Vulnerability analysis and robustification of power grid state estimation

Ming Jin

with Prof. Karl Henrik Johansson and Prof. Javad Lavaei

ISMP2018 BORDEAUX

Ukraine power grid cyberattack (2015)

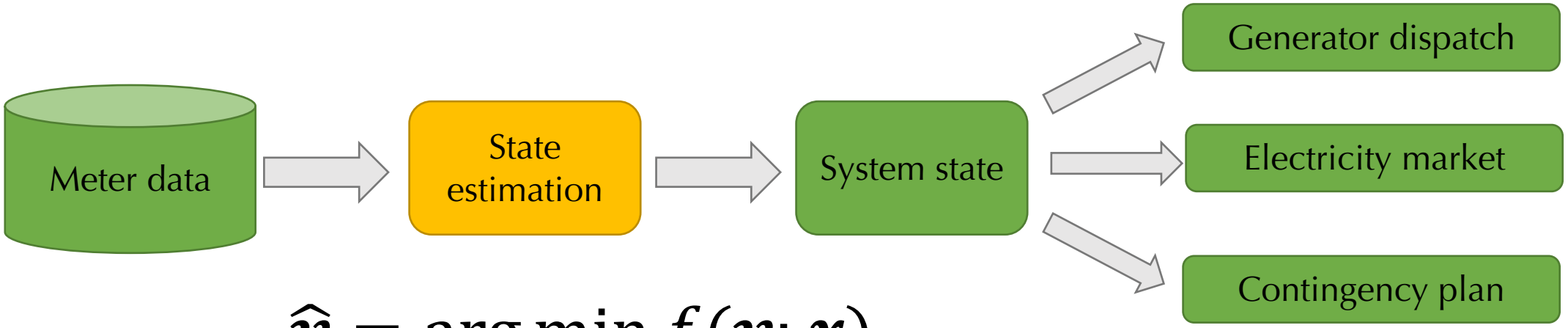
The Ukrainian Power Grid Was Hacked Again

Experts say the country appears to be a "testbed" for cyber attacks that could be used around the world.

Vulnerability analysis

- Better understand the behavior of complex control systems
- Better understand the limitations of state estimation
- Motivate/inspire defense mechanism

Vulnerability of power grid state estimation



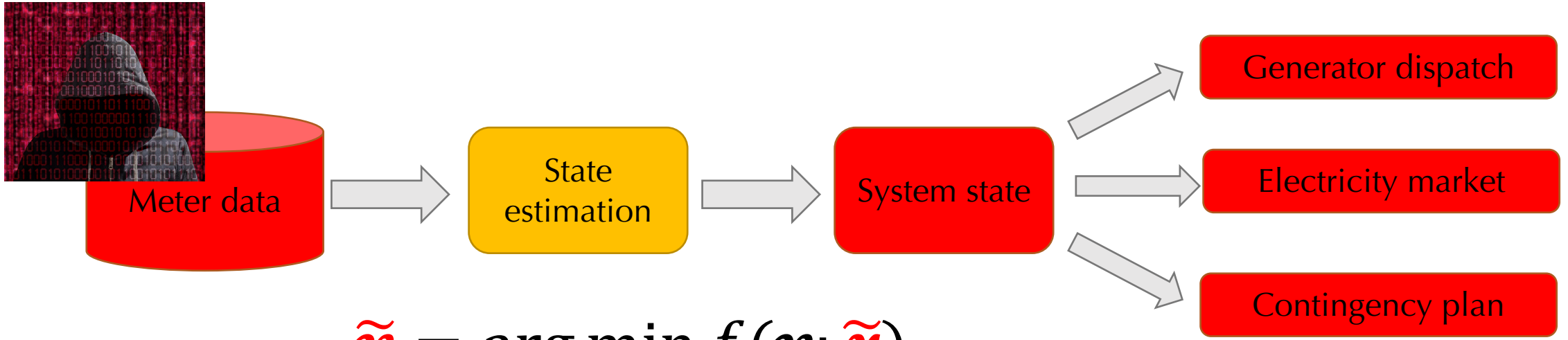
$$\hat{\boldsymbol{v}} = \arg\min f(\boldsymbol{v}; \boldsymbol{x})$$

$\hat{\boldsymbol{v}}$: estimated system state

$\boldsymbol{x}$: meter data

f: regression function

Vulnerability of power grid state estimation



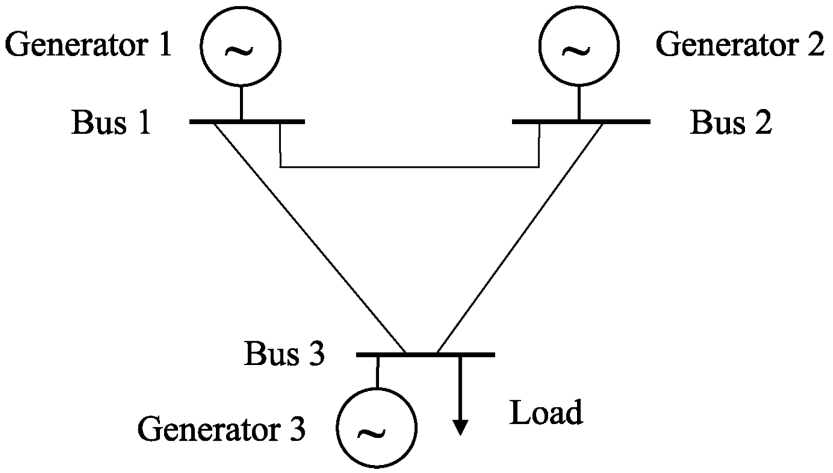
$$\tilde{\boldsymbol{v}} = \arg\min f(\boldsymbol{v}; \tilde{\boldsymbol{x}})$$

$\tilde{\boldsymbol{v}}$: false system state

$\tilde{\boldsymbol{x}}$: false data

f: regression function

Power system modeled as a graph

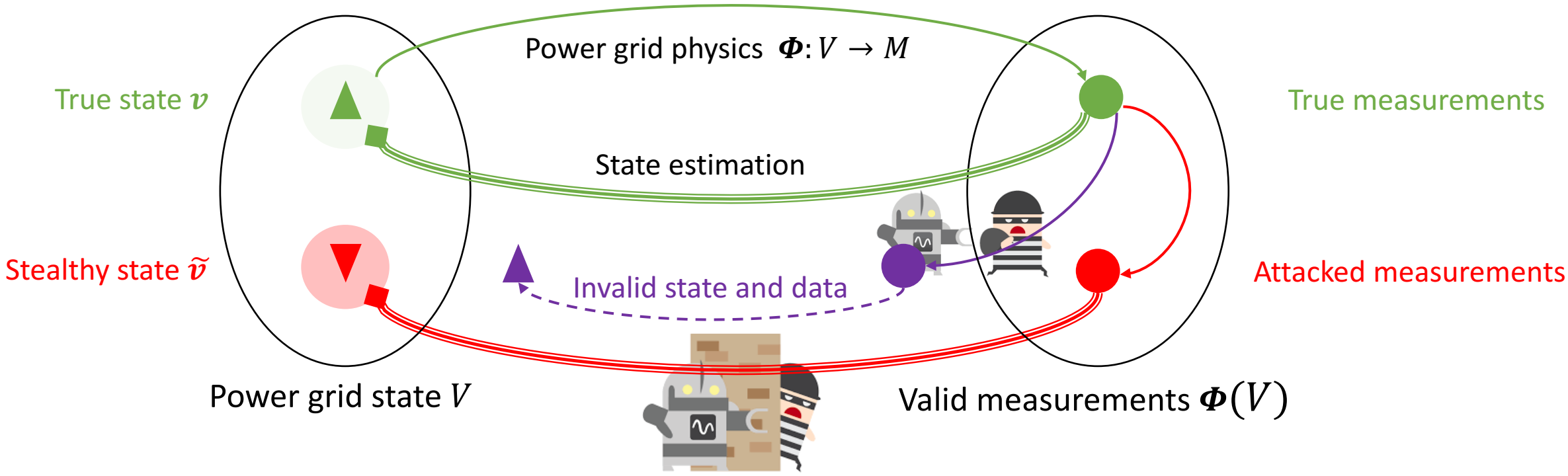


- A power system $\mathcal{G} = (\mathcal{N}, \mathcal{L})$
  - Transmission lines, buses, and transformers
- Complex voltage: $\boldsymbol{v} = [v_1, v_2, \ldots, v_n]^T \in \mathbb{C}^n$
- Sensor measurement is **quadratic function** of the state:

