Are Machine Learning Detectors Sufficient? Exploring Cyberattacks and Defense Strategies in Smart Grids

Bo Liu Ph.D.
Assistant Professor
Washington State University Tri-cities
bo.liu1@wsu.edu

WASHINGTON STATE UNIVERSITY
World Class. Face to Face.

Content

- Introduction to Smart Grids
- Cyberattacks in Smart Grids
- Defense Strategies in Smart Grids
- Highly Stealthy Cyberattacks

Introduction to Smart Grids

National goals of renewable generation

2015: 70 GW of wind and 20 GW of solar

2020: Enable 50% DER in distribution system[1]
Enable 10% Wind (113 GW)[2] and 2% Solar (50 GW)[3]

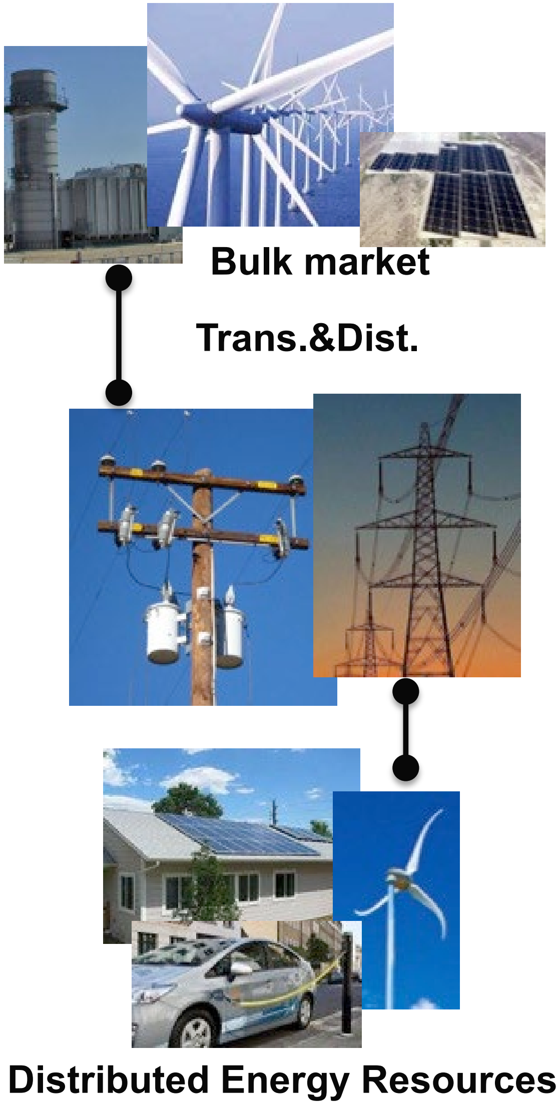
2030: Enable 20% Wind (224 GW)[2] and 14% Solar (330 GW)[3]

2035: Enable 35% Variable Generation[1]
10% of Grid Flexibility comes from Loads, EVs, DER[1]
80% Clean Electricity[4]

Vision of power system

Build a sustainable, secure, and reliable electricity grid that drives a clean-energy economy

Renewable energy brings variability and uncertainty to the power system operation



1. EERE Strategic Plan - http://energy.gov/eere/downloads/eere-strategic-plan
2. Wind Vision Report - http://energy.gov/eere/wind/wind-vision
3. SunShot Vision Study http://www.energy.gov/eere/sunshot/sunshot-vision-study
4. President's Climate Action Plan - https://www.whitehouse.gov/sites/default/files/image/president27sclimateactionplan.pdf

Introduction to Smart Grids

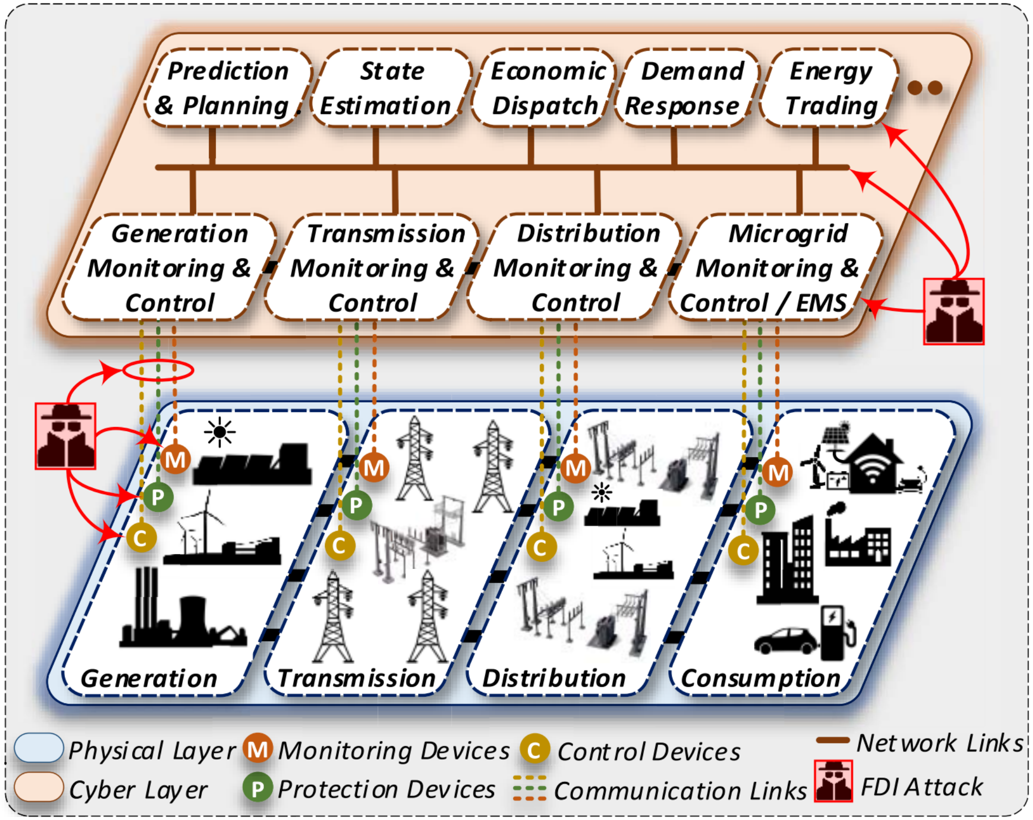
Cyber-physical Smart Grid

Control Center

- Monitoring functions
- Control functions

Physical System

- Generation
- Transmission
- Distribution
- Consumption



Vulnerabilities of smart grids[1]

[1] A. S. Musleh et al "A Survey on the Detection Algorithms for False Data Injection Attacks in Smart Grids," IEEE Transactions on Smart Grid, 2020

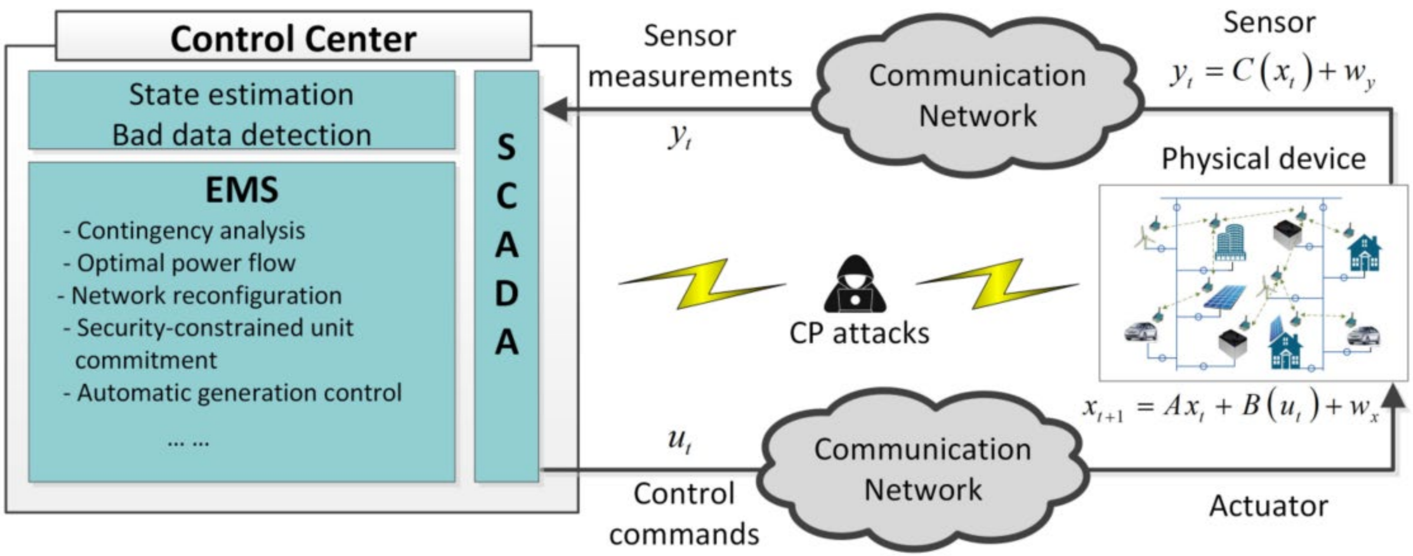
Introduction to Smart Grids

– Information & communication technology and IoT technology

– Huge attack surface for cyber-physical attacks

– 362 power interruption reports related to cyber-physical attacks between 2011 and 2014

– DNP3-SA or IEC-61850 protocols are used in the communication, not all packets are encrypted during communication



Introduction to Smart Grids

Supervisory Control and Data Acquisition (SCADA) System

– SCADA supervises the whole system in real-time

– Collects, analyzes, and visualizes the power system data

– Functions in EMS generate control commands, then SCADA sends these commands to remote substation control devices

Local substation processors

– Remote Terminal Unite (RTU)

– Programmable Logic Controller (PLC)

