University of Glasgow

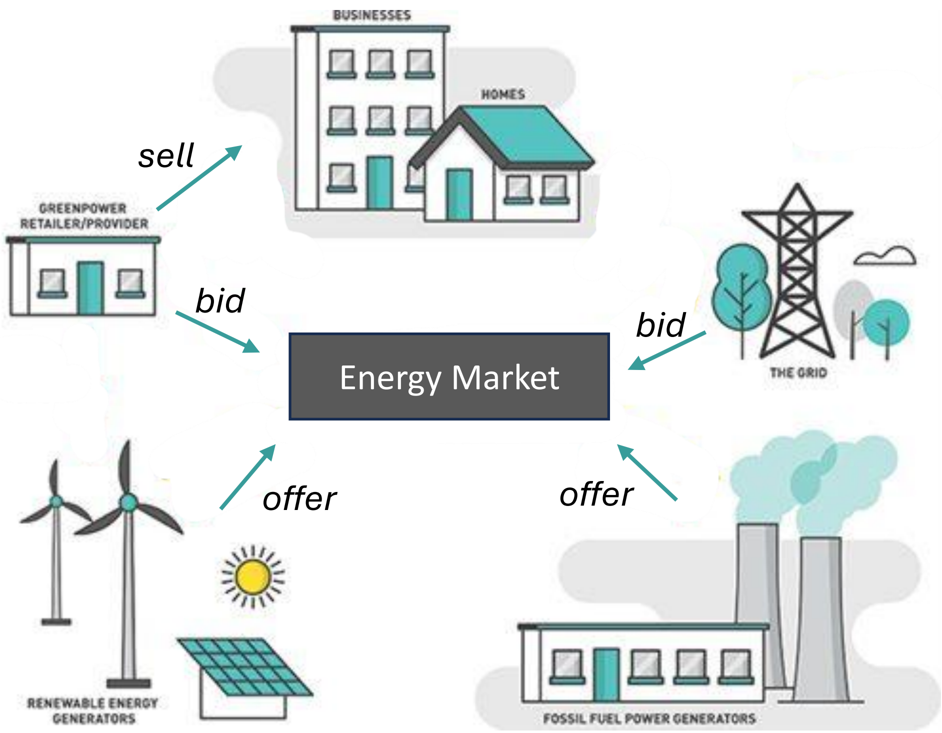
Knowledge & Data Engineering Systems

SMLT: A Synthetic Dataset for Stealthy Manipulation of Energy Market via False Data Injection Attacks

Ghadeer Alsharif, Christos Anagnostopoulos, Angelos Marnerides and Mathaios Panteli

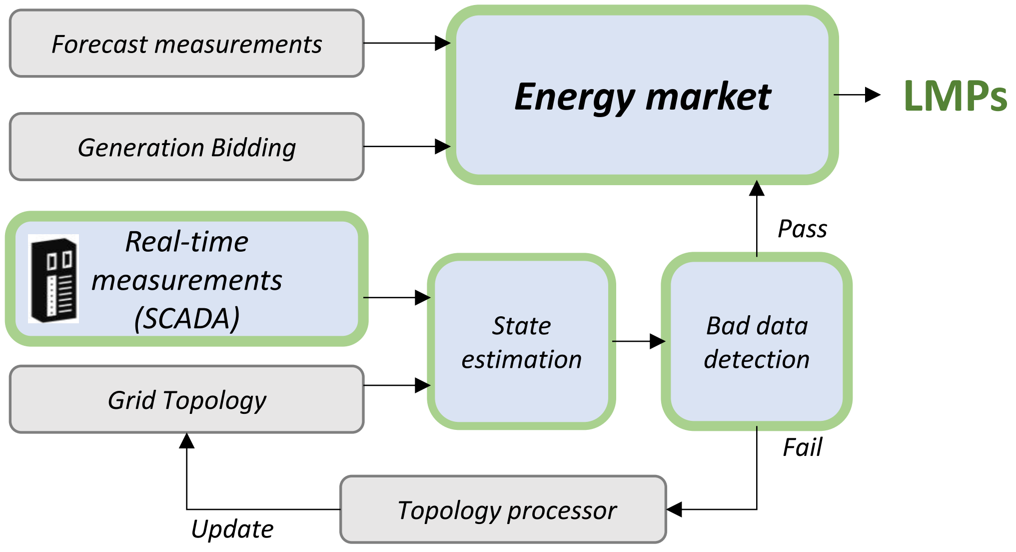
Introduction

- **The Evolution of the Electricity Market:**
  - The electricity industry has undergone a transition towards a competitive framework where participants can bid and offer energy within a dynamic pool.
  - This shift has been driven by the adoption of *Locational Marginal Prices (LMPs)* as the primary mechanism for determining market dynamics.



