Stochastic defense against ideal grid attacks

Daniel Bienstock

Columbia University

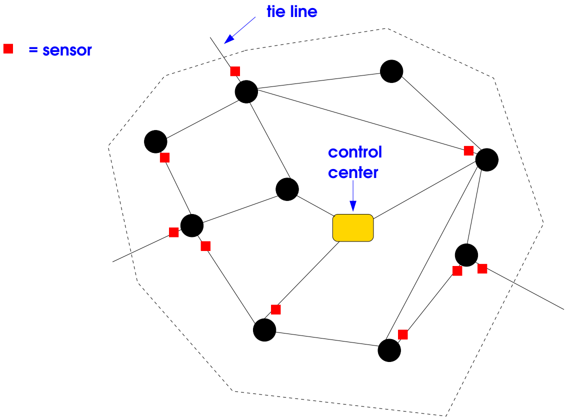
June 2018

"Cyber-physical" attacks

Fact or fiction?

1. An adversary carries out a physical alteration of a grid (example: disconnecting a power line)
2. This is coordinated with a modification of sensor signals ("data replay") – a **hack**
3. The goal is to disguise, or keep completely hidden, the nature of the attack and its likely consequences
4. Power industry: it will never happen ("we would know what happened")
5. Really?

Control centers, RTUs, PMUs, state estimation



Control centers, RTUs, PMUs

- Control center performs a regulatory and economic role
- Sensors report to control center
- Control center issues commands to (in particular) smaller generators
- Sensors: RTUs (old), PMUs (new – and expensive)
- RTUs report **once every four seconds**
- PMUs report
  - **30 to 100 times a second**
  - PMUs report (AC) voltage and current (plus more ...)
- Anecdotal: PMUs overwhelming human operators

State estimation

A data-driven procedure to estimate relevant grid parameters

- Even with PMUs, data is sketchy and noisy
- Statistical procedure: "state estimation" (at control center)

DC power flow equations:

$$B\theta \;=\; P^g - P^d$$

$\boldsymbol{B}$ = susceptance matrix, $\boldsymbol{\theta}$ = phase angles, $\boldsymbol{P^g}$, $\boldsymbol{P^d}$ generation and load vectors

- Sensors provide information that fit **some** of the $\boldsymbol{\theta}, \boldsymbol{P^d}, (\boldsymbol{P^g}?)$ parameters
- State estimation: least squares procedure to estimate the rest, plus more

Prior basic research

- All, or mostly, DC-based
- Intelligent procedures for enriching state estimation so as to detect and reconstruct attacks
- Unavoidable: a model for attacking behavior is essential
- Liu, Ning Reiter (2009), Kim and Poor (2011),
- Deka, Baldick, Vishwanath (2015)
- Soltan, Yannakakis, Zussman (2015 -)
- **Warning:** watch out for those assumptions!

Soltan, Yannakakis, Zussman 2017

- Attacker disconnects lines plus alters sensor output in an (unknown) zone of the grid
- As a result, the equation

$$\boldsymbol{B\theta = P^g - P^d}$$

  is wrong because $\boldsymbol{B}$ is incorrect and measurements $\boldsymbol{\theta}$ are (sparsely) false
- A statistical procedure to try to "fit" a correction to $\boldsymbol{B}$ and a discovery of false data
- Important: testing done using **AC** phase angles $\boldsymbol{\theta}$
- Some assumptions

Today: load change, signal hacking – all AC

- An attacker causes physical changes in the network: in particular **load** changes (no generator changes)
- Attacker also hacks the signal flow: the output of some sensors is altered
- Goal of the attacker is twofold:
  1. Hide the location of the attack and even the fact that an attack happened
  2. **Cause line overloads that remain hidden**
- We assume **full PMU deplyoment**. Everything is **AC** based.

AGC, primary and secondary response

What happens when generation - loads spontaneously changes (i.e. a net imbalance)?

