Sparsity and Coordination Constraints on Stealth Data Injection Attacks

Xiuzhen Ye

Department of Automatic Control and Systems Engineering

This dissertation is submitted for the degree of

Doctor of Philosophy

March 28, 2023

Abstract

In this thesis, data injection attacks (DIAs) to smart grid is studied from two perspectives: centralized and decentralized systems.

The fundamental limits of the data injection attacks are characterized by the information measures. Specifically, two metrics, mutual information and the Kullback-Leibler (KL) divergence, quantifies the disruption caused by the attacks and the corresponding stealthiness, respectively.

From the perspective of centralized system, a unique attacker constructs the attacks that jointly minimize the mutual information acquired from the measurements about the state variables and the KL divergence between the distribution of measurements with and without attacks. One of the main contributions in the centralized attack construction is the sparsity constraints. Two scenarios where the attacks between different locations are independent and correlated are studied, respectively. In independent attacks, the challenge of the combinatorial character of identifying the support of the sparse attack vector is circumvented by obtaining the closed-form solution to single measurement attack problem followed by a greedy construction that leverages the insight distilled. In correlated attacks, the challenge is tackled by incorporating an additional measurement that yields sequential sensor selection problem. The sequential procedure allows the attacker to identify the additional sensor first and character the corresponding covariances between the additional measurement and the compromised measurements. Following the studies on sparse attacks, a novel metric that describes the vulnerability of the measurements on smart grids to data integrity attacks is proposed. The new metric, coined vulnerability index (VuIx), leverages information theoretic measures to assess the attack effect on the fundamental limits of the disruption and detection tradeoff. The result of computing the VuIx of the measurements in the system yields an ordering of the measurements vulnerability based on the level of the exposure to data integrity attacks. The assessment on the measurements vulnerability of IEEE test systems observes that power injection measurements are overwhelmingly more vulnerable to data integrity attacks than power flow measurements.

From the perspective of decentralized system, the attack constructions are determined by a group of attackers in a cooperative manner. The interaction between the attackers is formulated as a game with a normal form. The uniqueness of the Nash Equilibrium (NE) is characterized in different games where the attackers have different objectives. Closed-form expression for the best response of the attackers in different games are obtained and followed by best response dynamics that leads to the NEs. The sparsity constraint is considered in decentralized system where the attackers have limited access to sensors. The attack construction with sparsity constraints in decentralized system is also formulated as a game with a normal form. The uniqueness of the NE and the closed-form expression for the best response are obtained.

Keywords: Data injection attacks, information theoretic security, sparsity, measurement vulnerability, decentralized attacks

Acknowledgements

I would like to express my deep gratitude and appreciation to my supervisory team, led by Dr Iñaki Esnaola and supported by Professor Robert F. Harrison and Professor Samir M. Perlaza for their support, guidance, and encouragement throughout my PhD study. I would also like to thank my parents for their understanding to my research. I am also grateful to my friends in general and in the University of Sheffield particularly for their inspiration and help both professionally and socially.

Declaration

I, Xiuzhen Ye, declare that the work presented in this thesis is my original research. All the results in this thesis that are not of my own work are properly accredited and referenced.

Xiuzhen Ye
Sheffield, September 20, 2023.

Contents

List of Figures

List of Tables

List of Symbols

Notation	Description
$\mathbb{R}^m$	real numbers of dimension m
$\mathbb{R}_+^m$	positive real numbers of dimension m
$\mathbb{Z}^m$	integer numbers of dimension m
$\mathbb{Z}_+^m$	positive integer numbers of dimension m
$\mathcal{S}_+^n$	the set of positive semi-definite matrices of dimension $n \times n$
$\mathcal{S}_{++}^n$	the set of positive definite matrices of dimension $n \times n$
A^m	random vector of dimension m
$\mathbf{a}$	realization of A^m
A_i	the component of vector A^m in position i
$\mathbf{0}$	vector with all components are zero
$\mathbf{1}$	vector with all components are one
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$
$\mathbf{I}_m$	identical matrix with size $m \times m$
$\mathrm{tr}\,(\mathbf{I}_m)$	trace of the matrix $\mathbf{I}_m$
$\mathbf{H} \in \mathbb{R}^{m \times n}$	matrix with dimension $m \times n$
$\mathbf{H}^\intercal$	transpose of matrix $\mathbf{H}$
$\mathrm{rank}\,(\mathbf{H})$	rank of the matrix $\mathbf{H}$
$(\mathbf{H})_{ij}$	the entry of matrix $\mathbf{H}$ in the i-th row and the j-th column
$\hat{\mathbf{x}}$	estimate of $\mathbf{x}$
$\|A^m\|_{\ell_2}$	the ℓ_2 norm of vector A^m
$\mathcal{H}_0/\mathcal{H}_1$	null hypothesis/alternative hypothesis
$\mathcal{X}_k^2$	chi-squared distribution with k degrees of freedom
$\mathbb{E}\,[X]$	expectation of X

List of Symbols

Notation	Description
P_{D}	probability of attack detection
$\mathbb{1}_{\{\cdot\}}$	the indicator function for event given in $\{\cdot\}$
α	probability of Type-I error in hypothesis testing
β	probability of Type-II error in hypothesis testing
$\mathrm{card}\,(\mathcal{A})$	the cardinality of a set $\mathcal{A}$
$\mathrm{supp}(\mathbf{a})$	support of the vector $\mathbf{a}$
$\|\mathbf{a}\|_{\ell_0}$	the ℓ_0 norm of the vector $\mathbf{a}$
$\|\mathbf{a}\|_{\ell_2}$	the ℓ_2 norm of the vector $\mathbf{a}$
$\mathbf{e}_i$	elementary vector with a one in position i
$\mathrm{diag}(\mathbf{A})$	the vector formed by the diagonal entries of the matrix $\mathbf{A}$
$\|\mathbf{A}\|$	determinant of the matrix $\mathbf{A}$
$\lambda_i(\mathbf{A})$	the i-th eigenvalue of the matrix $\mathbf{A}$ in descending order
$X^n \bigoplus Y^n$	Minkowski sum of X^n and Y^n
$X^n \bigotimes Y^n$	Kronecker product of X^n and Y^n
BR^p	best response correspondence for the game $\mathcal{G}_p$
$\mathrm{abs}\,(x)$	absolute value of a number x

Abbreviations

AWGN	Additive White Gaussian Noise
AC	Alternating Current
BRD	Best Response Dynamics
CS	Compressed Sensing
DIAs	Data Injection Attacks
DC	Direct Current
EMS	Energy Management System
FC	Frequency Control
KL divergence	Kullback-Leibler divergence
LRT	Likelihood Ratio Test
LS	Least Squares
LMP	Locational Marginal Price
LNRT	Largest Normalized Residual Test
MIMO	Multiple-Input Multiple-Output
MSE	Mean Square Error
MMSE	Minimum Mean Square Error
NE	Nash Equilibrium
NEs	Nash Equilibriums
NTP	Network Topology Processor
OPF	Optimal Power Flow
pmf	Probability Mass Function
pdf	Probability Density Function
PFC	Power Flow Control
RT	Residual Test
SG	Smart Grid
SCADA	Supervisory Control and Data Acquisition
STCPA	the state and topology cyber-physical attack
TEP	Topology Error Processing
V2G	Electric Vehicle to Grid
VuIx	Vulnerability index
WLS	Weighted Least Squares

Chapter 1

Introduction

1.1 Background and Motivation

Power system is a critical infrastructure for the functioning of both industry and daily activities. The traditional electricity grids have been upgraded into smart grids (SG) by monitoring and control processes that are supported by Supervisory Control and Data Acquisition (SCADA) systems and more recently by advanced communication and control technologies.

The increasing interconnectivity between communications, control systems, and smart grids gives rise to numerous benefits. While the implementation of advanced communication and control procedures improves system operation, this cyber layer exposes the systems to malicious attacks that exploit the vulnerabilities of the sensing and communication infrastructure.

One of the main threats faced by smart grid is data injection attacks (DIAs) that alter the state estimate of the system obtained from different estimation methods by compromising the system measurements without triggering bad data detection mechanisms set by the system operator [1–3]. State estimate is the key in the decision on optimal power flow (OPF), frequency control (FC) and energy management system (EMS), etc. [4] Hence, compromised measurements fed to the state estimation (SE) damage efficient, scalable, and secure operation of smart grids [5].

A large body of literature studies the case in which attack detection is performed by a residual test (RT) [6] under the assumption that state estimation is deterministic both in centralized and decentralized scenarios [7–10]. In this setting, attack construction that requires access to a small set of measurements yields l_0-norm minimization problems, which are in general hard to solve. In [11], it is shown that the operator can secure a small fraction of measurements to make undetectable attack constructions significantly harder.

The unprecedented data acquisition capabilities facilitate the efficient operation of the smart grid but also increase the threats posed by DIAs given the fact that accurate stochastic models of the system can be generated. This problem is cast in a Bayesian framework in [12]. In this Bayesian paradigm, the attack detection can be formulated as the likelihood ratio test (LRT) [13]. Alternatively, machine learning methods [14] can be employed to learn the geometry of the data generated by the systems. Data analytics are increasingly important in the operation of smart grid and they are the central to advanced estimation, control,

and management of the smart grid [15]. For this reason, it is essential to study attack constructions in fundamental terms to understand the impact over a wide range of data analysis paradigms.

Stealth data injection attacks within Bayesian framework were first introduced in [16] and then generalized in [17]. In this research, the attack construction uses information theoretic measures: (a) the mutual information between the state variables and the measurements under attacks; and (b) Kullback-Leibler (KL) divergence between the distributions of measurements with attacks and without attacks. The rationale of measuring the disruption of the attack in terms of mutual information stems from the fact that it characterizes in fundamental terms the amount of information, understood as evidence, collected by the observations about the state variables. That being the case, by minimizing the mutual information the attacker limits the information that the measurements contain about the state variables, and ultimately, disrupts the state estimation in a fundamental sense [18]. The rationale for minimizing the KL divergence between the distributions as means to minimize the probability of attack detection stems from the Chernoff-Stein Lemma [19, Th. 11.8.3]. Within this framework, the attack is constructed with probability distribution function that jointly minimizes the the mutual information and KL divergence with a weighting parameter that governs the tradeoff between these two objectives.

The state variables are assumed to follow a Gaussian distribution in [12,16,17,20]. From a practical point of view, the adoption of Gaussian random vectors as the data injection attack vectors is validated given the data shared by Electricity North West Limited [21,22]. However, both the stealth attacks constructed in [16] and [17] require that the attacker tampers with all the measurements in the system, which is not feasible in most scenarios. Information theoretic attack constructions that incorporate sparsity constraints and the construction that effectively exploits the correlation between attack variables are still an open problem that requires novel approaches.

Apart from the *centralized attacks* where there is only one attacker, in *decentralized attacks* with multiple attackers operating over a larger number of processes poses the framework for the exploration of game theoretic techniques [23]. A comprehensive description of existing game theoretic applications in smart grids is given in [24]. From the perspective of DIAs, in [25], centralized data injection attacks are studied in a game theoretic setting in which the operator performs least square (LS) estimation. Decentralized attack construction with interaction between several attackers is studied with a game formulation in a normal form where the utility captures the main objectives in attack constructions. The game formulation results in a potential game where the existence of a Nash Equilibrium (NE) is claimed and the convergence of best response dynamics (BRD) to a NE. However, the case in which attackers disrupt the state estimation process in an uncoordinated way is still not well understood. Furthermore, the impact of making the statistical structure of the state variables available to attackers in decentralized settings has not been studied either.

1.2 Contributions

The following are the main contributions of this thesis:

(1) An information theoretic independent sparse attack constructions. The attack constructions assume that the attack vector consists of independent entries, and therefore, requires no communication between different attacked locations. This thesis proposed a cost function that combines the mutual information and the KL divergence that is amenable to sparse attack constructions. This thesis theoretically characterized the single observation attack case by proving that the resulting cost function is convex and obtaining the optimal attack construction for this case. This thesis distilled the insight obtained from the single measurement attack case to propose a k-sparse attack via a greedy algorithm that overcomes the combinatorial challenge posed by the sensor selection problem [20].

(2) An information theoretic correlated sparse attack constructions. The attack constructions are formulated as the design of a multi-objective optimization problem that aims to minimize the mutual information while limiting the Kullback-Leibler divergence. A heuristic greedy algorithm for the correlated attack construction are proposed where correlation between the attack vector entries results in larger disruption and smaller probability of detection at the expense of coordination between different locations [26].

(3) A novel metric that describes the vulnerability of the measurements in smart grid to data integrity attacks. The new metric, coined vulnerability index (VuIx), leverages information theoretic measures to assess the attack effect on the fundamental limits of the disruption and detection tradeoff. The result of computing the VuIx of the measurements in the system yields an ordering of their vulnerability based on the level of exposure to data integrity attacks. This new framework is used to assess the measurement vulnerability of IEEE 9-bus and 30-bus test systems and it is observed that power injection measurements are overwhelmingly more vulnerable to data integrity attacks than power flow measurements. A detailed numerical evaluation of the VuIx values for IEEE test systems is provided [27].

(4) Decentralized stealth attack constructions with coordination between the attackers. The objectives of the attacks are to minimize the mutual information between the state variables and measurements while constraining the Kullback-Leibler divergence between the distribution of the measurements under attacks and the distribution of the measurements without attacks. The attack constructions are formulated as random Gaussian attacks. The proposed information metrics adopted measure the disruption and attack detection both globally and locally. The decentralized attack constructions are formulated in a framework of normal games. The global and local information metrics yield games with global and local objectives in disruption and attack detection. This thesis proved the games are potential games and the convexity of the potential functions followed by the uniqueness and the achievability of the Nash Equilibrium, accordingly. This thesis proposed a best response dynamics to achieve the Nash Equilibrium of the games. This thesis numerically evaluates the performance of the proposed decentralized stealth random attacks on IEEE test systems and show it is feasible to exploit coordination with game theoretic techniques in decentralized attack constructions.

(5) Decentralized stealth attack constructions with coordination and sparsity constraints. Specifically, the attack constructions are formulated as random Gaussian attacks that minimize the mutual information between the state variables and the measurements while constraining KL divergence between the distribution of the measurements under attacks and the distribution of the measurements without attacks. The sparsity constraints

limit the number of measurements that are potentially compromised. This thesis assumes each attacker has access to a set of measurements and the sets of measurements form a partition of the set of measurements in the systems. The attackers minimize the information theoretic cost of launching a random attack to one of the measurements that it has access to in a coordinated fashion. The decentralized sparse attacks with partition is modelled in a game form that yields a potential game. The uniqueness and achievability of the Nash Equilibrium in the game are obtained. A best response dynamics is proposed to achieve the NE.

1.3 Outline

This thesis is divided into five parts as follows:

• **Part 1**. These two chapters describe the system model of classical DIAs and attack detection methods as well as DIAs that is mainly studied in this thesis. It also establishes the difference between centralized and decentralized systems.

- **Chapter 2**. This chapter presents the mathematical formulation of the system model and establishes the classical DIAs and attack detection methods. This chapter also presents the system model of DIAs and the corresponding optimal attack detection.

- **Chapter 3**. This chapter presents main results on DIAs in the literature in centralized systems and decentralized systems.

• **Part II**. These three chapters present the main results on DIAs with sparsity constraints and the analysis on measurement vulnerability in centralized systems.

- **Chapter 4**. This chapter presents the main results for independent sparse DIAs.

- **Chapter 5**. This chapter presents the main results for correlated sparse DIAs.

- **Chapter 6**. This chapter presents the main results for the analysis on measurement vulnerability.

• **Part III**. These two chapters develop the interaction between multiple attackers with a game framework in decentralized systems.

- **Chapter 7**. This chapter presents the main results for stealth DIAs in a decentralized system.

- **Chapter 8**. This chapter presents the main results for stealth DIAs with sparsity constraints in a decentralized system.

• **Part IV**.

- **Chapter 9** summaries the conclusions and future work of this thesis.

• **Part V**. The Appendix contains fundamental concepts on information theory that are used along this thesis and the proofs of the main results from Chapter 4 to Chapter 8.

- **Appendix A**. This appendix contains the analytic expression of mutual information between two random variables with Gaussian distributions.

- **Appendix B**. This appendix contains the analytical expression of KL divergence between two random variables with Gaussian distributions.

- **Appendix C**. This appendix contains the analytical solution of optimal single sensor attack construction.

- **Appendix D**. This appendix contains the analytical expression of the difference between the information theoretic cost with different covariance matrix of the random attack vector.

- **Appendix E**. This appendix contains the proof of the convexity of the equivalent optimization problem with respect to the variance of a random attack variable in the sequential sensor selection procedure.

