Data Quality and its Effects on Market Functions

Lang Tong, Robert J. Thomas, Liyan Jia, and Jinsub Kim

School of Electrical and Computer Engineering
Cornell University, Ithaca, New York

CERTS Review, August 8, 2012

Project Overview

Objectives:

The main objective of this research is to investigate how data quality affects market functions of current and future grids:

- develop models and measures of data quality;
- quantify risks of bad and malicious data;
- develop techniques to isolate and mitigate effects of bad and malicious data.

Project Overview

Objectives:

The main objective of this research is to investigate how data quality affects market functions of current and future grids:

- develop models and measures of data quality;
- quantify risks of bad and malicious data;
- develop techniques to isolate and mitigate effects of bad and malicious data.

Summary of results:

- Quantifying impacts of worst analog data on real-time LMP.
- Geometric characterization of real-time LMP.
- Topology error and joint data and topology attacks.
- Robustness of nonlinear state estimation against linear data attack.

Outline

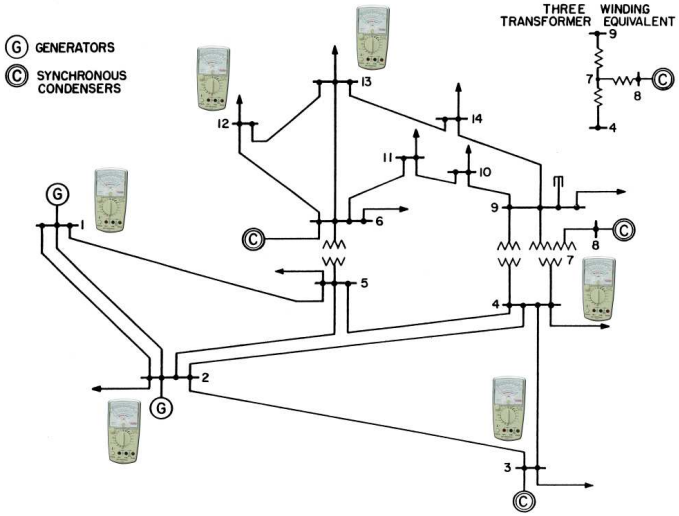
Outline

Outline

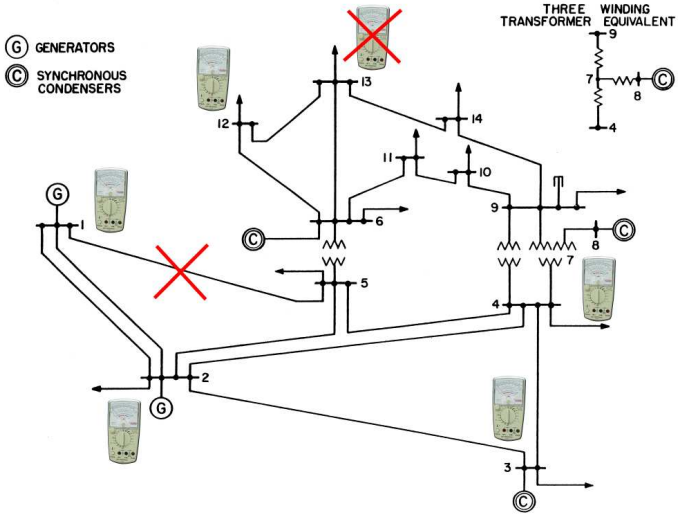
Outline

Outline

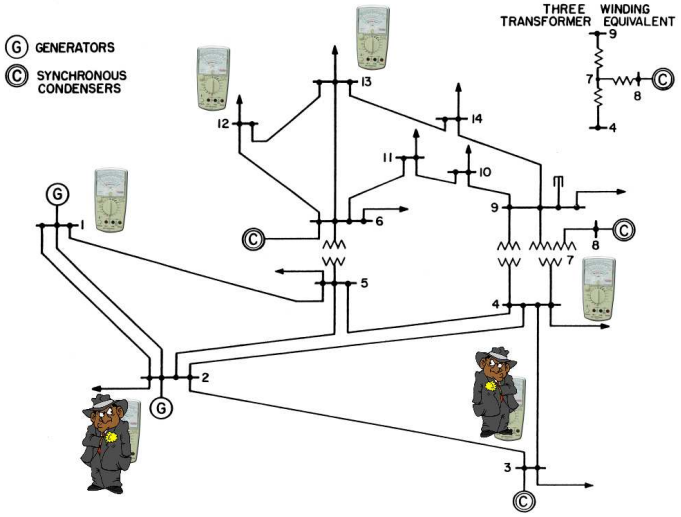
The good, the bad, the malicious, ...



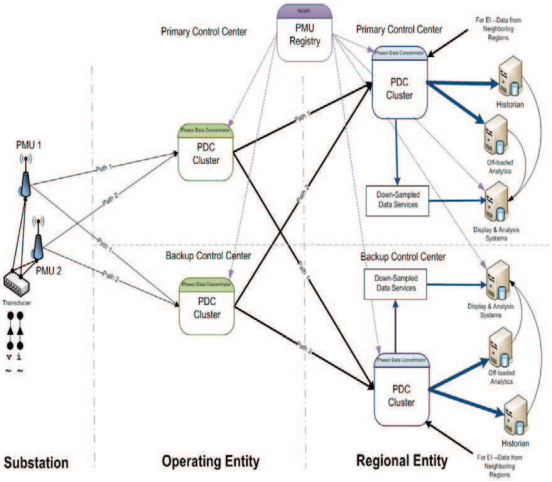
The good, the bad, the malicious, ...

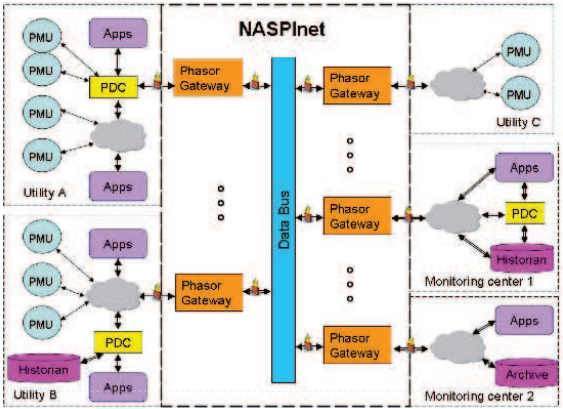


The good, the bad, the malicious, ...

