False Data Injection Attacks against State Estimation in Electric Power Grids

Yao Liu and Peng Ning
Department of Computer Science
North Carolina State University

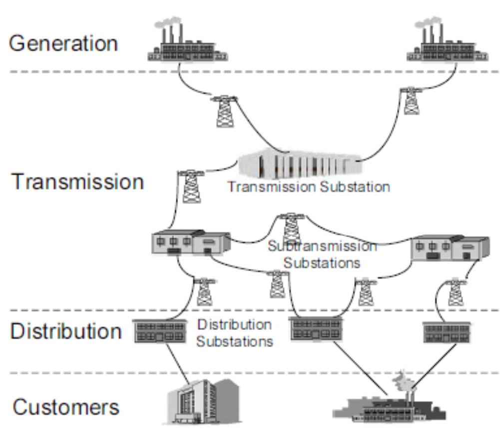
Michael K. Reiter
Department of Computer Science
University of North Carolina, Chapel Hill

Published on *CCS'09*

Presenter: Jinkyu Koo
Sep. 15, 2010

PURDUE UNIVERSITY

Introduction (1/2)



- System monitoring is necessary to ensure the reliable operation of power grids.
- The meter measurements
  - Bus voltages, bus real and reactive power injections, and branch reactive power flows in every subsystem of a power grid.
- These measurements are typically transmitted to a control center.

PURDUE UNIVERSITY

Introduction (2/2)

- State estimation
  - the process of estimating unknown state variables in a power grid based on the meter measurements.
- The output of state estimation is used to control the power grid components
  - e.g., to increase the yield of a power generator to maintain the reliable operation even if some faults may occur next.
- It is possible for an attacker to compromise meters to introduce malicious measurements.
- Bad measurements may result in catastrophic consequences such as blackouts in large geographic areas.

PURDUE UNIVERSITY

State estimation (1/2)

$$\mathbf{z} = \underline{\mathbf{h}(\mathbf{x})} + \mathbf{e}, \qquad \text{to find an estimate } \hat{\mathbf{x}} \text{ of } \mathbf{x}$$

state variables $\mathbf{x} = (x_1, x_2, ..., x_n)^T$

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

measurement errors $\mathbf{e} = (e_1, e_2, ..., e_m)^T$

$\mathbf{h}(\mathbf{x}) = (h_1(x_1, x_2, ..., x_n), ..., h_m(x_1, x_2, ..., x_n))^T$

Nonlinear
= difficult!!!

Approximate as a linear model

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

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

less accurate, but simpler

PURDUE UNIVERSITY

State estimation (2/2)

- Least-square estimation

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

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

σ_i^2 is the variance of the i-th meter $(1 \le i \le m)$

the presence of bad measurements is assumed if $\|\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}\| > \tau$

measurement residual

Of course, with a fixed probability of false alarm

How can bad guys avoid being detected?

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

$\mathbf{z} = (z_1, ..., z_m)^T$ original measurements

$\mathbf{a} = (a_1, ..., a_m)^T$ malicious data added

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

$\mathbf{c}$ estimation error injected by the attacker

Suppose the original measurements can pass the bad measurement detection

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

if the attacker can use $\mathbf{Hc}$ as the attack vector $\mathbf{a}$ (i.e., $\mathbf{a} = \mathbf{Hc}$)

$$\begin{aligned} \|\mathbf{z_a} - \mathbf{H}\hat{\mathbf{x}}_{\mathbf{bad}}\| &= \|\mathbf{z} + \mathbf{a} - \mathbf{H}(\hat{\mathbf{x}} + \mathbf{c})\| \\ &= \|\mathbf{z} - \mathbf{H}\hat{\mathbf{x}} + (\mathbf{a} - \mathbf{Hc})\| \\ &= \|\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}\| \le \tau. \end{aligned}$$

undetected!

False data injection attacks

- Assume the attacker knows the matrix **H** of the target power system, and can manipulate some of meter measurements.
  - Is this possible?
- Random false data injection attacks
  - the attacker aims to find any attack vector as long as it can result in a wrong estimation of state variables.
- Targeted false data injection attacks
  - the attacker aims to find an attack vector that can inject a specific error into certain state variables.

Scenario I – Limited access to meters

- The attacker is restricted to accessing *k* specific meters due to, for example, different physical protection of meters.