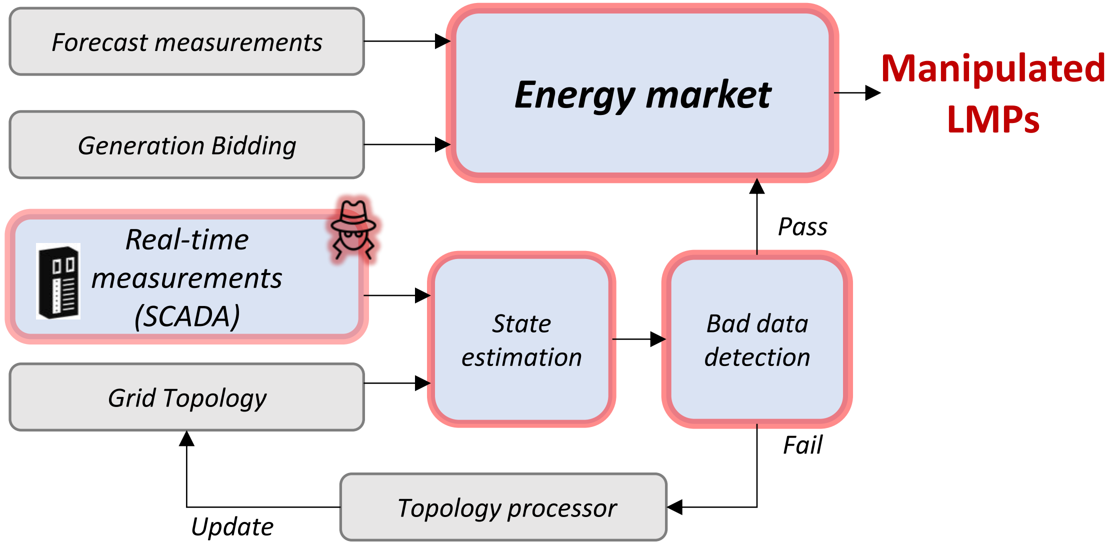
Introduction

- **The Evolution of the Electricity Market:**
  - LMPs reflect the marginal cost of supplying an electricity unit at specific locations within the grid, at any given point in time.
  - LMPs facilitate efficient resource allocation, congestion management, and market equilibrium



Motivation

- **Stealthy False Data Injection Attacks in the Energy Market:**
  - Malicious actors target data transmitted from Remote Terminal Units (RTUs) to the SCADA system.
  - **Objective**: Manipulate market outcomes for financial gain.
  - **Persistence**: Attacks designed to persist over an extended period for long-term gains.
  - **Impact**: Manipulation of state estimation results, skewing LMPs.
  - **Consequences**: Financial losses, inefficient resource allocation, and reduced system efficiency.



Motivation

- *Current research Focus has mainly been either on threat models or physical protection of state estimation*
- *Anomaly detection using model-based grid representations based on predefined behaviors & known attack patterns.*
- *Data-driven detection models are a promising approach for identifying electricity market cyber attacks in real time.*
- *Barriers to AI-Based Anomaly Detection:*
  - *Lack of publicly available datasets for LMP manipulation scenarios.*
  - *Existing benchmark systems simulate markets but: Rarely include adversarial scenarios and Lack labeled time-series data for systematic evaluation*

Contribution

- **S**tealthy **M**anipulated **L**MP **T**imeseries**; SMLT** dataset
  1. *First open-source dataset for stealthy FDIA attacks in electricity markets*
  2. *Incorporates 8 manipulation cases (transmission ratings, system parameters, topology, demand profiles)*
  3. *Hourly resolution time series (up to 20 weeks) with ground-truth labels*
  4. *Open-source FDIA simulation framework built on Matpower*
  5. *In-depth spatio-temporal analysis of LMP manipulation + case study*