- AC frequency changes proportionally (to first order) near-instantaneously
- **Primary response.** (very quick) Inertia in generators contributes electrical energy to the system
- **Secondary response.** (seconds) Suppose **estimated** generation **shortfall** $= \boldsymbol{\Delta P}$. Then:

Generator $\boldsymbol{g}$ changes output by $\alpha_{\boldsymbol{g}} \boldsymbol{\Delta P}$

AGC, primary and secondary response

What happens when generation - loads spontaneously changes (i.e. a net imbalance)?

- AC frequency changes proportionally (to first order) near-instantaneously
- **Primary response.** (very quick) Inertia in generators contributes electrical energy to the system
- **Secondary response.** (seconds) Suppose **estimated** generation **shortfall** $= \boldsymbol{\Delta P}$. Then:

Generator $\boldsymbol{g}$ changes output by $\boldsymbol{\alpha_g \Delta P}$

- $\boldsymbol{\sum_g \alpha_g = 1}$, $\boldsymbol{\alpha \geq 0}$,

AGC, primary and secondary response

What happens when generation - loads spontaneously changes (i.e. a net imbalance)?

- AC frequency changes proportionally (to first order) near-instantaneously
- **Primary response.** (very quick) Inertia in generators contributes electrical energy to the system
- **Secondary response.** (seconds) Suppose **estimated** generation **shortfall** = $\boldsymbol{\Delta P}$. Then:

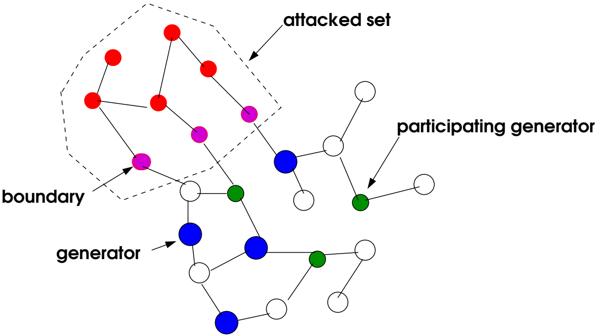
Generator $\boldsymbol{g}$ changes output by $\boldsymbol{\alpha_g \Delta P}$

- $\sum_g \alpha_g = 1$, $\alpha \geq 0$, $\alpha > 0$ for "participating" generators
- Preset participation factors
- $\boldsymbol{\Delta P}$ sensed by control center, which issues commands

Ideal attack: setup

- For this talk, PMUs everywhere: at both ends of each line
- Attacker has been in the system long enough to learn the system
- Attacker chooses, in advance, a non-generator, sparse set $\mathcal{A}$ of buses to attack and in particular a line uv to overload
- In near real-time, the attacker learns the current loads **and their stochastics**
- In near real-time, the attacker solves computational problem that diagrams the attack on $\mathcal{A}$
- This will specify the load changes and the signal distortion
- Post-attack, attacker cannot recompute much and only relies on adding "noise" to the computed distorted signals

Undetectable attack: The attacker's perspective



Undetectable attack: decisions for the attacker (abridged!)

- For every bus in $\mathcal{A}$, a "true" and "reported" complex voltage (magnitude and angle) $\boldsymbol{V_k^{\mathrm{T}}}$ and $\boldsymbol{V_k^{\mathrm{R}}}$
- True and reported voltages **must** agree on the boundary of $\mathcal{A}$!
- True and reported **currents** for lines within $\mathcal{A}$
- Voltages and currents on all other lines (true and reported are identical)
- Two power flow solutions; each must satisfy AC power flow line
- A generation change consistent with secondary response

Undetectable attack: formulation (abridged!)

$$\text{Max } (p_{uv}^{\mathrm{T}})^2 + (q_{uv}^{\mathrm{T}})^2 \quad (1a)$$

s.t.

