Cyber Security of the Smart Grid

Dr. Deepa Kundur

University of Toronto

Background: State Estimation

Operating States of a Power System

- operating conditions of a power system at a particular time can be determined if the following are known:
  - network model
  - complex phasor voltages at every bus
- the power system can move into one of the three possible states:
  - normal
  - emergency
  - restorative

Normal State

- all the loads in the system can be supplied power by the existing generators without violating any operating constraints; constraint examples: maximum generation levels, min/max bus voltages, etc.
- A system in the normal state can be classified as:
  - secure: the system can remain in normal state following the occurrence of a critical contingency (line outage, generator outage)
  - insecure: the system cannot remain in normal state with the occurrence of a critical contingency

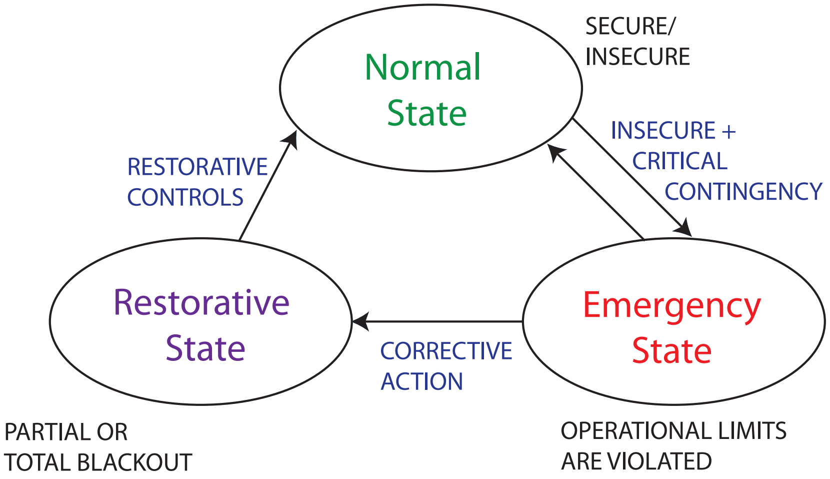
Emergency State

- violation of some of the operating constraints while the power system continues to supply power to all the loads
- must bring it back to normal state using corrective actions

Restorative State

- in emergency state corrective control measures applied; examples:
  - disconnecting loads
  - disconnecting lines
  - disconnecting transformers
  - disconnecting other equipment . . .

  in order to stabilize/eliminate operating limit violations in reduced configuration
- restore the load versus generator balance to supply power to all the loads



Normal State

SECURE/ INSECURE

INSECURE + CRITICAL CONTINGENCY

Emergency State

OPERATIONAL LIMITS ARE VIOLATED

CORRECTIVE ACTION

Restorative State

PARTIAL OR TOTAL BLACKOUT

RESTORATIVE CONTROLS

Power System Security Analysis

1. continuous monitoring of the system conditions
2. identification of the operating state
3. determination of the necessary preventative action in case the system state is found to be insecure

State Estimation

- State Estimator: facilitate accurate and efficient monitoring of the operational constraints; estimates transmission line loadings, bus voltages
- conducted in SCADA system leading to the establishment of an Energy Management System
- provides information to analyze contingencies and determine corrective actions
- acts like a filter between the raw measurements received from the system and all the application functions that require the most reliable data base for current state of the system

State Estimator Functions

From Abur and Expósito (2004):

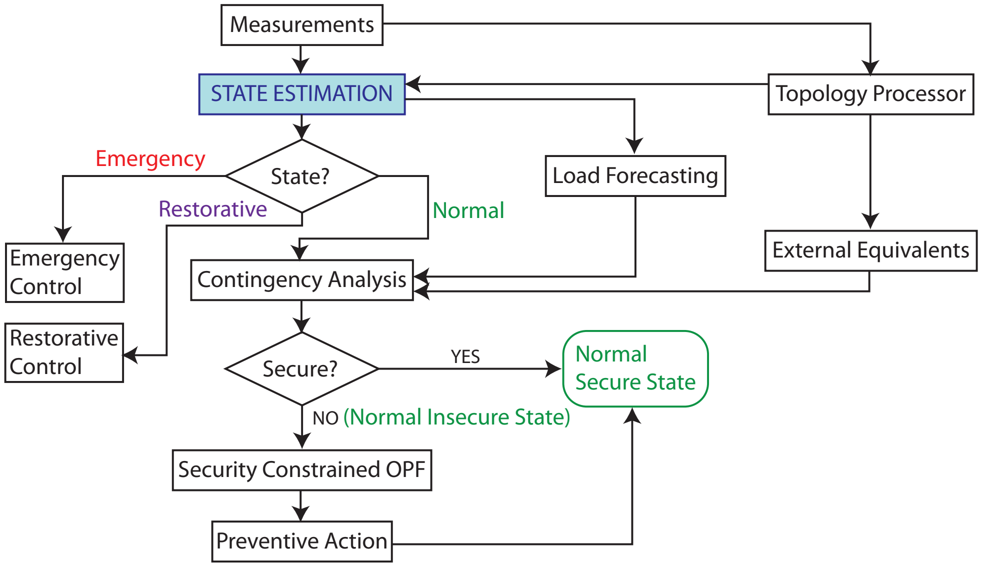
- Topology processor: gathers status data about the circuit breakers and switches and configures the single-line diagram of the system.
- Observability analysis: determines if a state estimation solution for the entire system can be obtained using the available set of measurements (Q: Is the problem underdetermined?); identifies the unobservable branches and the observable islands in the system if any exist.
- State estimation solution: determines the optimal estimate for the system state, which is composed of complex bus voltages in the entire power system, based on the network model and the gathered measurements from the system; also provides the best estimates for all the line flows, loads, transformer taps and generator outputs.

