Cyber-Security Analysis of State Estimators in Electric Power Systems

André Teixeira[1], Saurabh Amin[2], Henrik Sandberg[1], Karl H. Johansson[1], and Shankar Sastry[2]

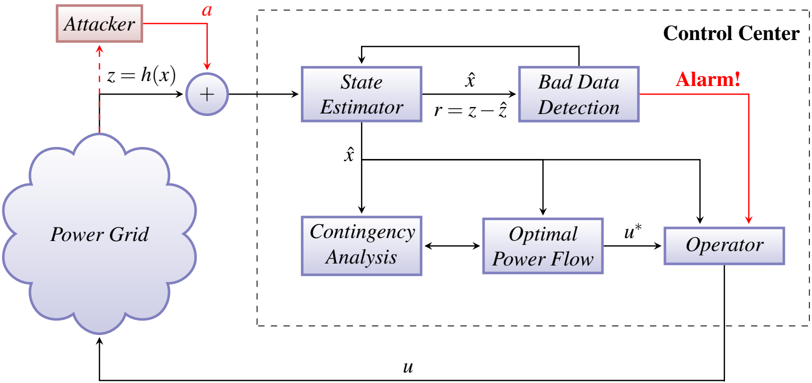
ACCESS Linnaeus Centre, KTH-Royal Institute of Technology[1]
TRUST Center, UC Berkeley[2]

Conference on Decision and Control
December 17th, 2010

Outline

Outline

- Normal failures have huge impact - US-Canada 2003 Blackout
- What about intentional failures?

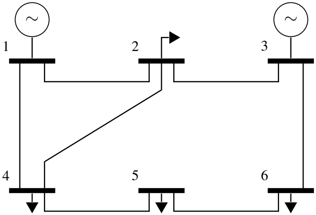


- Most of the theory developed from the 70's to the 90's assumes the data corruption comes from "nature" $\Rightarrow$ noise
  - A framework to analyze this system under malicious data corruption is lacking!

- Questions
  - Can malicious attackers generate stealthy deception attacks, with perfect model knowledge? [Liu et al. 2009]
  - Can malicious attackers generate stealthy deception attacks, without perfect model knowledge? [This paper]
  - How to reasonably model the attacker? [This paper]
  - How "hard" is it to perform stealthy deception attacks? [Sandberg et al. 2010, Dán et al. 2010]
  - How to deploy protective resources? [Bobba et al. 2010, Dán et al. 2010]
- Objectives
  - Provide a (comprehensive) framework to analyze control systems under malicious data corruption.

Outline

- **Steady-State Model:**
  $z = h(x) + \epsilon$
  Ex.: $P_{14} = V_1 V_4 b_{14} \sin(\theta_1 - \theta_4)$
  measurements: $z \in \mathbb{R}^m$
  state: $x \in \mathbb{R}^n$
  nonlinear model: $h(x)$
  Gaussian noise: $\epsilon \sim \mathcal{N}(0, R)$



- **Nonlinear Weighted Least-Squares:**

$$\hat{x} = \arg\min_{x \in \mathbb{R}^n} \frac{1}{2} r(x)^\top R^{-1} r(x),$$

where $r(x) = z - h(x)$ is the measurement residual

  - *Local Linear Approximation around origin ($z = Hx + \epsilon$):*

$$\hat{x} = \left[H^\top R^{-1} H\right]^{-1} H^\top R^{-1} z$$

$H = \frac{\partial h}{\partial x}(\hat{x}^0)$ - the Jacobian matrix (tall and sparse)

- **Normalization:**

$$\begin{aligned} \bar{z} &= R^{-1/2} z \\ \bar{\epsilon} &= R^{-1/2}\epsilon \\ \bar{H} &= R^{-1/2} H \end{aligned} \quad \Rightarrow \quad \begin{aligned} \hat{x} &= \bar{H}^{\dagger}\bar{z} \\ \hat{z} &= \bar{H}\bar{H}^{\dagger}\bar{z} = \bar{K}\bar{z} \\ \bar{r} &= (I-\bar{K})\bar{z} = \bar{S}(\bar{H}x+\bar{\epsilon}) = \bar{S}\bar{\epsilon} \\ \bar{\epsilon} &\sim \mathcal{N}(0, I) \end{aligned}$$

- **Main useful concepts:**
  - $\bar{K}$ is the orthogonal projector onto $\mathrm{Im}(\bar{H})$, since $\bar{K}\bar{K} = \bar{K} = \bar{K}^{\top}$
  - $\bar{S} = (I - \bar{K})$ is the orthogonal projector onto $\mathrm{Ker}(\bar{H}^{\top})$
  - $\mathrm{Im}(\bar{H}) \perp \mathrm{Ker}(\bar{H}^{\top}) \Rightarrow \bar{S}a = 0\, \forall a \in \mathrm{Im}(\bar{H})$ [Clements et al. 81, Liu et al. 09]