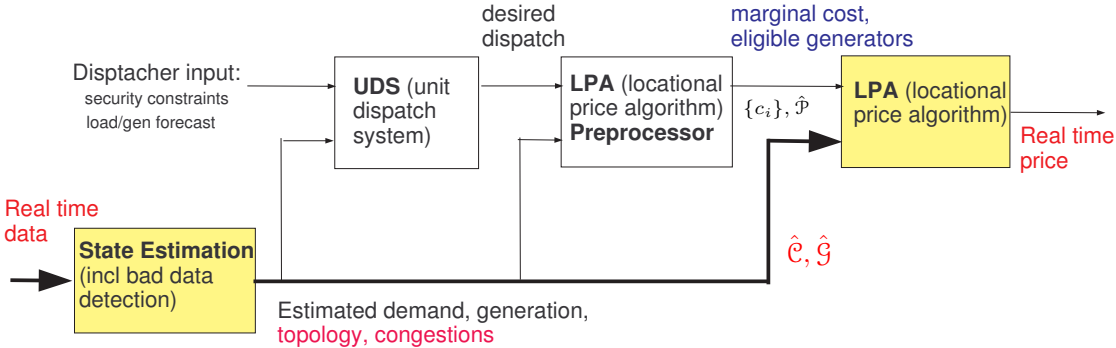


... and it is a cyber-physical system



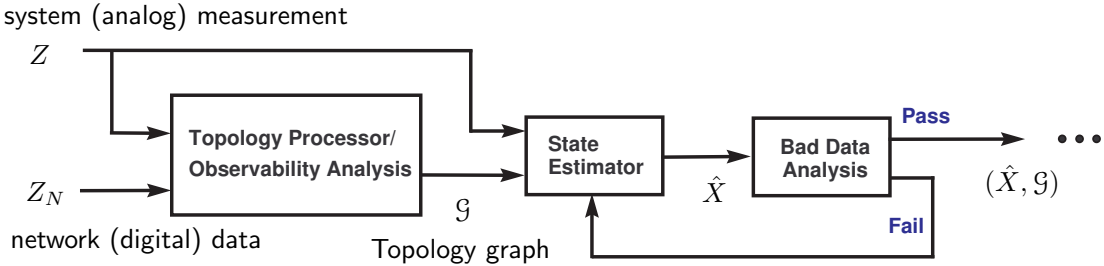


Data in real-time market operation*

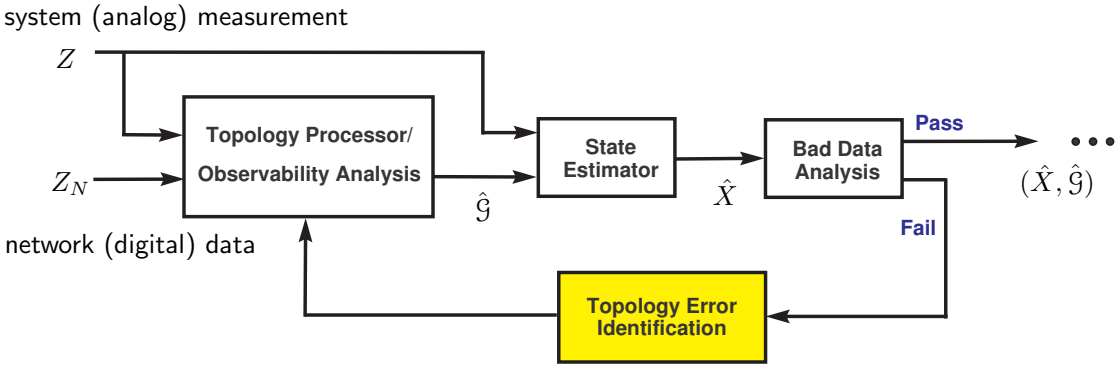


*A. Ott, IEEE TPS, May 2003.

Standard state estimation: schematics



Generalized state estimation: schematics



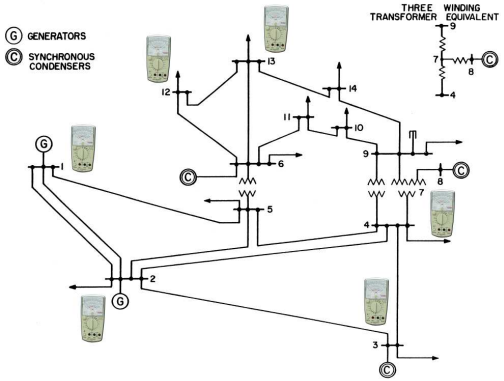
Power system state estimation: model

Static state space model:

$$\begin{aligned} Z_N &= X_N + W_N, X_N \leftrightarrow \mathcal{G} = (\mathcal{V}, \mathcal{E}) \\ Z &= h(X, \mathcal{G}) + W \end{aligned}$$

where

- X: system state (voltage phasors).
- X_N: network state (breaker/switch states).
- Z, Z_N: system and network measurements
- W, W_N: measurement errors.



State estimation: algorithm

- Static state space model:

$$\begin{aligned} Z_N &= X_N + W_N, \qquad X_N \leftrightarrow \mathcal{G} = (\mathcal{V}, \mathcal{E}) \\ Z &= h(X; \mathcal{G}) + W \end{aligned}$$

- The weighted least squares (WLS) estimator:

$$Z_N \to \hat{\mathcal{G}}, \hat{X}_{\mathsf{WLS}}(z) = \arg\min_x (z - h(x; \hat{\mathcal{G}}))^\intercal \Sigma_W^{-1} (z - h(x; \hat{\mathcal{G}})).$$

State estimation: algorithm

- Static state space model:

$$\begin{aligned} Z_N &= X_N + W_N, \quad X_N \leftrightarrow \mathcal{G} = (\mathcal{V}, \mathcal{E}) \\ Z &= h(X; \mathcal{G}) + W \end{aligned}$$

- The weighted least squares (WLS) estimator:

$$Z_N \to \hat{\mathcal{G}}, \hat{X}_{\text{WLS}}(z) = \arg\min_x (z - h(x; \hat{\mathcal{G}}))^\intercal \Sigma_W^{-1} (z - h(x; \hat{\mathcal{G}})).$$

- The generalized weighted least squares (GWLS) estimator:

$$(\hat{X}_{\text{GWLS}}, \hat{\mathcal{G}}_{\text{GWLS}}) = \arg\min_{(x,\mathcal{G})} \left\{ (z - h(x; \mathcal{G}))^\intercal \Sigma_W^{-1} (z - h(x; \mathcal{G})) \right\}$$

State estimation: algorithm

- Static state space model:

$$\begin{aligned} Z_N &= X_N + W_N, \quad X_N \leftrightarrow \mathcal{G} = (\mathcal{V}, \mathcal{E}) \\ Z &= h(X; \mathcal{G}) + W \end{aligned}$$

- The weighted least squares (WLS) estimator:

$$Z_N \to \hat{\mathcal{G}}, \hat{X}_{\text{WLS}}(z) = \arg\min_x (z - h(x; \hat{\mathcal{G}}))^\intercal \Sigma_W^{-1} (z - h(x; \hat{\mathcal{G}})).$$

- The generalized weighted least squares (GWLS) estimator:

$$(\hat{X}_{\text{GWLS}}, \hat{\mathcal{G}}_{\text{GWLS}}) = \arg\min_{(x,\mathcal{G})} \left\{ (z - h(x; \mathcal{G}))^\intercal \Sigma_W^{-1} (z - h(x; \mathcal{G})) \right\}$$

- The generalized state estimator (Monticelli-Wu) is a practical approach involving bad data detection and heuristic searches.