State Estimator Functions

From Abur and Expósito (2004):

- Bad data processing: detects the existence of gross errors in the measurement set; identifies and eliminates bad measurements provided that there is enough redundancy in the measurement configuration
- Parameter and structural error processing: estimates various network parameters such as transmission line model parameters, tap changing transformer parameters, shunt capacitor or reactor parameters; detects structural errors in the network configuration and identifies the erroneous breaker status provided that there is enough measurement redundancy

Adapted from Abur and Expósito (2004):



Measurements

STATE ESTIMATION

Topology Processor

Emergency

State?

Load Forecasting

Restorative

Normal

Emergency Control

External Equivalents

Contingency Analysis

Restorative Control

Secure?

YES

Normal Secure State

NO (Normal Insecure State)

Security Constrained OPF

Preventive Action

Paper Overview:

False Data Injection Attacks against State Estimation in Electric Power Grids

by

Yao Liu, Peng Ning and Michael K. Reiter

Overview Contributions

- introduce false data injection attacks
  - demonstrate how an opponent can make use of information on the power system to create malicious errors into state estimation variables
  - bypass bad data detection methods
- consider different scenarios: constrained to adding attack errors to k specific meters or any subset of k meters
- construct attack vectors and assess them via simulations on IEEE test systems

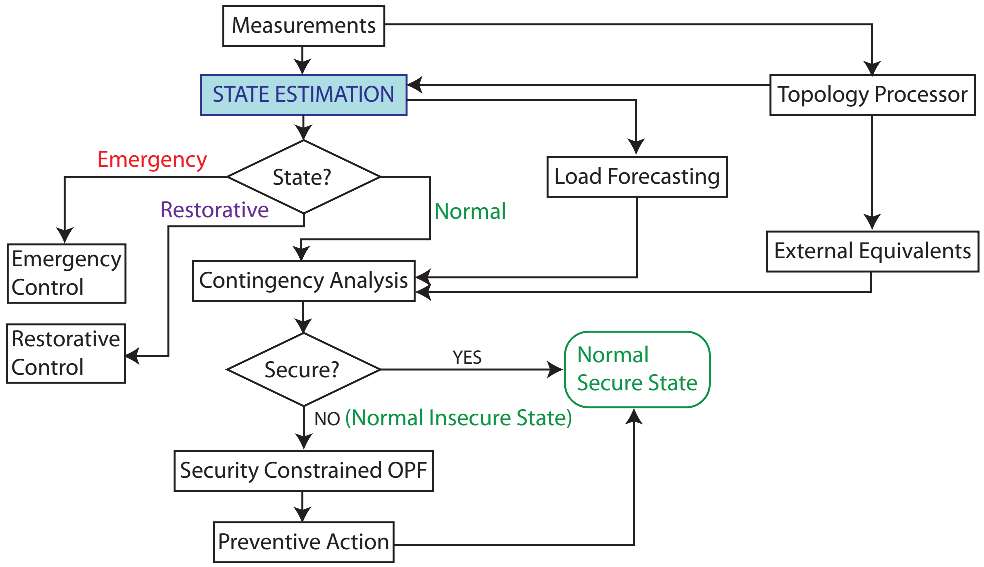
How can one corrupt measurements?

- the physical availability of meters makes it possible for an opponent to tamper them and change the value that they sense:
  - can fabricate measurements completely
  - can add a bias to meter measurements
  - can conduct a denial-of-service and not report meter measurements

What can potentially occur if meter measurements are inaccurate?

- state estimation results will be incorrect
- incorrect decision-making resulting in incorrect control operations leading to devastating effects . . .

Adapted from Abur and Expósito (2004):



Bad Data Detection

- algorithms exist that attempt to detect "bad" data and them remove them from state estimation

Q: Can bad data detection algorithms be overcome?

A: Possibly, if an opponent has knowledge of the configuration of the power system.

State Estimation Problem in this Paper

- Estimate power system state variable

$$\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$$

based on the meter measurements

$$\mathbf{z} = (z_1, z_2, \ldots, z_m)^T$$

where $n, m \in \mathbb{Z}^+$ and $x_j, z_i \in \mathbb{R}$ for $j = 1, 2, \ldots, n$ and $i = 1, 2, \ldots, m$.

State Estimation Problem in this Paper

- Specifically, assuming

$$\mathbf{e} = (e_1, e_2, \ldots, e_m)^T$$

where $e_i \in \mathbb{R}$, $i = 1, 2, \ldots, m$ are measurement errors; the state variables are related to the measurements as follows:

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{e}$$

where $\mathbf{h}(\mathbf{x}) = (h_1(x_1, x_2, \ldots, x_n), \ldots, h_m(x_1, x_2, \ldots, x_n))^T$, the state estimation problem is to find an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ that is the best fit of the measurement $\mathbf{z}$ above.

Linearization: DC Power Flow Model

- The original generally nonlinear relationship

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{e}$$

through linearization becomes

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{e}$$

where $\mathbf{H} = (h_{i,j})_{m\times n}$.

- Note: physically **H** is dependent on the topology and line impedances of the power system and is constant.