- **Appendix F**. This appendix contains the analytical expression of the optimal variance of a random attack variable in the sequential sensor selection procedure.

- **Appendix G**. This appendix contains the analytical expression of the mutual information between a n-dimensional random vector with Gaussian distribution and a one dimension random variable with Gaussian distribution.

- **Appendix H**. This appendix contains the analytical expression of the KL divergence between two one dimensional random variables with Gaussian distributions.

- **Appendix I**. This appendix contains the proof of the convexity of the cost function in game $\mathcal{G}_3$.

- **Appendix J**. This appendix contains the equivalent cost function for one attacker in game $\mathcal{G}_1$.

- **Appendix K**. This appendix contains the equivalent cost function for one attacker in game $\mathcal{G}_2$.

- **Appendix L**. This appendix contains the equivalent cost function for one attacker in game $\mathcal{G}_3$.

- **Appendix M**. This appendix contains the analytical solution of the best response for one attacker in game $\mathcal{G}_1$.

- **Appendix N**. This appendix contains the analytical solution of the best response for one attacker in game $\mathcal{G}_2$.

- **Appendix O**. This appendix contains the analytical solution of the best response for one attacker in game $\mathcal{G}_3$.

1.4 Disseminated Results

The results from this research are disseminated in the following:

- **Journals**

  - **X. Ye**, I. Esnaola, S. M. Perlaza, R. F. Harrison, "Stealth Data Injection Attacks with Sparsity Constraints". IEEE Transaction on Smart Grid(Early Access), 2023.

- **Conferences**

  - **X. Ye**, I. Esnaola, S. M. Perlaza, R. F. Harrison, "Information theoretic data injection attacks with sparsity constraints", in *Proc. 2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, virtual conference, Nov. 11 - 13, 2020, pp. 1-6.

- **Preprints and drafts**

  - **X. Ye**, I. Esnaola, S. M. Perlaza, R. F. Harrison, "An information theoretic vulnerability metric for data integrity attacks on smart grids". This work has been submitted to IET Smart Grid on Nov. 4, 2022.

  - **X. Ye**, I. Esnaola, S. M. Perlaza, R. F. Harrison, "Decentralized Data Injection Attacks on Cyber-physical systems". To be submitted to IEEE Transactions on Information Forensics and Security.

- **INRIA Technical Reports**

  - **X. Ye**, I. Esnaola, S. M. Perlaza, R. F. Harrison, "Stealth Data Injection Attacks with Sparsity Constraints", Technical Report, Inria, Centre de Recherche de Sophia Antipolis Méditérranée, Sophia Antipolis, Sep., 2022.

- **Oral Presentation**

  - **X. Ye**, I. Esnaola, S. M. Perlaza, R. F. Harrison, "Information theoretic data injection attacks with sparsity constraints", in *Proc. 2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, virtual conference, Nov., 2020.

  - **X. Ye**, I. Esnaola, "Stealth Data Injection Attacks with Sparsity Constraints", in *ACSE PGR Symp. (Departmental PhD Symp.)*, Sheffield, UK, Mar. 2021.

- **Poster Presentation**

- **X. Ye**, I. Esnaola, S. M. Perlaza, R. F. Harrison, "Stealth Data Injection Attacks with Sparsity Constraints", in *ACSE PGR Symp. (Departmental PhD Symp.)*, Sheffield, UK, Mar. 2021. Best Poster Award.

Chapter 2

Data Injection Attacks

This chapter introduces the system model and the data injection attacks (DIAs) in smart grid. Section 2.1 introduces the mathematical formulation of a power system. Section 2.2 presents the classical DIAs and the attack detection procedures. The chapter concludes with Section 2.3, which focuses on DIA constructions in a Bayesian framework.

2.1 Mathematical Formulation

2.1.1 Observation Model

Fig. 2.1 depicts a general two-port π-model for the branches in the considered IEEE test systems. Note that $g_{ij} + jb_{ij}$ is the admittance of the series branch connecting bus i and bus j and $g_{si} + jb_{si}$ is the admittance of the shunt branch connected at bus i, the measurements of the power system in (2.3) can be expressed in terms of the vector of the state variables in (2.4). Specifically, the active power injection and reactive power injection at bus i are

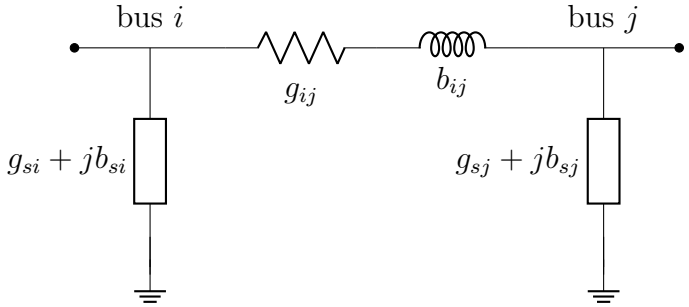

Figure 2.1: Two-port π-model of a network branch.

$$P_i = V_i \sum_{j \in \mathcal{N}_i} V_j (G_{ij} \cos\theta_{ij} + B_{ij} \sin\theta_{ij}), \tag{2.1a}$$

$$Q_i = V_i \sum_{j \in \mathcal{N}_i} V_j (G_{ij} \sin\theta_{ij} - B_{ij} \cos\theta_{ij}), \tag{2.1b}$$

respectively [4], where $\theta_{ij} \triangleq \theta_i - \theta_j$ and $G_{ij} + jB_{ij}$ is the entry of the complex bus admittance matrix in the i-th row and the j-th column. The active power flow and reactive power flow from bus i to bus j are

$$P_{ij} = V_i^2(g_{si} + g_{ij}) - V_iV_j(g_{ij}\cos\theta_{ij} + b_{ij}\sin\theta_{ij}), \tag{2.2a}$$
$$Q_{ij} = -V_i^2(b_{si} + b_{ij}) - V_iV_j(g_{ij}\sin\theta_{ij} - b_{ij}\cos\theta_{ij}), \tag{2.2b}$$

A standard topology of the IEEE test system is formed by general two-port branch networks as in Fig. 2.1. Specifically, Figure. 2.2 depicts the standard topology of the IEEE 14 bus test system where there are 14 buses in the system. Any two physically connected buses, e.g., bus i and bus j, form a Two-port π-model of a network branch in Fig. 2.1. The measurements from bus i are the active power injection P_i and reactive power injection Q_i as in (2.1) as well as the active power flow P_{ij} and reactive power flow Q_{ij} from bus i to bus j as in (2.2). The entries of the measurement vector of the system denoted as Y^m are the active power

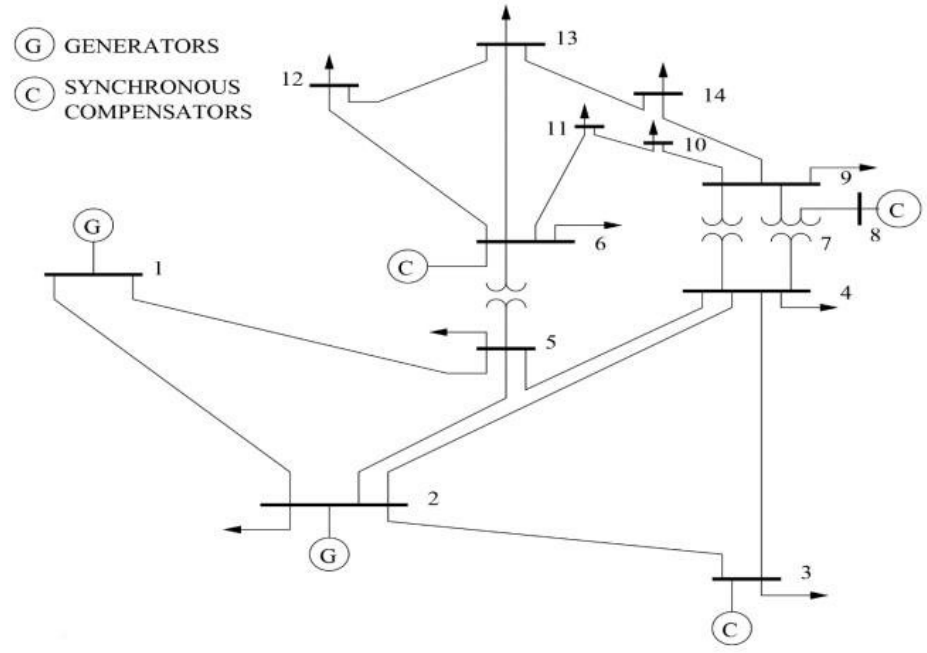

Figure 2.2: Topology of the IEEE 14 bus test system.

injection, reactive power injection, active power flow and reactive power flow from all the buses. To specify the vector of the measurements Y^m, consider the following definition.

Definition 1 (Vector of measurements)**.** *Consider a power system with $N \in \mathbb{Z}_+$ buses. Let $\mathcal{N}_i$ be the set of buses that are physically connected to bus i and $Y^m \in \mathbb{R}^m$ be the vector of measurements in the system such that*

$$Y^m \triangleq [\ldots, P_{ij}, \ldots, Q_{ij}, \ldots, P_1, P_2, \ldots, P_N, Q_1, Q_2, \ldots, Q_N]^\mathsf{T}, \tag{2.3}$$

where $P_i \in \mathbb{R}$ and $Q_i \in \mathbb{R}$, with $i \in \{1, 2, \ldots, N\}$, are the active power injection and reactive power injection at bus i, respectively; and P_{ij} and Q_{ij}, with $i \in \{1, 2, \ldots, N\}$ and $j \in \mathcal{N}_i$, are the active power flow and the reactive power flow from bus i to bus j, respectively.

Note that $P_i \in \mathbb{R}_+$ and $Q_i \in \mathbb{R}_+$ imply that bus i consumes active power and reactive power from the main grid, respectively [4]. Note that $P_i \in \mathbb{R}_-$ and $Q_i \in \mathbb{R}_-$ imply that bus i injects active power and reactive power to the main grid, respectively [4]. Similarly, $P_{ij} \in \mathbb{R}_+$ and $Q_{ij} \in \mathbb{R}_+$ imply that bus i transfers active power and reactive power to bus j, respectively [4]. Note that $P_{ij} \in \mathbb{R}_-$ and $Q_{ij} \in \mathbb{R}_-$ imply that bus i takes active power and reactive power from bus j, respectively [4]. The measurement vector Y^m is transmitted to estimate the state variables of the system denoted as $\mathbf{x}$. To specify the vector of the state variables of a power system, consider the following definition.

Definition 2 (Vector of state variables)**.** *Consider a power system with $N \in \mathbb{Z}_+$ buses and assume bus 1 is chosen as the reference bus such that the phase angle of bus 1 is set to the arbitrary value zero. Let $\mathbf{x} \in \mathbb{R}^n$ be the vector of state variables of the system such that*

$$\mathbf{x} \triangleq [\theta_2, \theta_3, \ldots, \theta_N, V_1, V_2, \ldots, V_N]^\mathsf{T}, \tag{2.4}$$

where $\theta_i \in [-\pi, \pi)$ and $V_i \in \mathbb{R}$, with $i \in \{1, 2, \ldots, N\}$, are the phase angle and voltage magnitude of bus i, respectively.

Note that there are $2N - 1$ state variables in a N bus system where N variables are bus voltage magnitudes and $N - 1$ variables are phase angles.

In general, the observation model in which the operation state of a power system is described by a state vector $\mathbf{x} \in \mathbb{R}^n$ and observed through the acquisition function $F\colon \mathbb{R}^n \to \mathbb{R}^m$ such that

$$Y^m = F(\mathbf{x}) + Z^m, \tag{2.5}$$

where $Y^m \in \mathbb{R}^m$ is the random vector of measurements provided by the *Supervisory Control and Data Acquisition* (SCADA) system and corrupted by *additive white Gaussian noise (AWGN)* $Z^m \in \mathbb{R}^m$ introduced by the sensors, c.f., [4, 5]. The noise is described by the random vector $Z^m \triangleq (Z_1, Z_2, \ldots, Z_m)^\mathsf{T} \in \mathbb{R}^m$ in (2.5) such that [4, 5]

$$Z^m \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_m), \tag{2.6}$$

where $\mathbf{0} = (0, 0, \ldots, 0)^\mathsf{T}$ and $\sigma^2 \in \mathbb{R}_+$ is the variance of the noise. For all $i \in \{1, 2, \ldots, m\}$, the random variable Z_i that corresponds to the noise added to measurement i satisfies $Z_i \sim \mathcal{N}(0, \sigma^2)$.

2.1.2 Observation Model with Linearized Dynamics

The observation model $Y^m = F(\mathbf{x}) + Z^m$ in (2.5) denotes the relationship between the measurements Y^m given in (2.3) and the state variables of the system $\mathbf{x}$ in (2.4). Note that the relationships between power and the state of the buses in the system given in (2.1) and (2.2) are nonlinear. The case where the nonlinearity is considered results in *Alternating Current (AC)* model as described in Section 2.1.1. This work linearizes the nonlinear observation functions $F\colon \mathbb{R}^n \to \mathbb{R}^m$ at a valid operation point that leads to significantly simplified linearized observation model. Let $\bar{\mathbf{x}}$ be one of the valid operation points. The Jacobian matrix of the observation model denoted by $\mathbf{H} \in \mathbb{R}^{m \times n}$ at operation point $\bar{\mathbf{x}} \in \mathbb{R}^n$ is

$$\mathbf{H}_{\mathbf{x}=\bar{\mathbf{x}}} \triangleq \frac{\partial}{\partial \mathbf{x}} F(\mathbf{x})|_{\mathbf{x}=\bar{\mathbf{x}}}. \tag{2.7}$$

Specifically, for the vector of measurements Y^m in (2.3) and the vector of the state variables $\mathbf{x}$ in (2.4), the Jacobian matrix is given by

$$\mathbf{H}_{\mathbf{x}=\bar{\mathbf{x}}} = \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}_1} P_{ij} & \frac{\partial}{\partial \mathbf{x}_2} P_{ij} & \cdots & \frac{\partial}{\partial \mathbf{x}_n} P_{ij} \\ \ddots & \ddots & \ddots & \ddots \\ \frac{\partial}{\partial \mathbf{x}_1} Q_{ij} & \frac{\partial}{\partial \mathbf{x}_2} Q_{ij} & \cdots & \frac{\partial}{\partial \mathbf{x}_n} Q_{ij} \\ \ddots & \ddots & \ddots & \ddots \\ \frac{\partial}{\partial \mathbf{x}_1} P_i & \frac{\partial}{\partial \mathbf{x}_2} P_i & \cdots & \frac{\partial}{\partial \mathbf{x}_n} P_i \\ \ddots & \ddots & \ddots & \ddots \\ \frac{\partial}{\partial \mathbf{x}_1} Q_i & \frac{\partial}{\partial \mathbf{x}_2} Q_i & \cdots & \frac{\partial}{\partial \mathbf{x}_n} Q_i \end{bmatrix}_{|\mathbf{x}=\bar{\mathbf{x}}}, \tag{2.8}$$

where $\bar{\mathbf{x}} \triangleq (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$. Particularly, from (2.4), the Jacobian matrix at the operation point $\bar{\mathbf{x}}$ is

$$\mathbf{H}_{\mathbf{x}=\bar{\mathbf{x}}} = \begin{bmatrix} \frac{\partial}{\partial \theta_c} P_{ij} & \cdots & \cdots & \frac{\partial}{\partial V_l} P_{ij} \\ \ddots & \ddots & \ddots & \ddots \\ \frac{\partial}{\partial \theta_c} Q_{ij} & \cdots & \cdots & \frac{\partial}{\partial V_l} Q_{ij} \\ \ddots & \ddots & \ddots & \ddots \\ \frac{\partial}{\partial \theta_c} P_i & \cdots & \cdots & \frac{\partial}{\partial V_l} P_i \\ \ddots & \ddots & \ddots & \ddots \\ \frac{\partial}{\partial \theta_c} Q_i & \cdots & \cdots & \frac{\partial}{\partial V_l} Q_i \end{bmatrix}_{|\mathbf{x}=\bar{\mathbf{x}},} \tag{2.9}$$

where $c \in \{2, 3, \ldots, N\}$ and $l \in \{1, 2, \ldots, N\}$; and the entries corresponding to the measurements of active power flow and reactive power flow are [4]

$$\frac{\partial}{\partial \theta_l} P_{ij} = \begin{cases} V_i V_j (g_{ij} \sin\theta_{ij} - b_{ij} \cos\theta_{ij}), & l = i \\ -V_i V_j (g_{ij} \sin\theta_{ij} - b_{ij} \cos\theta_{ij}), & l = j \\ 0, & l \neq i \text{ and } l \neq j \end{cases} \tag{2.10a}$$

