Malicious Data Attacks on Smart Grid State Estimation: Attack Strategies and Countermeasures

Authors: O. Kosut, J. Liyan, R. J. Thomas and L. Tong

Presenter: Daehyun. Choi

Submitted in Partial Fulfillment of the Course Requirements for
ECEN 689: Cyber Security of the Smart Grid
Instructor: Dr. Deepa Kundur

Outline

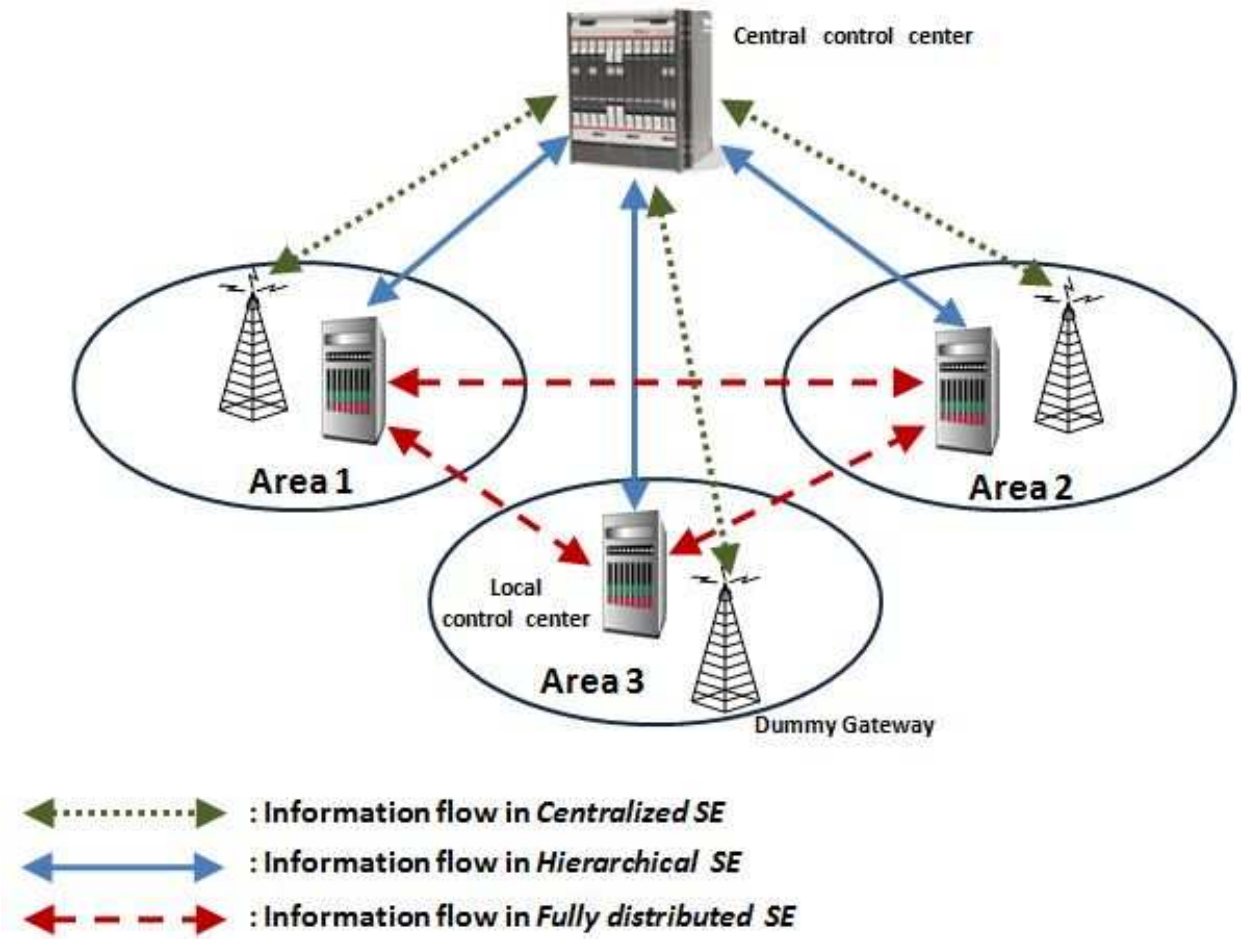
Outline

Power System State Estimation

- **State estimation** is to process redundant measurements to obtain the best estimate of the current state of a power system

- The main task of the the state estimator is conducted in a control center, with the following three functions
  - **Observability analysis**(for unique solution)
  - **State estimation**(using a weighted least square algorithm)
  - **Bad data processing**(consisting of bad data detection and identification)

State Estimation Communication Topology



Motivation and Goal

- If an adversary hacks into the power grid and generates fake meter data, the control center may be misled by the state estimator.