Linear State Estimation Problem

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{e}$$

where $\mathbf{H} = (h_{i,j})_{m\times n}$

- State estimation problem: How can we find the best fit $\hat{\mathbf{x}}$ for a given **z**?
- We have **n** unknowns and **m** equations.
- Typically $\mathbf{m} \gg \mathbf{n} \implies$ OVERDETERMINED SYSTEM

Common Estimation Approaches

- maximum likelihood criterion: selects estimation of **x** that produces (with the greatest probability) the observed data **z** for a given statistical model of **e**; formulated as a maximization of a likelihood function
- weighted least-square criterion: minimizes weighted sum of squares of errors for solving each equation of the overdetermined system; squared error for ith Eq: $(z_i - \sum_j h_{i,j}x_j)^2$
- minimum variance criterion: solution that minimizes the variance of the estimate of **x** assuming a statistical model for **e**

Common Estimation Approaches

Assuming that **e** is zero mean and Gaussian, all estimators result in the SAME solution:

$$\hat{\mathbf{x}} = (\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}\mathbf{z}$$

where

$$\mathbf{W} = \begin{bmatrix} \sigma_1^{-2} & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^{-2} & 0 & \cdots & 0 \\ 0 & 0 & \vdots & \cdots & 0 \\ \vdots & \cdots & & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_m^{-2} \end{bmatrix}$$

and σ_i^2 is the variance of the ith meter's measurement noise, for $i = 1, 2, \ldots, m$.

Bad Data Detection

- bad measurements may be incidental, accidental or malicious
- techniques to detect bad measurements and then remove them from state estimation have been developed; exploit the fact that there is a statistical inconsistency amongst good and bad measurements
- exploit measurement residual:

$$\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}$$

where
$\mathbf{z} \equiv$ observed measurements and
$\mathbf{H}\hat{\mathbf{x}} \equiv$ estimated measurements

Bad Data Detection

- Using the L_2-norm of the residual and comparing it to a threshold:
  - No bad data detected if:

$$\|\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}\| \leq \tau$$

  - Bad data detected if:

$$\|\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}\| > \tau$$

Bad Data Detection

Q: Can $\|\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}\| > \tau$ detect false data inject attacks?

A: Not always.

Attacks of Interest to Power Systems Community

Attacks on:

- timeliness of measurement data: attacks could involve disrupting communications routing, applying a denial-of-service attack on a critical communication link, flooding the network with bogus data to create congestion and slow down communications, etc.
- accuracy of measurement data: attacks could involve injecting fabricated measurement readings into the network, modifying measurement readings, changing the timestamp of measurement readings, etc.

Attack Model

Two goals:

- Random attack: attacker aims to find any attack vector (measurement bias) as long as it can result in a wrong estimation of $\mathbf{x}$
- Targeted attack: attacker aims to find an attack vector that can inject a *specific* error into certain state variables (and either no error or any possible error into the remaining state variables)

Note: the random attack is easier to apply, but the targeted attack has the potential to cause more damage; trade-off between ease of attack and impact

Attack Model

Two restrictions on resources:

- Limited Access to Meters: opponent can only access and inject attack data into specific predefined meters; for example, some meters may be more susceptible to attack due to proximity or (lack of) physical protection
- Limited Resources to Compromise Meters: opponent can access up to any k meters (within the total set of meters); attacker is limited in the resources required to compromise meters, but is free to select which meters are compromised

Attack Model

- The vector of observed measurements that may contain malicious data is given by

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

where $\mathbf{z} = (z_1, z_2, \ldots, z_m)^T$ is a vector of original measurements and $\mathbf{a} = (a_1, a_2, \ldots, a_m)^T$ is the attack vector which is added to the original meter readings

- If $a_i \neq 0$, then the ith meter has been compromised.
- The original measurement z_i has been replaced with the phony $z_{a_i} = z_i + a_i$.

Impact on State Estimate

- Let $\hat{\mathbf{x}}$ be the state estimate of x using the original measurement $\mathbf{z}$.
- Let $\hat{\mathbf{x}}_{bad}$ be the state estimate of x using the malicious measurement $\mathbf{z}_a = \mathbf{z} + \mathbf{a}$.
- We can represent

$$\hat{\mathbf{x}}_{bad} = \hat{\mathbf{x}} + \mathbf{c}$$

where $\mathbf{c} \neq \mathbf{0}$ and is called the estimation error injected by the opponent

Theorem 1 (Liu et al., 2009)

Suppose the original measurements $\mathbf{z}$ can pass the bad data detection. The malicious measurements $\mathbf{z}_a = \mathbf{z} + \mathbf{a}$ can pass the bad measurement detection if $\mathbf{a}$ is a linear combination of column vectors of $\mathbf{H}$; i.e., $\mathbf{a} = \mathbf{H}\mathbf{c}$.

Proof: Since $\mathbf{z}$ can pass the detection, we can assume $\|\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}\| \leq \tau$ and:

$$\begin{aligned} \|\mathbf{z}_a - \mathbf{H}\mathbf{x}_{bad}\| &= \|\mathbf{z} + \mathbf{a} - \mathbf{H}(\hat{\mathbf{x}} + \mathbf{c})\| = \|\mathbf{z} + \mathbf{a} - \mathbf{H}\hat{\mathbf{x}} - \mathbf{H}\mathbf{c}\| \\ &= \|\mathbf{z} - \mathbf{H}\hat{\mathbf{x}} + \underbrace{\mathbf{a} - \mathbf{H}\mathbf{c}}_{=0}\| = \|\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}\| \leq \tau \end{aligned}$$

Therefore, $\mathbf{z}_a$ can also pass bad data detection.

