A FRAMEWORK FOR ANALYZING THE IMPACT OF DATA INTEGRITY/QUALITY ON ELECTRICITY MARKET OPERATIONS

A Dissertation

by

DAE HYUN CHOI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Le Xie
Committee Members,	Chanan Singh
	Srinivas Shakkottai
	Natarajan Gautam
Head of Department,	Chanan Singh

May 2014

Major Subject: Electrical Engineering

ABSTRACT

This dissertation examines the impact of data integrity/quality in the supervisory control and data acquisition (SCADA) system on real-time locational marginal price (LMP) in electricity market operations. Measurement noise and/or manipulated sensor errors in a SCADA system may mislead system operators about real-time conditions in a power system, which, in turn, may impact the price signals in real-time power markets. This dissertation serves as a first attempt to analytically investigate the impact of bad/malicious data on electric power market operations. In future power system operations, which will probably involve many more sensors, the impact of sensor data integrity/quality on grid operations will become increasingly important.

The first part of this dissertation studies from a market participant's perspective a new class of malicious data attacks on state estimation, which subsequently influences the result of the newly emerging look-ahead dispatch models in the real-time power market. In comparison with prior work of cyber attack on static dispatch where no inter-temporal ramping constraint is considered, we propose a novel attack strategy, named ramp-induced data (RID) attack, with which the attacker can manipulate the limits of ramp constraints of generators in look-ahead dispatch. It is demonstrated that the proposed attack can lead to financial profits via malicious capacity withholding of selected generators, while being undetected by the existing bad data detection algorithm embedded in today's state estimation software.

In the second part, we investigate from a system operator's perspective the sensitivity of locational marginal price (LMP) with respect to data corruption-induced state estimation error in real-time power market. Two data corruption scenarios are

considered, in which corrupted continuous data (e.g., the power injection/flow and voltage magnitude) falsify power flow estimate whereas corrupted discrete data (e.g., the on/off status of a circuit breaker) do network topology estimate, thus leading to the distortion of LMP. We present an analytical framework to quantify real-time LMP sensitivity subject to continuous and discrete data corruption via state estimation. The proposed framework offers system operators an analytical tool to identify economically sensitive buses and transmission lines to data corruption as well as find sensors that impact LMP changes significantly.

This dissertation serves as a first step towards rigorous understanding of the fundamental coupling among cyber, physical and economical layers of operations in future smart grid.

DEDICATION

To my family

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest respect and gratitude to my advisor Professor Le Xie for his wonderful guidance and support over the years. Dr. Xie has truly been a friend and teacher for me during the past four years at Texas A&M. His valuable research skills, vast knowledge of power system and information technology, and deep insight of seeing the big picture will be extremely useful for my future career. Moreover, his great personality and kindness for always considering my family have given me a more comfortable environment to focus on my research. Without his guidance and persistent help throughout my Ph.D. process, I would have never completed this dissertation.

Next, I wish to thank my committee members, Dr. Chanan Singh, Dr. Srinivas Shakkottai and Dr. Natarajan Gautam, for the interest they showed in my defense, and for their suggestions and time. I would also like to thank Dr. Soummya Kar at Carnegie Mellon University for the fruitful collaboration and encouragement.

I am also grateful to current and past members of Dr. Xie's research group, my office mates and friends: Chen Yang, Anupam A. Thatte, Yingzhong Gu, Omar A. Urquidez, Yun Zhang, Yang Chen, Fan Zhang, James Carroll, Meichen Chen, and Haiwang Zhong. I thank them for the many discussions and for making my Ph.D. journey enjoyable. I also wish to acknowledge some of my Korean friends: Byunghak Kim, Jaewon Yoo, and Sangwoo Park for technical discussions and fun days at Texas A&M University.

Special thanks to my family for their endless love and support. I would like to thank my mother and parents-in-law for their constant support and encouragement, and my sister for taking care of my mother while I am studying at Texas A&M

University. I also wish to express my gratitude to my dear wife, Jeonghwa Moon, for her unconditional love and sharing. She always has encouraged me to research my project and has stayed with me in joy and sorrow. Her devotion was an impetus to completing my dissertation. My two sons, Jaden Wonjun Choi and Jason Wonwoo Choi, have become a source of happiness. Especially, I devote my work to my father who passed away in 2007.

TABLE OF CONTENTS

Page

ABSTRACT ii

DEDICATION iv

ACKNOWLEDGEMENTS v

TABLE OF CONTENTS vii

LIST OF FIGURES ix

LIST OF TABLES xii

1. INTRODUCTION 1

1.1 Motivation and Overview 1
1.2 Malicious Data Attack on Economic Dispatch 4
1.2.1 Prior Work 4
1.2.2 Main Contributions 5
1.3 Sensitivity Analysis of LMP to Data Corruption 6
1.3.1 Prior Work 6
1.3.2 Main Contributions 7
1.4 Dissertation Outline 9

2. BACKGROUND 11

2.1 Power System State Estimation 11
2.1.1 State Estimation Algorithms 11
2.1.2 Bad Data Processing 17
2.1.3 Topology Error Processing 20
2.2 Real-Time Power Market 24
2.2.1 Ex-Ante Market 25
2.2.2 Ex-Post Market 26

3. MALICIOUS TEMPORAL DATA ATTACK ON TIME-COUPLED LOOK-AHEAD DISPATCH 28

3.1 Introduction 28
3.2 Background 30
3.2.1 DC State Estimation Model 30
3.2.2 Look-Ahead Dispatch Model 32
3.3 Overview of Spatial Data Attack on Static Dispatch 35

3.3.1 Ex-Post LMP Formulation . . . 35
3.3.2 Attack Model and Undetectability . . . 36
3.3.3 Attack Procedure Using Virtual Bidding Mechanism . . . 37
3.3.4 Attack Strategy . . . 38
3.4 Statement of Temporal Data Attack Problem . . . 41
3.5 Formulation of the Ramp-Induced Data Attack . . . 45
3.5.1 Requirements and Procedure for a Successful RID Attack . . . 45
3.5.2 Proposed Attack Strategy . . . 46
3.5.3 Attack Performance Metrics . . . 52
3.6 Numerical Example . . . 54
4. LMP SENSITIVITY ANALYSIS TO DATA CORRUPTION-INDUCED ESTIMATION ERROR . . . 62
4.1 Introduction . . . 62
4.2 Preliminaries . . . 66
4.2.1 AC State Estimation Model . . . 66
4.2.2 Real-Time Electricity Pricing Model . . . 69
4.3 Impact Analysis of LMP Subject to Power Flow Estimate Errors . . . 71
4.3.1 Problem Formulation . . . 71
4.3.2 LMP Sensitivity to Continuous Sensor Data Corruption . . . 73
4.3.3 Numerical Example . . . 80
4.3.4 Discussions . . . 95
4.4 Impact Analysis of LMP Subject to Network Topology Estimate Errors . . . 96
4.4.1 Introduction . . . 96
4.4.2 Preliminaries . . . 98
4.4.3 Derivation of LMP Sensitivity to Network Topology Error . . . 101
4.4.4 Simulation Results . . . 108
5. CONCLUSIONS . . . 115
5.1 Malicious Data Attack on Look-ahead Dispatch . . . 115
5.1.1 Summary . . . 115
5.1.2 Future Work . . . 116
5.2 LMP Sensitivity Analysis to Data Corruption-Induced Estimation Error . . . 117
5.2.1 Summary . . . 117
5.2.2 Future Work . . . 119
REFERENCES . . . 121

LIST OF FIGURES

FIGURE Page

1.1 Information flow among cyber-physical components. 3

2.1 Contrast of communication architecture for hierarchical and fully distributed state estimation. 13

2.2 Branch status error and substation configuration error. 22

2.3 Electricity market operation. 25

3.1 A three-layered framework illustrating cyber data attack. 29

3.2 Statistical signal processing framework illustrating the relationship among sensor data, state estimation, and economic dispatch. 42

3.3 Conceptual diagrams illustrating a ramp-induced data attack. 44

3.4 Conceptual diagrams illustrating a binding unit attack. 48

3.5 IEEE 14-bus test system with three attack cases. 55

3.6 LMP of static and look-ahead dispatch without attack and with Case I,II and III attacks. 57

3.7 $P_{g_3}^{\max} - P_{g_3}^{*}$ of static and look-ahead dispatch without attack and with Case I,II and III attacks. 59

4.1 Illustrating the impact of corrupted continuous and discrete SCADA sensor data on state estimation and SCED. 63

4.2 A three-layered framework illustrating the coupling of the physical power system, state estimation, and SCED. 65

4.3 IEEE 14-bus system with a given measurement configuration. 82

4.4 Sensitivities of Ex-ante prices with respect to (a) real power injection measurements, (b) reactive power injection measurements, (c) real power flow measurements, (d) reactive power flow measurements, and (e) voltage magnitude measurements. Line 3-4 is congested and P_{g_3} is binding at $\hat{P}_{g_3}^{\min}$ in the IEEE 14-bus system. 86

4.5 Sensitivities of Ex-post prices with respect to (a) real power injection measurements, (b) reactive power injection measurements, (c) real power flow measurements, (d) reactive power flow measurements, and (e) voltage magnitude measurements. Line 6-12 is congested and the corresponding line flow is binding at the capacity limit of line 6-12 in the IEEE 14-bus system. 87

4.6 Comparison of sensitivities between the perturbation method ($\epsilon = 0.01$) and the proposed method with respect to (a) P_3 corruption, (c) Q_3 corruption, (e) $P_{5,6}$ corruption, and (g) $Q_{5,6}$ corruption in Fig. 4.6(a), (c), (e), (g). Comparison of price deviations between economic dispatch and the proposed method with respect to (b) P_3 corruption, (d) Q_3 corruption, (f) $P_{5,6}$ corruption, and (h) $Q_{5,6}$ corruption in Fig. 4.6(b), (d), (f), (h). 89

4.7 LMP differences between with and without corrupted data when z_8 is corrupted in Fig. 4.4(c). 90

4.8 LMP differences between with and without corrupted data in Fig. 4.4 (a) P_3, Q_3, and V_3 corruptions (b) $P_{5,6}$ and $Q_{5,6}$ corruptions. 91

4.9 Comparison of LMP sensitivities at bus 3 in Fig. 4.4(a) with varying variances of injection measurements P_3 and P_{11}. 91

4.10 IEEE 118-bus system. 92

4.11 Sensitivities of Ex-ante prices with respect to (a) real power injection measurements, (b) reactive power injection measurements, (c) real power flow measurements, (d) reactive power flow measurements, and (e) voltage magnitude measurements. Line 15-17 is congested and $P_{g_{19}}$ is binding at $\hat{P}_{g_{19}}^{\max}$ in the IEEE 118-bus system. 93

4.12 Illustration of a linear relationship between $\Delta\boldsymbol{\pi}_l^k$ and $\mathbf{v}_l^k$. 106

4.13 IEEE 14-bus system including bus-breaker model. 109

4.14 LMP results in Fig. 4.13: (a) comparison of LMPs between with and without line exclusion error; (b) comparison of LMP sensitivities obtained by SCED and the proposed approach. 110

4.15 Impact of a varying gap between the energy costs of marginal units on LMP sensitivity. 111

4.16 Comparison of LMP sensitivities with four different branch exclusion errors under the line 5-6 congestion. 112

4.17 Comparison of LMP sensitivities with four different congestion patterns under the line 4-5 exclusion. 114

LIST OF TABLES

TABLE Page

3.1 Notations. 32

3.2 Comparison between RID attack and spatial attack. 52

3.3 Generator parameters of the IEEE 14-bus test system. 56

3.4 Attack performance in static and look-ahead dispatch. 58

3.5 Impact of ramp rate and measurement variance on the attack performance in Case I. 58

3.6 Attack performance with varying attack magnitude in Case I. 60

4.1 Notations. 67

4.2 Generator parameters of the IEEE 14-bus system. 80

4.3 Nomenclature. 99

4.4 Generator parameters of the IEEE 14-bus system. 109

1. INTRODUCTION*

1.1 Motivation and Overview

Managing and understanding the electric grid data is an increasing concern for Independent System Operators (ISOs), utilities, Load Serving Entities (LSEs) and market participants. As more and more sensing devices such as smart meters and synchrophasors are deployed in the power system, it is prudent for future electric grid operators to understand the fundamental impact of data quality on resilient physical and market operations. Furthermore, data integrity is closely related to cybersecurity that may occur in the power industry. For instance, data manipulated by an intelligent adversary may result in misleading system operators and smart grid control systems, thus resulting in the severe malfunction of physical and financial grid operations. To meet the challenges in data quality and data integrity, we propose a framework to analyze and design robust cyber-physical electric energy systems with the presence of bad/malicious data. In this dissertation, we introduce and test a resilient operating paradigm against bad/malicious data for smart grid market operations.

Currently, the operations of the physical power systems and the wholesale level electricity markets are controlled by Regional Transmission Organizations (RTOs) (e.g., ISO-New England, PJM, ERCOT and CAISO). RTOs conduct state estimation and economic dispatch in the two representative software systems, which are Energy Management System (EMS) and Market Management System (MMS), respectively.

*This section is in part a reprint of the material in the papers: D.-H. Choi and L. Xie, "Ramp-Induced Data Attacks on Look-Ahead Dispatch in Real-Time Power Markets", *IEEE Transactions on Smart Grid* vol. 4, no. 3, pp. 1235-1243, September 2013 and D.-H. Choi and L. Xie, "Sensitivity Analysis of Real-Time Locational Marginal Price to SCADA Sensor Data Corruption", *IEEE Transactions on Power Systems* with DOI: 10.1109/TPWRS.2013.2293634.

Given the stronger coupling among EMS/MMS, heterogeneous cyber devices (e.g., PMUs, smart meters), and communication networks (e.g., NASPInet [1]), and physical power systems, the future smart grid must cope with a variety of data integrity attacks in this cyber-physical system. Furthermore, future power system operations will probably involve data acquisition and processing by a large number of sensors. Therefore, maintaining sensor data quality and data integrity will become increasingly important for reliable and economical grid operations.

The main goal of this dissertation is to investigate the impact of bad data and malicious data attacks on real-time electricity market operations via state estimation. A key part of this research is to understand the coupling among the following three components: (1) bad/malicious data; (2) state estimation; and (3) economic dispatch. In this dissertation, data represent measurements collected by massively deployed sensors in supervisory control and data acquisition (SCADA) systems. Bad data refer to sensors' measurements that have large errors beyond typical confidence interval of random Gaussian noise. Bad data usually occur due to meter malfunction, telecommunication system failure, and unexpected communication noise. In comparison with naturally generated bad data, malicious data are designed and injected into normal sensor data by an intelligent adversary that is capable of optimizing the effectiveness and the stealth nature of the attack. State estimation and economic dispatch are core functions in EMS and MMS, respectively. State estimation converts redundant SCADA sensor measurements and other available information into an estimate of the state of an interconnected electric power system.Economic dispatch generates the optimal generation dispatch and nodal price, otherwise known as locational marginal price (LMP). In real-time electricity markets, security constrained economic dispatch (SCED) is conducted to calculate LMP in the two main real-time pricing models: the ex-ante and ex-post models [2, 3, 4].

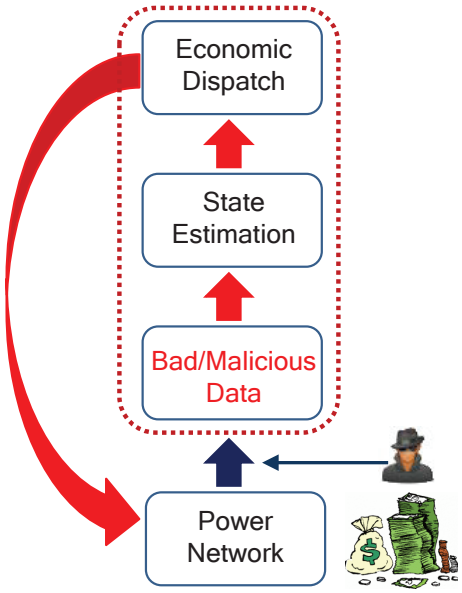


Figure 1.1: Information flow among cyber-physical components.

We first consider from *a market participant's perspective* a potential attack problem in which the attacker stealthily manipulates an estimate of the system state by injecting false data into the sensor's measurements. This manipulated estimate is fed into economic dispatch module in MMS and subsequently changes operational constraints embedded in economic dispatch formulation. As a result, the attacker generates the wrong optimal generation outputs and changes the nodal prices in the desired direction. Second, we examine from *a system operator' perspective* the impact of bad/malicious data on the nodal price. The development of an analytical framework to quantify such impact is necessary because it enables system operators to identify economically sensitive buses with respect to data corruption as well as prioritize sensor data quality upgrades in view of robust market operations. Fig. 1.1 shows the information flow among the physical power system and aforementioned three components. It should be noted from this figure that bad and malicious data have an adverse chain-effect on economic dispatch process via state estimation.

1.2 Malicious Data Attack on Economic Dispatch

1.2.1 Prior Work

A large body of literature has been accumulated recently on the subject of cyber security in power grids, ranging from risk mitigation [5], generation control security (e.g., automatic generation control (AGC) attack [6, 7]), control security in distribution system [8], and privacy protection [9, 10, 11, 12]. A concise summary paper is presented in [13], including risk assessment methodology, power system control application and cyber infrastructure security. Meanwhile, many researchers have been studying false data injection attacks, which malfunction the state estimator by injecting false data into sensors. For the subject of false data injection attacks, two major categories of work have been presented:

- *Vulnerability analysis of state estimation:* a false data injection attack against DC state estimation was formulated and analyzed in [14, 15]. Efficient algorithm to find sparse attacks and Phasor Measurement Units (PMUs) placement algorithm to prevent sparse attacks were developed in [16, 17]. Various attack strategies against distributed state estimation were proposed and their impacts were quantified [18]. As work following [18], a fully distributed attack detection scheme based on local measurements was proposed to detect such attacks. A distributed joint detection-estimation approach to malicious data attack was presented in [19]. In [20], it was shown that the attacker can hack the power grid without the knowledge of the power network topology, which can be estimated using linear independent component analysis (ICA). A false data injection attack against AC state estimation was proposed in [21] and its effect on AC state estimation performance was investigated in more detail [22]. Graph theory based algorithm was proposed, which determines the type and number of

measurements compromised by the attacker to bypass bad data detection.

- *Financial risk analysis in electricity market operations:* this area examined the economic impact of false data injection attacks on electricity market operations. In [23], a conceptual four-layer framework consisting in the physical, communication/control, market layer, and a cyber security layer was proposed and the impact of cyber attack on marker layer was assessed. Undetectable and profitable attack strategies, which exploit virtual bidding mechanism, were proposed in [24]. In [25], a more general malicious data attack problem was formulated in the real-time electricity market. In [26], the scenario for the attacker and defender was modeled as a zero-sum game between them, and simulation results showed the effectiveness of attack on the real-time market prices. Metrics were developed to evaluate the economic impact of data integrity attacks under economic dispatch process using optimal power flow methods in [27].

However, the proposed attacks were characterized in static economic dispatch without modeling inter-temporal constraints.

1.2.2 Main Contributions

In Section 3, we present a new class of false data injection attacks on state estimation, which may lead to financial arbitrage in real-time power markets with an emerging *look-ahead* dispatch model.

Motivated by the increasing penetration of variable resources such as wind and solar [28], look-ahead dispatch has been implemented by major Independent System Operators (ISOs)/Regional Transmission Organizations (RTOs) in the past few years in order to improve the market dispatch efficiency [29, 30, 31]. Look-ahead dispatch is different from conventional static dispatch in that it calculates the optimal dispatch in an extended period of time, taking into account inter-temporal ramp

rates of generators of different technologies. In recent work [24, 25], the proposed attacks against state estimation were characterized in *static* economic dispatch without modeling inter-temporal ramping constraint.

In comparison with this prior work of cyber attack on static dispatch, we propose a novel attack strategy with which the attacker can manipulate, in look-ahead dispatch, the limits of ramp constraints of generators. The main contributions of this section are two-fold:

- We formulate a malicious ramp-induced data (RID) attack problem in look-ahead dispatch. The attacker could stealthily change the ramp constraint limits of generators through manipulating sensors data, aiming at increasing the nodal price by withholding capacity of generator.

- We propose a RID attack strategy with which the attacker could make a profit without being detected by RTOs in the real-time electricity market. The attack strategy for undetectability and profitability is formulated as an optimization problem in which the attacker computes the attack vector injected into sensors. The feasibility of such cyber attacks and their economic impact on real-time electricity market operations are illustrated in the IEEE 14-bus system.

1.3 Sensitivity Analysis of LMP to Data Corruption

1.3.1 Prior Work

Real-time market LMPs are primarily affected by a system's physical conditions, which are the results of state estimation routine. A study of LMP sensitivity with respect to system physical conditions was first conducted by Conejo et al. [32]. In this work, the LMP sensitivity problem was formulated in nonlinear programming based on the AC optimal power flow (ACOPF) model. It provided a generalized

platform for calculating the sensitivity of LMP with respect to changes in various parameters such as load, generator cost, voltage limit, generation power limit, and network topology. Sensitivity studies have also been conducted with linear programming based on the DC optimal power flow (DCOPF) model with a DCOPF-based algorithm [33], the probabilistic model [34], and the continuous locational marginal pricing approach [35]. All previous work has focused mainly on the impact of physical load variations on LMP sensitivity. More recently, some work has studied the economic impact of cyber data attacks on real-time power market operations. This recent work has demonstrated that *continuous* data corruption from an adversary can bypass the Chi-squares bad data detection [14, 36], consequently leading to LMP distortion due to state estimation error [24, 25]. In addition, it has been recently demonstrated that malicious topology data attacks due to *discrete* data corruption [37, 38] could completely bypass topology error processing in state estimation software. As a result, they may generate an undetectable misconfigured network topology, which, in turn, results in misleading state estimation and economic dispatch. However, no analytic study for quantifying the impact of undetectable continuous and discrete data corruption on LMP sensitivity has been done yet.

1.3.2 Main Contributions

Section 4 focuses on the development of an analytical framework for answering the following question:

- How much does LMP change at every bus given a set of SCADA measurements with corrupted data?

Specifically, we investigate the sensitivity of real-time LMP with respect to state estimation error due to continuous (e.g., the power injection/flow and voltage magnitude) and discrete (e.g., the on/off status of a circuit breaker) data corruption.

In the first part of Section 4, we present an analytical framework to quantify LMP sensitivity with respect to changes in *continuous* sensor data. This framework consists of a unified LMP sensitivity matrix subject to sensor data corruption, which combines two sensitivity matrices: the first with LMP sensitivity at any bus to any estimate, and the second with sensitivity of any estimate to data at any sensor. Thus, this unified sensitivity matrix reflects a coupling among the sensor data, an estimation of the power system states, and the real-time LMP. The proposed framework offers system operators an *online* analytical tool to:

1. assess the impact of corrupted data at any sensor on LMP variation at any bus
2. identify buses with LMPs highly sensitive to data corruption
3. find sensors that impact LMP changes significantly and influentially
4. evaluate the impact of SCADA data accuracy on real-time LMP.

The results of the proposed sensitivity based analysis are illustrated and verified with IEEE 14-bus and 118-bus systems with both Ex-ante and Ex-post real-time pricing models.

In the second part of Section 4, we formulate and analyze the impact of power transmission network topology error due to *discrete* data corruption on real-time electricity market prices. We consider the scenario in which the undetected false status of circuit breakers from topology error processing may lead to wrong modeling of real-time network topology, which, in turn, misleads the results of state estimation and real-time economic dispatch. In particular, we focus on the economic impact of this circuit breaker-induced network topology error on LMP. The main contributions of this second part is twofold:

- We propose a simple analytical LMP sensitivity index that explains the relationship between the change in network topology (i.e., the left-hand side of constraints in SCED) and LMP.

- The proposed analytical approach is illustrated and verified in the IEEE 14-bus system. The proposed sensitivity index provides system operators an analytical tool to identify economically sensitive transmission lines and circuit breakers, whose status error will significantly impact the real-time LMPs. The validity of the derived sensitivity index is verified and illustrated with numerical examples in the IEEE-14 bus system.

1.4 Dissertation Outline

The rest of this dissertation is organized as follows. Section 2 presents an overview of power system state estimation and real-time power market. The weighted lease squares (WLS) formulation for power system state estimation is introduced along with bad data processing and topology error processing techniques. Distributed state estimation and bad data processing methods are briefly reviewed. In real-time power market, two major real-time pricing models are considered: ex-ante and ex-post models. Security constrained economic dispatch (SCED) model is formulated in each pricing model, which relies on state estimation results.

In Section 3, malicious data attack against state estimation in time-coupled look-ahead dispatch is considered from a market participant's perspective. We first review spatial data attack on static dispatch through virtual bidding transaction in day-ahead and real-time markets. In comparison with the existing spatial data attack, we propose a temporal data attack strategy with which the attacker can manipulate, in look-ahead dispatch, the limits of ramp constraints of selected generators through malicious capacity withholding and make a profit while being undetected by the bad

data detection algorithm.

In Section 4, the sensitivity of real-time LMP to sensor data corruption is studied from a system operator's perspective. We develop an analytical tool to assess the ex-ante and ex-post LMP sensitivity subject to state estimation error due to the corruption in two types of sensor data: continuous (e.g., the power injection/flow and voltage magnitude) and discrete (e.g., the on/off status of a circuit breaker). The developed tool enables system operators to identify economically vulnerable buses, transmission lines and circuit breakers to bad data and malicious data attack as well as find sensors that impact LMP sensitivity significantly.

In Section 5, we summarize our main contributions in this dissertation and provide some future research directions.

2. BACKGROUND*

2.1 Power System State Estimation

2.1.1 State Estimation Algorithms

State estimation is one of the key functions in control centers' energy management systems (EMSs). A state estimator converts redundant meter readings and other available information obtained from a supervisory control and data acquisition (SCADA) system into the estimate of the state of an interconnected power system [39] and distribution system [40]. The computed estimation solution can be utilized by applications in EMS such as optimal power flow, contingency analysis, automatic generation control (AGC) and security constrained economic dispatch (SCED).

The pioneering work on state estimation was done by Schweppe et al. [41]. A centralized state estimator embedded in a control center was developed to monitor static operating conditions of a power system, based on all measurements collected by deployed sensors in a SCADA system. The practical implementation of state estimation was considered in [42]. A large number of technical literature related to centralized state estimation can be found in [43]. Recently, the emergence of large power systems with increased complexity suggests the need for decentralized estimation and control in wide area power system operations in order to reduce the computational burden at a central control center. Several approaches to decentralized state estimation have been proposed in the literature (see [44, 45], for example, for a treatment of decentralized iterative algorithms for system analysis and optimization).

*This section is in part a reprint of the material in the papers: L. Xie, D.-H. Choi, S. Kar, and H. Vincent Poor, "Fully Distributed State Estimation for Wide-Area Monitoring Systems", *IEEE Transactions on Smart Grid* vol. 3, no. 3, pp. 1154-1169, September 2012 and D.-H. Choi and L. Xie, "Fully Distributed Bad Data Processing for Wide Area State Estimation", *2011 Second International Conference on Smart Grid Communications*, October 2011.

In [46] and [47] a *star-like* hierarchical state estimation method was proposed. More recently the two-level state estimation for multi-area power system has been studied in [48, 49, 50, 51], driven by the capability and need to conduct WAMS. The local state estimation obtained at the first level is coordinated at a higher level via synchronized phasor measurements. A survey on multi-area state estimation is summarized in [52]. Most recently a multilevel state estimator (feeder, substation, transmission system organization, and regional levels) is described for the purpose of monitoring large-scale interconnected power systems [53]. However, as the measurement number and sampling rate increase, hierarchical state estimation approaches may suffer from communication bottlenecks and computational reliability issues inherent in system architecture with one single coordination center. A parallel and distributed state estimation was envisioned in [54]. By leveraging the naturally decoupled characteristic of weighted least squares (WLS) estimation, state estimation problem is decomposed into each area's local estimator with a coupling constraints optimization technique to ensure convergence of the boundary buses' estimates. Numerical results illustrate that the distributed algorithm could not only speed up the computational time, but also yields acceptable accuracy. However, local observability of each control area is always required in the aforementioned algorithms. In other words, all the local control areas need to have enough measurement redundancy in order to compute the locally decoupled weighted least squares estimate (excluding the boundary bus measurements). This assumption may not always hold due to (1) the increasing vulnerability of measurements subject to potential bad/malicious data, and (2) the emergence of smaller control areas such as micro-grids. In recent work [55], a fully distributed static state estimation algorithm with relaxed local observability is exploited. An iterative distributed state estimation scheme is proposed, under which the local control areas begin with their own estimates of the entire system, commu-

nicate their estimates with pre-specified neighboring control areas, and eventually make all local estimates converge to the centralized state estimation result. Fig. 2.1 contrasts the communication architecture of fully distributed state estimation with that of hierarchical state estimation in a multi-area power system.

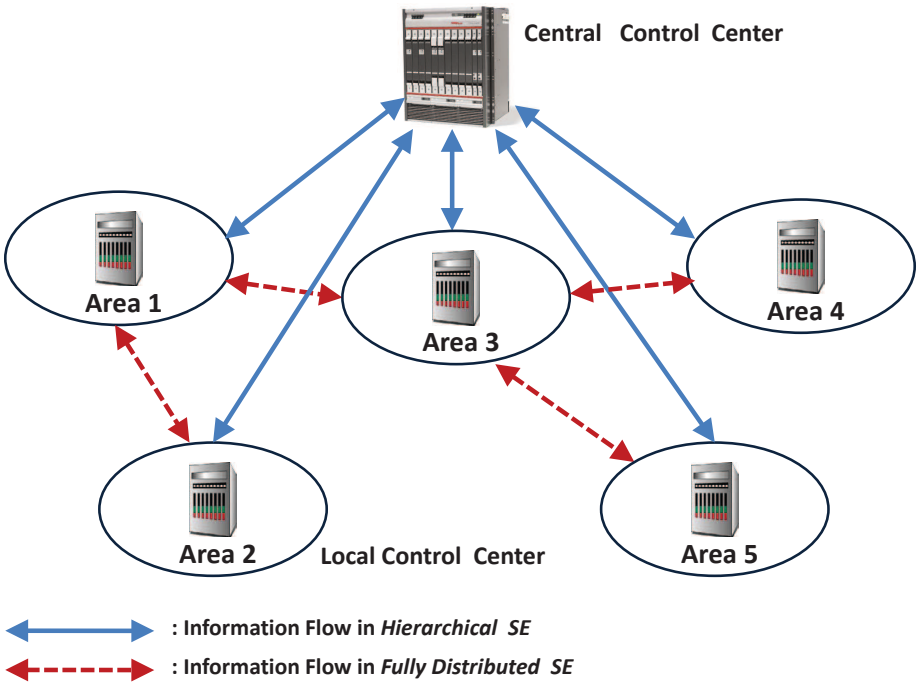


Figure 2.1: Contrast of communication architecture for hierarchical and fully distributed state estimation.

The weighted least squares (WLS) method is the most commonly applied one for estimating the operating conditions of a power system.The measurement model of the state estimation is formulated as follows:

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{e} \tag{2.1}$$

where $\mathbf{z}$ is measurement vector in the form of bus power injections and line power flows, bus voltage magnitudes and line current flow magnitudes, $\mathbf{x}$ is the state vector of the entire power system with phase angles and bus voltage magnitudes, $\mathbf{h}(\mathbf{x})$ is nonlinear measurement function relating measurements to states, and $\mathbf{e}$ is independent identically distributed (i.i.d.) Gaussian measurement error vector with zero mean and diagonal covariance matrix $\mathbf{R}$. The WLS will minimize the weighted least squares of measurement error to compute the optimal estimate of $\mathbf{x}$:

$$\text{minimize } J(\mathbf{x}) = \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} \tag{2.2}$$

$$\text{s.t. } \mathbf{r} = \mathbf{z} - \mathbf{h}(\mathbf{x}). \tag{2.3}$$

Using the Gauss-Newton method, the weighted least squares estimate vector $\hat{\mathbf{x}}$ is computed by the following iterative procedure:

$$\Delta \hat{\mathbf{x}}^{k+1} = [\mathbf{G}(\hat{\mathbf{x}}^k)]^{-1} \mathbf{H}^T(\hat{\mathbf{x}}^k) \mathbf{R}^{-1} \Delta \mathbf{z}^k \tag{2.4}$$

where $\mathbf{H}(\hat{\mathbf{x}}^k) = \left[\frac{\partial \mathbf{h}(\hat{\mathbf{x}}^k)}{\partial \hat{\mathbf{x}}^k}\right]$ is the Jacobian matrix at k-th iteration, and

$$\Delta \hat{\mathbf{x}}^{k+1} = \hat{\mathbf{x}}^{k+1} - \hat{\mathbf{x}}^k \tag{2.5}$$

$$\Delta \mathbf{z}^k = \mathbf{z} - \mathbf{h}(\hat{\mathbf{x}}^k) \tag{2.6}$$

$$\mathbf{G}(\hat{\mathbf{x}}^k) = \mathbf{H}^T(\hat{\mathbf{x}}^k) \mathbf{R}^{-1} \mathbf{H}(\hat{\mathbf{x}}^k). \tag{2.7}$$

The iteration process in (2.4) continues until the maximum of $|\Delta \hat{\mathbf{x}}^k|$ is less than a predetermined threshold, otherwise stops and yields the ultimate estimates. Here, $\mathbf{G}(\hat{\mathbf{x}}^k)$ is defined as gain matrix. If this gain matrix is full rank, the network is called as observable network, and an unique WLS estimation solution is computed finally.