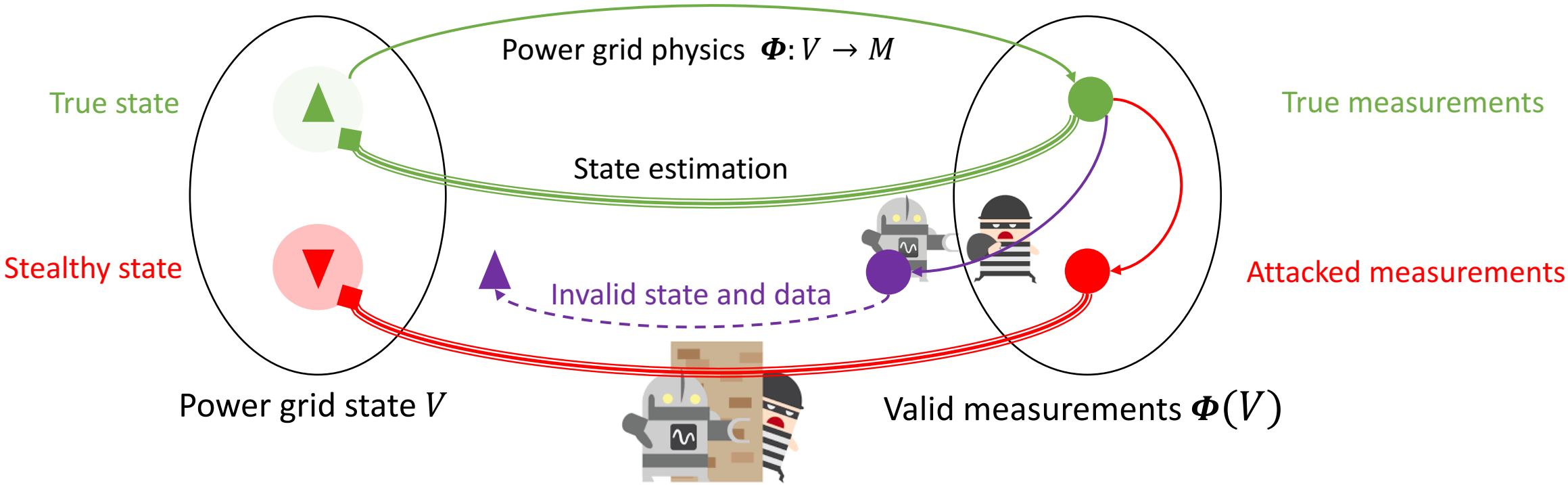
True data $m_i = \langle \boldsymbol{v}\boldsymbol{v}^*, \boldsymbol{M}_i \rangle + w_i$ noise

true state

Stealth attack



Stealth attack



Definition: an attack is **stealthy** if the attacked measurement lies in $\boldsymbol{\Phi}(V)$, the image of V under Φ.

- For $\boldsymbol{m} = \boldsymbol{\Phi}(\boldsymbol{v})$, An attack $\boldsymbol{b}$ is stealthy *if and only if* $\exists\ \tilde{\boldsymbol{v}} \in V$ such that $\boldsymbol{\Phi}(\tilde{\boldsymbol{v}}) = \boldsymbol{b} + \boldsymbol{m}$

Stealth attack quadratic measurement model

AC model

$$m_i = \langle \boldsymbol{v}\boldsymbol{v}^*, \boldsymbol{M}_i \rangle + w_i$$

Stealth attack

$$\tilde{m}_i = m_i + b_i = \langle \tilde{\boldsymbol{v}}\tilde{\boldsymbol{v}}^*, \boldsymbol{M}_i \rangle$$

adversarial data

False state

DC model

$$\boldsymbol{m} = \boldsymbol{H}\boldsymbol{v} + \boldsymbol{w}$$

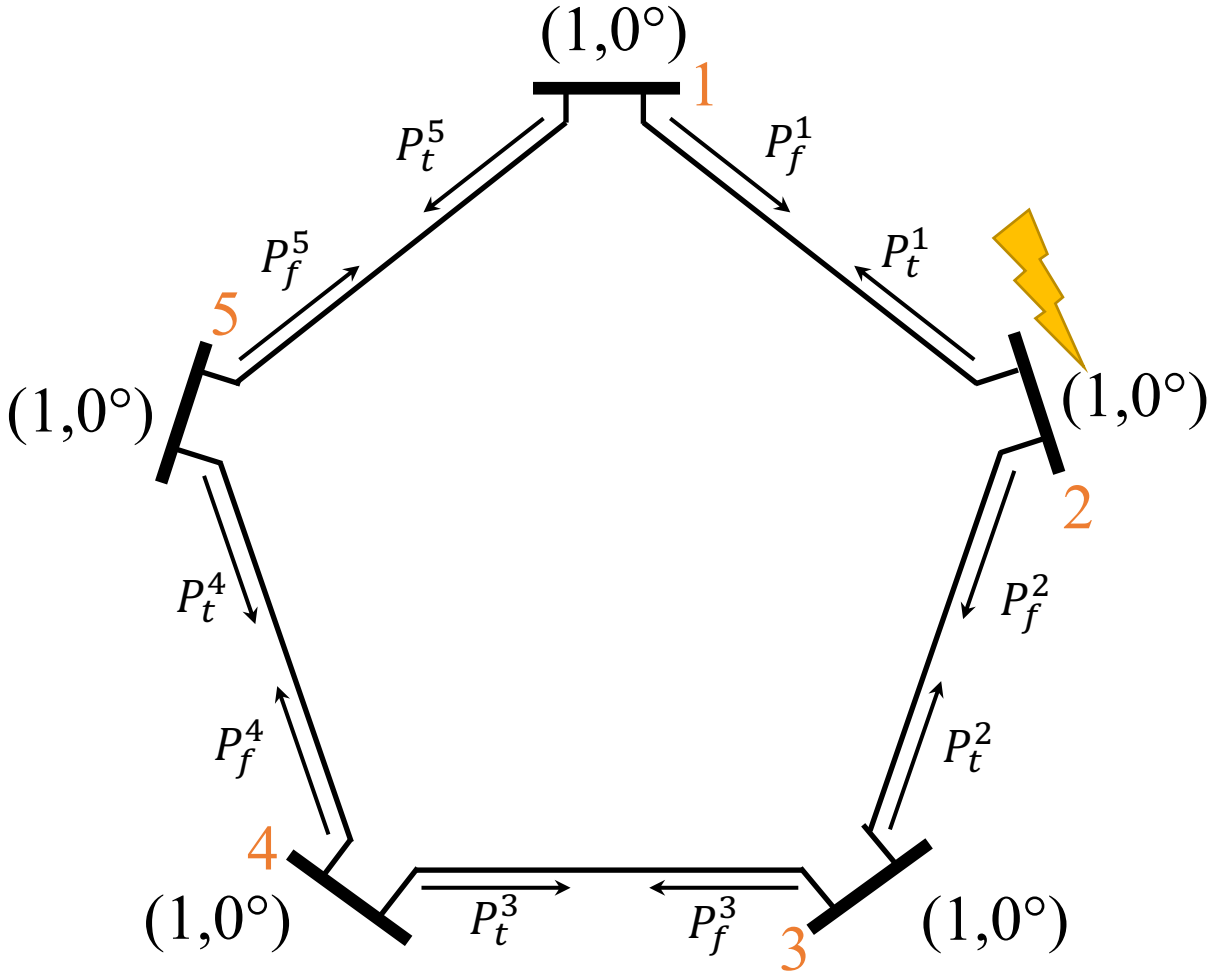
$$\tilde{\boldsymbol{m}} = \boldsymbol{m} + \boldsymbol{b} = \boldsymbol{H}\tilde{\boldsymbol{v}}$$