$$\frac{\partial}{\partial V_l} P_{ij} = \begin{cases} -V_j (g_{ij} \cos\theta_{ij} + b_{ij} \sin\theta_{ij}) + 2(g_{ij} + g_{si}) V_i, & l = i \\ -V_j (g_{ij} \cos\theta_{ij} + b_{ij} \sin\theta_{ij}), & l = j \\ 0, & l \neq i \text{ and } l \neq j \end{cases} \tag{2.10b}$$

and

$$\frac{\partial}{\partial \theta_l} Q_{ij} = \begin{cases} -V_i V_j (g_{ij} \cos\theta_{ij} + b_{ij} \sin\theta_{ij}), & l = i \\ V_i V_j (g_{ij} \cos\theta_{ij} + b_{ij} \sin\theta_{ij}), & l = j \\ 0, & l \neq i \text{ and } l \neq j \end{cases} \tag{2.10c}$$

$$\frac{\partial}{\partial V_l} Q_{ij} = \begin{cases} -V_j (g_{ij} \sin\theta_{ij} - b_{ij} \cos\theta_{ij}) - 2V_i (b_{ij} + b_{si}), & l = i \\ -V_i (g_{ij} \sin\theta_{ij} - b_{ij} \cos\theta_{ij}), & l = j \\ 0, & l \neq i \text{ and } l \neq j, \end{cases} \tag{2.10d}$$

respectively. The entries corresponding to the measurements of active power injection and reactive power injection are [4]

$$\frac{\partial}{\partial\theta_l}P_i=\begin{cases}\sum_{j\in\mathcal{N}_i}V_iV_j(-G_{ij}\sin\theta_{ij}+B_{ij}\cos\theta_{ij})-V_i^2B_{ii}, & l=i\\ V_iV_j(G_{ij}\sin\theta_{ij}-B_{ij}\cos\theta_{ij}), & l\in\mathcal{N}_i\\ 0, & l\neq i \text{ and } l\notin\mathcal{N}_i\end{cases} \tag{2.11a}$$

$$\frac{\partial}{\partial V_l}P_i=\begin{cases}\sum_{j\in\mathcal{N}_i}V_j(G_{ij}\cos\theta_{ij}+B_{ij}\sin\theta_{ij})+V_iG_{ii}, & l=i\\ V_i(G_{ij}\cos\theta_{ij}+B_{ij}\sin\theta_{ij}), & l\in\mathcal{N}_i\\ 0, & l\neq i \text{ and } l\notin\mathcal{N}_i\end{cases} \tag{2.11b}$$

and

$$\frac{\partial}{\partial\theta_l}Q_i=\begin{cases}\sum_{j\in\mathcal{N}_i}V_iV_j(G_{ij}\cos\theta_{ij}+B_{ij}\sin\theta_{ij})-V_i^2G_{ii}, & l=i\\ V_iV_j(-G_{ij}\cos\theta_{ij}-B_{ij}\sin\theta_{ij}), & l\in\mathcal{N}_i\\ 0, & l\neq i \text{ and } l\notin\mathcal{N}_i\end{cases} \tag{2.11c}$$

$$\frac{\partial}{\partial V_l}Q_i=\begin{cases}\sum_{j\in\mathcal{N}_i}V_j(G_{ij}\sin\theta_{ij}-B_{ij}\cos\theta_{ij})+V_iB_{ii}, & l=i\\ V_i(G_{ij}\sin\theta_{ij}-B_{ij}\cos\theta_{ij}), & l\in\mathcal{N}_i\\ 0, & l\neq i \text{ and } l\notin\mathcal{N}_i,\end{cases} \tag{2.11d}$$

respectively. Therefore, the observation model with linearized dynamics at the operation point $\bar{\mathbf{x}}$ is

$$\begin{bmatrix}\vdots\\ P_{ij}\\ \vdots\\ Q_{ij}\\ \vdots\\ P_i\\ \vdots\\ Q_i\\ \vdots\end{bmatrix}=\begin{bmatrix}\frac{\partial}{\partial\theta_l}P_{ij} & \cdots & \cdots & \frac{\partial}{\partial V_l}P_{ij}\\ \ddots & \ddots & \ddots & \ddots\\ \frac{\partial}{\partial\theta_l}Q_{ij} & \cdots & \cdots & \frac{\partial}{\partial V_l}Q_{ij}\\ \ddots & \ddots & \ddots & \ddots\\ \frac{\partial}{\partial\theta_l}P_i & \cdots & \cdots & \frac{\partial}{\partial V_l}P_i\\ \ddots & \ddots & \ddots & \ddots\\ \frac{\partial}{\partial\theta_l}Q_i & \cdots & \cdots & \frac{\partial}{\partial V_l}Q_i\end{bmatrix}\begin{bmatrix}\vdots\\ \theta_i\\ \vdots\\ V_i\\ \vdots\end{bmatrix}+Z^m. \tag{2.12}$$

Note that the Jacobian matrix is a description of the topology and the physical parameters of a system.

The linearized observation model in (2.12) is further simplified to *Direct Current (DC)* model by putting in place the following assumptions:

(1) All the voltage magnitudes are 1 at all buses, that is

$$[V_1,V_2,\ldots,V_N]^{\mathsf{T}}=\mathbf{1}. \tag{2.13}$$

(2) The shunt susceptances $g_{si}+jb_{si}$ of bus i and the series resistances g_{ij} on the power lines from bus i to bus j in Fig. 2.1 are assumed to be:

$$g_{si}+jb_{si}=0 \tag{2.14a}$$

$$g_{ij}=0. \tag{2.14b}$$

(3) The phase angles θ_{ij}, with $i \neq j$, between bus i and j, with $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$, satisfy

$$\theta_{ij} = 0. \tag{2.15}$$

Hence, the measurements only contain the active power and the state variables only contain the phase angles of the buses. The following definition characterizes the state variables for the DC model.

Definition 3 (State variables for the DC model)**.** *Consider a power system with $N \in \mathbb{Z}_+$ buses and assume bus 1 is chosen as the reference bus such that the phase angle of bus 1 is set to 0. Let $\mathbf{x} \in \mathbb{R}^n$ be the vector of state variables of the system such that*

$$\mathbf{x} \triangleq [\theta_2, \theta_3, \ldots, \theta_N]^\mathsf{T}, \tag{2.16}$$

where $\theta_i \in [-\pi, \pi)$, with $i \in \{2, 3, \ldots, N\}$, is the phase angle of bus i in the system.

The following definition characterizes the measurements for the DC model.

Definition 4 (Measurements for the DC model)**.** *Consider a power system with $N \in \mathbb{Z}_+$ buses. Let $Y^m \in \mathbb{R}^m$ be the vector of measurements in the system such that*

$$Y^m \triangleq [P_{ij}, \ldots, P_1, P_2, \ldots, P_N]^\mathsf{T}, \tag{2.17}$$

where $P_i \in \mathbb{R}$, with $i \in \{1, 2, \ldots, N\}$, is the active power injection to bus i. Let $\mathcal{N}_i$ be the set of buses that are physically connected to bus i. The active power $P_{ij} \in \mathbb{R}$, with $i \in \{1, 2, \ldots, N\}$ and $j \in \mathcal{N}_i$, is the power flow from bus i to bus j.

Under the assumption of DC model in (2.13), (2.14) and (2.15), it yields the vector of state variables and vector of measurements in (2.16) and (2.17), respectively. Therefore, the DC observation model is defined in the following.

Definition 5 (DC observation model)**.** *Consider a power system with $N \in \mathbb{Z}_+$ buses. Let $\mathbf{x} \in \mathbb{R}^n$ be the vector of state variables in* (2.16) *and $Y^m \in \mathbb{R}^m$ be the vector of measurements in* (2.17)*. The DC observation model is*

$$Y^m \triangleq \mathbf{H}\mathbf{x} + Z^m, \tag{2.18}$$

where the Jacobian matrix $\mathbf{H} \in \mathbb{R}^{m \times n}$ for the DC observation model is

$$\mathbf{H} = \begin{bmatrix} \frac{\partial}{\partial \theta_l} P_{ij} \\ \ddots \\ \frac{\partial}{\partial \theta_l} P_i \end{bmatrix}, \tag{2.19}$$

where θ_l and P_{ij} are in (2.16) *and* (2.17)*, respectively. The entries of $\mathbf{H}$ that correspond to active power flow measurements, that is, $\frac{\partial}{\partial \theta_l} P_{ij}$ in* (2.10a) *are simplified as*

$$\frac{\partial}{\partial \theta_l} P_{ij} = \begin{cases} V_i V_j (g_{ij} \sin\theta_{ij} - b_{ij} \cos\theta_{ij}) = -b_{ij}, & l = i \\ -V_i V_j (g_{ij} \sin\theta_{ij} - b_{ij} \cos\theta_{ij}) = b_{ij}, & l = j, \end{cases} \tag{2.20}$$

and the entries that correspond to active power injection to bus i, that is, $\frac{\partial}{\partial\theta_i}P_i$ in (2.11a) *are simplified as*

$$\frac{\partial}{\partial\theta_l}P_i = \begin{cases} \frac{\partial}{\partial\theta_i}\sum_{j\in\mathcal{N}_i}P_{ij} = -\sum_{j\in\mathcal{N}_i}b_{ij}, & l=i \\ \frac{\partial}{\partial\theta_j}\sum_{j\in\mathcal{N}_i}P_{ij} = \sum_{j\in\mathcal{N}_i}b_{ij}, & l\in\mathcal{N}_i \end{cases}. \tag{2.21}$$

The noise vector Z^m is as defined in (2.6).

2.2 Classical DIAs and Attack Detection

2.2.1 State Estimation

The measurements are collected via a SCADA system and processed by a state estimator at the control center to analyze the current state of the operation. The result of state estimation is used for the purpose of optimal power flow, frequency control, economic dispatch, etc. [4]

The aim of the state estimator is to obtain an estimate of the state variables that minimizes the cost of the function $c : \mathbb{R}^n \to \mathbb{R}$ as follows:

$$c(\mathbf{x}, \hat{\mathbf{x}}), \tag{2.22}$$

where $\hat{\mathbf{x}} \in \mathbb{R}^n$ is the estimate of $\mathbf{x}$. Given a vector of the measurements $\mathbf{y} \in \mathbb{R}^m$, one of the commonly used cost functions is given by

$$c(\mathbf{x}, \hat{\mathbf{x}}) \triangleq \|\mathbf{y} - F(\hat{\mathbf{x}})\|_{\ell_2}^2, \tag{2.23}$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is introduced in (2.5); and $\|\mathbf{y} - F(\hat{\mathbf{x}})\|_{\ell_2}$ is the ℓ_2 norm of $\mathbf{y} - F(\hat{\mathbf{x}})$. Minimizing the cost in (2.23) yields the least squares (LS) estimate as

$$\hat{\mathbf{x}}^* \triangleq \arg\min_{\hat{\mathbf{x}}} \|\mathbf{y} - F(\hat{\mathbf{x}})\|_{\ell_2}^2. \tag{2.24}$$

Iterative approaches such as Gauss-Newton method are adopted to obtain the LS estimate above [28, 29]. However, nonlinear LS estimate is computationally expensive and does not always converge to a solution.

The corollaries provide the LS estimate for the DC model in (2.18).

Corollary 0.1. *For the DC observation model in* (2.18)*, let $\mathbf{z} \in \mathbb{R}^m$ be a realization of the system noise and $\mathbf{y} \in \mathbb{R}^m$ be a the vector of measurements such that $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}$. The LS estimate is given by*

$$\hat{\mathbf{x}}^* \triangleq \arg\min_{\hat{\mathbf{x}}} \|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|_{\ell_2}^2 = (\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}\mathbf{y}, \tag{2.25}$$

where $\mathbf{y} \in \mathbb{R}^m$ is a vector of the measurements and the matrix $\mathbf{H} \in \mathbb{R}^{m\times n}$ is described in (2.19).

Specifically, for the case where the vector of noise Z^m in (2.18) satisfies $Z^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{ZZ})$, with $\mathbf{\Sigma}_{ZZ} \in \mathcal{S}_+^m$, the following corollary provides the weighted least squares (WLS) estimate.

Corollary 0.2. *For the DC observation model in* (2.18)*, let* $\mathbf{z} \in \mathbb{R}^m$ *be a realization of the system noise and* $\mathbf{y} \in \mathbb{R}^m$ *be a the vector of measurements such that* $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}$*. The WLS estimate is given by*

$$\hat{\mathbf{x}}^* \stackrel{\Delta}{=} \arg\min_{\hat{\mathbf{x}}} \|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|_{\ell_2}^2 = (\mathbf{H}^\mathsf{T}\mathbf{\Sigma}_{ZZ}^{-1}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}\mathbf{\Sigma}_{ZZ}^{-1}\mathbf{y}, \tag{2.26}$$

where $\mathbf{y} \in \mathbb{R}^m$ *is a vector of the measurements; the matrix* $\mathbf{H} \in \mathbb{R}^{m\times n}$ *is given in* (2.19)*; and the matrix* $\mathbf{\Sigma}_{ZZ} \in \mathcal{S}_+^m$ *is the covariance matrix for the vector of noise in* (2.18)*.*

2.2.2 Deterministic DIAs

DIAs, first introduced by [1, 2], are one of the main cyber threats that target the measurements. Specifically, consider a malicious attack vector $\mathbf{a} \in \mathbb{R}^m$ to the DC observation model described in Definition 5. The DC observation model that emerges when compromised by a malicious attack vector $\mathbf{a}$ is defined below.

Definition 6 (DC observation model under attacks)**.** *Let* $\mathbf{a} \in \mathbb{R}^m$ *be the malicious attack vector. The DC observation model under attacks is*

$$Y_A^m \stackrel{\Delta}{=} \mathbf{H}\mathbf{x} + Z^m + \mathbf{a}, \tag{2.27}$$

where $Y_A^m \in \mathbb{R}^m$ *denotes the vector of compromised measurements, the vector of state variables* $\mathbf{x} \in \mathbb{R}^n$ *is described in* (2.16)*, the Jacobian matrix* $\mathbf{H} \in \mathbb{R}^{m\times n}$ *is described in* (2.19)*; and the noise* $Z^m \in \mathbb{R}^m$ *is described in* (2.6)*.*

Let $\mathbf{z} \in \mathbb{R}^m$ be a realization of the system noise. Consequently, the vector of compromised measurements denoted by $\mathbf{y}_a$ is given by

$$\mathbf{y}_a \stackrel{\Delta}{=} \mathbf{H}\mathbf{x} + \mathbf{z} + \mathbf{a} \tag{2.28}$$

2.2.3 Residual-based Anomaly Detection

As part of the state estimation, the system operator launches bad data detection prior to state estimation. The commonly used bad data detection method is residual test [4,30] given by

$$r \stackrel{\Delta}{=} \|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|_{\ell_2}^2, \tag{2.29}$$

where r is the residual. The attack detection is cast as the following hypothesis testing problem

$$\mathcal{H}_o\colon r < r_0 \quad \text{there is no attack,} \tag{2.30a}$$
$$\mathcal{H}_1\colon r \geq r_0 \quad \text{measurements are compromised,} \tag{2.30b}$$

where $\mathcal{H}_o$ and $\mathcal{H}_1$ are the null hypothesis and alternative hypothesis, respectively; r_0 is a detection threshold set by the operator. At time step $t \in \mathbb{Z}_+$, the system operator acquires a vector of measurements $\mathbf{y}$ and decides whether the vector of measurements is produced following a no attack scenario as described in (2.18) or is the result of an attack.

When the noise vector Z^m is assumed to follow a zero mean multivariate Gaussian distribution, that is, $Z^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{ZZ})$ where $\mathbf{\Sigma}_{ZZ} \in \mathcal{S}_+^m$ is a covariance matrix such that only diagonal entries are nonzero, the residual test can be cast as normalized residual test as follows

$$r_n = (\mathbf{y} - \mathbf{H}\hat{\mathbf{x}})^\mathsf{T} \mathbf{\Sigma}_{ZZ}^{-1} (\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}), \tag{2.31}$$

where the normalized residual denoted by r_n follows a chi-squared distribution with $m - n$ degrees of freedom, that is,

$$r_n \sim \mathcal{X}^2_{m-n}. \tag{2.32}$$

Hence, it follows that the hypothesis testing problem in normalized residual test is given by

$$\mathcal{H}_o\text{: } r_n \in \mathcal{X}^2_{m-n}(\gamma) \quad \text{there is no attack,} \tag{2.33a}$$
$$\mathcal{H}_1\text{: } r_n \notin \mathcal{X}^2_{m-n}(\gamma) \quad \text{measurements are compromised,} \tag{2.33b}$$

where $\gamma \in [0, 1]$ is the significant level chosen by the operator.