– Intelligent Electronic Devices (IED)

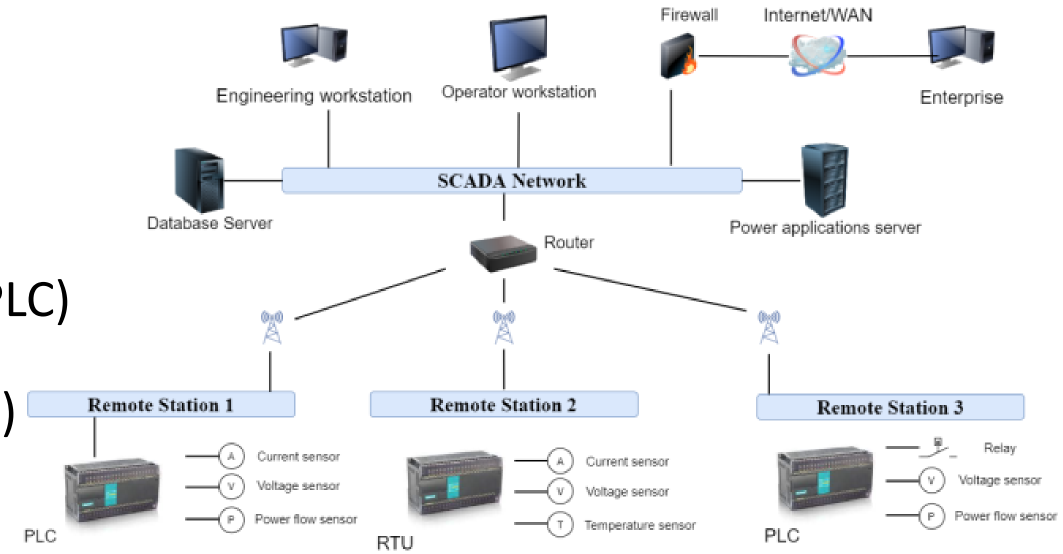
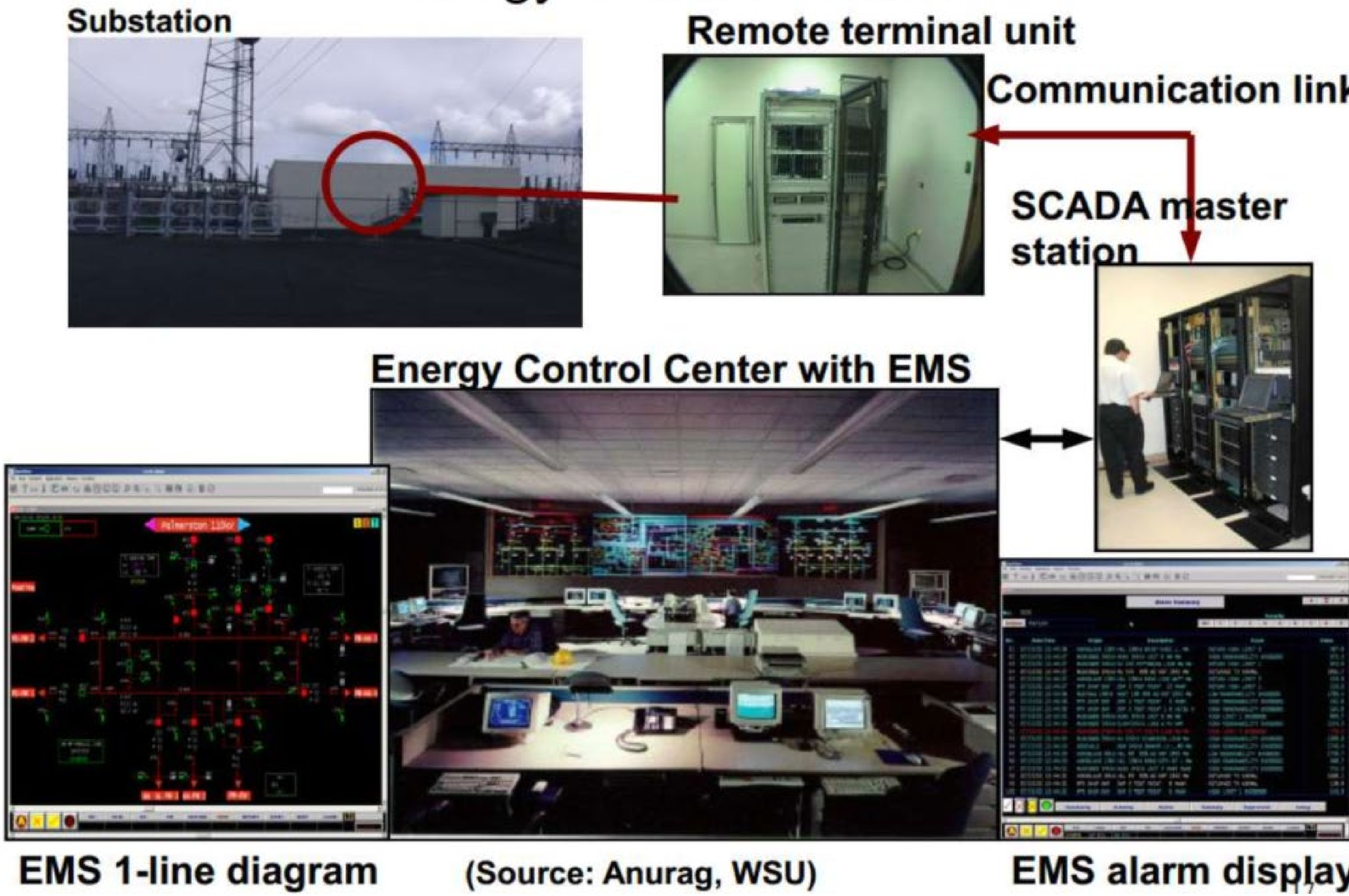


Figure 1.1: *SCADA system network.*

Energy Control Centers



(Source: Anurag, WSU)

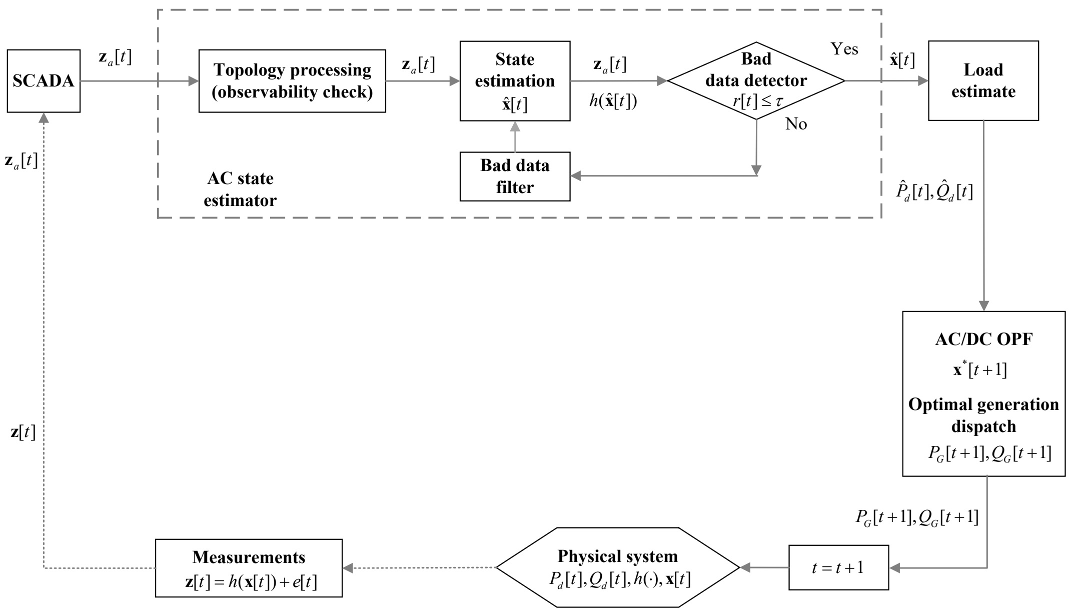
Power System State Estimation (SE)

– SE’s input is SCADA measurements and output is the voltage of each bus

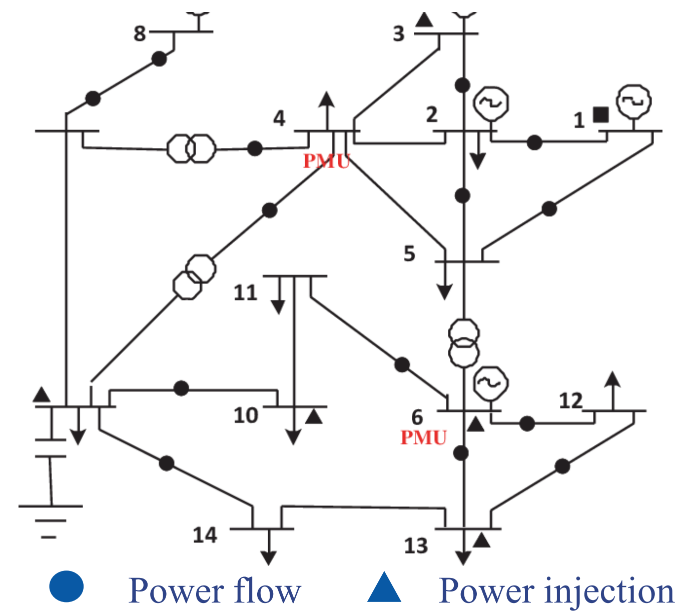
– Weighted Least Square (WLS) is the most common SE method

– Optimal voltage estimates due to measurement noises, errors, redundancy

– Bad data detection (BDD) in SE detects large measurement errors and attacks



Role of SE in power system operation



IEEE 14-bus power system

Blackout from Cyber Incidents

Timeline 2003-2015

2003 Northeast Blackout
Cause: Alarm system failure due to software bug
Consequence: 50 million customers lost electricity