However, the aforementioned AC state estimation method often faces the divergence of the estimation solution. In addition, real-time market operation such as security constrained economic dispatch (SCED) is performed based on a DC based power flow solution. Moreover, as phasor measurement units (PMUs) become more and more in future smart grid, the measurement functions of PMUs become linear. Hence, DC state estimation algorithm is applicable for state estimation with PMUs. With these reasons, DC state estimation is often preferred to AC state estimation.

We present the linearized DC state estimation problem with 1.0 per unit (p.u.) voltage magnitudes at all buses and j1.0 p.u. branch impedance. Then, the state vector $\mathbf{x}$ is considered as the voltage phase angle vector $\boldsymbol{\theta}$ for the entire power system. Therefore, the nonlinear measurement model for AC state estimation (2.1) is modified to

$$\overline{\mathbf{z}} = \overline{\mathbf{H}}\boldsymbol{\theta} + \mathbf{e} \tag{2.8}$$

where the measurement vector $\overline{\mathbf{z}}$ includes only power injections and line flows, and the elements in the linearized matrix $\overline{\mathbf{H}}$ rely on line susceptance, network topology and measurement configuration. Centralized state estimation computes the optimal estimate of $\boldsymbol{\theta}$ by minimizing the weighted least squares of measurement error:

$$\text{minimize } J(\boldsymbol{\theta}) = \mathbf{r}^T\mathbf{R}^{-1}\mathbf{r} \tag{2.9}$$

$$\text{s.t. } \mathbf{r} = \overline{\mathbf{z}} - \overline{\mathbf{H}}\boldsymbol{\theta}. \tag{2.10}$$

Then, if the system is observable (i.e., the linearized Jacobian matrix $\overline{\mathbf{H}}$ is full rank), the centralized weighted least squares estimate of $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\theta}_c = \left(\overline{\mathbf{H}}^T\mathbf{R}^{-1}\overline{\mathbf{H}}\right)^{-1}\overline{\mathbf{H}}^T\mathbf{R}^{-1}\overline{\mathbf{z}} = \overline{\mathbf{G}}^{-1}\overline{\mathbf{H}}^T\mathbf{R}^{-1}\overline{\mathbf{z}}. \tag{2.11}$$

Recently, a fully distributed DC state estimation algorithm in a multi-area power system is proposed, and the convergence of the proposed algorithm to a centralized estimation is proven mathematically in [55]. An interconnected multi-area power system is assumed to be partitioned into a total of N regions, each region n corresponding to a geographically non-overlapping control area. Each control area is allowed, if necessary, to exchange information with its neighboring areas. The measurement model for the multi-area state estimation is formulated as follows:

$$\overline{\mathbf{z}}_n = \overline{\mathbf{H}}_n \mathbf{x} + \mathbf{e}_n \tag{2.12}$$

where $\overline{\mathbf{z}}_n$ is measurement vector (including the boundary injection and flow measurements) in control area n, $\mathbf{x}$ is the state vector of the entire interconnected power system, $\overline{\mathbf{H}}_n$ is linear measurement matrix for control area n, and $\mathbf{e}_n$ is measurement error vector with zero mean in area n. Then, we assume that the vector of initial estimate of the states, $\mathbf{x}_n(0) \in \mathbb{R}^M$, is deterministic where M is a total number of buses. A sequence of estimate vectors, $\{\mathbf{x}_n(i)\}_{i\geq 0}$ is computed by each control area in a distributed iterative manner. The state estimate vector $\mathbf{x}_n(i+1)$ of the n-th control area at $(i+1)$-th iteration is a function of: its previous estimate vector; the communicated estimate vectors at i-th iteration from its neighboring control areas; and the local measurement vector $\mathbf{z}_n$. Based on the current state vector $\mathbf{x}_n(i)$, the exchanged data $\{\mathbf{x}_l(i)\}_{l\in\Omega_n}$, and the measurement vector $\overline{\mathbf{z}}_n$, we update the estimate of the states at the n-th control area by the following distributed iterative algorithm:

$$\mathbf{x}_n(i+1) = \mathbf{x}_n(i) - \left[\beta(i)\sum_{l\in\Omega_n}(\mathbf{x}_n(i) - \mathbf{x}_l(i)) - \alpha(i)\overline{\mathbf{H}}_n^T\left(\overline{\mathbf{z}}_n - \overline{\mathbf{H}}_n\mathbf{x}_n(i)\right)\right]. \tag{2.13}$$

In (2.13), Ω_n represents the neighborhood of n, $\{\alpha(i)\}, \{\beta(i)\}$ are appropriately

chosen time-varying weight sequences with $\alpha(i) = \frac{a}{(i+1)^{\tau_1}}$, $\beta(i) = \frac{b}{(i+1)^{\tau_2}}$. (Here, $a, b > 0$ are constants and the exponents τ_1, τ_2 satisfy $0 < \tau_1 \leq 1$, $0 \leq \tau_2 < \tau_1$). Algorithm (2.13) is *distributed* because for n-th control area it involves only the data from the sensors in its neighborhood Ω_n. Then, for each n, the estimate sequence $\{\mathbf{x}_n(i)\}$ converges a.s. (with probability one) to the centralized least squares estimator $\boldsymbol{\theta}_c$

2.1.2 Bad Data Processing

In power system state estimation, bad data refers to measurements that have large errors beyond typical confidence interval of random Gaussian noise. As an important component of state estimation, bad data processing consists of two procedures: detection and identification [43]. Bad data detection determines whether the measurement set contains bad data. Then, bad data identification is subsequently performed to find which measurements contain bad data. Representative methods for bad data detection and identification are the Chi-squares test and the Largest Normalized Residual Test. They are briefly explained as follows:

2.1.2.1 Chi-squares Test

Consider the estimation objective function

$$J(\widehat{\mathbf{x}}) = \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} \tag{2.14}$$

where $\mathbf{r} = \mathbf{z} - \mathbf{h}(\widehat{\mathbf{x}})$ is defined as the estimated residual vector. Since the measurement errors are normally distributed, the estimated objective function $J(\widehat{\mathbf{x}})$ obeys a chi-square distribution with m-n degrees of freedom, i.e., $J(\widehat{\mathbf{x}}) \sim \chi^2_{m-n}$. m and n represent the number of measurements and state variables, respectively. Therefore, bad data

will be detected if

$$J(\widehat{\mathbf{x}}) \geq \chi^2_{(m-n),p} \tag{2.15}$$

where p is the detection confidence probability.

2.1.2.2 The Largest Normalized Residual Test

We consider the linearized measurement model. Normalized residuals are used for identifying bad data after state estimation is performed. The measurement residual vector $\mathbf{r}$ could be represented as

$$\mathbf{r} = \mathbf{z} - \mathbf{H}\boldsymbol{\theta} = \mathbf{S}\mathbf{e} \tag{2.16}$$

where the residual sensitivity matrix $\mathbf{S}$ represents the relationship between the measurement residuals and the measurement errors:

$$\mathbf{S} = \mathbf{I} - \mathbf{H}\mathbf{G}\mathbf{H}^T\mathbf{R}^{-1} \tag{2.17}$$

where we define the gain matrix $\mathbf{G} = \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$. Therefore, normalized residual vector can be represented as

$$\mathbf{r^N} = \frac{|\mathbf{r}|}{\sqrt{\text{diag}(\mathbf{SR})}}. \tag{2.18}$$

If the measurement corresponding to the largest normalized residual is greater than a chosen identification threshold, that measurement is considered as bad data and is eliminated for another round of state estimation.

Conventionally bad data processing, along with other functions of state estimation, is performed at one control center. More recently the possibility of decentralized bad data processing has been exploited for the purpose of WAMPAC. A two-level

distributed bad data processing is discussed in [48, 56, 57]. The local level and coordination level are in charge of filtering local and boundary bad data, respectively. The first attempt for fully distributed bad data processing is presented in [58], which introduces the concept of error residual spread area. Measurement error spreads only to measurement residuals within error residual spread area. In other words, no measurement error in one error residual spread area contaminates measurement residuals in the other error residual spread areas. In this error residual spread area setting, a reduced model for distributed bad data processing was proposed in [59]. Independent bad data processing is performed for each error residual spread area by the existing bad data detection and identification techniques. A fully distributed bad data detection and identification scheme with distributed state estimator was proposed based on error residual area decomposition [60]. A heuristic measurement design algorithm for information exchange between local areas was proposed in [61]. By exchanging local measurements or estimates with neighboring local areas, each local area could achieve improved performance of bad data processing.

However, the aforementioned work of distributed bad data processing is primarily based on error residual spread decomposition. For a large-scale power system, the decomposition of administrative areas are typically not overlapped with the error residual spread decomposition. The research challenge is to design the information exchange scheme among administrative control areas so that through communication, a distributed bad data detection and identification scheme becomes possible. In [62], fully distributed bad data processing is formulated. This formulation is tailored for the natural decomposition of an interconnected power system based on administrative boundaries (e.g., New England and New York). In addition, an information exchange scheme is proposed for detecting and identifying bad data in a distributed manner, which requires minimum communication among the control areas. For an

error residual spread area m, the local measurement Jacobian matrix is denoted by $\mathbf{H}_m$. It is proven in [59] that the residual sensitivity matrix $\mathbf{S}$ ordered based on error residual spread areas exhibits block diagonal structure. Namely, the measurement error in error residual spread area m does not contaminate any measurement outside of area m. Moreover, the block matrix $\mathbf{S}_m$ is expressed as

$$\mathbf{S}_m = \mathbf{I}_m - \mathbf{H}_m[\mathbf{G}_m]^{-1}\mathbf{H}_m^T\mathbf{R}_m^{-1} \tag{2.19}$$

Using (2.19), normalized residual vector in mth error residual spread area is defined as

$$\mathbf{r_m^N} = \frac{|\mathbf{r_m}|}{\sqrt{\text{diag}(\mathbf{S}_m\mathbf{R}_m)}} \tag{2.20}$$

where $\mathbf{r_m^N}$ is the normalized residual vector at mth error residual spread area. $[\mathbf{G}_m]^{-1}$ is the inverse of the sub-block gain matrix corresponding to the mth error residual spread area. Note that the computation of $\mathbf{G}_m$ requires the knowledge from only mth error residual spread area. If the error residual spread area decomposition coincide with the administrative area decomposition, then each control area only needs to conduct an independent bad data processing without considering the impact of bad data from other areas. However, in most real-world power systems, the two decomposition does not overlap with each other. Therefore, data communication would be required among the administrative areas to conduct satisfactory bad data processing. A simple information exchange scheme is proposed for detection and identification of bad measurement data in [62].

2.1.3 Topology Error Processing

In comparison with bad data processing taking into account the detection and identification of errors in continuous data such as the power injection/flow and volt-

age/current magnitude, topology error processing filters the erroneous discrete data such as the on/off status of a circuit breaker (CB).

Topology processor transforms a bus/section detailed model into a bus/branch model using the status data of CB collected through SCADA system. The constructed bus/branch model is then utilized by major EMS applications such as state estimator, contingency analysis, optimal power flow and security constrained economic dispatch. However, due to the malfunction of CB, unreported CB maintenance and human operation error the incorrect status data of CB is often fed into topology processor, thus generating the wrong network topology. Topology error can be categorized into the following two types of errors [43]:

- *Line status error:* the wrongly reported CB status data related to a transmission line generate two types of line status errors: line inclusion and exclusion error. They correspond to the incorrect line inclusion and exclusion, respectively.

- *Substation configuration error:* the wrongly reported CB status data in a substation generate two types of substation configuration errors: bus split and merging error. They correspond to the incorrect bus split and merging due to the incorrect link between bus sections, respectively.

Fig. 2.2 illustrates these two types of topology errors in a two-bus system.

The problem for detecting topological errors in power systems was first addressed in [63]. The method for checking topology status at substation level was proposed in [64], based on linear programming with the variables of power flow estimate through CBs. In [65], it was shown that line topology errors can be detected and identified using a geometrically based test. In [66], the largest normalized residual test was employed to handle bus split error as well as line errors. A correlation index

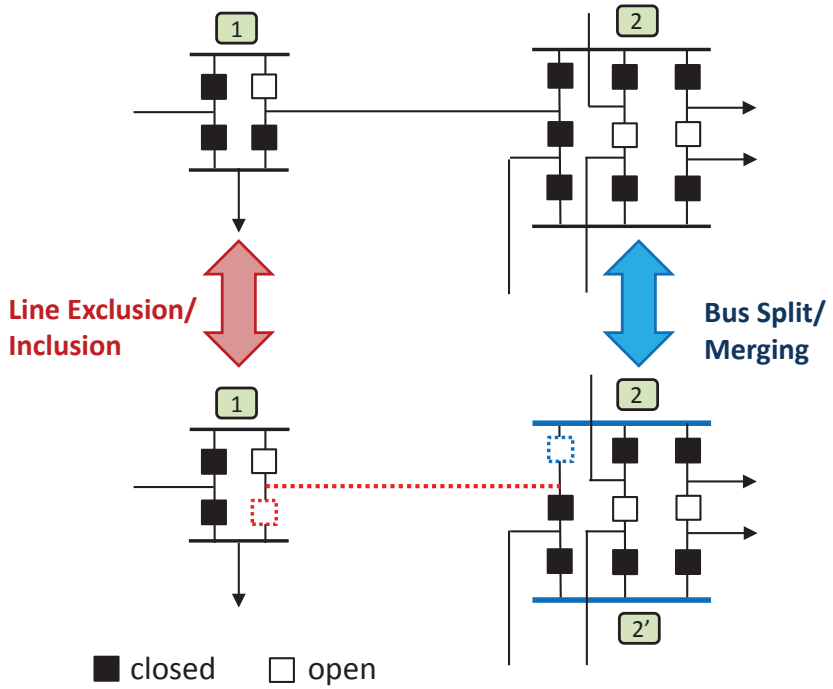


Figure 2.2: Branch status error and substation configuration error.

that explains the relationship between topology error and suspected digital measurement set was proposed in [67]. In [68], a modified two stage least absolute value (LAV) state estimator was developed, identifying substation configuration and line errors at the first and second stage, respectively. Computation-effective schemes for topology error processing were developed based on reduced model [69] and implicit model [70] to reduce the size of substation model for the least possible cost.

Topology error processing for line status error is conducted using measurement residual analysis similar to bad data processing. We consider a linearized DC power flow measurement model:

$$\mathbf{z} = \mathbf{J}\mathbf{x} + \mathbf{e} \tag{2.21}$$

where $\mathbf{x}$ is the state vector of the entire power system, $\mathbf{z}$ is measurement vector, $\mathbf{e}$

is independent identically distributed (i.i.d.) Gaussian random measurement error vector with zero mean and covariance matrix $\mathbf{R}$, and $\mathbf{J}$ is the *true* system Jacobian matrix of the state vector $\mathbf{x}$. For line status error, the measurement equation (2.21) is rewritten as

$$\mathbf{z} = \mathbf{J_E x} + \mathbf{E x} + \mathbf{e}$$

where $\mathbf{J} = \mathbf{J_E} + \mathbf{E}$. Here, $\mathbf{J_E}$ is the incorrect system Jacobian matrix due to topology errors. $\mathbf{E}$ is the system Jacobian error matrix. The measurement residual vector is expressed as

$$\mathbf{r} = \mathbf{z} - \mathbf{J_E}\hat{\mathbf{x}} = (\mathbf{I} - \mathbf{K_E})\mathbf{M f} \tag{2.22}$$

and the statistical properties of $\mathbf{r}$ are described in terms of

$$E(\mathbf{r}) = (\mathbf{I} - \mathbf{K_E})\mathbf{E x} \tag{2.23}$$

$$Cov(\mathbf{r}) = (\mathbf{I} - \mathbf{K_E})\mathbf{R} \tag{2.24}$$

Then, topology error detection is performed using the normalized residual vector as follows:

$$E(\mathbf{r}^N) = \Omega^{-\frac{1}{2}}(\mathbf{I} - \mathbf{K_E})\mathbf{M f} = \mathbf{S f} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{2.25}$$

where $\mathbf{M}$ is the measurement-to-branch incidence matrix, $\mathbf{f}$ is a vector of branch flow errors, $\mathbf{K_E} = \mathbf{J_E}(\mathbf{J_E}^T\mathbf{R}^{-1}\mathbf{J_E})^{-1}\mathbf{J_E}^T\mathbf{R}^{-1}$, $\Omega = \mathrm{diag}\{Cov(\mathbf{r})\}$, $\mathbf{S} = \Omega^{-\frac{1}{2}}(\mathbf{I} - \mathbf{K_E})\mathbf{M}$ is the sensitivity matrix for $\mathbf{r}^N$ with respect to branch flow errors $\mathbf{f}$, and η is the threshold of topology error detection. H_1 and H_0 correspond to the cases with and without topology error, respectively. On the other hand, topology error processing for substation configuration error is initially carried out using an augmented set

of measurement equations in terms of the state variables and the CB power flows. However, since this method use the whole set of substations in full detailed model it can increase computation complexity. Recently, reduced model [69] and implicit model [70] are proposed to decrease computation complexity with much reduced number of state variables.

2.2 Real-Time Power Market

In deregulated electricity markets, the nodal price at each bus is computed by RTOs. The electric power market consists of two-settlement system, day-ahead market and real-time spot market. Day-ahead market is often called as a forward market where clearing optimal generation output and prices at every hour are calculated for each hour of the next operation day based on generation offers, demand bids and scheduled bilateral transactions. Real-time spot market is a spot market where prices are calculated at every five minute based on the actual grid operation conditions.

The spot electricity nodal prices are obtained as the by-product of SCED. Based on the industry practice, SCED is usually conducted using a linear DCOPF-based model via linear programming (LP) due to the robustness of convergence and fast computation time of LP [33]. There are two representative pricing models for SCED: Ex-ante (e.g. in ERCOT, NY ISO) and Ex-post (e.g. in ISO New England, PJM, and Midwest ISO). In ex-ante pricing model, SCED is carried out every 10 to 15 minute prior to real time grid operation in order to determine the optimal generation output and nodal price while satisfying the balance of supply and demand and operational constraints. In ex-post pricing model, only prices are calculated by SCED at every five minute based on the actual system condition for settlement purposes.

On the other hand, the real-time spot market can be operated by two types of economic dispatch methods: static dispatch and look-ahead dispatch. The former

computes the optimal dispatch for the current operating interval whereas the latter calculates the optimal dispatch solution over multiple future time intervals. Fig. 2.3 illustrates a diagram of two-settlement electricity market operation including the conceptual difference of look-ahead and static dispatch methods.

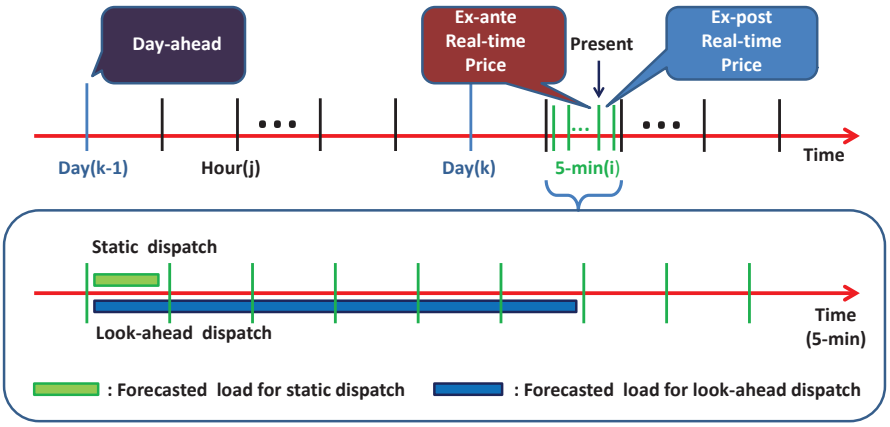


Figure 2.3: Electricity market operation.

2.2.1 Ex-Ante Market

In ex-ante real-time pricing models, the optimal generation output and LMP are computed before the actual deployment of dispatch orders. For the system operator, the Ex-ante dispatch is formulated as follows:

$$\min_{P_{g_i}} \sum_{i \in G} C_i(P_{g_i}) \tag{2.26}$$

s.t.

$$\sum_{i \in G} P_{g_i} = \sum_{n=1}^{N} D_n \tag{2.27}$$

$$\hat{P}_{g_i}^{\min} \leq P_{g_i} \leq \hat{P}_{g_i}^{\max} \quad \forall i \in G \tag{2.28}$$

$$F_l^{\min} \leq F_l \leq F_l^{\max} \quad \forall l = 1, \ldots, L \tag{2.29}$$

where

$$\hat{P}_{g_i}^{\max} = \min\{P_{g_i}^{\max}, \hat{P}_{g_i}(\mathbf{z}) + R_i \Delta T\} \tag{2.30}$$

$$\hat{P}_{g_i}^{\min} = \max\{P_{g_i}^{\min}, \hat{P}_{g_i}(\mathbf{z}) - R_i \Delta T\}. \tag{2.31}$$

In this formulation, the objective function is to minimize the total generation costs in (2.26). (2.27) is the system-wide energy balance equation. (2.28) is the physical capacity constraints of each generator embedded with its ramp constraints. (2.29) is the transmission line constraints. It should be noted from (2.30) and (2.31) that the physical capacity limits of each generator depend on the state estimation result based on sensor measurement $\mathbf{z}$. The calculated optimal dispatch is then allocated to corresponding market participant such as generation company and load serving entity.

2.2.2 Ex-Post Market

In ex-post real-time pricing models, LMPs are computed after the fact using real-time estimates for settlement purposes. Assuming no demand elasticity, the Ex-post dispatch is written as:

$$\min_{P_{g_i}} \sum_{i \in G} C_i(P_{g_i}) \tag{2.32}$$

s.t.

$$\sum_{i \in G} \Delta P_{g_i} = 0 \tag{2.33}$$

$$\Delta P_{g_i}^{\min} \leq \Delta P_{g_i} \leq \Delta P_{g_i}^{\max} \quad \forall i \in G \tag{2.34}$$

$$\Delta F_l \leq 0 \qquad \forall l \in \mathcal{CL}_+ \tag{2.35}$$

$$\Delta F_l \geq 0 \qquad \forall l \in \mathcal{CL}_- \tag{2.36}$$

where

$$\Delta P_{g_i} = P_{g_i} - \hat{P}_{g_i}(\mathbf{z}), \quad \Delta F_l = F_l - \hat{F}_l(\mathbf{z})$$

$$\mathcal{CL}_+ = \{l : \hat{F}_l(\mathbf{z}) \geq F_l^{\max}\}, \quad \mathcal{CL}_- = \{l : \hat{F}_l(\mathbf{z}) \leq F_l^{\min}\}$$

and $\Delta P_{g_i}^{\max}$ and $\Delta P_{g_i}^{\min}$ are usually chosen to be 0.1MWh and -2MWh, respectively. In comparison with ex-ante pricing model, ex-post pricing model relies on the state estimation result more.

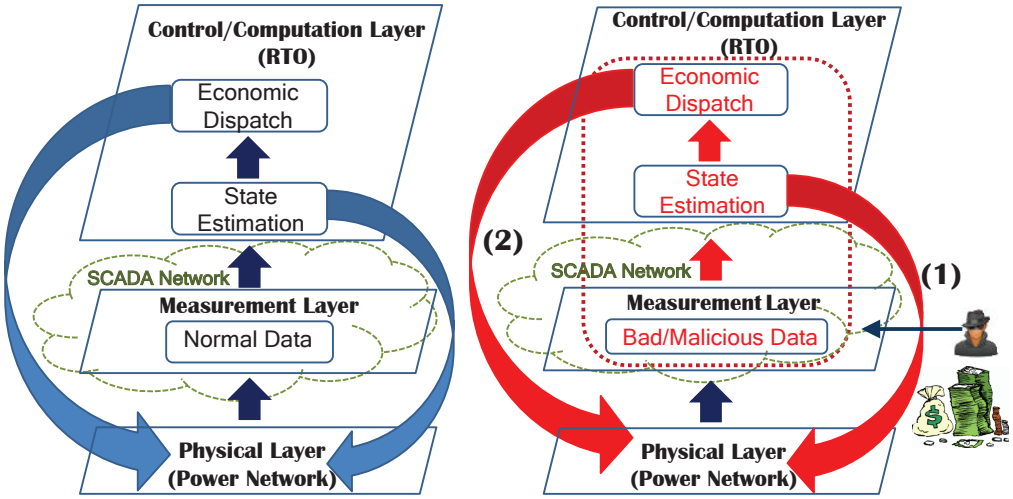
3. MALICIOUS TEMPORAL DATA ATTACK ON TIME-COUPLED LOOK-AHEAD DISPATCH*

3.1 Introduction

Critical infrastructure (e.g., the electricity grid) has been facing an increasing number of potential cyber attacks. Given the much stronger coupling between cyber and physical layers of smart grid, development of cyber security technology tailored for smart grid is of paramount importance.

The main objective of this section is to study the impact of cyber attacks on state estimation, which subsequently influence the result of the newly emerging *look-ahead dispatch model* in the real-time electricity market. Figs. 3.1(a),(b) illustrate the information flow in a three-layered framework (with physical, measurement, and control/computation layer) without and with such cyber attacks, respectively. The information includes the physical state such as the nodal power injection and flow and the dispatch instruction such as the optimal generation output and nodal price. Compared to Fig. 3.1(a), Fig. 3.1(b) describes that bad/malicious data injected into the measurement layer can lead to corrupted estimation of the states of the physical layer. Consequently, the attacker could distort the feedback information from control/communication layer back to the physical layer in two ways, leading to (1) physical insecurity in the power grid operations, and/or (2) financial misconduct in the power markets as shown in Fig. 3.1(b). This section contributes to topic (2) using a more realistic dispatch model in power markets.

In this section we present a new type of potential cyber attacks in more realistic



(a) Without cyber data attack (b) With cyber data attack

Figure 3.1: A three-layered framework illustrating cyber data attack.

economic dispatch model, i.e., *look-ahead* dispatch. Motivated by the increasing penetration of variable resources such as wind and solar, look-ahead dispatch has been implemented by major Independent System Operators (ISOs)/Regional Transmission Organizations (RTOs) in the past few years in order to improve the market dispatch efficiency. Look-ahead dispatch is different from conventional static dispatch in that it calculates the optimal dispatch in an extended period of time, taking into account inter-temporal ramp rates of generators of different technologies. In this section, an attack strategy is demonstrated, in which the attacker could withhold generation capacity for financial gain by stealthily manipulating the ramp constraint limits of generators in look ahead dispatch. It should be noted that the proposed attack strategy is different from the capacity withholding methods used for a generation company to report capacity noticeably lower than its maximum capacity based on learning algorithm (e.g., SA-Q-Learning algorithm) [71, 72]. In contrast, the proposed method is to inject *undetectable* malicious data into sensors in order to withhold capacity for

financial misconduct in real-time markets. The main contributions of this paper are two-fold:

1. We formulate a malicious ramp-induced data (RID) attack problem in look-ahead dispatch. The attacker could stealthily change the ramp constraint limits of generators through manipulating sensors' data, aiming at increasing the nodal price by withholding capacity of generator.

2. We propose a RID attack strategy with which the attacker could make a profit without being detected by RTOs in the real-time electricity market. Numerical examples are illustrated in the IEEE-14 bus system.

Following this introductory section, the rest of Section 3 is structured as follows. Section 3.2 provides the brief overview of DC state estimation and real-time power market with look-ahead dispatch model. Section 3.3 overview spatial data attack on static economic dispatch in ex-post pricing model. Two types of profitable and undetectable attack strategies based on virtual bidding mechanism are formulated, which lead to the distortion of ex-post LMPs. Section 3.4 states the proposed ramp-induced data attack problem in real-time power markets with look-ahead dispatch. The required attack conditions, attack procedure and strategy, and attack performance metrics are elaborated in more detail in Section 3.5, which is followed by illustrative examples based on the IEEE 14-bus test system in Section 3.6.

3.2 Background

The notations used in this paper are summarized in Table 3.1.

3.2.1 DC State Estimation Model

We consider the scenario in which the adversary attacks the linearized DC state estimation model. As shown in (2.8), DC state estimation model normally includes

the bus voltage phase angle θ as the system state. Given a fixed topology and choice of slack bus, there exists a bijective relationship between bus voltage phase angle and the vector of nodal power injection x. Since the LMPs are explicitly calculated from nodal power injections, the states in this section is defined as the vector of nodal power injection $\mathbf{x}$. Therefore, the linearized DC state estimation measurement model is rewritten as follows:

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{e} = \begin{bmatrix} \mathbf{I} \\ \\ \mathbf{H_d} \end{bmatrix} \mathbf{x} + \mathbf{e}, \tag{3.1}$$

where

- $\mathbf{x}$: state vector (nodal power injections)
- $\mathbf{z}$: measurement vector (power injection and flow measurements)
- $\mathbf{e}$: independent identically distributed (i.i.d.) Gaussian measurement error vector following $\mathcal{N}(0, \mathbf{R})$
- $\mathbf{H}$: the system factor matrix specifying the relationship between $\mathbf{x}$ and $\mathbf{z}$.

Here the matrix $\mathbf{H}$ is concatenated with two submatrices, $\mathbf{H_d}$ and $\mathbf{I}$, which denote the distribution factor matrix and the identity matrix, respectively.