- Hypothesis test:
  - H_0: No bad data is present (null hypothesis)
  - H_1: Bad data is present (alternative hypothesis)

- **Performance index test:**
  $J(\hat{x}) = \bar{\epsilon}^\top \bar{S} \bar{\epsilon} \sim \chi^2_{m-n}$:
  accept H_0 if $\|\bar{r}\|_2 \leq \sqrt{\tau_\chi(\alpha)}$

- **Largest normalized residual test:**
  $\bar{r}(\hat{x}) \sim \mathcal{N}(0, \bar{S})$, $D = \mathrm{diag}(\bar{S})$:
  accept H_0 if $\|D^{-1/2}\bar{r}\|_\infty \leq \tau_\mathcal{N}(\alpha)$

- $\alpha \in [0,\ 1]$ is the false alarm rate, *i.e.* $P(H_1|H_0)$.
- **General expression:** $\|Wr(\hat{x})\|_p < \tau$, for suitable W, p and τ.

Outline

- **Corrupted measurements:** $\bar{z}^a = \bar{z} + a$
- **Attacker Goals**
  - Convergence of the estimator (trivial for the linear case);
  - Stealthiness: $\|Wr(\hat{x}^a)\|_p < \tau$;
  - Induce a desired bias on a subset of measurements

- **Minimum "Effort" Attack Synthesis**

$$\min_a \|a\|_p$$
$$\text{s.t. } a \in \mathcal{G}, \ a \in \mathcal{U}$$

  - $\mathcal{G}$ - set of goals
  - $\mathcal{U}$ - class of stealthy attacks

- **Different metrics for "effort"**
  - $p = 0$: cardinality of a (# of measurements to be corrupted) - not convex
  - $p = 1$: may be used as a convex approximation of $p = 0$
  - $p = 2$: is related to measurement redundancy in the system
  - All quantify "how hard" it is to attack the estimator, for a given set of goals [Sandberg et al. 10]

- **Stealthy attacks with Perfect Model Knowledge**

$$a \in \mathrm{Im}(\bar{H}) \Rightarrow a \in \mathcal{U} \text{ [Clements et al. 81, Liu et al. 09]}$$

  - $a \in \mathrm{Im}(\bar{H}) \Leftrightarrow$
    $\exists c : a = \bar{H}c$
  - Guaranteed that $r(\bar{z}^a) = \bar{S}(\bar{z} + a) = \bar{S}\bar{z} = r(\bar{z})$
  - $P(H_1|H_1) = P(H_1|H_0)$

- **Stealthy attacks with Perturbed Model Knowledge**
  - Known model is $\tilde{H} = \bar{H} + \Delta\bar{H}$
  - Let the same policy be used: $a = \tilde{H}c$, for some c.
  - $\bar{r}(\bar{z}^a) = \bar{S}\bar{\epsilon} + \bar{S}a$
  - $\bar{S}a \neq 0 \Rightarrow P(H_1|H_1) \neq P(H_1|H_0)$: No perfect stealthiness
  - Relaxation - Allow for a maximum detection risk tolerated by the attacker, $\bar{\delta} : P(H_1|H_1) \leq P(H_1|H_0) + \bar{\delta}$. Depends on the detection scheme!
    - What is the class of attacks satisfying such condition?
- Solution steps:
  - Given a detection scheme, α, and $\bar{\delta}$, obtain $\lambda : \|\bar{S}a\|_p \leq \lambda \Rightarrow P(H_1|H_1) \leq P(H_1|H_0) + \bar{\delta}$
  - Given λ, obtain $\beta : \|a\|_p \leq \beta \Rightarrow \|\bar{S}a\|_p \leq \lambda$
  - Then $\|a\|_p \leq \beta \Rightarrow a \in \mathcal{U}$

Performance index test

- Under attack, $J_a(\hat{x}) \sim \chi^2_{m-n}(\lambda)$ where $\lambda = \|\bar{S}a\|_2^2$ (noncentrality parameter).
- $\bar{r}_a = \bar{S}a$ corresponds to the residual bias due to the attack (recall $\bar{r}(\bar{z}^a) = \bar{S}\bar{\epsilon} + \bar{S}a$)
- An attack is $\bar{\delta}$-stealthy if $P(H_1|H_1) = P(J_a > \tau_\chi(\alpha)) \leq P(H_1|H_0) + \bar{\delta}$:

$$\int_{\tau_\chi(\alpha)}^{\infty} g_\lambda(u)du \leq \alpha + \bar{\delta}. \tag{1}$$

Assumption

$P(H_1|H_1)$ increases monotonically with λ.

Proposition

Given α and $\bar{\delta}$, an attack is $\bar{\delta}$-stealthy regarding the performance index test if the following holds

$$\|\bar{r}_a\|_2^2 = \|\bar{S}a\|_2^2 \leq \bar{\lambda}(\alpha, \bar{\delta})$$

where $\bar{\lambda}(\alpha, \bar{\delta})$ is the maximum value of λ for which (1) is satisfied.