2003 Italy Blackout
Cause: Cascading failures between communication and power infrastructure
Consequence: 56 million customers affected

2007 Arizona Blackout
Cause: Unexpected activation of load shedding program
Consequence: 100K customers lost 400 MW load

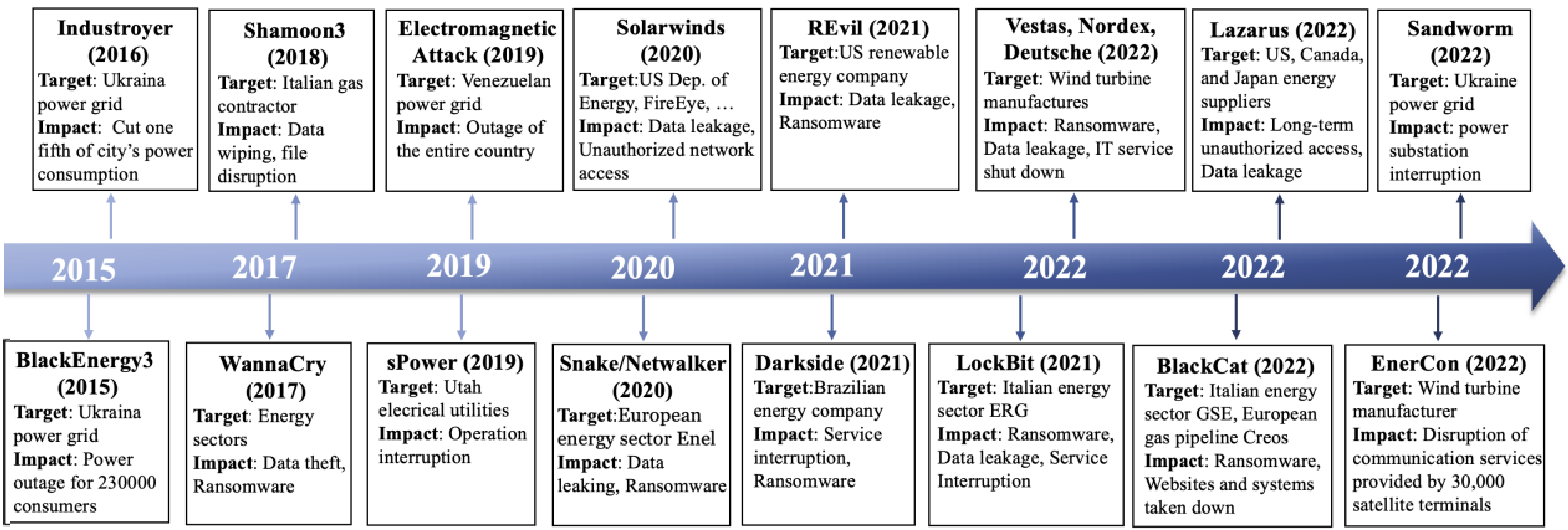
2008 Florida
Cause: Disabled relay protection during a diagnostic process
Consequence: 1 million customers lost 3650 MW load

2011 Southwest Blackout
Cause: Monitoring equipment failure at a substation
Consequence: 2.7 million customers lost electricity

2015 Ukraine Blackout
Cause: Remote cyber intrusions after the malware installation
Consequence: 225K customers lost electricity

Cyberattacks against Smart Grids

Timeline 2015-2022[1]



[1] M. Liu et al., "Enhancing Cyber-Resiliency of DER-Based Smart Grid: A Survey," IEEE Transactions on Smart Grid, 2024

Cyberattacks in Smart Grids

Cyberattacks in Smart Grids

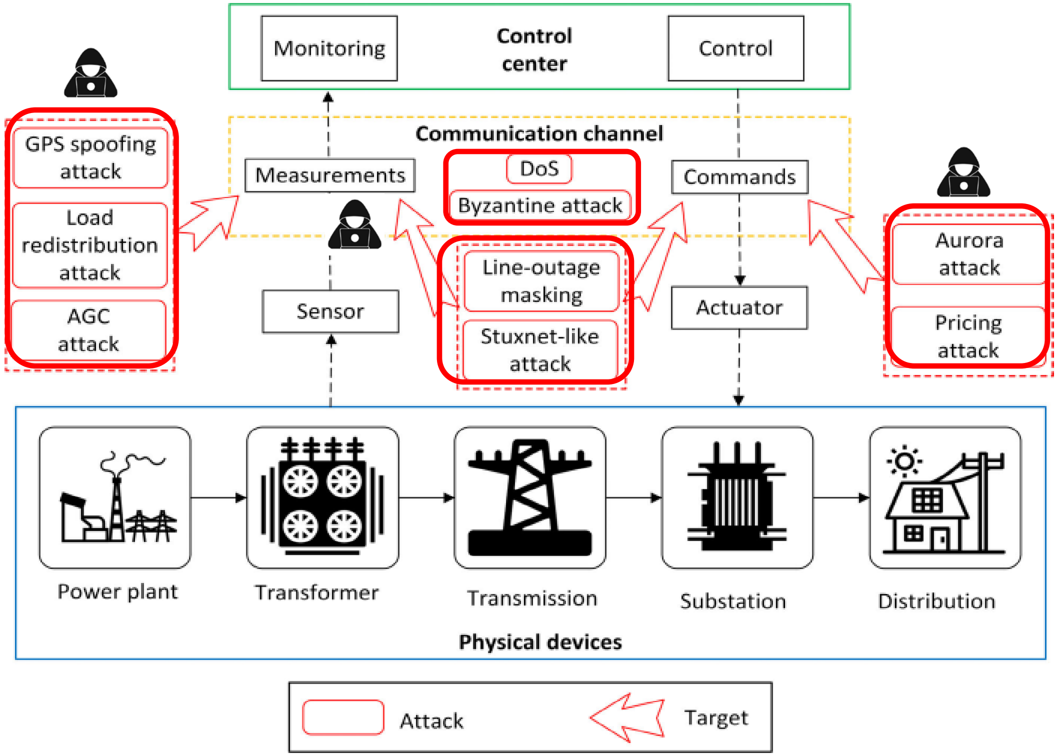
Cyberattack types

– Data availability attacks (DoS attacks)

– Control signal attacks

– Measurement attacks

– Control signal and measurement attacks



Cyberattacks in Smart Grids: Control Signal Attacks

Aurora attacks[1]

– Maliciously open and re-close the circuit breaker of a generator

– Generator protection is delayed to prevent unnecessary tripping

– Re-close the breaker before any protection device kicks in

– High torque and currents caused by re-close cause physical damage

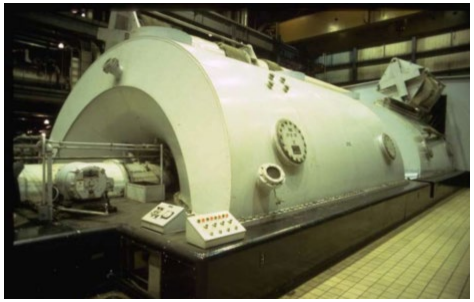
Pricing attacks[2]

– Demand-response is a control mechanism

– Manipulate the price signal, bid prices, and bid quantities

– Cause mismatch between generation and consumption

– Cause system operation stability issues and economic losses



[1] M. Zeller, ''Common questions and answers addressing the aurora vulnerability,'' Schweitzer Eng. Lab., 2011

[2] R. Tan, el al, ''Impact of integrity attacks on real-time pricing in smart grids,'' in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2013

Cyberattacks in Smart Grids: Control Signal Attacks

Windshark attacks against wind turbine

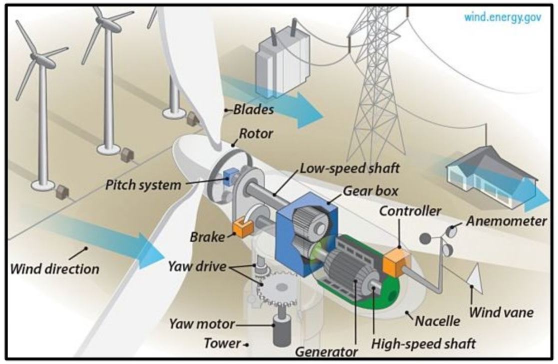
– Wind Energy is the predominant source of renewable energy

– SCADA system controls the wind turbine and substations

– IEC-61400-25 defines communication requirements for wind plant

– Operator can remotely set on/off/idle, and emergency shutdown

– Emergency shutdown (AKA hardstop) induces excessive wear and tear on critical mechanical components