The following lemma provides an attack construction that does not change the residual in (2.29).

Lemma 1. *[2] The vector of compromised measurements* $\mathbf{y}_a$ *in* (2.28) *does not change the residual in* (2.29) *if* $\mathbf{a}$ *is a linear combination of the column vectors of* $\mathbf{H}$*, that is,*

$$\mathbf{a} = \mathbf{H}\mathbf{c}, \tag{2.34}$$

where $\mathbf{c} \in \mathbb{R}^n$ *is an arbitrary nonzero vector.*

Proof. Let $\hat{\mathbf{x}}_a$ be the vector of state estimate obtained from $\mathbf{y}_a$. From Corollary 0.2, the following holds

$$\begin{aligned}
\hat{\mathbf{x}}_a =& (\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}\mathbf{y}_a && (2.35)\\
=& (\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}(\mathbf{y} + \mathbf{a}) && (2.36)\\
=& \hat{\mathbf{x}} + (\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}\mathbf{a}. && (2.37)
\end{aligned}$$

For $\mathbf{a} = \mathbf{H}\mathbf{c}$, let r_a be the residual under attacks. The following holds

$$\begin{aligned}
r_a =& \|\mathbf{y}_a - \mathbf{H}\hat{\mathbf{x}}_a\|_{\ell_2}^2 && (2.38)\\
=& \|\mathbf{y} + \mathbf{a} - \mathbf{H}\left(\hat{\mathbf{x}} + (\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}\mathbf{a}\right)\|_{\ell_2}^2 && (2.39)\\
=& \|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}} + \left(\mathbf{a} - \mathbf{H}(\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}\mathbf{H}\mathbf{c}\right)\|_{\ell_2}^2 && (2.40)\\
=& \|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}} + (\mathbf{H}\mathbf{c} - \mathbf{H}\mathbf{c})\|_{\ell_2}^2 && (2.41)\\
=& \|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|_{\ell_2}^2 && (2.42)\\
=& r && (2.43)
\end{aligned}$$

Therefore, the vector of compromised measurements $\mathbf{y}_a$ does not change the residual in (2.29). This completes the proof. □

Note that the injected error to the vector of state variables is

$$\begin{aligned}&\hat{\mathbf{x}}_a - \hat{\mathbf{x}} && (2.44)\\ =&\hat{\mathbf{x}} + (\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}\mathbf{a} - \hat{\mathbf{x}} && (2.45)\\ =&(\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}\mathbf{H}\mathbf{c}, && (2.46)\\ =&\mathbf{c} && (2.47)\end{aligned}$$

where (2.45) holds from taking (2.37) into (2.44).

2.3 DIAs within a Bayesian Framework

Consider the system model in Definition 5 within a Bayesian framework in which the state variables are modelled as random variables, that is

$$X^n \sim P_{X^n}, \tag{2.48}$$

where $X^n \in \mathbb{R}^n$ is a random vector that describes the state variables with a given distribution. The rationale for modelling state variables as a random process is to capture the complexity and dynamic nature of power systems. Moreover, with the unprecedented data acquisition capabilities available to cyberphysical systems, the attackers can even learn the statistical structure of the system and incorporate the underlying stochastic process to launch the attacks [17,20]. From the perspective of the operator, the introduction of stochastic descriptors opens the door to information theoretic quantifications of the measurement vulnerability.

The definition of DC random observation model is provided below.

Definition 7 (DC random observation model)**.** *Let $Y^m \in \mathbb{R}^m$ be the random vector of measurements and $X^n \in \mathbb{R}^n$ be the random vector of state variables. The DC random observation model is*

$$Y^m = \mathbf{H}X^n + Z^m, \tag{2.49}$$

where the Jacobian matrix $\mathbf{H} \in \mathbb{R}^{m\times n}$ is described in (2.19)*; and the random vector of noise Z^m is given in* (2.6)*.*

2.3.1 State Estimation

The aim of state estimator is to obtain an estimate $\hat{X}^n$ of the vector X^n from the vector of measurements Y^m. A widely used cost function in state estimation within the Bayesian framework is mean square error (MSE) given by

$$\mathbb{E}[\|X^n - \hat{X}^n\|_{\ell_2}^2]. \tag{2.50}$$

By adopting a linear estimate, the resulting state estimate is $\hat{X}^n = \mathbf{L}Y^m$, where $\mathbf{L} \in \mathbb{R}^{n\times m}$ is the linear estimation matrix. That being the case, the optimal estimator that achieves the minimum mean squared error (MMSE) is given by

$$\mathbf{M} = \underset{\mathbf{L}\in\mathbb{R}^{n\times m}}{\arg\min}\, \mathbb{E}[\frac{1}{n}\|X^n - \mathbf{L}Y^m\|_{\ell_2}^2], \tag{2.51}$$

where the expectation is taken with respect to X^n and Y^m.

Lemma 2. *Under the assumption that $X^n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{XX})$ and a vector of measurements $\mathbf{y}$, the MMSE estimate is given by*

$$\hat{\mathbf{x}}^*_{\text{MMSE}} = \mathbf{M}\mathbf{y}, \tag{2.52}$$

where,

$$\mathbf{M} = \boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}} + \sigma^2\boldsymbol{I}_m)^{-1}, \tag{2.53}$$

and $\sigma^2 \in \mathbb{R}_+$ is the variance of the system noise.

2.3.2 Random Attack Construction

Consider an additive random attack vector denoted by $A^m \in \mathbb{R}^m$ to DC random observation model in Definition 7. Note that the attack is modelled as a random process. The following definition provides the DC model when random attacks are present.

Definition 8 (DC model with random attacks)**.** *Let $Y^m_A \in \mathbb{R}^m$ be the random vector of measurements compromised by random attacks $A \in \mathbb{R}^m$ and $X^n \in \mathbb{R}^n$ be the random vector of state variables. The DC random observation model with random attacks is*

$$Y^m_A = \mathbf{H}X^n + Z^m + A^m, \tag{2.54}$$

where the Jacobian matrix $\mathbf{H} \in \mathbb{R}^{m\times n}$ is given by (2.19)*; the random vector of noise Z^m is described in* (2.6)*; and $A^m \sim P_{A^m}$ where P_{A^m} is determined by the attacker.*

The independence of the random attack vector and state variables implies that the attacker does not need to know the distribution of the state variables to construct the attacks.

In this setting, the goal of the attacker is to hinder the accuracy of state estimate without being detected. Therefore, two objectives are involved in the attack constructions: (1) the disruption caused to the state estimate; (2) the probability of detection. From the attacker's point of view, the goal is to maximize the disruption or minimize the probability of detection under certain constraints.

Let $\mathbf{a} \in \mathbb{R}^m$ denote an attack vector, it follows that

$$\mathbf{y}_a = \mathbf{H}\mathbf{x} + \mathbf{z} + \mathbf{a}. \tag{2.55}$$

The optimal MMSE state estimate without attacks is given in Lemma 2, that is,

$$\hat{\mathbf{x}}^* = \mathbf{M}\mathbf{y}, \tag{2.56}$$

where $\mathbf{M}$ is described in (2.53). Similarly, the MMSE state estimate under attack is

$$\begin{aligned}\hat{\mathbf{x}}^*_a =& \mathbf{M}\mathbf{y}_a\\ =& \mathbf{M}(\mathbf{y} + \mathbf{a})\\ =& \hat{\mathbf{x}} + \mathbf{M}\mathbf{a}.\end{aligned} \tag{2.57}$$

In this setting, the deviation of the state estimate is $\mathbf{M}\mathbf{a}$. On the other hand, the probability distribution under attacks, $P_{Y^m_A}$, determines the probability of attack detection. Let P_{D} be the probability of attack detection. Hence, it yields that

$$P_{\text{D}} \triangleq \int_{\mathcal{S}} \mathrm{d}P_{Y^m_A} = \mathbb{E}[\mathbb{1}_{\{L(\mathbf{y}_a)\geq\tau\}}], \tag{2.58}$$

where the integration domain $\mathcal{S}$ only contains the realizations of Y_a^m that yield a likelihood ratio value larger than τ and $\mathbb{1}_{\{\cdot\}}$ is the indicator function. Specifically, the integration domain is

$$\mathcal{S} \triangleq \{\mathbf{y}_a \in \mathbb{R}^m : L(\mathbf{y}_a) \geq \tau\}, \tag{2.59}$$

where $L(\mathbf{y}_a) \triangleq \frac{f_1(\mathbf{y}_a)}{f_0(\mathbf{y}_a)}$, and $f_i, i \in \{0,1\}$ is the probability density function of a probability distribution.

Let $\|\mathbf{Ma}\|_{\ell_2}^2$ denote the disruption caused by the attacks. It follows in the case where the attacker aims at maximizing the disruption with a constraint on the probability of detection, the objective function is given by

$$\begin{aligned} \max_{\mathbf{a}} \ & \|\mathbf{Ma}\|_{\ell_2}^2 \\ \text{s.t.} \ & P_{\mathrm{D}} \leq \tau_1, \end{aligned} \tag{2.60}$$

where τ_1 is the detection threshold set by the operator. Alternatively, in the case where the attacker aims at minimizing the probability of detection with a constraint on the disruption, the objective function is given by

$$\begin{aligned} \min_{\mathbf{a}} \ & P_{\mathrm{D}} \\ \text{s.t.} \ & \|\mathbf{Ma}\|_{\ell_2}^2 \geq \tau_2, \end{aligned} \tag{2.61}$$

where τ_2 is the induced distortion threshold set by the operator. In [12] and [31], the tradeoff between attack distortion and probability of detection is cast as an optimization problem given by (2.60). The tradeoff is studied in [32] and [33] in a dynamic setting.

2.3.3 Optimal Attack Detection

As a part of a security strategy, the operator implements an attack detection procedure prior to performing state estimation. Attack detection is cast as a hypothesis testing problem given by

$$\mathcal{H}_o\text{:} \quad \text{There is no attack,} \tag{2.62a}$$

$$\mathcal{H}_1\text{:} \quad \text{Measurements are compromised.} \tag{2.62b}$$

The system operator obtains a vector of measurements $\bar{Y}^m$ and decides whether it is produced following a no attack scenario as in (2.49) or whether the measurements have been compromised. In a Bayesian setting, the hypothesis test can be recast in terms of the probability density functions (pdf) induced by the distribution of the measurements. Let P_1 and P_0 be the distributions of the vector of measurements with attacks and without attacks, respectively. Hence, it follows that the hypotheses in (2.62) are

$$\mathcal{H}_o\text{:} \ \bar{Y}^m \sim P_0, \tag{2.63a}$$

$$\mathcal{H}_1\text{:} \ \bar{Y}^m \sim P_1. \tag{2.63b}$$

A test to determine what distribution generates the observation data is a deterministic test $T : \mathbb{R}^m \to \{0,1\}$. Given an observation vector $\bar{\mathbf{y}}$, let $T(\bar{\mathbf{y}}) = 0$ denote the case in which

the test decides $\mathcal{H}_o$ upon the observation of $\bar{\mathbf{y}}$; and $T(\bar{\mathbf{y}}) = 1$ the case in which the test decides $\mathcal{H}_1$. The performance of the test is assessed in terms of the Type-I error, denoted by $\alpha \triangleq \mathbb{P}\left[T\left(\bar{Y}^m\right) = 1\right]$, with $\bar{Y}^m \sim P_0$; and the Type-II error, denoted by $\beta \triangleq \mathbb{P}\left[T\left(\bar{Y}^m\right) = 0\right]$, with $\bar{Y}^m \sim P_1$. Table 2.1 summarizes the Type-I and Type-II error in the binary hypothesis testing. In a Neyman-Person, the decision rule aims to minimize the Type-II error β given

Table 2.1: Type-I error and Type-II error in binary hypothesis testing

	Accept $\mathcal{H}_o$	Reject $\mathcal{H}_o$
$\mathcal{H}_o$ is true	✓	Type-I error (false alarm)
$\mathcal{H}_1$ is true	Type-II error (miss)	✓

the constraint that the Type-I error α satisfies $\alpha \leq \alpha'$, with $\alpha' \in [0, 1]$. The following lemma provides the optimality of the *likelihood ratio test (LRT)* for the hypothesis test given in (2.63).

Lemma 3. *[34, Proposition II.D.1: Neyman-Pearson Lemma] Given the hypothesis testing problem in* (2.63)*, among all the tests that achieve probability of Type-I error α such that $\alpha \leq \alpha'$, the LRT given by*

$$T(\bar{\mathbf{y}}) = \mathbb{1}_{\{L(\bar{\mathbf{y}}) \geqslant \tau\}}, \tag{2.64}$$

where $T(\bar{\mathbf{y}})$ is the likelihood ratio, that is,

$$L(\bar{\mathbf{y}}) = \frac{f_1(\bar{\mathbf{y}})}{f_0(\bar{\mathbf{y}})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau, \tag{2.65}$$

achieves the minimum probability of Type-II error β, where τ is a decision threshold that achieves Type-I error $\alpha = \alpha'$, and the functions f_1 and f_0 are the probability density function (pdf) of P_1 and P_0, respectively.

Note that changing the value of τ results in changes to the tradeoff between Type-I and Type-II errors.

Chapter 3

State of the Art

In this chapter, the research results on DIAs from the perspectives of centralized and decentralized scenarios are summaried. In a centralized system, there exists a unique controller [35]. Specifically, in DIAs, the term *centralized attacks* refers to the case where DIAs are determined by one attacker, that is, the unique attacker decides the attack vector $\mathbf{a} \in \mathbb{R}^m$. On the other hand, the nature of the configuration of smart grid, e.g., microgrid integration [36], electric vehicle to grid (V2G) [37], yields a decentralized system in power flow control (PFC), energy management system (EMS) [38], etc. [39] In decentralized scenario, DIAs is referred to as *decentralized attacks* where several attackers determined the attack vector jointly [13]. The main research on DIAs includes three aspects and the corresponding challenges are listed in Table 3.1.

Table 3.1: Main Research on DIAs

Research on DIAs	Aims	Challenges
DIAs constructions	Construction of attack vector without being detected	Sparsity constraints [9, 20, 40–44]
		Incomplete system information [45–51]
		Falsified topology [52–55]
		AC power flow Model [56]
		Attacks within Bayesian framework [17, 26]
		Decentralized attacks [31]
The impacts of DIAs	Analysis of the impacts on power system	Economic attacks [57–60]
		Load redistribution attacks [61, 62]
		Energy deceiving attacks [63]
Defense strategies against DIAs	Defence strategies for system operators	Vulnerable measurements protection [11, 42, 64]
		PMU-based protection [65–67]
		Other defence mechanism [68–73]

The main research results and contributions of this thesis are listed in Table 3. This thesis proposed independent and correlated attacks in centralized systems as well as the measurement vulnerability analysis. In decentralized systems, decentralized attacks where

the attackers coordinate in a game formula are proposed. The sparsity constraints in decentralized attacks are considered.

Table 3.2: Contributions of this thesis

	One attacker	Results	Convergence
Centralized systems	Y	Independent attacks	Y
		Correlated attacks	Y
		Measurement vulnerability analysis	N/A
Decentralized systems	N	Decentralized attacks	Y
		Decentralized attacks with sparsity	Y

Consider the system model under attack vector $\mathbf{a} \in \mathbb{R}^m$ in Definition 6, that is,

$$Y_A^m = \mathbf{H}\mathbf{x} + Z^m + \mathbf{a}, \tag{3.1}$$

where $Y_A^m \in \mathbb{R}^m$ is the random vector of compromised measurements, the matrix $\mathbf{H} \in \mathbb{R}^{m \times n}$ is the Jacobian matrix described in (2.19), the vector $\mathbf{x} \in \mathbb{R}^n$ is the vector of state variables and the noise Z^m is introduced in (2.6).

This chapter introduces and reviews the main results on centralized attacks with the challenges that include: sparsity constraints in Section 3.1, incomplete system information in Section 3.2, falsified topology in Section 3.3, and AC power flow model in Section 3.4. The scenario with multiple attackers is described in Section 3.6.

3.1 Sparse Attacks

The attack vector $\mathbf{a} \in \mathbb{R}^m$ in (3.1) that relies on having access to all the measurements satisfies

$$\text{card}\left(\text{supp}\left(\mathbf{a}\right)\right) = m, \tag{3.2}$$

where $\text{card}\left(\text{supp}\left(\mathbf{a}\right)\right)$ denotes the cardinality of the set $\text{supp}(\mathbf{a})$; and $\text{supp}(\mathbf{a})$ is the support of the vector $\mathbf{a}$ defined as follows:

$$\text{supp}(\mathbf{a}) \stackrel{\Delta}{=} \left\{i : \mathbb{P}\left[\mathbf{a}_i = 0\right] = 0\right\}. \tag{3.3}$$

Given that the operator is likely to have access control policies in place [74], the assumption that the attack construction has access to all the measurements in the system is costly and limits the feasibility of the attack.