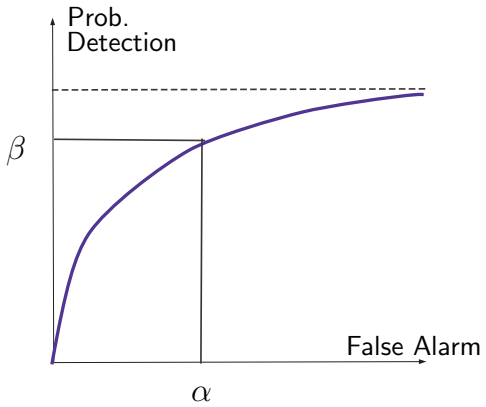
State estimation and bad data detection

- State estimation: Obtain state estimate $\hat{X}$ (and topology graph $\hat{\mathcal{G}}$) using a (generalized) state estimator.
- Bad data detection based on residue error:

$$||z - h(\hat{X}; \hat{\mathcal{G}})||^2_{\Sigma_W^{-1}} \underset{\text{good}}{\overset{\text{bad}}{\gtrless}} \tau$$

where the choice of τ determines the operating point.

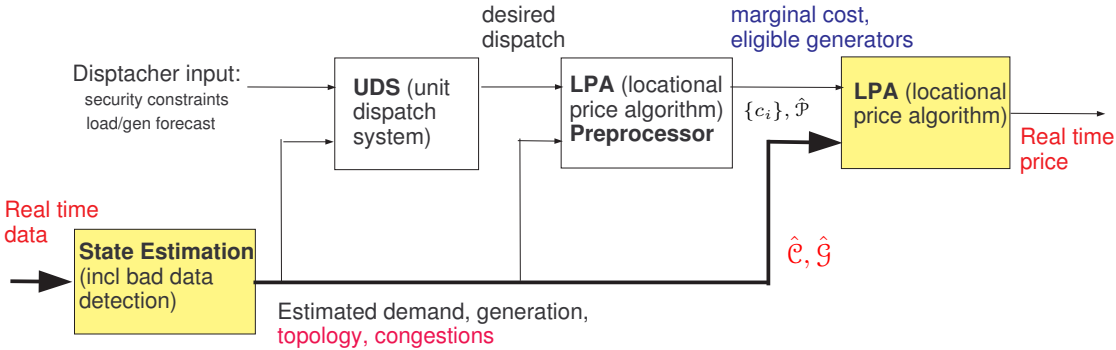
- Bad data identification: If bad data are detected, identify their locations.



Detector operating characteristic

Outline

Real-time LMP†



†A. Ott, IEEE TPS, May 2003.

Real-time LMP algorithm

- The real time LMP is computed via the incremental DC-OPF:

$$
\begin{array}{lll}
\text{minimize} & \sum_i \hat{c}_i \Delta p_i - \sum_j \hat{c}_j \Delta d_j & \\
\text{subjcet to} & \sum_i \Delta p_i = \sum_j \Delta d_j & (\lambda) \\
& \Delta p_{\min} \le \Delta p_i \le \Delta p_{\max} & i \in \hat{\mathcal{P}}; \\
& \Delta d_{\min} \le \Delta d_i \le \Delta d_{\max} & \\
& \sum S_{ki} \Delta p_i - \sum S_{kj} \Delta d_j \le 0; & k \in \hat{\mathcal{C}} \qquad (\mu_k)
\end{array}
$$

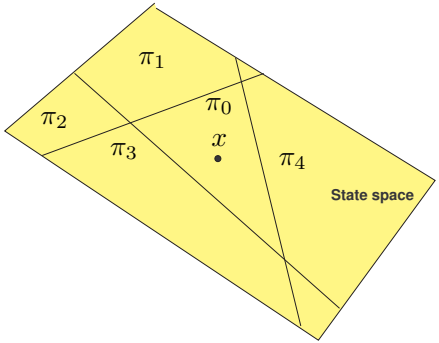
- The LMP at bus i is given by

$$\hat{\pi}_i = \hat{\lambda}^* - \sum_{j \in \hat{\mathcal{C}}} S_{ji} \hat{\mu}_j^*$$

From data to LMP: a geometrical view

- The state space is partitioned by hyperplanes into price regions
  - Each hyperplane corresponds to a congested line:

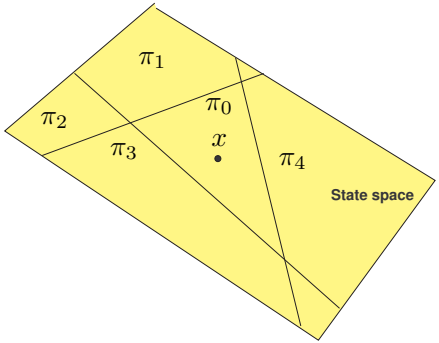
    congestion on line $i \Leftrightarrow f_i^{\mathsf{T}} x > c_i$



From data to LMP: a geometrical view

- The state space is partitioned by hyperplanes into price regions
  - Each hyperplane corresponds to a congested line:

    congestion on line $i \Leftrightarrow f_i^{\mathsf{T}} x > c_i$
  - Each price region corresponds to a congestion pattern.

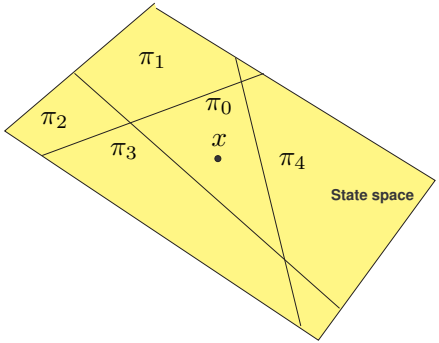


From data to LMP: a geometrical view

- The state space is partitioned by hyperplanes into price regions
  - Each hyperplane corresponds to a congested line:

    congestion on line $i \Leftrightarrow f_i^{\mathsf{T}} x > c_i$

  - Each price region corresponds to a congestion pattern.
  - The LMP value in each region is determined by shift factors S_{ij}.