The state estimation problem is to find the optimal estimate of $\mathbf{x}$ to minimize the weighted least square of measurement error:

$$\text{minimize } J(\mathbf{x}) = \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} \tag{3.2}$$

$$\text{s.t. } \mathbf{r} = \mathbf{z} - \mathbf{H}\mathbf{x}, \tag{3.3}$$

where $\mathbf{r}$ is the estimated residual vector. If the system is observable (i.e., the system factor matrix $\mathbf{H}$ is full rank), the unique weighted least squares estimate of $\mathbf{x}$ is given

by

$$\hat{\mathbf{x}}(\mathbf{z}) = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{z} = \mathbf{B}\mathbf{z}. \tag{3.4}$$

Table 3.1: Notations.

i	Index for generators i
n	Index for buses n
l	Index for transmission line l
K	Total number of sampling period
N	Total number of buses
L	Total number of transmission lines
M	Total number of measurements
G	Set of generation units
G_M	Set of marginal units
$\underline{G}_M^c$	Set of binding units with lower marginal cost than marginal unit
$\overline{G}_M^c$	Set of binding units with higher marginal cost than marginal unit
D	Set of demands
$\hat{D}_n[k]$	nth bus fixed demand at time k
$P_{g_i}[k]$	Scheduled ith generator power at time k
$F_l[k]$	Transmission flow at line l at time k
R_i	Ramp rate of generator i
ΔT	Dispatch interval
$P_{g_i}^{\min}$, $P_{g_i}^{\max}$	Min/max generation limits for generator i
$F_l^{\min}$, $F_l^{\max}$	Min/max flow limits at line l

3.2.2 Look-Ahead Dispatch Model

The electric power market consists of two-settlement system, day-ahead and real-time spot markets. Recently, due to limited predictability in day-ahead and high inter-temporal variability of renewable resources (e.g., wind and solar), RTOs are

upgrading real-time market clearing engine from static dispatch to look-ahead dispatch models for more flexible operations in support of high penetration of variable resources. For the system operator, look-ahead dispatch is formulated as follows,

$$\min_{P_{g_i}[k]} \sum_{k=1}^{K} \sum_{i \in G} C_i(P_{g_i}[k]) \tag{3.5}$$

s.t.

$$\sum_{i \in G} P_{g_i}[k] = \sum_{n=1}^{N} \hat{D}_n[k] \qquad \forall k = 1, \ldots, K \tag{3.6}$$

$$|P_{g_i}[k] - P_{g_i}[k-1]| \leq R_i \Delta T \qquad \forall k = 1, \ldots, K \tag{3.7}$$

$$P_{g_i}^{\min} \leq P_{g_i}[k] \leq P_{g_i}^{\max} \qquad \forall k = 1, \ldots, K \tag{3.8}$$

$$F_l^{\min} \leq F_l[k] \leq F_l^{\max} \qquad \forall k = 1, \ldots, K. \tag{3.9}$$

In this formulation, the objective function is to minimize the total generation costs in (3.5). (3.6) is the system-wide energy balance equations. (3.7) and (3.8) are the ramp constraints and the physical capacity constraints of each generator, respectively. (3.9) is the transmission line constraints. In this section, we define one-step look-ahead dispatch with $K = 1$ as static dispatch. The Lagrangian function of the

aforementioned look-ahead dispatch is written as

$$\begin{aligned}
\mathcal{L} = & \sum_{k=1}^{K}\sum_{i\in G} C_i(P_{g_i}[k]) - \sum_{k=1}^{K}\lambda[k]\left[\sum_{i\in G} P_{g_i}[k] - \sum_{n=1}^{N}\hat{D}_n[k]\right] \\
& + \sum_{k=1}^{K}\sum_{i\in G}\left[\omega_{i,\max}[k](P_{g_i}[k] - P_{g_i}[k-1] - R_i\Delta T)\right] \\
& + \sum_{k=1}^{K}\sum_{i\in G}\left[\omega_{i,\min}[k](P_{g_i}[k-1] - P_{g_i}[k] - R_i\Delta T)\right] \\
& + \sum_{k=1}^{K}\sum_{i\in G}\left[\tau_{i,\max}[k](P_{g_i}[k] - P_{g_i}^{\max})\right] \\
& + \sum_{k=1}^{K}\sum_{i\in G}\left[\tau_{i,\min}[k](P_{g_i}^{\min} - P_{g_i}[k])\right] \\
& + \sum_{k=1}^{K}\sum_{l=1}^{L}\left[\mu_{l,\max}[k](F_l[k] - F_l^{\max})\right] \\
& + \sum_{k=1}^{K}\sum_{l=1}^{L}\left[\mu_{l,\min}[k](F_l^{\min} - F_l[k])\right],
\end{aligned}$$

where all the Lagrangian multipliers at time k ($\lambda[k]$, $\omega_{i,\max}[k]$, $\omega_{i,\min}[k]$, $\tau_{i,\max}[k]$, $\tau_{i,\min}[k]$, $\mu_{l,\max}[k]$, and $\mu_{l,\min}[k]$) are positive. According to the definition of the nodal price [73], and assuming that bus 1 is the slack bus, the locational marginal price (LMP) for each bus n ($n = 2, \ldots, N$) at time k is given by

$$\lambda_n[k] = \lambda[k] - \mathbf{H_n^d}^T(\mu_{\max}[k] - \mu_{\min}[k]), \tag{3.10}$$

where $\lambda[k]$ is the LMP for the slack bus 1 at time k, $\mathbf{H_n^d} = [\frac{\partial F_1}{\partial \hat{D}_n}, \ldots, \frac{\partial F_L}{\partial \hat{D}_n}]^T$, $\mu_{\max}[k] = [\mu_{1,\max}[k], \ldots, \mu_{L,\max}[k]]^T$, and $\mu_{\min}[k] = [\mu_{1,\min}[k], \ldots, \mu_{L,\min}[k]]^T$.

Alternatively, by the first-order KKT condition of look-ahead dispatch formula-

tion, the LMP for each generator i connected to bus n is written as

$$\begin{aligned}\lambda_i[k] = \frac{\partial C_i(P_{g_i}[k])}{\partial P_{g_i}[k]} - \mathbf{H}_{\mathbf{n}}^{\mathbf{d}\,T}(\mu_{\max}[k] - \mu_{\min}[k]) \\ + (\tau_{i,\max}[k] - \tau_{i,\min}[k]) + (\omega_{i,\max}[k] - \omega_{i,\max}[k+1]\mathbb{1}_A[k]) \\ + (\omega_{i,\min}[k+1]\mathbb{1}_A[k] - \omega_{i,\min}[k])\,, \end{aligned} \tag{3.11}$$

where $\mathbb{1}_A[k]$ is the indicator function based on the set $A = \{1 \leq k \leq K-1\}$. In other words, $\mathbb{1}_A[k]$=1 when $k \in A$, otherwise (i.e., $k \in A^c = \{k = K\}$) $\mathbb{1}_A[k]$=0. We can observe from (3.11) that the Lagrangian multipliers, $\omega_{i,\max}[k+1]$ and $\omega_{i,\min}[k+1]$, corresponding to the ramp constraints at the future time $k+1$ influence the LMPs calculation at the current time k. However, the LMP formulation in static dispatch (one-step look-ahead) does not capture future constraints.

3.3 Overview of Spatial Data Attack on Static Dispatch

In this section, we overview the problem of false data injection attack on static dispatch in ex-post pricing model [24]. A key result is to develop a profitable and undetectable attack strategy in which the attacker uses virtual bidding mechanism to result in consistent financial arbitrage between day-ahead and ex-post real-time prices at selected pairs of nodes.

3.3.1 Ex-Post LMP Formulation

From ex-post pricing formulation in Subsection 2.2.2, the Lagrangian of this formulation is defined as

$$\begin{aligned}\mathcal{L} = \sum_{i \in G} C_i(P_{g_i}) - \lambda \sum_{i \in G} \Delta P_{g_i} + \sum_{i \in G} \mu_{i,\max}\left(\Delta P_{g_i} - \Delta P_{g_i}^{\max}\right) + \\ \sum_{i \in G} \mu_{i,\min}\left(\Delta P_{g_i}^{\min} - \Delta P_{g_i}\right) + \sum_{l \in \mathcal{CL}_+} \eta_l \Delta F_l + \sum_{l \in \mathcal{CL}_-} \zeta_l\left(-\Delta F_l\right)\end{aligned} \tag{3.12}$$

where all Lagrangian multipliers positive or equal to zero. We define $\eta_l = 0$ if $l \notin \mathcal{CL}_+$, $\zeta_l = 0$ if $l \notin \mathcal{CL}_-$. Then, the nodal price at each bus n ($n = 2, \ldots, N$) is given by

$$\lambda_n = \lambda + \sum_{l=1}^{L} (\eta_l - \zeta_l) \frac{\partial F_l}{\partial D_n}. \tag{3.13}$$

We can write (3.13) in a matrix form. Let us define $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_L]'$ and $\boldsymbol{\zeta} = [\zeta_1, \ldots, \zeta_L]'$. Since $\partial F_l / \partial D_n = H^d_{ln}$ where H^d_{ln} is the element on the lth row and nth column of $\mathbf{H^d}$, (3.13) can be rewritten as

$$\lambda_n = \lambda + \mathbf{H_n^{d}}^T (\boldsymbol{\eta} - \boldsymbol{\zeta}), \tag{3.14}$$

where $\mathbf{H_n^d}$ is the nth column of $\mathbf{H^d}$ matrix. Then, the price gap at two nodes n_1 and n_2 is expressed as

$$\lambda_{n_1} - \lambda_{n_2} = \left(\mathbf{H_{n_1}^d} - \mathbf{H_{n_2}^d}\right)^T (\boldsymbol{\eta} - \boldsymbol{\zeta}). \tag{3.15}$$

3.3.2 Attack Model and Undetectability

We consider the additive attack measurement model:

$$\mathbf{z_a} = \mathbf{Hx} + \mathbf{e} + \mathbf{a}, \tag{3.16}$$

where $\mathbf{a}$ is the attack vector, which leads to the corrupted measurement vector $\mathbf{z_a}$. The new residual vector $\mathbf{r_a}$ can be decomposed into two terms, corresponding to without and with attack, respectively:

$$\mathbf{r_a} = \mathbf{r} + (\mathbf{I} - \mathbf{HB})\mathbf{a} \tag{3.17}$$

, and by triangular inequality of the L_2-norm $||\cdot||_2$,

$$\begin{aligned} ||\mathbf{r_a}||_2 &= ||\mathbf{r} + (\mathbf{I} - \mathbf{HB})\mathbf{a}||_2 \\ &\leq ||\mathbf{r}||_2 + ||(\mathbf{I} - \mathbf{HB})\mathbf{a}||_2 < \eta, \end{aligned} \tag{3.18}$$

where η is the bad data detection threshold. For bypassing the bad data detection algorithm, the attacker aims at constructing the attack vector $\mathbf{a}$ so that the value of $||(\mathbf{I}-\mathbf{HB})\mathbf{a}||_2$ added to $||\mathbf{r}||_2$ still makes the above undetectable condition hold true.

3.3.3 Attack Procedure Using Virtual Bidding Mechanism

The attacker is assumed to exploit the virtual bidding mechanism for making a profit. In some RTOs, virtual bidding activities are legitimate financial instruments in electricity markets with some advantages such as mitigation of supplier market power and market efficiency improvement due to the convergence of the prices in day-ahead and real-timer markets. A market participant purchase/sell a certain amount of virtual power at location in day-ahead forward market and sell/purchase the exact same amount of power in the subsequent real-time market. The attack procedure exploiting virtual bidding mechanism can be summarized as follows:

Step 1) In day-ahead forward market, the attacker buys and sells virtual power P_v at buses n_1 and n_2 with day-ahead prices $\lambda_{n_1}^{DA}$ and $\lambda_{n_2}^{DA}$, respectively.

Step 2) The attacker injects the malicious attack vector $\mathbf{a}$ into sensor measurement vector $\mathbf{z}$ to manipulate the nodal price of ex-post market. Therefore, the corrupted measurement vector $\mathbf{z_a}$ is written as $\mathbf{z_a} = \mathbf{z} + \mathbf{a}$.

Step 3) In ex-post real-time market, the attacker again sells and buys virtual power P_v at buses n_1 and n_2 at real-time prices λ_{n_1} and λ_{n_2}, respectively.

Therefore, the profit obtained through this virtual trading transaction is written as

$$\text{Profit} = \left(\lambda_{n_1} - \lambda_{n_1}^{DA}\right) P_v + \left(\lambda_{n_2}^{DA} - \lambda_{n_2}\right) P_v = (\underbrace{\lambda_{n_1} - \lambda_{n_2} + \lambda_{n_2}^{DA} - \lambda_{n_1}^{DA}}_{p}) P_v. \quad (3.19)$$

In (3.19), the following p is defined as a profit signal

$$p = \lambda_{n_1} - \lambda_{n_2} + \lambda_{n_2}^{DA} - \lambda_{n_1}^{DA}. \quad (3.20)$$

Using (3.15), (3.20) changed by a malicious data injection can be written as

$$p(\mathbf{z_a}) = \left(\mathbf{H_{n_1}^d} - \mathbf{H_{n_2}^d}\right)^T (\eta(\mathbf{z_a}) - \xi(\mathbf{z_a})) + \lambda_{n_2}^{DA} - \lambda_{n_1}^{DA}. \quad (3.21)$$

It should be noted that Lagrangian multipliers $\eta(\mathbf{z_a})$ and $\xi(\mathbf{z_a})$ are manipulated by corrupted measurement vector $\mathbf{z_a}$.

3.3.4 Attack Strategy

In this subsection, two cases are considered where the subset of compromised sensors is fixed and only a limited number of measurement sensors could be compromised.

Case I: predetermined subset of compromised sensors

The goal of the attacker is to find a profitable input $\mathbf{a}$ when the subset of compromised sensors is fixed. Let us define the set

$$L_+ = \{l : H_{ln_1}^d > H_{ln_2}^d\}, \qquad L_- = \{l : H_{ln_1}^d < H_{ln_2}^d\}.$$

Then, $p(\mathbf{z_a})$ can be decomposed into

$$\begin{aligned} p(\mathbf{z_a}) &= \sum_{l\in L_+} \left(H^d_{ln_1} - H^d_{ln_2}\right)\left(\eta_l(\mathbf{z_a}) - \zeta_l(\mathbf{z_a})\right) \\ &+ \sum_{l\in L_-} \left(H^d_{ln_2} - H^d_{ln_1}\right)\left(\zeta_l(\mathbf{z_a}) - \eta_l(\mathbf{z_a})\right) + \lambda^{DA}_{n_2} - \lambda^{DA}_{n_1}. \end{aligned} \tag{3.22}$$

By the fact that $\eta_l(\zeta_l)$ is nonnegative and it is 0 if the line is not congested, the sufficient conditions for $p(\mathbf{z_a}) > 0$ are as follows:

(C1) $\lambda^{DA}_{n_2} > \lambda^{DA}_{n_1}$.

(C2) $\hat{F}'_l < F^{\max}_l$ if $l \in L_-$.

(C3) $\hat{F}'_l > F^{\min}_l$ if $l \in L_+$.

The attacker could satisfy (C1) easily in the day-ahead market. That is, the attacker knows day-ahead prices at specific buses. The attacker needs to change the measurement in order to make sure that (C2) and (C3) hold true. We give the following definition:

Definition 1. *An attack input* $\mathbf{a}$ *is called* δ*-profitable if the following inequalities hold*

$$\mathbb{E}[\hat{F}'_l] \leq F^{max}_l - \delta, \quad \forall l \in L_-, \tag{3.23}$$

$$\mathbb{E}[\hat{F}'_l] \geq F^{min}_l - \delta, \quad \forall l \in L_+, \tag{3.24}$$

where $\mathbb{E}[\hat{F}'_l] = F^* + \mathbf{H^d Ba}$ *and* F^* *is the result of the ex-ante dispatch.*

Therefore, the attacker's strategy is to find an ϵ feasible $\mathbf{a}$ such that the margin δ is maximized. The attack method can be formulated as

$$\max_{\mathbf{a}\in\text{span}(\mathcal{A})} \delta \tag{3.25}$$

s.t.

$$||(\mathbf{I}-\mathbf{HB})\mathbf{a}||_2 \leq \epsilon \tag{3.26}$$

$$\mathbb{E}[\hat{F}_l'] \leq F_l^{\max} - \delta, \quad \forall l \in L_-, \tag{3.27}$$

$$\mathbb{E}[\hat{F}_l'] \geq F_l^{\min} - \delta, \quad \forall l \in L_+, \tag{3.28}$$

$$\delta > 0 \tag{3.29}$$

where the set $\mathcal{A}$ represents the attack vector space, which describes the attack pattern related to the type and number of compromised sensors. We can see from the above optimization problem that the objective function and all the constraints are convex. Therefore, this problem is a convex programming problem and can be solved efficiently.

Case II: limited resources to compromise sensors

We consider a case where the attacker is able to choose the set of sensors to compromise. However, since the attacker has limited capabilities, the total number of compromised sensors cannot exceed certain threshold κ. Consequently, the attacker need to construct an optimal attack vector to system and to choose the optimal set of sensors to compromise at the same time.

To this end, we can write the optimization problem as

$$\max_{\mathbf{a}\in\text{span}(\mathcal{A})} \delta \tag{3.30}$$

s.t.

$$||(\mathbf{I}-\mathbf{HB})\mathbf{a}||_2 \leq \epsilon \tag{3.31}$$

$$\mathbb{E}[\hat{F}_l'] \leq F_l^{\max} - \delta, \quad \forall l \in L_-, \tag{3.32}$$

$$\mathbb{E}[\hat{F}_l'] \geq F_l^{\min} - \delta, \quad \forall l \in L_+, \tag{3.33}$$

$$\delta > 0 \tag{3.34}$$

$$||\mathbf{a}||_0 \leq \kappa, \tag{3.35}$$

Here, $||\cdot||_0$ implies the zero norm, which is defined as the number of nonzero elements in a vector. The nonzero elements of $\mathbf{a}$ correspond to the sensors the attacker wishes to compromise. In this formulation we note that $\mathbf{a}$ does not necessarily lie in the span of $\mathcal{A}$, but instead $\mathbf{a}$ must have no more than κ nonzero elements.

3.4 Statement of Temporal Data Attack Problem

A general problem of cyber attack against state estimation in economic dispatch can be illustrated in statistical signal processing framework in Fig. 3.2. It provides a graphical interpretation for the relationship among sensor's measurement, state estimation, and economic dispatch. The state set $\boldsymbol{\mathcal{S}}$ is partitioned into a finite number of nodal price subsets $\lambda_i[1]$. The operation of economic dispatch is implicitly included in the state set $\boldsymbol{\mathcal{S}}$. The measurement space $\boldsymbol{\mathcal{M}}$ is the collection of all realizable sensor's measurements $\mathbf{z}$. The set of probability measure $\boldsymbol{\mu}$ provides a mathematical basis for describing the randomness of measurements. In the power system state

estimation literature, the probability measure normally follows the Gaussian distribution. These random measurement errors can be filtered by the existing bad data processing algorithm. The objective of the attacker is to move the estimate from a certain nodal price subset to a desired nodal price subset by corrupting original measurements $\mathbf{z}$ into $\mathbf{z_a}$ while avoiding the bad data detection. Detailed attack model and formulation are described in Section 3.5.

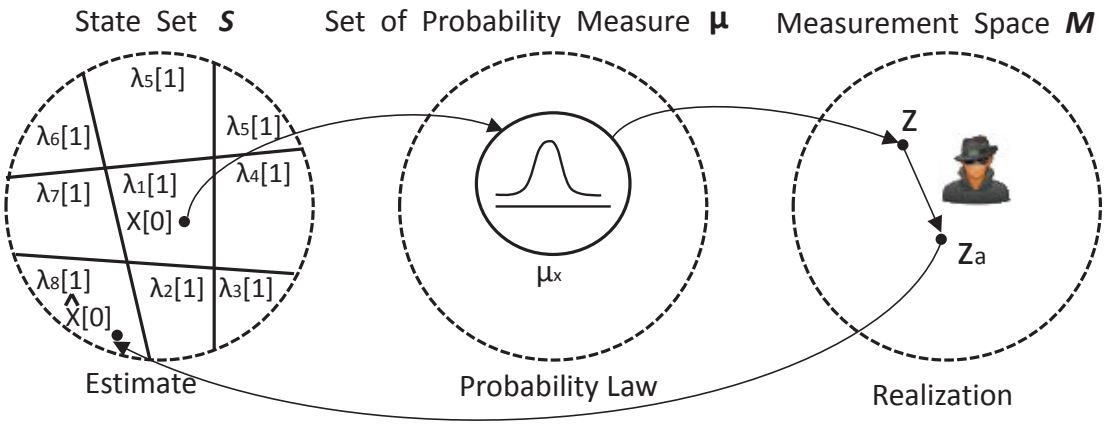


Figure 3.2: Statistical signal processing framework illustrating the relationship among sensor data, state estimation, and economic dispatch.

In the above framework, a potential cyber attack in look-ahead dispatch is described as follows. The ith unit's initial generation power $P_{g_i}[0]$ embedded in (3.7) is replaced, at every dispatch interval, by its corresponding estimate $\hat{P}_{g_i}(\mathbf{z})$, which is processed and delivered by the state estimator. Therefore, in static dispatch the

generation power of unit i at $k = 1$ becomes bounded by

$$P_{g_i}^{\max}[1] = \min\{P_{g_i}^{\max}, P_{g_{i,\mathrm{R}}}^{\max}(\mathbf{z})\} \tag{3.36}$$

$$P_{g_i}^{\min}[1] = \max\{P_{g_i}^{\min}, P_{g_{i,\mathrm{R}}}^{\min}(\mathbf{z})\}, \tag{3.37}$$

where the maximum and minimum limits of the ramp constraints, $P_{g_{i,\mathrm{R}}}^{\max}(\mathbf{z})$ and $P_{g_{i,\mathrm{R}}}^{\min}(\mathbf{z})$, are

$$P_{g_{i,\mathrm{R}}}^{\max}(\mathbf{z}) = \hat{P}_{g_i}(\mathbf{z}) + R_i\Delta T, \quad P_{g_{i,\mathrm{R}}}^{\min}(\mathbf{z}) = \hat{P}_{g_i}(\mathbf{z}) - R_i\Delta T. \tag{3.38}$$

If the attacker manipulates the estimate $\hat{P}_{g_i}(\mathbf{z})$ by injecting false data into $\mathbf{z}$ so that the capacity limits of unit i at $k = 1$ are binding to stealthily changed ramp constraint limits, the optimal generation dispatch and nodal price might be miscalculated by RTOs. In this section we define this type of attack as a ramp-induced data (RID) attack in a potential class of malicious inter-temporal data attacks.

Fig. 3.3 illustrates the RID attack, which withholds capacity of a marginal unit (a part-loaded generator). Left and right diagrams describe the generation characteristics of the marginal unit without and with the attack, respectively. W is the feasible range of generation limited by the ramp rate of the marginal unit, and ΔL is an incremental (in this figure) or decremental system load from $k = 0$ to $k = 1$. We note that as $\hat{P}_{g_i}[0]$ (for simplicity, we omit $\mathbf{z}$, instead emphasize the time) is manipulated by the attacker, ΔL can deviate, upwards or downwards, from the range of W, leading to capacity withholding or capacity withdrawing, respectively. The right diagram in Fig. 3.3 shows that if $\hat{P}_{g_i}[0]$ is decreased to $\hat{P}_{g_i,\mathbf{a}}[0]$ by the attacker at $k = 0$ so that ΔL deviates upwards from the range of W, the attacker succeeds in withholding capacity, resulting in a new dispatch output $\hat{P}^*_{g_i,\mathbf{a}}[1]$ at $k = 1$. As

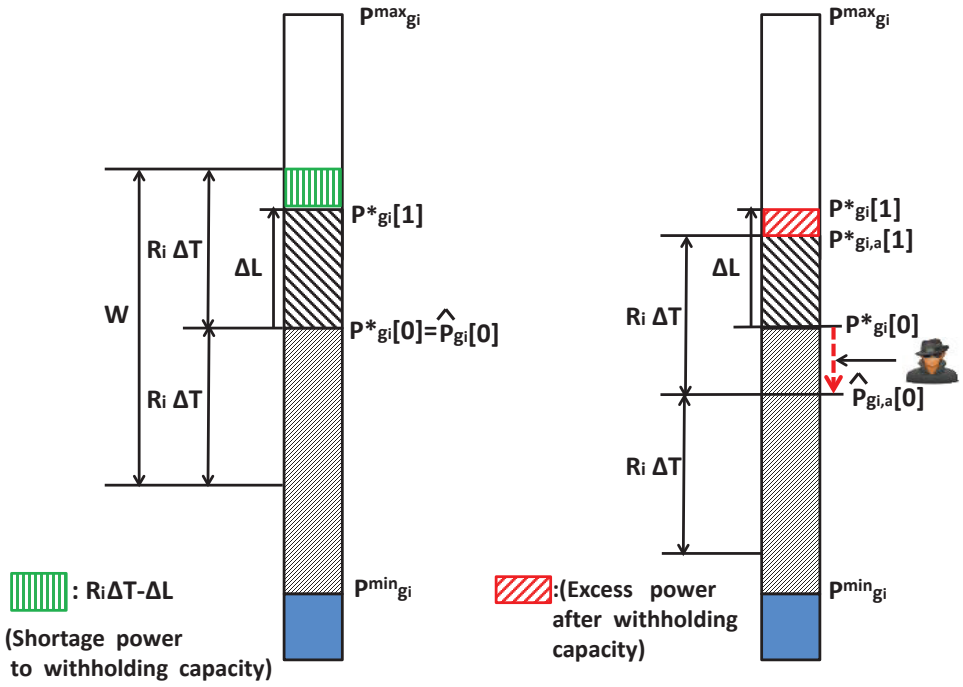


Figure 3.3: Conceptual diagrams illustrating a ramp-induced data attack.

a result, the infra-marginal unit (the unit with the next higher marginal cost) is dispatched to supply the excess demand, consequently leading to a uniformly higher market price.

Remark 1. *Define* $\hat{P}_{g_i,\mathbf{a}}[0] - P^*_{g_i}[0]$ *as the contribution of the attacker to changing the nodal price. The RID attack fails (i.e., the nodal price remains unchanged) if the value of this contribution belongs to the following interval:*

$$\Delta L - R_i \Delta T \leq \hat{P}_{g_i,\mathbf{a}}[0] - P^*_{g_i}[0] \leq \Delta L + R_i \Delta T. \tag{3.39}$$

The feasible region of $\hat{P}_{g_i}[0]$ *based on constraint* (3.39) *is defined as the price-invulnerable region.*

3.5 Formulation of the Ramp-Induced Data Attack

3.5.1 Requirements and Procedure for a Successful RID Attack

From the analysis above, in order to implement a RID attack with profits, the attacker is required to have the knowledge of

(R1) the system topology (e.g., distribution factor matrix), which remains constant at every dispatch interval

(R2) the ramp rates of the targeted generators

(R3) the amount of changing system load between two consecutive dispatch intervals

The system topology for the targeted power system in Requirement (R1) can be simply obtained off-line by an internal intruder in a control center or estimated by linear independent component analysis (ICA) technique proposed in [20]. For Requirement (R2), typical ramp rates are estimable for typical generators. Requirement (R3) is feasible since the attacker can estimate an amount of changing system load from RTOs' website. With these assumptions, the procedure of the proposed RID attack is summarized as follows:

Step 1) The attacker synchronizes the attack time with the start time at every dispatch interval. This step is necessary for injected false data to mislead economic dispatch via the state estimator.

Step 2) The attacker determines sensors to compromise and computes the attack vector using the proposed attack strategy formulated in the next subsection.

Step 3) The attacker injects the attack vector into sensors' measurements at the attack time set in Step 1). Then, these corrupted measurements are transmitted to the state estimator via SCADA network.

Step 4) The state estimator based on received false measurements may lead to distorted generation output estimates. They are utilized for setting the ramp constraints in look-ahead dispatch.

Step 5) Consequently, the manipulated ramp constraints result in the attacker's desired dispatch instruction. Then, it is sent to the dispatchable generators.

Step 6) For the continuous attack, the procedure goes back to Step 2).

3.5.2 Proposed Attack Strategy

In this subsection we formulate a ramp-induced data attack strategy. The power system is assumed to have sufficient transmission capacity. As the first step toward understanding the impact of cyber attack on *temporal* ramp-constrained economic dispatch, we exclude the impact of *spatial* transmission congestion on the market clearning prices. In practice, temporal ramp constraints are coexisting with spatial transmission flow constraints. Therefore, for a successful RID attack in congested networks the attacker should know the targeted power system very well and as much as the system operator knows, however this scenario is unrealistic. Developing a feasible RID attack strategy in congested networks is beyond the scope of this paper and referred to as a future work.

The proposed attacks are classified into the following three types:

- *Marginal unit attack*: a injection measurement sensor associated with the marginal unit is compromised.
- *Binding unit attack*: injection measurement sensors associated with the binding units are compromised.
- *Coordinated attack*: injection measurement sensors associated with the binding units as well as the marginal unit are compromised.

Here a binding unit represents two types of units: an intra-marginal unit with the lower marginal cost or an infra-marginal unit with the higher marginal cost than a marginal unit. Fig. 3.4 shows a conceptual diagram for a binding unit attack. The following proposed attack strategy and simulation results focus on intra-marginal unit attack belonging to binding unit attack.

Remark 2. *When there is no network transmission congestion, it is well acknowledged that static dispatch involves a single marginal unit and multiple binding units that produce their minimum or maximum outputs. On the other hand, look-ahead dispatch may involve multiple marginal units even if there is no congestion in the the transmission network. In this paper the marginal unit attack is associated with the marginal unit in static dispatch.*

For achieving undetectability and profitability, the attacker computes the attack vector $\mathbf{a}$ by compromising sensors $i \in G_M$ or $j \in \underline{G}_M^c$, which is the solution of the following optimization problem:

$$\max_{\mathbf{a} \in \text{span}(\mathcal{A})} \delta \tag{3.40}$$

s.t.

$$||(\mathbf{I} - \mathbf{HB})\mathbf{a}||_2 \leq \epsilon \tag{3.41}$$

$$\alpha \mathcal{C}_M(\mathbf{a}) + \beta \mathcal{C}_B(\mathbf{a}) \leq \Delta L - R_i \Delta T - \delta \tag{3.42}$$

$$\delta > 0 \tag{3.43}$$

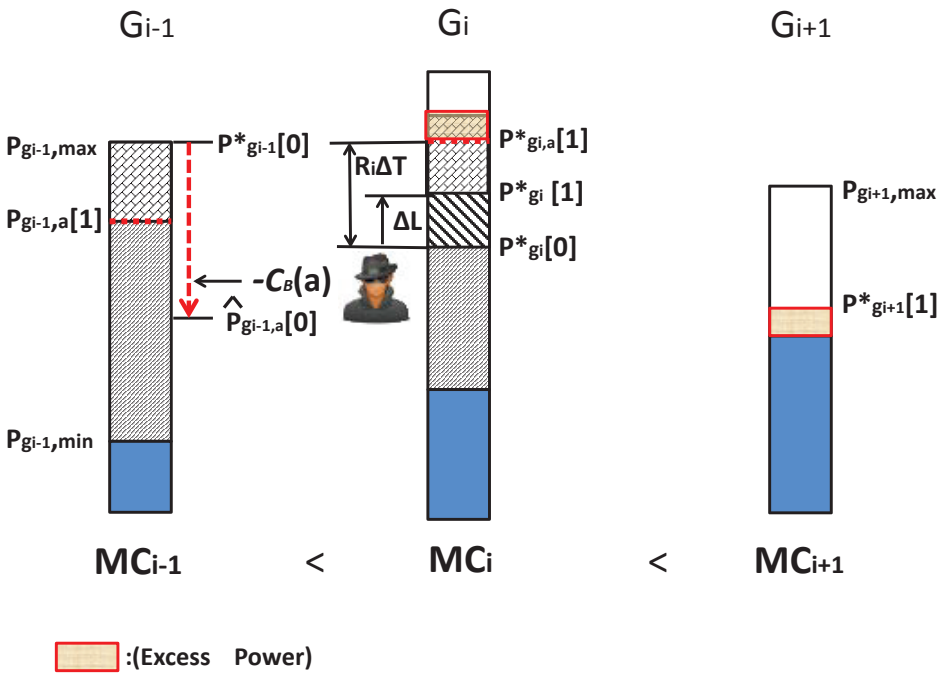


Figure 3.4: Conceptual diagrams illustrating a binding unit attack.

where

$$\mathcal{C}_M(\mathbf{a}) = \mathbf{B}_i\mathbf{a}, \quad \mathcal{C}_B(\mathbf{a}) = \sum_{j \in \underline{G}_M^c} [\mathbf{B}_j\mathbf{a} + R_j\Delta T].$$

$\mathcal{C}_M(\mathbf{a})$ and $\mathcal{C}_B(\mathbf{a})$ are the contributions of the attacker to changing the nodal price, corresponding to the marginal unit and binding unit attacks, respectively. In order to derive these two types of the attack contribution terms, we define the contributions of the marginal unit and binding unit attacks in the expected sense as

$$\mathcal{C}_M(\mathbf{a}) = E[d_i^M(\mathbf{a})] \tag{3.44}$$

$$\mathcal{C}_B(\mathbf{a}) = E[d^B(\mathbf{a})] \tag{3.45}$$

where

$$d_i^{(M)}(\mathbf{a}) = \hat{P}_{g_{i,\mathbf{a}}}[0] - P_{g_i}^*[0] \tag{3.46}$$

$$d^{(B)}(\mathbf{a}) = \sum_{j \in \underline{G}_M^c} (\hat{P}_{g_{j,\mathbf{a}}}[0] + R_j\Delta T - P_{g_j}^{\max}[0]). \tag{3.47}$$

Here, $\hat{P}_{g_{i,\mathbf{a}}}[0]$ is the manipulated estimate of generation power at generation bus i. Then,

$$\begin{aligned} \mathcal{C}_M(\mathbf{a}) &= E[d_i^{(M)}(\mathbf{a})] \\ &= E[\hat{P}_{g_i,\mathbf{a}}[0]] - P_{g_i}^*[0] \\ &\overset{(a)}{=} E[\mathbf{B}_i(\mathbf{H}\mathbf{x} + \mathbf{e} + \mathbf{a})] - P_{g_i}^*[0] \\ &\overset{(b)}{=} \mathbf{B}_i\mathbf{a} \end{aligned} \tag{3.48}$$

where $\mathbf{B}_i$ is the row vector of matrix $\mathbf{B}$, which corresponds to the injection mea-

surement sensor of generator i. (a) follows from $\hat{P}_{g_i,\mathbf{a}}[0] = \mathbf{B}_i\mathbf{z}$. (b) follows from $\mathbf{B}_i\mathbf{H}$=[0 . . . 0 1 0 . . . 0] where 1 is the ith element of vector $\mathbf{B}_i\mathbf{H}$ and $E[\mathbf{x}_i] \approx P^*_{g_i}[0]$ together with $E[\mathbf{e}] = 0$. Similarly,

$$\begin{aligned}
\mathcal{C}_B(\mathbf{a}) &= E[d^{(B)}(\mathbf{a})] \\
&= \sum_{j \in \underline{G}^c_M} [E[\hat{P}_{g_{j,\mathbf{a}}}[0]] + R_j\Delta T - P^{\max}_{g_j}[0]] \\
&= \sum_{j \in \underline{G}^c_M} [\mathbf{B}_j\mathbf{a} + P^*_{g_j}[0] + R_j\Delta T - P^{\max}_{g_j}[0]] \\
&\overset{(c)}{=} \sum_{j \in \underline{G}^c_M} [\mathbf{B}_j\mathbf{a} + R_j\Delta T]
\end{aligned} \tag{3.49}$$

where (c) follows from $P^*_{g_j}[0] = P^{\max}_{g_j}[0]$.