Attack constructions that have access to a subset of the measurements are referred to as *sparse attacks*. Typically, the sparsity constraints are incorporated to the attack construction formulation by imposing additional constraints to the optimization problem describing the attack construction. Specifically, attack constructions that require access to $k < m$ measurements on the system yields k-sparse attack constructions, that is,

$$\text{card}\left(\text{supp}\left(\mathbf{a}\right)\right) = k < m. \tag{3.4}$$

In view of this, the attacker constructs a sparse attack vector that chooses k indices from m measurements and puts nonzero mass on the indices selected for the attack. Lemma 1 states that the attack vector $\mathbf{a}$ that satisfies the constraint $\mathbf{a} = \mathbf{Hc}$ is undetectable for all $\mathbf{c} \in \mathbb{R}^n$ by a residual-based detection. Let $\|\mathbf{a}\|_{\ell_0}$ be the ℓ_0 norm of the vector $\mathbf{a}$, that is, total number of nonzero entries of vector $\mathbf{a}$. The k-sparse attack construction problem is cast as follows [2]:

$$\begin{aligned} &\min_{\mathbf{a}\in\mathbb{R}^m} \|\mathbf{a}\|_{\ell_0} \\ &\text{s.t. } \mathbf{a} = \mathbf{Hc}, \mathbf{a} \neq \mathbf{0}, \\ &\qquad \mathbf{a}_i = 0 \ \text{ for all } i \in \{1, 2, \ldots, m\} \setminus \mathcal{C}, \end{aligned} \tag{3.5}$$

where $\|\mathbf{a}\|_{\ell_0}$ is the ℓ_0 norm of vector $\mathbf{a}$ and $\mathcal{C}$ is the set of measurements that the attacker has access to. However, the optimization problem in (3.5) is NP-hard (Nondeterministic Polynomial) [75, Th. 2]. A problem is NP-hard if the problem is not solvable in polynomial time but can be verified in polynomial time [76]. As in subset sum problem [77] and travelling salesman problem [78], the problem in (3.5) is a combinatorial optimization problem where a nonzero vector $\mathbf{c}$ yields a configuration of combinatorial of the columns of Jacobian matrix $\mathbf{H}$.

A brute force approach to the optimization problem in (3.5) is not practical due to the fact that power systems often operate with high dimensional state variable descriptions that limit the choice of practical computationally feasible approaches [79]. Indeed, the number of support choices for the attack grows exponentially with the dimension of the system, and therefore, the optimization domain quickly becomes intractable. Heuristic algorithms and greedy algorithms such as matching pursuit [80] and orthogonal matching pursuit [81] are widely used as a powerful tool to circumvent NP-hard problems and provide good performance by trading off optimality for low complexity [82,83]. For instance, a heuristic approach proposed in [2] performs column transformations to the matrix $\mathbf{H}$ to reduce the number of nonzeros in the resulting transformed $\mathbf{H}$ until the sparsity constraint $k < m$ is satisfied. Essentially, this approach induces sparsity by finding a linear combination of the columns of $\mathbf{H}$ to solve the minimization problem in (3.5). Given the fact that the Jacobian matrix $\mathbf{H}$ is usually a sparse and rank deficient matrix in real power systems [4,5], the heuristic approach is less time consuming. However, a successful attack construction is not guaranteed even if the attack exists, nor does it guarantee the minimal support for the constructed attack vector.

To manipulate a predetermined number of state variables $s \in \{1, 2, \ldots, n\}$, the study in [75] proposed a *least effort attack* model, i.e., the objective of the adversary is to identify the minimum number of measurements that attacker has to compromise. The problem is formulated as

$$\begin{aligned} &\min_{\mathbf{c}\in\mathbb{R}^n} \|\mathbf{Hc}\|_{\ell_0} \\ &\text{s.t. } \|\mathbf{c}\|_{\ell_0} = s. \end{aligned} \tag{3.6}$$

The rationale for formulating the aim of manipulating s state variables according to the constraint in (3.6) comes from (2.44). In (2.44), the vector $\mathbf{c}$ is the difference between the

state estimate with attacks and without attacks. Hence, constraining the ℓ_0 norm of $\mathbf{c}$ to s yields that the number of state variables is s.

The following theorem characterizes the number of measurements to be compromised given s state variables to manipulate.

Theorem 4. *[75, Th. 1] Let $\Gamma = \{i_1, i_2, \ldots, i_s\}$ be the set of state variables to manipulate, that is, $\mathbf{c}_{i_j} \neq 0$, with $j \in \{1, 2, \ldots, s\}$, and $\mathcal{N}_i$ be the set of buses that are physically connected to bus i. For the attack vector $\mathbf{a} = \mathbf{Hc}$, the following holds*

$$\|\mathbf{a}\|_{\ell_0} = s + 3\sum_{j=i_1}^{i_s} \mathrm{card}\left(\mathcal{Q}_j\right) - \sum_{j=i_1}^{i_{s-1}} \mathrm{card}\left(\mathcal{E}_j\right) - \mathrm{card}\left(\mathcal{W}\right), \tag{3.7}$$

where $\|\mathbf{a}\|_{\ell_0}$ is the ℓ_0 norm of $\mathbf{a}$; $\mathrm{card}\left(\mathcal{Q}_j\right)$ is the cardinality of $\mathcal{Q}_j$; $\mathcal{Q}_j \triangleq \{i : i \in \mathcal{N}_j \setminus \Gamma\}$, $\mathcal{E}_j \triangleq \{i : i \in \mathcal{N}_j, i \in \mathcal{N}_p, p \in \Gamma, p > j\}$ and $\mathcal{W} \triangleq \{i : i \in \Gamma, \mathrm{card}\left(\mathcal{Q}_i\right) = 0\}$.

Theorem 4 characterizes the number of the measurements that need to be compromised in order to manipulate s state variables. The optimal set of measurements is identified by a linear transformation-based approach in [75], but the calculation of the transformation is heavy due to the possible column exchanges of $\mathbf{H}$. Alternatively, a heuristic-based approach for large systems is proposed.

In fact, a large body of literature in solving NP-hard problems focuses on heuristic algorithms or greedy algorithms. A heuristic algorithm is proposed in [40] where an upper bound on the number of the measurements that need to be compromised is given for a successful sparse attack, that is, an upper bound on $\|\mathbf{a}\|_{\ell_0}$ for the optimization problem in (3.5). The minimal number of measurements need to be compromised is also studied in [41] and [42]. The necessary and sufficient condition on the matrix $\mathbf{H}$ for the attack vector to bypass under residual detection is also provided therein. Meanwhile, in [42], the same insight is obtained via a graph theoretic analysis.

Solving this problem is challenging in general owing to the combinatorial nature of support selection for the attack vector. In fact, minimizing ℓ_0 norm in (3.5) is hard and no algorithms are found yet to solve the ℓ_0 norm minimization problem efficiently [84]. Alternatively, the ℓ_0 norm minimization problem is relaxed to a convex ℓ_1 minimization under certain conditions in [82, 85–87] and [88]. Specifically, the approach that relaxes the ℓ_0 norm to ℓ_l norm minimization problem in (3.5) is as follows:

$$\begin{aligned} &\min_{\mathbf{a}\in\mathbb{R}^m} \|\mathbf{a}\|_{\ell_1} \\ &\text{s.t. } \mathbf{a} = \mathbf{Hc}, \mathbf{a} \neq \mathbf{0}, \\ &\qquad \mathbf{a}_i = 0 \ \text{ for all } i \in \{1, 2, \ldots, m\} \setminus \mathcal{C}, \end{aligned} \tag{3.8}$$

where $\|\mathbf{a}\|_{\ell_1}$ denotes the ℓ_1 norm of $\mathbf{a}$.

Generally, the ℓ_1 norm minimization problem is not equivalent to the ℓ_0 norm minimization. However, under certain conditions, the ℓ_1 relaxation can yield the same solution as that of the ℓ_0 minimization but determining when the equivalence holds is not trivial. The main result in [43] shows that the ℓ_1 relaxation approach achieves the exact solution to the ℓ_0 norm minimization under the assumption that no power injection to the buses is

measured. The theorem is based on a polyhedral combinatorics argument rather than the mutual coherence and restricted isometry property (RIP) given in [85]. This relaxation for sparse attack constructions is also studied in [89] and [44]. In [44], the ℓ_1 norm optimization problem is formulated for both the centralized and distributed attack construction settings. Specifically, the sparse attack is formulated as a standard *compressed sensing* (CS) problem as follows:

$$\begin{aligned} &\min_{\mathbf{c}\in\mathbb{R}^n} \|\mathbf{c}\|_{\ell_1} \\ &\text{s.t. } \mathbf{a} = \mathbf{H}\mathbf{c}. \end{aligned} \tag{3.9}$$

In [89], two security indices are proposed to identify the vulnerable measurements. The k-sparse attack construction is studied via the convex optimization problem given by

$$\begin{aligned} &\min_{\mathbf{c}\in\mathbb{R}^n} \|\mathbf{H}\mathbf{c}\|_{\ell_1} \\ &\text{s.t. } \sum_i \mathbf{H}_{j,i}\mathbf{c}_i = 1, \end{aligned} \tag{3.10}$$

where $\mathbf{H}_{j,i}$ denotes the entry of the matrix $\mathbf{H}$ in the j-th row and the i-th column; and $\mathbf{c}_i$ denotes the i-th entry of the vector $\mathbf{c}$. Note that to maintain the stealthiness of the attack, that is, $\mathbf{a} = \mathbf{H}\mathbf{c}$, the cost in (3.10) is the same as in (3.5). The constraint in (3.10) guarantees that a solution $\mathbf{c}^*$ to (3.10) can be re-scaled to obtain an attack vector $\mathbf{a}^* = \mathbf{a}_k\mathbf{H}\mathbf{c}^*$ such that the malicious data injected into measurement k is $\mathbf{a}_k$ where $\mathbf{a}_k$ is the kth entry of vector $\mathbf{a}$. It follows that the vector of compromised measurements is $\mathbf{y}_a = \mathbf{y} + \mathbf{a}^*$.

There is a large body of ℓ_1 relaxation approaches widely employed to propose strategic protection schemes. In general, the research on attack constructions with sparsity constraints gives the insight on the vulnerability of measurements. Given that the operator is likely to have access control policies in place [74], research on sparse attack constructions helps the operator evaluate the vulnerability of the measurements and allocate encryption devices on the measurements sensibly.

The first study on measurement vulnerability to DIAs is presented in [89] where the security indices help identifying the vulnerable measurements. The analysis of vulnerable measurements gives an insight on securing a subset of corresponding sensors. Strategic defensive and protection mechanisms against DIAs that choose and protect a subset of measurements are proposed in [11], [12] and [90].

Sparsity is one of the main constraints in attack constructions. This is what the thesis is primarily concerned with the study of sparsity constraints in the construction of attack vectors.

3.2 Attack Constructions with Incomplete Information

Lemma 1 states that attack constructions that are undetectable by a residual test obey the structure given by $\mathbf{a} = \mathbf{H}\mathbf{c}$, where $\mathbf{c}$ is the vector of difference between the state estimate under attacks and without attacks as shown in (2.47). In this setting, it is assumed that the attacker has perfect knowledge of the Jacobian matrix $\mathbf{H} \in \mathbb{R}^{m\times n}$ that is determined by the

system topology, system parameters and the operation point. However, an important and practical scenario is that the attacker has limited information or imperfect knowledge about the Jacobian matrix, for example, the lack of real-time knowledge with respect to various grid parameters and attributes such as the position of circuit breaker switches, transformer tap changers, limited physical access to the facilities in the system, etc. The relaxation of the perfect knowledge about the Jacobian matrix $\mathbf{H}$ yields *attacks with incomplete information*, that is, the Jacobian matrix $\bar{\mathbf{H}} \in \mathbb{R}^{m\times n}$ that the attacker has access to during the construction of the attack vector is

$$\bar{\mathbf{H}} = \mathbf{H} + \Delta\mathbf{H}, \tag{3.11}$$

where $\Delta\mathbf{H} \in \mathbb{R}^{m\times n}$ is the Jacobian matrix modelling error that describes the model mismatch due to imperfect knowledge or mismatched information.

The research on attack constructions with line admittance uncertainty is firstly studied in [46]. The resulting attack construction denoted by $\bar{\mathbf{a}} \in \mathbb{R}^m$ is as follows:

$$\begin{aligned}
\bar{\mathbf{a}} &= \bar{\mathbf{H}}\mathbf{c} && (3.12)\\
&= (\mathbf{H} + \Delta\mathbf{H})\mathbf{c} && (3.13)\\
&= \mathbf{H}\mathbf{c} + \Delta\mathbf{H}\mathbf{c}. && (3.14)
\end{aligned}$$

Let $\bar{\mathbf{x}}_a \in \mathbb{R}^n$ be the vector of state estimate obtained from the compromised measurements with attack vector in (3.12). From Corollary 0.1, the LS estimate is

$$\begin{aligned}
\bar{\mathbf{x}}_a &= \hat{\mathbf{x}} + (\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}\bar{\mathbf{a}} && (3.15)\\
&= \hat{\mathbf{x}} + (\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}(\mathbf{H}\mathbf{c} + \Delta\mathbf{H}\mathbf{c}) && (3.16)\\
&= \hat{\mathbf{x}} + \mathbf{c} + (\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}\Delta\mathbf{H}\mathbf{c} && (3.17)\\
&= \hat{\mathbf{x}} + \bar{\mathbf{c}}, && (3.18)
\end{aligned}$$

where,

$$\bar{\mathbf{c}} \triangleq \mathbf{c} + (\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}\Delta\mathbf{H}\mathbf{c} \tag{3.19}$$

is the actual injected error to the state variables. Let $\bar{\mathbf{r}}_a \in \mathbb{R}^m$ be the residual vector under attacks in (3.12). The following holds

$$\begin{aligned}
\bar{\mathbf{r}}_a &= \bar{\mathbf{y}}_a - \mathbf{H}\bar{\mathbf{x}}_a && (3.20)\\
&= \mathbf{y} + \bar{\mathbf{a}} - \mathbf{H}(\hat{\mathbf{x}} + \bar{\mathbf{c}}) && (3.21)\\
&= \mathbf{y} - \mathbf{H}\hat{\mathbf{x}} + \bar{\mathbf{a}} - \mathbf{H}\bar{\mathbf{c}} && (3.22)\\
&= \mathbf{r} + \mathbf{H}\mathbf{c} + \Delta\mathbf{H}\mathbf{c} - \mathbf{H}(\mathbf{c} + (\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}\Delta\mathbf{H}\mathbf{c}) && (3.23)\\
&= \mathbf{r} + \Delta\mathbf{H}\mathbf{c} - \mathbf{H}(\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}\Delta\mathbf{H}\mathbf{c} && (3.24)\\
&= \mathbf{r} + (\mathbf{I}_m - \Theta)\Delta\mathbf{H}\mathbf{c}, && (3.25)
\end{aligned}$$

where the vector $\mathbf{r} \triangleq \mathbf{y} - \mathbf{H}\hat{\mathbf{x}}$ is the residual without attacks; and the matrix Θ is defined as $\Theta \triangleq \mathbf{H}(\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}\mathbf{H}^\mathsf{T}$. From (3.25), the residual vector $\bar{\mathbf{r}}_a$ depends not only on the Jacobian modelling error $\Delta\mathbf{H}$ but also the actual Jacobian matrix $\mathbf{H}$.

From (2.29), the residual based bad data detection evaluates the following inequality

$$\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\ell_2}^2 \leq r_0, \tag{3.26}$$

where r_0 is the residual threshold defined in (2.30) and set by the system operator. Similarly, the attack detection for the attacks with incomplete information in (3.14) is to assess the following inequality

$$\|\bar{\mathbf{y}}_a - \mathbf{H}\bar{\mathbf{x}}_a\|_{\ell_2}^2 = \|\mathbf{y} - \mathbf{H}\mathbf{x} + (\mathbf{I}_m - \Theta)\Delta\mathbf{H}\mathbf{c}\|_{\ell_2}^2 \leq r_0. \tag{3.27}$$

Note that with $\Delta\mathbf{H} = \mathbf{0}$, that is, the attacker has access to the exact Jacobian matrix $\mathbf{H}$, the attack vector in (3.14) boils down to the case $\bar{\mathbf{a}} = \mathbf{H}\mathbf{c}$ and the detection in (3.27) boils down to the detection in (3.26) where the attack construction does not change the residual. With $\Delta\mathbf{H}\mathbf{c} = \mathbf{0}$, the attacker can also launch an attack with inaccurate information about $\mathbf{H}$ while not changing the residual. Furthermore, even with $\Delta\mathbf{H}\mathbf{c} \neq \mathbf{0}$, there exist attack constructions that do not change the residual, namely when (3.27) holds.