$$\forall k \in \mathcal{A}^C \cup \text{bd}(\mathcal{A}),\ |V_k^{\mathrm{T}}| = |V_k^{\mathrm{R}}|,\ \theta_k^{\mathrm{T}} = \theta_k^{\mathrm{R}} \quad (1b)$$

$$\forall k \in \mathcal{A},\quad -(P_k^{d,\mathrm{R}} + jQ_k^{d,\mathrm{R}}) = \sum_{km \in \delta(k)} (p_{km}^{\mathrm{R}} + jq_{km}^{\mathrm{R}}), \quad (1c)$$

$$-(P_k^{d,\mathrm{T}} + jQ_k^{d,\mathrm{T}}) = \sum_{km \in \delta(k)} (p_{km}^{\mathrm{T}} + jq_{km}^{\mathrm{T}}), \quad (1d)$$

$$\forall k \in \mathcal{A}^C \setminus \mathcal{R}: \quad \hat{P}_k^g - \hat{P}_k^d + j(\hat{Q}_k^g - \hat{Q}_k^g) = \sum_{km \in \delta(k)} (p_{km}^{\mathrm{T}} + jq_{km}^{\mathrm{T}}) \quad (1e)$$

$$\forall k \in \mathcal{R}: \quad P_k^g - \hat{P}_k^d + j(Q_k^g - \hat{Q}_k^g) = \sum_{km \in \delta(k)} (p_{km}^{\mathrm{T}} + jq_{km}^{\mathrm{T}}) \quad (1f)$$

$$P_k^g - \hat{P}_k^g = \alpha_k \Delta \qquad \text{(AGC response)} \quad (1g)$$

operational limits on all buses, generators and lines (other than **uv**) (1h)

all $\boldsymbol{p}_{\boldsymbol{km}}^{\mathbf{T}}, \boldsymbol{q}_{\boldsymbol{km}}^{\mathbf{T}}$ related to $|\boldsymbol{V}_{\boldsymbol{k}}^{\mathbf{T}}|, |\boldsymbol{V}_{\boldsymbol{m}}^{\mathbf{T}}|, \boldsymbol{\theta}_{\boldsymbol{k}}^{\mathbf{T}}, \boldsymbol{\theta}_{\boldsymbol{m}}^{\mathbf{T}}$ through AC power flow laws (1i)

Ideal attack: follow-up

Following the attack, for any bus $\in \mathcal{A} - \text{bd}(\mathcal{A})$ the attacker reports (at each time t) a complex voltage value

$$\tilde{V}_k \;=\; (|V_k^{\mathrm{R}}| + \boldsymbol{\nu_k(t)})e^{j(\theta_k^{\mathrm{R}} + \boldsymbol{\phi_k(t)})}$$

Here,

$$\mathrm{E}(\boldsymbol{\nu_k(t)}) = \mathrm{E}(\boldsymbol{\phi_k(t)}) = 0,$$

Ideal attack: follow-up

Following the attack, for any bus $\in \mathcal{A} - \text{bd}(\boldsymbol{\mathcal{A}})$ the attacker reports (at each time t) a complex voltage value

$$\tilde{V}_k \;=\; (|V_k^{\mathrm{R}}| + \boldsymbol{\nu_k(t)})e^{j(\theta_k^{\mathrm{R}} + \boldsymbol{\phi_k(t)})}$$

Here,

$$\mathrm{E}(\boldsymbol{\nu_k(t)}) = \mathrm{E}(\boldsymbol{\phi_k(t)}) = 0,$$

(consistent with zero expected load change)

Ideal attack: follow-up

Following the attack, for any bus $\in \mathcal{A} - \text{bd}(\mathcal{A})$ the attacker reports (at each time t) a complex voltage value

$$\tilde{V}_k \ = \ (|V_k^{\mathrm{R}}| + \boldsymbol{\nu_k(t)})e^{j(\theta_k^{\mathrm{R}} + \boldsymbol{\phi_k(t)})}$$