False Data Injection Attack

Definition: False Data Injection Attack

- attack in which the attack vector $\mathbf{a}$ equals $\mathbf{H}\mathbf{c}$, where $\mathbf{c}$ is an arbitrary non-zero vector.

- A False Data Injection Attack allows an opponent
  - to bypass bad data detection
  - while changing the result of state estimation by effectively corrupting the measurements.

False Data Injection Attack

$$\mathbf{a} = \mathbf{H}\mathbf{c} = \sum_{j=1}^{n} c_j \mathbf{h}_j$$

- Let $\mathbf{H} = (\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n)$ where $\mathbf{h}_i = (h_{1,i}, \ldots, h_{m,i})^T$ is the ith column vector of $\mathbf{H}$.
- Therefore, in a false data injection attack $\mathbf{a}$ is a linear combination of the column vectors of $\mathbf{H}$.

Attack Model

Two restrictions on resources:

- Limited Access to Meters: opponent can only access and inject attack data into specific predefined meters; for example, some meters may be more susceptible to attack due to proximity or (lack of) physical protection

- Limited Resources to Compromise Meters: opponent can access up to any k meters (within the total set of meters); attacker is limited in the resources required to compromise meters, but is free to select which meters are compromised

Limited Access to Meters

Scenario I: Limited Access to Meters

- opponent can only access and inject attack data into specific predefined meters; for example, some meters may be more susceptible to attack due to proximity or (lack of) physical protection
- assume attacker has access to k specific meters from the set:

$$\mathcal{I}_m = \{i_1, i_2, \ldots, i_k\}$$

- Thus, attacker can modify measurement z_{i_j} where $i_j \in \mathcal{I}_m$.

False Data Injection Attack in Scenario I

- To launch a false data injection attack, the attack vector $\mathbf{a}$ must obey the following restrictions:

1. $\mathbf{a} \neq \mathbf{0}$.
2. $\mathbf{a} = (a_1, \ldots, a_m)^T$ such that $a_i = 0$ for $i \notin \mathcal{I}_m$.
3. $\mathbf{a}$ is a linear combination of the column vectors of $\mathbf{H}$; that is, $\mathbf{a} = \mathbf{Hc}$ for any non-zero vector $\mathbf{c}$.

Scenario I: Limited Access to Meters

- Random false data injection attack: The state estimation error vector $\mathbf{c}$ can be of any value; that is, as long as a valid attack vector $\mathbf{a}$ can be found, then this attack is fulfilled.
- Targeted false data injection attack – Constrained Case: The state estimation error vector $\mathbf{c}$ has to be a specific value for a certain set of state elements and zero for the remaining state elements.
- Targeted false data injection attack – Unconstrained Case: The state estimation error vector $\mathbf{c}$ has to be a specific value for a certain set of state elements and any value for the remaining state elements.

Random False Data Injection Attack

Random False Data Injection Attack

- $\mathbf{a} = (a_1, \ldots, a_m)^T = \mathbf{Hc}$ for any non-zero vector $\mathbf{c}$ with $a_i = 0$ for $i \notin \mathcal{I}_m$.
- The state estimation error vector $\mathbf{c}$ can be of any value; that is, as long as a valid attack vector $\mathbf{a}$ can be found, then this attack is fulfilled.
- IDEA: Transform the problem such that it is independent of $\mathbf{c}$.

- Let $\mathbf{P} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$.
- Let $\mathbf{B} = \mathbf{P} - \mathbf{I}$.
- Consider

$$\begin{aligned}
\mathbf{a} &= \mathbf{Hc} \\
\mathbf{Pa} &= \mathbf{PHc} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{Hc} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}(\mathbf{H}^T\mathbf{H})\mathbf{c} = \mathbf{Hc} \\
\mathbf{Pa} &= \mathbf{Hc} = \mathbf{a} \\
\mathbf{Pa} - \mathbf{a} &= \mathbf{0} \\
(\mathbf{P} - \mathbf{I})\mathbf{a} &= \mathbf{0} \\
\mathbf{Ba} &= \mathbf{0}
\end{aligned}$$

Random False Data Injection Attack

Therefore,

$$\mathbf{a} = \mathbf{Hc} \iff \mathbf{Ba} = \mathbf{0}$$

where $\mathbf{c} \in \mathbb{R}^n$.

A random false data injection attack can be constructed if an attack vector $\mathbf{a}$ is found that fulfills:

- $\mathbf{a} \neq \mathbf{0}$
- $\mathbf{Ba} = \mathbf{0}$.
- $\mathbf{a} = (0, \ldots, 0, a_{i_1}, 0, \ldots, 0, a_{i_2}, 0, \ldots, 0, a_{i_k}, 0, \ldots, 0)^T$ where $a_{i_j} \in \mathbb{R}$ and $\mathcal{I}_m = \{i_1, i_2, \ldots, i_k\}$.