$\mathcal{I}_m = \{i_1, ..., i_k\}$ the set of indices of those meters

the attacker needs to find a non-zero attack vector

$$\mathbf{a} = (a_1, ..., a_m)^T \text{ such that } a_i = 0 \text{ for } i \notin \mathcal{I}_m \text{ and}$$

$$\mathbf{a} = \mathbf{Hc}$$

Scenario I: Random false data injection attacks

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

$$\begin{aligned} \mathbf{a} = \mathbf{Hc} \quad &\Leftrightarrow \quad \mathbf{Pa} = \underline{\mathbf{PHc}} \Leftrightarrow \mathbf{Pa} = \underline{\mathbf{H}}\mathbf{c} \Leftrightarrow \mathbf{Pa} = \underline{\mathbf{a}} \\ &\Leftrightarrow \quad \mathbf{Pa} - \mathbf{a} = \mathbf{0} \Leftrightarrow (\mathbf{P} - \mathbf{I})\mathbf{a} = \mathbf{0} \\ &\Leftrightarrow \quad \mathbf{Ba} = \mathbf{0}. \end{aligned}$$

$$\mathbf{a} = (0, ..., 0, a_{i_1}, 0, ..., 0, a_{i_2}, 0, ..., 0, a_{i_k}, 0, ..., 0)^T$$

$$\mathbf{B}' = (\mathbf{b}_{i_1}, ..., \mathbf{b}_{i_k}) \qquad \mathbf{a}' = (a_{i_1}, ..., a_{i_k})^T$$

$$\mathbf{Ba} = \mathbf{0} \Leftrightarrow \mathbf{B}'\mathbf{a}' = \mathbf{0} \qquad [b_1 \quad b_2 \quad b_3]\begin{bmatrix} a_1 \\ 0 \\ a_3 \end{bmatrix} = [b_1 \quad b_3]\begin{bmatrix} a_1 \\ a_3 \end{bmatrix}$$

If # of equations is more than # of unknowns, no solution exists.

Attacker needs to compromise more than a certain number of meters, to inject an error into the state estimation.

Scenario I: Targeted false data injection attacks (1/2)

$$\mathcal{I}_v = \{i_1, ..., i_r\}$$

the set of indexes of the r target state variables

- Constrained Case:

$$\mathbf{c} \quad \begin{cases} \text{the chosen value when } i \in \mathcal{I}_v \\ 0 \text{ when } i \notin \mathcal{I}_v \end{cases}$$

  - Attacker can substitute **c** back into the relation **a = Hc**, and check if $a_i = 0 \text{ for } \forall i \notin \mathcal{I}_m$
  - If yes, the attacker succeeds in constructing the (only) attack vector **a**. Otherwise, the attack is impossible.

Scenario I: Targeted false data injection attacks (2/2)

- Unconstrained Case:
  - the other elements c_j for $j \notin \mathcal{I}_v$ can be any values.

$$\mathbf{c_s} = (c_{j_1}, ..., c_{j_{n-r}})^T$$
$$\mathbf{H_s} = (\mathbf{h}_{j_1}, ..., \mathbf{h}_{j_{n-r}})$$
$$\mathbf{b} = \sum_{j\in\mathcal{I}_v} \mathbf{h}_j c_j$$
$$\mathbf{P_s} = \mathbf{H_s}(\mathbf{H_s^T H_s})^{-1}\mathbf{H_s^T}$$
$$\mathbf{B_s} = \mathbf{P_s} - \mathbf{I}$$
$$\mathbf{y} = \mathbf{B_s b}$$

$$\begin{aligned}
\mathbf{a} = \mathbf{Hc} \quad &\Leftrightarrow \quad \mathbf{a} = \sum_{i\notin\mathcal{I}_v} \mathbf{h}_i c_i + \sum_{j\in\mathcal{I}_v} \mathbf{h}_j c_j = \mathbf{H_s c_s} + \mathbf{b} \\
&\Leftrightarrow \quad \mathbf{P_s a} = \underline{\mathbf{P_s H_s}}\mathbf{c_s} + \mathbf{P_s b} \\
&\Leftrightarrow \quad \mathbf{P_s a} = \underline{\mathbf{H_s}}\mathbf{c_s} + \mathbf{P_s b} \\
&\Leftrightarrow \quad \mathbf{P_s a} = \underline{\mathbf{a} - \mathbf{b}} + \mathbf{P_s b} \\
&\Leftrightarrow \quad (\mathbf{P_s} - \mathbf{I})\mathbf{a} = (\mathbf{P_s} - \mathbf{I})\mathbf{b} \\
&\Leftrightarrow \quad \mathbf{B_s a} = \mathbf{B_s b} \Leftrightarrow \mathbf{B_s a} = \mathbf{y}.
\end{aligned}$$

PURDUE UNIVERSITY

Scenario II – Limited resources to compromise meters

- The attacker is limited in the resources required to compromise meters. For example, the attacker only has resources to compromise up to k meters (out of all the meters).
  - compromise up to k meters.
- Unlike Scenario I, there is no restriction on what meters can be chosen.