- Different from the **DC attack model (linear measurements)**: (Liu et al., 2010) (Kosut et al., 2010) (Sandberg et al., 2010) (Dan and Sandberg, 2010) (Yuan et al., 2011) (Sou et al., 2013) (Hendrickx et al., 2014) (Liu et al., 2010) (Kosut et al., 2010) (Sandberg et al., 2010) (Dan and Sandberg, 2010) (Yuan et al., 2011) (Sou et al., 2013) (Hendrickx et al., 2014) (Soltan et al., 2016, 2017)
- AC attack model: (Rahman and Mohsenian-Rad, 2013) (Hug and Giampapa, 2012)

Is it possible to find a *sparse* & *stealthy* attack?

Motivation case study

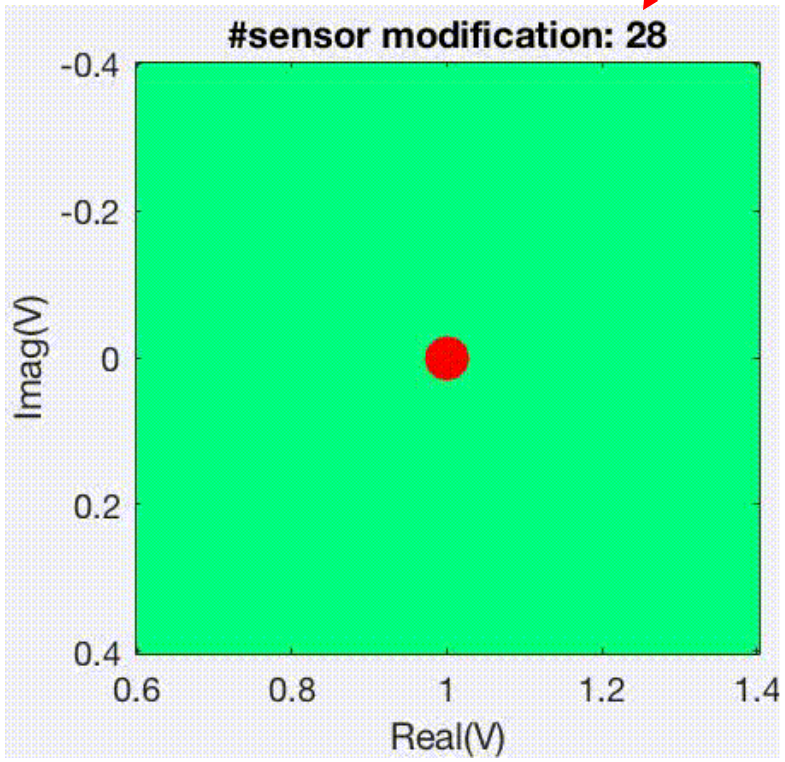
- 5 buses
- Attack the state at bus 2
- Total 28 measurements
  - 20 branch real and reactive power flows
  - 8 (nodal real and reactive power injections) or (nodal real and voltage magnitudes)



Feasibility region

- Grid search over the real/imaginary values of spurious state $\tilde{v}_2$

Attack budget: number of nonzero sensor modifications allowed



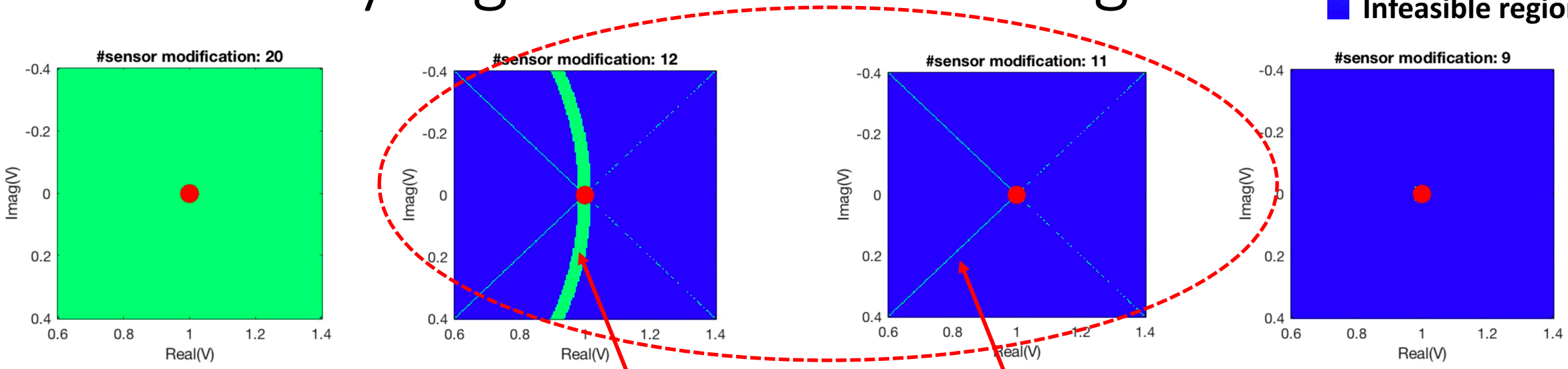
- True state
- **Feasible region:** a spurious state that corresponds to an attack "within the budget"
- **Infeasible region:** a spurious state that corresponds to an attack "out of the budget"