The set $\mathcal{A}$ represents the attack vector space, which describes the attack pattern related to the type and number of compromised sensors. $\Delta L - R_i\Delta T$ is the minimum amount of power which the attacker should reduce at $k = 0$ in order to withhold the capacity of unit i at $k = 1$. Constraint (3.41) assures undetectability as the parameter ϵ is tuned with an appropriate value. Constraint (3.42) assures profitability since it enables unit i to bind at the limit of the up-ramp constraint, leading to the increasing nodal price. Therefore, the attacker aims to maximize the margin δ in order to make a financial gain via capacity withholding with a high probability. The binary values of α and β in (3.42) determine the following three types of attacks:

1. $\alpha = 1$, $\beta = 0$: Marginal unit attack

2. $\alpha = 0$, $\beta = 1$: Binding unit attack

3. $\alpha = 1$, $\beta = 1$: Coordinated attack.

This optimization problem has the number of decision variables equal to the number of compromised sensors and *only* three linear inequality constraints. In

the proposed attack model, the number of decision variables (i.e., the number of compromised sensors) would not increase significantly. This results from the fact that the attacker wishes to compromise less sensors for reducing his effort and is sometimes incapable of compromising most sensors if they are highly protected against the cyber attacks. Therefore, with a modest number of decision variables and inequality constraints the computational complexity of this problem is quite manageable. In particular, since the proposed attack strategy is formulated in Linear Programming (LP), it can be solved efficiently using off-the-shelf LP method even in a large-scale system.

Remark 3. *Compared to the capacity withholding mentioned above, capacity withdrawing can benefit a load serving entity (LSE) by manipulating the down-ramp constraint limit. This type of the attack is feasible when constraint* (3.42) *is replaced with*

$$\alpha \mathcal{C}_M(\mathbf{a}) + \beta \mathcal{C}_B(\mathbf{a}) \geq \Delta L + R_i \Delta T + \delta \tag{3.50}$$

where

$$\mathcal{C}_M(\mathbf{a}) = \mathbf{B}_i \mathbf{a}, \quad \mathcal{C}_B(\mathbf{a}) = \sum_{j \in \overline{G}_M^c} [\mathbf{B}_j \mathbf{a} - R_j \Delta T].$$

Remark 4. *Table 3.2 summarizes the characteristics of the RID attack, as well as the spatial attack proposed in [24]. Specifically, we note the vulnerability index. This quantifies the vulnerability of the targeted power system subject to each type of attack. If variables* ΔL *and* F_l^{ante} *(power flow at the Ex-ante market) become closer to constants* $R_i \Delta T$ *and* F_l^{max}*, respectively, the power system becomes more and more vulnerable to both attacks.*

Table 3.2: Comparison between RID attack and spatial attack.

	RID Attack	Spatial Attack
Potential Attacker	Generation Company	Third Party
Bidding Method	Generation Bidding	Virtual Bidding
Market Structure	RT (Time-coupled only)	DA/RT
RT Pricing Model	Ex-ante	Ex-post
Line Congestion	No	Yes
Vulnerability Index	$\Delta L - R_i \Delta T$	$F_l^{\max} - F_l^{\text{ante}}$
Target Sensors	Injection Sensors	Flow Sensors

* RT: Real-time, DA: Day-ahead

3.5.3 Attack Performance Metrics

The performance of the proposed RID attack is evaluated using the following performance metrics:

3.5.3.1 Attack Profitability

Assuming that the power injection measurement sensor at generator i is compromised, we define the attack profit efficiency (PE) of generator i as the ratio of the profit with attack to without attack:

$$\text{PE}(i) = \frac{P^*_{g_i,\mathbf{a}}[1](\lambda_i^{(a)} - c_i)}{P^*_{g_i}[1](\lambda_i^{(b)} - c_i)} \times 100\ (\%). \tag{3.51}$$

Here, $\left(\lambda_i^{(a)}, P^*_{g_i,\mathbf{a}}[1]\right)$ and $\left(\lambda_i^{(b)}, P^*_{g_i}[1]\right)$ are two pairs of the nodal price and optimal generation dispatch with and without attack, respectively. c_i is the marginal cost for generator i.

3.5.3.2 Attack Undetectability

The system operator normally performs the Chi-squares test for detecting bad data in the measurements. Bad (or malicious) data will bypass if

$$J(\hat{\mathbf{x}}) \leq \chi^2_{(m-s),p} := \eta_\chi, \tag{3.52}$$

where p is the detection confidence probability, and m and s represent the number of measurements and state variables, respectively.

3.5.3.3 Attack Vulnerability

Since the measurement noise follows a Gaussian distribution, the manipulated estimate of the state at generator i is also a Gaussian random variable

$$\hat{\mathbf{x}}_\mathbf{i}(\mathbf{z_a}) \sim \mathcal{N}(\mathbf{P}^*_\mathbf{i}[\mathbf{0}] + \mathbf{B}_i\mathbf{a}, \mathbf{B}_i\mathbf{R}\mathbf{B}_i^T). \tag{3.53}$$

The probability of the distorted estimate $\hat{\mathbf{x}}_\mathbf{i}(\mathbf{z_a})$ being within the price-invulnerable region defined in Remark 1 is expressed as in terms of $Q(\cdot)$ functions

$$\begin{aligned} \mathbb{P}_i(\mathbf{a}) &= \mathbb{P}(\mathbf{l}(i) \leq \hat{\mathbf{x}}_\mathbf{i}(\mathbf{z_a}) \leq \mathbf{u}(i)) \\ &= Q(\mathbf{l}(i)) - Q(\mathbf{u}(i)), \end{aligned} \tag{3.54}$$

where the complementary Gaussian cumulative distribution function $Q(x)$ is defined as

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\xi^2}{2}\right) d\xi \tag{3.55}$$

and

$$\mathbf{l}(i) = \frac{\Delta L - R_i \Delta T - \mathbf{B}_i \mathbf{a}}{\sqrt{\mathbf{B}_i \mathbf{R} \mathbf{B}_i^T}} \tag{3.56}$$

$$\mathbf{u}(i) = \frac{\Delta L + R_i \Delta T - \mathbf{B}_i \mathbf{a}}{\sqrt{\mathbf{B}_i \mathbf{R} \mathbf{B}_i^T}}. \tag{3.57}$$

We define $\mathbb{P}_i(\mathbf{a})$ as the price-invulnerable probability (PIP) with respect to generator i. From (3.54), (3.55), (3.56) and (3.57), we specify the relationship among the ramp rate $R_i \Delta T$, the diagonal measurement covariance matrix $\mathbf{R}$, and the PIP as follows:

1. The increase of the $R_i \Delta T$ leads to the increase of the PIP.

2. The decrease of the values of the diagonal elements in $\mathbf{R}$ leads to the increase of the PIP.

In other words, the deployment of more accurate sensors and generators with a faster ramp rate enables the power system to become more robust to the RID attack.

3.6 Numerical Example

In this section the economic impact of the proposed RID attack on the real-time electricity market operation is illustrated in the IEEE 14-bus system as shown in Fig. 3.5. Measurement configuration includes nodal power injection measurements at all generation and load buses, and power flow measurements at one end of each transmission line. This system has a total of 34 measurements including 14 power injection and 20 power flow measurements, which assure the system observability. Table 3.3 shows the five generators' operating characteristics, including unit type (generation bus number), physical capacity limit, ramp rate and marginal cost (MC).

In a decongested power network where generators are set up with their marginal costs and ramp rates, only the amount of changing system load between two consec-

utive dispatch intervals (ΔL) affects the proposed attack performance. Other factors such as the size of network as well as the number and locations of generators do not influence the feasibility of the attack. Therefore, our case study in the IEEE-14 bus system is good enough to evaluate the performance of the proposed attacks.

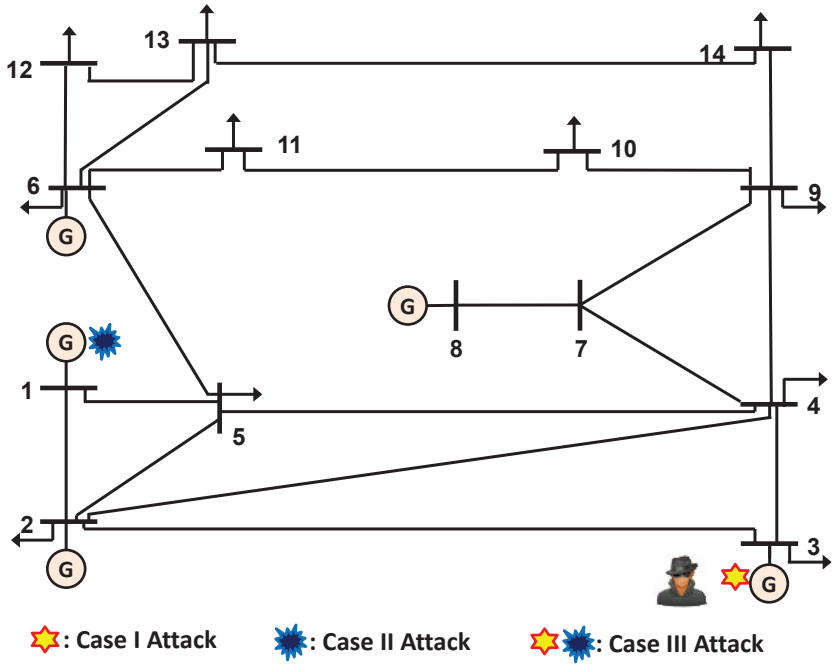


Figure 3.5: IEEE 14-bus test system with three attack cases.

In this section, three cases are simulated in the IEEE-14 bus system:

- Case I: Marginal unit attack.
- Case II: Binding unit attack.
- Case III: Coordinated attack.

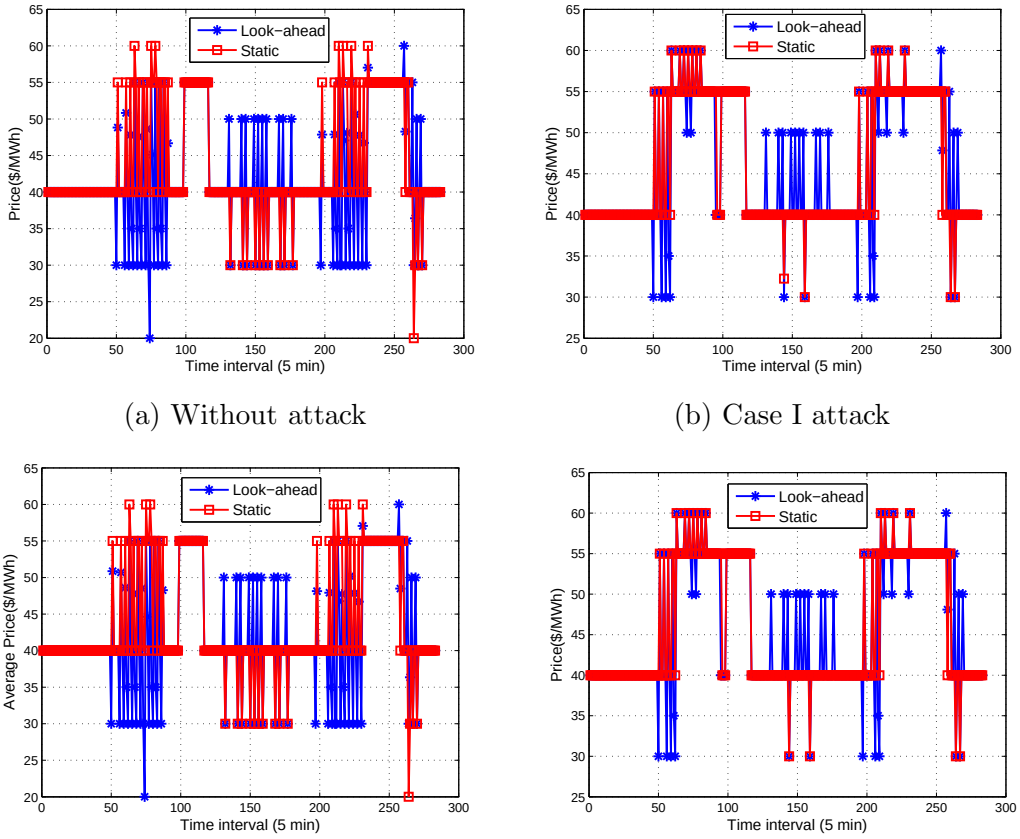
The performance of the proposed RID attack is evaluated based on the one day load profile with a 5-min resolution. This load profile is obtained by interpolating a

Table 3.3: Generator parameters of the IEEE 14-bus test system.

Unit Type	$P_{\min}$	$P_{\max}$	Ramp Rate	MC
Coal(1)	0MW	200MW	10MW/5min	30$/MWh
Wind(2)	0MW	300MW	150MW/5min	20$/MWh
Nuclear(3)	0MW	300MW	8MW/5min	40$/MWh
Coal(6)	50MW	250MW	15 MW/5min	55$/MWh
Oil(8)	60MW	150MW	60 MW/5min	60$/MWh

15-min daily data in the ERCOT website. The load is scaled down to be consistent with the IEEE 14-bus test system's peak load data. The common goal of all three cases is to withhold the capacity of generator 3 for the purpose of making a profit. A power injection sensor at generation bus 3 is compromised in Case I whereas a power injection sensor at generation 1 is compromised in Case II. Case III represents the coordinated attack, which compromises both sensors targeted in Case I and Case II.

Figs. 3.6 show the comparison of the LMPs between static ($K = 1$) and look-ahead dispatch ($K = 6$) without attack and with attack in Cases I, II and III. Due to no network transmission congestion, the prices in these figures denote the uniform LMPs for all the buses at every dispatch interval. In Fig. 3.6(a), the LMPs in look-ahead dispatch are oscillating around 40$/MWh more than the ones in static dispatch. This phenomenon is due to the fact that the binding of generator 3 at the up- or down-ramp constraints at time $k+1$ makes its corresponding Lagrangian multiplier, $\omega_{3,\max}[k+1]$ or $\omega_{3,\min}[k+1]$, become positive. As shown in equation (3.11), this leads to different LMPs at time k than the ones from static dispatch. We observe from Figs. 3.6(b),(c),(d) that the LMPs in both dispatch models tend to increase with attack. This observation implies that the attacker successfully withholds the capacity of generator 3 by lowering its up-ramp constraint limit through the reduction of the



(a) Without attack

(b) Case I attack

(c) Case II attack

(d) Case III attack

Figure 3.6: LMP of static and look-ahead dispatch without attack and with Case I,II and III attacks.

initial estimate $\hat{P}_{g_3,\mathbf{a}}[0]$. Consequently, this leads to the shift of the marginal unit to another one with a more expensive marginal cost.

Table 3.4: Attack performance in static and look-ahead dispatch.

Case	Static (PE(3))	Look-ahead (PE(3))	$J(\eta_\chi = 37.6)$
I	131.9	148.9	28.2
II	101.2	102.6	35.5
III	108.9	113.8	31.5

Table 3.5: Impact of ramp rate and measurement variance on the attack performance in Case I.

	Ramp Rate (MW/5min)				Measurement Variance (σ^2)			
	8	10	12	14	0.0005	0.005	0.05	0.5
Static (PE(3))	131.9	119.7	106.4	100.5	123.2	129.1	130.3	136.9
Look-ahead (PE(3))	148.9	123.5	108.5	103.1	143.5	144.75	146.1	152.8
PIP	0.017	0.021	0.037	0.044	0.056	0.041	0.034	0.021

Table 3.4 shows the attack performance of Cases I, II and III in both static and look-ahead dispatch. The second and third columns of this table indicate the attack profit efficiency at generation bus 3. We can observe from the comparison of these two columns several facts. First, the PE values in all three cases of both dispatch models are larger than 100. It indicates that the attacker makes an additional profit using the proposed attack strategy. Second, for all three cases, the PE in look-ahead dispatch is higher than in static dispatch. This observation might result from the fact

that the attack leads to more increase of the nodal price in look-ahead dispatch than in static dispatch. Lastly, among three cases, Case I and Case II attacks yield the largest and smallest PE, respectively. The PE in Case III is between Case I and Case II. This result is natural since Case II and Case III attacks require an extra effort for withholding the binding unit's capacity as well as the marginal unit's capacity so that both attacks fail with a higher probability than Case I attack.

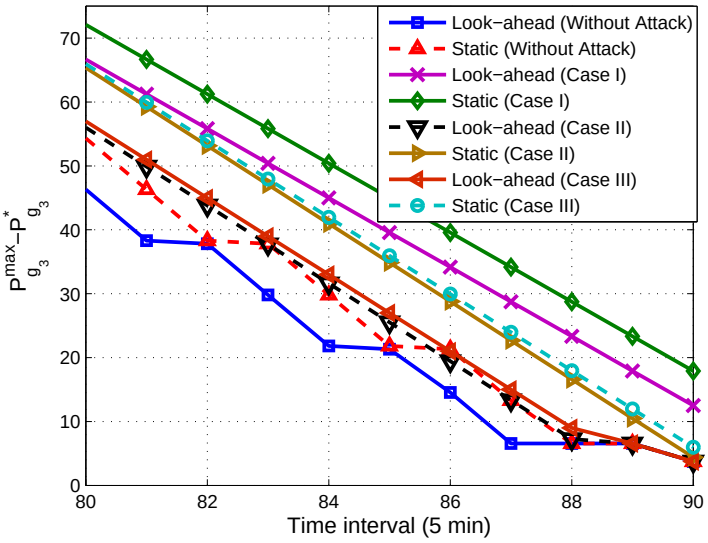


Figure 3.7: $P_{g3}^{\max} - P_{g3}^{*}$ of static and look-ahead dispatch without attack and with Case I,II and III attacks.

Fig. 3.7 shows the amount of generator 3's capacity which all three attacks withhold between 80 and 90 time intervals. As expected, it is verified that Case I, Case III, and Case II attacks withhold capacity the most in a descending order. This fact also justifies the third observation mentioned above. The values of the estimated objective functions for all three cases are shown in the last column of Table 3.4. Based

on the measurement configuration with m=34 and s=14, the threshold (η_χ) of the Chi-squares test with a 99% confidence level is set to 37.6. For undetectability, the parameter ϵ in (3.41) is set to 3. Therefore, all three attacks in both dispatch models succeed in avoiding the Chi-squares bad data detection.

Table 3.5 shows Case I attack performance with the varying ramp rate of generator 3 and measurement variance of sensors. We first observe from this table that as the ramp rate of generator 3 increases the PE in both dispatch models decreases. Another observation is that the decrease of measurement variance leads to the decrease of the attack profit. These observations imply that the nodal prices become less manipulable, which is verified with the increasing PIP in Table 3.5.

Table 3.6: Attack performance with varying attack magnitude in Case I.

	Attack Relative Magnitude (ARM %)			
	0.25	0.5	0.75	1
Static (PE(3))	111.8	120.8	126.4	126.9
Look-ahead (PE(3))	112.2	125.8	127.6	137.7
J	21.1	25.4	29.2	33.1
PIP	0.433	0.344	0.259	0.188

Table 3.6 shows the sensitivity of Case I attack performance with respect to the attack magnitude. In this table, the attack relative magnitude (ARM) is defined as $\frac{||\mathbf{a}||_\infty}{||\mathbf{z}||_\infty} \times 100$ where $||\cdot||_\infty$ denotes an infinity norm. We observe from this table that the increase of the ARM leads to more profit (the third and fourth rows) in both dispatch models. However, the estimated objective function J (the fifth row) used for the Chi-squares bad data test increases and the PIP (the last row) decreases. This implies that as the ARM increases the attack becomes more vulnerable to the

bad data detection and fails with an increasing probability.

4. LMP SENSITIVITY ANALYSIS TO DATA CORRUPTION-INDUCED ESTIMATION ERROR*

4.1 Introduction

State estimation is one of the key applications for power system energy management systems (EMSs). The impact of bad data on power systems has been intensively investigated in recent decades in power system state estimation literature. Measurement noise and/or manipulated sensor errors in a supervisory control and data acquisition (SCADA) system may mislead system operators about real-time conditions in a power system, which in turn may impact the price signals in real-time power markets. This section attempts to provide a novel analytical framework with which to investigate the impact of bad sensor data on electric *power market* operations. In future power system operations, which will probably involve many more sensors, the impact of sensor data quality on grid operations will become increasingly important.

Locational marginal price (LMP) is the core variable in market operations [74]. In real-time power markets, LMP is obtained as the by-product of security constrained economic dispatch (SCED) in either of the two main pricing models: Ex-ante (e.g. in ERCOT, NY ISO) and Ex-post (e.g. in ISO New England, PJM, and Midwest ISO). Both pricing models are built on the power flow and network topology results given by the state estimator, which uses two types of sensor data: 1) continuous (e.g., the power injection/flow and voltage magnitude); and 2) discrete (e.g., the on/off status of a circuit breaker).

*

Section 4 focuses on the impact of both continuous and discrete sensor data on real-time LMP. Fig. 4.1 illustrates how the corrupted SCADA sensor data impact the state estimation as well as the security constrained economic dispatch in energy management systems (EMSs) and market management systems (MMSs). The two lines **(a)** and **(b)** in Fig. 4.1 represent the flow of manipulated network topology estimate and power flow estimate, corresponding to the corruption of discrete and continuous data, respectively.

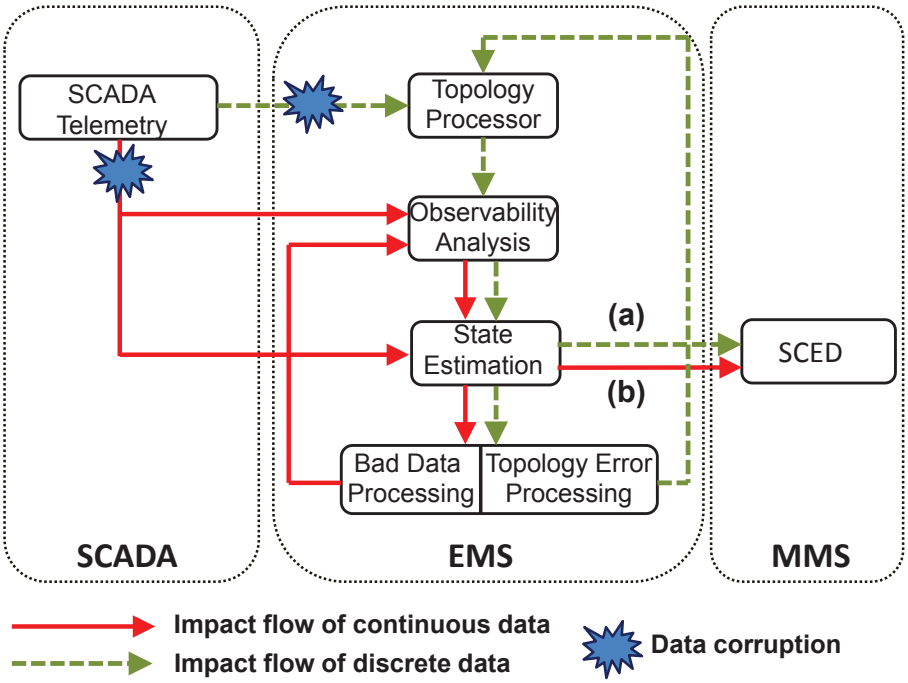


Figure 4.1: Illustrating the impact of corrupted continuous and discrete SCADA sensor data on state estimation and SCED.

The main part of Section 4 is categorized into two sections. In Section 4.3, we conduct the impact analysis of real-time LMP subject to power flow estimation error due to continuous data corruption (dotted line (b) in Fig. 4.1). In Sections 4.4, we

focus on investigating the impact of undetectable circuit breaker-induced network topology errors on real-time LMP (dotted line (a) in Fig. 4.1).

Section 4.3 is motivated by a desire to study the effect of *continuous* data corruption on LMP via *state estimation* and provides an analytical framework for answering the following questions:

1. How much does LMP change at every bus given a set of SCADA measurements with corrupted continuous data?

2. What is the impact of continuous data accuracy on LMP sensitivity at each bus?

Here, data corruption refers to both natural noise and man-made attacks. In this section, we focus on a sensitivity analysis of real-time LMP in ex-ante and ex-post pricing models subject to corrupted continuous data fed into the state estimator. Fig. 4.2 illustrates that via state estimation, SCADA measurement $\mathbf{z}$ may impact the results of a pair of Ex-ante nodal price and optimal generation dispatch $\{\boldsymbol{\pi}\left(\hat{\mathbf{x}}_A(\mathbf{z})\right), P_g^*\left(\hat{\mathbf{x}}_A(\mathbf{z})\right)\}$ and the Ex-post price $\boldsymbol{\pi}\left(\hat{\mathbf{x}}_P(\mathbf{z})\right)$.

The novelty aspect of Section 4.3 is that it provides system operators with an analytical tool for assessing the financial risks of bad/malicious data in light of secure market operations. To this end, a unified LMP sensitivity matrix subject to data corruption is developed, describing the coupling between LMP, the estimation of power system states, and the sensor data. This matrix offers system operators an *online* tool to: 1) quantify the impact of corrupted data at any sensor on LMP variation at any bus; 2) identify buses with LMP highly sensitive to data corruption; 3) find significant and influential sensors with regards to LMP change; and 4) study the effect of data accuracy on LMP sensitivity.

In Section 4.4, we derive a mathematical LMP sensitivity index that illustrates

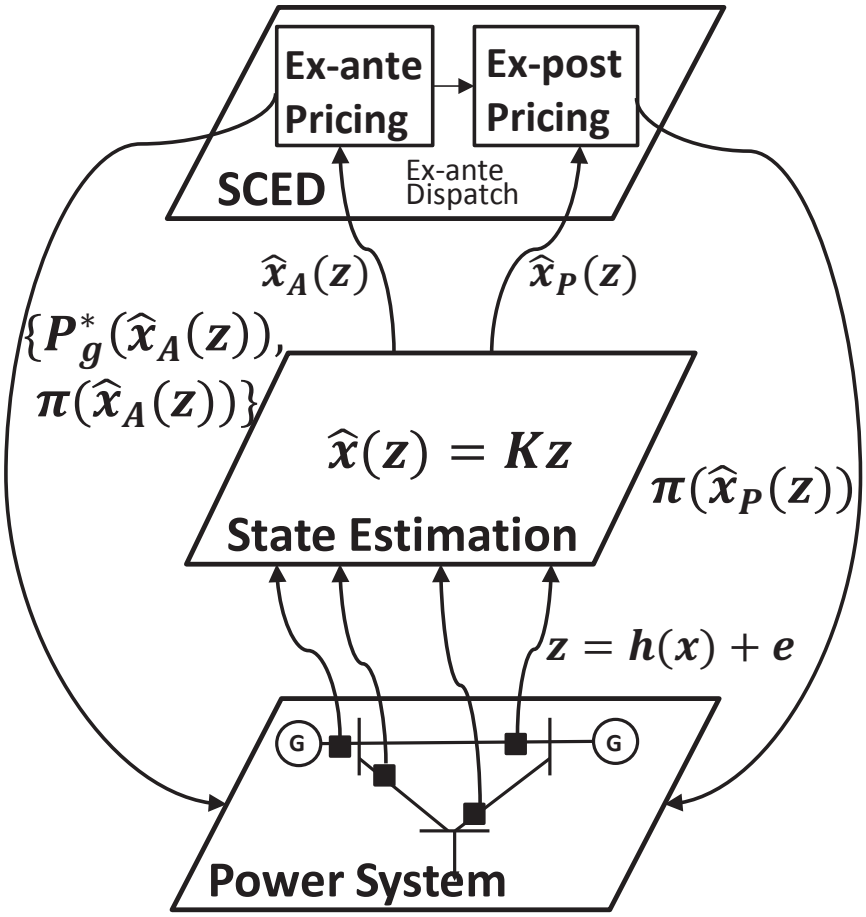


Figure 4.2: A three-layered framework illustrating the coupling of the physical power system, state estimation, and SCED.

the impact of power system topology error on LMP change. We assume that the corruption in discrete data is not detected by topology error processing, thus leading to *undetectable* topology change. Similar to the novelty in Section 4.3, the derived LMP sensitivity index in Section 4.4 provides system operators with an analytical tool to identify the most economically impacting transmission line/circuit breakers on LMP as well as find the economically sensitive bus to topology error.

The remainder of this section is organized as follows. We briefly review AC state estimation and two representative real-time pricing models in Section 4.2. In Section 4.3, we formulate the problem of LMP sensitivity analysis subject to continuous data corruption, and derive such quantifying sensitivity metric. Using the derived metric, we present numerical examples that illustrate the impact of different types of SCADA sensors on LMP in IEEE 14-bus and 118-bus systems with both the Ex-ante and Ex-post pricing models. Section 4.4 provides a LMP sensitivity index with respect to network topology error. The derived sensitivity index is verified and illustrated in the IEEE 14-bus system.

4.2 Preliminaries

The main notations used throughout this section are summarized in Table 4.1. Bold symbols represent vectors or matrices. Hat symbols represent estimates of true parameter value.