PURDUE UNIVERSITY

Scenario II: Random false data injection attacks

- Brute-force approach:
  - attacker may try all possible **a**'s consisting of k unknown elements and $m - k$ zero elements.
- If there exists a non-zero solution of a such that **Ba = 0**, the attacker succeeds in constructing an attack vector. Otherwise, the attack vector does not exist.

Scenario II: Targeted false data injection attacks

- Constrained Case:
  - All elements of **c** are fixed.
  - So the attacker can substitute c into the relation **a = Hc**.
  - If the resulting **a** has k non-zero elements, the attacker succeeds in constructing the attack vector. Otherwise, the attacker fails.
- Unconstrained Case:
  - Attacker needs to find a attack vector **a** of k non-zero elements that satisfies the relation $\mathbf{B_s a} = \mathbf{y}$.
  - Known as a NP complete problem, called Minimum Weight Solution for Linear Equation Problem.

Experimental results

- Configuration of the IEEE test systems, including the IEEE 9-bus, 14-bus, 30-bus, 118-bus, and 300-bus systems.
  - particularly matrix **H**
- MATLAB package for solving power flow problems.
- DELL PC running Windows XP, which has a 3.0 GHz Pentium 4 processor and 1 GB memory.
- Two evaluation metrics:
  - the probability that the attacker can successfully construct an attack vector given the *k* specific meters
    - for each k, we randomly choose k specific meters to attempt an attack vector construction
    - repeat this process 100 times

$$success\ probability\ p_k \text{ as } p_k = \frac{\#\ successful\ trials}{\#\ trials}.$$

  - the execution time required to either construct an attack vector or conclude that the attack is infeasible.

PURDUE UNIVERSITY

Experimental results: Scenario I (1/2)

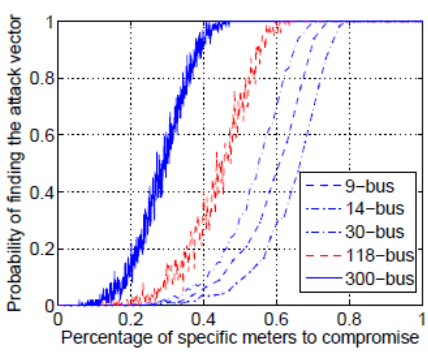


Figure 2: Probability of finding an attack vector for random false data injection attacks

larger systems have higher p_k

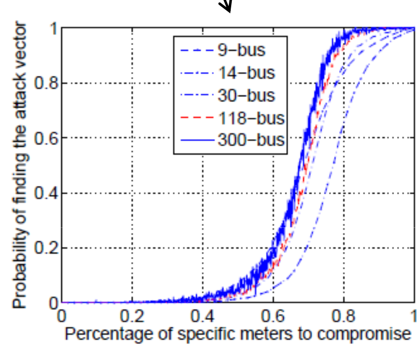


Figure 3: Probability of finding an attack vector for compromising a single state variable in targeted false data injection attacks (unconstrained case)

in general lower than that in Figure 2

PURDUE UNIVERSITY

Experimental results: Scenario I (2/2)

- Constrained case:
  - pick 6 sets of meters for the IEEE 118-bus and 300-bus systems.
  - In each set, there are 350 meters and 700 meters, respectively.
  - check the number of individual target state variables that can be affected by each set of meters without affecting the estimation of the remaining state variables.
  - The results show that the attacker can affect 8–11 and 13–16 individual state variables in the IEEE 118-bus and 300-bus systems, respectively.

Table 1: Timing results in Scenario I (ms)

Test system	Random attack	Targeted attack (unconstrained)
IEEE 9-bus	0.17–2.4	0.21–2.2
IEEE 14-bus	0.16–5.6	0.26–11.3
IEEE 30-bus	0.35–14.9	0.24–31.4
IEEE 118-bus	0.34–867.9	0.42–1,874.5
IEEE 300-bus	0.55–8,549.6	0.73–18,510

Conclusion

- We show that an attacker can take advantage of the configuration of a power system to launch such attacks to bypass the existing techniques for bad measurement detection.
- Security protection of the electric power grid must be revisited when there are potentially malicious attacks.

PURDUE UNIVERSITY