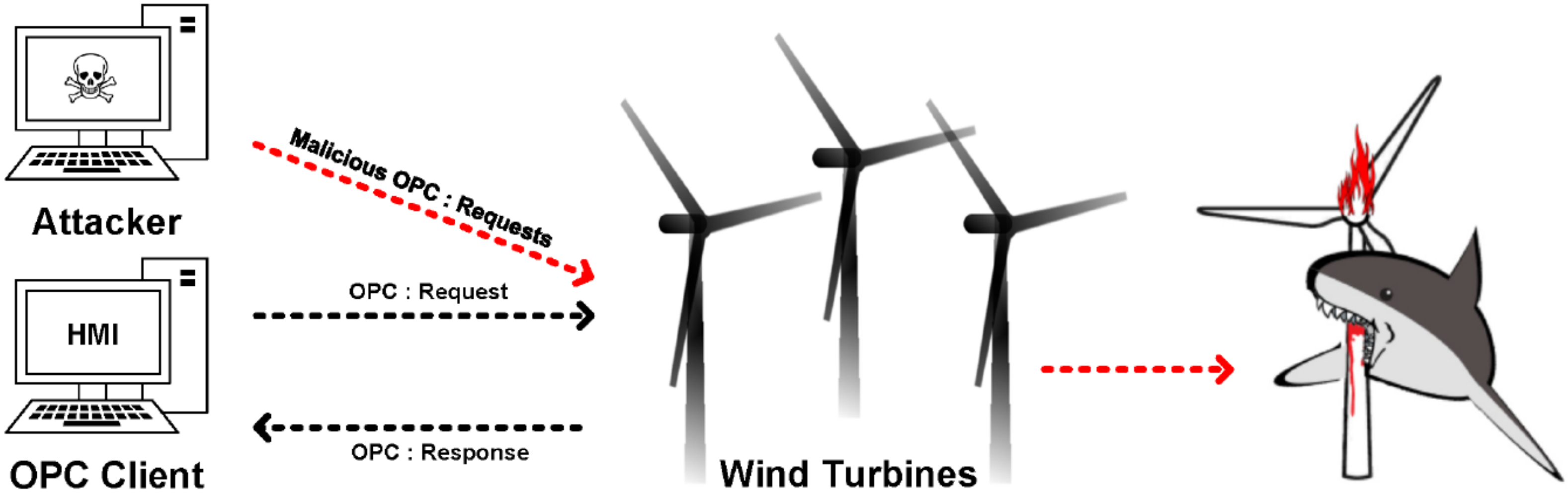
Nacelle

Cyberattacks in Smart Grids: Control Signal Attacks

Windshark attacks against wind turbine

– Python based tool to hijack the turbines and damage them.

– They can list the IP address to target and send commands to the turbine such as emergency shutdown.



"Hard-stop to death"

Jason Staggs et al. "Wind farm security: attack surface, targets, scenarios, and mitigation." International Journal of Critical Infrastructure Protection

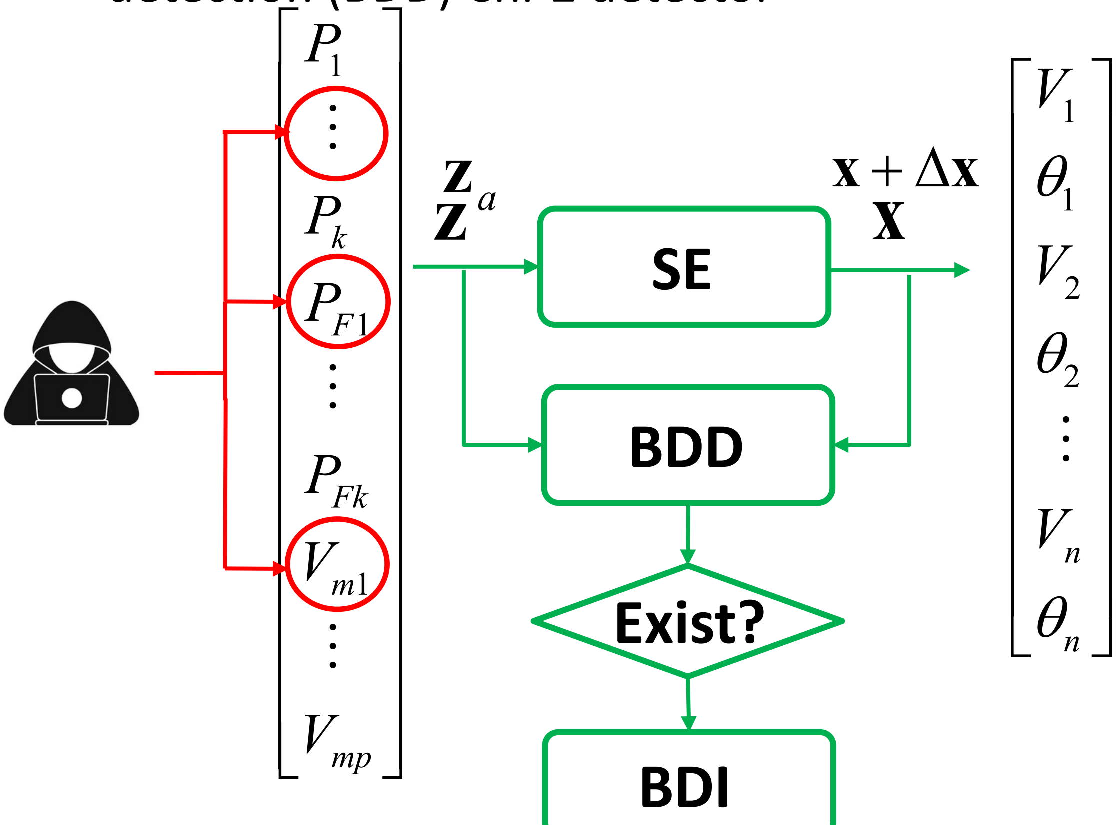
Cyberattacks in Smart Grids: Measurement Attacks

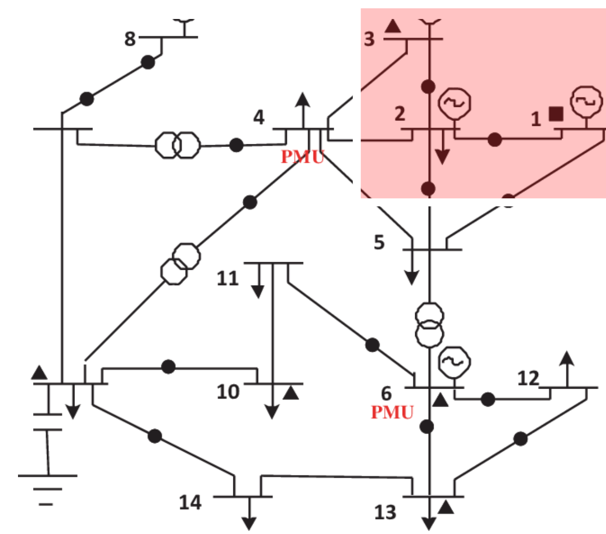
False Data Injection (FDI) attacks

– Manipulate SCADA measurements received by system operators

– Cause a bias $\Delta\mathbf{x}$ in the operator’s estimated voltage

– Follow the physical law of power systems to remain stealthy to bad data detection (BDD) Chi-2 detector





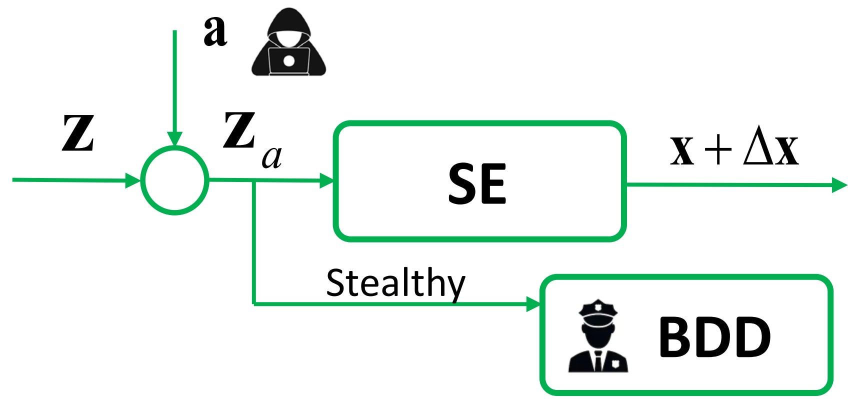
IEEE 14-bus power system

Cyberattacks in Smart Grids: Measurement Attacks

FDI attacks mathematic model

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

$$\mathbf{a} = h(\mathbf{x} + \Delta\mathbf{x}) - h(\mathbf{x})$$



- $\mathbf{x}, \mathbf{z}$ is the state and measurements before attacks
- $\mathbf{x} + \Delta\mathbf{x}, \mathbf{z}_a$ is the state and measurements after attacks
- $h(.)$ is non-linear power flow equations

FDI attack knowledge requirements

- System topology, represented by $h(.)$
- Transmission line impedance, represented by $h(.)$
- Bus voltage $\mathbf{x}$