Dataset Construction

- ***Baseline System: NPCC****[1]*

- ***Cyberattack Scenarios:***
  - *Transmission Line Rating Attack [2]*
  - *Critical Parameter Attack [3]*
  - *Cyber-Topology Attack. [4]*
  - *Ramp-Induced Data Attack [5]*
  - *Load-Altering Attack [6]*
  - *Aggregator-Based Strategic Curtailment [7]*

[1] Zhang, Q. and Li, F., 2023. A Dataset for Electricity Market Studies on Western and Northeastern Power Grids in the United States. Scientific Data, 10(1), p.646.
[2] Ye, H., Ge, Y., Liu, X. and Li, Z., 2015. Transmission line rating attack in two-settlement electricity markets. IEEE Transactions on Smart Grid, 7(3), pp.1346-1355.
[3] Xu, H., Lin, Y., Zhang, X. and Wang, F., 2020. Power system parameter attack for financial profits in electricity markets. IEEE Transactions on Smart Grid, 11(4), pp.3438-3446.
[4] Liang, G., Weller, S.R., Zhao, J., Luo, F. and Dong, Z.Y., 2017. A framework for cyber-topology attacks: Line-switching and new attack scenarios. IEEE Transactions on Smart Grid, 10(2), pp.1704-1712.
[5] Choi, D.H. and Xie, L., 2013. Ramp-induced data attacks on look-ahead dispatch in real-time power markets. IEEE Transactions on Smart Grid, 4(3), pp.1235-1243.
[6] Mohsenian-Rad, A.H. and Leon-Garcia, A., 2011. Distributed internet-based load altering attacks against smart power grids. IEEE Transactions on Smart Grid, 2(4), pp.667-674.
[7] Ruhi, N.A., Dvijotham, K., Chen, N. and Wierman, A., 2017. Opportunities for price manipulation by aggregators in electricity markets. IEEE Transactions on Smart Grid, 9(6), pp.5687-5698.