4.2.1 AC State Estimation Model

The measurement model for state estimation is formulated as

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{e}. \tag{4.1}$$

Table 4.1: Notations.

a_i	Linear cost coefficient for generator i
b_i	Quadratic cost coefficient for generator i
$C_i(\cdot)$	Energy cost for generator i
P_{g_i}	Scheduled generator power output for generator i
L_{d_i}	Fixed demand at bus i
$P_{q_i}^{\min}, P_{q_i}^{\max}$	Min/max generation limits for generator i at Ex-ante dispatch
$F_l^{\min}, F_l^{\max}$	Min/max flow limits for transmission line l at Ex-ante dispatch
S_{li}	Generation shift factor of transmission line l to bus i
$\Delta P_{q_i}^{\max}, \Delta P_{q_i}^{\min}$	Min/max incremental generation limits for generator i at Ex-post dispatch
R_i	Ramp rate of generator i
ΔT	Dispatch interval
π_i	Locational marginal price at bus i
λ	Shadow price of the system energy balance equation
τ_i	Shadow price of the capacity constraint for generator i
μ_l	Shadow price of the transmission line constraint for transmission line l
N_b	Total number of buses
N_m	Total number of sensor measurements
N_l	Total number of transmission lines
$\mathcal{CL}_+, \mathcal{CL}_-$	Sets of positively and negatively congested lines at Ex-ante dispatch
$\mathcal{S}_v$	Set of voltage magnitude measurements
$\mathcal{S}_{ri}$	Set of real power injection measurements
$\mathcal{S}_{ai}$	Set of reactive power injection measurements
$\mathcal{S}_{rf}$	Set of real power flow measurements
$\mathcal{S}_{af}$	Set of reactive power flow measurements
$\mathbf{I}_k$	$k \times k$ identity matrix
$\mathbf{1}_k, \mathbf{0}_k$	$k \times 1$ column vectors with all ones and all zeros, respectively

Here $\mathbf{z} = [\mathbf{z}_r^T \ \mathbf{z}_a^T \ \mathbf{z}_v^T]^T$ is the $N_m \times 1$ measurement vector that consists of real power injection and the flow vector $\mathbf{z}_r = [\mathbf{z}_{ri}^T \ \mathbf{z}_{rf}^T]^T$, the reactive power injection and flow vector $\mathbf{z}_a = [\mathbf{z}_{ai}^T \ \mathbf{z}_{af}^T]^T$, and the bus voltage magnitude vector $\mathbf{z}_v$. $\mathbf{x} = [\boldsymbol{\theta}^T \ \mathbf{V}^T]^T$ is the state vector that consists of the $(N_b - 1) \times 1$ bus voltage phase angle vector $\boldsymbol{\theta}$ excluding a slack bus and the $N_b \times 1$ voltage magnitude vector $\mathbf{V}$. $\mathbf{h}(\mathbf{x})$ is the $N_m \times 1$ nonlinear vector valued measurement function relating measurements to states, and $\mathbf{e}$ is the $N_m \times 1$ independent identically distributed (i.i.d.) Gaussian measurement error vector with zero mean and diagonal covariance matrix $\mathbf{R}$. The state estimator computes the optimal estimate of $\mathbf{x}$ by minimizing the weighted least squares of measurement error:

$$\text{minimize } J(\mathbf{x}) = \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} \tag{4.2}$$

$$\text{s.t. } \mathbf{r} = \mathbf{z} - \mathbf{h}(\mathbf{x}). \tag{4.3}$$

Using the Gauss-Newton method, the weighted least squares estimate vector $\hat{\mathbf{x}}$ is computed by the following iterative procedure [43]:

$$\Delta \hat{\mathbf{x}}^{k+1} = [\mathbf{G}(\hat{\mathbf{x}}^k)]^{-1} \mathbf{H}^T(\hat{\mathbf{x}}^k) \mathbf{R}^{-1} \Delta \mathbf{z}^k \tag{4.4}$$

where $\mathbf{H}(\hat{\mathbf{x}}^k) = \left[\frac{\partial \mathbf{h}(\hat{\mathbf{x}}^k)}{\partial \hat{\mathbf{x}}^k}\right]$ is the $N_m \times (2N_b - 1)$ Jacobian matrix at k-th iteration, and

$$\Delta \hat{\mathbf{x}}^{k+1} = \hat{\mathbf{x}}^{k+1} - \hat{\mathbf{x}}^k \tag{4.5}$$

$$\Delta \mathbf{z}^k = \mathbf{z} - \mathbf{h}(\hat{\mathbf{x}}^k) \tag{4.6}$$

$$\mathbf{G}(\hat{\mathbf{x}}^k) = \mathbf{H}^T(\hat{\mathbf{x}}^k) \mathbf{R}^{-1} \mathbf{H}(\hat{\mathbf{x}}^k). \tag{4.7}$$

The iteration process in (4.4) continues until the maximum of $|\Delta\hat{\mathbf{x}}^k|$ is less than a predetermined threshold, otherwise stops and yields the ultimate estimates.

4.2.2 Real-Time Electricity Pricing Model

State estimation results impact real-time market operations. In this subsection, we present two real-time pricing models based on the DCOPF model. For simplicity, we assume that each bus has one generator and one load.

4.2.2.1 The Ex-ante Model

In ex-ante real-time market models, LMPs are computed before the actual deployment of dispatch orders. For the system operator, the Ex-ante dispatch is formulated as follows:

$$\min_{P_{g_i}} \sum_{i=1}^{N_b} C_i(P_{g_i}) \tag{4.8}$$

s.t.

$$\lambda : \sum_{i=1}^{N_b} P_{g_i} = \sum_{i=1}^{N_b} L_{d_i} \tag{4.9}$$

$$\boldsymbol{\tau} : \hat{P}_{g_i}^{\min} \leq P_{g_i} \leq \hat{P}_{g_i}^{\max} \qquad \forall i = 1, \ldots, N_b \tag{4.10}$$

$$\boldsymbol{\mu} : F_l^{\min} \leq \sum_{i=1}^{N_b} S_{li}(P_{g_i} - L_{d_i}) \leq F_l^{\max} \; \forall l = 1, \ldots, N_l \tag{4.11}$$

where

$$\hat{P}_{g_i}^{\max} = \min\{P_{g_i}^{\max}, \hat{P}_{g_i}(\mathbf{z}) + R_i \Delta T\}$$
$$\hat{P}_{g_i}^{\min} = \max\{P_{g_i}^{\min}, \hat{P}_{g_i}(\mathbf{z}) - R_i \Delta T\}.$$

In this formulation, the objective function is to minimize the total generation costs in (4.8). (4.9) is the system-wide energy balance equation. (4.10) is the physical capacity constraints of each generator embedded with its ramp constraints. (4.11) is the transmission line constraints.

4.2.2.2 The Ex-post Model

In ex-post real-time market models, LMPs are computed after the fact using real-time estimates for settlement purposes. Assuming no demand elasticity, the Ex-post dispatch is written as:

$$\min_{P_{g_i}} \sum_{i=1}^{N_b} C_i(P_{g_i}) \tag{4.12}$$

s.t.

$$\lambda : \sum_{i=1}^{N_b} P_{g_i} = \sum_{i=1}^{N_b} \hat{P}_{g_i}(\mathbf{z}) \tag{4.13}$$

$$\boldsymbol{\tau} : \hat{P}_{g_i}^{\min} \leq P_{g_i} \leq \hat{P}_{g_i}^{\max} \qquad \forall i = 1, \ldots, N_b \tag{4.14}$$

$$\boldsymbol{\mu}_{\max} : \sum_{i=1}^{N_b} S_{li}(P_{g_i} - L_{d_i}) \leq \hat{F}_l(\mathbf{z}) \quad \forall l \in \mathcal{CL}_+ \tag{4.15}$$

$$\boldsymbol{\mu}_{\min} : \sum_{i=1}^{N_b} S_{li}(P_{g_i} - L_{d_i}) \geq \hat{F}_l(\mathbf{z}) \quad \forall l \in \mathcal{CL}_- \tag{4.16}$$

where

$$\hat{P}_{g_i}^{\max} = \hat{P}_{g_i}(\mathbf{z}) + \Delta P_{g_i}^{\max}, \ \hat{P}_{g_i}^{\min} = \hat{P}_{g_i}(\mathbf{z}) + \Delta P_{g_i}^{\min}.$$

The above formulation is expressed with different notation than a general Ex-post model in order to emphasize that the state estimation solution has a direct impact on the Ex-post model.

4.3 Impact Analysis of LMP Subject to Power Flow Estimate Errors

4.3.1 Problem Formulation

For all buses ($i = 1, \ldots, N_b$) and measurements ($j = 1, \ldots, N_m$), the $N_b \times 1$ vector of LMPs can be expressed in a composite function form:

$$\mathbf{LMP} = \boldsymbol{\pi}(\hat{\mathbf{x}}(\mathbf{z}))$$

where

$$\boldsymbol{\pi} = [\pi_1, \pi_2, \ldots, \pi_{N_b}]^T \tag{4.17}$$

$$\pi_i = f_i(\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_{N_m}) \tag{4.18}$$

$$\hat{x}_j = g_j(z_1, z_2, \cdots, z_{N_m}). \tag{4.19}$$

π_i represents the LMP at bus i. z_j and $\hat{x}_j$ are the measurement and its corresponding estimate at sensor j, respectively. $f_i(\cdot)$ is the vector function that describes the relationship between any estimate and LMP at bus i. $g_j(\cdot)$ is the vector function that describes the relationship between any measurement and estimate at sensor j.

The primary goal of this section is to compute LMP sensitivity at any bus i subject to a measurement change at any sensor j throughout the entire transmission network.

$$\frac{\partial \pi_i}{\partial z_j} = \mathbf{\Lambda}_{(i,j)}. \tag{4.20}$$

By chain rule, for all i and j, (4.20) is written as

$$\frac{\partial \pi_i}{\partial z_j} = \frac{\partial \pi_i}{\partial \hat{x}_1}\frac{\partial \hat{x}_1}{\partial z_j} + \frac{\partial \pi_i}{\partial \hat{x}_2}\frac{\partial \hat{x}_2}{\partial z_j} + \cdots + \frac{\partial \pi_i}{\partial \hat{x}_{N_m}}\frac{\partial \hat{x}_{N_m}}{\partial z_j}. \tag{4.21}$$

In (4.21), the estimate $\hat{x}_j$ is chosen as an intermediate variable for computing the partial derivative of π_i with respect to z_j. This variable is used to set the bounds for: 1) minimum and maximum generation capacity in (4.10), (4.14); 2) the system balance equation in (4.13); and 3) the positive and negative transmission line capacity in (4.15), (4.16). Equation (4.21) can be expressed in matrix form as shown in (4.22).

$$\begin{aligned}\boldsymbol{\Lambda}_{(N_b\times N_m)} &= \frac{\partial \boldsymbol{\pi}}{\partial \mathbf{z}} \\ &= \frac{\partial \boldsymbol{\pi}}{\partial \hat{\mathbf{x}}}\frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{z}} = \underbrace{\begin{bmatrix} \frac{\partial \pi_1}{\partial \hat{x}_1} & \frac{\partial \pi_1}{\partial \hat{x}_2} & \cdots & \frac{\partial \pi_1}{\partial \hat{x}_{N_m}} \\ \frac{\partial \pi_2}{\partial \hat{x}_1} & \frac{\partial \pi_2}{\partial \hat{x}_2} & \cdots & \frac{\partial \pi_2}{\partial \hat{x}_{N_m}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \pi_{N_b}}{\partial \hat{x}_1} & \frac{\partial \pi_{N_b}}{\partial \hat{x}_2} & \cdots & \frac{\partial \pi_{N_b}}{\partial \hat{x}_{N_m}} \end{bmatrix}}_{\boldsymbol{\Lambda}_A} \underbrace{\begin{bmatrix} \frac{\partial \hat{x}_1}{\partial z_1} & \frac{\partial \hat{x}_1}{\partial z_2} & \cdots & \frac{\partial \hat{x}_1}{\partial z_{N_m}} \\ \frac{\partial \hat{x}_2}{\partial z_1} & \frac{\partial \hat{x}_2}{\partial z_2} & \cdots & \frac{\partial \hat{x}_2}{\partial z_{N_m}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \hat{x}_{N_m}}{\partial z_1} & \frac{\partial \hat{x}_{N_m}}{\partial z_2} & \cdots & \frac{\partial \hat{x}_{N_m}}{\partial z_{N_m}} \end{bmatrix}}_{\boldsymbol{\Lambda}_B}.\end{aligned} \tag{4.22}$$

The sensitivity $\boldsymbol{\Lambda}_{(i,j)}$ in (4.20) is the element at the ith row and jth column of the $N_b \times N_m$ sensitivity matrix $\boldsymbol{\Lambda}$. The matrix $\boldsymbol{\Lambda}$ is written as the multiplication form of two matrices with different types of sensitivities: the $N_b \times N_m$ matrix $\boldsymbol{\Lambda}_A = \frac{\partial \boldsymbol{\pi}}{\partial \hat{\mathbf{x}}}$ quantifies the *economic* impact of any estimate on any LMP, and the $N_m \times N_m$ matrix $\boldsymbol{\Lambda}_B = \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{z}}$ quantifies the *cyber* impact of any sensor measurement on any estimate. The derivations of $\boldsymbol{\Lambda}_A$ and $\boldsymbol{\Lambda}_B$ are described in more detail in the next subsection.

4.3.2 LMP Sensitivity to Continuous Sensor Data Corruption

4.3.2.1 Sensitivity of LMPs to Estimated States

We first derive the sensitivity matrix $\mathbf{\Lambda}_A$ using the Ex-ante model. To this end, the perturbation approach developed in [32] is applied to the Ex-ante model in Subsection 4.2.2. The Lagrangian function of the Ex-ante dispatch is written as

$$\begin{aligned}\mathcal{L} = &\sum_{i=1}^{N_b} C_i(P_{g_i}) - \lambda \left(\sum_{i=1}^{N_b} [P_{g_i} - L_{d_i}] \right) \\ &+ \sum_{j=1}^{2N_b} \tau_j \left(\sum_{i=1}^{N_b} A_{ji} P_{g_i} - \hat{C}_j \right) \\ &+ \sum_{l=1}^{2N_l} \mu_l \left(\sum_{i=1}^{N_b} S_{li} [P_{g_i} - L_{d_i}] - D_l \right)\end{aligned}$$

where A_{ji}, S_{li}, $\hat{C}_j$ and D_l are the elements of the following matrices

$$\mathbf{A}_{(2N_b \times N_b)} = \left[\; A_{ji} \; \right] = \begin{bmatrix} \mathbf{I}_{N_b} \\ -\mathbf{I}_{N_b} \end{bmatrix} \tag{4.23}$$

$$\mathbf{B}_{(2N_l \times N_b)} = \left[\; S_{li} \; \right] = \begin{bmatrix} \mathbf{S} \\ -\mathbf{S} \end{bmatrix} \tag{4.24}$$

$$\hat{\mathbf{C}}_{(2N_b \times 1)} = \left[\; \hat{C}_j \; \right] = \begin{bmatrix} \hat{\mathbf{P}}_g^{\max} \\ -\hat{\mathbf{P}}_g^{\min} \end{bmatrix} \tag{4.25}$$

$$\mathbf{D}_{(2N_l \times 1)} = \left[\; D_l \; \right] = \begin{bmatrix} \mathbf{F}^{\max} \\ -\mathbf{F}^{\min} \end{bmatrix}. \tag{4.26}$$

Here, **S** is the generation shift factor matrix, and

$$\hat{\mathbf{P}}_g^{\max(\min)} = [\hat{P}_{g_1}^{\max(\min)}, \ldots, \hat{P}_{g_{N_b}}^{\max(\min)}]^T \tag{4.27}$$

$$\mathbf{F}^{\max(\min)} = [F_1^{\max(\min)}, \ldots, F_{N_l}^{\max(\min)}]^T. \tag{4.28}$$

As in [32], unbinding inequality constraints are excluded in our sensitivity analysis. Let us define B_g and B_f as the number of binding constraints associated with generation capacity and line capacity, respectively. Then, the KKT conditions of the Ex-ante problem are written as

$$\text{(i)} \ \frac{\partial C_i(P_{g_i})}{\partial P_{g_i}} - \lambda + \sum_{j=1}^{B_g} \tau_j A_{ji} + \sum_{l=1}^{B_f} \mu_l S_{li} = 0 \qquad \forall i = 1, \ldots, N_b$$

$$\text{(ii)} \ \sum_{i=1}^{N_b} P_{g_i} = \sum_{i=1}^{N_b} L_{d_i}$$

$$\text{(iii)} \ \sum_{i=1}^{N_b} A_{ji} P_{g_i} = \hat{C}_j \qquad \forall j = 1, \ldots, B_g$$

$$\text{(iv)} \ \sum_{i=1}^{N_b} S_{li}[P_{g_i} - L_{d_i}] = D_l \qquad \forall l = 1, \ldots, B_f.$$

after which the above KKT equations are perturbed with respect to P_{g_i}, L_{d_i}, $\hat{C}_j$, λ, τ_j, and μ_j as follows:

$$\text{(i)} \ \underbrace{\frac{\partial}{\partial P_{g_i}}\left(\frac{\partial C_i(P_{g_i})}{\partial P_{g_i}}\right)}_{M_i} dP_{g_i} - d\lambda + \sum_{j=1}^{B_g} A_{ji} d\tau_j + \sum_{l=1}^{B_f} S_{li} d\mu_l = 0 \quad \forall i = 1, \ldots, N_b$$

$$\text{(ii)} \ \sum_{i=1}^{N_b} dP_{g_i} = \sum_{i=1}^{N_b} dL_{d_i}$$

$$\text{(iii)} \ \sum_{i=1}^{N_b} A_{ji} dP_{g_i} = d\hat{C}_j \qquad \forall j = 1, \ldots, B_g$$

$$\text{(iv)} \ \sum_{i=1}^{N_b} S_{li} dP_{g_i} = \sum_{i=1}^{N_b} S_{li} dL_{d_i} \qquad \forall l = 1, \ldots, B_f.$$

It should be noted that the variables D_l, A_{ji}, and S_{li} in the KKT equations are not perturbed. This is due to the fact that 1) the limits of line flow constraint limits in the Ex-ante model are not updated by the state estimator, and 2) the network topology is not affected by corrupted analog data. These perturbation equations can be expressed in matrix form:

$$\underbrace{\begin{bmatrix} \mathbf{M} & -\mathbf{1}_{N_b} & \Upsilon \\ \mathbf{1}_{N_b}^T & 0 & \mathbf{0} \\ \Upsilon^T & \mathbf{0} & \mathbf{0} \end{bmatrix}}_{\Xi} \begin{bmatrix} d\mathbf{P}_g \\ d\lambda \\ d\boldsymbol{\tau}_s \\ d\boldsymbol{\mu}_s \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{U_1}^T & \mathbf{U_2}^T \end{bmatrix}}_{\Phi} \begin{bmatrix} d\mathbf{L}_d \\ d\hat{\mathbf{C}}_s \end{bmatrix} \tag{4.29}$$

where

$$\mathbf{M}_{(N_b \times N_b)} = \text{diag}(M_1, \ldots, M_{N_b}) \tag{4.30}$$

$$\Upsilon_{(N_b \times [B_g + B_f])} = \begin{bmatrix} \mathbf{A}_s^T & \mathbf{B}_s^T \end{bmatrix} \tag{4.31}$$

$$\mathbf{U_1}_{(N_b \times [N_b + 1 + B_g + B_f])} = \begin{bmatrix} \mathbf{0} & \mathbf{1}_{N_b} & \mathbf{0} & \mathbf{B}_s^T \end{bmatrix} \tag{4.32}$$

$$\mathbf{U_2}_{(B_g \times [N_b + 1 + B_g + B_f])} = \begin{bmatrix} \mathbf{0} & \mathbf{0}_{B_g} & \mathbf{I}_{B_g} & \mathbf{0} \end{bmatrix}. \tag{4.33}$$

Taking the inverse of $\boldsymbol{\Xi}$ on both sides of (4.29),

$$\begin{bmatrix} d\mathbf{P}_g \\ d\lambda \\ d\boldsymbol{\tau}_s \\ d\boldsymbol{\mu}_s \end{bmatrix} = \underbrace{\boldsymbol{\Xi}^{-1}\boldsymbol{\Phi}}_{\boldsymbol{\Lambda}_p} \begin{bmatrix} d\mathbf{L}_d \\ d\hat{\mathbf{C}}_s \end{bmatrix}. \tag{4.34}$$

The subscript s of the variables in (4.29), (4.31), and (4.32) represents the subvector (submatrix) of the original vector (matrix) that corresponds to the binding constraints. The matrix $\boldsymbol{\Lambda}_p$ in (4.34) is partitioned into two sensitivity matrices—$\boldsymbol{\Lambda}_{\mathbf{L}_d}$ and $\boldsymbol{\Lambda}_{\hat{\mathbf{C}}_s}$:

$$\boldsymbol{\Lambda}_p = \left[\begin{array}{c|c} \boldsymbol{\Lambda}_{\mathbf{L}_d} & \boldsymbol{\Lambda}_{\hat{\mathbf{C}}_s} \end{array}\right] = \left[\begin{array}{c|c} \frac{\partial \mathbf{P}_g}{\partial \mathbf{L}_d} & \frac{\partial \mathbf{P}_g}{\partial \hat{\mathbf{C}}_s} \\ \frac{\partial \lambda}{\partial \mathbf{L}_d} & \frac{\partial \lambda}{\partial \hat{\mathbf{C}}_s} \\ \frac{\partial \boldsymbol{\tau}_s}{\partial \mathbf{L}_d} & \frac{\partial \boldsymbol{\tau}_s}{\partial \hat{\mathbf{C}}_s} \\ \frac{\partial \boldsymbol{\mu}_s}{\partial \mathbf{L}_d} & \frac{\partial \boldsymbol{\mu}_s}{\partial \hat{\mathbf{C}}_s} \end{array}\right]. \tag{4.35}$$

Using the sensitivities of two shadow prices with respect to $\hat{\mathbf{C}}_s$ $\left(\frac{\partial \lambda}{\partial \hat{\mathbf{C}}_s}, \frac{\partial \boldsymbol{\mu}_s}{\partial \hat{\mathbf{C}}_s}\right)$ in $\boldsymbol{\Lambda}_{\hat{\mathbf{C}}_s}$ and according to the definition of LMP, we finally construct the matrix $\boldsymbol{\Lambda}_A$.

On the other hand, in the Ex-post model, (4.29) can be extended as follows:

$$\begin{bmatrix} \mathbf{M} & -\mathbf{1}_{N_b} & \Upsilon \\ \mathbf{1}_{N_b}^T & 0 & \mathbf{0} \\ \Upsilon^T & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} d\mathbf{P}_g \\ d\lambda \\ d\boldsymbol{\tau}_s \\ d\boldsymbol{\mu}_s \end{bmatrix} = \begin{bmatrix} \mathbf{U_1}^T & \mathbf{U_2}^T & \mathbf{U_3}^T \end{bmatrix} \begin{bmatrix} d\hat{\mathbf{P}}_g \\ d\hat{\mathbf{C}}_s \\ d\hat{\mathbf{D}}_s \end{bmatrix}. \tag{4.36}$$

$\hat{\mathbf{D}}_s$ is the subvector of $\hat{\mathbf{D}}$ (the real power flow estimate vector) that corresponds to the binding constraints, and

$$\mathbf{U_3}_{(B_f \times [N_b+1+B_g+B_f])} = \begin{bmatrix} \mathbf{0} & \mathbf{0}_{B_f} & \mathbf{0} & \mathbf{I}_{B_f} \end{bmatrix}. \tag{4.37}$$

Compared to (4.35), $\boldsymbol{\Lambda}_p$ in the Ex-post model is written as

$$\boldsymbol{\Lambda}_p = \left[\begin{array}{c|c|c} \boldsymbol{\Lambda}_{\hat{\mathbf{P}}_g} & \boldsymbol{\Lambda}_{\hat{\mathbf{C}}_s} & \boldsymbol{\Lambda}_{\hat{\mathbf{D}}_s} \end{array} \right] = \left[\begin{array}{c|c|c} \frac{\partial \mathbf{P}_g}{\partial \hat{\mathbf{P}}_g} & \frac{\partial \mathbf{P}_g}{\partial \hat{\mathbf{C}}_s} & \frac{\partial \mathbf{P}_g}{\partial \hat{\mathbf{D}}_s} \\ \frac{\partial \lambda}{\partial \hat{\mathbf{P}}_g} & \frac{\partial \lambda}{\partial \hat{\mathbf{C}}_s} & \frac{\partial \lambda}{\partial \hat{\mathbf{D}}_s} \\ \frac{\partial \boldsymbol{\tau}}{\partial \hat{\mathbf{P}}_g} & \frac{\partial \boldsymbol{\tau}}{\partial \hat{\mathbf{C}}_s} & \frac{\partial \boldsymbol{\tau}}{\partial \hat{\mathbf{D}}_s} \\ \frac{\partial \boldsymbol{\mu}}{\partial \hat{\mathbf{P}}_g} & \frac{\partial \boldsymbol{\mu}}{\partial \hat{\mathbf{C}}_s} & \frac{\partial \boldsymbol{\mu}}{\partial \hat{\mathbf{D}}_s} \end{array} \right]. \tag{4.38}$$

4.3.2.2 Sensitivity of State Estimation to Continuous Sensor Data

Sensitivity analysis of state estimation subject to SCADA measurements was pioneered by Stuart and Herget [75], who investigated the effect of power system modeling errors on weighted least squares (WLS) state estimation. A more rigorous sensitivity analysis method, based on the same perturbation approach illustrated

in [32], has been proposed by Mínguez and Conejo [76]. This method has been formulated in a general optimization problem that allows for the sensitivity analysis of alternative state estimation methods with different objective functions, such as the least absolute value (LAV) from the weighted least squares. It should be noted that, in this section, the sensitivity analysis is based on WLS state estimation. However, one can apply it to various state estimation methods by using the method proposed in [76].

We first derive the matrix $\mathbf{\Lambda}_B$ that illustrates the sensitivities of the real power injection and real flow measurement estimates with respect to the changes in all types of measurements. In equation (4.4), the matrix $\mathbf{\Psi}(\hat{\mathbf{x}}^k)$ is defined and partitioned as

$$\mathbf{\Psi}(\hat{\mathbf{x}}^k) = [\mathbf{G}(\hat{\mathbf{x}}^k)]^{-1}\mathbf{H}^T(\hat{\mathbf{x}}^k)\mathbf{R}^{-1} = \left[\begin{array}{c} \mathbf{\Psi}_{\hat{\boldsymbol{\theta}}}(\hat{\mathbf{x}}^k) \\ \hline \mathbf{\Psi}_{\hat{\mathbf{V}}}(\hat{\mathbf{x}}^k) \end{array}\right] \tag{4.39}$$

where $\mathbf{\Psi}_{\hat{\boldsymbol{\theta}}}(\hat{\mathbf{x}}^k)$ and $\mathbf{\Psi}_{\hat{\mathbf{V}}}(\hat{\mathbf{x}}^k)$ represent the sensitivities of the voltage phase angle estimates and the magnitudes with respect to all perturbed measurements at the k-th iteration, respectively. Therefore, (4.4) can be rewritten as

$$\left[\begin{array}{c} d\hat{\boldsymbol{\theta}}^{k+1} \\ \hline d\hat{\mathbf{V}}^{k+1} \end{array}\right] = \left[\begin{array}{c} \mathbf{\Psi}_{\hat{\boldsymbol{\theta}}}(\hat{\mathbf{x}}^k) \\ \hline \mathbf{\Psi}_{\hat{\mathbf{V}}}(\hat{\mathbf{x}}^k) \end{array}\right] d\mathbf{z}. \tag{4.40}$$

It should be noted that the DCOPF-based SCED is formulated with linearized real power injection and a flow estimation solution [77]. Using the linear equations in the upper partition of equation (4.40) and the matrix $\mathbf{\Psi}_{\hat{\boldsymbol{\theta}}}$ computed with the converged

estimate $\hat{\mathbf{x}}$, we have the following sensitivity equation:

$$d\hat{\mathbf{z}}_r = \begin{bmatrix} \mathbf{B}^S_{P\theta} \\ \mathbf{B}_{P\theta} \\ \mathbf{B}_{F\theta} \end{bmatrix} d\hat{\boldsymbol{\theta}} = \begin{bmatrix} \mathbf{B}^S_{P\theta} \\ \mathbf{B}_{P\theta} \\ \mathbf{B}_{F\theta} \end{bmatrix} \boldsymbol{\Psi}_{\hat{\boldsymbol{\theta}}} d\mathbf{z} = \mathbf{K} d\mathbf{z} \tag{4.41}$$

where

$$\mathbf{K} = \begin{bmatrix} \mathbf{B}^S_{P\theta} \\ \mathbf{B}_{P\theta} \\ \mathbf{B}_{F\theta} \end{bmatrix} \boldsymbol{\Psi}_{\hat{\boldsymbol{\theta}}}. \tag{4.42}$$

$d\hat{\mathbf{z}}_r$ is the perturbed estimate vector of the real power injection and the flow measurements. The matrix $\mathbf{B}_{P\theta} = \mathbf{A}_r \mathbf{B}_d \mathbf{A}_r^T$ is defined as the $(N_b - 1) \times (N_b - 1)$ reduced node-to-node susceptance matrix that explains the relationship between real power injections at any bus except the slack bus and the phase angles. Here $\mathbf{B}_d = diag(s_1, s_2, \ldots, s_{N_l})$ is the $N_l \times N_l$ diagonal branch susceptance matrix and $\mathbf{A}_r$ is the $(N_b - 1) \times N_l$ reduced node-to-branch incidence matrix without a slack bus. According to the law of conservation of power, the $1 \times (N_b - 1)$ matrix $\mathbf{B}^S_{P\theta} = -\mathbf{1}^T_{(N_b-1)} \mathbf{B}_{P\theta}$ is derived, and it explains the relationship between real power injections at the slack bus and the phase angles. The matrix $\mathbf{B}_{F\theta} = \mathbf{B}_d \mathbf{A}_r^T$ specifies the relationship between real power flows and the phase angles. Using (4.42), we compute the matrix $\boldsymbol{\Lambda}_B = \mathbf{K}$.

Remark 5. *The proposed sensitivity analysis is beneficial in several ways. Through on-line predictions of LMP variations, system operators can identify economically sensitive buses with respect to data corruption. This could help system operators prioritize sensor data quality upgrades in view of robust market operations. From a practical perspective, the proposed approach can be easily integrated into applica-*

tions in existing energy management systems (EMSs) or market management systems (MMSs).

Remark 6. *The sensitivity index* (4.20) *can be used to provide the following* system-wide *metrics that alert system operators to the jth most and kth least influential sensors on LMP, on average, with respect to a total of N_b buses and T dispatch intervals:*

$$j = \arg\max_{j} \left(\sum_{i=1}^{N_b} \sum_{t=1}^{T} \left| \mathbf{\Lambda}_{(i,j)}^{(t)} \right| / N_b T \right)$$
$$k = \arg\min_{k} \left(\sum_{i=1}^{N_b} \sum_{t=1}^{T} \left| \mathbf{\Lambda}_{(i,k)}^{(t)} \right| / N_b T \right).$$

4.3.3 Numerical Example

In this subsection, we illustrate and verify the proposed approach to quantifying the sensitivities of real-time LMP with respect to changes in sensor data. The proposed sensitivity analysis is applied to IEEE 14-bus and 118-bus systems. System data for the IEEE 14-bus system are taken from the MATPOWER 4.0 IEEE 14-bus test case file. Table 4.2 shows the generator parameters in the IEEE 14-bus system.

Table 4.2: Generator parameters of the IEEE 14-bus system.

Bus	$P_{g_i}^{\min}$(MW)	$P_{g_i}^{\max}$(MW)	a_i(\$/MWh)	b_i(\$/(MW)^{2}h)
1	0	332.4	20	0.043
2	0	140	20	0.25
3	0	100	40	0.01
6	0	100	40	0.01
8	0	100	40	0.01

In this simulation, the measurement configuration consists of 8 voltage magnitude measurements, 8 pairs of real and reactive power injection measurements, and 12 pairs of real and reactive power flow measurements. V_i is the measurement of voltage magnitude at bus i, P_i and Q_i are the measurements of real and reactive power injection at bus i, respectively, and $P_{i,j}$ and $Q_{i,j}$ are the measurements of real and reactive power flow from bus i to bus j, respectively. Fig. 4.3 shows the IEEE 14-bus system with a measurement configuration that consists of the following five measurement sets:

$$\mathcal{S}_v = \{V_2, V_3, V_7, V_8, V_{10}, V_{11}, V_{12}, V_{14}\} \tag{4.43}$$

$$\mathcal{S}_{ri} = \{P_2, P_3, P_7, P_8, P_{10}, P_{11}, P_{12}, P_{14}\} \tag{4.44}$$

$$\mathcal{S}_{ai} = \{Q_2, Q_3, Q_7, Q_8, Q_{10}, Q_{11}, Q_{12}, Q_{14}\} \tag{4.45}$$

$$\mathcal{S}_{rf} = \{P_{1,2}, P_{2,3}, P_{4,2}, P_{4,7}, P_{4,9}, P_{5,2}, P_{5,4}, P_{5,6}, \tag{4.46}$$

$$P_{6,13}, P_{7,9}, P_{11,6}, P_{12,13}\} \tag{4.47}$$

$$\mathcal{S}_{af} = \{Q_{1,2}, Q_{2,3}, Q_{4,2}, Q_{4,7}, Q_{4,9}, Q_{5,2}, Q_{5,4}, Q_{5,6}, \tag{4.48}$$

$$Q_{6,13}, Q_{7,9}, Q_{11,6}, Q_{12,13}\}. \tag{4.49}$$

In this measurement configuration, the locations of the voltage magnitude measurements are consistent with those of the real and reactive power injection measurements. For each measurement set, the measurement index is numbered from one to the total number of measurements in each set.