This can be further reduced as follows:

Let $\mathbf{a} = (0,\ldots,0,a_{i_1},0,\ldots,0,a_{i_2},0,\ldots,0,a_{i_k},0,\ldots,0)^T$.
Let $\mathbf{B} = (\mathbf{b}_1,\ldots,\mathbf{b}_m)$ where $\mathbf{b}_i$ is a $m \times 1$ column vector for $i = 1,\ldots,m$.

$$\mathbf{B}\mathbf{a} =$$
$$(\ldots,\mathbf{b}_{i_1},\ldots,\mathbf{b}_{i_2},\ldots,\mathbf{b}_{i_k},\ldots)(0,\ldots,0,a_{i_1},0,\ldots,0,a_{i_2},0,\ldots,0,a_{i_k},0,\ldots,0)^T$$
$$= \underbrace{(\mathbf{b}_{i_1},\mathbf{b}_{i_2},\ldots,\mathbf{b}_{i_k})}_{=\mathbf{B}'}\underbrace{(a_{i_1},a_{i_2},\ldots,a_{i_k})}_{=\mathbf{a}'}^T = \mathbf{B}'\mathbf{a}'$$

Random False Data Injection Attack

Therefore,
$$\mathbf{B}\mathbf{a} = \mathbf{0} \quad \Longleftrightarrow \quad \mathbf{B}'\mathbf{a}' = \mathbf{0}$$

- Note that $\mathbf{a} = \mathbf{0}$ is a solution to the above equations, but it is not a valid attack vector.
- For there to be a non-zero, solution $\mathbf{B}'$ must be rank deficient.

Dimensionality

- Recall, $\mathbf{P} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$ and $\mathbf{B} = \mathbf{P} - \mathbf{I}$.
- Therefore, the dimensions of $\mathbf{B}$ is equal to the dimensions of $\mathbf{P}$:

$$\mathbf{H}_{(m\times n)}((\mathbf{H}^T)_{(n\times m)}\mathbf{H}_{(m\times n)})^{-1}(\mathbf{H}^T)_{(n\times m)} = \mathbf{P}_{m\times m}$$

- $\mathbf{B}$ therefore has dimensions: $m \times m$.
- $\mathbf{B}'$ therefore has dimensions: $m \times k$; since we are removing all columns but k from the set $\mathcal{I}_m = \{i_1, i_2, \ldots, i_k\}$.

Dimensionality and Rank

- Note: If the rank$(\mathbf{B}') = k$, then $\mathbf{a}' = \mathbf{0}$ is a unique solution and no attack vector exists.
- It is possible that an attack vector does not exist if k is too small; that is, the set of meters that have been compromised is too small to enable an attack that will not be detected by bad data detection.
- It can be shown, however, that an attack vector exists if $k \geq m - n + 1$ ...

Theorem 2 (Liu et al., 2009)

If the attacker can compromise k specific meters, where $k \geq m - n + 1$, there always exist attack vectors $\mathbf{a} = \mathbf{Hc}$ such that $\mathbf{a} \neq \mathbf{0}$ and $a_i = 0$ for $i \notin \mathcal{I}_m$.

Proof:
Please see paper for proof.

- makes use of matrix projection theory, rank theory and notions of eigenvalues

Constrained Targeted False Data Injection Attack

Targeted False Data Injection Attack – Constrained

- The state estimation error vector $\mathbf{c}$ has to be a specific value for a certain set of state elements and <u>zero</u> for the remaining state elements.
- Let $\mathcal{I}_v = \{j_1, j_2, \ldots, j_r\}$ where $r < n$ denote the set of indices of the r <u>target state variables</u> chosen by opponent.
  - That is, the opponent has selected to add specific biases only to $x_{j_1}, x_{j_2}, \ldots, x_{j_r}$.

Recall the salient relationships:

- $\hat{\mathbf{x}} = (\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}\mathbf{z}$
- $\mathbf{z}_a = \mathbf{z} + \mathbf{a}$
- $\hat{\mathbf{x}}_{bad} = (\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}\mathbf{z}_a$
- If $\mathbf{a} = \mathbf{Hc}$, then $\hat{\mathbf{x}}_{bad} = \hat{\mathbf{x}} + \mathbf{c}$:

$$\begin{aligned}
\hat{\mathbf{x}}_{bad} &= (\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}\mathbf{z}_a \\
&= (\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}(\mathbf{z} + \mathbf{a}) \\
&= (\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}\mathbf{z} + (\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}\mathbf{a} \\
&= (\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}\mathbf{z} + (\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}(\mathbf{H}\mathbf{c}) \\
&= \underbrace{(\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}\mathbf{z}}_{=\hat{\mathbf{x}}} + \underbrace{(\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}(\mathbf{H}^T\mathbf{W}\mathbf{H})}_{=\mathbf{I}}\mathbf{c} \\
&= \hat{\mathbf{x}} + \mathbf{c}
\end{aligned}$$

Targeted False Data Injection Attack – Constrained

Overall, the attack vector $\mathbf{a}$ must be constructed such that:

- $\hat{\mathbf{x}}_{bad} = \hat{\mathbf{x}} + \mathbf{c}$; in other words, $\mathbf{a} = \mathbf{H}\mathbf{c}$
- $\mathbf{c} = (0, \ldots, 0, c_{j_1}, 0, \ldots, 0, c_{j_2}, 0, \ldots, 0, c_{j_r}, 0, \ldots, 0)^T$
- $\mathbf{a} = (0, \ldots, 0, a_{i_1}, 0, \ldots, 0, a_{i_2}, 0, \ldots, 0, a_{i_k}, 0, \ldots, 0)^T$
- $\mathbf{a} \neq \mathbf{0}$

Note: It may not be possible that such an $\mathbf{a}$ exists!

