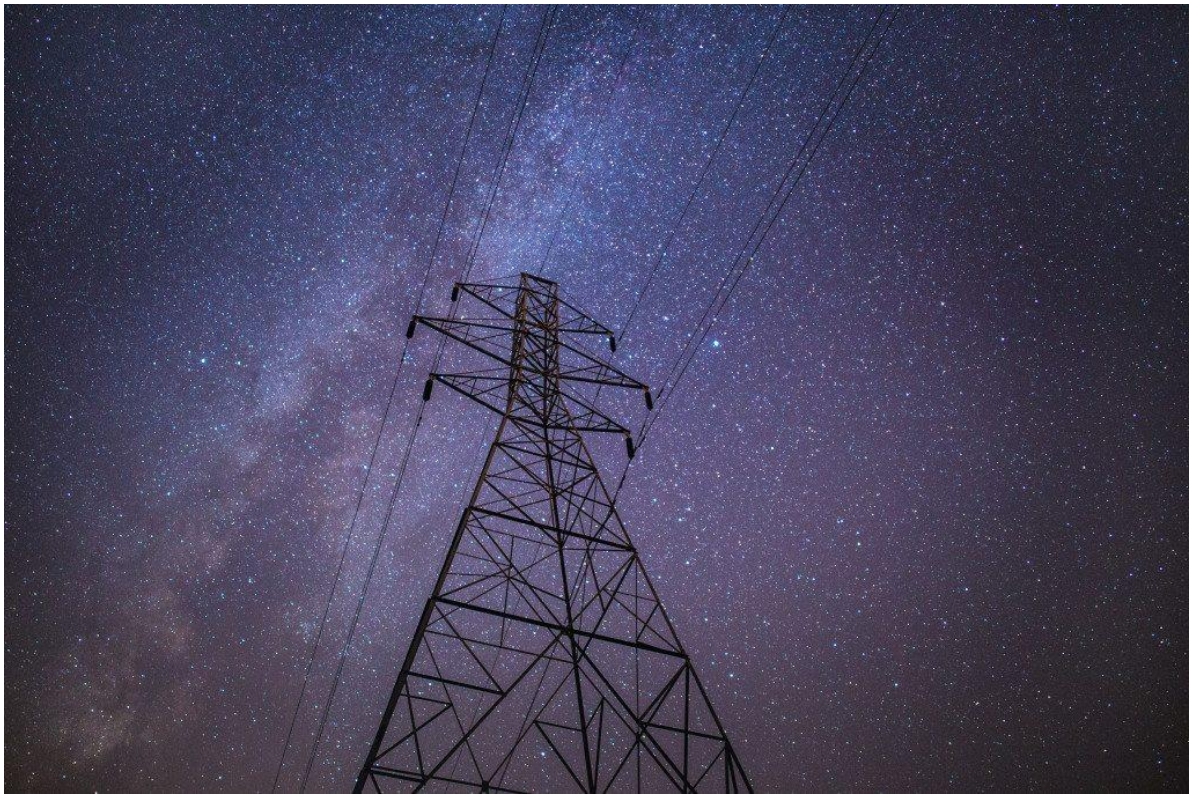
Saleh Soltan, Gil Zussman, "EXPOSE the Line Failures following a Cyber-Physical Attack on the Power Grid", to appear in IEEE Transactions on Control of Network Systems, 2018.

S. Soltan, P. Mittal, and H. V. Poor, "Bayesian Regression for Robust Power Grid State Estimation Following a Cyber-Physical Attack," to appear in Proc. IEEE PES-GM'18, Aug. 2018.

PRINCETON UNIVERSITY

Thank You!

ssoltan@princeton.edu
http://ssoltan.mycpanel.princeton.edu/



This work was supported in part by DTRA grant HDTRA1-13-1-0021, DARPA RADICS under contract #FA-8750-16-C-0054, funding from the U.S. DOE OE as part of the DOE Grid Modernization Initiative, U.S. DOE under Contract No. DEAC36-08GO28308 with NREL, and NSF under grant CCF-1703925 and CCF-1423100.