We assume that all measurements are corrupted by additive Gaussian noises with equal variances σ^2=0.00001. Finally, for all buses i, j, and k, we compute LMP sensitivities with respect to the five types of measurements—real/reactive power

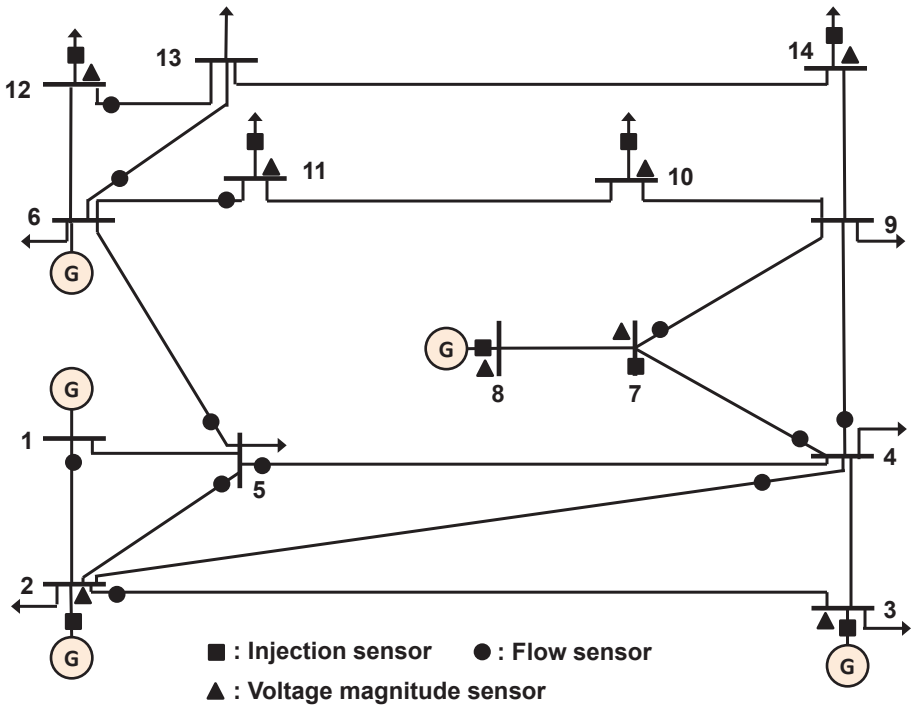


Figure 4.3: IEEE 14-bus system with a given measurement configuration.

injection, real/reactive power flow, and voltage magnitude—as follows:

$$\frac{\partial \pi_i}{\partial P_j}, \quad \frac{\partial \pi_i}{\partial Q_j}, \quad \frac{\partial \pi_i}{\partial P_{j,k}}, \quad \frac{\partial \pi_i}{\partial Q_{j,k}}, \quad \frac{\partial \pi_i}{\partial V_j}. \tag{4.50}$$

Units for the sensitivities $\left\{\frac{\partial \pi_i}{\partial P_j}, \frac{\partial \pi_i}{\partial P_{j,k}}\right\}$, $\left\{\frac{\partial \pi_i}{\partial Q_j}, \frac{\partial \pi_i}{\partial Q_{j,k}}\right\}$, and $\left\{\frac{\partial \pi_i}{\partial V_j}\right\}$ are (\$/MWh)/(puMW), (\$/MWh)/(puMVAr), and (\$/MWh)/(puV), respectively.

These figures provide information about the directions of the post-corruption LMPs as well as their sensitivities with respect to each type of measurement at a given dispatch time. In this simulation, after the Ex-ante dispatch problem has been solved, there exist two binding generation capacity constraints: P_{g_3} and P_{g_8} are binding at $\hat{P}_{g_3}^{\min}$ and $\hat{P}_{g_8}^{\max}$, respectively. We assume that the corruption of the measurements impacts the binding constraint associated with P_{g_3}. In other words, the corrupted

measurements affect $\hat{P}_{g_3}^{\min}$ (an intermediate variable in (4.21)), subsequently leading to changes in all the LMPs. We randomly choose seven buses (buses 1, 2, 3, 4, 5, 10, 13) out of the fourteen to differentiate clearly the LMP sensitivities among the various buses. The absolute values of the LMP sensitivities at buses 3 and 5 are the largest and smallest, setting the upper and lower bounds for sensitivity at the fourteen buses. We obtain from the simulation results the following observations:

(O1) *Sensitivity grouping property*: all buses can be categorized into two sensitivity groups. In each group, buses obtain sensitivities with the same sign, but of different magnitude and subject to all types of measurements. Group I includes buses 1, 2, 3 and 5, and Group II buses 4, 10 and 13. For example, in Fig. 4.4(b) the corruption of z_2 yields positive sensitivities for Group I and negative sensitivities for Group II, whereas the corruption of z_6 yields the reverse: negative sensitivities for Group I and positive sensitivities for Group II. This grouping property enables system operators to predict rapidly the direction of LMP's distortion in response to sensor data corruption.

(O2) *Identification of buses that are economically sensitive to data corruption*: buses incident to both ends of the congested line have the highest LMP sensitivities with respect to sensor data corruption. For example, bus 3 in Group I and bus 4 in Group II incident to congested line 3-4 have the largest absolute sensitivities in each group. In particular, it should be noted that the largest sensitivities are associated with bus 3. This implies that bus 3 is the most financially vulnerable to any corruption in sensor measurement.

(O3) *Identification of influential sensors on LMP*: the sensor most influential on LMP change is identified in each measurement group. In Fig. 4.4(a),(b), the sensors with z_2 (P_3 and Q_3) have the most significant impact on LMP. This is

due to the fact that the change of the intermediate variable $\hat{P}_{g_3}^{\min}$ is dominantly affected by P_3 and Q_3, subsequently leading to more change in LMP. This effect is also verified in the figure on p. 93 (a),(b) based on the IEEE-118 bus system. In Fig. 4.4(c),(d) and (e), the sensors with z_8, z_{11} and z_3 ($P_{5,6}$, $Q_{11,6}$ and V_7) are the most influential, respectively. In addition, it should be noted that the localized effects on increasing sensitivity of measurements adjacent to the congested line and/or the intermediate variable do not always hold true. For example, z_{11} ($P_{11,6}$) is farther away from both the congested line and the intermediate variable than z_5 ($P_{4,9}$); however, in Fig. 4.4(c), data corruption in the former leads to a higher sensitivity than in the latter. This non-localized data effect motivates system operators to use our developed tool for identifying which sensors impact LMP sensitivity.

(O4) *Impact of different types of sensor data on LMP*: through a comparison of all the figures, LMP appears to be more sensitive to real power injection/flow measurements than to reactive power injection/flow and voltage magnitude measurements. In order to compare the sensitivities of different units fairly, a normalized LMP sensitivity $|z_j|\frac{\partial \pi_i}{\partial z_j}$ is defined, which is incorporated into the following proposed metric:

$$\Omega_k^i = \sum_{j=1}^{|\mathcal{S}_k|} \left| |z_j| \frac{\partial \pi_i}{\partial z_j} \right| / |\mathcal{S}_k| \tag{4.51}$$

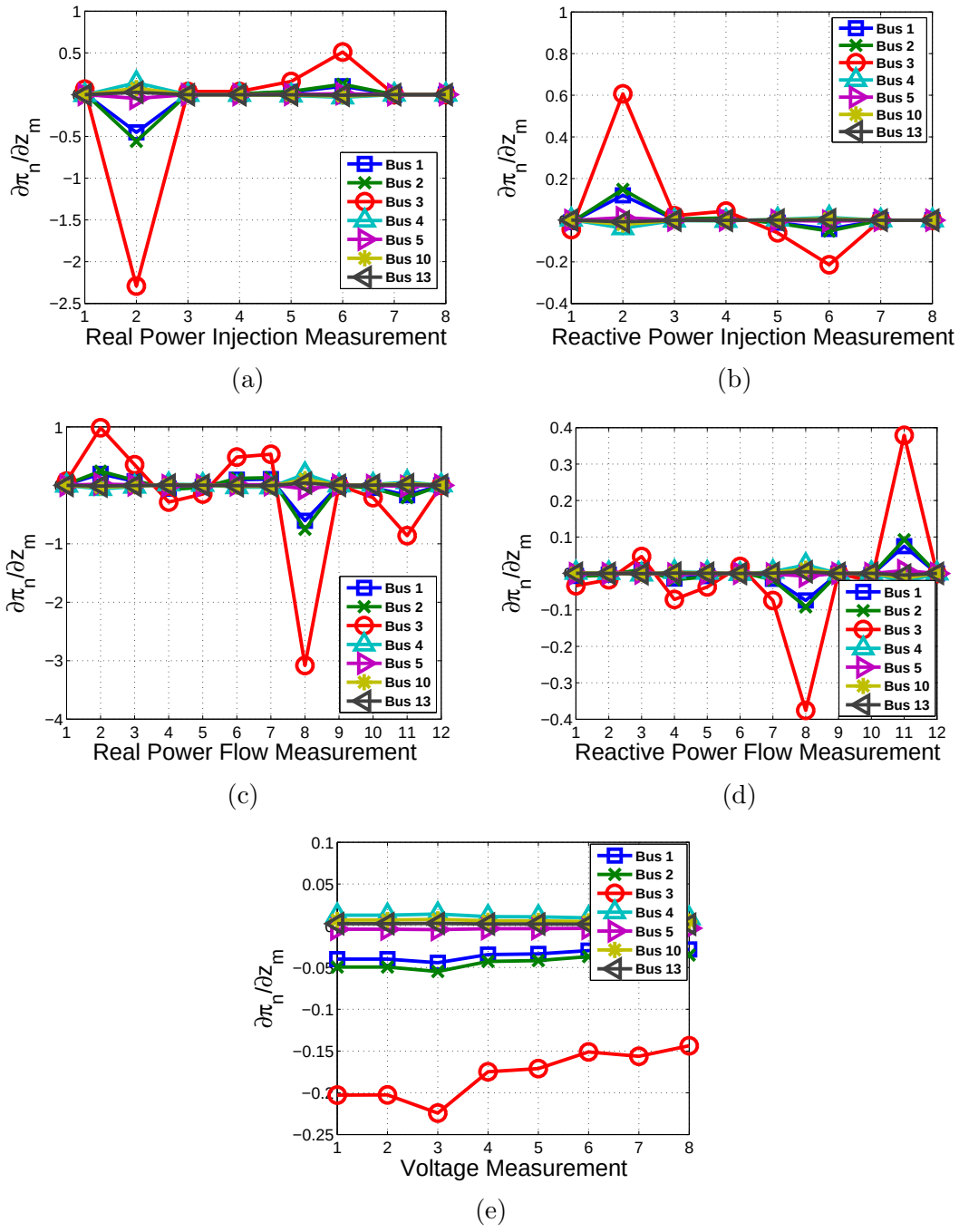
where Ω_k^i is the average of the absolute normalized sensitivities at bus i with respect to any measurement z_j in the set $\mathcal{S}_k$ ($k = v, ri, ai, rf, af$). The cardinality of the set $|\mathcal{S}_k|$ means the number of elements in $\mathcal{S}_k$. For example, at bus 3, we compute $\Omega_{ri}^3 = 0.474$, $\Omega_{rf}^3 = 0.253$, $\Omega_v^3 = 0.175$, $\Omega_{af}^3 = 0.013$, and

$\Omega^3_{ai} = 0.012$, which is consistent with our expectation that real power injection and flow measurements have a more significant impact on LMP sensitivity than other measurements. This is due to the fact that DCOPF-based SCED is conducted based on a linearized state estimation solution that is more influenced by real power measurements than by reactive power and voltage magnitude measurements, as illustrated in (4.40) and (4.41).

(O5) In Fig. 4.4(e), LMP sensitivities at all buses affected by corrupted voltage magnitude measurements fluctuate more smoothly than the ones affected by other types of corrupted measurements. In other words, all voltage magnitude measurements impact LMP variations almost evenly. In addition, the non-localized effect mentioned in (O3) is also verified between z_2 (V_3) and z_3 (V_7).

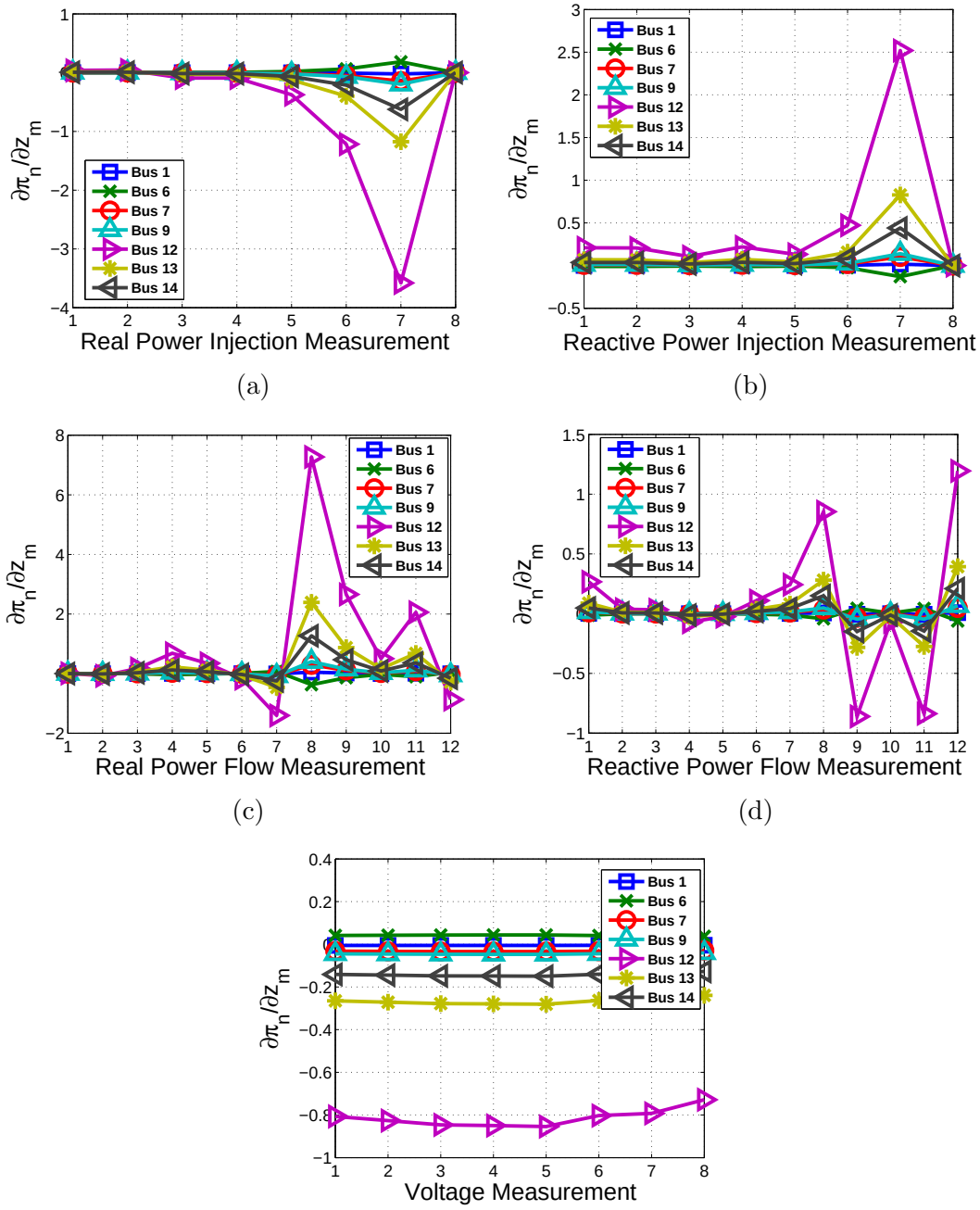
Fig. 4.5 provides snapshots of the Ex-post LMP sensitivities at arbitrarily chosen buses (buses 1, 6, 7, 9, 12, 13 and 14) with respect to the aforementioned five types of sensor measurements. In this simulation, line 6-12 is assumed to be congested at both Ex-ante dispatch and Ex-post dispatch. $\hat{P}_{6,12}$ is chosen as an intermediate variable to compute LMP sensitivity. We can observe from Fig. 4.5 the same phenomena as in Fig. 4.4: (O1) Group 1 for buses 1 and 6, and Group 2 for buses 7, 9, 12, 13 and 14; (O2) buses 6 and 12 incident to the congested line have the largest absolute value of LMP sensitivity in each group; (O3) in Fig. 4.5(a) and (b), the sensors with z_7 (P_{12} and Q_{12}) have the most significant impact on LMP, and in Fig. 4.5(c), (d) and (e), the sensors with z_8, z_{12} and z_5 ($P_{5,6}$, $Q_{12,13}$ and V_{10}) are the most influential, respectively; and (O4) & (O5) real power measurements have a stronger impact on LMP sensitivity than the reactive power and voltage magnitude measurements, and the voltage magnitude measurements influence LMP sensitivity almost evenly.

In addition, these analytical LMP sensitivity results have been checked using the



(a) (b) (c) (d) (e)

Figure 4.4: Sensitivities of Ex-ante prices with respect to (a) real power injection measurements, (b) reactive power injection measurements, (c) real power flow measurements, (d) reactive power flow measurements, and (e) voltage magnitude measurements. Line 3-4 is congested and P_{g_3} is binding at $\hat{P}_{g_3}^{\min}$ in the IEEE 14-bus system.

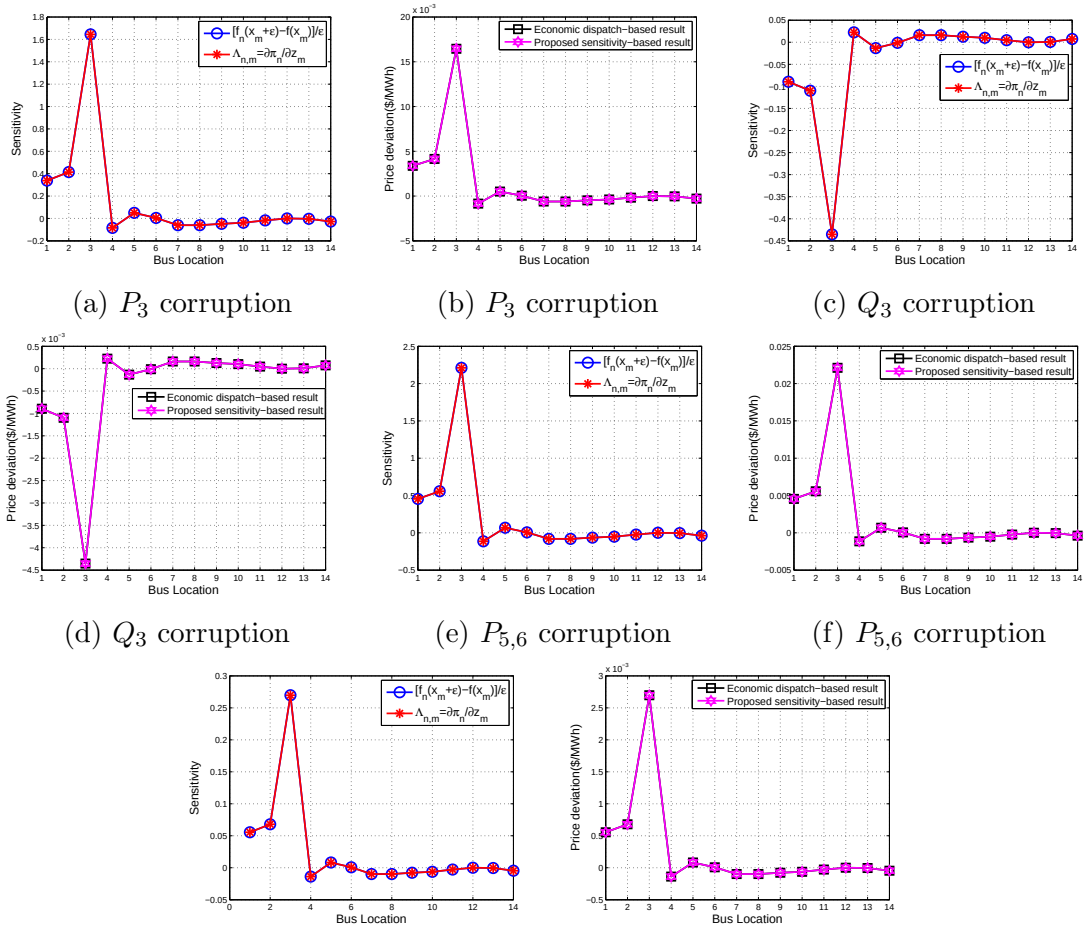


(a) (b) (c) (d) (e)

Figure 4.5: Sensitivities of Ex-post prices with respect to (a) real power injection measurements, (b) reactive power injection measurements, (c) real power flow measurements, (d) reactive power flow measurements, and (e) voltage magnitude measurements. Line 6-12 is congested and the corresponding line flow is binding at the capacity limit of line 6-12 in the IEEE 14-bus system.

perturbation method $\left(f'(x) \approx \frac{[f(x+\epsilon)-f(x)]}{\epsilon}\right)$. It has been verified that the sensitivity results obtained from the proposed analytical method are consistent with those from the perturbation method. Fig. 4.6 shows LMP sensitivities and corresponding actual price deviations with respect to the four randomly chosen measurements in the IEEE 14-bus system. These results are obtained in the same setup as the Ex-ante simulation in this section. From Fig. 4.6(a), (c), (e), (g), we observe that the sensitivities from the perturbation method ($\frac{[f_n(x_m+\epsilon)-f_n(x_m)]}{\epsilon}$ with $\epsilon = 0.01$) are consistent with those from the proposed analytical method ($\Lambda_{n,m} = \frac{\partial \pi_n}{\partial z_m}$). Fig. 4.6(b), (d), (f), (h) shows actual price deviations computed by economic dispatch and the proposed method ($\Lambda_{n,m} \times \epsilon$) when $\epsilon = 0.01$, respectively. As we can see, both results are consistent with each other.

Fig. 4.7 shows actual Ex-ante LMP and how they differ when they have or do not have corrupted data at all buses. It is assumed that the magnitude of z_8 is corrupted by 2% in Fig. 4.4(c). In the Chi-squares test [43] within a 99% confidence level, the estimated objective functions and the bad data detection threshold are computed. $J(\hat{x}) = 15.69$ and $J^{(b)}(\hat{x}) = 30.17$ correspond to the values of the estimated objective functions without and with corrupted data, respectively, and $\chi^2 = 38.93$ is the value of the bad data detection threshold. It should be noted that since $J^{(b)}(\hat{x}) = 30.17 < \chi^2 = 38.93$, the corrupted measurement z_8 bypasses the bad data detection engine, which could then lead to LMP distortion. As expected, Fig. 4.7 justifies the result of our sensitivity analysis in two main ways. First, the prices at buses 1, 2, 3, and 5 in Group I change in a positive direction whereas the prices at the buses in Group II change in a negative direction. This observation explains the grouping property specified in (O1). Second, the descending order of the magnitudes of the actual price deviations is in accordance with that of sensitivity magnitudes. For example, Fig. 4.4(c) shows that buses 3, 2, 1 and 5 in Group I are in descending order of



(a) P_3 corruption (b) P_3 corruption (c) Q_3 corruption

(d) Q_3 corruption (e) $P_{5,6}$ corruption (f) $P_{5,6}$ corruption

(g) $Q_{5,6}$ corruption (h) $Q_{5,6}$ corruption

Figure 4.6: Comparison of sensitivities between the perturbation method ($\epsilon = 0.01$) and the proposed method with respect to (a) P_3 corruption, (c) Q_3 corruption, (e) $P_{5,6}$ corruption, and (g) $Q_{5,6}$ corruption in Fig. 4.6(a), (c), (e), (g). Comparison of price deviations between economic dispatch and the proposed method with respect to (b) P_3 corruption, (d) Q_3 corruption, (f) $P_{5,6}$ corruption, and (h) $Q_{5,6}$ corruption in Fig. 4.6(b), (d), (f), (h).

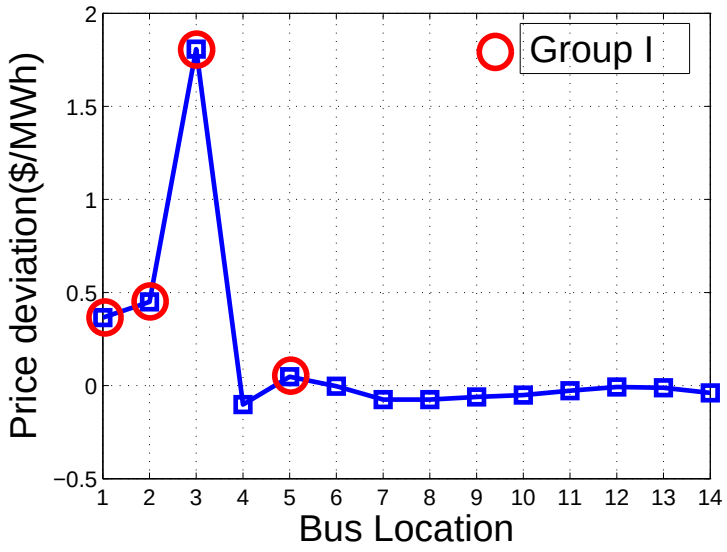
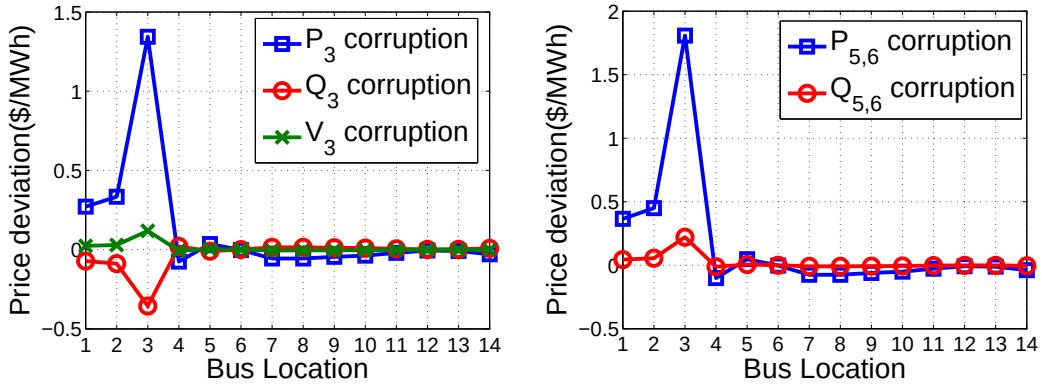


Figure 4.7: LMP differences between with and without corrupted data when z_8 is corrupted in Fig. 4.4(c).

sensitivity magnitudes, which is consistent with the descending order of the actual price deviations at those same buses in Fig. 4.7.

Fig. 4.8 shows the Ex-ante LMP deviations that are caused by the undetectable same amount of corruption in each measurement group $\{P_3, Q_3, V_3\}$ and $\{P_{5,6}, Q_{5,6}\}$. These figures show that real power injection and flow measurements have a more significant impact on LMP than other measurements. This fact justifies observation (O4).

Fig. 4.9 shows the impact of sensor measurement accuracy on LMP sensitivity. In this figure, four plots represent LMP sensitivities at bus 3 in Fig. 4.4(a), with consistently varying variances of the two injection measurements z_2 (P_3) and z_6 (P_{11}). These sensitivities are measured at four different variance levels; σ^2=0.00005, 0.0001, 0.0005, and 0.001. We can observe from Fig. 4.9 that, as the measurement variance decreases (i.e., the measurement accuracy increases), the corresponding LMP sen-



(a) (b)

Figure 4.8: LMP differences between with and without corrupted data in Fig. 4.4 (a) P_3, Q_3, and V_3 corruptions (b) $P_{5,6}$ and $Q_{5,6}$ corruptions.

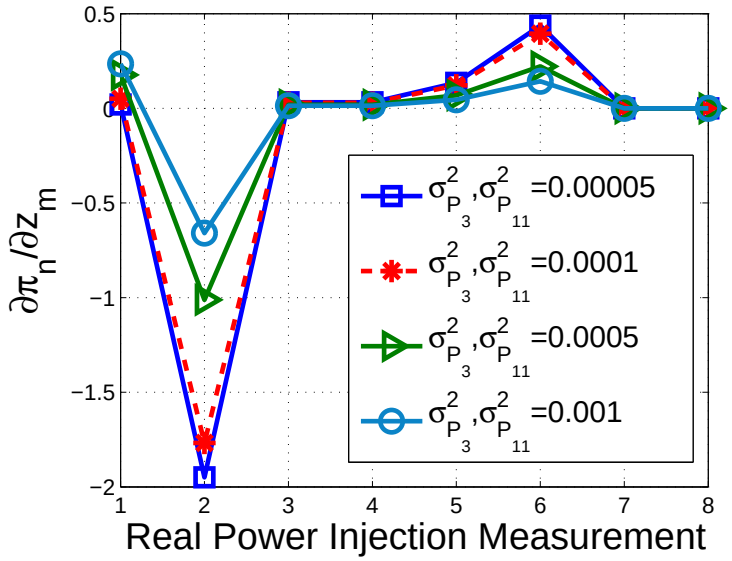


Figure 4.9: Comparison of LMP sensitivities at bus 3 in Fig. 4.4(a) with varying variances of injection measurements P_3 and P_{11}.

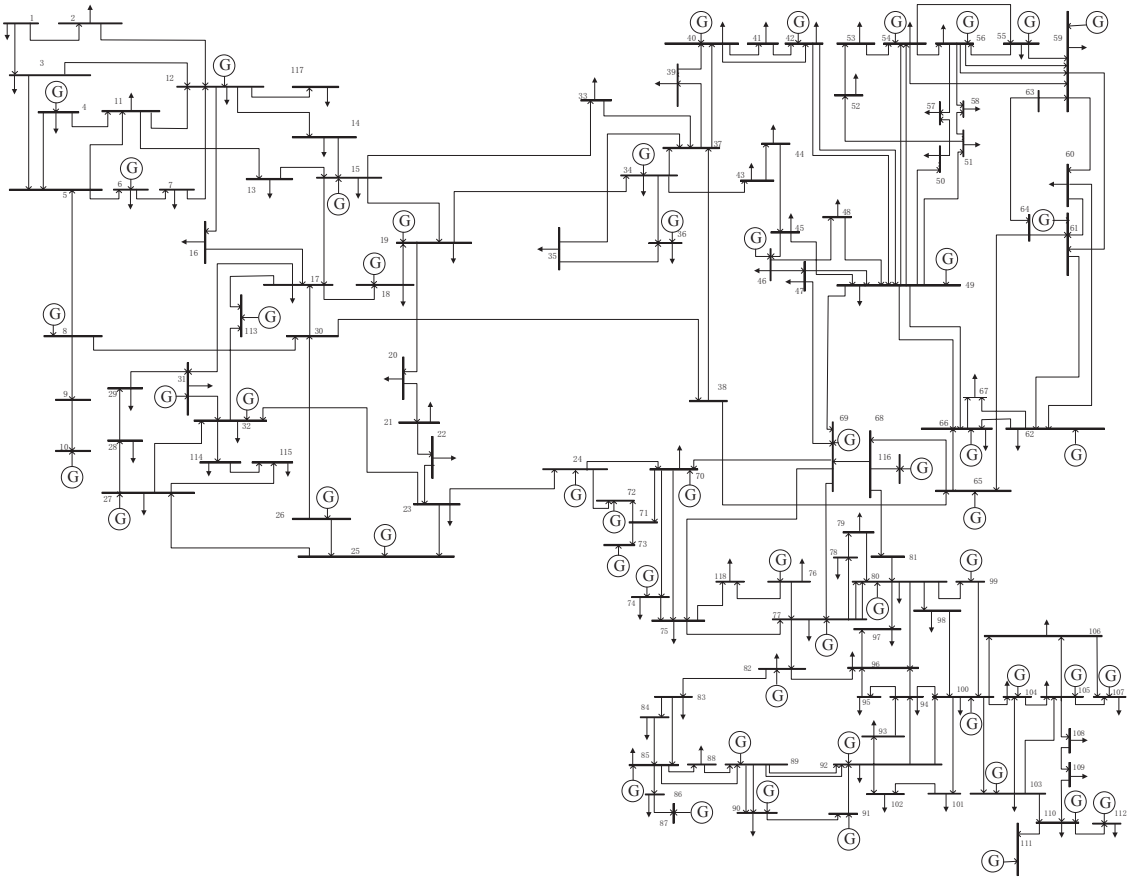


Figure 4.10: IEEE 118-bus system.