Feasibility region and attack budget

True state

Feasible region

Infeasible region



Entire space

Magnitude measurement

Taylor expansion of the quadratic measurement

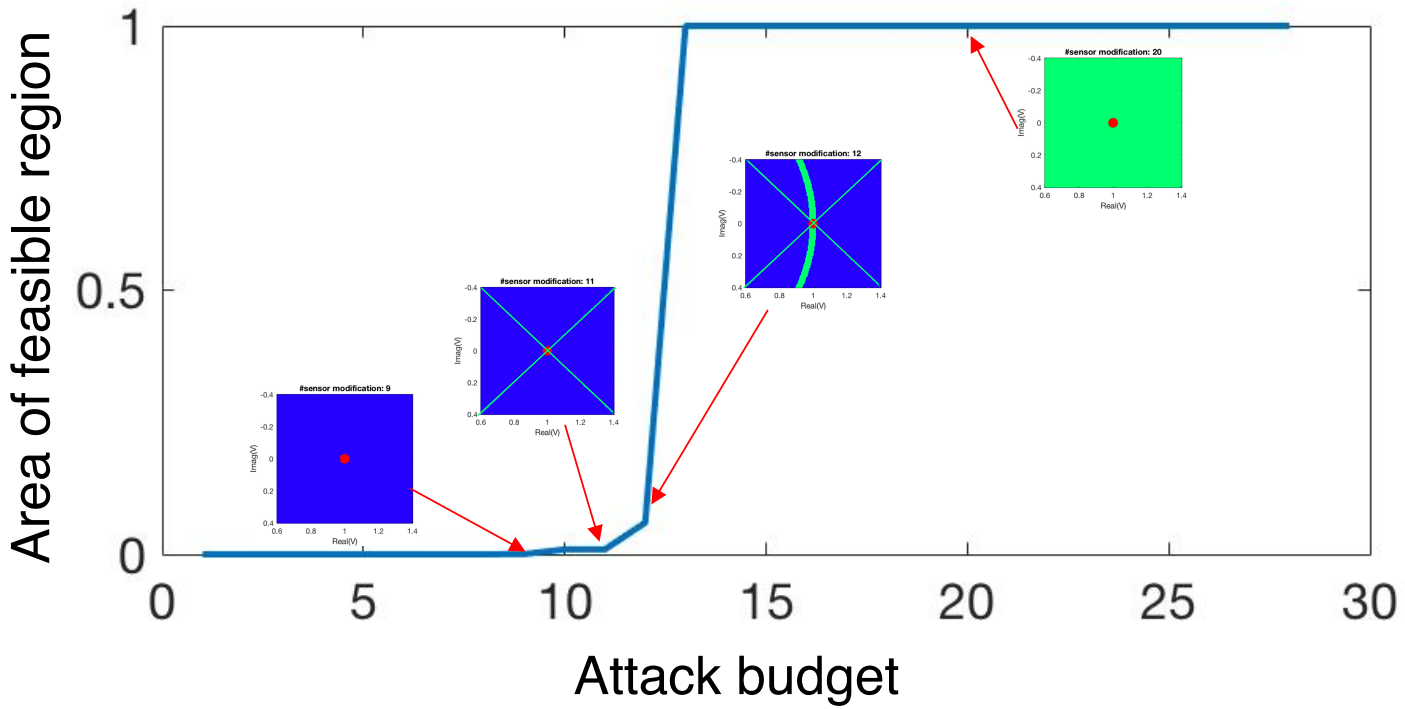
Infeasible problem

AC model $m_i = \langle \boldsymbol{v}\boldsymbol{v}^*, \boldsymbol{M}_i \rangle + w_i$

DC model $\boldsymbol{m} = \boldsymbol{H}\boldsymbol{v} + \boldsymbol{w}$

- AC-DC **compare:** the feasibility region for DC would be either the entire subspace or empty

Feasibility region and attack budget

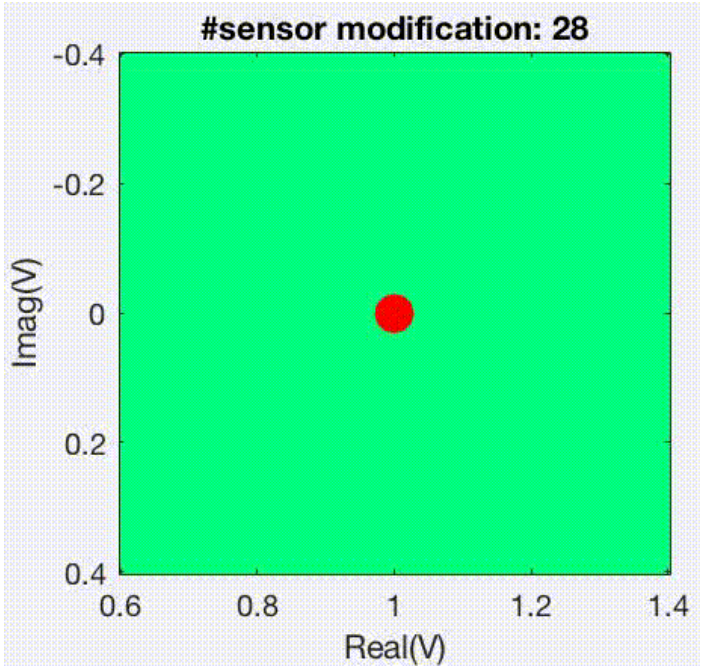


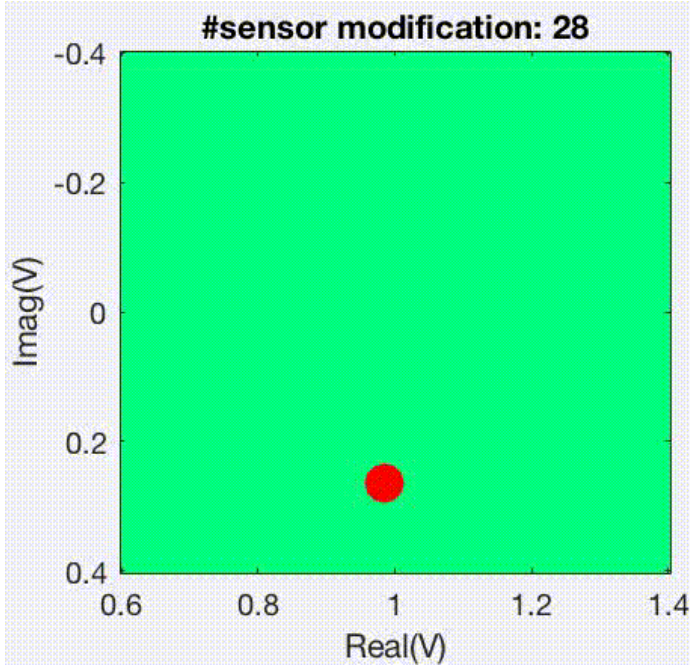
AC model $m_i = \langle \boldsymbol{v}\boldsymbol{v}^*, \boldsymbol{M}_i \rangle + w_i$

DC model $\boldsymbol{m} = \boldsymbol{H}\boldsymbol{v} + \boldsymbol{w}$

- AC-DC **compare:** the "phase transition" happens around 2*number of buses

Feasibility region for a more complex case

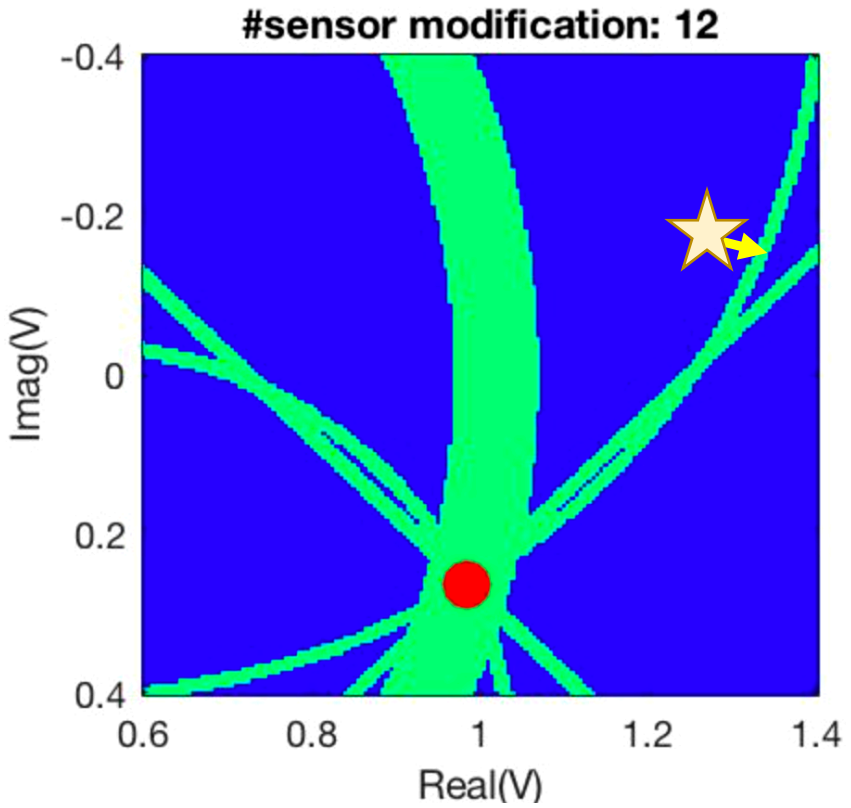




- True state
- **Feasible region:** a spurious state that corresponds to an attack "within the budget"
- **Infeasible region:** a spurious state that corresponds to an attack "out of the budget"