In [46, Th. 1], with $\Delta\mathbf{H} \neq \mathbf{0}$, the attacker can still implement successful DIAs with complete knowledge about the admittance when making at least one transmission between two buses disconnected so that it can pass the attack detection, that is, the inequality in (3.27) still holds. However, the error injected to the state estimate is limited as the injection vector $\mathbf{c}$ cannot be arbitrary as in the general case and needs to satisfy the constraint $\Delta\mathbf{H} = \mathbf{0}$. Besides, a measure of vulnerability of the system topology to DIAs is proposed to compare different topologies and identify more robust topologies when incomplete information is available to the attacker. Dropping the restriction of the knowledge studied in [46], a local load redistribution attack is proposed in [47] where the attacker only needs to obtain the information of the local attacking region to launch successful attacks. The result in [47] is extended to an AC state estimation based on a few measurements in the attacking region associated with boundary buses without knowing the full topology and parameter information in [48].

With the aim of minimizing the network information, an efficient strategy for determining the optimal feasible attacking region is proposed in [49]. A subspace method for DIAs with full and partial measurement models is proposed in [50]. Specifically, by exploiting the information that the measurements contain, an estimated system subspace is obtained and attack strategies that leverage this information are proposed accordingly. Conditions for the existence of an unobservable subspace attack are obtained. The impact of the attacks with incomplete information on the electricity market is also studied in [51] where the state of the system is modelled with stochastic uncertainty.

DIAs with incomplete information relax the assumption that the attackers have full and instantaneous access to grid topology and state. Unfortunately, incomplete information available to the attackers does not make power systems immune to DIAs [46]. Studies on attacks with incomplete information show that the attackers can launch valid attacks with limited information about the system [47, 48].

3.3 DIAs with Falsified Topology

Consistency and integrity of the information about system topology is a fundamental prerequisite for a safe and economic operation. In practice, the system topology can be changed

for the purpose of transmission line maintenance or outages. An accurate description of the system topology is obtained through the digital information of switches and breakers, that is, the on-off states of switches and breakers. The state of the switches and breakers is transmitted to the *Network Topology Processor* (NTP) that constructs the system topology [91]. Cyber topology attacks compromise the observations on the switches and breakers state which yields *falsified topologies*. Note that in cyber topology attacks, buses and transmission lines are not physically attacked, that is, the physical connection of the system does not change, but the digital information of switches and breakers is modified by the attacker.

In DIAs with falsified topology, to evade the topology error identification by the *topology error processing* (TEP), the attacker needs to simultaneously compromise the measurements and modify digital information about the switches and the breakers state. In other words, the attacker needs to maintain the consistency of the measurements and the falsified topology. For example, it is not possible that when a breaker of a power transmission line is off, the power flow measurement in this line is to be nonzero.

Let $d \in \mathbb{Z}_+$ be the number of switches and line breakers. Specifically, the switches and line breakers state are denoted by $\mathbf{s} \in \{0,1\}^d$ which indicates the on and off states of various switches and line breakers. The switches and line breakers state $\mathbf{s}$ corresponds to a graph $\mathcal{T} \triangleq (\mathcal{V}, \mathcal{B})$ where $\mathcal{V}$ is the set of buses and $\mathcal{B}$ is the set of connected buses pairs. Consequently, from (2.5), the vector of measurements is

$$Y^m = \bar{F}(\mathbf{x}) + Z^m, \tag{3.28}$$

where the function $\bar{F}$ is determined by the graph $\mathcal{T}$ and the system parameters; the system noise Z^m is as defined in (2.6).

The DIAs model with cyber topology attacks that modify the states of the switches and line breakers is

$$\bar{\mathbf{s}} = \mathbf{s} + \Delta\mathbf{s} \tag{3.29a}$$

$$\mathbf{y}_a = \mathbf{y} + \bar{\mathbf{a}}, \tag{3.29b}$$

where the vector $\Delta\mathbf{s}$ represents the modifications of the switches and line breakers that yield a modified topology $\bar{\mathcal{T}} \triangleq (\mathcal{V}, \bar{\mathcal{B}})$; the vector $\bar{\mathbf{s}}$ is the modified states; and the vector $\mathbf{y}_a \in \mathbb{R}^{\bar{m}}$ is the vector of compromised measurements with attacks $\bar{\mathbf{a}} \in \mathbb{R}^{\bar{m}}$. Note that the modified topology $\bar{\mathcal{B}}$ corresponds to an $\bar{m}$-dimensional vector of measurements. Note that it holds that $\bar{m} \neq m$ when the number of measurements is different after the topologies are falsified.

Let $\bar{\mathbf{H}}$ be the Jacobian matrix with the topology $\bar{\mathcal{T}}$ under the topology attacks $\Delta\mathbf{s}$. From Corollary 0.1, the LS estimate with falsified topology in (3.29) is

$$\hat{\mathbf{x}}_a^* = \arg\min_{\hat{\mathbf{x}}} \|\mathbf{y}_a - \bar{\mathbf{H}}\hat{\mathbf{x}}\|_{\ell_2}^2 = (\bar{\mathbf{H}}^\mathsf{T}\bar{\mathbf{H}})^{-1}\bar{\mathbf{H}}^\mathsf{T}\mathbf{y}_a, \tag{3.30}$$

where $\|\mathbf{y}_a - \bar{\mathbf{H}}\hat{\mathbf{x}}\|_{\ell_2}$ is the ℓ_2 norm of the vector $\mathbf{y}_a - \bar{\mathbf{H}}\hat{\mathbf{x}}$. Hence, the residual is

$$r = \|\mathbf{y}_a - \bar{\mathbf{H}}\hat{\mathbf{x}}_a^*\|_{\ell_2}^2, \tag{3.31}$$

where r is the resulting residual.

The first study on DIAs with falsified topology is in [52] with the aim of conveying falsified topology information to the operator. An attack to modify $\mathcal{T}$ to $\bar{\mathcal{T}}$ with the attack vector $\mathbf{a}$ is undetectable if

$$\mathbf{y} + \mathbf{a} \in \text{Col}(\bar{\mathbf{H}}), \text{for all } \mathbf{y} \in \text{Col}(\mathbf{H}), \tag{3.32}$$

where $\mathbf{H}$ and $\bar{\mathbf{H}}$ are the Jacobian matrix for $\mathcal{T}$ and $\bar{\mathcal{T}}$, respectively, and $\text{Col}(\mathbf{H})$ is the column space of $\mathbf{H}$. The following theorem states the necessary and sufficient condition for the existence of DIAs with falsified topology.

Theorem 5. *[52, Th. 3.2] There exists an undetectable attack that modifies $\mathcal{T}$ and $\bar{\mathcal{T}}$ with the subspace $\mathcal{A}$ of feasible attack vectors $\mathbf{a}$ if and only if $\text{Col}(\mathbf{H}) \subset \text{Col}(\bar{\mathbf{H}}, \mathcal{A})$, where $\text{Col}(\mathbf{H})$ is .*

Note that the assumption for Theorem 5 is that the attacker can observe all measurements and the state of all the switches and breakers. For an attacker with limited information, a heuristic method of undetectable attack is proposed in [52]. A comprehensive coordinated attack scenarios on cyber topology and DIAs that covers line-addition attack, line-removal attack and line-switching attack are proposed in [53]. The economic impact of this coordinated attack is studied in [54] and [92]. Research on the DIAs under the coordination with physical attacks where the attackers physically disconnect a power line and launch a topology-preserving attack using DIAs to mask the physical damage is referred as a *state and topology cyber-physical attack* (STCPA) [93]. The STCPA can cause severe overload on other power lines due to line outages [94, 95]. A bilevel model for analysing coordinated cyber-physical attacks is proposed in [96] to study the most damaging and undetectable physical attacks.

DIAs with falsified topology are typical coordinated attacks which can result in severe disturbance on system operation. This coordinated attack affects the state estimate with a falsified topology as well as the electricity market, for instance, via *locational marginal price* (LMP) [55].

3.4 Attack Constructions under AC Power Flow Model

The AC power flow model is described in Section 2.1.1 where the state variables are observed through the nonlinear function $F : \mathbb{R}^n \to \mathbb{R}^m$, that yields the observation model

$$Y^m = F(\mathbf{x}) + Z^m, \tag{3.33}$$

where the vector of measurements Y^m is corrupted by the noise vector Z^m introduced in (2.6).

Due to the nonlinearity between the measurements and the state variables, it is much more challenging to characterize analytically the formula of a successful attack vector. The attack construction in Lemma 1, that is, $\mathbf{a} = \mathbf{Hc}$, is an attack construction for a linear observation model where the system is assumed to be in an operation point under the assumptions in (2.13), (2.14) and (2.15). In other words, the Jacobian matrix $\mathbf{H} \in \mathbb{R}^{m \times n}$ is a fixed matrix that does not change over time. However, for the AC model, the relaxation of the assumptions yields a dynamic Jacobian matrix that depends on the operation point

where the system operates. Furthermore, for the AC model the LS state estimate in (2.24) becomes

$$\hat{\mathbf{x}}^* = \arg\min_{\hat{\mathbf{x}}} \|\mathbf{y} - F(\hat{\mathbf{x}})\|_{\ell_2}^2, \tag{3.34}$$

which is usually solved by iterative normal equation methods [4]. It is shown in [56] that the DIAs targeting the DC state estimator are easily detected when the operator actually adopts nonlinear state estimator.

However, attack constructions for which it holds that

$$\mathbf{a} = F(\mathbf{x}_a) - F(\mathbf{x}) \tag{3.35}$$

do not change the residual. The residual under attack in (3.35) is

$$\begin{aligned}
\|\mathbf{r}_a\|_{\ell_2}^2 =& \|\mathbf{y}_a - F(\hat{\mathbf{x}}_a)\|_{\ell_2}^2 && (3.36)\\
=& \|\mathbf{y}_a - F(\hat{\mathbf{x}}_a) + F(\mathbf{x}) - F(\mathbf{x})\|_{\ell_2}^2 && (3.37)\\
=& \|\mathbf{y} + \mathbf{a} - F(\hat{\mathbf{x}}_a) + F(\mathbf{x}) - F(\mathbf{x})\|_{\ell_2}^2 && (3.38)\\
=& \|\mathbf{y} - F(\mathbf{x}) + \mathbf{a} - F(\hat{\mathbf{x}}_a) + F(\mathbf{x})\|_{\ell_2}^2 && (3.39)\\
=& \|\mathbf{r} + \mathbf{a} - F(\hat{\mathbf{x}}_a) + F(\mathbf{x})\|_{\ell_2}^2 && (3.40)\\
=& r \leq r_0. && (3.41)
\end{aligned}$$

Therefore, the attack in (3.35) can pass the residual based bad data detection. However, the nonlinear state estimate is often obtained via a gradient descent methods for which convergence is not guaranteed, and therefore, the exact operation point cannot be analytically characterized.

3.5 Random Attacks within a Bayesian Framework

As discussed in Section 2.3, unprecedented data acquisition capabilities help to generate stochastic models for the system. Moreover, data analysis on the system depend on the reliability of the observations that are used with a variety of estimation, statistics, and machine learning tools that provide the operator with different insight about the system. However, the cybersecurity threats to the smart grid are not well understood yet. Therefore, practical insight needs to come forth combining technologies from information theory, statistics, and machine learning [14], etc. In view of this, it is essential to assess attacks in fundamental terms to understand the impact over a wide range of data analysis paradigms.

In the context of *stealth attack*, information theoretic attacks are first introduced in [16] and then generalized in [17]. Information loss caused by the attacks, that is, attack disruption is measured by mutual information (MI) that is denoted by $I(X^n; Y_A^m)$ and describes how much information the measurements Y_A^m contain about the state variables X^n [97]. The probability of detection is measured in terms of Kullback-Leibler (KL) divergence between the distributions of the measurements with attacks and without attacks denoted by $D(P_{Y_A^m} \| P_{Y^m})$. Note that $D(P_{Y_A^m} \| P_{Y^m})$ is relevant to probability of detection as $P_D \approx 1 - \exp\{-D(P_{Y_A^m} \| P_{Y^m})\}$ [97]. Hence, this information theoretic formulation yields

the following attack construction:

$$\begin{aligned} &\min_{P_A^m} I(X^n; Y_A^m) \\ &\text{s.t. } D(P_{Y_A^m} \| P_{Y^m}) \leq \delta, \end{aligned} \tag{3.42}$$

where δ is the constraint for $D(P_{Y_A^m} || P_{Y^m})$ set by the attacker; and P_A^m is the distribution of the attack vector A^m.

The tradeoff between mutual information and KL divergence is summed up as an optimization problem [16], [98] given by

$$\min_{P_A^m} I(X^n; Y_A^m) + D(P_{Y_A^m} \| P_{Y^m}). \tag{3.43}$$

The Lagrangian $\mathcal{L}$ associated with (3.42) is given by

$$\mathcal{L}(P_A^m, \lambda) = I(X^n; Y_A^m) + \lambda(D(P_{Y_A^m} \| P_{Y^m}) - \delta). \tag{3.44}$$

It follows that

$$\begin{aligned} \frac{\partial \mathcal{L}(P_A^m, \lambda)}{\partial P_A^m} &= \frac{\partial}{\partial P_A^m}(I(X^n; Y_A^m) + \lambda(D(P_{Y_A^m} \| P_{Y^m}) - \delta)) \\ &= \frac{\partial}{\partial P_A^m}(I(X^n; Y_A^m) + \lambda D(P_{Y_A^m} \| P_{Y^m})). \end{aligned} \tag{3.45}$$

The saddle point of the Lagrangian is the same as that of $I(X^n; Y_A^m) + \lambda D(P_{Y_A^m} || P_{Y^m})$. Thus, instead of summing the two measures up directly, the optimization problem in (3.42) is equivalent to [17]

$$\min_{P_A^m} I(X^n; Y_A^m) + \lambda D(P_{Y_A^m} \| P_{Y^m}), \tag{3.46}$$

where the optimization domain is the set of all possible m-dimensional Gaussian probability distributions; and $\lambda \geq 1$ is a weighting parameter that prioritizes remaining undetected over minimizing the amount of information obtained by the system measurements. By increasing the value of λ the attacker decreases the probability of detection at the expense of increasing the amount of information obtained by the system operator via the measurements. The generalized cost function in (3.46) with weighting parameter achieves an arbitrarily low probability of detection for the attacker.

The vector of random state variables X^n in Definition 7 is assumed to follow a multivariate Gaussian distribution with a null mean vector given by [17]

$$X^n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{XX}), \tag{3.47}$$

where $\mathbf{\Sigma}_{XX} \in S_+^n$ is a covariance matrix. Assume the vector of random state variables A^m in the random attack model (2.54) is a multivariate Gaussian distribution that satisfies

$$A^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{AA}), \tag{3.48}$$

where $\mathbf{0} = (0, 0, \ldots, 0)^\mathsf{T}$ and $\mathbf{\Sigma}_{AA} \in \mathcal{S}_+^m$ are the mean vector and the covariance matrix of the random vector A^m. Narrowing down the distribution of state variables and the random attacks to a Gaussian distribution yields the optimization problem in (3.46) as follows [17]

$$\min_{\mathbf{\Sigma}_{AA} \in \mathcal{S}_+^m} (1-\lambda) \log |\mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_{AA}| - \log |\sigma^2 \mathbf{I}_m + \mathbf{\Sigma}_{AA}| + \lambda \text{tr}(\mathbf{\Sigma}_{YY}^{-1} \mathbf{\Sigma}_{AA}), \tag{3.49}$$

where the optimization domain $\mathcal{S}_+^m$ is a convex set. The following proposition characterizes the convexity of the optimization problem in (3.49).

Proposition 1. *[17] Let $\lambda \geq 1$. The optimization problem in* (3.49) *is convex.*

The following theorem characterizes the solution to (3.49).

Theorem 6. *[17] Let $\lambda \geq 1$. The solution to the optimal Gaussian attack is given by*

$$\bar{P}_{A^m} \sim \mathcal{N}(\mathbf{0}, \bar{\mathbf{\Sigma}}), \tag{3.50}$$

with

$$\mathbf{\Sigma}_{AA}^* = \lambda^{-1/2}\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}. \tag{3.51}$$

Remark 1. *The construction of the stealth attack vector in Theorem 6 is not sparse, indeed all the entries of the realizations of the random attack vector are nonzero with probability one, that is, $\mathbb{P}\left[\text{card}\left(\text{supp}\left(A^m\right)\right) = m\right] = 1$. Note that this thesis defines the support of the attack vector A^m in* (3.3).