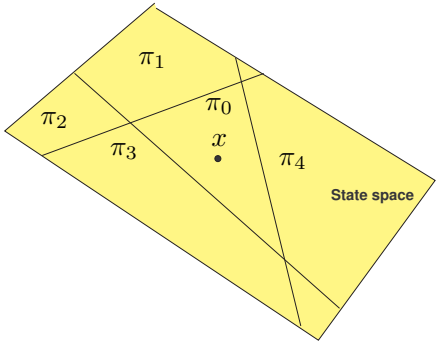


From data to LMP: a geometrical view

- The state space is partitioned by hyperplanes into price regions
  - Each hyperplane corresponds to a congested line:

    congestion on line $i \Leftrightarrow f_i^{\mathsf{T}} x > c_i$

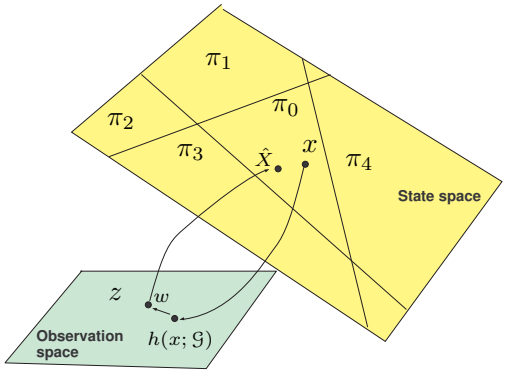
  - Each price region corresponds to a congestion pattern.
  - The LMP value in each region is determined by shift factors S_{ij}.
  - LMP value in each region is not affected by data!



From data to LMP: a geometrical view

- The state space is partitioned by hyperplanes into price regions
- The real time LMP forms a Markov chain:

$$Z \to \hat{X} \to \pi$$

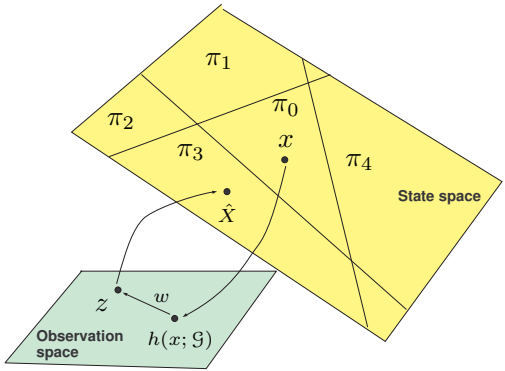


From data to LMP: a geometrical view

- The state space is partitioned by hyperplanes into price regions
- The real time LMP forms a Markov chain:

$$Z \to \hat{X} \to \pi$$

- If data are to affect price, they have to move $\hat{X}$ to a different price region.

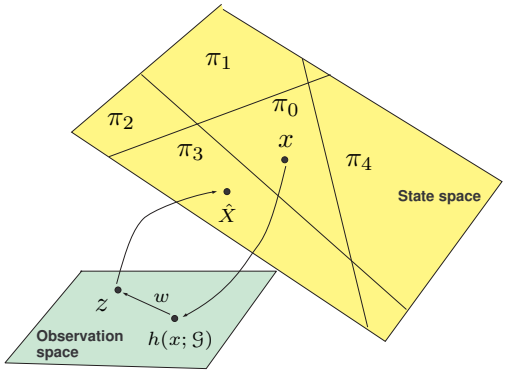


From data to LMP: a geometrical view

- The state space is partitioned by hyperplanes into price regions
- The real time LMP forms a Markov chain:

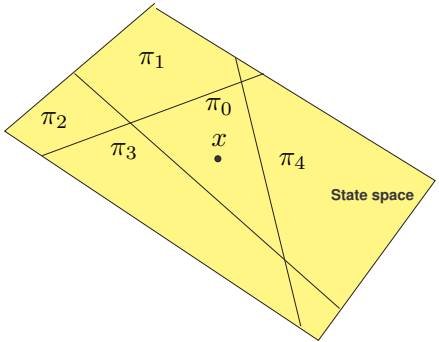
$$Z \to \hat{X} \to \pi$$

- If data are to affect price, they have to move $\hat{X}$ to a different price region.
- What happens with topology error?



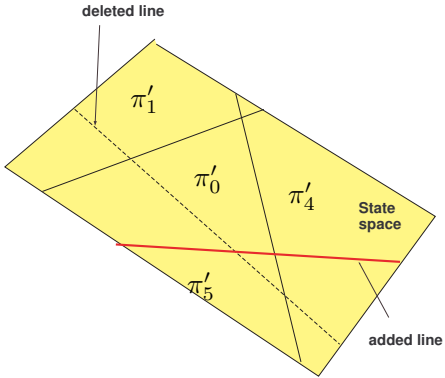
From data to LMP with topology error

- The state space is partitioned by hyperplanes into price regions, and each region corresponds to a congestion pattern.



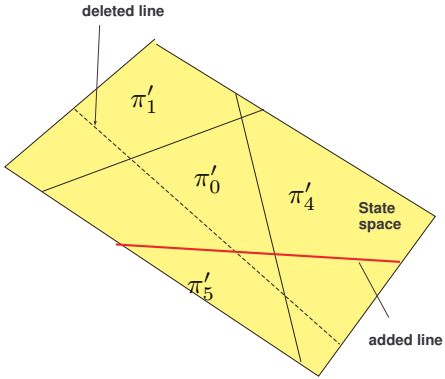
From data to LMP with topology error

- The state space is partitioned by hyperplanes into price regions, and each region corresponds to a congestion pattern.
- A change in topology may add or delete hyperplanes.



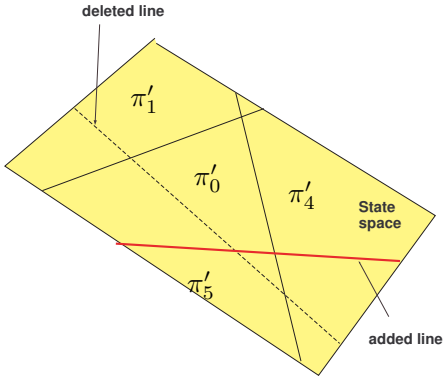
From data to LMP with topology error

- The state space is partitioned by hyperplanes into price regions, and each region corresponds to a congestion pattern.
- A change in topology may add or delete hyperplanes.
- But LMP values in all regions are function of shift factor S_{ij}.