Targeted False Data Injection Attack – Constrained

To determine if a constrained attack exists:

1. Compute: $\mathbf{a} = \mathbf{H}\mathbf{c}$.
2. Check whether: $a_i \neq 0$ for $i \notin \mathcal{I}_m$.
   - If true, then biases need to be added to measurement devices that haven't been corrupted. Therefore, it is not possible to successfully construct or apply a constrained targeted false data injection attack.
   - Otherwise, the constrained targeted false data inject attack exists and the corresponding attack vector is:

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

Unconstrained Targeted False Data Injection Attack

Targeted False Data Inj Attack – Unconstrained

- The state estimation error vector $\mathbf{c}$ has to be a specific value for a certain set of state elements and any value for the remaining state elements.
- Let $\mathcal{I}_v = \{j_1, j_2, \ldots, j_r\}$ where $r < n$ denote the set of indices of the r target state variables chosen by opponent.
  - That is, the opponent has selected to add specific biases to $x_{j_1}, x_{j_2}, \ldots, x_{j_r}$.
  - The other states may or may not exhibit biases.

Targeted False Data Inj Attack – Unconstrained

Overall, the attack vector $\mathbf{a}$ must be constructed such that:

- $\hat{\mathbf{x}}_{bad} = \hat{\mathbf{x}} + \mathbf{c}$; in other words, $\mathbf{a} = \mathbf{H}\mathbf{c}$
- $\mathbf{c} = (x_{1a}, \ldots, x_{1b}, c_{i_1}, x_{2a}, \ldots, x_{2b}, c_{i_2}, x_{3a}, \ldots, x_{3b}, c_{i_r}, x_{(r+1)a}, \ldots, x_{(r+1)b})^T$ where $x_{ns} \in \mathbb{R}$.
- $\mathbf{a} = (0, \ldots, 0, a_{i_1}, 0, \ldots, 0, a_{i_2}, 0, \ldots, 0, a_{i_k}, 0, \ldots, 0)^T$
- $\mathbf{a} \neq \mathbf{0}$

- Let $\mathbf{H}_s = (\mathbf{h}_{j_1}, \ldots, \mathbf{h}_{j_{n-r}})$ where $j_t \notin \mathcal{I}_v$ for $1 \leq t \leq n-r$.
- Let $\mathbf{c}_s = (c_{j_1}, \ldots, c_{j_{n-r}})^T$ where $j_t \notin \mathcal{I}_v$ for $1 \leq t \leq n-r$.
- Let $\mathbf{b} = \sum_{j \in \mathcal{I}_v} \mathbf{h}_j c_j$.
- Let $\mathbf{P}_s = \mathbf{H}_s(\mathbf{H}_s^T\mathbf{H}_s)^{-1}\mathbf{H}_s^T$.
- Let $\mathbf{B}_s = \mathbf{P}_s - \mathbf{I}$.
- Let $\mathbf{y} = \mathbf{B}_s\mathbf{b}$.

$$\begin{aligned}
\mathbf{a} &= \mathbf{H}\mathbf{c} \\
&= \sum_{i \notin \mathcal{I}_v} \mathbf{h}_i c_i + \sum_{j \in \mathcal{I}_v} \mathbf{h}_j c_j = \mathbf{H}_s\mathbf{c}_s + \mathbf{b} \quad \Longleftrightarrow \quad \mathbf{H}_s\mathbf{c}_s = \mathbf{a} - \mathbf{b} \\
\mathbf{P}_s\mathbf{a} &= \mathbf{P}_s(\mathbf{H}_s\mathbf{c}_s + \mathbf{b}) = \mathbf{P}_s\mathbf{H}_s\mathbf{c}_s + \mathbf{P}_s\mathbf{b} = \mathbf{H}_s(\mathbf{H}_s^T\mathbf{H}_s)^{-1}\mathbf{H}_s^T\mathbf{H}_s\mathbf{c}_s + \mathbf{P}_s\mathbf{b} \\
&= \mathbf{H}_s\underbrace{(\mathbf{H}_s^T\mathbf{H}_s)^{-1}}\underbrace{(\mathbf{H}_s^T\mathbf{H}_s)}\mathbf{c}_s + \mathbf{P}_s\mathbf{b} = \mathbf{H}_s\mathbf{c}_s + \mathbf{P}_s\mathbf{b} = \mathbf{a} - \mathbf{b} + \mathbf{P}_s\mathbf{b}
\end{aligned}$$

Therefore,

$$(\mathbf{P}_s - \mathbf{I})\mathbf{a} = (\mathbf{P}_s - \mathbf{I})\mathbf{b} \quad \Longleftrightarrow \quad \mathbf{B}_s\mathbf{a} = \mathbf{B}_s\mathbf{b} \quad \Longleftrightarrow \quad \mathbf{B}_s\mathbf{a} = \mathbf{y}$$

Targeted False Data Inj Attack – Unconstrained

Therefore,

$$\mathbf{a} = \mathbf{H}\mathbf{c} \quad \Longleftrightarrow \quad \mathbf{B}_s\mathbf{a} = \mathbf{y}$$

- Since $\mathbf{a} = (0, \ldots, 0, a_{i_1}, 0, \ldots, 0, a_{i_2}, 0, \ldots, 0, a_{i_k}, 0, \ldots, 0)^T$.
- Let $\mathbf{B}'_s = (\mathbf{b}_{s_{i_1}}, \ldots, \mathbf{b}_{s_{i_k}})$.
- Let $\mathbf{a}' = (a_{i_1}, \ldots, a_{i_k})^T$.

$$\mathbf{a} = \mathbf{H}\mathbf{c} \quad \Longleftrightarrow \quad \mathbf{B}_s\mathbf{a} = \mathbf{y} \quad \Longleftrightarrow \quad \mathbf{B}'_s\mathbf{a}' = \mathbf{y}$$