The attack detection based on LRT in (2.64) yields the probability of detection given by

$$P_D \stackrel{\Delta}{=} \mathbb{E}\left[\mathbb{1}_{\{L(Y_{A^m}) \geq \tau\}}\right]. \tag{3.52}$$

The following lemma particularizes the above expression to the optimal attack construction in (3.51).

Lemma 7. *[17, Lemma 1] The probability of detection of LRT in* (2.64) *with threshold τ for the attack construction in* (3.51) *is given by*

$$P_D(\lambda) = \mathbb{P}\left[(U^p)^\mathsf{T}\Delta U^p \geq \lambda\left(2\log\tau + \log\left|\boldsymbol{I}_p + \lambda^{-1}\Delta\right|\right)\right], \tag{3.53}$$

where $p = \text{rank}(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T})$, $U^p \in \mathbb{R}^p$ is a vector of random vaiables with distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$, and $\Delta \in \mathbb{R}^{p\times p}$ is a diagonal matrix with entries given by $(\Delta)_{ii} = \lambda_i(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T})\lambda_i(\mathbf{\Sigma}_{YY}^{-1})$ where $\lambda_i(\mathbf{A})$ with $i \in \{1, 2, \ldots, p\}$ denotes the i-th eigenvalue of the matrix $\mathbf{A}$ in descending order.

The following theorem provides a sufficient condition for λ to achieve a desired probability of attack detection.

Theorem 8. *Let $\tau > 1$ be the decision threshold of the LRT. For any $t > 0$ and $\lambda \geq \max(\lambda^*(t), 1)$, the probability of attack detection satisfies*

$$P_D(\lambda) \leq e^{-t}, \tag{3.54}$$

where $\lambda^*(t)$ is the only positive solution of λ satisfying

$$2\log\tau - \frac{1}{2\lambda}\text{tr}(\Delta^2) - 2\sqrt{\text{tr}(\Delta^2)t} - 2\|\Delta\|_\infty t = 0, \tag{3.55}$$

where $\|\Delta\|_\infty$ is the infinity norm of Δ.

Note that in [16, 17], the construction of the stealth attack requires that the attacker has access to all the measurements, indeed all the components of the attack realizations are nonzero with probability one, that is, $\mathbb{P}\left[\text{card}\left(\text{supp}\left(A^m\right)\right) = m\right] = 1$. Incorporating sparsity constraints with information theoretic attacks is still an open problem that requires novel approaches.

Apart from stealth attack constructions, information theoretic tools are adopted in sensor placement and smart sensor privacy, etc. In [99], a sensor placement strategy that considers the amount of information acquired by the sensors is discussed. Information theoretic privacy guarantees for smart meter users are studied for general random processes in [100].

3.6 Decentralized Attacks

In a decentralized system, a central controller does not exist. The DIAs are constructed by several attackers that have access to the measurements in the power system. This scenario is referred to as *decentralized attacks*. In this scenario, the attackers decide the attack vector $\mathbf{a} \in \mathbb{R}^m$ jointly. The aim of each attacker is to individually construct the attack vector that maximizes the damage they inflict to the system, e.g., distortion to the state estimate, while staying undetected. All the attackers have the same interest, which reveals an alignment of actions among the attackers in decentralized attacks.

Let $\mathcal{K} = \{1, 2, \dots K\}$ be the set of attackers and $\mathcal{C}_i \in \{1, 2, \dots, m\}$ be the set of measurements that the attacker $i \in \mathcal{K}$ can compromise. Let also the vector $\mathbf{a}_i \in \mathcal{A}_i$ be the attack vector by attacker i such that

$$\mathcal{A}_i = \{\mathbf{a}_i \in \mathbb{R}^m : (\mathbf{a}_i)_j = 0 \text{ for all } j \notin \mathcal{C}_i\}. \tag{3.56}$$

The overall attack vector injected to the system is formed by the attack vectors of all the attackers. The system model for the case with decentralized attacks is, in general, the same as in the centralized case described in Section 2.2 and Section 2.3, respectively. The main differences are:

• Each attacker has access to different sets of measurements, that is, $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_K$ are proper sets and form a partition of the set $\mathcal{M} \triangleq \{1, 2, \dots, m\}$.

• The attack vector is determined by all the attackers, that is,

$$\mathbf{a} = \sum_{i \in \mathcal{K}} \mathbf{a}_i. \tag{3.57}$$

Decentralized attacks are proposed in distributed systems [101], blockchain-based decentralized finance attacks [102], etc. In [31], the decentralized attack constructions are developed in a game framework. This is of particular interest of this thesis since this thesis explores both the sparsity and coordination constraints in data injection attacks. In a game setting, both the sparsity and coordination constraints can be adopted naturally. Therefore, in this section, the following results are particularly visited. The cooperative interaction among the attackers in the system is modelled by a game in a normal form in [103]:

$$\mathcal{G} = (\mathcal{K}, \{\mathcal{A}_i\}_{i \in \mathcal{K}}, \Phi), \tag{3.58}$$

where $\mathcal{K}$ is the set of players, the set $\mathcal{A}_i$ is the set of vectors that attacker i can take and u_i is the utility for attacker i.

A function $\Phi(\mathbf{a})$ is chosen as the utility function in the game in [31] considering that an attack is said to be successful if it induces a nonzero distortion and it is not detectable. Specifically, the utility is

$$\Phi(\mathbf{a}) = \mathbb{P}_{\mathrm{ND}}(\mathbf{a})\mathbf{x}_a^{\mathsf{T}}\mathbf{x}_a, \tag{3.59}$$

where $\mathbb{P}_{\mathrm{ND}}(\mathbf{a})$ is the probability of non-detection given an attack vector $\mathbf{a}$, the vector $\mathbf{x}_a$ is the excess distortion induced by the attack $\mathbf{a}$. Under the assumption that $X^n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{XX})$, from Lemma 2, the excess distortion $\mathbf{x}_a$ is

$$\mathbf{x}_a = \mathbf{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}} + \sigma^2\mathbf{I}_m)^{-1}\mathbf{a}, \tag{3.60}$$

where the matrix $\mathbf{H}$ is in (2.49) and the real σ is the variance of the system noise.

The following proposition presents the analytical expression of the probability of non-detection $\mathbb{P}_{\mathrm{ND}}(\mathbf{a})$.

Proposition 2. *[31] For all $\mathbf{a} \in \mathcal{A}$, it holds that*

$$\mathbb{P}_{\mathrm{ND}}(\mathbf{a}) = \frac{1}{2}\mathrm{erfc}\left(\frac{\frac{1}{2}\mathbf{a}^{\mathsf{T}}\mathbf{\Sigma}_{YY}^{-1}\mathbf{a} + \log\tau}{\sqrt{2\mathbf{a}^{\mathsf{T}}\mathbf{\Sigma}_{YY}^{-1}\mathbf{a}}}\right), \tag{3.61}$$

where $\mathbf{\Sigma}_{YY} \triangleq \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}} + \sigma^2\boldsymbol{I}_m$ and τ is the threshold for the LRT in (2.64) *in Lemma 3.*

The benefit $\Phi(\mathbf{a})$ obtained by attacker $i \in \mathcal{K}$ does not only depend on its own attack vector $\mathbf{a}_i$ but also on the attack vectors by all the other attackers as shown in (3.57). Therefore, given the attack vector by all the other attackers except attacker i results in

$$\mathbf{a}_{-i} = \sum_{j\in\mathcal{K}, j\neq i} \mathbf{a}_j, \tag{3.62}$$

the attacker i aims to construct an attack vector $\mathbf{a}_i$ to maximize the benefit $\Phi(\mathbf{a})$, that is,

$$\mathbf{a}_i \in \mathrm{BR}_i(\mathbf{a}_{-i}), \tag{3.63}$$

where BR_i is the best response correspondence such that

$$\mathrm{BR}_i(\mathbf{a}_{-i}) = \arg\max_{\mathbf{a}_i\in\mathcal{A}_i} \Phi\left(\mathbf{a}_i + \mathbf{a}_{-i}\right). \tag{3.64}$$

A game solution that is particularly relevant for this analysis is the *Nash Equilibrium* [103].

Definition 9 (Nash Equilibrium)**.** *The attack vector $\mathbf{a}$ is an NE of the game if and only if it is a solution of the fixed point equation*

$$\mathbf{a} = \mathrm{BR}\left(\mathbf{a}\right), \tag{3.65}$$

with BR *as the global best response correspondence, that is,*

$$\mathrm{BR}\left(\mathbf{a}\right) = \mathrm{BR}_1\left(\mathbf{a}_{-1}\right) + \mathrm{BR}_2\left(\mathbf{a}_{-2}\right) + \ldots + \mathrm{BR}_K\left(\mathbf{a}_{-K}\right). \tag{3.66}$$

The following propositions highlight the properties of the game [31].

Proposition 3. *The game in* (3.58) *with utility function in* (3.59) *is a potential game.*

Proposition 4. *The game in* (3.58) *with utility function in* (3.59) *possesses at least one NE.*

The following lemma characterizes the achievability of NE attacks [31, Lemma 1].

Lemma 9. *Any best response dynamic (BRD) in the game converges to an attack vector that is an NE.*

The following theorem bounds the number of NEs in the game [31, Th. 1].

Theorem 10. *Let* $\mathcal{A}_{\mathrm{NE}}$ *be the set of all the DIAs that form NEs. The cardinality of the set* $\mathcal{A}_{\mathrm{NE}}$ *of NE of the game satisfies*

$$2 \geq \mathrm{card}\left(\mathcal{A}_{\mathrm{NE}}\right) \leq C \cdot \mathrm{rank}(\mathbf{H}), \tag{3.67}$$

where $C < \infty$ *is a constant that depends on* τ *and* $\mathbf{H}$ *is in* (2.49).

The results in [31] provide the maximum distortion MMSE attacks. However, the decentralized methods are proposed in multiple research perspectives such as resilient secondary control again DIAs [104], dynamic estimator [105], attack detection [106], control for neural networks subject to cyber attacks [107], etc.

3.7 Summary

DIAs are the main threats faced by modern power system with the unprecedented data acquisition and transition capabilities, more generally, faced by Cyber-physical systems. Instead of studying the defense strategies and the impacts of DIAs, the research on DIAs constructions with constraints give the insights on the sensor vulnerability and the protection strategies in the first place. Sparsity constraints are one of the main constraints in DIAs. Sparse attack constructions can be developed in a decentralized scenario where the coordination between attackers can be explored. This is the focus of this research.

Chapter 4

Independent Sparse Stealth Attacks

This chapter presents the main results on the independent sparse stealth attacks in centralized systems. Specifically, this chapter has developed independent random attack constructions with sparsity constraints that operate on a Bayesian framework. The attacker minimizes the mutual information between the state variables and the compromised measurements to disrupt procedure that uses measurements. Simultaneously, the KL divergence between the distributions of measurements with attacks and without attacks is minimized to guarantee the stealthiness of the data injection events. The attack construction is cast as an optimization problem that jointly minimizes the weighted sum of mutual information and KL divergence. A closed form expression of the optimal measurement to be targeted by the attacker when a single measurement is compromised is obtained. Following this result, a general k-sparse attack construction that leverages the insight distilled in the single measurement attack case is proposed. The k-sparse attack construction is based on a greedy procedure that sequentially selects the measurements to attack by minimizing the cost in terms of the optimal decision at each step. This chapter has numerically assessed the performance of the proposed independent attack constructions on the IEEE test systems and observed that mutual information decreases linearly while the probability of attack detection exhibits a threshold effect when a critical number of measurements are compromised.

4.1 Bayesian Framework for State Estimation

4.1.1 State Variables and Attack Model

The observation model with linearized dynamics within Bayesian framework is given by Definition 7, that is,

$$Y^m = \mathbf{H}X^n + Z^m, \tag{4.1}$$

where the Jacobian matrix $\mathbf{H} \in \mathbb{R}^{m\times n}$ is defined in (2.19) that is determined by the system components and the topology; the vector $X^n \in \mathbb{R}^n$ is the vector of random state variables that describe the phase angle of the buses as defined in (2.16); the vector $Y^m \in \mathbb{R}^m$ is the vector of random measurements defined in (2.49) that are corrupted by Additive White Gaussian Noise (AWGN) introduced by the sensors [4, 5]. Such noise is modelled by the

random vector $Z^m \in \mathbb{R}^m$ that follows a multivariate Gaussian distribution, that is,

$$Z^m \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_m), \tag{4.2}$$

where σ^2 is the noise variance.

In a Bayesian framework, the state variables are described by a random vector X^n with a given distribution. The choice of the distribution has been studied in the literature. It is shown in [22] that the bus voltages of a low voltage system in the Northwest of England are well described by a multivariate Gaussian distribution from the voltage data provided by Electricity North West Limited (ENWL). In [108], the measurements from real power grids follow a joint Gaussian distribution. Therefore, in this study, the vector of state variables X^n is assumed to follow a multivariate Gaussian distribution with a null mean vector and a covariance matrix $\mathbf{\Sigma}_{XX}$, that is,

$$X^n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{XX}), \tag{4.3}$$

where $\mathbf{\Sigma}_{XX}$ is the covariance matrix such that $\mathbf{\Sigma}_{XX} \in S_+^n$ where S_+^n is the set of positive semi-definite matrices of dimension $n \times n$.

Hence, the vector of random measurements Y^m in (4.1) follows a multivariate Gaussian distribution with a null mean vector and a covariance matrix $\mathbf{\Sigma}_{YY}$, that is,

$$Y^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{YY}), \tag{4.4}$$

where

$$\mathbf{\Sigma}_{YY} \triangleq \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T} + \sigma^2 \mathbf{I}_m. \tag{4.5}$$

The resulting measurements are corrupted by a vector of malicious random attack $A^m \in \mathbb{R}^m$ with distribution P_{A^m}, that is, $A^m \sim P_{A^m}$. Consequently, the observation model under attacks is given in definition 8, that is,

$$Y_A^m = \mathbf{H}X^n + Z^m + A^m. \tag{4.6}$$

In this study, the attack construction aims at minimizing the mutual information between the random state variables X^n and the compromised random measurements denoted by Y_A^m, that is,

$$\min_{P_{A^m}} I(X^n; Y_A^m). \tag{4.7}$$

Hence, it is assumed that

$$A^m \sim \mathcal{N}(\boldsymbol{\mu}_A, \mathbf{\Sigma}_{AA}), \tag{4.8}$$

where $\boldsymbol{\mu}_A \in \mathbb{R}^m$ is the mean vector of the random attack vector A^m.

The choice in (4.8) is justified by the fact that when $Z^m + A^m$ in (4.6) follows a Gaussian distribution, the mutual information between the state variables X^n and the compromised measurements Y_A^m in (4.7) is minimized [109]. Hence, from the Lévy-Cramér decomposition theorem [110,111], it holds that for the sum $Z^m + A^m$ to be Gaussian, given that Z^m satisfies (4.2), then, A^m must also be Gaussian distributed.

Remark 2. *The fact that the Gaussian distribution achieves the minimum in* (4.7) *holds for any* P_{X^n}*. The assumption in* (4.8) *implies that the attack construction does not require access to the realizations of the state variables, but rather to the mean and second order statistics.*

From (4.4), (4.6) and (4.8), the compromised random measurements Y_A^m satisfies that

$$Y_A^m \sim \mathcal{N}(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{Y_AY_A}), \tag{4.9}$$

where

$$\boldsymbol{\Sigma}_{Y_AY_A} \triangleq \mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T} + \sigma^2\mathbf{I}_m + \boldsymbol{\Sigma}_{AA}. \tag{4.10}$$

4.1.2 Attack Detection

As described in Section 2.3.3, the attack detection with Bayesian framework is cast as a hypothesis testing problem. Specifically, given the distributions of the measurements without and with attacks in (4.4) and (4.9), respectively, the hypotheses are

$$\mathcal{H}_0\text{: } \bar{Y}^m \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{YY}), \tag{4.11}$$
$$\mathcal{H}_1\text{: } \bar{Y}^m \sim \mathcal{N}(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{Y_AY_A}). \tag{4.12}$$

From Lemma 3, the optimal detection method is likelihood ratio test (LRT), that is,

$$T(\bar{\mathbf{y}}) = \mathbb{1}_{\{L(\bar{\mathbf{y}})\geqslant\tau\}}, \tag{4.13}$$

where $T(\bar{\mathbf{y}})$ is the likelihood ratio given by

$$L(\bar{\mathbf{y}}) = \frac{f_{Y_A^m}(\bar{\mathbf{y}})}{f_{Y^m}(\bar{\mathbf{y}})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau, \tag{4.14}$$

where the functions $f_{Y_A^m}$ and f_{Y^m} are the pdf of $\mathcal{N}(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{Y_AY_A})$ in (4.9) and the pdf of $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{YY})$ in (4.4), respectively, the parameter τ is the decision threshold set by the operator that meets the false alarm constraint.