From data to LMP with topology error

- The state space is partitioned by hyperplanes into price regions, and each region corresponds to a congestion pattern.
- A change in topology may add or delete hyperplanes.
- But LMP values in all regions are function of shift factor S_{ij}.



LMPs may all change due to the deletion of a single line!

Outline

Data quality models

We consider a simple perturbation model

$$\begin{aligned} Z &= h(X;\mathcal{G}) + \Delta + W, \quad \Delta \in \mathcal{A} \\ Z_N &= X_N + \Delta_N, \quad \Delta_N \in \mathcal{A}_N \end{aligned}$$

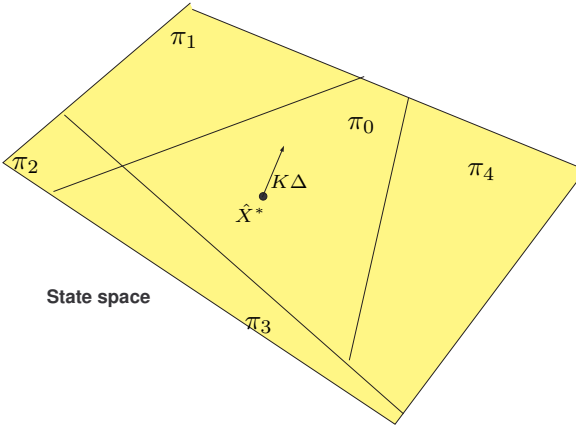
where $\mathcal{A}, \mathcal{A}_N$ characterize the nature of the perturbation.

- Some examples:
  1. Static perturbation: $\mathcal{A} = \{a : \|a\|_0 \leq T\}$, $\mathcal{A}_N = \{\text{single line change}\}$.
  2. Data dependent perturbation: $\Delta = Q(Z) \in \mathcal{A}$.
  3. Dynamic perturbation: $\Delta_t = Q(Z_t, Z_{t-1}, \cdots)$.
- We focus on the worst case analysis for (1-2).

Data attack: constructing the worst data

- Under the DC model, $z = Hx + \Delta + w$, the bad data Δ moves the WLS estimate

$$\hat{X}^* \to \hat{X}_{\text{WLS}} = Kz = \hat{X}^* + K\Delta$$



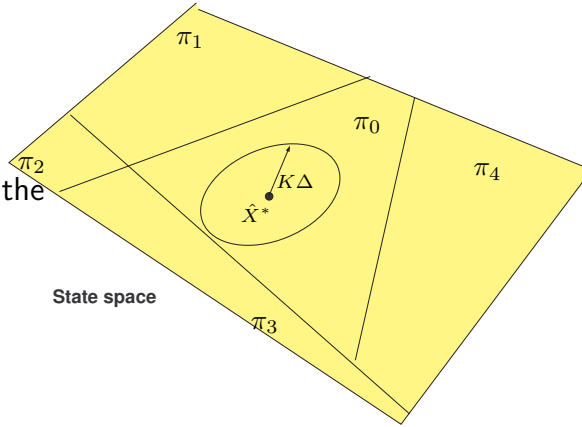
Data attack: constructing the worst data

- Under the DC model, $z = Hx + \Delta + w$, the bad data Δ moves the WLS estimate

$$\hat{X}^* \to \hat{X}_{\text{WLS}} = Kz = \hat{X}^* + K\Delta$$

- Thus Δ are not detectable if $\hat{X}_{\text{WLS}}$ passes the bad data test,

$$\|z - H\hat{X}_{\text{WLS}}\|^2_{\Sigma_W^{-1}} \quad \leq \quad \tau$$



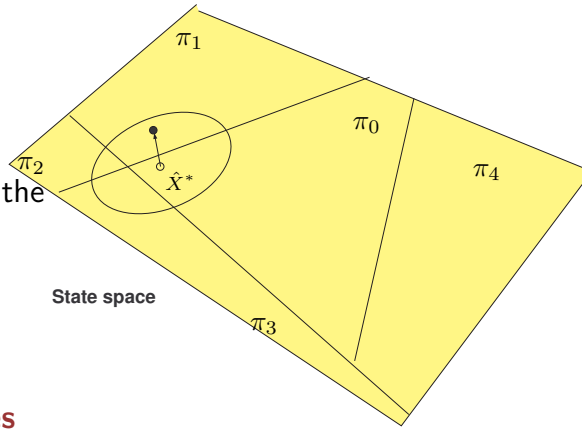
Data attack: constructing the worst data

- Under the DC model, $z = Hx + \Delta + w$, the bad data Δ moves the WLS estimate

$$\hat{X}^* \to \hat{X}_{\text{WLS}} = Kz = \hat{X}^* + K\Delta$$

- Thus Δ are not detectable if $\hat{X}_{\text{WLS}}$ passes the bad data test,

$$\|z - H\hat{X}_{\text{WLS}}\|^2_{\Sigma_W^{-1}} \quad \leq \quad \tau$$



- **LMP change is possible only for states near boundaries!**

Joint topology and data attack

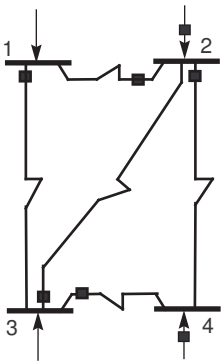
- Under the actual topology,

$$\begin{aligned} z_N &= x_N \leftrightarrow \mathcal{G} = (\mathcal{V}, \mathcal{E}) \\ z &= Mf = MYA^{\mathsf{T}}x + w, \end{aligned}$$

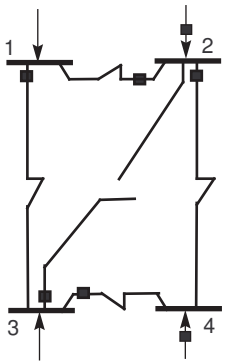
  where Y is the admittance matrix, A^{T} the reduced branch-to-node incidence matrix.