- The main goals of this paper are to
  1. consider the problem of constructing malicious data attack of smart grid state estimation($\rightarrow$ **attacker view**)
     - when the adversary can perform an unobservable attack(**Minimum Size Unobservable Attack**)
     - when the adversary cannot or does not execute an unobservable attack(**Minimum Residue Energy Attack**)
  2. propose the countermeasures that detect the presence of such attacks in **a single central control center**($\rightarrow$ **control center view**)

Related Works

- Y. Liu, P.Ning and M. K. Reiter "False data injection attacks against state estimation in electric power grids " in ACM Conference on Computer and Communications Security, pp.21-32, 2009.

- D. Gorinevsky, S. Boyd and S. Poll "Estimation of faults in DC electrical power systems" Proc. 2009 American Control Conf., pp.4334-4339, June 2009.

- A key difference from the above two works is
  - the state is **random variable** $\longrightarrow$ **Bayesian framework!!**

Outline

A Bayesian Framework and MMSE Estimation

- Assume a linearized DC power flow model

$$z = Hx + a + e$$

where $e \sim \mathcal{N}(0, \Sigma_e)$, $a \in \mathcal{A}_k = \{a \in R^m : \|a\|_0 \leq k\}$.

- When $a = 0$, MMSE estimator of the state vector x is

$$\hat{x}(z) = \arg_{\hat{x}} \min E(\|x - \hat{x}(z)\|^2) = Kz$$

where $K = \Sigma_x H^T (H\Sigma_x H^T + \Sigma_e)^{-1}$.

- With attack vector a, the mean square error is

$$\underbrace{\min_{\hat{x}} E(\|x - \hat{x}(z)\|^2)}_{without \quad attack} + \underbrace{\|Ka\|_2^2}_{with \quad attack}$$

Outline

Unobservable Attacks

According to Liu's paper,

- if there exists a nonzero k-sparse a for which $a = Hc$, then

$$z = Hx + a + e = H(x + c) + e$$

- An attack vector a is **unobservable** if it has the form $a = Hc$.

Theorem 1: A k-sparse attack vector a comprises an unobservable attack if and only if the network becomes unobservable when the k meters associated with the nonzero entries of a are removed from the network.

Remark: Here, the k meters belong to critical k-tuple measurements class.

Minimum Size Unobservable Attacks

When the adversary can perform an unobservable attack,

- The power system is modeled as an undirected graph (V, E)
  - V represents the set of buses and E is the set of transmission lines.
- For a set of lines $\mathcal{A} \subseteq E$, let $g(\mathcal{A})$ and $h(\mathcal{A})$ be the set of meters in $\mathcal{A}$ and be the number of connected components in the graph $(V, E \backslash A)$

Theorem 2: For all $\mathcal{A} \subseteq E$, removing an arbitrary subset of $g(\mathcal{A})$ of size $|g(\mathcal{A})|$-$h(\mathcal{A})$+2 makes the system unobservable. Moreover, the minimum size unobservable attack can be found by minimizing $|g(\mathcal{A})|$-$h(\mathcal{A})$+2 over $\mathcal{A}$.

Remark: By Theorem 2, the adversary can find unobservable attacks. If $|g(\mathcal{A})|$-$h(\mathcal{A})$+2 can be minimized over all sets of edges $\mathcal{A}$, the adversary can also find attacks using as few meters as possible.

Minimum Residue Energy Attack

When the adversary can't perform an unobservable attack,

- The estimation residue error is given by

$$r = Gz = GHx + \underbrace{Ga}_{\text{from attack}} + Ge, \quad G \triangleq I - HK$$

- Indeed, the attacker wish to
  - damage the control center $\rightarrow$ increase $\|Ka\|_2^2$
  - be less detectable at the control center $\rightarrow$ decrease $\|Ga\|_2^2$

- Therefore, we consider the following problem:

$$\min_{a \in \mathcal{A}_k} \|Ga\|_2^2 \quad \text{subject to} \quad \|Ka\|_2^2 \geq C.$$

Outline

Two classical bad data detectors

Based on the residual error $r = z - H\hat{x}$,

- The Chi-Squares test is given by

$$r^T \Sigma_e^{-1} r \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

- The largest normalized residual test is given by

$$\max_i \frac{r_i}{\sigma_{r_i}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

where σ_{r_i} is the standard deviation of the ith residual error r_i.

Generalized Likelihood Ratio Test(GLRT)

- Consider a formulation of the detection problem at the control center.

$$\mathcal{H}_0 : z \sim \mathcal{N}(0, \Sigma_z)$$
$$\mathcal{H}_1 : z \sim \mathcal{N}(a, \Sigma_z), \quad a \in \mathcal{A}_k \backslash 0$$

  where $\Sigma_z \triangleq H\Sigma_x H^T + \Sigma_e$.