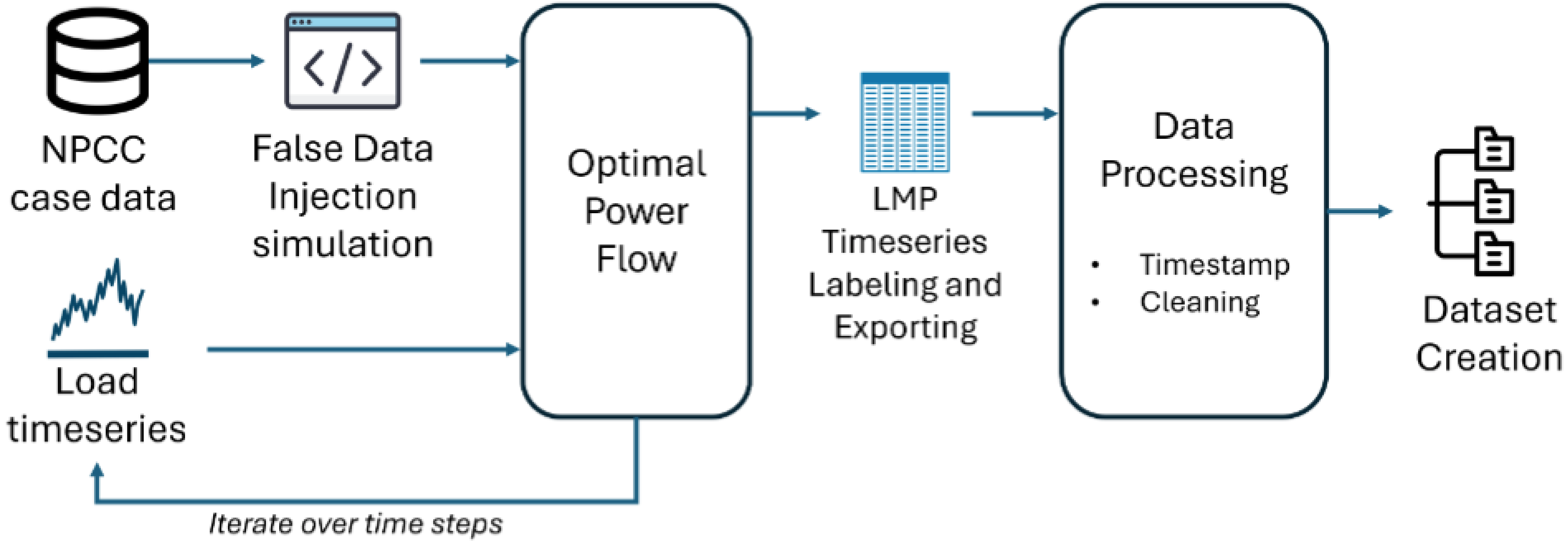
Dataset Construction

- ***Summary of attack cases and their validation outcomes***

	Scenario	$\mathbf{a}_{\mathrm{nom}}$	α (p.u.)	Target Bus	**Duration**	**BDD**	ΔLMP	**Profit**(per week)
Case 1	S1	$L_{\mathbf{rate},109}$	0.14	Bus 115	Week	Pass	0.19\$	127.1 \$/MWh
Case 2	S1	$L_{\mathbf{rate},109}$	0.2	Bus 115	Week	Pass	1.7\$	1146 \$/MWh
Case 3	S2	R_{181} X_{181}	2 1.5	Bus 128	Week	Pass	2.47\$	416.37 \$/MWh
Case 4	S3	$L_{\mathbf{breaker},109}$	-	Bus 115	Week	Pass	−3.24\$	-545.8\$/MWh
Case 5	S4	$G_{\mathbf{pmax},13}$ $G_{\mathbf{ramp},13}$	0.2	Bus 50	Week	Pass	3.18\$	534.5\$/MWh
Case 6	S5	P_{115}	1.2	Bus 115	Peak hours	Pass	0.93\$	630.95\$/MWh
Case 7	S1, S5	$P_{115}, L_{\mathbf{rate},109}$	1.2 0.2	Bus 115	Peak hours	Pass	1.74\$	293.2\$/MWh
Case 8	S6	$G_{\mathbf{pmax},15,16,19,20}$	0.02	Bus 56	Peak hours	Pass	1.15\$	193.5\$/MWh

Dataset Construction

- ***Overview of the SMLT dataset development framework***



Empirical Observations

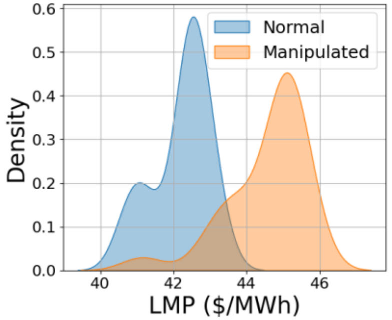
Knowledge & Data Engineering Systems

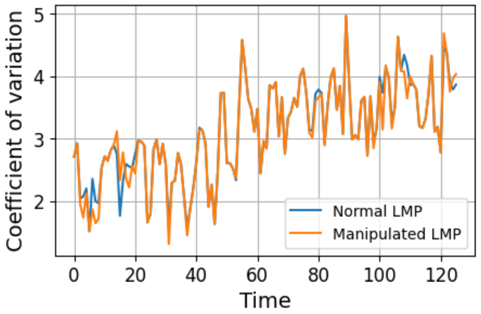
***RQ1**: What is the impact of stealthy FDIAs on the distribution of LMP data?*

Fixed window (one day)

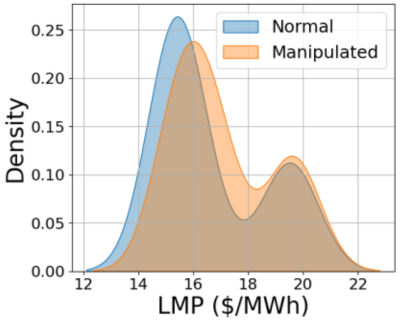
Consecutive time steps

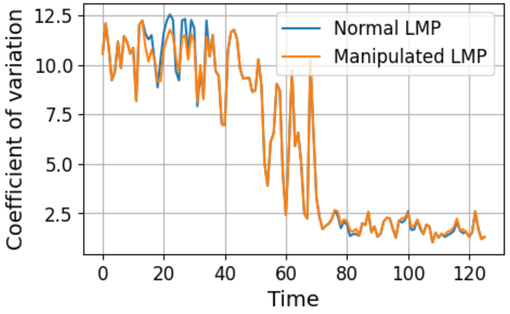
Case 3





Case 6





Statistical Result

Case	Visibility(%)		Detectability	
	Value	Qual.	Value	Qual.
Case 1	16.67%	−−	0.0005	-
Case 2	42.32%	+	0.0010	+
Case 3	33.36%	+	0.0003	−−
Case 4	84.71%	++	0.0014	+
Case 5	16.69%	-	0.0014	+
Case 6	9.73%	−−	0.0005	−−
Case 7	50.54%	++	0.0008	-
Case 8	18.86%	-	0.0135	++