- Under the bad data perturbed topology $z_N \to \bar{\mathcal{G}} = (\bar{\mathcal{V}}, \bar{\mathcal{E}})$,

$$\begin{aligned} \bar{z}_N &= x_N + \Delta_N \\ \bar{z} &= M\bar{f} = MY\bar{A}^{\mathsf{T}}\bar{x} + w \end{aligned}$$



Actual topology



Bad data perturbed topology

Joint topology and data attack

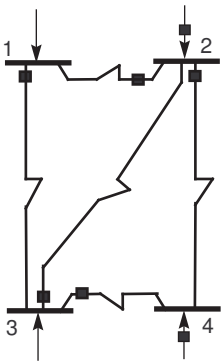
- Under the actual topology,

$$\begin{aligned} z_N &= x_N \leftrightarrow \mathcal{G} = (\mathcal{V}, \mathcal{E}) \\ z &= Mf = MYA^{\mathsf{T}}x + w, \end{aligned}$$

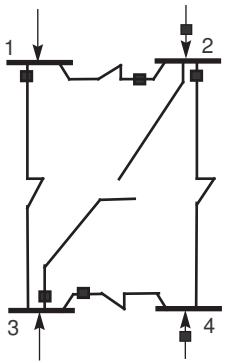
where Y is the admittance matrix, A^{T} the reduced branch-to-node incidence matrix.

- Under the bad data perturbed topology $z_N \to \bar{\mathcal{G}} = (\bar{\mathcal{V}}, \bar{\mathcal{E}})$,

$$\begin{aligned} \bar{z}_N &= x_N + \Delta_N \\ \bar{z} &= M\bar{f} = MY\bar{A}^{\mathsf{T}}\bar{x} + w \end{aligned}$$



Actual topology



Bad data perturbed topology

- Find undetectable bad and sparse data such that $\bar{z} = z + \Delta$ and $\bar{z}_N = z_N + \Delta_N$.

Single branch topology change

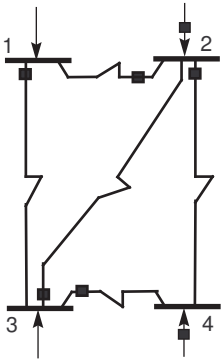
- Fix the state the same in both (noiseless) systems

$$\begin{aligned} z &= MYA^{\mathsf{T}}x \quad \xrightarrow{\Delta} \quad \bar{z} = MY\bar{A}^{\mathsf{T}}\bar{x} \\ \Delta &= \bar{z} - z = MY(\bar{A} - A)^{\mathsf{T}}x \end{aligned}$$

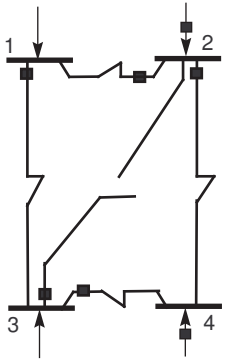
- If $\mathcal{G}$ differ from $\bar{\mathcal{G}}$ by only one branch, then

$$\Delta z = \rho_i m_i$$

where ρ_i is the power flow on the ith branch, and m_i is a column of M. And it is sparse!



Actual topology



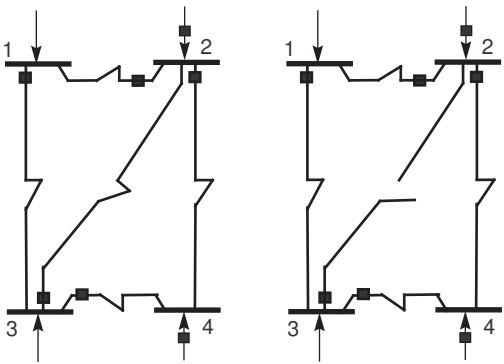
Bad data perturbed topology

Single branch topology change

- By jointly change the system (analog) and network (digital) data

$$z \to \bar{z} = z + \Delta z, \quad \mathcal{G} \to \bar{\mathcal{G}}$$

the bad data is not detectable.



Actual topology

Bad data perturbed topology

Single branch topology change

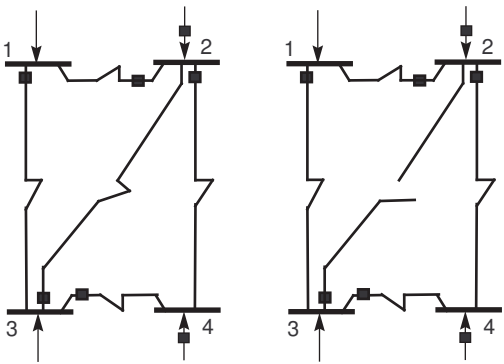
- By jointly change the system (analog) and network (digital) data

$$z \to \bar{z} = z + \Delta z, \quad \mathcal{G} \to \bar{\mathcal{G}}$$

the bad data is not detectable.

- In general, a set of lines can be added or deleted by

$$\Delta z = \sum_i \rho_i m_i$$



Actual topology

Bad data perturbed topology

Single branch topology change

- By jointly change the system (analog) and network (digital) data

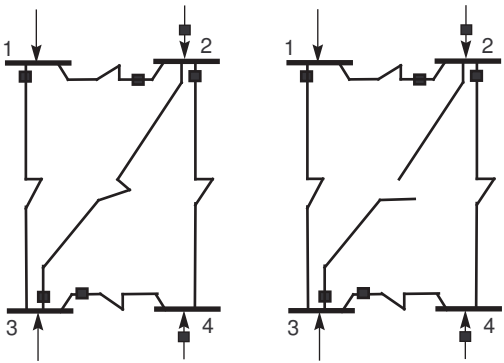
$$z \to \bar{z} = z + \Delta z, \quad \mathcal{G} \to \bar{\mathcal{G}}$$

  the bad data is not detectable.

- In general, a set of lines can be added or deleted by

$$\Delta z = \sum_i \rho_i m_i$$

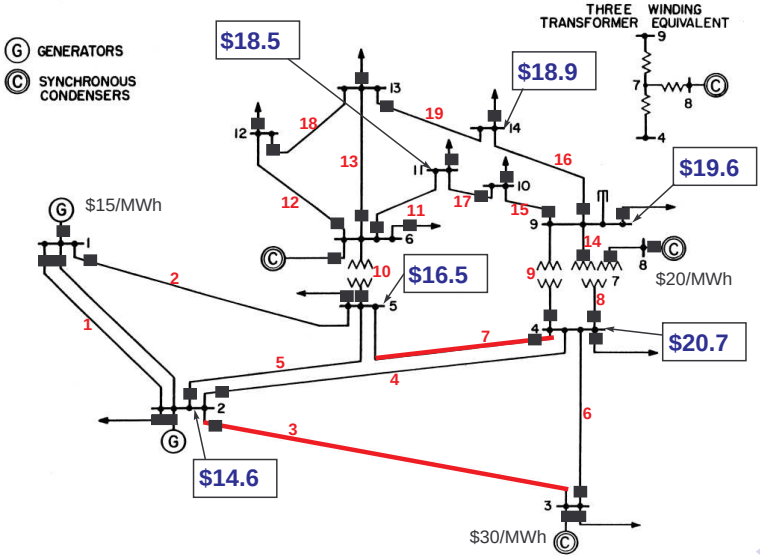
- Intuition: redistribute flow of the altered branches.



