A Survey on False Data Injection Attack and Detection in Smart Grid

Presenter: Yessica Saez

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

Presentation Overview

- Explore and understand the problem of constructing false data injection attacks from the adversary's point of view.
- Present and analyze countermeasures to malicious data attacks developed by different researchers.

Motivation

- Bad data processing represents one of the main tasks of the state estimator.
- Bad data injection attacks need to be analyzed from both the attacker and the operator's point of view.

Background: State Estimation, Traditional Bad Data Detection Techniques, and Malicious Attacks

Linear State Estimation Problem

- The equation relating the measurements $\mathbf{z}$ observed by the control center and the state vector $\mathbf{x}$ is:

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

$\boldsymbol{z} \in \mathbb{R}^{\boldsymbol{m}}$ = sensor measurements for m active power flow branches.

$\mathbf{x} \in \mathbb{R}^{\mathbf{n}}$ = power system state variable that need to be estimated.

$\mathbf{e} \in \mathbb{R}^{\mathbf{m}}$ = random error measurements which are assumed to be jointly Gaussian with covariance matrix $\sum_{\mathrm{e}}$. (i.e. $\mathrm{N}(0, \sum_{\mathrm{e}})$).

$\mathrm{h}(\mathrm{x})$ = power flow model for the transmitted active power in transmission lines.

Linear State Estimation Problem

The linear approximation of the nonlinear relationship:

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

is accurate and can be described as:

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

Where the constant Jacobian mxn matrix H is defined as:

$$\mathrm{H} = \left.\frac{\partial \mathrm{h}(\mathrm{x})}{\partial \mathrm{x}}\right|_{\mathrm{x}=0}$$

Linear State Estimation Problem

- When $m > n$, as is the typical case, state estimation involves solving an overdetermined system of linear equations.

It can be solved as a weighted least square, maximum likelihood criterion and as minimum variance criterion problem. In all the cases the solution to $\mathbf{z} = \mathrm{H}\mathbf{x} + \mathbf{e}$ arrive to the following estimate

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

Where $\boldsymbol{W}$ is a diagonal matrix whose elements are the reciprocal of the variances of the measurement errors.

Traditional Bad Data Detection Techniques

TEXAS A&M UNIVERSITY

-

A traditional approach proposed by power system researchers for detecting bad measurements consists of computing the measurement residual vector defined as:[1]

$$r = \boldsymbol{z} - H\hat{\boldsymbol{x}}$$

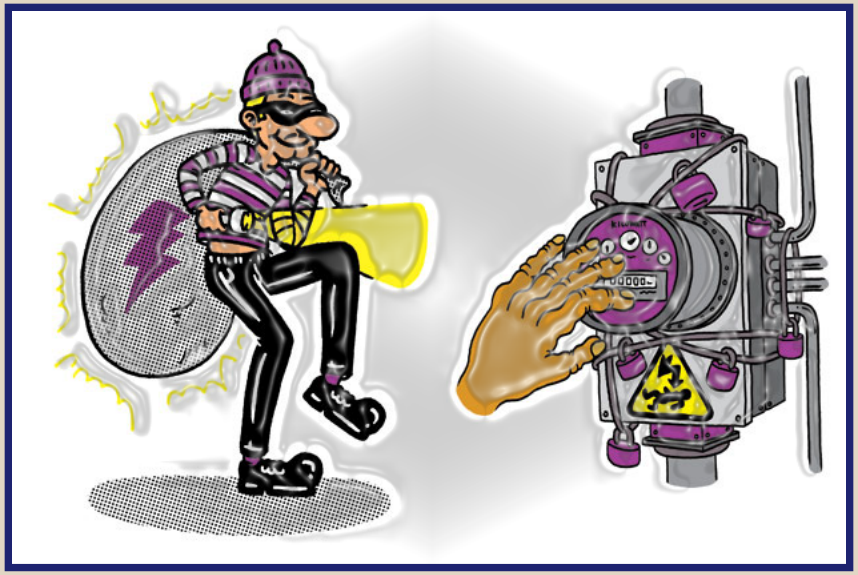
The l_2-norm of the measurement residual vector $\|\boldsymbol{r}\|$ is used to detect gross errors in the measurements. Specifically, $\|\boldsymbol{r}\|$ is compared with a threshold γ. If $\|\boldsymbol{r}\| > \gamma$, presence of bad measurements is assumed.

