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

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

Cyber Attacks
- Supervisory Control and Data Acquisition (SCADA) system
- Commands
- Data
Components of Power Grid SCADA

Physical Attack Target

Cyber Attack Target

Power Grid Physical Infrastructure

Supervisory Control and Data Acquisition (SCADA) system

PMU: Phasor Measurement Unit
PDC: Phasor Data Concentrators

IEC: Intelligent Electronic Devices
DC: Data Concentrators

PMU:

PDC:

Comm. Net.

Control Center

Commands

Data
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
AC Power Flows

- Present the grid by a connected graph $G = (N, E)$
- In the phasor domain
  - $V_i = |V_i| e^{i\theta_i}$
    - $|V_i|$ is the Voltage magnitude
    - $\theta_i$ is the phase angle
- Transmission line $(i, k)$ is characterized by series admittance $y_{ik} = g_{ik} + i b_{ik}$
- 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 $\rightarrow$ 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 \(\Rightarrow y_{ik} \approx ib_{ik}\)
  - \(\theta_{ik} \ll 1\) \(\Rightarrow\) \(\cos(\theta_{ik}) \approx 1\) and \(\sin(\theta_{ik}) \approx \theta_{ik}\)

- The power flow equations reduce to

\[
\sum_{k} P_{ik} = P_{i}
\]

- The DC power flows only considers active powers
DC Power Flows (Matrix Form)

The DC power flow can be written in matrix form:

\[ Y D^T \hat{\theta} = \hat{f} \]
\[ A \hat{\theta} = \hat{p} \]

\( D \in \{-1,0,1\}^{n \times m} \): 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 susceptance values,

and \( A = D Y D^T \): the admittance matrix of the grid
Assumptions and Objective

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

$\triangleright \quad \tilde{\theta}^*_H = \tilde{\theta}'_H$

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

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

$H$ : an induced subgraph of $G$ that represents the attacked area
$\overline{H}$: $G \setminus H$
$F$ : 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:
    \[
    \tilde{\theta}_H^* = \tilde{\theta}_H' + \tilde{z}
    \]
  - **Data replay:** the attacker replays measurements from previous hours/days instead of the actual measurements coming from the attacked area:
    \[
    \tilde{\theta}_H^* = \tilde{\theta}_H''
    \]
    in which \( A\tilde{\theta}'' = \tilde{p}'' \) and \( \tilde{p}_H'' = \tilde{p}_H \).

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

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

Reduction from 3-partition problem

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

$$A^{(F)} \begin{bmatrix} \tilde{\theta}'_H \\ \tilde{\theta}'_H \end{bmatrix} = A \tilde{\theta}$$

Lemma. Given $A$, $\tilde{\theta}$ and $\tilde{\theta}^*$, 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 $\tilde{\theta}'_{H_0}$ such that

$$A^{(F)} \begin{bmatrix} \tilde{\theta}'_{H_0} \\ \tilde{\theta}^*_{H_0} \end{bmatrix} = A \tilde{\theta}$$
Attacked Area Approximation
Data Distortion Attack

- For any $i \in \text{int}(H), A_i \tilde{\theta}^* = p_i$
- For any $i \in V \setminus \text{int}(H), A_i \tilde{\theta}^* \neq p_i$
- $S_0 := G[\text{supp}(A \tilde{\theta}^* - \bar{p})]$
- $V_H \subseteq \text{int}(S_0)$

\[ \text{int}(H) \Rightarrow \text{green} \]

\[ S_0 \Rightarrow \text{yellow} \text{ and orange} \]

\[ \text{int}(S_0) \Rightarrow \text{orange} \]

\[ \text{int}(S) := \text{nodes in } S \text{ such that their neighbors are also in } S \]
Data Replay Attack

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

- $S_0 := G[\text{supp}(A\hat{\theta}^* - \hat{p})]$ does not contain the attacked area in this case

Data replay attack

Data distortion attack

$\text{int}(S) := \text{nodes in } S \text{ that their neighbors are also in } S$

$\partial(S) := \text{nodes in } S \text{ that have neighbors also in } \overline{S}$
ATtacked Area Containment (ATAC)

- Provide multiple areas that may contain the attacked area

\[
\begin{align*}
\text{int}(\overline{H}) & \quad \text{green} \\
\text{int}(H) & \quad \text{red} \\
\partial(H) & \quad \text{orange} \\
\partial(\overline{H}) & \quad \text{yellow}
\end{align*}
\]

\[
\begin{align*}
G_1 & := C_1 \cup C_2 \\
G_2 & := C_4 \cup C_5 \\
G_3 & := C_3 \\
G_4 & := C_6 \cup C_7
\end{align*}
\]

At least one of the \( S_0, S_i := G \setminus G_i \) contains the attacked area.