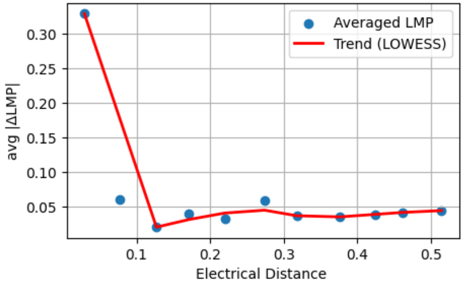
University of Glasgow

Empirical Observations

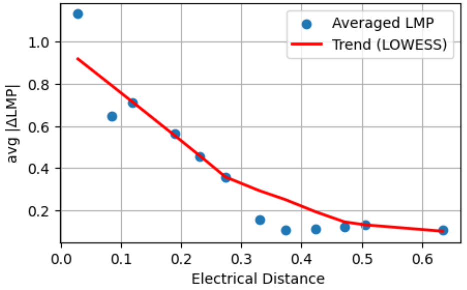
Knowledge & Data Engineering Systems

***RQ2:** How does the impact of an attack propagate throughout the system?*

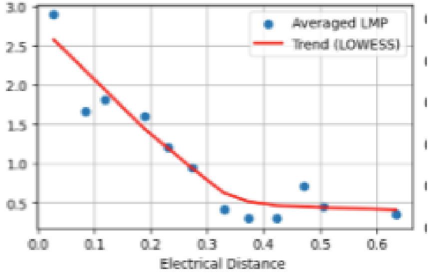
Case 3



Case 6



Case 7



Case	Visibility(%)		Detectability		Spreadability	
	Value	Qual.	Value	Qual.	Value	Qual.
Case 1	16.67%	−−	0.0005	-	0.1650	-
Case 2	42.32%	+	0.0010	+	0.3668	+
Case 3	33.36%	+	0.0003	−−	0.0267	−−
Case 4	84.71%	++	0.0014	+	0.2275	+
Case 5	16.69%	-	0.0014	+	0.9137	++
Case 6	9.73%	−−	0.0005	−−	0.2077	-
Case 7	50.54%	++	0.0008	-	0.5780	++
Case 8	18.86%	-	0.0135	++	0.0465	−−

Case Study

- ***Geometric Entropy Minimization based model*** *[8]*

Case	GEM before drift					GEM after drift				
	DR	PR	F1	AUC	FAR	DR	PR	F1	AUC	FAR
1	0.137	0.958	0.240	0.566	0.005	0.881	0.265	0.408	0.692	0.496
2	0.976	0.692	0.810	0.802	0.372	0.958	0.272	0.423	0.718	0.523
3	0.006	0.500	0.012	0.500	0.005	0.943	0.276	0.428	0.721	0.502
4	0.988	0.897	0.941	0.946	0.097	0.991	0.280	0.437	0.737	0.517
5	0.994	0.898	0.944	0.949	0.097	0.988	0.280	0.437	0.736	0.515
6	0.185	0.409	0.254	0.438	0.308	0.956	0.247	0.393	0.747	0.463
7	0.769	0.476	0.588	0.715	0.338	0.989	0.240	0.387	0.746	0.497
8	0.110	0.769	0.192	0.549	0.011	0.945	0.163	0.278	0.728	0.489

Lessons Learned

- ***Insights from empirical analysis & case study with the SMLT dataset***
  1. *Market-level data matters*
     - *LMP data is a strong signal for detecting FDIA attacks, even when BDD fails.*
  2. *FDIA impacts propagate*
     - *Localized attacks spread system-wide, affecting neighboring buses.*
  3. *Duration & timing are critical*
     - *Short, peak-hour attacks are harder to detect due to natural LMP volatility.*
  4. *Need for drift-aware models*
     - *LMPs are non-stationary; adaptive models must handle regime shifts & drift*

University of Glasgow

Knowledge & Data Engineering Systems

Thank you!

Ghadeer Alsharif

g.alsharif.1@research.gla.ac.uk

GitHub