First introduced in 2009 by Liu et al.

- Show than an adversary can exploit the configuration of the power system to launch such attacks to create malicious errors into certain state variables.
- Demonstrate how attackers can bypass existing techniques for bad measurement detection.

False Data Injection Attacks

Practical Implications



http://www.businessweek.com/articles/2012-03-08/smart-meters-help-brazil-zap-electricity-theft

- Requires the attacker to have knowledge of the power system.
- Adversaries have to manipulate some meters (or meter measurements) before they are used for state estimation.

False Data Injection Attacks

Basic Principle

- z = vector of original measurements.
- z_a = vector of observed measurements that may contain malicious data.
- a = malicious data added to the original measurements ("attack vector").
- x_{bad} = estimate of x using malicious measurements z_a.
- x = estimate of x using original measurements z.
- c = estimation error injected by the attacker.

Thus,

$z_a = z + a$ and $\hat{x}_{bad} = \hat{x} + c$

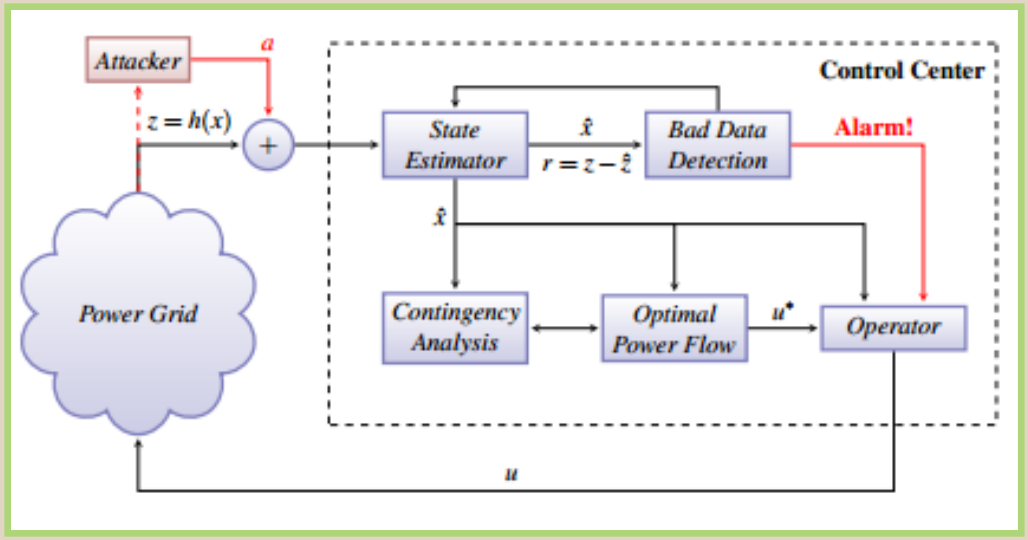


Fig. 1. [10] The state estimator under a cyber attack

False Data Injection Attacks

Theorem 1 (Liu et al., 2009)