Targeted False Data Inj Attack – Unconstrained

It can be shown that:

- if $\text{rank}(\mathbf{B}'_s) = \text{rank}(\mathbf{B}'_s|\mathbf{y})$, then there exist an infinite number of solutions for $\mathbf{a}'$ given by:

  $$\mathbf{a}' = \mathbf{B}'^{-}_s\mathbf{y} + (\mathbf{I} - \mathbf{B}'^{-}_s\mathbf{B}_s)\mathbf{d}$$

  where $\mathbf{B}'^{-}_s$ is the 1-inverse of matrix $\mathbf{B}'_s$; that is,

  $$\mathbf{B}'_s\mathbf{B}'^{-}_s\mathbf{B}'_s = \mathbf{B}'_s$$

  and $\mathbf{d}$ is an arbitrary non-zero $k \times 1$ vector.

- otherwise there is no solution for $\mathbf{a}'$.

Why?

Note: $\text{rank}(\mathbf{B}_s') = \text{rank}(\mathbf{B}_s'|\mathbf{y})$, is required to ensure a consistent system of m equations and n unknowns.

- For there to be a solution for $\mathbf{a}'$ then, $\mathbf{y}$ must be in the range of $\mathbf{B}_s'$ (a.k.a. column-space of $\mathbf{B}_s'$)

$$\mathbf{B}_s'\mathbf{a}' = \mathbf{y}$$

Why?

Note: Since $\mathbf{B}_s'\mathbf{B}_s'^{-}\mathbf{B}_s' = \mathbf{B}_s'$,

$$\begin{aligned}
\mathbf{B}_s'\mathbf{a}' &= \mathbf{y}\\
\mathbf{B}_s'\mathbf{B}_s'^{-}\mathbf{B}_s'\mathbf{a}' &= \mathbf{y}\\
\mathbf{B}_s'\mathbf{B}_s'^{-}\underbrace{\mathbf{B}_s'\mathbf{a}'}_{=\mathbf{y}} &= \mathbf{y}\\
\mathbf{B}_s'\mathbf{B}_s'^{-}\mathbf{y} &= \mathbf{y}\\
\mathbf{B}_s'\underbrace{\mathbf{B}_s'^{-}\mathbf{y}}_{=\mathbf{a}'} &= \mathbf{y}
\end{aligned}$$

Therefore, one solution is given by:

$$\mathbf{a}' = \mathbf{B}_s'^{-}\mathbf{y}$$

Why

Note: If $\mathbf{a}' = \mathbf{B}_s'^{-}\mathbf{y}$ is a solution to $\mathbf{B}_s'\mathbf{a}' = \mathbf{y}$, then so is

$$\mathbf{a}' = \mathbf{B}_s'^{-}\mathbf{y} + \underbrace{(\mathbf{I} - \mathbf{B}_s'^{-}\mathbf{B}_s')\mathbf{d}}_{\in\ nullspace(\mathbf{B}_s'^{-})}$$

for an arbitrary vector $\mathbf{d}$ because $(\mathbf{I} - \mathbf{B}_s'^{-}\mathbf{B}_s')$ spans the nullspace of $\mathbf{B}_s'$:

$$\begin{aligned}
\mathbf{B}_s'(\mathbf{I} - \mathbf{B}_s'^{-}\mathbf{B}_s') &= \mathbf{B}_s' - \mathbf{B}_s'\mathbf{B}_s'^{-}\mathbf{B}_s'\\
&= \mathbf{B}_s' - \underbrace{\mathbf{B}_s'\mathbf{B}_s'^{-}\mathbf{B}_s'}_{=\mathbf{B}_s'}\\
&= \mathbf{B}_s' - \mathbf{B}_s' = \mathbf{0}
\end{aligned}$$

Therefore, if $\text{rank}(\mathbf{B}_s') = \text{rank}(\mathbf{B}_s'|\mathbf{y})$, an infinite number of attack vectors can be constructed as follows:

$$\mathbf{a}' = \mathbf{B}_s'^{-}\mathbf{y} + (\mathbf{I} - \mathbf{B}_s'^{-}\mathbf{B}_s)\mathbf{d}$$

where $\mathbf{d}$ is an arbitrary $k \times 1$ vector such that $\mathbf{a}' \neq 0$.

Limited Resources to Compromise Meters

Scenario II: Limited Resources to Compromise Meters

- opponent can access up to any k meters (within the total set of meters); attacker is limited in the resources required to compromise meters, but is free to select which meters are compromised
- if up to any k meters can be compromised, an opponent must find an attack vector $\mathbf{a} \neq 0$ with at most k non-zero elements.
- A $m \times 1$ vector with at most k non-zero elements is denoted a k-sparse vector.

Scenario II: Limited Resources to Compromise Meters

- In comparison to Scenario I: Limited Access to Meters, the construction of this attack is easier.
- Existence of attack vector in Scenario I $\Longrightarrow$ Existence of attack vector in Scenario II

False Data Injection Attack in Scenario II

- To launch a false data injection attack, the attack vector $\mathbf{a}$ must obey the following restrictions:

1. $\mathbf{a} \neq \mathbf{0}$.
2. $\mathbf{a}$ must be a k-sparse vector.
3. $\mathbf{a}$ is a linear combination of the column vectors of $\mathbf{H}$; that is, $\mathbf{a} = \mathbf{Hc}$ for any non-zero vector $\mathbf{c}$.