Control Signal and Measurement Attacks

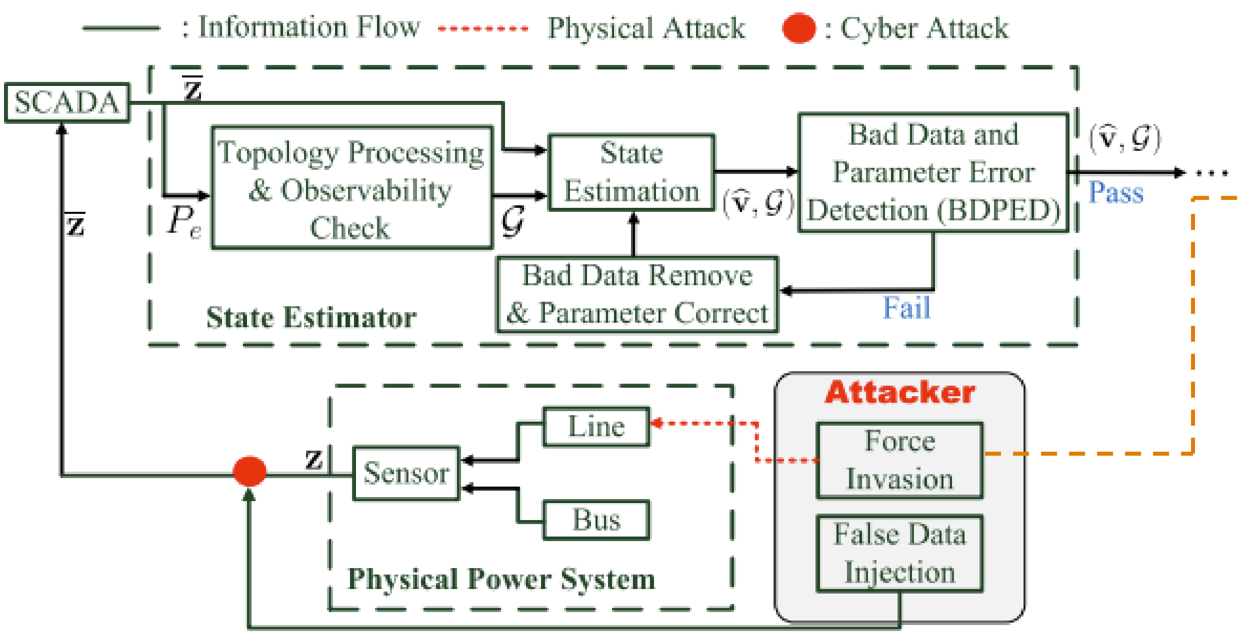
Line outage masking attacks

– Physically disconnect some lines from the attacked area

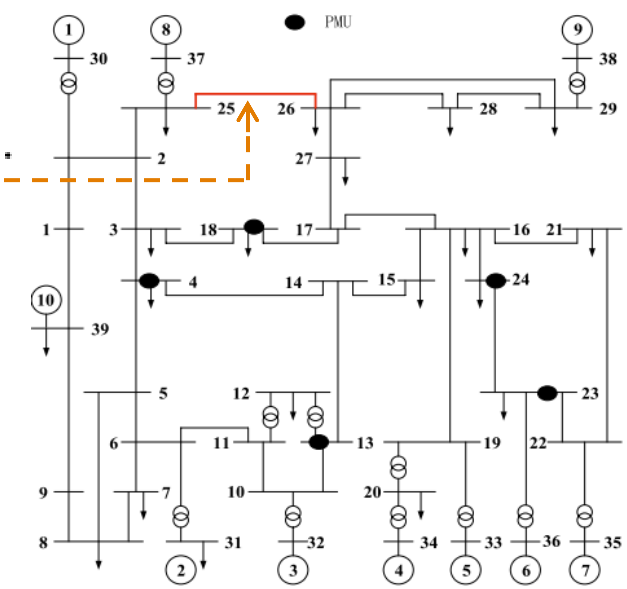
– Mask the measurements within the attacked area by DoS or FDI attacks

– Cause immediate failure and block the operator’s awareness at the same time

– Lead to cascading failures



Line outage masking attack blocks



IEEE 39-bus power system

[1] X. Liu, et al. "Masking Transmission Line Outages via False Data Injection Attacks," in IEEE Transactions on Information Forensics and Security, 2016

Defense Strategies in Smart Grids

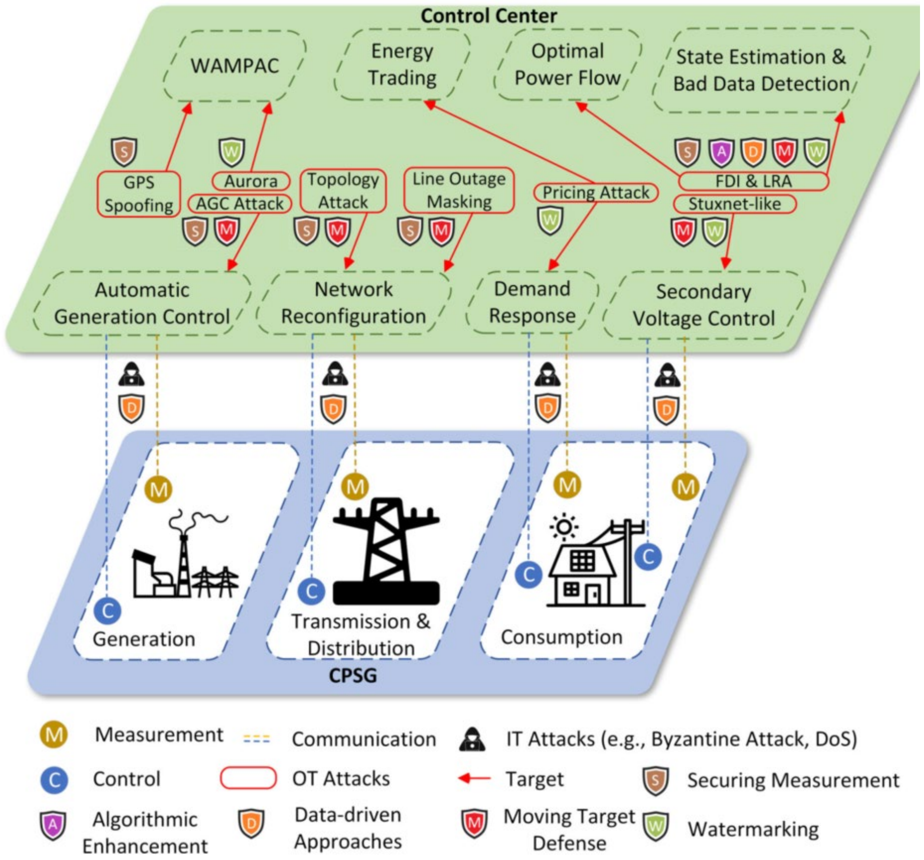
Defense Strategies in Smart Grids

Defense Strategies

– Securing measurement sensors

– Moving target defense

– Data-driven approaches



Defense Strategy : Securing Measurement Sensors

- Cyberattacks require write access to control/measurement signal
- A natural approach is to select and protect critical control/measurement signal
- Protecting all sensors *vs* protecting a few sensors
- Optimally select and protect sensors through graph analysis and optimizations

$$\begin{aligned} \underset{\mathcal{P}\subseteq\mathcal{M}}{\text{minimize}} \quad & |\mathcal{P}| \\ \text{subject to} \quad & rank\left(\mathbf{H}_{\{\mathcal{P}\},*}\right) = rank\left(\mathbf{H}_{\{\mathcal{P}\},\{\mathcal{I}\backslash\mathcal{D}\}}\right) + |\mathcal{D}|, \end{aligned}$$

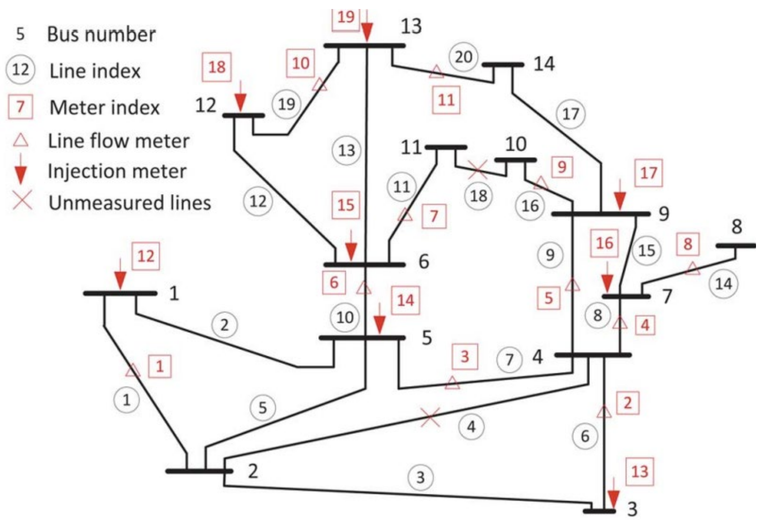


Fig. 2. A measurement placement for the IEEE 14-bus testcase.

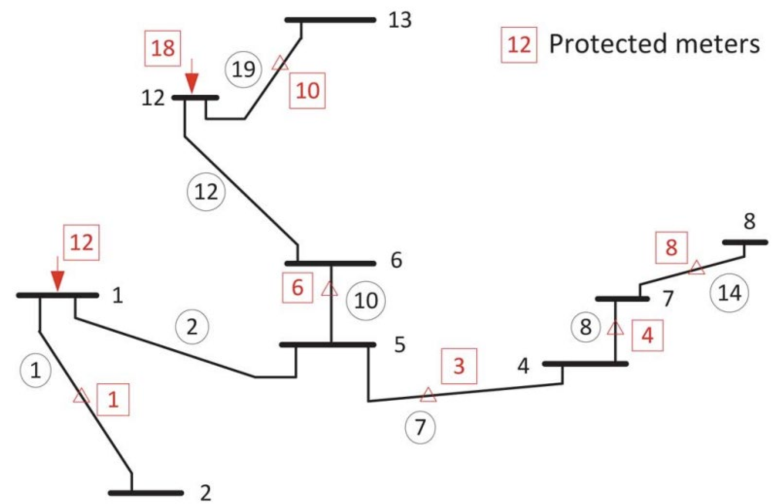


Fig. 3. An illustration of MMST from the IEEE 14-bus testcase.