True state on the **left**: $v_2 = 1 + 0i$, **right**: $v_2 = 0.9852 + 0.2640i$

Simple(?) grid search



True state

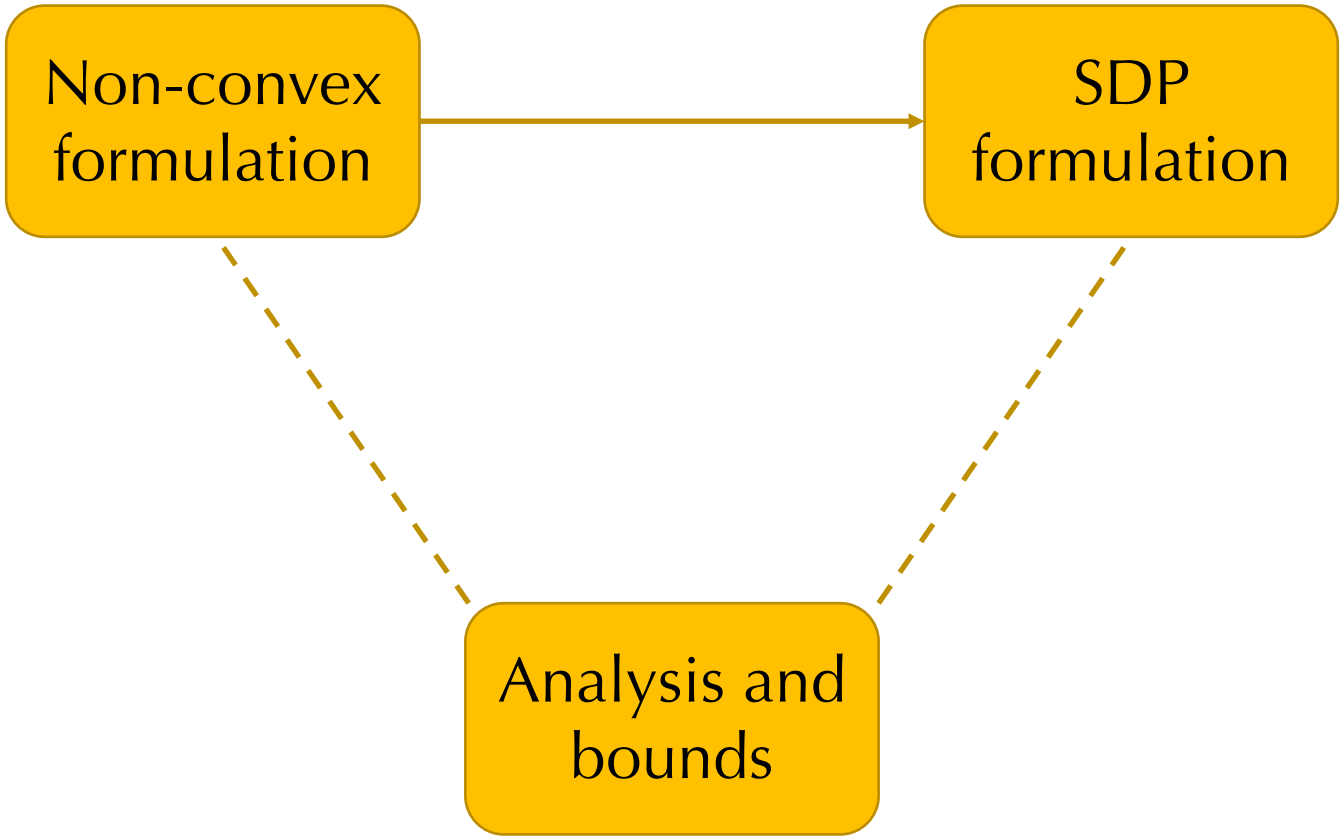
Target state

Feasible region: a spurious state that corresponds to an attack "within the budget"

Infeasible region: a spurious state that corresponds to an attack "out of the budget"

"Curse of dimensionality": requires searching over (#discretization levels)^{#bus*2} number of points (>10^200 for a 100-bus system)

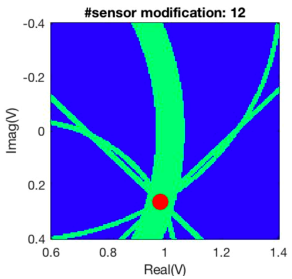
Mathematical program



False data injection attack

Assumption: the attacker can access grid topology and measurements (not necessarily able to modify them)

$$\min_{\tilde{\boldsymbol{v}} \in \mathbb{C}^{n_b}, \boldsymbol{b} \in \mathbb{R}^{n_m}} \left\| \tilde{\boldsymbol{v}} - \boldsymbol{v}_{\mathrm{tg}} \right\|_2^2 \quad \cdots\cdots \text{“Closeness” to target}$$



$$\text{s.t.} \quad b_j = \left\langle \tilde{\boldsymbol{v}} \tilde{\boldsymbol{v}}^*, \boldsymbol{M}_j \right\rangle - m_j \quad \cdots\cdots \text{Stealth attack}$$

$$\left\| \boldsymbol{b} \right\|_0 \leq K \quad \cdots\cdots \text{Limited budget } K$$

- Two challenges: quadratic equality & cardinality constraint
- Cannot be solved efficiently due to nonconvexity

Convexification

- **Equivalent reformulation:**
  - Replace $\tilde{\boldsymbol{v}}\tilde{\boldsymbol{v}}^*$ with $\boldsymbol{W} \in \mathbb{C}^{n_b \times n_b}$
  - The augmented matrix $\mathbf{Z} = \begin{bmatrix} 1 & \tilde{\boldsymbol{v}}^* \\ \tilde{\boldsymbol{v}} & \boldsymbol{W} \end{bmatrix}$ is PSD and rank 1.
- **Relaxation:** relax the rank 1 constraint
- **Penalty:** add penalty for the rank and cardinality constraint.

$$\begin{aligned} \min\ & \langle \boldsymbol{W}, \boldsymbol{M}_{tg} \rangle + 2Re\{\boldsymbol{v}_{tg}^* \tilde{\boldsymbol{v}}\} + \alpha \|\boldsymbol{b}\|_1 + \langle \boldsymbol{W}, \boldsymbol{M}_{pen} \rangle \\ \text{s.t.}\quad & b_j = \langle \boldsymbol{W}, \boldsymbol{M}_j \rangle - m_j \\ & \begin{bmatrix} 1 & \tilde{\boldsymbol{v}} \\ \tilde{\boldsymbol{v}} & \boldsymbol{W} \end{bmatrix} \succcurlyeq 0 \end{aligned}$$

Observable region

For a complex vector $\boldsymbol{v}$ and matrix $\boldsymbol{M}$, define

$$\underline{\boldsymbol{v}} = \begin{bmatrix} \Re(\boldsymbol{v}) \\ \Im(\boldsymbol{v}) \end{bmatrix}, \quad \underline{\boldsymbol{M}} = \begin{bmatrix} \Re(\boldsymbol{M}) & -\Im(\boldsymbol{M}) \\ \Im(\boldsymbol{M}) & \Re(\boldsymbol{M}) \end{bmatrix}$$

Definition: a state $\boldsymbol{v}$ belongs to the **observable region** $\mathcal{O}(\mathcal{M})$ for a given set of measurements $\mathcal{M}$, if the following matrix has full column rank:

$$\boldsymbol{J}(\boldsymbol{v}) = \left[\left(\underline{\boldsymbol{M}}_1 + \underline{\boldsymbol{M}}_1^{\mathsf{T}} \right) \underline{\boldsymbol{v}} \quad \cdots \quad \left(\underline{\boldsymbol{M}}_m + \underline{\boldsymbol{M}}_m^{\mathsf{T}} \right) \underline{\boldsymbol{v}} \right]^{\mathsf{T}}$$