- Let $f(z|a)$ be the Gaussian density function with mean a and covariance Σ_z,

$$f(z|a) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma_z|} \exp\left(-\frac{1}{2}(z-a)^T\Sigma_z^{-1}(z-a)\right)$$

- Then, the GLRT is expressed as

$$L(z) \triangleq \frac{\max_{a\in\mathcal{A}_k} f(z|a)}{f(z|a=0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

Generalized Likelihood Ratio Test(GLRT)(cont'd)

- The GLRT is rewritten by

$$\min_{a \in \mathcal{A}_k} a^T \Sigma_z^{-1} a - 2 z^T \Sigma_z^{-1} a \underset{H_1}{\overset{H_0}{\gtrless}} \eta$$

- Finally, the GLRT reduces to solving

$$\begin{aligned} \text{minimize} \quad & a^T \Sigma_z^{-1} a - 2 z^T \Sigma_z^{-1} a \\ \text{subject to} \quad & \|a\|_0 \leq k \\ & (\|a\|_1 \leq \gamma \quad \text{if } k \text{ is larger}). \end{aligned}$$

Simulation Results

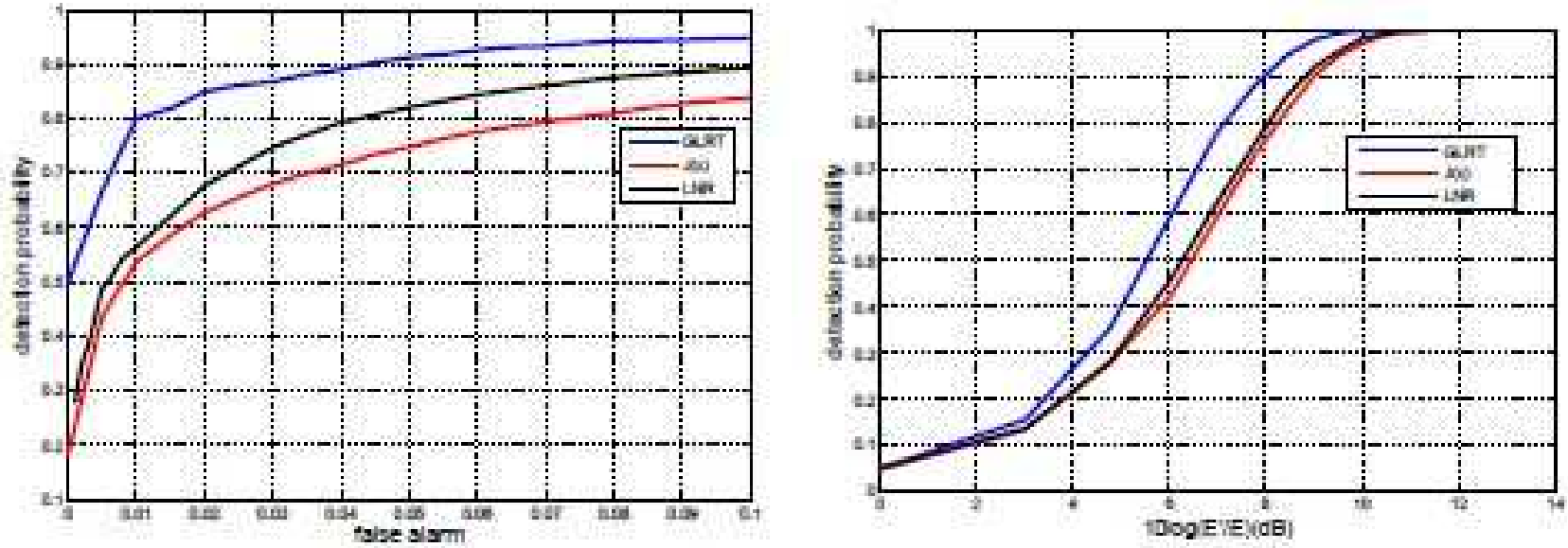


Figure: Above: ROC Performance of GLRT for the 2 sparsity case. MSE with attack is 8db. SNR=10db. Below: AOC Performance of GLRT for the 2 sparsity case. False alarm rate is 0.05. SNR=10dB.

O. Kosut, L. Jia, R. J. Thomas and L. Tong "Malicious Data Attacks on Smart Grid State Estimation: Attack Strategies and Countermeasures " Proc. IEEE International Conference on Smart Grid Communications, Gaithersburg, Maryland, pp.220-225, October 2010.

Outline

Critical Assessment

- Considered malicious data attack problem using a graph theoretic approach(Network observability analysis).
- GLRT shows a good performance for a malicious attack detection, but not for **identification.**
- Needed for a new countermeasure in a (fully) distributed state estimation.
  - a new countermeasure need to be carefully designed considering **error residual spread area.**(Next slide..)