[1] S. Bi, et al "Graphical Methods for Defense Against False-Data Injection Attacks on Power System State Estimation," IEEE Transactions on Smart Grid, 2014

Defense Strategy: Moving Target Defense

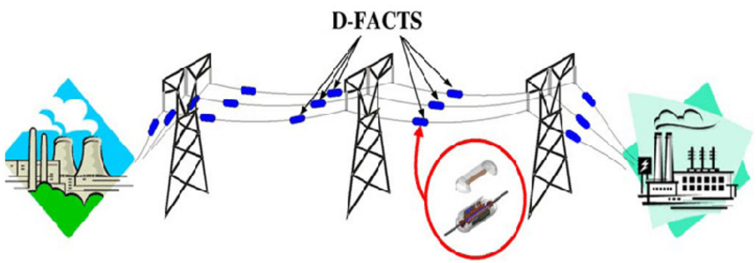
– Actively change the system configuration to invalidate the attacker's knowledge about true system configuration

– Distributed flexible AC transmission system (D-FACTS) devices are attached to transmission lines

– Line impedance can be controlled by D-FACTS devices using encrypted communication

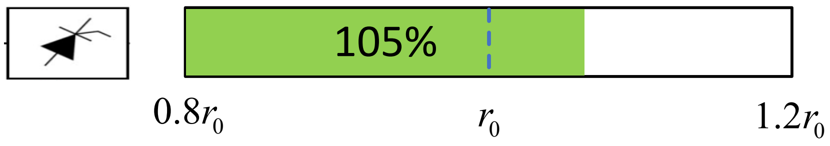
MTD Planning

– Location of D-FACTS devices



MTD Operation

– Setpoints of D-FACTS devices



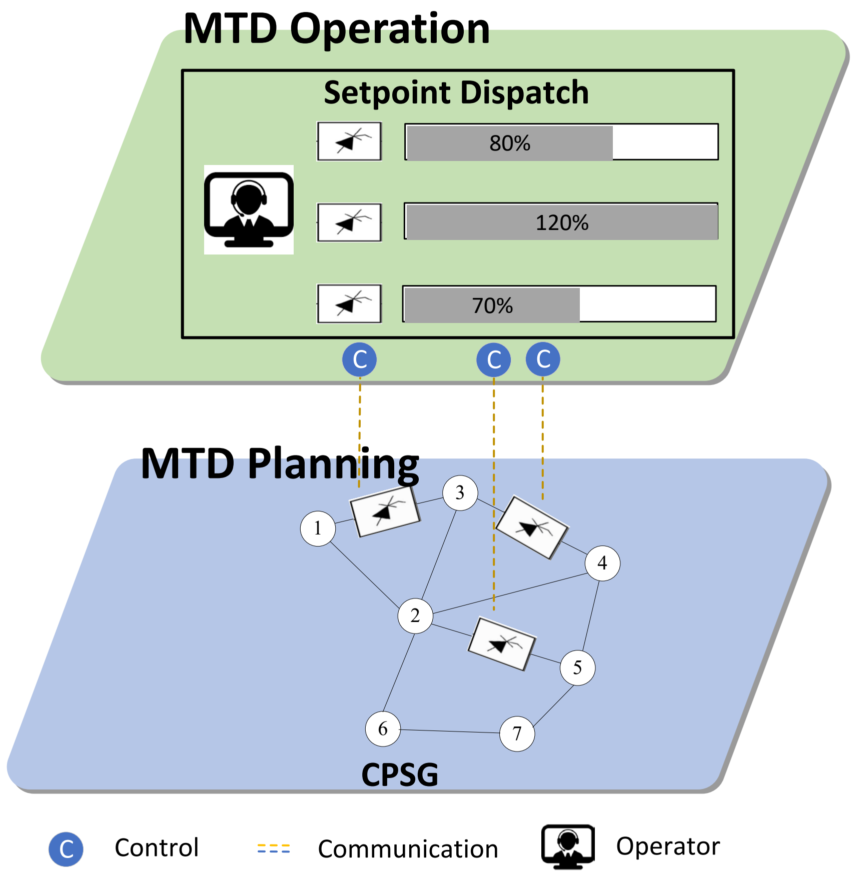
r_0 Original line impedance

Defense Strategy: Moving Target Defense

Moving Target Defense

Metrics

- Cost
- Detection
- Stability
- Stealthiness



Operation Cost

Voltage Stability

Hidden Operation

Capital Cost

Detection Effectiveness

Hidden Placement

- How could operator optimally place D-FACTS devices considering cost and detection?
- How could operator optimally determine setpoints considering operation cost?
- How could operator adjust D-FACTS setpoints to ensure the voltage stability?
- How could operator design a hidden MTD to smart attackers?

Defense Strategy: Moving Target Defense Planning

MTD Planning Objectives

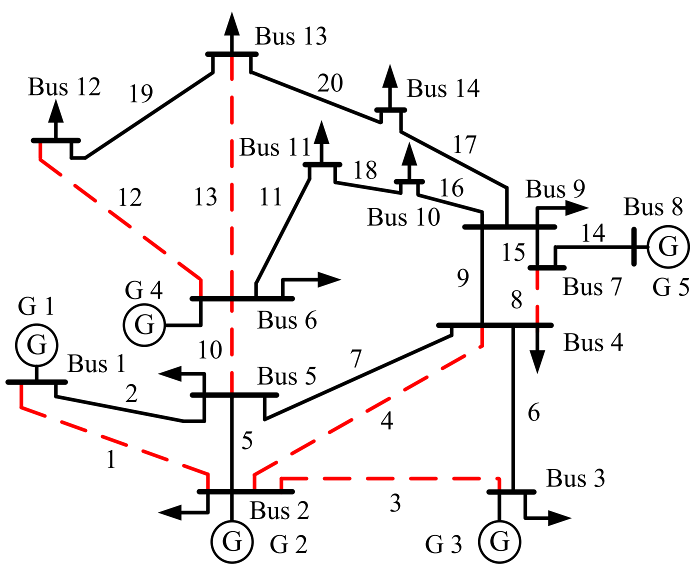
– Maximum detection effectiveness
– Minimum number of D-FACT devices
– Economic benefits of D-FACTS devices

graph theory →

Planning Requirements

– Black graph: spanning tree
– Red graph: no loops
– D-FACTS on lines with high PLIS

MTD Planning Solution



Optimal planning for the IEEE 14-bus system

Planning Algorithm

Algorithm 1: D-FACTS Placement for the Incomplete MTD

Input: The edge-weighted graph $G\{L, \mathcal{E}\}$ of a power grid topology
Output: *DF*: set of D-FACTS lines; *NDF*: set of non-D-FACTS lines

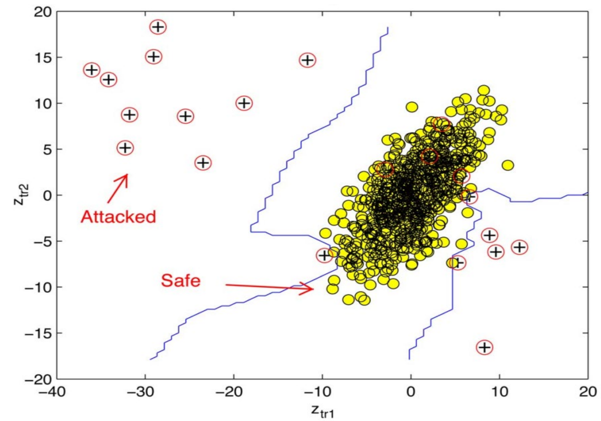
Initialization: $\mathcal{E}_{lp} = \emptyset$ // set of edges in a loop
NDF = find the MST in G // NDF candidates
$DF = \mathcal{E} - NDF$
Generate a graph G_{DF} composed of DF lines and all nodes
while G_{DF} has loops
  Add all edges in the first loop to set $\mathcal{E}_{lp}$
  Arrange edges in $\mathcal{E}_{lp}$ in ascending order of their weights
  for each edge ε in $\mathcal{E}_{lp}$ // start from the lowest-weight edge
    $\varepsilon.\omega = \varepsilon.\omega \times \lambda$ // decrease the positive weight ($\lambda < 1$)
    NDF = find the MST in weight-updated G
    $DF = \mathcal{E} - NDF$
    Update G_{DF} using new DF lines
    if G_{DF} has no loops
      return *DF*, *NDF*
    else if the same loop $\mathcal{E}_{lp}$ still exists in G_{DF}
      $\varepsilon.\omega = \varepsilon.\omega \div \lambda$ //restore ε, try the next edge in loop
    else
      break // loop $\mathcal{E}_{lp}$ doesn't exist, move to the next loop
    end if
  end for
end while
return *DF*, *NDF*

B. Liu et. al. "Optimal D-FACTS Placement in Moving Target Defense Against False Data Injection Attacks," in IEEE Transactions on Smart Grid, 2020

Defense Strategy: Data-driven Approaches

Machine Learning Detectors

- Collect historical measurements
- PCA dimension reduction
- Supervised classification methods



Characteristic of normal and attacked data [1]

Deep Learning Detectors

- High Dimension measurements (Convolutional Neural Network)
- Measurement pattern (Recurrent Neural Network)
- Imbalanced data (Generative Adversarial Network/ Autoencoder)
- Power system configuration (Graph Neural Network)

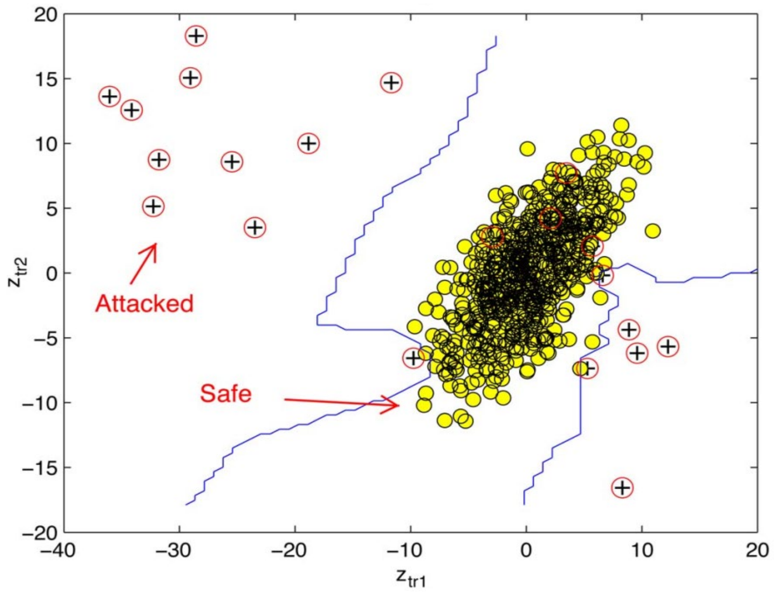
[1] M. Esmalifalak, et.al, “Detecting Stealthy False Data Injection Using Machine Learning in Smart Grid,” IEEE Syst. J., Sep. 2017.