- Known results [Galántai 06]:

Definition

Let M_1 and M_2 be subspaces of $\mathcal{C}^m$. The smallest principal angle $\gamma_1 \in [0, \pi/2]$ between M_1 and M_2 is defined by

$$\cos(\gamma_1) = \max_{u \in M_1} \max_{v \in M_2} |u^H v|$$
$$\text{subject to } \|u\| = \|v\| = 1$$

Lemma

Let $\mathcal{P}_1, \mathcal{P}_2 \in \mathbb{R}^{m \times m}$ be orthogonal projectors of M_1 and M_2, respectively. Then the following holds

$$\|\mathcal{P}_1 \mathcal{P}_2\|_2 = \cos(\gamma_1)$$

- Applying the previous results we have:

Proposition

Let γ_1 be the smallest principal angle between Ker$(\bar{H}^\top)$ and Im$(\tilde{H})$. The residual increment due to a deception attack, $\bar{r}_a$, following the policy $a = \tilde{H}c$ satisfies

$$\|\bar{r}_a\|_2 \leq \cos\gamma_1 \|a\|_2.$$

Proof.

Recall $\bar{r}(\bar{z}^a) = \bar{S}\bar{z}^a = \bar{S}\bar{z} + \bar{S}a = \bar{r} + \bar{r}_a$.
$a = \tilde{H}c \Rightarrow a \in \mathrm{Im}(\tilde{H}) \Rightarrow a = \tilde{K}a$.
$\bar{r}_a = \bar{S}\tilde{K}a \Rightarrow \|\bar{r}_a\|_2 \leq \|\bar{S}\tilde{K}\|_2\|a\|_2$. □

Theorem

Given the perturbed model $\tilde{H}$, the false-alarm probability α and the maximum admissible risk $\bar{\delta}$, an attack following the policy $a = \tilde{H}c$ is stealthy regarding the performance index test if

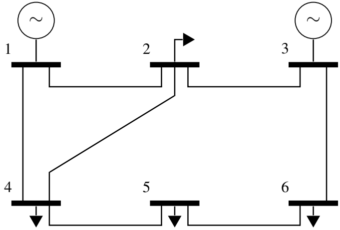
$$\|a\|_2 \leq \beta(\alpha, \bar{\delta}) \ ,$$

where $\beta(\alpha, \bar{\delta}) = \frac{\sqrt{\bar{\lambda}(\alpha, \bar{\delta})}}{\cos \gamma_1}$.

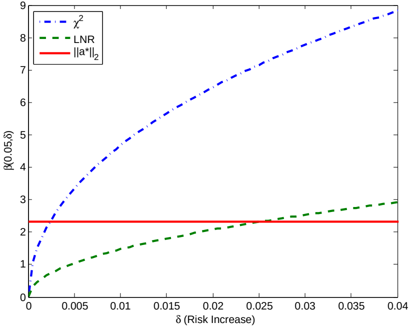
Outline

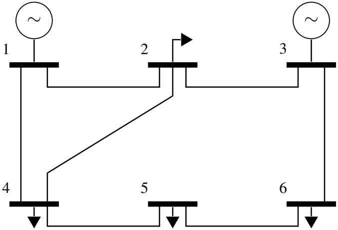
- Consider the 6 bus system with the following branch parameters:

Branch	From bus	To bus	Reactance (pu)	Parameter Error
$b1$	1	4	0.370	-20%
$b2$	1	2	0.518	+20%
$b3$	6	5	1.05	-20%
$b4$	6	3	0.640	-20%
$b5$	5	4	0.133	-20%
$b6$	4	2	0.407	-20%
$b7$	3	2	0.300	+20%



- The attacker's model $\tilde{H}$ has the correct topology and a $\pm 20\%$ error in the parameters.
- The parameter errors were numerically computed so that $\|\bar{S}\tilde{K}\|_2 = \cos\gamma_1$ is maximized.
- Objective: induce a unit bias in z_{b_1}, *i.e.* have $a_{b_1} = 1$, without being detected.





- Upper bound on the attack vector as a function of the detection risk.
- The solid line represents the 2-norm of the optimal attack vector a^* constrained by $a_{b_1} = 1$
- The curves denoted as χ^2 and LNR represent the value of $\beta(0.05, \delta)$ for the performance index test and largest normalized residual test.

Outline

Final Remarks

- The proposed framework can also be applied to other structured uncertain models such as models
  - with missing rows/measurements;
  - with missing columns;
  - obtained from data analysis.
- The optimization framework for attack synthesis enables the embedding of constraints such as
  - encrypted measurements;
  - pseudo-measurements;
  - finite resources;
- The proposed framework has been applied to a real SCADA/EMS software - submitted to the IFAC World Congress 2011