Connected components of \( G \setminus S_0 \):
\[ S_0 = G [\text{supp}(A\hat{\theta}^* - \hat{p})] \]
Line Failures Detection

- Assume $S_0, S_1, \ldots, S_t$ are the subgraphs from ATAC
- Assume that $S$ contains $H \rightarrow \tilde{\theta}_S^* = \tilde{\theta}'_S$
- Brute force search algorithm
  \[
  \min_{F, \tilde{y}} \| A_{G|S} \tilde{\theta}_S^* + A_{G|\tilde{S}}^{(F)} \tilde{y} - \tilde{p} \|_2
  \]
- Not efficient $\rightarrow$ specially that we don’t know if $S$ contains the attacked area or not
- Solution $\tilde{x}$ and $\tilde{y}$ to the following linear program can detect the phase angles and line failures

\[
\text{Fewest number of line failures} \rightarrow \min \| \tilde{x} \|_1 \quad \text{s.t.} \quad \begin{align*}
A_{S|G}(\tilde{\theta} - \tilde{\theta}') &= D_S \tilde{x} \\
A_{S|\tilde{S}}(\tilde{\theta}_S - \tilde{y}) + A_{S|\tilde{S}}(\tilde{\theta}_S - \tilde{\theta}_S^*) &= D_S \tilde{x}
\end{align*}
\]

\[
A_{\tilde{S}|G}(\tilde{\theta} - \tilde{\theta}') = 0 \rightarrow \begin{align*}
A_{\tilde{S}|S}(\tilde{\theta}_S - \tilde{y}) + A_{\tilde{S}|\tilde{S}}(\tilde{\theta}_S - \tilde{\theta}_S^*) &= 0
\end{align*}
\]

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

## Conditions and Limitations

<table>
<thead>
<tr>
<th>External Conditions</th>
<th>Internal Conditions</th>
<th>Attack Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matching</td>
<td>Acyclic</td>
<td>None</td>
</tr>
<tr>
<td>Matching</td>
<td>Planar</td>
<td>Less than half of the edges in each cycle are failed</td>
</tr>
<tr>
<td>Partial Matching</td>
<td>Acyclic</td>
<td>Less than half of the edges connected to an internal node are failed</td>
</tr>
<tr>
<td>Partial Matching</td>
<td>Planar</td>
<td>Two above conditions</td>
</tr>
</tbody>
</table>

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

\[
\begin{align*}
\text{min } & \| W \tilde{x} \|_1 \quad \text{s.t.} \\
A_{S|S} (\tilde{\theta}_S - \tilde{y}) + A_{S|\bar{S}} (\tilde{\theta}_{\bar{S}} - \tilde{\theta}_{\bar{S}}^*) &= D_S \tilde{x} \\
A_{\bar{S}|S} (\tilde{\theta}_S - \tilde{y}) + A_{\bar{S}|\bar{S}} (\tilde{\theta}_{\bar{S}} - \tilde{\theta}_{\bar{S}}^*) &= 0
\end{align*}
\] (***)

- Confidence of the solution

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

- Generate random weights, solve (***)
  - check if for $F = \text{supp}(\tilde{x})$ and $\tilde{y}$, $\| A_{G|S} \tilde{\theta}_{\bar{S}}^* + A_{G|S}^{(F)} \tilde{y} - \tilde{p} \|_2$ 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 $S_0, S_1, ..., S_t$ 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 $c(F, \hat{y}) < 99.9$ and $counter < T$
  5. Generate a random weight matrix $W$
  6. Solve (**) 
  7. Return a solution with the highest confidence

\[
\begin{align*}
\min & \quad \| W \hat{x} \|_1 \quad \text{s.t.} \\
A_{S|S}(\hat{\theta}_S - \hat{y}) + A_{S|\bar{S}}(\hat{\theta}_{\bar{S}} - \hat{\theta}_S^*) &= D_S \hat{x} \\
A_{S|S}(\hat{\theta}_S - \hat{y}) + A_{S|\bar{S}}(\hat{\theta}_{\bar{S}} - \hat{\theta}_S^*) &= 0
\end{align*}
\]
Numerical Results

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

IEEE 300-Bus
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


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.