Scenario II: Limited Resources to Compromise Meters

- Random false data injection attack: The state estimation error vector **c** can be of any value; that is, as long as a valid attack vector **a** can be found, then this attack is fulfilled.
- Targeted false data injection attack – Constrained Case: The state estimation error vector **c** has to be a specific value for a certain set of state elements and zero for the remaining state elements.
- Targeted false data injection attack – Unconstrained Case: The state estimation error vector **c** has to be a specific value for a certain set of state elements and any value for the remaining state elements.

Random False Data Injection Attack

Random False Data Injection Attack

- Constructing attack vectors is not so prescriptive due to the constraint that **a** is k-sparse.
- Brute force search for attack vectors will require searching through all possible k-sparse vectors which is time consuming as there are on the order of $\binom{m}{k}$ possible k-sparse vectors.
- IDEA: since **a** is a linear combination of column vectors of **H**, then conduct column transformations on **H** to reduce the number of non-zero elements in each row; use the column vectors with no more than k non-zero entries as attack vectors.

Constrained Targeted False Data Injection Attack

Targeted False Data Injection Attack – Constrained

- The state estimation error vector $\mathbf{c}$ has to be a specific value for a certain set of state elements and zero for the remaining state elements.
- Let $\mathcal{I}_v = \{j_1, j_2, \ldots, j_r\}$ where $r < n$ denote the set of indices of the r target state variables chosen by opponent.
  - That is, the opponent has selected to add specific biases only to $x_{j_1}, x_{j_2}, \ldots, x_{j_r}$.

Targeted False Data Injection Attack – Constrained

Overall, the attack vector $\mathbf{a}$ must be constructed such that:

- $\hat{\mathbf{x}}_{bad} = \hat{\mathbf{x}} + \mathbf{c}$; in other words, $\mathbf{a} = \mathbf{H}\mathbf{c}$
- $\mathbf{c} = (0, \ldots, 0, c_{j_1}, 0, \ldots, 0, c_{j_2}, 0, \ldots, 0, c_{j_r}, 0, \ldots, 0)^T$
- $\mathbf{a}$ is a k-sparse vector.
- $\mathbf{a} \neq \mathbf{0}$

Note: It may not be possible that such an $\mathbf{a}$ exists!

Targeted False Data Injection Attack – Constrained

To determine if a constrained attack exists:

1. Compute: $\mathbf{a} = \mathbf{H}\mathbf{c}$.
2. Check whether: $\mathbf{a}$ is k-sparse.
   - If true, the constrained targeted false data inject attack exists and the corresponding attack vector is:
   $$\mathbf{a} = \mathbf{H}\mathbf{c}$$
   - Otherwise, biases need to be added to more than k measurement devices, which is beyond the opponent's resources. Therefore, it is not possible to successfully construct or apply a constrained targeted false data injection attack.

Unconstrained Targeted False Data Injection Attack

Targeted False Data Inj Attack – Unconstrained

- The state estimation error vector $\mathbf{c}$ has to be a specific value for a certain set of state elements and any value for the remaining state elements.
- Let $\mathcal{I}_v = \{j_1, j_2, \ldots, j_r\}$ where $r < n$ denote the set of indices of the r target state variables chosen by opponent.
  - That is, the opponent has selected to add specific biases to $x_{ji_1}, x_{j_2}, \ldots, x_{j_r}$.
  - The other states may or may not exhibit biases.

Targeted False Data Inj Attack – Unconstrained

Overall, the attack vector $\mathbf{a}$ must be constructed such that:

- $\hat{\mathbf{x}}_{bad} = \hat{\mathbf{x}} + \mathbf{c}$; in other words, $\mathbf{a} = \mathbf{Hc} \iff \mathbf{B}_s\mathbf{a} = \mathbf{y}$
- $\mathbf{c} = (x_{1a}, \ldots, x_{1b}, c_{i_1}, x_{2a}, \ldots, x_{2b}, c_{i_2}, x_{3a}, \ldots, x_{rb}, c_{i_r}, x_{(r+1)a}, \ldots, x_{(r+1)b})^T$ where $x_{ns} \in \mathbb{R}$.
- $\mathbf{a}$ must be a k-sparse vector.
- $\mathbf{a} \neq \mathbf{0}$

A search procedure can be used such as the Matching Pursuit algorithm (employed by authors).

Simulations

- Consider the **H** matrices for standard IEEE test systems:
  - IEEE 9-bus
  - IEEE 14-bus
  - IEEE 30-bus
  - IEEE 118-bus
  - IEEE 300-bus

Simulations

- Evaluation metrics:
  - probability that opponent can construct an attack vector:

    $$\frac{\text{number of successful trials}}{\text{number of trials}}$$

    versus percentage of meters under attackers' control:

    $$\frac{k}{\text{m}}$$
  - execution time to construct attack vector or conclude it is infeasible

References

A. Abur and A.G. Expósito, *Power System State Estimation: Theory and Implementation*, CRC Press, 2004.

Y. Liu, P. Ning and M.K. Reiter, "False Data Injection Attacks against State Estimation in Electric Power Grids," *Proc. 16th ACM Conference on Computer and Communications Security (CCS '09)* , Chicago, IL, pp. 21-32, November 2009.

Y. Liu, P. Ning and M.K. Reiter, "Generalized False Data Injection Attacks against State Estimation in Electric Power Grids," *ACM Transactions on Information and System Security*, vol. 14, no. 1, May 2011.

■