Administrative Area vs Error Residual Spread Area

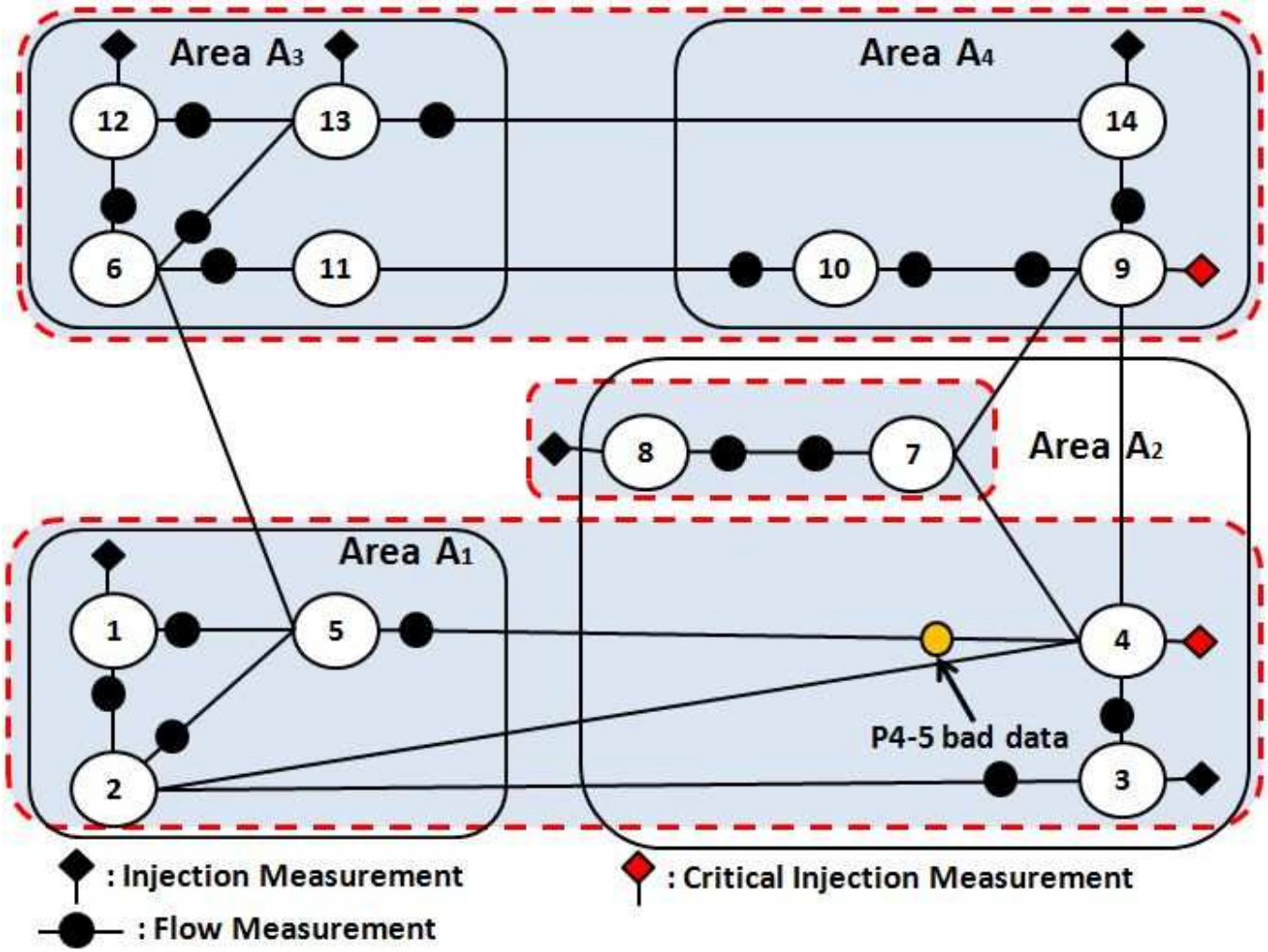


Figure: IEEE 14-bus system

Outline

Conclusions and Future works

Conclusions

- present adversarial strategies for malicious data attacks
- propose countermeasure(GLRT) for the control center

Future works

- propose a novel countermeasure for malicious attack detection and identification
- a novel countermeasure should be implementable for large-scale AC power networks even in a distributed state estimation framework.

References

- O. Kosut, L. Jia, R. J. Thomas and L. Tong "Malicious Data Attacks on Smart Grid State Estimation: Attack Strategies and Countermeasures " Proc. IEEE International Conference on Smart Grid Communications, Gaithersburg, Maryland, pp.220-225, October 2010.

- O. Kosut, L. Jia, R. J. Thomas and L. Tong "On Malicious Data Attacks on Power System State Estimation" Proc. UPEC 2010, Cardiff, Wales, UK, August 2010.

- O. Kosut,"Adversaries In Networks", Ph.D. dissertation, Cornell University, August 2010.