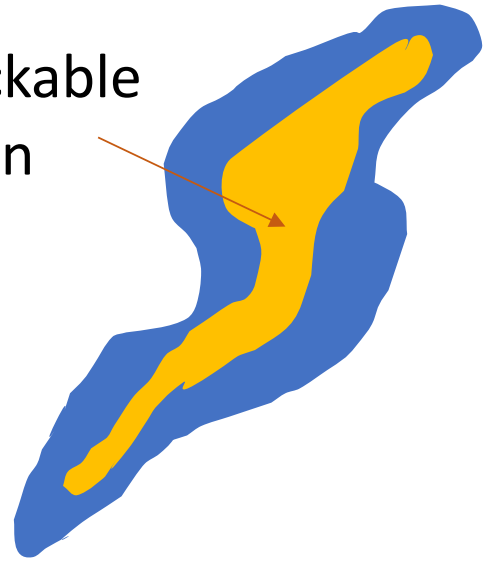
- Guarantees invertability in the neighborhood of the measurement function $\boldsymbol{\Phi}: V \rightarrow M$
- State-estimator agnostic

Attackable region

Definition: a state $\boldsymbol{v}_{at}$ belongs to the **attackable region** $\boldsymbol{\mathcal{A}}(\mathcal{M}, K)$ for a given set of measurements $\mathcal{M}$, if there exists some $\boldsymbol{v}_{tg}$, $\boldsymbol{M}_{pen}$ and attack $\boldsymbol{b}$ under budget K such that $(\boldsymbol{v}_{at}, \boldsymbol{W} = \boldsymbol{v}_{at}\boldsymbol{v}_{at}^*)$ is the unique and optimal solution of the following program:

$$
\begin{aligned}
g(\boldsymbol{b}) = \ & \min \ \langle \boldsymbol{W}, \boldsymbol{M}_{tg} \rangle + 2Re\{\boldsymbol{v}_{tg}^* \tilde{\boldsymbol{v}}\} + \langle \boldsymbol{W}, \boldsymbol{M}_{pen} \rangle \\
& \text{s.t.} \quad b_j = \langle \boldsymbol{W}, \boldsymbol{M}_j \rangle - m_j \\
& \quad \begin{bmatrix} 1 & \tilde{\boldsymbol{v}} \\ \tilde{\boldsymbol{v}} & \boldsymbol{W} \end{bmatrix} \succcurlyeq 0
\end{aligned}
$$

- Obviously, $\boldsymbol{\mathcal{A}}(\mathcal{M}, K_1) \subseteq \boldsymbol{\mathcal{A}}(\mathcal{M}, K_2)$ if $K_1 \leq K_2$



Target state is attackable

Theorem: If the target state $\boldsymbol{v}_{tg}$ is observable, then it is also attackable with enough budget K.

Design the rank penalty
$\boldsymbol{M}_{pen} = -I + \epsilon \boldsymbol{v}_{tg}\boldsymbol{v}_{tg}^{*} + \boldsymbol{L}_0$, where $\boldsymbol{L}_0$ satisfies:

$$\boldsymbol{L}_0 \succcurlyeq 0 \ , \ \mathrm{rank}(\boldsymbol{L}_0) = n-1 \ , \ \boldsymbol{v}_{tg} \in \mathrm{null}(\boldsymbol{L}_0)$$

$$g(\boldsymbol{b}) = \min \langle \boldsymbol{W}, \boldsymbol{M}_{tg} \rangle + 2Re\{\boldsymbol{v}_{tg}^{*}\tilde{\boldsymbol{v}}\} + \langle \boldsymbol{W}, \boldsymbol{M}_{pen} \rangle$$
$$\text{s.t.} \quad b_j = \langle \boldsymbol{W}, \boldsymbol{M}_j \rangle - m_j$$
$$\begin{bmatrix} 1 & \tilde{\boldsymbol{v}} \\ \tilde{\boldsymbol{v}} & \boldsymbol{W} \end{bmatrix} \succcurlyeq 0$$

Proof sketch

- Due to observability, strong duality holds
- Define $\boldsymbol{H}(\boldsymbol{\xi}) = \boldsymbol{L}_0 + \sum_i \xi_i \boldsymbol{M}_i$, using KKT condition

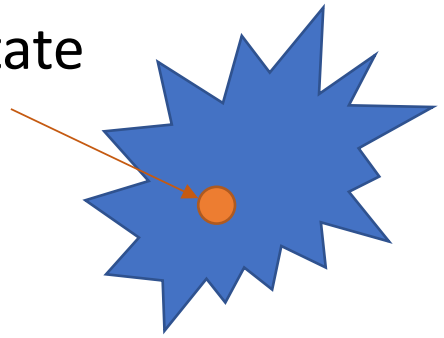
$$H(\boldsymbol{0}) = \boldsymbol{L}_0 \succcurlyeq 0 \quad , \quad \boldsymbol{H}(\boldsymbol{0})\boldsymbol{W} = 0 \quad , \quad \boldsymbol{H}(0)\boldsymbol{v}_{tg} = 0$$

- Perturbation analysis for uniqueness

Target state is attackable

Theorem: The intersection between the attackable region and the observable region forms an open set.

- Proof uses an eigenvalue perturbation argument that depends on the second smallest eigenvalue of $\boldsymbol{L}_0$



Theorem: For the designed $\boldsymbol{M}_{pen}$ and α in a suitable range, the solution satisfies the rank-1 condition: $\boldsymbol{W} = \tilde{\boldsymbol{v}}\tilde{\boldsymbol{v}}^*$.

- Parameter α controls the sparsity, ϵ controls the rank penalty

Performance bounds

Theorem: For $\alpha \geq 2\|\partial g(\boldsymbol{b}^*)\|_\infty$, the difference between the SDP solution $\widehat{\boldsymbol{b}}$ and the original nonconvex solution $\boldsymbol{b}^*$ is bounded by:

$$-2\alpha\|\Delta_S\|_1 \leq g(\widehat{\boldsymbol{b}}) - g(\boldsymbol{b}^*) \leq \alpha\left(\|\Delta_S\|_1 - \|\Delta_{S^c}\|_1\right)$$

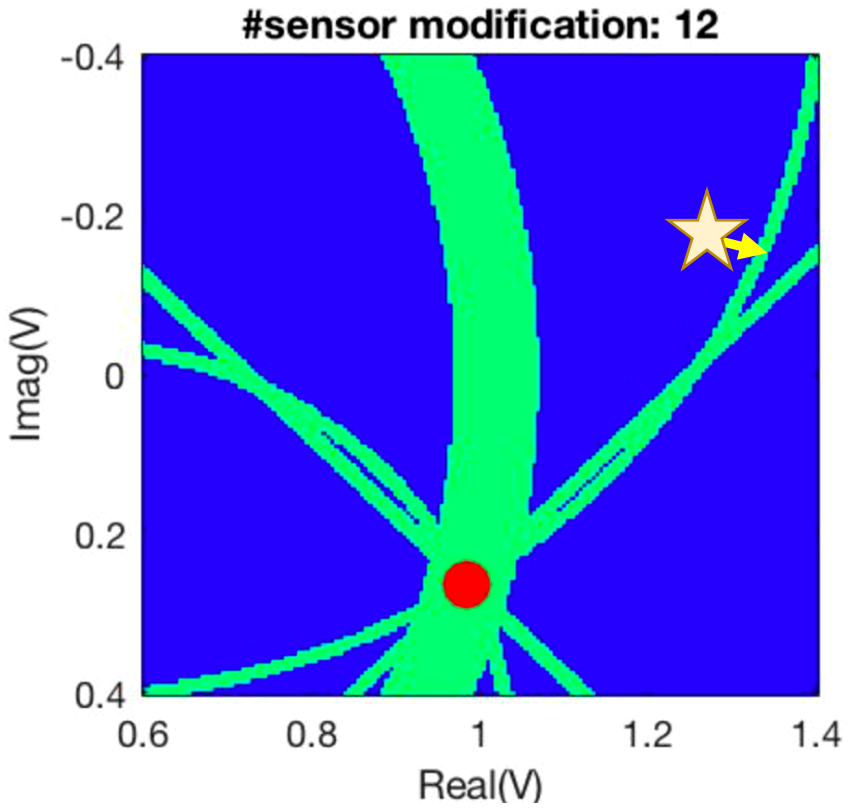
where $\Delta = \widehat{\boldsymbol{b}} - \boldsymbol{b}^*$ and S is the support of $\boldsymbol{b}^*$.