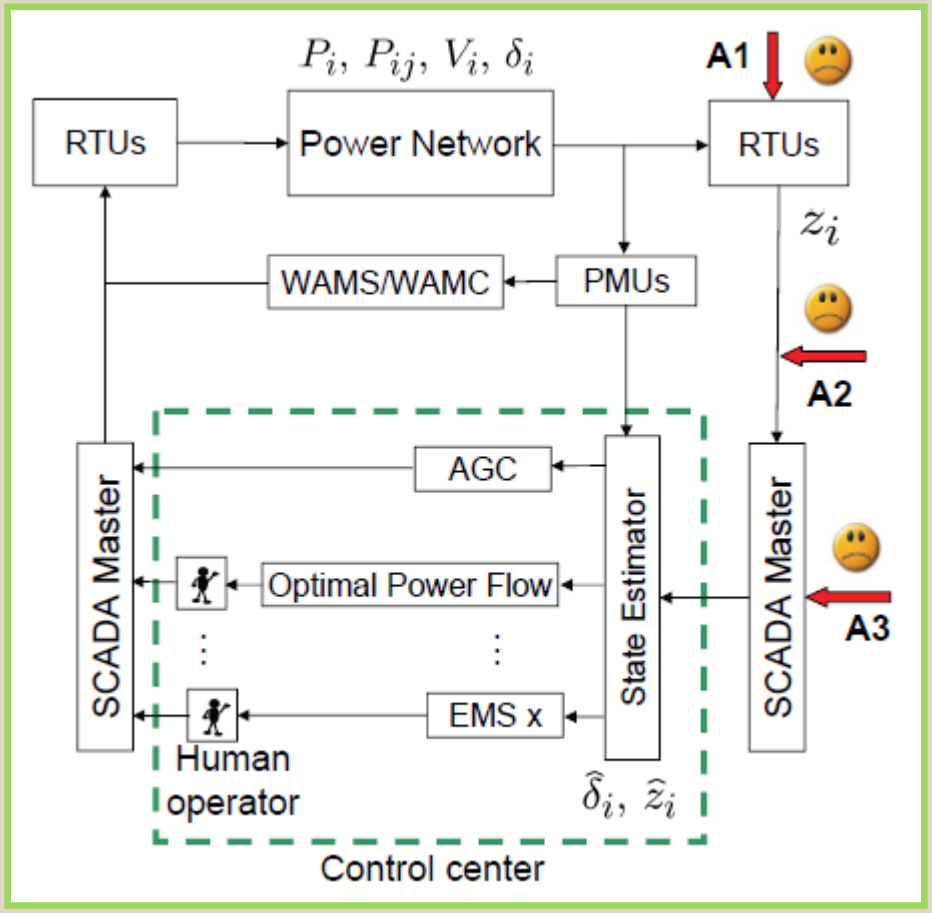
Suppose the original measurements z can pass the bad measurement detection. Then the malicious measurements z_a can pass the bad data measurement detection if a is a linear combination of column vectors of H; i.e., $a = Hc$.

Proof: consider the resulting l_2-norm of the measurement residual vector:

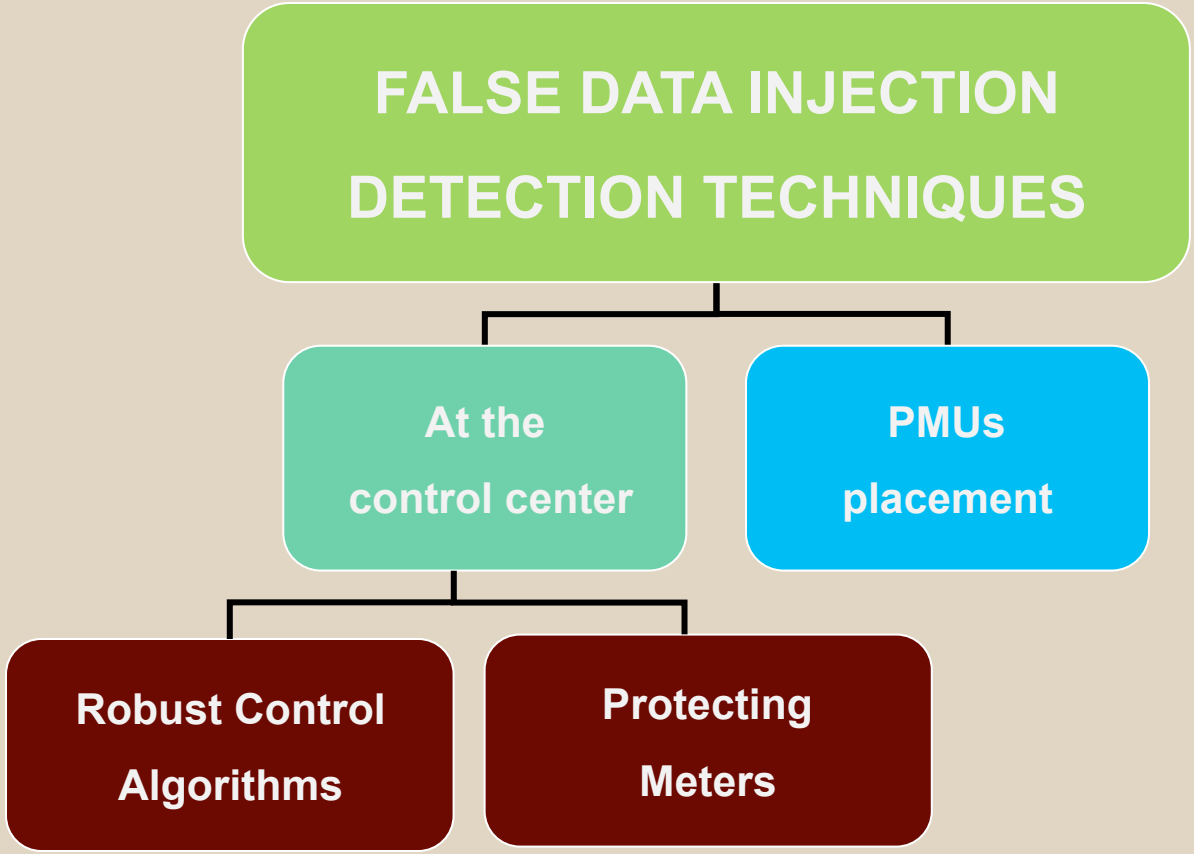
$$\begin{aligned} \|\boldsymbol{r}\| &= \|\mathbf{z_a} - \mathrm{H}\,\hat{\mathbf{x}}_{\mathbf{bad}}\| \\ &= \|\mathbf{z} + \mathbf{a} - \mathrm{H}\,(\hat{\mathbf{x}} + \mathbf{c})\| \\ &= \|\mathbf{z} + \mathbf{a} - \mathrm{H}\hat{\mathbf{x}} - \mathrm{H}\mathbf{c}\| \\ &= \|\mathbf{z} - \mathrm{H}\hat{\mathbf{x}}\| \leq \gamma \qquad ; \text{When } \mathbf{a} = \mathrm{H}\mathbf{c} \end{aligned}$$