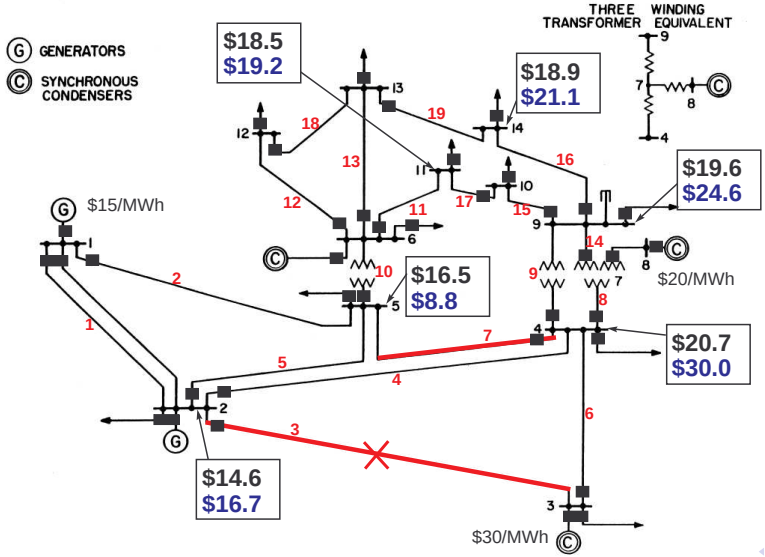
Actual topology

Bad data perturbed topology

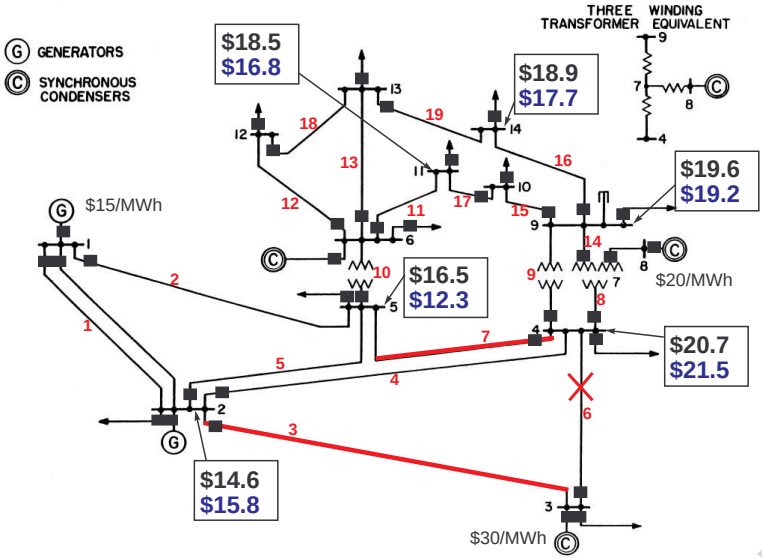
Example: IEEE 14-bus network



Single branch change: IEEE 14-bus network



Example: IEEE 14-bus network



Single branch change: IEEE 14-bus network

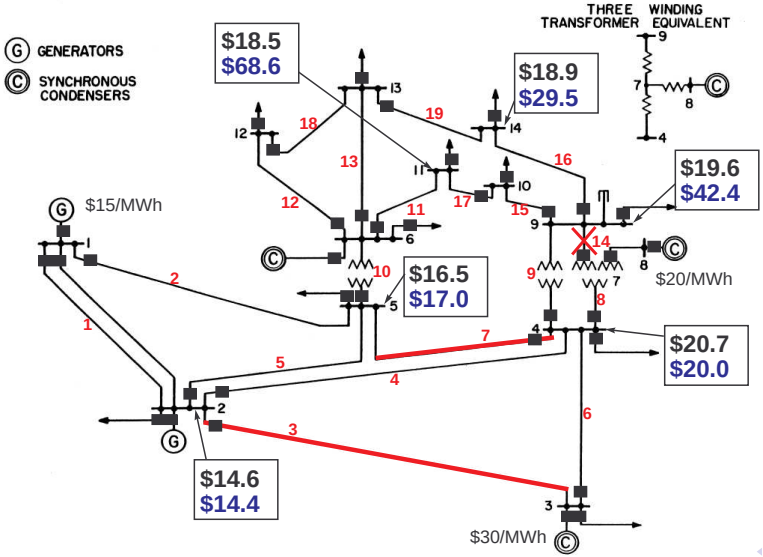
Line 6 removed

	line 3	line 7	line 11
2	0.00	0.07	-0.00
3	-1.00	0.07	-0.00
4	0.00	0.58	-0.02
5	0.00	-0.24	0.01
6	0.00	0.01	0.18
7	0.00	0.45	-0.12
8	0.00	0.45	-0.12
9	0.00	0.37	-0.17
10	0.00	0.31	-0.28
11	0.00	0.16	-0.54
12	0.00	0.03	0.16
13	0.00	0.06	0.13
14	0.00	0.24	-0.04

No attack

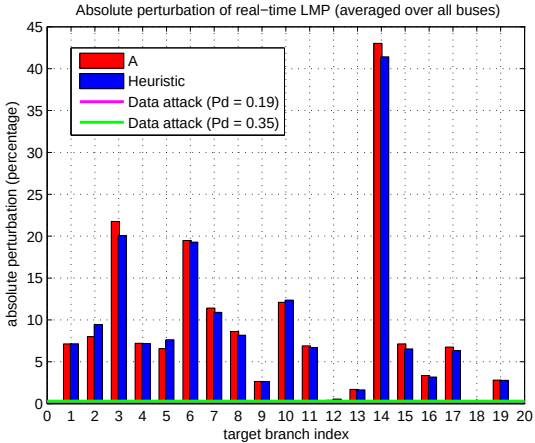
	line 3	line 7	line 11
2	0.03	0.08	0.00
3	-0.55	0.30	-0.01
4	-0.15	0.51	-0.02
5	-0.10	-0.30	0.01
6	-0.12	-0.05	0.19
7	-0.14	0.37	-0.11
8	-0.14	0.37	-0.11
9	-0.14	0.30	-0.17
10	-0.13	0.24	-0.28
11	-0.13	0.10	-0.54
12	-0.12	-0.03	0.16
13	-0.12	0.00	0.13
14	-0.13	0.17	-0.04