Highly Stealthy Cyberattack

Data-driven Approaches

Machine Learning Detectors

- Collect historical measurements
- PCA dimension reduction
- Supervised classification methods



Characteristic of normal and attacked data

Research Question

- What is the limitation of these ML detectors?
- How could the attacker smartly construct an FDI attack?

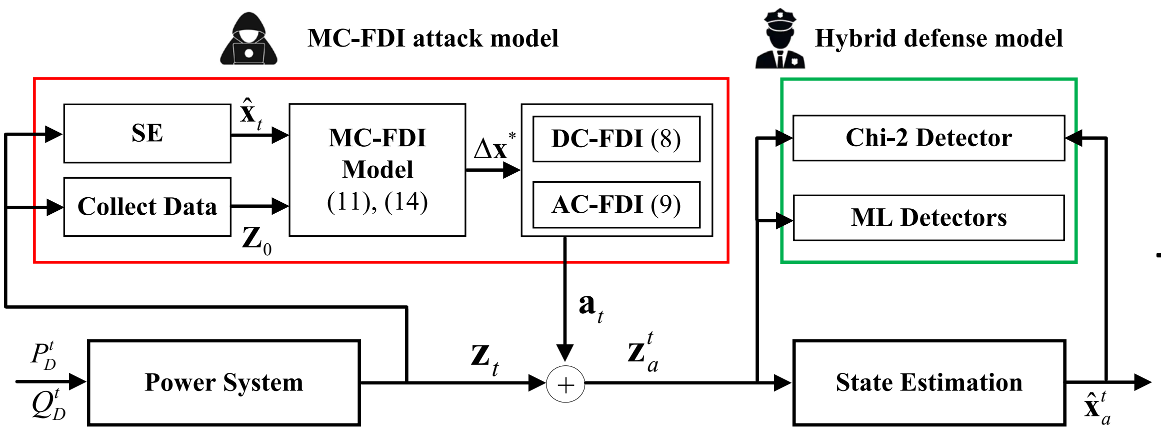
Highly Stealthy + **High Impact**

Matrix-Completion (MC)-FDI Attack

– First FDI attack designed to maintain stealthiness against machine learning (ML) detectors and the BDD

– Apply MC to make compromised measurements consistent with the temporal correlation of historical measurements

– Maximize the incremental voltage to ensure a sufficient negative impact on the power system operation



– BDD Chi-2 Detector

Physical law of power system

– ML detectors

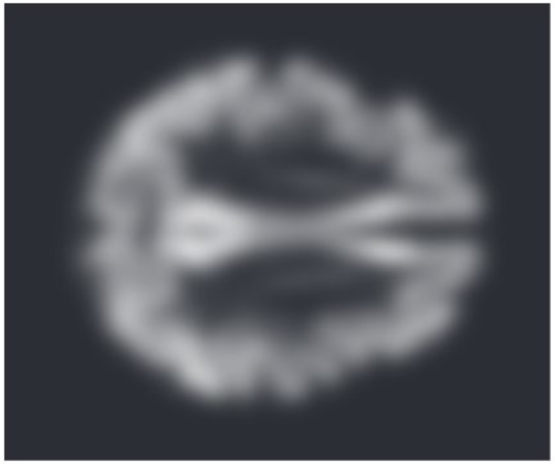
Consistent with historical measurements

B. Liu, H. Wu, Q. Yang, H. Zhang, Y. Liu, and Y. Zhang, "Matrix-Completion-based False Data Injection Attacks against Machine Learning Detectors," *IEEE Transactions on Smart Grid, Sept. 2023.*

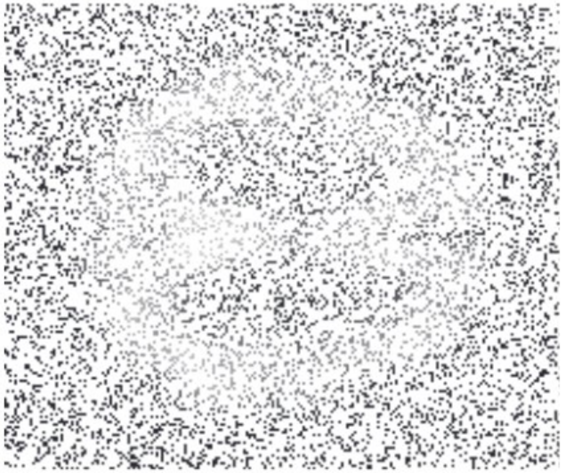
Preliminaries: Matrix Completion (MC)

– A technology aiming to estimate the unknown elements in an incomplete matrix that has a low-rank property

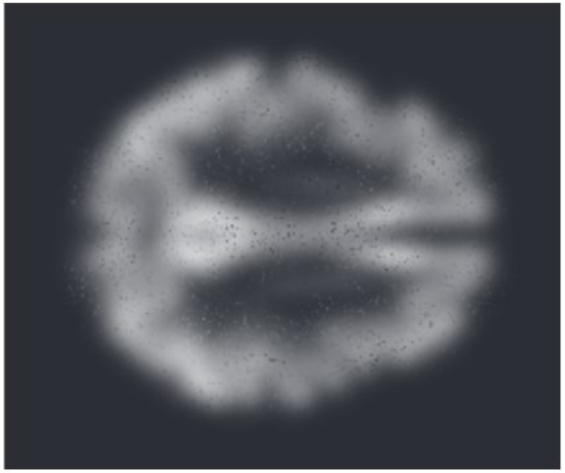
– Original matrix and incomplete matrix



(a) Original matrix



(b) Incomplete matrix $\mathbf{M}$



(c) Completed matrix $\mathbf{X}$

– Formulation

min (**Rank of matrix X**)

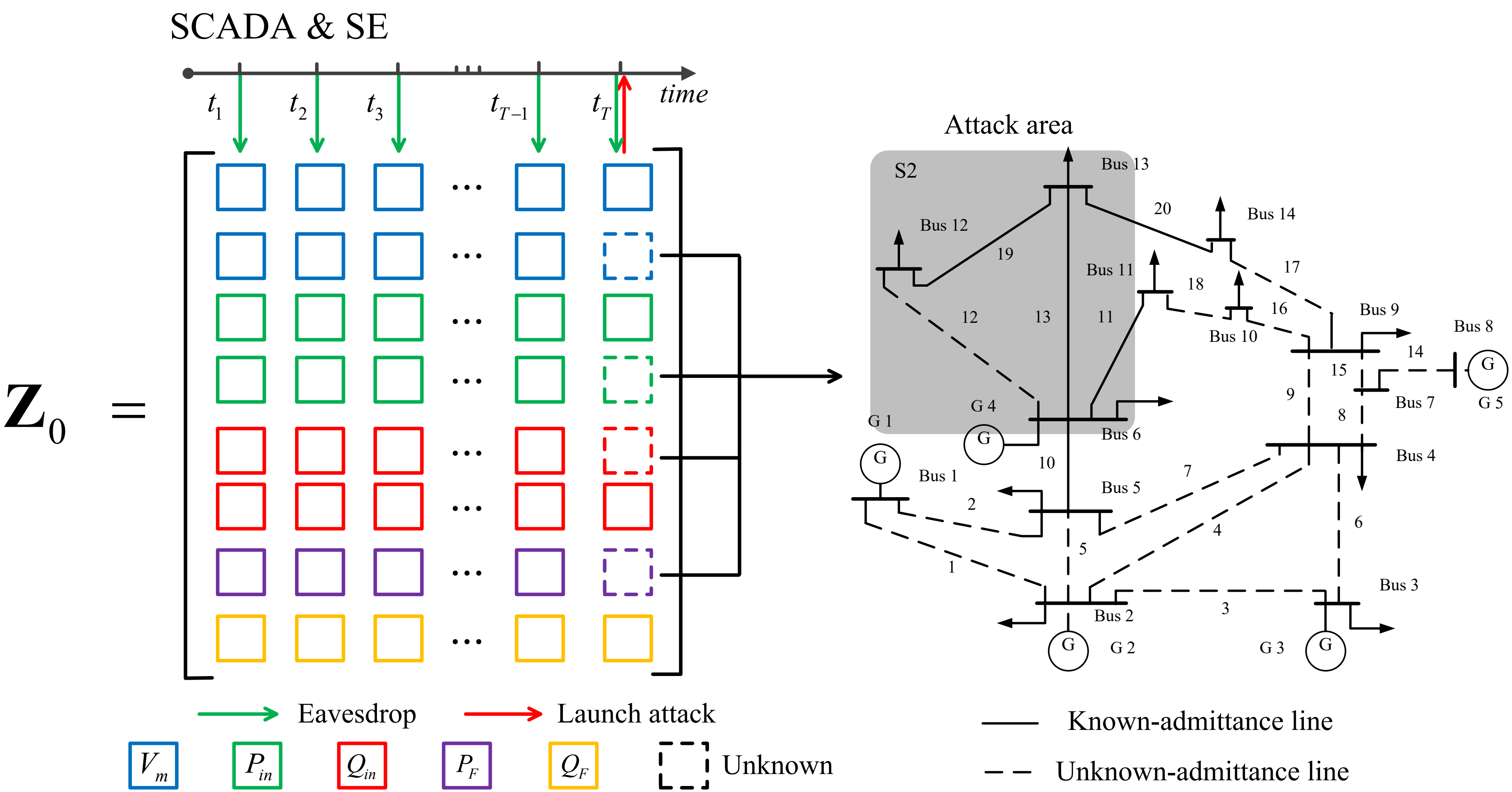
Elements in **X** = **Known elements** in **M**

$$\min_{\mathbf{X}\in\mathbb{R}^{n_1\times n_2}} \quad \|\mathbf{X}\|_* \qquad s.t. \quad \mathbf{X}_\Psi = \mathbf{M}_\Psi$$

– **Research Question:** How can we use MC to construct highly stealthy FDI?