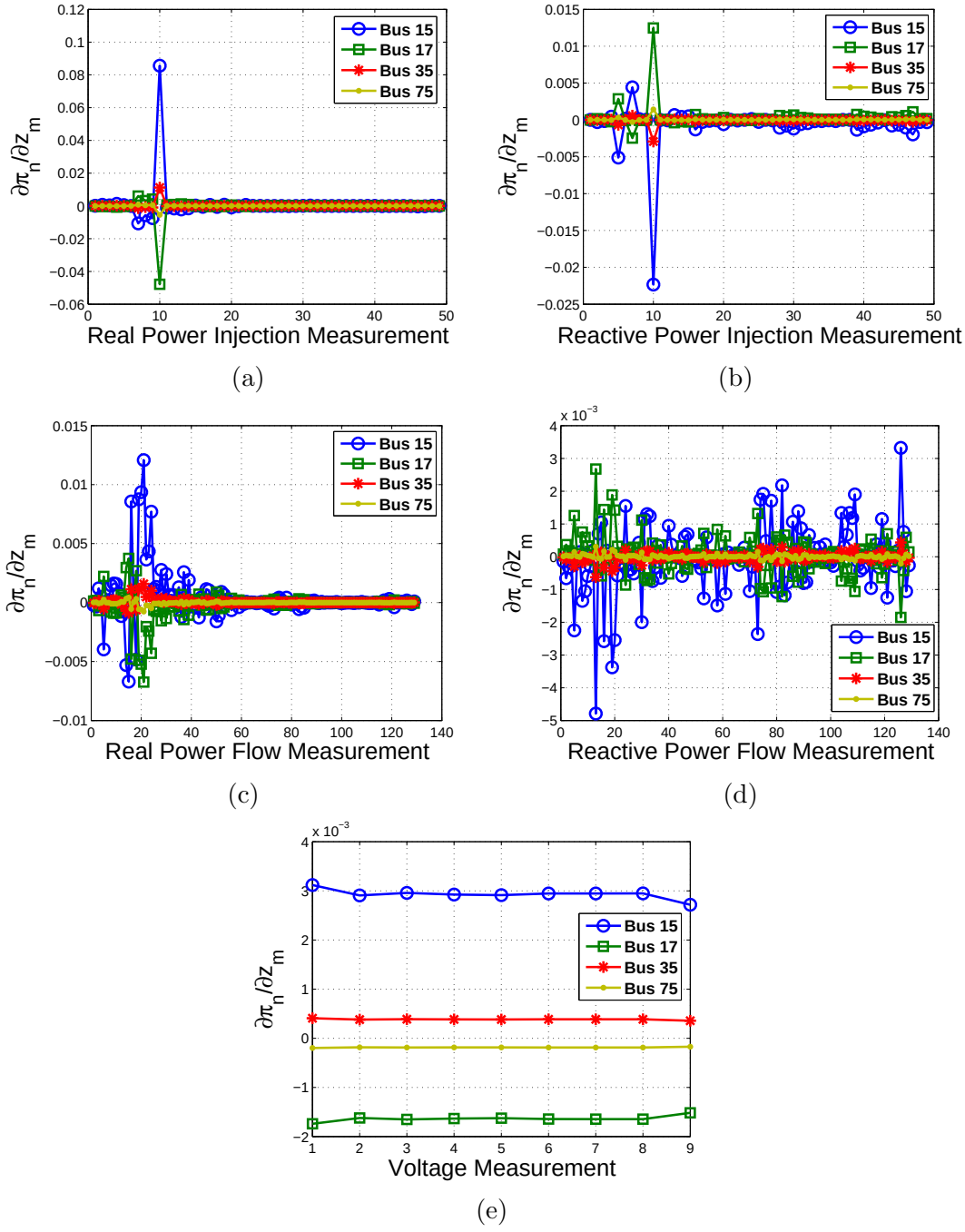


Figure 4.11: Sensitivities of Ex-ante prices with respect to (a) real power injection measurements, (b) reactive power injection measurements, (c) real power flow measurements, (d) reactive power flow measurements, and (e) voltage magnitude measurements. Line 15-17 is congested and $P_{g_{19}}$ is binding at $\hat{P}_{g_{19}}^{\max}$ in the IEEE 118-bus system.

sitivity increases. In other words, more accurate sensors lead to more change in LMP while sensor data remain corrupted. This shows the coupling between state estimation accuracy and LMP calculation. Based on this observation, one possible guideline for mitigating the financial risk from data corruption is to make it a high priority to protect accurate sensors.

For the IEEE 118-bus system, with 54 generation buses and 186 transmission lines as shown in Fig. 4.10, we assume that real and reactive power injection measurements are placed at 49 generator buses, voltage magnitude measurements at 9 generator buses, and real and reactive flow measurements at 129 lines. Therefore, this system has a total of 365 measurements. System data for the IEEE 118-bus system are taken from the MATPOWER 4.0 IEEE 118-bus test case file.

Fig. 4.11 show the Ex-ante LMP sensitivities at buses 15, 17, 35, and 75 in the IEEE 118-bus system with line 15-17 congestion with respect to the five different types of measurement. The magnitudes of the sensitivities at buses 15 and 17 are the highest in each sensitivity group. $\hat{P}_{19}^{\max}$ is chosen as an intermediate variable to compute LMP sensitivity. As expected, all observations from Fig. 4.4 are also verified in the larger IEEE 118-bus system: 1) sensor grouping property (Group 1: buses 15 and 35, and Group 2: buses 17 and 75); 2) identification of the most economically sensitive buses in each group (buses 15 and 17) and the most influential sensors (e.g., z_{10} with P_{19} and Q_{19} in Fig. 4.11(a), (b)) on LMP change; and 3) the impact of different types of sensor data on LMP (e.g., the more significant impact of real power measurements than other types of measurements).

4.3.4 Discussions

4.3.4.1 PMU Implementation

Recently, with Phasor Measurement Units (PMUs) being increasingly deployed in power systems, novel hybrid state estimation methods based on traditional SCADA and PMU measurements have been intensively investigated [78, 79, 80, 81]. The state estimation measurement model (4.1) presented in this dissertation can be easily expanded into a hybrid model (e.g., equation (12) in [79]). Using the same steps illustrated in Subsection 4.3.2.2, our proposed approach quantifies LMP sensitivity with respect to PMU as well as SCADA measurements.

4.3.4.2 Dependent Sensor Measurements

Recent studies have speculated that state estimation measurements with a substation may be correlated, and proposed as a response a novel dependent weighted least square (DWLS) state estimation method that considers measurement dependencies [82]. These methods differ from the traditional weighted least square (WLS) method based on independent measurements only in that it computes the nondiagonal covariance matrix of dependent measurements using the point estimate technique in the WLS formulation. By simply replacing diagonal covariance matrix $\mathbf{R}$ in (4.2) with a computed nondiagonal covariance matrix, our proposed formulation is also applicable to any LMP sensitivity analysis subject to dependent sensor measurements.

4.3.4.3 LMP Sensitivity to Network Topology Errors

Section 4.3 is limited to the study of LMP sensitivity with respect to changes in the analog data corruption-induced power flow estimate. A very important extension of our work here would be to study LMP sensitivity with respect to network topology estimate changes caused by corrupted digital sensor data such as the on/off status

of a circuit breaker. A key part of this task would be to investigate how such LMP sensitivity is analytically expressed as a function of topology information such as the generation shift factor matrix and corrupted digital data. As an initial step in this direction, recent work of ours in Section 4.4 proposed a simple LMP sensitivity index that accounts for the relationship between changes in network topology and LMP during single transmission line congestion. This index quantifies the *economic* impact of network topology errors on LMP and is expressed in terms of the energy costs of marginal units and the congested line-related generation shift factors at any bus and marginal units. Here, a marginal unit is defined as a unit that generates power in a range somewhere between its minimum and maximum capacity. Future work should include the development of a sensitivity index that evaluates the *cyber* impact of digital data on network topology estimate.

4.4 Impact Analysis of LMP Subject to Network Topology Estimate Errors

4.4.1 Introduction

Locational marginal price (LMP) is the key variable to competitive wholesale electricity markets. In real-time power markets, LMP is obtained as the by-product of security constrained economic dispatch (SCED). SCED is formulated with the state estimation solution (e.g., the estimated power flow and network topology) based on sensor data in supervisory control and data acquisition (SCADA) system. Therefore, data corruption may result in the significant change of real-time LMP via *state estimation*. The subject of this section is to examine the impact of *power network topology estimation error* on *real-time market operations*.

The transmission-level network model of a power system, namely, the bus/branch model is built by topology processor using discrete data (e.g., the on/off status of

circuit breakers (CBs)) collected by SCADA sensors. Bad CB status data are generally detected and identified by topology error processing using continuous data (e.g., power injection/flow and voltage magnitude), and afterwards topology processor rebuilds a corrected bus/branch model. Based on this corrected model, the state estimation and economic dispatch process compute the optimal estimate of the entire power system' state and the dispatch instruction such as the optimal generation output and LMP, respectively.

In this section, we assume that topology data attacks succeed, and focus on their economic impacts on real-time market operations. The main contributions of this section is twofold:

- We propose a simple analytical index that explains the relationship between the change in network topology (i.e., the left-hand side of constraints in SCED) and LMP.

- The proposed analytical approach is illustrated and verified in the IEEE 14-bus system. Using the proposed approach, we identify the most economically impacting transmission line/CBs and sensitive bus to topology error.

The proposed LMP sensitivity index provides system operators an online analysis tool to predict the effects of the induced topology errors on real-time LMP. From a practical perspective, this tool can be easily integrated into applications in the existing MMS.

This section is organized as follows. Subsection 4.4.2 provides the brief overview of models for state estimation, topology error processing, and real-time power market. A LMP sensitivity index with respect to topology error is derived in Subsection 4.4.3. Subsection 4.4.4 verifies and illustrates the derived sensitivity index in the IEEE 14-bus system.

4.4.2 Preliminaries

The main notations used throughout this section are summarized in Table 4.3. Bold symbols represent vectors or matrices.

4.4.2.1 Topology Error Processing

We consider a state estimation model based on a linearized DC power flow. The measurements taken by each sensor are written by

$$\mathbf{z} = \mathbf{J}\mathbf{x} + \mathbf{e} \tag{4.52}$$

where $\mathbf{x}$ is the state vector of the entire power system, $\mathbf{z}$ is measurement vector, $\mathbf{e}$ is independent identically distributed (i.i.d.) Gaussian random measurement error vector with zero mean and covariance matrix $\mathbf{R}$, and $\mathbf{J}$ is the *true* system Jacobian matrix of the state vector $\mathbf{x}$. Then, the weighted least squares estimate of $\mathbf{x}$ is calculated by

$$\hat{\mathbf{x}}(\mathbf{z}) = (\mathbf{J}^T\mathbf{R}^{-1}\mathbf{J})^{-1}\mathbf{J}^T\mathbf{R}^{-1}\mathbf{z}. \tag{4.53}$$

Topology error processing detects and identifies topology errors based on measurement residuals. The wrongly reported circuit breaker status data generate two types of topology errors: (i) line status error; and (ii) substation configuration error. The former represents an incorrect exclusion/inclusion of transmission lines from the network model whereas the latter a split/merging error of buses at the substation. In this section, we focus on line status error and substation configuration error is beyond the scope of this dissertation. For line status error, the measurement equation (4.52) is rewritten as

$$\mathbf{z} = \mathbf{J}_\mathbf{E}\mathbf{x} + \mathbf{E}\mathbf{x} + \mathbf{e}$$

Table 4.3: Nomenclature.

n	Index for buses n
l	Index for transmission line l
C_i	Energy cost for generator i
G	Set of generators
P_{g_n}	Total real power output at bus n
p_i	Real power output at generator i
L_{d_n}	Fixed demand at bus n
$p_i^{\min}$, $p_i^{\max}$	Min/max generation limits for generator i
$F_l^{\min}$, $F_l^{\max}$	Min/max flow limits at line l
$H_{l,n}$	Distribution factor of transmission line l to bus n
λ	Shadow price of the system energy balance equation
$\boldsymbol{\tau}$	Shadow price vector of the capacity constraints for generators
$\boldsymbol{\mu}$	Shadow price vector of the constraints for transmission lines
N_b	Total number of buses
N_l	Total number of transmission lines
$\mathbf{1}_k$	$k \times 1$ column vectors with all ones

where $\mathbf{J} = \mathbf{J_E} + \mathbf{E}$. Here, $\mathbf{J_E}$ is the incorrect system Jacobian matrix due to topology errors. $\mathbf{E}$ is the system Jacobian error matrix. Then, topology error detection is performed using the normalized residual vector as follows:

$$E(\mathbf{r}^N) = \Omega^{-\frac{1}{2}}(\mathbf{I} - \mathbf{K_E})\mathbf{M}\mathbf{f} = \mathbf{S}\mathbf{f} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{4.54}$$

where $\mathbf{M}$ is the measurement-to-branch incidence matrix, $\mathbf{f}$ is a vector of branch flow errors, $\mathbf{K_E} = \mathbf{J_E}(\mathbf{J_E}^T\mathbf{R}^{-1}\mathbf{J_E})^{-1}\mathbf{J_E}^T\mathbf{R}^{-1}$, $\Omega = \text{diag}\{Cov(\mathbf{r})\}$, $\mathbf{S} = \Omega^{-\frac{1}{2}}(\mathbf{I} - \mathbf{K_E})\mathbf{M}$ is the sensitivity matrix for $\mathbf{r}^N$ with respect to branch flow errors $\mathbf{f}$, and η is the threshold of topology error detection. H_1 and H_0 correspond to the cases with and without topology error, respectively. In this section, we assume that the attack proposed in [37] successfully changes network topology estimate while bypassing topology error detection (4.54).

4.4.2.2 Real-Time Power Market Model

Real-time power market consists of the two main pricing models: Ex-ante (e.g. in ERCOT, NY ISO) and Ex-post (e.g. in ISO New England, PJM, and Midwest ISO). Since Ex-ante and Ex-post models rely on the network topology and the cost functions of generators, our results illustrated in the next section are applicable to both models. In this section we consider a real-time Ex-ante market model where LMPs are computed before the actual deployment of dispatch orders. For the system operator, the Ex-ante dispatch is formulated as follows,

$$\min_{p_i} \sum_{i \in G} C_i \cdot p_i \tag{4.55}$$

s.t.

$$\lambda : \sum_{n=1}^{N_b} P_{g_n} = \sum_{n=1}^{N_b} L_{d_n} \tag{4.56}$$

$$\boldsymbol{\tau} : p_i^{\min} \leq p_i \leq p_i^{\max} \qquad \forall i \in G \tag{4.57}$$

$$\boldsymbol{\mu} : F_l^{\min} \leq \sum_{n=1}^{N_b} H_{l,n}(P_{g_n} - L_{d_n}) \leq F_l^{\max} \ \forall l = 1, \ldots, N_l \tag{4.58}$$

In this formulation, the objective function is to minimize the total generation costs in (4.55). (4.56) is the system-wide energy balance equation. (4.57) is the physical capacity constraints of each generator. (4.58) is the transmission line constraints. λ, $\boldsymbol{\tau}$, and $\boldsymbol{\mu}$ are the dual variables associated with the aforementioned equality and inequality constraints. $\boldsymbol{\tau}$ and $\boldsymbol{\mu}$ are expressed as $\boldsymbol{\tau} = [\boldsymbol{\tau}_{\max}^T, \boldsymbol{\tau}_{\min}^T]^T$ and $\boldsymbol{\mu} = [\boldsymbol{\mu}_{\max}^T, \boldsymbol{\mu}_{\min}^T]^T$ where subscript max(min) represents max(min) inequality constraint. $H_{l,n}$ is the element at the lth row and nth column of the $N_l \times N_b$ distribution factor matrix $\mathbf{H}$. This matrix explains the sensitivity of branch flows to nodal injection powers. The real-time LMP vector $\boldsymbol{\pi}$ is computed using the following equation:

$$\boldsymbol{\pi} = \lambda \mathbf{1}_{N_b} - \mathbf{H}^T \left[\boldsymbol{\mu}_{\max} - \boldsymbol{\mu}_{\min}\right]. \tag{4.59}$$

4.4.3 Derivation of LMP Sensitivity to Network Topology Error

In this subsection, we derive a simple sensitivity index to quantify the impact of network topology errors on LMP. This derivation is based on the following assumptions:

(A1) Only one single transmission line is congested for both cases with and without topology error.

(A2) Network congestion patterns and marginal units remain unchanged with topol-

ogy error.

(A3) The value of λ (LMP at slack bus) remains unchanged with topology error.

In (A2), a marginal unit is defined as a unit that generates power between its minimum and maximum capacity. The above assumptions would hold true under the situation in which other lines except a congested line have sufficient transmission capacity. It should be noted that these assumptions do not capture all the possible scenarios in actual operation. However, a large number of scenarios would fit into these assumptions under normal to lighted loading situations. Future work will expand the analysis to a broader set of scnarios.

We first present Proposition 1 where the shadow price associated with a congested line is expressed as the ratio of the gap of the energy costs of marginal units to the gap of distribution factors that correspond to the intersections of marginal units and the congested transmission line. It serves as the theoretical basis for the main result of this section (equation (4.75)), which does not require extensive numerical simulations in order to determine the sensitivity of LMP to network topology errors.

Proposition 1. *Let i and j be two marginal units with $C_j > C_i$, belonging to different buses. Then, the shadow price for the congested transmission line l is expressed as*

$$\mu_l = \frac{\Delta C(j,i)}{\Delta H_l(i,j)} \tag{4.60}$$

where $\Delta C(j,i) = C_j - C_i$ and $\Delta H_l(i,j) = H_{l,i} - H_{l,j}$.

Proof. The shadow price of a congested transmission line is defined as the change of total dispatch cost via relaxing the transmission constraint by one unit. Thereore,

the shadow price for the congested transmission line l can be written as

$$\mu_l = -\sum_{i \in G} C_i \Delta p_i, \tag{4.61}$$

which satisfies the following two constraints:

$$\sum_{n=1}^{N_b} \Delta P_{g_n} = 0 \tag{4.62}$$

$$\sum_{n=1}^{N_b} H_{l,n} \Delta P_{g_n} = 1. \tag{4.63}$$

(4.61) is the increasing total generation cost. (4.62) represents that the overall demand still needs to be balanced. (4.63) is the line flow equation obtained by relaxing the constraint of the transmission line l by 1MW. Then, using (4.61) and (4.62),

$$\begin{aligned} \mu_l &\overset{(a)}{=} -C_i \Delta P_{g_i} - C_j \Delta P_{g_j} \\ &\overset{(b)}{=} -C_i \Delta P_{g_i} + C_j \Delta P_{g_i} \\ &= (C_j - C_i) \Delta P_{g_i} \end{aligned} \tag{4.64}$$

where (a) follows from the property that a single transmission line congestion yields two marginal units [83], thus setting the variable ΔP_i associated with any other unit to be zero, and (b) follows from (4.62). Similarly, (4.63) can be rewritten as

$$\begin{aligned} 1 &= H_{l,i} \Delta P_{g_i} + H_{l,j} \Delta P_{g_j} \\ &= H_{l,i} \Delta P_{g_i} - H_{l,j} \Delta P_{g_i} \\ &= (H_{l,i} - H_{l,j}) \Delta P_{g_i}. \end{aligned} \tag{4.65}$$

Finally, the combination of (4.64) and (4.65) provides the following desired result,

$$\mu_l = \frac{[C_j - C_i]}{[H_{l,i} - H_{l,j}]} = \frac{\Delta C(j,i)}{\Delta H_l(i,j)}. \tag{4.66}$$

□

Proposition 1 together with (A1)-(A3) implies the following corollaries.

Corollary 1. *Consider the situation under (A1)-(A3). Suppose that the line l is congested. Then, the LMP sensitivity index with respect to the line k status error ($k \neq l$) is written as*

$$\Delta \boldsymbol{\pi}_l^k = \Delta C(j,i) \mathbf{v}_l^k \tag{4.67}$$

where

$$\Delta \boldsymbol{\pi}_l^k = \left[\Delta \pi_{l,1}^k, \ldots, \Delta \pi_{l,N_b}^k\right]^T \tag{4.68}$$

$$\mathbf{v}_l^k = [v_{l,1}^k, \ldots, v_{l,N_b}^k]^T \tag{4.69}$$

$$v_{l,n}^k = \frac{\tilde{H}_{l,n}^k}{\Delta \tilde{H}_l^k(i,j)} - \frac{H_{l,n}}{\Delta H_l(i,j)}. \tag{4.70}$$

Proof. For simplicity, the shadow price corresponding to only a positive line congestion is considered in (4.59). Under assumption (A1)-(A2), the LMPs vectors without and with the line k status error are written as

$$\boldsymbol{\pi}_l = \lambda \mathbf{1}_{N_b} - \mu_l \mathbf{H}_l^T \tag{4.71}$$

$$\tilde{\boldsymbol{\pi}}_l^k = \tilde{\lambda}^k \mathbf{1}_{N_b} - \tilde{\mu}_l^k \tilde{\mathbf{H}}_l^{kT} \tag{4.72}$$

where $\mathbf{H}_l$ is the lth row vector of the distribution factor matrix $\mathbf{H}$. In (4.72), a tilde symbol over characters refers to topology error. Then, under assumption (A3) (i.e.,

$\lambda = \tilde{\lambda}^k$) the LMP sensitivity vector that illustrates the differences between LMPs without and with topology error is written as

$$\Delta\boldsymbol{\pi}_l^k = \boldsymbol{\pi}_l - \tilde{\boldsymbol{\pi}}_l^k = \tilde{\mu}_l^k \tilde{\mathbf{H}}_l^{kT} - \mu_l \mathbf{H}_l^T. \tag{4.73}$$

From Proposition 1, the shadow price corrupted by topology error is expressed as

$$\tilde{\mu}_l^k = \frac{[C_j - C_i]}{\left[\tilde{H}_{l,i}^k - \tilde{H}_{l,j}^k\right]} = \frac{\Delta C(j,i)}{\Delta \tilde{H}_l^k(i,j)}. \tag{4.74}$$

Finally, substituting (4.66) and (4.74) into (4.73),

$$\begin{aligned} \Delta\boldsymbol{\pi}_l^k &= \left[\frac{\Delta C(j,i)}{\Delta \tilde{H}_l^k(i,j)}\right] \tilde{\mathbf{H}}_l^{kT} - \left[\frac{\Delta C(j,i)}{\Delta H_l(i,j)}\right] \mathbf{H}_l^T \\ &= \Delta C(j,i) \left[\frac{\tilde{\mathbf{H}}_l^{kT}}{\Delta \tilde{H}_l^k(i,j)} - \frac{\mathbf{H}_l^T}{\Delta H_l(i,j)}\right] \\ &= \Delta C(j,i) \mathbf{v}_l^k. \end{aligned} \tag{4.75}$$

That is, the sensitivity of LMP at any bus to any line status error is written as the multiplication form of two independent functions which depend on: (i) the energy costs of marginal units; and (ii) congested line-related distribution factors at any bus and marginal units, respectively. □

Corollary 2. *For any buses n and m ($n \neq m$),*

(a) If $v_{l,n}^k > 0$, LMP at bus n with topology error decreases, otherwise it remains the same or increases.

(b) $|v_{l,n}^k| > |v_{l,m}^k|$ implies that LMP sensitivity at bus n is higher than at bus m.

(c) The increase (decrease) of $\Delta C(j,i)$ leads to the increase (decrease) of LMP sen-

sitivity at any bus.

By Corollary 2(a), buses can be categorized into three groups with positive, negative and zero sensitivities. This grouping property enables system operators to make a quick prediction for the direction of post-LMPs by topology error. By Corollary 2(b), economically sensitive buses to topology error can be identified through the comparison of $|v^k_{l,n}|$. Corollary 2(c) allows system operators to assess the impact of the energy costs of marginal units on LMP sensitivity. Furthermore, it may provide guidelines for making a bidding strategy of market participants such as generation company. Fig. 4.12 illustrates a linear relationship scaled by $\Delta C(j,i)$ between $\Delta\boldsymbol{\pi}^k_l$ and $\mathbf{v}^k_l$, as well as sensitivity grouping, identification of economically sensitive buses and impact of varying $\Delta C(j,i)$ on LMP sensitivity, all of which are mentiond in Corollary 2.

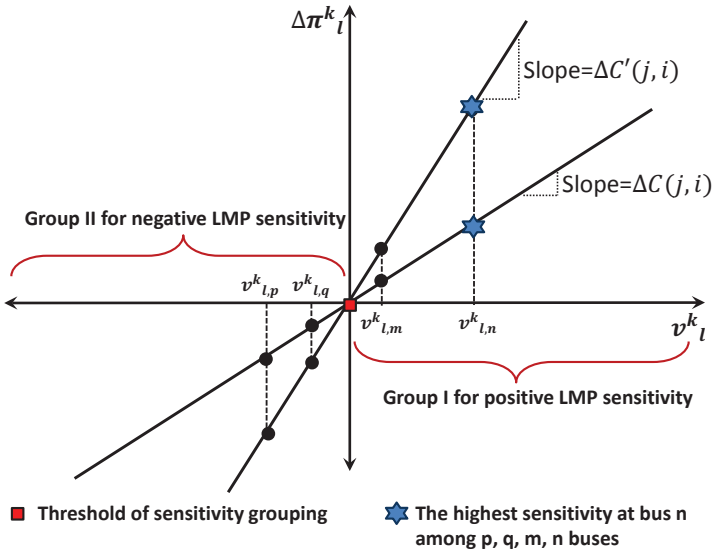


Figure 4.12: Illustration of a linear relationship between $\Delta\boldsymbol{\pi}^k_l$ and $\mathbf{v}^k_l$.

Remark 7. *Let us relax the assumption (A2). That is, topology error generates new marginal units p and q with cost functions C_p and C_q with $C_q > C_p$ and a new congested line $\tilde{l}$. Then, LMP sensitivity equations are categorized into four different cases as follows:*

Case 1: *both* unchanged *congestion pattern and marginal units (This case corresponds to Corollary 1)*

$$\Delta\boldsymbol{\pi}_l^k = \Delta C(j,i)\left[\frac{\widetilde{\mathbf{H}}_l^{kT}}{\Delta\widetilde{H}_l^k(i,j)} - \frac{\mathbf{H}_l^T}{\Delta H_l(i,j)}\right] \tag{4.76}$$

Case 2: unchanged *congestion pattern but* changed *marginal units*

$$\Delta\boldsymbol{\pi}_l^k = \Delta C(q,p)\left[\frac{\widetilde{\mathbf{H}}_l^{kT}}{\Delta\widetilde{H}_l^k(p,q)}\right] - \Delta C(j,i)\left[\frac{\mathbf{H}_l^T}{\Delta H_l(i,j)}\right] \tag{4.77}$$

Case 3: changed *congestion pattern but* unchanged *marginal units*

$$\Delta\boldsymbol{\pi}_{l,\tilde{l}}^k = \Delta C(j,i)\left[\frac{\widetilde{\mathbf{H}}_{\tilde{l}}^{kT}}{\Delta\widetilde{H}_{\tilde{l}}^k(i,j)} - \frac{\mathbf{H}_l^T}{\Delta H_l(i,j)}\right] \tag{4.78}$$

Case 4: *both* changed *congestion pattern and marginal units*

$$\Delta\boldsymbol{\pi}_{l,\tilde{l}}^k = \Delta C(q,p)\left[\frac{\widetilde{\mathbf{H}}_{\tilde{l}}^{kT}}{\Delta\widetilde{H}_{\tilde{l}}^k(p,q)}\right] - \Delta C(j,i)\left[\frac{\mathbf{H}_l^T}{\Delta H_l(i,j)}\right]. \tag{4.79}$$

The results in Remark 8 can be used to derive an analytical form for the profit through virtual bidding mechanism.

Remark 8. *LMP sensitivity index can be used to provide the following* system-wide *metrics that alert system operators to the kth most and jth least influential*

transmission line on LMP, on average, with respect to a total of N_b buses and T dispatch intervals:

$$k = \arg\max_{k} \left(\sum_{n=1}^{N_b} \sum_{t=1}^{T} \left| \Delta \boldsymbol{\pi}_{l,n}^{k,t} \right| / N_b T \right)$$
$$j = \arg\min_{j} \left(\sum_{n=1}^{N_b} \sum_{t=1}^{T} \left| \Delta \boldsymbol{\pi}_{l,n}^{j,t} \right| / N_b T \right).$$

and to the n_1th most and n_2th least influential bus on LMP, on average, with respect to a total of N_l transmission lines and T dispatch intervals

$$n_1 = \arg\max_{n_1} \left(\sum_{k=1}^{N_l} \sum_{t=1}^{T} \left| \Delta \boldsymbol{\pi}_{l,n_1}^{k,t} \right| / N_l T \right)$$
$$n_2 = \arg\min_{n_2} \left(\sum_{k=1}^{N_l} \sum_{t=1}^{T} \left| \Delta \boldsymbol{\pi}_{l,n_2}^{k,t} \right| / N_l T \right).$$

4.4.4 Simulation Results

In this subsection, we illustrate and verify the proposed analytical results in quantifying the impact of network topology errors on LMP in the IEEE 14-bus system. Fig. 4.13 shows the detailed bus-breaker model for the IEEE 14-bus system. In this figure, one scenario is illustrated where the misconfigured status of the circuit breaker at bus 5 leads to the (dotted) line 4-5 exclusion error as long as the line 5-6 is congested. It is assumed that this misconfiguration occurs due to a natural error or man-made attack and hence the corrupted network topology information is fed into economic dispatch module without being detected by topology error processing. Table 4.4 shows generator parameters in the IEEE 14-bus system.

Figs. 4.14 show the LMPs from the scenario illustrated in Fig. 4.13. Fig. 4.14(a) shows LMPs at all buses with and without the line exclusion error, respectively.

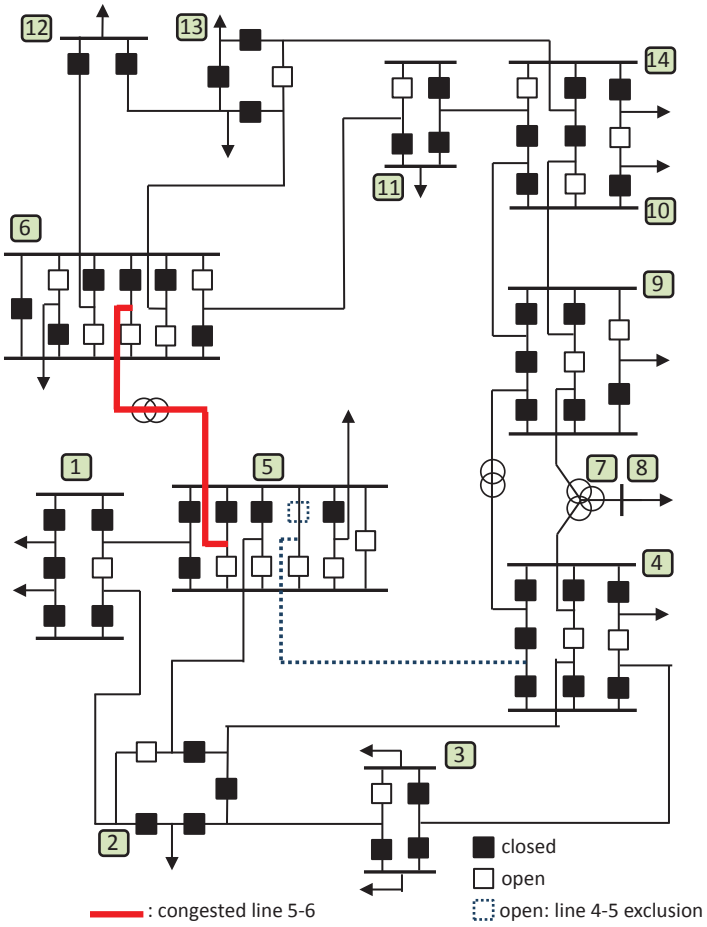
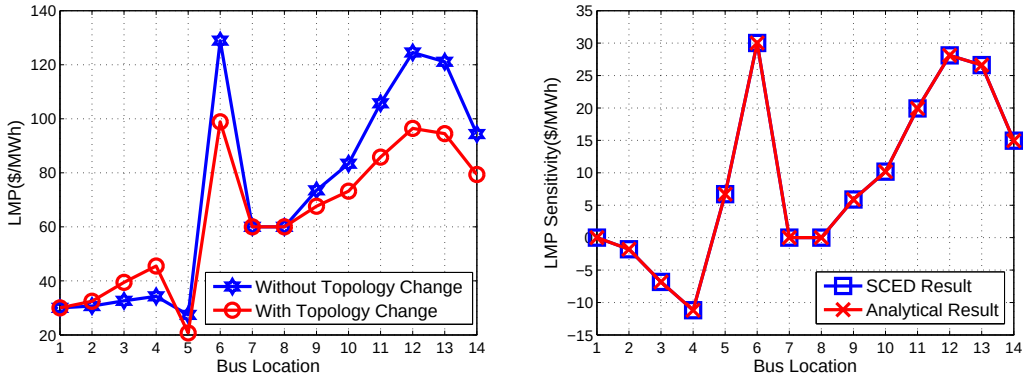


Figure 4.13: IEEE 14-bus system including bus-breaker model.