- Trade-off between attack sparsity and the attack outcome
- Proof based on the convexity of $g(\boldsymbol{b})$ and results from M-estimation

$$g(\boldsymbol{b}) = \min \langle \boldsymbol{W}, \boldsymbol{M}_{tg} \rangle + 2Re\{\boldsymbol{v}_{tg}^* \tilde{\boldsymbol{v}}\} + \langle \boldsymbol{W}, \boldsymbol{M}_{pen} \rangle$$
$$\text{s.t.} \quad b_j = \langle \boldsymbol{W}, \boldsymbol{M}_j \rangle - m_j$$
$$\begin{bmatrix} 1 & \tilde{\boldsymbol{v}} \\ \tilde{\boldsymbol{v}} & \boldsymbol{W} \end{bmatrix} \succcurlyeq 0$$

Experiments

Motivation case revisited



True state

Target state

Feasible region: a spurious state that corresponds to an attack "within the budget"

Infeasible region: a spurious state that corresponds to an attack "out of the budget"

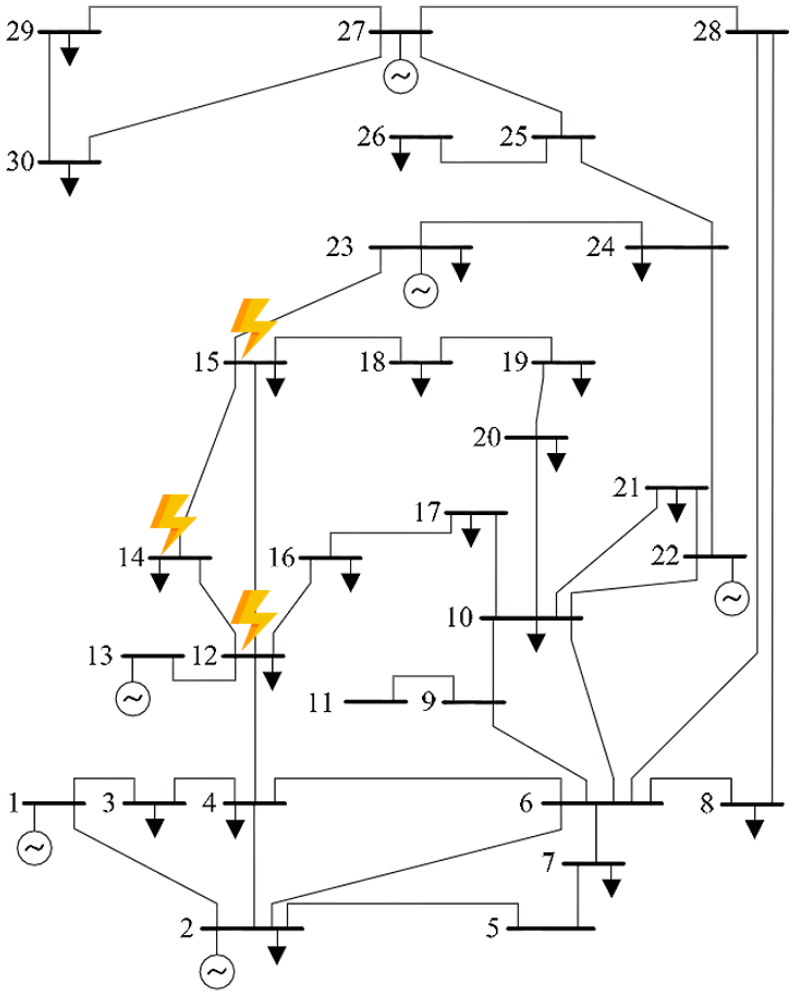
Solved exactly by SDP relaxation!

Experiment on IEEE 30-bus

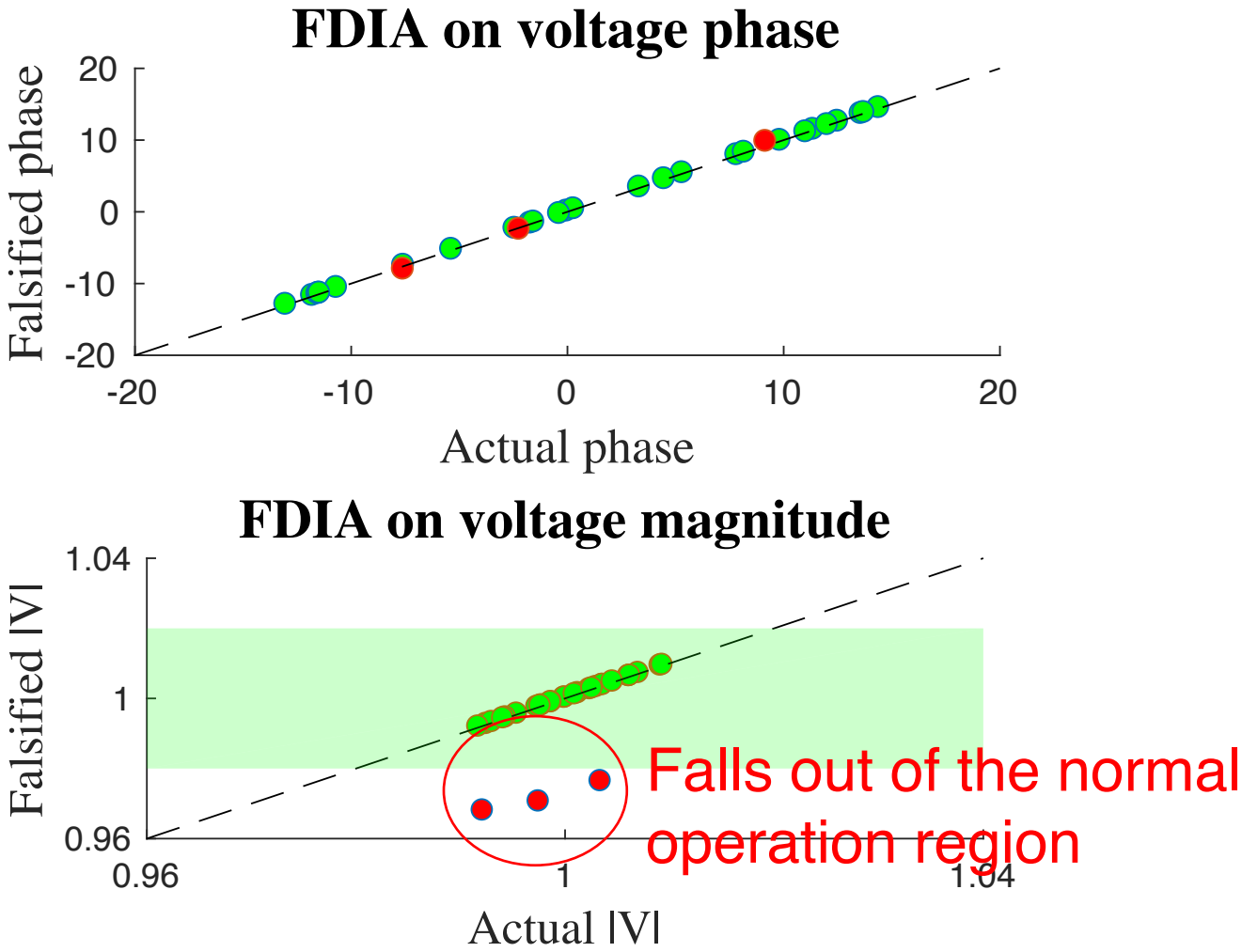
- Target attack:

$$h(\tilde{\boldsymbol{v}}) = \left\| \tilde{\boldsymbol{v}} - \boldsymbol{v}_{tg} \right\|_2^2$$