Highly Stealthy MC-FDI Attack

Matrix Formulation



Highly Stealthy MC-FDI Attack

MC-FDI Mathematic Model

min (**Rank of matrix** $\mathbf{Z}_a$)

$$\min_{\Delta\mathbf{x}} \quad \|\mathbf{Z}_a\|_* - \lambda\|\Delta\mathbf{x}\|_1$$

Known $\mathbf{Z}_a$ **= normal measurements** $\mathbf{Z}_0$

$$s.t. \quad \mathbf{Z}_a(i) = \mathbf{Z}_0(i) \qquad i \in idx_0^t$$

Unknown $\mathbf{Z}_a$ **= compromised measurements** $\mathbf{Z}_0$

$$\mathbf{Z}_a(i) = \mathbf{Z}_0(i) + \mathbf{a} \qquad i \in idx_a^t$$

Linearized FDI attack equation

$$\mathbf{a} = \mathbf{H}(\mathbf{x}_T)\Delta\mathbf{x}$$

Voltage deviation constraints (target buses)

$$\Delta\mathbf{x}_{lb}(i) \le \Delta\mathbf{x}(i) \le \Delta\mathbf{x}_{ub}(i) \qquad i \in idx_a^{bus}$$

Voltage deviation constraints (non-target buses)

$$\Delta\mathbf{x}(i) = \mathbf{0} \qquad i \notin idx_a^{bus}$$

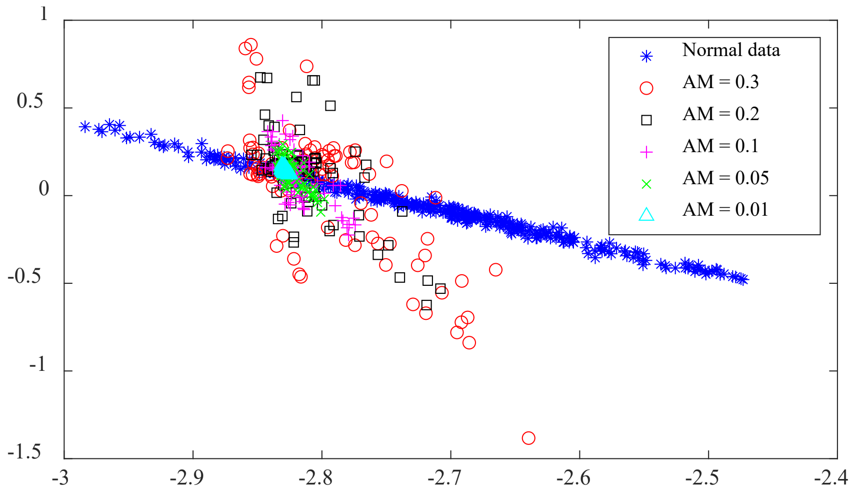
Highly Stealthy MC-FDI Attack

Experiment Results

– Stealthiness of MC-FDI attacks against ML detectors

– 500 Traditional FDI and 100 MC-FDI attacks

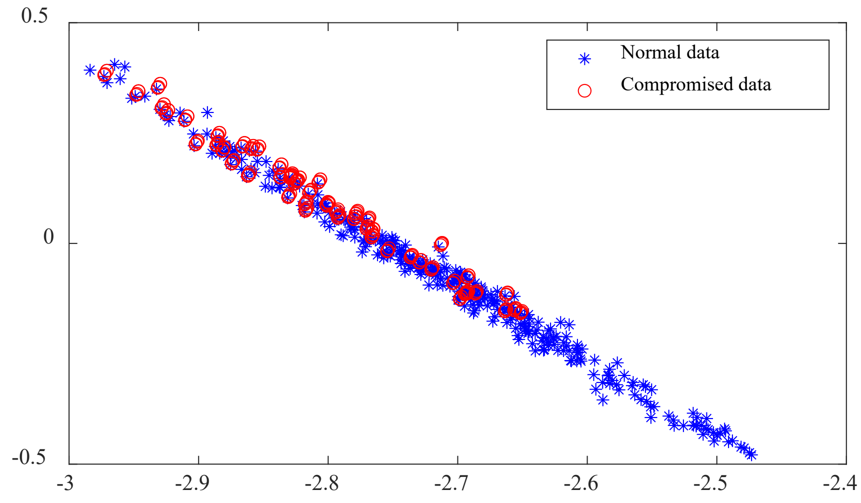
Traditional FDI



PERFORMANCE OF ML DETECTORS ON DETECTING FDI ATTACKS

Detector	AM	Norm	Precision	Recall	F1
SVM	0.1	0.015	0.95	0.35	0.51
	0.3	0.049	0.97	0.74	0.84
ANN	0.1	0.015	0.94	0.52	0.67
	0.3	0.049	0.96	0.78	0.86
LR	0.1	0.015	1.00	0.26	0.42
	0.3	0.049	1.00	**0.70**	0.82

MC-FDI



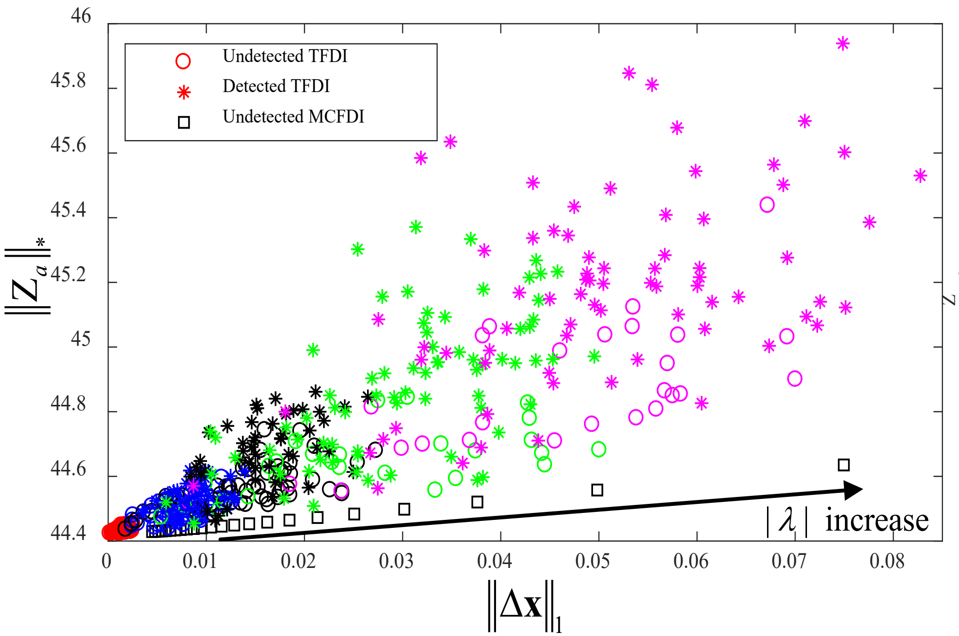
PERFORMANCE OF ML DETECTORS ON DETECTING MC-FDI ATTACKS

Detector	λ	Norm	Precision	Recall	F1
SVM	3.1	0.057	0.67	0.04	0.08
	3.2	0.140	0.78	0.07	0.13
ANN	3.1	0.057	0.67	0.12	0.20
	3.2	0.140	0.77	0.20	0.32
LR	3.1	0.057	1.00	0.02	0.04
	3.2	0.140	1.00	**0.06**	0.11

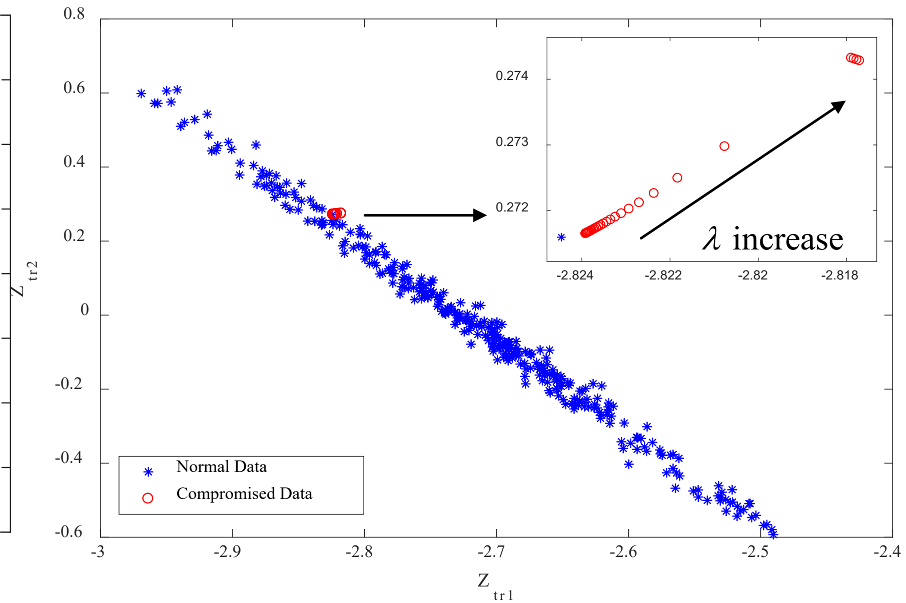
Highly Stealthy MC-FDI Attack

– Impact of weight λ on MC-FDI attacks

$$\min_{\Delta \mathbf{x}} \quad \|\mathbf{Z}_a\|_* - \lambda \|\Delta \mathbf{x}\|_1$$



(a) MC-FDI attacks in norm-norm space



(b) Stealthiness of MC-FDI attacks

Detection of MC-FDI attacks?

– Securing Measurement Sensors

– Moving Target Defense

Thanks