Table 4.4: Generator parameters of the IEEE 14-bus system.

Bus	$P_{\min}$	$P_{\max}$	Marginal Cost
1	0MW	330MW	30\$/MWh
2	0MW	140MW	20\$/MWh
3	0MW	100MW	40\$/MWh
6	0MW	100MW	55\$/MWh
8	0MW	100MW	60\$/MWh



(a) (b)

Figure 4.14: LMP results in Fig. 4.13: (a) comparison of LMPs between with and without line exclusion error; (b) comparison of LMP sensitivities obtained by SCED and the proposed approach.

These LMPs are obtained as the by-product of SCED formulated in Section 4.4.2. It should be noted in this scenario that the exclusion of the line 4-5 keeps both marginal units (at buses 1 and 8) and congestion pattern (the line 5-6 congestion) unchanged. Fig. 4.14(b) shows two LMP sensitivity plots for all buses with respect to the line 4-5 exclusion. Each plot is obtained using a different approach, which is based on SCED and the proposed analytical approach in Corollary 1. We emphasize again that in comparison with SCED approach the proposed approach computes LMP sensitivities using the derived sensitivity index without further economic redispatch. This could lead to reduced computational time compared with exhaustive numerical simulations. We can observe from Fig. 4.14(b) that the result of the proposed approach is consistent with that of SCED. This observation also holds true in other line exclusion cases under different network congestions. However, due to limited space, the validity of the test results for all other cases is not shown in this section. We note that LMP sensitivities in all subsequent figures are computed in the proposed approach.

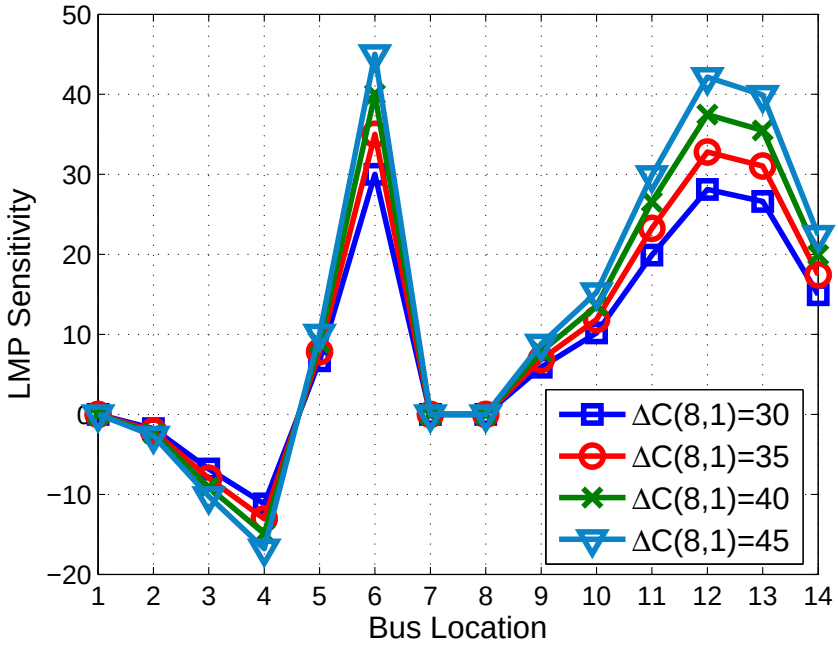


Figure 4.15: Impact of a varying gap between the energy costs of marginal units on LMP sensitivity.

Fig. 4.15 illustrates the impact of a varying gap between the energy costs of marginal units on LMP sensitivity. The results in this figure are based on the same system condition as in Fig. 4.13 so that marginal units are connected to buses 1 and 8, respectively. The energy cost of generator at bus 8 in Table 4.4 is assumed to increase from 60\$/MWh to 75\$/MWh with a step size of 5\$/MWh, thus changing the value of $\Delta C(8,1)$ from 30 to 45. We can observe from Fig. 4.15 that as the gap between the energy costs of marginal units increases, the absolute value of LMP sensitivity at any bus increases as well. This observation justifies Corollary 2(c).

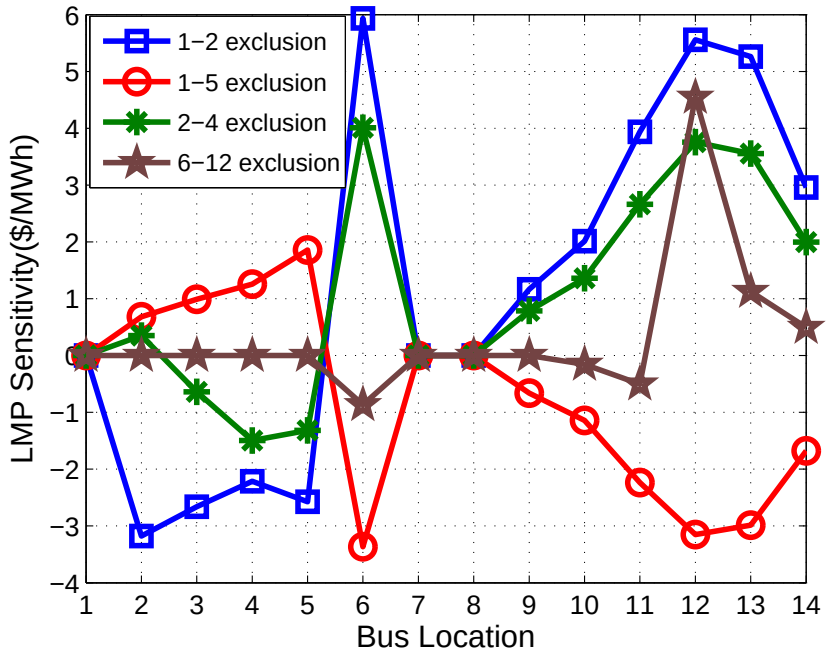


Figure 4.16: Comparison of LMP sensitivities with four different branch exclusion errors under the line 5-6 congestion.

Fig. 4.16 shows LMP sensitivities with four different line exclusion errors under

the identical congestion pattern where the line 5-6 is congested. For a clear comparison of sensitivities, we randomly choose four different lines (lines 1-2, 1-5, 2-4 and 6-12) out of twenty lines and then exclude each line from the network model to evaluate the impact of the line exclusion on LMP. First, we observe from Fig. 4.16 that LMP sensitivities at all buses with respect to the line 1-2 exclusion are higher than those with respect to other line exclusions. Therefore, the line 1-2 among chosen four lines has the most significant impact on LMP sensitivity at any bus. From a cybersecurity perspective, sensors collecting the status of CBs associated with the line 1-2 should be protected against bad data or malicious cyber attack with a high priority. Second, the most economically sensitive bus to topology error is identified in each line exclusion. For example, bus 6 has the highest sensitivity to the exclusion of the lines 1-2, 1-5 and 2-4 whereas bus 12 to the exclusion of the line 6-12. Lastly, we verify that buses are grouped according to the sign of sensitivity. For the line 1-2 exclusion, buses (6, 9∼14) obtain positive sensitivities, buses (2∼5) negative sensitivities and buses (1, 7∼8) have zero sensitivities. In particular, using the sign of sensitivity system operators are capable of predicting a market participant' profit or loss. For example, the sensitivities at generation bus 6 to the exclusions of lines 1-2 and 2-4 are positive so that post LMPs decrease, consequently providing a generation company a financial loss. On the other hand, since the sensitivities at the same bus to the exclusions of lines 1-5 and 6-12 are negative, a generation company makes a profit with increasing LMP.

Fig. 4.17 shows the sensitivities with respect to the line 4-5 exclusion under four different congestion patterns (the congestions of the lines 1-5, 2-4, 4-9 and 5-6). In this figure, the impact of different congested lines on LMP sensitivity is quantified for all buses. For example, it is observed that the line 2-4 congestion among the chosen four line congestions leads to the highest sensitivity at bus 4.

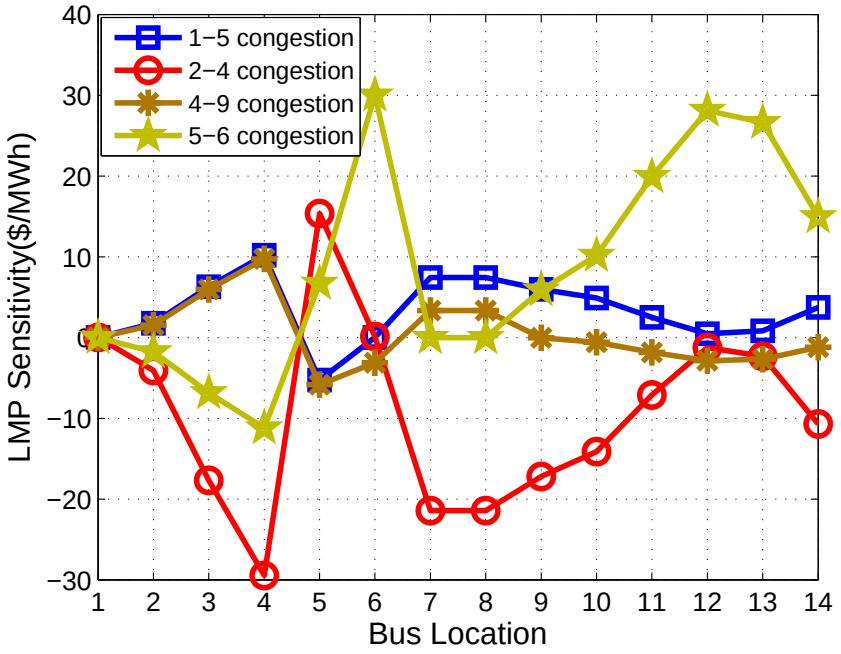


Figure 4.17: Comparison of LMP sensitivities with four different congestion patterns under the line 4-5 exclusion.

5. CONCLUSIONS

This dissertation studies the impact of power system sensor data quality on economic dispatch via state estimation in real-time power markets. This impact analysis is conducted from the perspectives of a market participant in Section 3 and a system operator in Section 4, respectively. In this section, we summarize the main contributions in the above two sections and also point out some potential future work.

5.1 Malicious Data Attack on Look-ahead Dispatch

5.1.1 Summary

Section 3 is concerned about temporal cyber data attacks on state estimation and their effects on time-coupled look-ahead dispatch. With the assumption of no network transmission congestion, from a market participant's perspective we formulate a malicious ramp-induced data (RID) attack problem in look-ahead dispatch. As illustrated in Figs. 3.3 and 3.4, we consider the scenarios where a potential attacker (e.g., generation company) could stealthily change the ramp constraint limits of generators through manipulating sensors' data without being detected by bad data detection, aiming at increasing the nodal price by withholding capacity of generator. This attack scenario is different from the traditional capacity withholding scenarios in which a generation company strategically reports capacity noticeably lower than its maximum capacity. The main result in Section 3 is an undetectable and profitable RID attack method in (3.40)–(3.43). Using malicious sensor data computed by this method, the attacker could manipulate the ramp constraint limits of generators for withholding generation capacity, subsequently leading to making a profit in the real-time power market. This proposed attack method is formulated in Linear Programming (LP) so that it can be solved efficiently using an off-the-shelf LP method

even in a large-scale system. Three attack cases based on the proposed method ((1) marginal unit attack, (2) binding unit attack, and (3) coordinated attack) are simulated and their performance is compared in the IEEE 14-bus system. Simulation results demonstrate the undetectability and profitability in three attack cases.

5.1.2 Future Work

In Section 3, we propose potential temporal data attacks against state estimation and economic dispatch and quantify their undetectability and profitability in real-time power markets. However, countermeasures to mitigate the financial risks from such data attacks are not developed in this dissertation work. To develop these countermeasures, there are possible directions in the future:

- **Robust PMU Placement Algorithm against Cyber Attacks**

Phasor measurement units (PMUs) are high-security sensors that provide fast, accurate, and time-stamped measurements that facilitate wide-area monitoring and control for large power systems [1]. The high cost of PMUs and corresponding communication facilities has been accelerating research into the optimal PMU placement problem with the minimum number of PMUs. However, most of the current research on this problem utilizes PMUs for only enhancing power system observability. We believe that robust PMU placement methods enable system operators to monitor, detect cyber attacks, and mitigate physical and economic attack risk. By applying observability-based PMU placement algorithms to the operational level of energy and market management systems, cyber-secure reslient power system and market operations can be assured. The proposed methods can be considered in conjunction with their impacts on the security aspects of communication in power control networks as well as the physical and economical operations in power systems. This could lead to fruitful collaboration with researchers in power system engineering and

communications.

• **Data Integrity-Resilient Multi-Area State Estimation**

Multi-area state estimation is becoming increasingly popular as the interconnected power system consists of multiple subsystems. An interconnected power system is operated by multiple control centers based on their administrative boundaries (e.g., New England, New York). However, malicious measurement data in one administrative area may contaminate the performance indices of other administrative areas due to the physical coupling among the subsystems. Therefore, it is important to assess which region may be affected by any given malicious data. The concept of error residual spread area (ERSA) is proposed to decompose the whole system into several non-overlapping regions, in which the corrupted data only contaminate measurements within each region. ERSA decomposition relies on network topology and measurement configuration. The proposed ERSA decomposition will localize and prevent data integrity attacks from spreading to a remote location. We plan to develop a new architecture and algorithm for attack-resilient multi-area state estimation based on ERSA decomposition. Research in this area will have great potential for interdisciplinary collaborations with researchers in power system engineering and statistical signal processing.

5.2 LMP Sensitivity Analysis to Data Corruption-Induced Estimation Error

5.2.1 Summary

In Section 4, from a system operator's perspective we present an analytical framework to quantify real-time LMP sensitivity subject to sensor data corruption via state estimation. Two data corruption scenarios are considered: in Section 4.3 corrupted continuous data (e.g., the power injection/flow and voltage magnitude) falsify power

flow estimates and in Section 4.4 corrupted discrete data (e.g., the on/off status of a circuit breaker) falsify network topology estimates.

In Section 4.3, we develop an analytical framework for calculating LMP sensitivity in response to small variations in SCADA and/or PMU continuous measurement data. Corrupted sensor data are shown to deviate power system state estimation from their actual values, which subsequently leads to the distortion of real-time market LMPs. The main results in Section 4.3 are two sensitivity matrices: the first with LMP sensitivity at any bus to any estimate in (4.35), (4.38), and the second with sensitivity of any estimate to data at any sensor in (4.42). A unified matrix that combines these two matrices in multiplication form enables system operators to quantify the impact on LMP of data at any sensor at any bus throughout the entire transmission network. Our simulation results suggest that the proposed sensitivity matrix can provide system operators with a quick and accurate method to identify the buses most vulnerable to measurement errors. In addition, we verify that more accurate sensors impact LMP much more significantly.

In Section 4.4, we examine the impact of circuit breaker-induced network topology errors on real-time LMP in electric power systems. The main result in Section 4.4 is Proposition 1, which states that the shadow price for the congested transmission line can be written as the ratio of the gap of the energy costs of marginal units to the gap of distribution factors that correspond to the intersections of marginal units and the congested transmission line. Using the result in Proposition 1, we derive an analytical index in Corollary 1 to compute LMP sensitivity with respect to network topology error, particularly line status error, in the power system. This index can be modified to three different cases according to the change of locations of marginal units and congested line after topology error occurs as shown in Remark 7. Similar to the sensitivity matrix derived in Section 4.3, the proposed sensitivity index also provides

system operators an analytical tool to identify economically sensitive transmission lines and circuit breakers, whose status error will significantly impact the real-time LMPs. The proposed sensitivity index is illustrated and tested in the IEEE 14-bus system.

5.2.2 Future Work

Built upon the preliminary work in Section 4, I plan to design new operational tools for future grid. This is achieved through engineering multi-scale control/communication systems. This research will be conducted along the following specific project:

- **Data Quality-Aware Multi-Scale Decision Making in Future Grid**

High-quality power grid data are essential to the control and management of large electric power systems in a reliable and economical manner. By leveraging the preliminary work on the impact analysis of data quality in Section 4, I plan to further develop a unified decision-making framework to investigate the impact of multi-scale spatial data quality at the transmission and distribution level on energy management system (EMS) and market management system (MMS) operations. Data in the transmission level refer to sensing for physical grid condition (e.g., power flow, voltage phasor, network topology), electricity market operation (e.g., market participants bidding price), and control command (e.g., open/close of circuit breakers, AGC signal). Data in the distribution level refer to smart metering (e.g., individual energy consumption), distributed renewable solar and wind generation, and energy storage/electric vehicle (e.g., location of sources, storage, and sinks of electric energy). Our prior work on two-level (EMS, MMS) sensitivity analysis based on a KKT condition perturbation approach will be extended to a three-level sensitivity approach with added distributed energy management system (DMS) level.

The proposed sensitivity index is the combination of the coupled partial derivatives quantifying the sensitivity of parameters in each system. This multi-level sensitivity approach may be extended to lower levels such as microgrids, building energy management system (BEMS) and home energy management system (HEMS). The developed framework can provide system operators with visualization tools on the level of multi-scale spatial data quality as well as a unified view on the impact of data quality on operations of EMS and MMS.

REFERENCES

[1] P. T. Myrda and K. Koellner, "NASPInet - The internet for Synchrophasors," in *Proceedings of the 43rd Hawaii International Conference on System Sciences*, Mar 2010.

[2] T. Zheng and E. Litvinov, "On ex post pricing in the real-time electricity market," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 153–164, Feb 2011.

[3] T. Zheng and E. Litvinov, "Ex post pricing in the co-optimized energy and reserve market," *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1528–1538, Nov 2006.

[4] F. Li, Y. Wei, and S. Adhikari, "Improving an unjustified common practice in ex post LMP calculation," *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp. 1528–1538, May 2010.

[5] D. Kundur, X. Feng, S. Liu, T. Zourntos, and K. Butler-Purry, "Towards a framework for cyber attack impact analysis of the electric smart grid," in *Proceedings of First IEEE Smart Grid Communication Conference*, Oct 2010.

[6] P. M. Esfahani, M. Vrakopoulou, K. Margellos, J. Lygeros, and G. Andersson, "Cyber attack in a two-area power system: Impact identification using reachability," in *Proceedings of American Control Conference*, pp. 962–967, June 2010.

[7] S. Sridhar and G. Manimaran, "Data integrity attacks and their impacts on SCADA control system," in *Proceedings of IEEE Power and Energy Society General Meeting*, pp. 1–6, July 2010.

[8] R. Anderson and S. Fuloria, "Who controls the off switch?," in *Proceedings of First IEEE Smart Grid Communication Conference*, Oct 2010.

[9] Y. Kim, E. C.-H. Ngai, and M. B. Srivastava, "Cooperative state estimation for preserving privacy of user behaviors in smart grid," in *Proceedings of Second IEEE Smart Grid Communication Conference*, Oct 2011.

[10] L. Sankar, S. Kar, R. Tandon, and H. V. Poor, "Competitive privacy in the smart grid: An information-theoretic approach," in *Proceedings of Second IEEE Smart Grid Communication Conference*, Oct 2011.

[11] S. R. Rajagopalan, L. Sankar, S. Mohajer, and H. V. Poor, "Smart meter privacy: A utility-privacy framework," in *Proceedings of Second IEEE Smart Grid Communication Conference*, Oct 2011.

[12] S. Wang, L. Cui, J. Que, D.-H. Choi, X. Jiang, S. Cheng, and L. Xie, "A randomized response model for privacy preserving smart metering," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1317–1324, Sept 2012.

[13] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber-physical system security for the electric power grid," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 210–224, Oct 2012.

[14] Y. Liu, M. K. Reiter, and P. Ning, "False data injection attacks against state estimation in electric power grids," in *Proceedings of the 16th ACM Conference on Computer and Communications Security*, Nov 2009.

[15] O. Kosut, L. Jia, R. Thomas, and L. Tong, "Malicious data attacks on smart grid state estimation: Attack strategies and Countermeasures," in *Proceedings of First IEEE Smart Grid Communication Conference*, Oct 2010.

[16] T. T. Kim and H. V. Poor, "Strategic protection against data injection attacks on power grids," *IEEE Transactions on Smart Grid*, vol. 3, no. 2, pp. 326–333, June 2011.

[17] A. Giani, E. Bitar, M. Garcia, M. McQueen, P. Khargonekar, and K. Poolla, "Smart grid data integrity attacks: Characterizations and Countermeasures," in *Proceedings of Second IEEE Smart Grid Communication Conference*, Oct 2011.

[18] O. Vuković and G. Dán, "On the security of distributed power system state estimation under targeted attacks," in *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, pp. 666–672, Mar 2013.

[19] A. Tajer, S. Kar, H. V. Poor, and S. Cui, "Distributed joint cyber attack detection and state recovery in smart grids," in *Proceedings of Second IEEE Smart Grid Communication Conference*, Oct 2011.

[20] M. Esmalifalak, H. A. Nguyen, R. Zheng, and Z. Han, "Stealth false data injection using independent component analysis in smart grid," in *Proceedings of Second IEEE Smart Grid Communication Conference*, Oct 2011.

[21] G. Hug and J. A. Giampapa, "Vulnerability assessment of ac state estimation with respect to false data injection cyber-attacks," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1362–1370, Sept 2012.

[22] L. Jia, R. J. Thomas, and L. Tong, "On the nonlinearity effects on malicious data attack on power system," in *Proceedings of IEEE Power and Energy Society General Meeting*, July 2012.

[23] M. Negrete-Pincetic, F. Yoshida, and G. Gross, "Towards quantifying the impacts of cyber attacks in the competitive electricity market environment," in *Proceedings of IEEE Bucharest Power Tech Conference*, June 2009.

[24] L. Xie, Y. Mo, and B. Sinopoli, "Integrity data attacks in power market operations," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 659–666, Dec 2011.

[25] L. Jia, R. J. Thomas, and L. Tong, "Malicious data attack on real-time electricity market," in *Proceedings of 2011 International Conference on Acoustics, Speech and Signal Processing*, pp. 5952–5955, May 2011.

[26] M. Esmalifalak, G. Shi, Z. Han, and L. Song, "Bad data injection attack and defense in electricity market using game theory study," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 160–169, Mar 2013.

[27] A. Giani, R. Bent, M. Hinrichs, M. McQueen, and K. Poolla, "Metrics for assessment of smart grid data integrity attacks," in *Proceedings of IEEE Power and Energy Society General Meeting*, July 2012.

[28] L. Xie, P. M. S. Carvalho, L. A. F. M. Ferreira, J. Liu, B. H. Krogh, N. Popli, and M. D. Ilić, "Wind integration in power systems: Operational challenges and possible solutions," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 1890–1908, Jan 2011.

[29] A. Ott, "Unit commitment in the PJM day-ahead and real-time markets," in *FERC Technical Conference on Increasing Market and Planning Efficiency Through Improved Software and Hardware*, Washington DC, June 2010.

[30] ERCOT, "Functional description of core market management system (MMS) applications for look-ahead SCED," *White paper*, 2011.

[31] CAISO, "Business Practice Manuals (BPM) Library: Market Operations, Version 11," Aug 2010.

[32] A. J. Conejo, E. Castillo, R. Mínguez, and F. Milano, "Locational marginal price sensitivities," *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 2026–2033, Nov 2005.

[33] F. Li and R. Bo, "DCOPF-based LMP simulation: Algorithm, comparison with ACOPF, and sensitivity," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1475–1485, Nov 2007.

[34] F. Li, "Continuous locational marginal pricing (CLMP)," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1638–1646, Nov 2007.

[35] R. Bo and F. Li, "Probabilistic LMP forecasing considering load uncertainty," *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1279–1289, Aug 2009.

[36] O. Kosut, L. Jia, R.-J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 645–658, Dec 2011.

[37] J. Kim and L. Tong, "On topology attack of a smart grid," in *2013 IEEE PES Innovative Smart Grid Technologies (ISGT)*, Washington, DC, Feb 2013.

[38] A. Ashok and M. Govindarasu, "Cyber attacks on power system state estimation through topology errors," in *Proceedings of IEEE Power and Energy Society General Meeting*, July 2012.

[39] F. F. Wu, "Power system state estimation: A survey," *International Journal of Electrical Power and Energy Systems*, vol. 12, no. 2, pp. 80–87, Apr 1990.

[40] C. N. Lu, J. H. Teng, and W.-H. E. Liu, "Distribution state estimation," *IEEE Transactions on Power Systems*, vol. 10, no. 1, pp. 229–240, Feb 1995.

[41] F. C. Schweppe, J. Wildes, and D. B. Rom, "Power system static state estimation, Parts I, II and III," *IEEE Transactions on Power Apparatus and Systems*, vol. 89, no. 1, pp. 120–135, Jan 1970.

[42] J. J. Allemong, L. Radu, and A. M. Sasson, "A fast and reliable state estimation algorithm for AEP's new control center," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, no. 4, pp. 933–944, Apr 1982.

[43] A. Abur and A. G. Expósito, *Power System State Estimation. Theory and Implementation.* New York: Marcel Dekker, 2004.

[44] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: numerical methods*. Nashua: Athena Scientific, 1997.

[45] Y. Saad, *Iterative Methods for Sparse Linear Systems*. Philadelphia: SIAM, 2003.

[46] T. V. Cutsem, J. L. Horward, and M. Ribbens-Pavella, "A two-level static state estimator for electric power systems," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-100, no. 8, pp. 3722–3732, Aug 1981.

[47] T. V. Cutsem and M. Ribbens-Pavella, "Critical survey of hierarchical methods for state estimation of electric power systems," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-102, no. 10, pp. 247–256, Oct 1983.

[48] G. N. Korres, "A distributed multiarea state estimation," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 73–84, Feb 2011.

[49] T. Yang, H. Sun, and A. Bose, "Transition to a two-level linear state estimator−Part I: Algorithm," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 46–53, Feb 2011.

[50] T. Yang, H. Sun, and A. Bose, "Transition to a two-level linear state estimator−Part II: Algorithm," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 54–62, Feb 2011.

[51] A. Gómez-Expósito and A. de la Villa Jaén, "Two-level state estimation with local measurement pre-processing," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 676–684, May 2009.

[52] A. Gómez-Expósito, A. de la Villa Jaén, C. Gómez-Quiles, P. Rousseaux, and T. V. Cutsem, "A taxonomy of multi-area state estimation methods," *Electric Power Systems Research*, vol. 81, no. 4, pp. 1060–1069, Apr 2011.

[53] A. Gómez-Expósito, A. Abur, A. de la Villa Jaén, and C. Gómez-Quiles, "A multilevel state estimatin paradigm for smart grids," *Proceedings of the IEEE*, vol. 99, no. 6, pp. 952–976, June 2011.

[54] D. M. Falcao, F. F. Wu, and L. Murphy, "Parallel and distributed state estimation," *IEEE Transactions on Power Systems*, vol. 10, no. 2, pp. 724–730, May 1995.

[55] L. Xie, D.-H. Choi, S. Kar, and H. V. Poor, "Fully distributed state estimation for wide-area monitoring systems," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1317–1324, Sept 2012.

[56] W. Jiang, V. Vittal, and G. T. Heydt, "A distributed state estimator utilizing synchronized phasor measurements," *IEEE Transactions on Power Systems*, vol. 22, no. 2, pp. 563–571, May 2007.

[57] L. Zhao and A. Abur, "Multiarea state estimation using synchronized phasor measurements," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 611–617, May 2005.

[58] T. A. Clements, G. R. Krumpholz, and P. W. Davis, "Power system state estimation residual analysis: an algorithm using network topology," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-100, no. 4, pp. 1779–1787, April 1981.

[59] H. N. Korres and G. C. Contaxis, "A reduced model for bad data processing in state estimation," *IEEE Transactions on Power Systems*, vol. 6, no. 2, pp. 550–557, May 1991.

[60] S. Y. Lin and C. H. Lin, "An implementable distributed state estimator and distributed bad data processing schemes for electric power systems," *IEEE Transactions on Power Systems*, vol. 9, no. 3, pp. 1227–1284, Aug 1994.

[61] G. M. Huang and J. Lei, "Measurement design of data exchange for distributed multi-utility operation," in *Proceedings of IEEE Power and Energy Society General Meeting*, Jan 2002.

[62] D.-H. Choi and L. Xie, "Fully distributed bad data processing for wide area state estimation," in *Proceedings of Second IEEE International Conference on Smart Grid Communications*, Oct 2011.

[63] R. L. Lugtu, D. F. Hackett, K. C. Liu, and D. D. Might, "Power system state estimation: Detection of topological errors," *IEEE Transactions on Power Systems*, vol. PAS-99, no. 6, pp. 2406–2412, Nov 1980.

[64] M. R. Irving and M. J. Sterling, "Substation data validation," *IEE Proceedings - Generation, Transmission and Distribution*, vol. 129, no. 3, pp. 119–122, May 1982.

[65] K. A. Clements and P. W. Davis, "Detection and identification of topology errors in electric power systems," *IEEE Transactions on Power Systems*, vol. 3,

no. 4, pp. 1748–1753, Nov 1988.

[66] F. F. Wu and W. H. E. Liu, "Detection of topology errors by state estimation," *IEEE Transactions on Power Systems*, vol. 4, no. 1, pp. 176–183, Aug 1989.

[67] A. Simões-Costa and J. A. Leao, "Identification of topology errors in power system state estimation," *IEEE Transactions on Power Systems*, vol. 8, no. 4, pp. 1531–1538, Nov 1993.

[68] A. Abur, H. Kim, and M. K. Çelik, "Identifying the unknown circuit breaker statuses in power networks," *IEEE Transactions on Power Systems*, vol. 10, no. 4, pp. 2029–2037, Nov 1995.

[69] A. Gómez-Expósito and A. Villa, "Reduced substation models for generalized state estimation," *IEEE Transactions on Power Systems*, vol. 16, no. 4, pp. 839–846, Nov 2001.

[70] A. Villa and A. Gómez-Expósito, "Implicitly constrained substation model for state estimation," *IEEE Transactions on Power Systems*, vol. 17, no. 3, pp. 850–856, Aug 2002.

[71] H. Li and L. Tesfatsion, "Capacity withholding in restructured wholesale power markets: An agent-based test bed study," in *Proceedings of Power System Conference and Exposition*, Mar 2009.

[72] A. Tellidou and A. Bakirtzis, "Agent-based analysis of capacity withholding and tacit collusion in electricity markets," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1735–1742, Nov 2007.

[73] F. F. Wu, P. Varaiya, P. Spiller, and S. Oren, "Folk theorems on transmission access: proofs and counterexamples," *Journal of Regulatory Economics*, vol. 10, no. 1, pp. 5–23, Jul 1996.

[74] W. W. Hogan, "Contract networks for electric power transmission," *Journal of Regulatory Economics*, vol. 4, no. 3, pp. 211–242, Sept 1992.

[75] T. A. Stuart and C. J. Herget, "A sensitivity analysis of weighted least squares state estimation for power systems," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-92, no. 5, pp. 1696–1701, Sep 1973.

[76] R. Mínguez and A. J. Conejo, "State estimation sensitivity analysis," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1080–1091, Aug 2007.

[77] A. L. Ott, "Experience with PJM market operation, sysem design, and implementation," *IEEE Transactions on Power Systems*, vol. 18, no. 2, pp. 528–534, May 2003.

[78] M. Zhou, V. A. Centeno, J. S. Thorp, and A. G. Phadke, "An alternative for including phasor measurements in state estimators," *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1930–1937, Nov 2006.

[79] T. S. Bi, X. H. Qin, and Q. X. Yang, "A novel hybrid state estimator for including synchronized phasor measurements," *Electric Power Systems Research*, vol. 78, no. 8, pp. 1343–1352, Feb 2008.

[80] G. Valverde, S. Chakrabarti, E. Kyriakides, and V. Terzija, "A constrained formulation for hybrid state estimation," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1102–1109, Aug 2011.

[81] K. Das, J. Hazra, D. P. Seetharam, R. K. Reddi, and A. K. Sinha, "Real-time hybrid state estimation incorporating SCADA and PMU measurements," in *Proceedings of Third IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies*, pp. 1–8, Oct 2012.

[82] E. Caro, A. J. Conejo, and R. Mínguez, "Power system state estimation considering measurement dependencies," *IEEE Transactions on Power Systems*, vol. 24, no. 4, pp. 1875–1885, Nov 2009.

[83] D. Kirschen and G. Strbac, *Fundamentals of Power System Economics*. New York: Wiley, 2004.