Power Network Block Diagram

Fig. 2.[7] A schematic block diagram of a power network, a SCADA system, and a control center. Noisy measurements (z_i) of power flows (Pi, Pij) are sent over the SCADA system to the state estimator where estimates of for example the bus phase angels ($\hat{\delta} i$) are computed.



False Data Injection Detection Techniques



TEXAS A&M UNIVERSITY

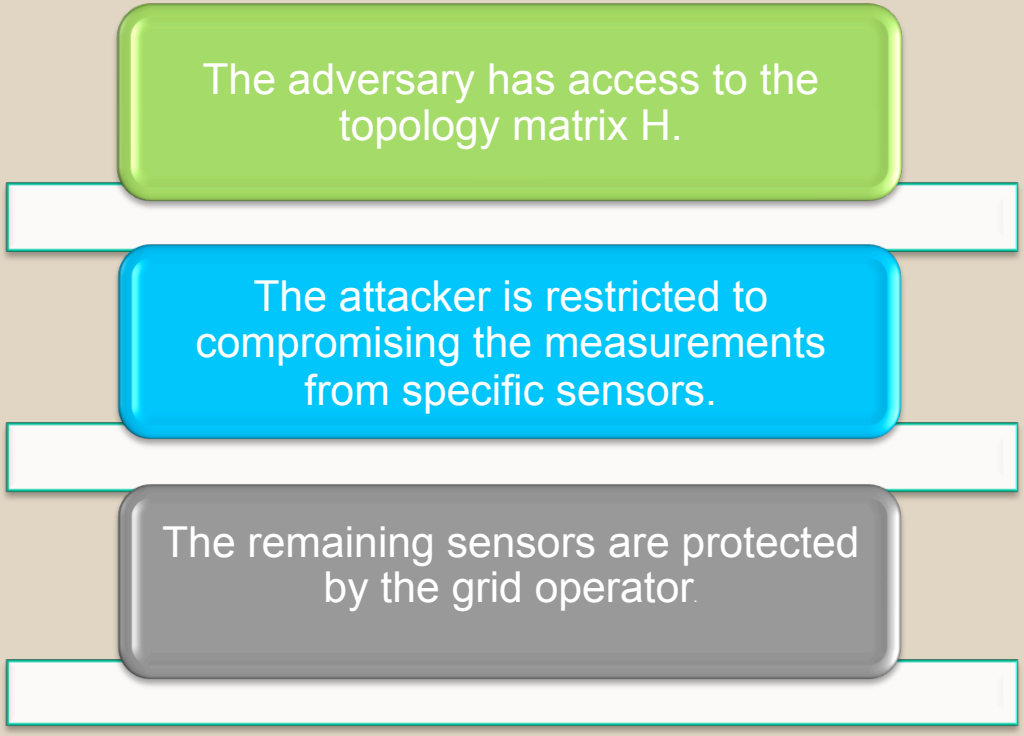
Overview Contributions

- Develop defense strategies for protecting DC state estimation against false data injection attacks proposed by Liu. et.al.
- For a given topology, explore the feasibility of detecting false data injection attacks without having to protect measurements from all sensors while having ways to independently verify or measure the value of a carefully chosen set of state variables.
- Present approaches to identify the set of sensors and states variables to be protected.

Motivation

- It may not be feasible to protect all sensor measurements.
- It is necessary to use application awareness in order to reduce the burden of protecting all sensor measurements.
- For a given topology some sensor measurements influence more states variables than others and hence it might provide better cost to benefit ratio when protected.
- Some state variables are dependent on more sensor measurements than others and hence independently verifying their estimates might limit the attacker's ability in manipulating sensor measurements without being detected.

Adversary Model



What it takes to defend against bad data attacks?

M= set of measurement indices.

V = set of state variable indices.

$I_{\widehat{M}} = M \backslash I_M$= set of indices of measurements protected by the grid operator.

$I_{\widehat{V}} = \mathrm{V} \backslash I_V$ = set of indices state variables that the operator can verify.

Thus,

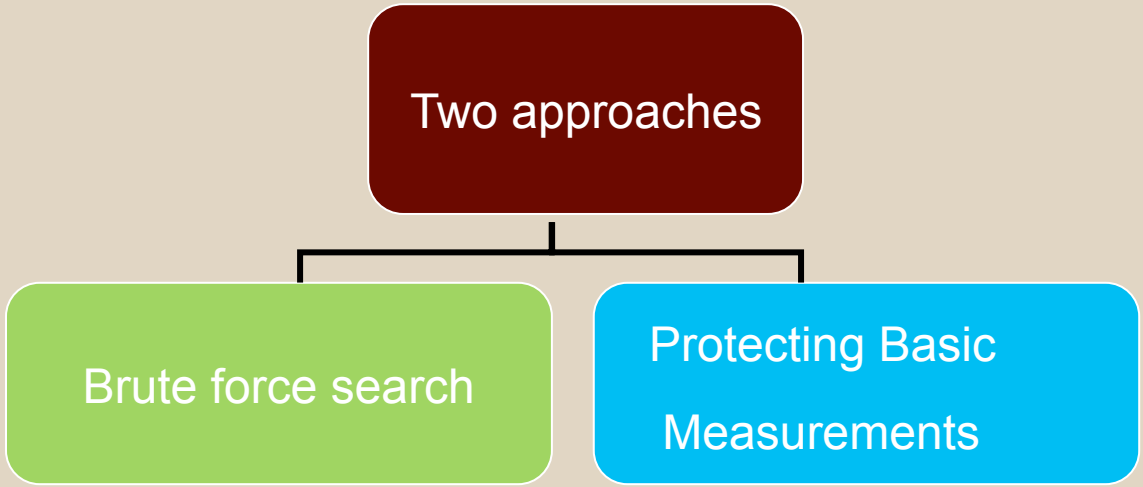
- a_i for $i \in I_{\widehat{M}}$ in $a = (a_1, a_2, \dots, a_m)^T$ are zero
- c_j for $j \in I_{\widehat{V}}$ in $c = (c_1, c_2, \dots, c_n)^T$ are zero

Thus, to launch a false data attack without being detected, the attacker need to find an attack vector a such that:

$$a = Hc$$
$$a_i = 0 \text{ for } i \in I_{\widetilde{M}}$$
$$c_j = 0 \text{ for } j \in I_{\widetilde{V}}$$

Operator's point of view

The operator needs to identify the set of sensors $I_{\hat{V}}$ and $I_{\hat{M}}$ such that the attacker cannot find an attack vector that does satisfy the constraints require to launch such attacks.