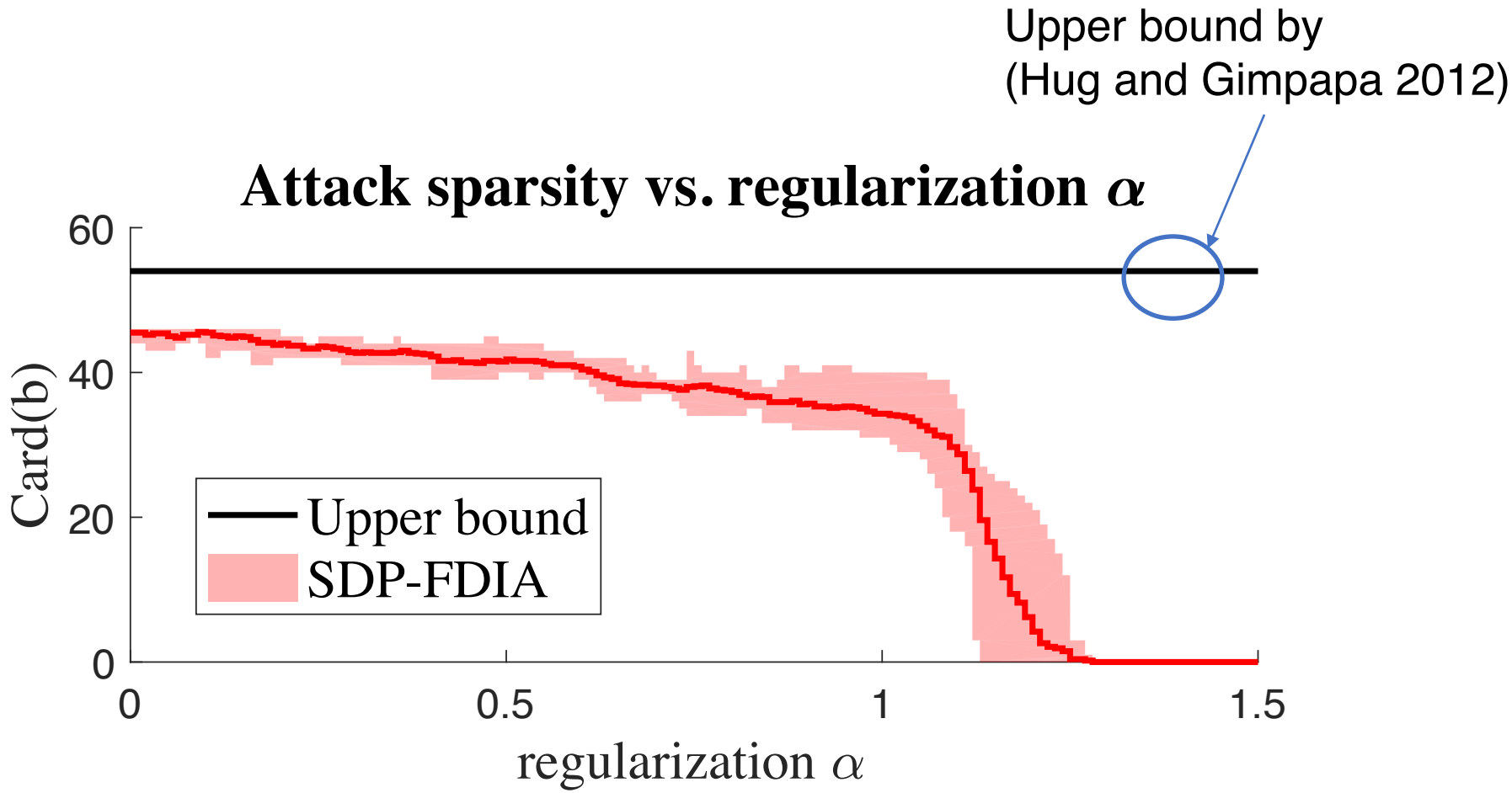
- Solve for the stealth attack $\boldsymbol{b}$ using SDP relaxation
- Estimate the spurious state using Gauss-Newton in MATPOWER / SDP relaxation / Least absolute value regression
- Check for BDD



Attacker “tricks” the operator to believe a potential “voltage sag”



Sparsity controlled by the l1 penalty weight α

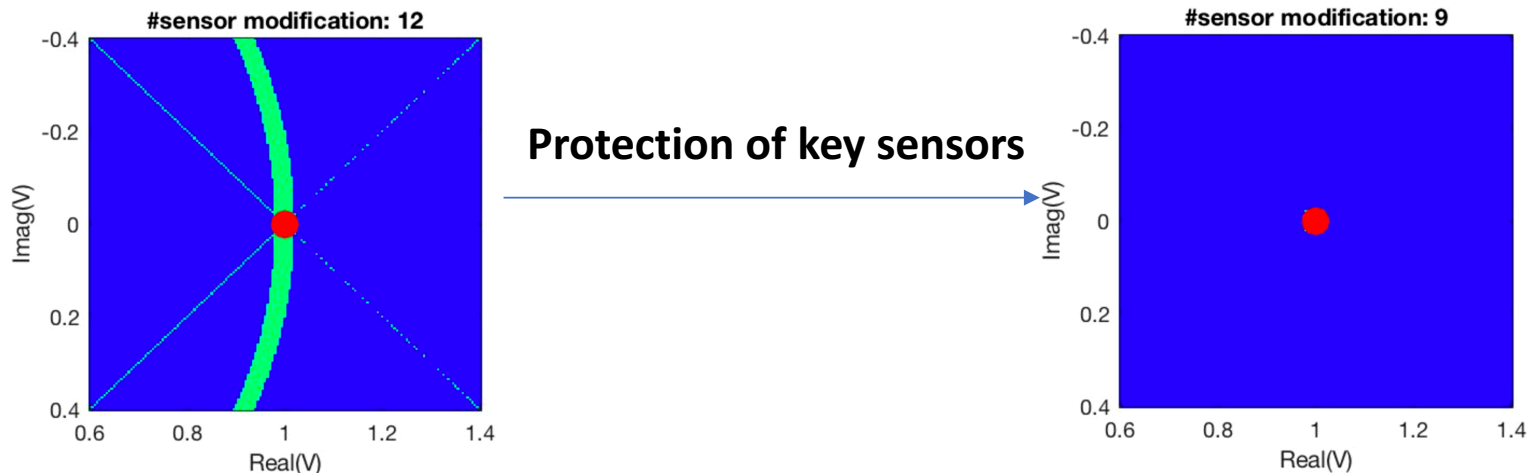


Attack by tampering a small set of sensors

system	α	Rank(Z)	Card(b)	Upper bound	Pass BDD
6-bus	.4	1	18	40	Yes
14-bus	.2	1	16	46	Yes
30-bus	1.16	1	21	54	Yes
39-bus	1.82	1	18	36	Yes
57-bus	.5	1	30	92	Yes

Conclusion & future directions

- Take away messages:
  - Power system AC state estimation is vulnerable to cyber attack
  - The highly nonconvex problem can be solved efficiently using convex relation based on SDP
- Study a protection scheme to protect sensors such that the problem becomes infeasible



Q&A

References:

[1] Ming Jin, Javad Lavaei, and Karl Johansson, A Semidefinite Programming Relaxation under False Data Injection Attacks against Power Grid AC State Estimation, 55th Annual Allerton Conference on Communication, Control, and Computing, 2017.

[2] Ming Jin, Javad Lavaei, and Karl Henrik Johansson, Power Grid AC-based State Estimation: Vulnerability Analysis Against Cyber Attacks, to appear in IEEE Transactions on Automatic Control, 2018.