Here,

$$\mathrm{E}(\boldsymbol{\nu_k(t)}) = \mathrm{E}(\boldsymbol{\phi_k(t)}) = 0,$$

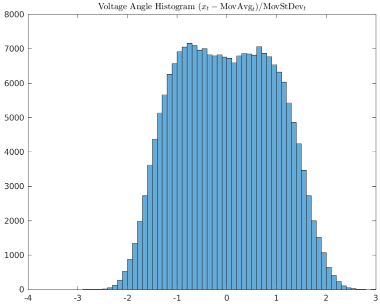
(consistent with zero expected load change)
and

$$\mathrm{var}(\boldsymbol{\nu(t)})$$

agrees with observed covariances

Noise is not just noise

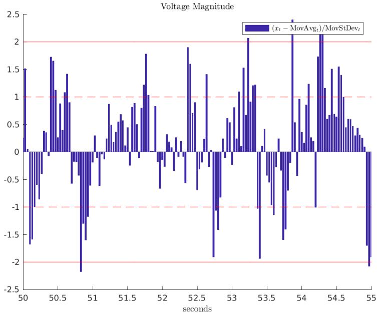
From real time series, voltage angle deviation histogram



Kolmogorov-Smirnoff gaussianity test strongly rejected, always

Noise is not just noise

From real time series, voltage magnitude deviations



Strong and nontrivial correlation structure

Learning noise

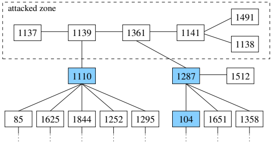
Theorem. (Co)variance of time series can be learned

- In real time
- In streaming fashion
- Under evolving stochasticity

Bienstock, Shukla, Yun, *Non-Stationary Streaming PCA*, Proc. 2017 NIPS Time Series Workshop.

A large-scale example

From `case2746wp` (that has 2746 buses) from the Matpower case library



Undetectable attack with strong overloads on branches $\mathbf{(1361, 1141)}$ and $\mathbf{(1138, 1141)}$.

Defense, 0

- Defender is likely to know that "something" happened (and quickly). But sensor data is noisy and "something" may be inconsequential
- We want a defensive action that is easily implementable in terms of today's grid operation
- Should not lead to false positives
- Solution: change the power flows in a way that the attacker cannot anticipate, and identify inconsistent signals. How?
- A solution: use responding generators – "pseudo AGC"

Defense, 0

- Defender is likely to know that "something" happened (and quickly). But sensor data is noisy and "something" may be inconsequential
- We want a defensive action that is easily implementable in terms of today's grid operation
- Should not lead to false positives
- Solution: change the power flows in a way that the attacker cannot anticipate, and identify inconsistent signals. How?
- A solution: use responding generators – "pseudo AGC"
  Mathematical statement: choose a **random** set of participating generators, change their output by random amounts so as to obtain a **random**, but **valid**, power flow solution

Defense, 0'

Following attack, and in suspicion of an attack

- Defender only has access to **reported** data, which is accurate in the non-attacked zone. But the defender **does not** know the attacked zone.
- **(repeatedly)** Defender chooses a random subset of the AGC-responding generators, and
- Defender computes a random power flow solution where the chosen generators are allowed to change (up or down) their output, within limits. The power flow solution must satisfy e.g. voltage constraints.
- Defender seeks to make the changes in generation large subject to above constraints. "AGC" $\Rightarrow$ generation change

Defense, 1

But attacker cannot anticipate this random action. **Therefore:**

Defense, 1

But attacker cannot anticipate this random action. **Therefore:**

- Reported currents, and implied power flows, will have **near-constant** values within attacked zone
- But outside of attacked zone, with high-probability (?) most lines will see significant changes in current and power flows

Above example (`case2746wp`) has over **3500** lines, but in a few iterations we reduce the number of suspicious lines to $< \mathbf{100}$.

Defense, 1

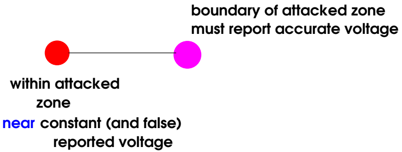
But attacker cannot anticipate this random action. **Therefore:**

- Reported currents, and implied power flows, will have **near-constant** values within attacked zone
- But outside of attacked zone, with high-probability (?) most lines will see significant changes in current and power flows

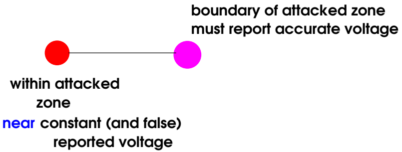
Above example (`case2746wp`) has over **3500** lines, but in a few iterations we reduce the number of suspicious lines to $< \mathbf{100}$.