TEXAS A&M UNIVERSITY

Overview Contributions

- Introduce a Bayesian problem formulation of the bad data injection attacks.
- Develop countermeasures to malicious data attacks at the control center in the form of attack detector.

TEXAS A&M UNIVERSITY

Motivation

- The control center might take advantage of historical data to preserve and track its believe state of the system.
- The Bayesian formulation can captures the prior information that the control center has about the likely state of the system.

TEXAS A&M UNIVERSITY

Problem Formulation: A Bayesian Framework

Consider a DC power flow model:

$$z = Hx + a + e$$

$z \in \mathbb{R}^m$ = vector of power flow measurements.

$x \in \mathbb{R}^n$ = power system state variable which Gaussian with $N(0, \sum_x)$.

$e \in \mathbb{R}^m$ = random error measurements which is assumed to be jointly Gaussian with covariance matrix $\sum_e$. (i.e. $N(0, \sum_e)$).

a= vector of malicious data injected by an adversary.

It is assumed that the adversary can control at most k meters (i.e. a is a vector with at most k nonzero entries $\|a\|_0 \leq k$). Thus, a is said to have sparsity k if $\|a\|_0 = k$.

- If there is no attack, i.e. $\boldsymbol{a=0}$, $(\boldsymbol{z},\boldsymbol{x})$ are jointly Gaussian. Thus, the minimum mean square error (MMSE) estimator of the vector $\boldsymbol{x}$ is a linear estimator given by:

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

Where $K = \sum_{\mathrm{x}} H^T (H \sum_{\mathrm{x}} H^T + \sum_e)^{-1}$

The MMSE in the absence of attack is given by:

$$\varepsilon_0 = \min_{\hat{x}} E(\|x - \hat{x}(z)\|^2) = \mathrm{Tr}(\sum_{\mathrm{x}} - K_x H \sum_{\mathrm{x}})$$

- In the presence of an attack, the mean square error can be written as:

$$E(\|x - K(Hx + e + a)\|^2) = Tr[(I - KH)\sum_x (I - KH)^T + K\sum_e K^T + Kaa^T K^T]$$

- Note that the only term dependent on a is the last one, which can be written as $\|Ka\|_2^2$.

- Thus , to increase the MSE at the state estimator, the adversary needs to increase the "energy" of the attack.

TEXAS A&M UNIVERSITY

Attack Scenario: Minimum residue energy attack

- This attack explores the problem of finding the worst attack in the regime that the adversary can not perform an unobservable attack.

What choice does the adversary have?

- Select an attack vector that is particularly damaging to the control center's state estimation without being easily detectable.

TEXAS A&M UNIVERSITY

- Given the naive MMSE state estimator $\hat{x} = Kz$, the estimation residue error is given by:

$$r = z - H\hat{x}$$
$$r = (I - HK)z$$

Substituting the measurement model $z = Hx + a + e$ and we define $G = (I - HK)$ we have:

$$r = GHx + Ga + Ge$$

Where Ga is the only term from the attack.

TEXAS A&M UNIVERSITY

- Therefore the attacker can consider the following equivalent problems:

$$\max_{a \in A_k} \|Ka\|_2^2 \text{ subject to } \|Ga\|_2^2 \leq n$$

or equivalently,

$$\min_{a \in A_k} \|Ga\|_2^2 \text{ subject to } \|Ka\|_2^2 \geq C$$

TEXAS A&M UNIVERSITY

Operator's Point of View: Statistical Model and Attack Hypothesis

Assuming a Bayesian model where the control center state variables are random with multivariate Gaussian distribution $\mathrm{x} \sim \mathrm{N}(0, \sum_{\mathrm{x}})$.

Consider the following composite binary hypothesis:

$$H_0: a = 0 \text{ versus } H_1: a \in A_k \backslash \{0\}$$

Given observation $\boldsymbol{z} \in \mathbb{R}^{\boldsymbol{m}}$, we wish to design a detector $\delta: \mathbb{R}^{\boldsymbol{m}} \rightarrow \{\mathbf{0}, \mathbf{1}\}$ with $\delta(\mathrm{z})=1$ indicating a detection of attack (H_1) and $\delta(\mathrm{z})=0$ the null hypothesis.

TEXAS A&M UNIVERSITY

- **Generalized Likelihood Ratio Detector with l_1 Norm Regularization**

  - Propose a detector based on the generalized likelihood ratio test **(GLRT)**.