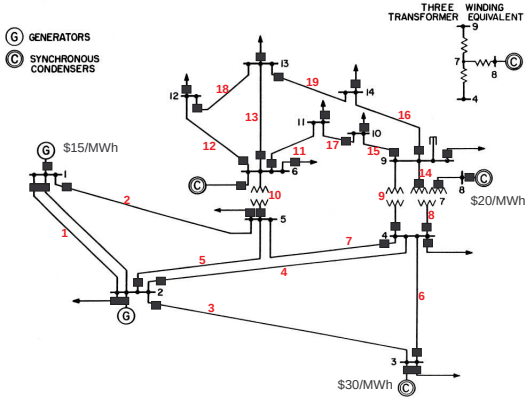
Example: IEEE 14-bus network



Example: data attack vs. topology attack

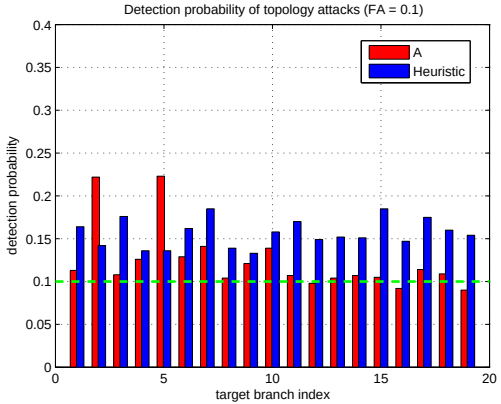
- Real-time LMP perturbation.

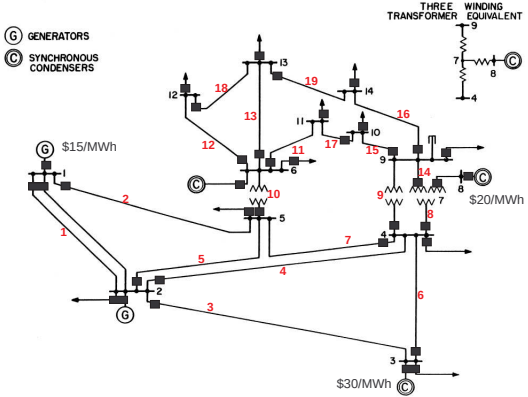




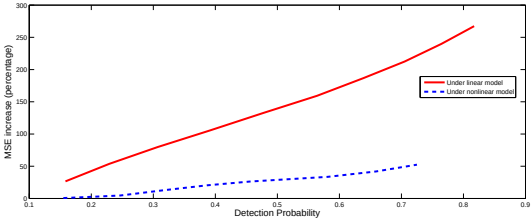
Detectability

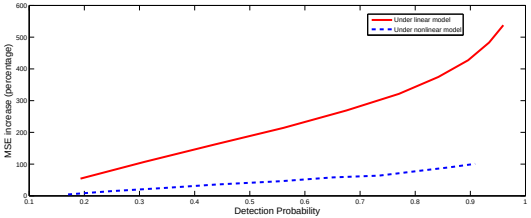
- FA of residual test is 0.1. AC model/ AC state estimator.
- Detection probability:

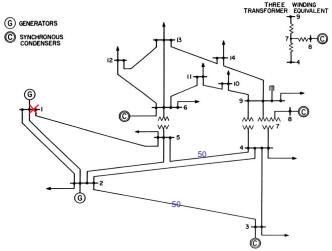


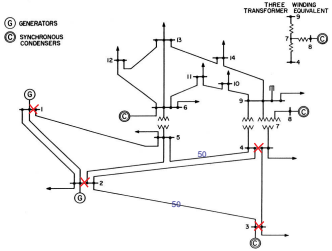


Robustness of nonlinear state estimation: MSE

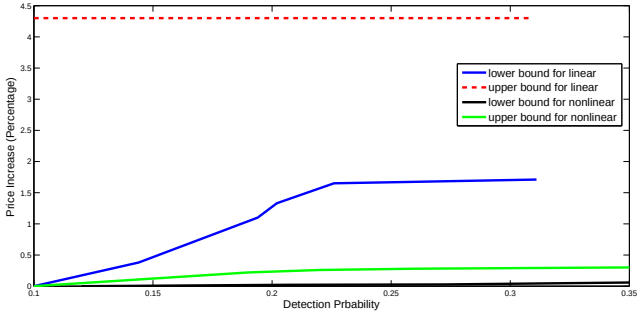


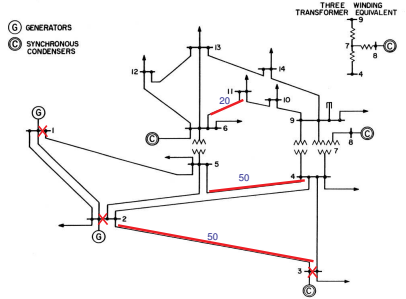




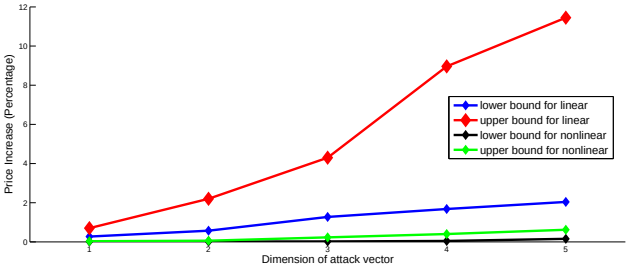


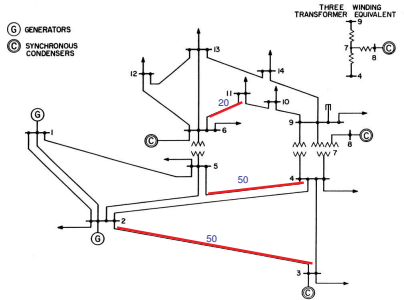
Robustness of nonlinear state estimation: LMP





Robustness of nonlinear state estimation: LMP





Project summary and future work

Project summary

1. Developed system and network and data quality models.
2. Obtained geometric characterization of real-time LMP.
3. Showed that bad analog and bad topology data affect LMP differently.
4. Constructed a simple undetectable joint topology and data attack.
5. Evaluated the effectiveness of worst data constructed from DC model.

Future work

- Optimal use of local information to construct joint data and topology attacks.
- Impacts of data on unit dispatch system (UDS) and LPA preprocessing.
- Dynamic models and more sophisticated bad data detection.