Good, but not good enough

Defense, 2

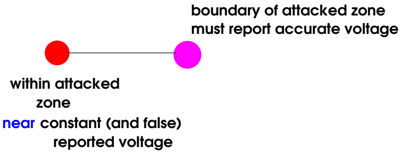

boundary of attacked zone
must report accurate voltage
within attacked
zone
near constant (and false)
reported voltage

Defense, 2



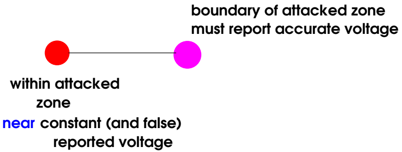
On a line going from boundary to interior of attacked zone

Defense, 2



On a line going from boundary to interior of attacked zone **reported** current will be wrong

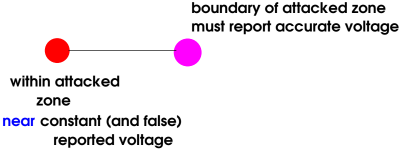
Defense, 2



On a line going from boundary to interior of attacked zone **reported** current will be wrong

because voltage at boundary bus is changing with our “AGC”

Defense, 2



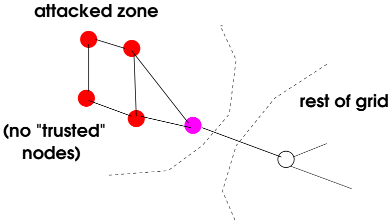
On a line going from boundary to interior of attacked zone **reported** current will be wrong

because voltage at boundary bus is changing with our "AGC" but voltage at interior bus is changing by very small amounts

In above example, **one** iteration identifies boundary lines with **no** false positives

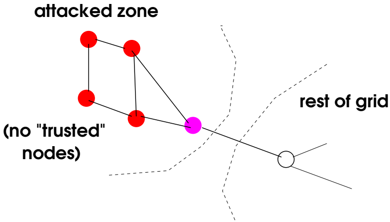
Publication: D. Bienstock and M. Escobar, *Computing undetectable grid attacks, and stochastic defenses*, 2018 SIAM Network Science Workshop. Journal version forthcoming.

A dead zone



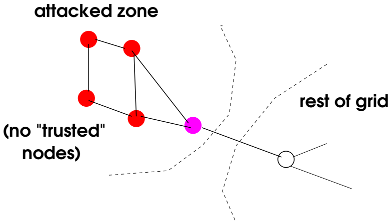
- More difficult/impossible to alter voltages in attacked zone

A dead zone



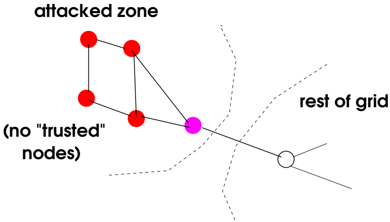
- More difficult/impossible to alter voltages in attacked zone
- More drastic maneuver:

A dead zone



- More difficult/impossible to alter voltages in attacked zone
- More drastic maneuver: use "AGC" to alter *frequency*?
- **Theorem:** under DC model if network is **2-node connected** there are no dead zones

A dead zone



- More difficult/impossible to alter voltages in attacked zone
- More drastic maneuver: use "AGC" to alter *frequency*?
- **Theorem:** under DC model if network is **2-node connected** there are no dead zones but there could be symmetry

- Can the attacker defend against our defense?

- Can the attacker defend against our defense?
- Can we defend against the attacker's defense against our defense?

- Can the attacker defend against our defense?
- Can we defend against the attacker's defense against our defense?
- **Moment** learning
- **Advantage: defender.** Attacker cannot unroll previously generated signals

Thu.Jun.28.000601.2018@blacknwhite