  - The distribution of the measurement z under the two hypothesis differ only by their means:

$$H_0: z \sim \mathrm{N}(0, \sum_{\mathrm{z}}) \text{ and } H_1: z \sim \mathrm{N}(\mathrm{a}, \sum_{\mathrm{z}}), \mathrm{a} \in A_k \backslash \{0\}$$

Where $\sum_{\mathrm{z}} \triangleq H \sum_x H^T + \sum_e$.

TEXAS A&M UNIVERSITY

- The **GLRT** is given by:

$$L(z) \triangleq \frac{\max_{a \in A_k} f(z/a)}{f(z/a=0)} \underset{H_0}{\overset{H_1}{\gtrless}} \tau,$$

Where $f(z/a)$ is the Gaussian density function with mean a and covariance $\sum_z$. and the threshold τ is chosen from under null hypothesis for a certain false alarm rate. This is equivalent to:

$$\min_{a \in A_k} a^T \sum_z^{-1} a - 2z^T \sum_z^{-1} a \underset{H_0}{\overset{H_1}{\gtrless}} \tau$$

- Thus the **GLRT** reduces to solving:

$$min\ a^T \sum_z^{-1} a - 2z^T \sum_z^{-1} a$$

Subject to $\|a\|_0 \leq k$

For larger k the above problem can be approximated by the following convex optimization:

$$min\ a^T \sum_z^{-1} a - 2z^T \sum_z^{-1} a$$

Subject to $\|a\|_1 \leq \gamma$

Critical Assessments

- ✓ The scenario when measurements from the protected sensors are not available so that the operator might not be able to detect malicious attacks could be explored. What may this scenario involve?

- ✓ It might be useful to develop attack scenarios without imposing higher burden on the adversary (i.e. the attacker does know z).

- ✓ It still remains unclear how much damage can be done to the power network when malicious attacks are launched. It would be worthwhile to analyze the next step to be taken whenever the control center takes erroneous decisions as a consequence of false data injection attacks on state estimation.

Future Work

- Extend results to state estimation using **AC** power flow models.
- Explore the possibility of adapting network anomaly detection techniques to identify false data injection attacks.
- Develop computationally efficient algorithms for the **GLRT** detector and the design of adaptive optimal minimum residue energy attack.
- Finding explicit mechanisms to secure some of the measurements, developing possible more efficient attacking strategies, and seeking computable upper bounds to better evaluate the performance of the design algorithms.

References

[1] A. Abur and A. G. Exposito, "Power System State Estimation: Theory and Implementation", Marcel Dekker, Inc., 2004.

[2] Y. Liu, P. Ning, and M. Reiter, "False Data Injection Attacks Against State Estimation in Electric Power Grids", in Proceedings of the 16th ACM Conference on Computer and Communications Security, Chicago, Illinois, 2009, pp. 21-32.

[3] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, “Limiting false data attacks on power system state estimation,” in Proc. Conf. Information Sciences and Systems, Princeton, NJ, Mar. 2010, pp. 1–7.

[4] R. B. Bobba, K. M. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. J. Overbye, "Detecting False Data Injection Attacks on DC State Estimation", in Preprints of the First Workshop on Secure Control Systems, CPSWEEK 2010, 2010.

[5] 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.

[6] 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

[7] H. Sandberg, A. Teixeira, and K. H. Johansson, “On security indices for state estimators in power networks,” in Preprints of the First Workshop on Secure Control Systems, CPSWEEK 2010, 2010.

[8] G. Dan and H. Sandberg, "Stealth Attacks and Protection Schemes for State Estimators in Power Systems", in 2010 First IEEE International Conference on Smart Grid Communications, 2010, PP.214-219.

References

[9] T. T. Kim and H. V. Poor, "Strategic Protection Against Data Injection Attacks on Power Grids," *IEEE Transactions on Smart Grid*, 2011, PP. 326 – 333.

[10]Teixeira. A, Dán G., Sandberg H., Johansson K. H. , "A Cyber Security Study of a SCADA Energy Management System: Stealthy Deception Attacks on the State Estimator," World Congress, Volume # 18 | Part# 1, 2011, pp. 11271-11277.

Questions?