4.2 Information Theoretic Metrics

The aim of the attacker is twofold. First, it aims to inflict a data integrity attack that disrupts all processes that use the measurements of the system, and secondly, it aims to guarantee the stealth of the attack. Hence, instead of assuming a particular state estimation procedure, this chapter adopts the methodology in [17] to construct stealth attacks that minimize the amount of information acquired by the measurements about the state variables. In doing so, the attacker targets a universal utility metric consisting of a weighted sum of two terms [19]: (a) the mutual information between the state variables and the measurements; and (b) the Kullback Leibler (KL) divergence between the probability distribution functions of the measurements with and without attacks. By minimizing this metric, the attacker guarantees a stealthy attack that impinges upon any procedure using the measurements.

4.2.1 Disruption Measure

The central purpose of DIAs is to disrupt the procedure where the measurements are used. DIAs within Bayesian Framework do not assume any particular state estimation procedure.

Instead, the state variables and the attacks are modelled as random variables. Mutual information is a measure of the amount of information between two random variables. The random attack construction within Bayesian framework in (4.6) allows the attacker to choose the distribution of the attack vector A^m. For that reason, the disruption of the attack is characterized by the mutual information in fundamental terms by characterizing the amount of information between the state variables and the compromised measurements. Therefore, the attacker chooses the distribution of the random attack vector such that

$$\min_{P_{A^m}} I(X^n; Y_A^m), \tag{4.15}$$

where X^n is in (4.3) and Y_A^m is in (4.9). In fact, the motivation to use information measures is to provide general performance metrics that apply to a wide range of estimation, control, and analysis procedures. Any reduction on mutual information necessarily implies a degradation of performance in any process that uses the measurements. For instance, the performance of the MMSE estimator in (2.51) can be linked to the mutual information via I-MMSE results that bridge the classical estimation paradigm to the information theoretic one [18].

The following proposition presents the analytical expression of mutual information with X^n in (4.3) and Y_A^m in (4.9).

Proposition 5. *The mutual information between the random variable* $X^n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{XX})$ *and* $Y_A^m \sim \mathcal{N}(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{Y_AY_A})$ *is*

$$I(X^n; Y_A^m) = \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_{XX}||\boldsymbol{\Sigma}_{Y_AY_A}|}{|\boldsymbol{\Sigma}|}, \tag{4.16}$$

where $\boldsymbol{\Sigma}_{XX}$ *is in* (4.3)*;* $\boldsymbol{\Sigma}_{Y_AY_A}$ *is in* (4.10) *and* $\boldsymbol{\Sigma}$ *is the covariance matrix of the joint distribution of* X^n *and* Y_A^m*, that is,* $(X^n; Y_A^m) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ *with*

$$\boldsymbol{\Sigma} \triangleq \begin{pmatrix} \boldsymbol{\Sigma}_{XX} & \boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T} \\ \mathbf{H}\boldsymbol{\Sigma}_{XX} & \mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T} + \sigma^2 \boldsymbol{I}_m + \boldsymbol{\Sigma}_{AA} \end{pmatrix}. \tag{4.17}$$

Proof. The proof is presented in Appendix A. □

Corollary 10.1. *The mutual information between the vector of random variables* $X^n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{XX})$ *and* $Y_{A^m} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{Y_AY_A})$

$$I(X^n; Y_A^m) = \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_{XX}||\boldsymbol{\Sigma}_{Y_AY_A}|}{|\boldsymbol{\Sigma}|}, \tag{4.18}$$

where $\boldsymbol{\Sigma}_{XX}$ *is in* (4.3)*;* $\boldsymbol{\Sigma}_{Y_AY_A}$ *is in* (4.10) *and* $\boldsymbol{\Sigma}$ *is in* (4.17)*.*

4.2.2 Detection Metric

Apart from the disruption that is captured by mutual information, the stealthiness of the attacks is guaranteed by minimizing the Kullback-Leibler (KL) divergence. The KL divergence between two probability distributions is a measure of the statistical difference between the distributions. The rationale for minimizing the KL divergence between the distributions

as means to minimize the probability of attack detection stems from the Chernoff-Stein Lemma [19, Th. 11.8.3]. That is, for the LRT in (4.13) in Lemma 3, for any probability of Type I error $\alpha < \frac{1}{2}$, the logarithm of the averaged minimum value of probability of Type II error β asymptotically converges to the inverse of the KL divergence between the distributions of the two hypothesis. Therefore, minimizing the asymptotic detection probability is equivalent to maximizing the probability of Type II error that is achieved by

$$\min_{P_{A^m}} D(P_{Y_A^m} \| P_{Y^m}), \tag{4.19}$$

where $P_{Y_A^m}$ and P_{Y^m} are the distribution of Y_A^m in (4.9) and the distribution of Y^m in (4.4), respectively.

The following proposition presents the analytical expression of the KL divergence.

Proposition 6. *The KL divergence between $Y_A^m \sim \mathcal{N}(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{Y_AY_A})$ and $Y^m \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{YY})$ is*

$$D(P_{Y_A^m} \| P_{Y^m}) = \frac{1}{2}\left(\log \frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{Y_AY_A}|} - m + tr\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{Y_AY_A}\right) + tr\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\mu}_A\boldsymbol{\mu}_A^{\mathsf{T}}\right)\right), \tag{4.20}$$

where the mean vector $\boldsymbol{\mu}_A$ and the matrix $\boldsymbol{\Sigma}_{Y_AY_A}$ are in (4.9)*, the matrix $\boldsymbol{\Sigma}_{YY}$ is in* (4.4)*.*

Proof. The proof of Proposition 6 is presented in Appendix B. □

Corollary 10.2. *The KL divergence between m-dimensional multivariate Gaussian distributions $Y_{A^m} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{Y_AY_A})$ and $Y^m \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{YY})$ is given by*

$$D(P_{Y_A^m} \| P_{Y^m}) = \frac{1}{2}\left(\log \frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{Y_AY_A}|} - m + \mathrm{tr}(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{Y_AY_A})\right), \tag{4.21}$$

where the matrices $\boldsymbol{\Sigma}_{YY}$ and $\boldsymbol{\Sigma}_{Y_AY_A}$ are in (4.4) *and* (4.9)*, respectively.*

The following lemma shows that the optimal Gaussian attack construction is with a null mean vector.

Lemma 11. *The optimal Gaussian attack construction is with a null mean vector, that is,*

$$A^m \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{AA}), \tag{4.22}$$

where $\boldsymbol{\Sigma}_{AA} \in \mathcal{S}_+^m$.

Proof. From Proposition 5, the mutual information in (4.16) is not a function of the mean vector $\boldsymbol{\mu}_A$, that is, the mean vector is arbitrary. From Proposition 6, the following holds

$$\begin{aligned} D(P_{Y_A^m} \| P_{Y^m}) =& \frac{1}{2}\left(\log \frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{Y_AY_A}|} - m + \mathrm{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{Y_AY_A}\right) + \mathrm{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\mu}_A\boldsymbol{\mu}_A^{\mathsf{T}}\right)\right) && (4.23)\\ =& \frac{1}{2}\left(\log \frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{Y_AY_A}|} - m + \mathrm{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{Y_AY_A}\right) + \boldsymbol{\mu}_A^{\mathsf{T}}\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\mu}_A\right) && (4.24)\\ \geq& \frac{1}{2}\left(\log \frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{Y_AY_A}|} - m + \mathrm{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{Y_AY_A}\right)\right), && (4.25) \end{aligned}$$

where (4.24) follows from the fact that

$$\begin{aligned} \operatorname{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\mu}_A\boldsymbol{\mu}_A^{\mathsf{T}}\right) =& \operatorname{tr}\left(\boldsymbol{\mu}_A^{\mathsf{T}}\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\mu}_A\right) && (4.26)\\ =& \boldsymbol{\mu}_A^{\mathsf{T}}\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\mu}_A, && (4.27) \end{aligned}$$

and (4.25) follows from $\boldsymbol{\Sigma}_{YY}^{-1} \in \mathcal{S}_+^m$. Note that the equality in (4.25) holds when $\boldsymbol{\mu}_A = \mathbf{0}$. Therefore, for all $\boldsymbol{\Sigma}_{AA} \in \mathcal{S}_+^m$, the optimal mean is $\boldsymbol{\mu}_A = \mathbf{0}$. This completes the proof. □

Remark 3. *The assumption in* (4.22) *boils down the attack construction to simply characterize the covariance matrix* $\boldsymbol{\Sigma}_{AA}$.

The Gaussian attack construction $A^m \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{AA})$ incorporates mutual information in (4.18) and the KL divergence in (4.21) which results in the construction of *stealth attacks* [17]. Specifically, the construction is given by the solution to the following optimization problem:

$$\min_{P_{A^m}} I(X^n; Y_A^m) + \lambda D(P_{Y_A^m} \| P_{Y^m}), \tag{4.28}$$

where $\lambda \geq 1$ is the weighting parameter that determines the tradeoff between attack disruption and probability of detection. Note that the optimization in (4.28) searches for the distribution of the random attack vector over the set of Gaussian multivariate distributions of m dimensions. Equivalently, for Lemma 11, it chooses the optimal covariance matrix for the distribution of the attack. It is shown in Theorem 6 that the optimal Gaussian attack is given by $\bar{P}_{A^m} = \mathcal{N}(\mathbf{0}, \bar{\boldsymbol{\Sigma}})$ where

$$\bar{\boldsymbol{\Sigma}} = \lambda^{-\frac{1}{2}}\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}. \tag{4.29}$$

From remark 1, the attack construction in (4.29) is such that

$$\mathbb{P}\left[\operatorname{card}\left(\operatorname{supp}\left(A^m\right)\right) = m\right] = 1, \tag{4.30}$$

where $\operatorname{supp}(A^m)$ is the support of vector A^m defined in (3.3). This implies that the attacker has to compromise all the measurements, which is costly and unrealistic. In the next section, the attack constructions with sparsity constraint are proposed.

4.3 Sparse Stealth Attack Formulation

4.3.1 Attack Construction with Sparsity Constraints

The attack implementation requires access to the sensing infrastructure of the industrial control system (ICS) operating the power system. DIAs usually exploit the vulnerabilities existing in the field zone by comprising remote terminal units or local secondary level control systems, or alternatively, by getting access to the SCADA system coordinating the control zone of the ICS. For that reason, attack constructions that are required to intrude the least amount of monitoring and data acquisition infrastructure are of particular interest from a security standpoint. In view of this, this chapter studies sparse attacks that require access to a limited number of sensors, that is, the attack construction problem is posed with sparsity

constraints by setting the domain as the set of distributions over the attack vector that put non-zero mass on at most $k \leq m$ attack vector entries, that is,

$$\tilde{\mathcal{P}}_k \stackrel{\Delta}{=} \{P_{A^m} : \text{card}\,(\text{supp}(A^m)) = k\}\,. \tag{4.31}$$

The same information theoretic metrics as in (4.28) are considered. The resulting k-sparse stealth attack construction is therefore posed as the optimization problem:

$$\min_{P_{A^m} \in \tilde{\mathcal{P}}_k} I(X^n; Y_A^m) + \lambda D(P_{Y_A^m} \| P_{Y^m}). \tag{4.32}$$

Solving this problem is hard in general owing to the combinatorial nature of the attack vector support selection.

4.3.2 Gaussian Sparse Stealth Attack Construction

The optimization domain including the sparsity constraint in (4.32) implies an additional difficulty in the construction of stealth attacks with respect to the construction proposed in (4.28) [17]. This additional difficulty lies on the combinatorial problem arising from the selection of at most k out of m dimensions of the attack vector to form the support of A^m, that is, an additional optimization constraint of the form $\text{card}\,(\text{supp}(A^m)) = k$, with $k \leqslant m$. To tackle this difficulty, this chapter exploits the structure that the Gaussian attack embeds into the sparse attack problem formulation to propose novel attack construction algorithms with verifiable performance guarantees. From Lemma 11 it follows that the optimal Gaussian distribution is a null mean vector. Hence, the attacker chooses the distribution over the set of multivariate Gaussian distributions given by

$$\mathcal{P}_k \stackrel{\Delta}{=} \left\{P_{A^m} \sim \mathcal{N}(\mathbf{0}, \bar{\mathbf{\Sigma}}) : \text{card}\,(\text{supp}(A^m)) = k\right\}, \tag{4.33}$$

where $\bar{\mathbf{\Sigma}} \in \mathcal{S}_+^m$ is the covariance matrix of the multivariate Gaussian distribution. The resulting k-sparse stealth Gaussian attack construction is therefore posed as the optimization problem:

$$\min_{P_{A^m} \in \mathcal{P}_k} I(X^n; Y_A^m) + \lambda D(P_{Y_A^m} \| P_{Y^m}). \tag{4.34}$$

Hence, writing the objectives of the optimization problem in (4.34), that is, mutual information $I(X^n; Y_A^m)$ in 4.18 and the KL divergence $D(P_{Y_A^m} \| P_{Y^m})$ in 4.21 in terms of the covariance matrix of the attack random vector A^m in (4.22), that is, $A^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{AA})$ leads to observing that, up to a constant additive term, it is equivalent to the expression

$$J(\mathbf{\Sigma}_{AA}) \stackrel{\Delta}{=} (1-\lambda) \log |\mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_{AA}| - \log |\sigma^2 \mathbf{I}_m + \mathbf{\Sigma}_{AA}| + \lambda \text{tr}(\mathbf{\Sigma}_{YY}^{-1} \mathbf{\Sigma}_{AA}), \tag{4.35}$$

where $\lambda \geq 1$ is introduced in (4.34), the matrix $\mathbf{\Sigma}_{YY}$ is in (4.4), the real $\sigma \in \mathbb{R}_+$ is in (4.2).

Hence, the optimization problem in (4.34) is equivalent to the following optimization problem:

$$\min_{\mathbf{\Sigma}_{AA} \in S_+^m} J(\mathbf{\Sigma}_{AA}). \tag{4.36}$$

However, the sparsity constraint in (4.34) is not specified in (4.36). In order to write the optimization domain of the problem in (4.34) in terms of the covariance matrix of the

random attack vector, it suffices to observe that the sparsity constraint denoted in (4.33) translates into a constraint on the number of nonzero entries in the diagonal of the covariance matrix of the attack vector. More specifically, the optimization domain becomes:

$$\mathcal{S}_k \stackrel{\Delta}{=} \left\{ \mathbf{S} \in S_+^m : \|\mathrm{diag}(\mathbf{S})\|_0 = k \right\}, \tag{4.37}$$

where $\mathrm{diag}(\mathbf{S})$ denotes the vector formed by the diagonal entries of $\mathbf{S}$.

The following lemma presents the equivalent expression of the optimization problem in (4.34).

Lemma 12. *The optimization problem in* (4.34) *is equivalent to*

$$\min_{\mathbf{\Sigma}_{AA} \in \mathcal{S}_k} J(\mathbf{\Sigma}_{AA}), \tag{4.38}$$

where the optimization domain $\mathcal{S}_k$ *is in* (4.37) *and the cost function* J*:* $\mathcal{S}_k \to \mathbb{R}_+$ *is in* (4.35)*.*

Proof. The proof follows by noting that in the Gaussian setting, the optimization problem in (4.34) is equivalent to (4.36) up to a constant additive term and the optimization domain is specified by (4.37). □

4.4 Independent Sparse Stealth Attacks

This section tackles the case in which the entries of the attack vector A^m are independent. More specifically, the focus is on product probability measures of the form

$$P_{A^m} = \prod_{i=1}^{m} P_{A_i}, \tag{4.39}$$

where A_i denotes the i-th entry of the vector A^m; for all $i \in \{1, 2, \ldots, m\}$, the probability density function of A_i denoted by P_{A_i} is Gaussian with zero mean and variance v_i, that is, $A_i \sim \mathcal{N}(0, v_i)$.

The assumption of independence relaxes the correlation requirements between the entries of the attack vector. As a result, the set of covariance matrices given by (4.37), with $k \leqslant m$, that arises from considering Gaussian attacks becomes the following set

$$\tilde{\mathcal{S}}_k \stackrel{\Delta}{=} \bigcup_{\mathcal{K}} \left\{ \mathbf{S} \in S_+^m : \mathbf{S} = \sum_{i \in \mathcal{K}} v_i \mathbf{e}_i \mathbf{e}_i^{\mathsf{T}} \text{with } v_i \in \mathbb{R}_+ \right\}, \tag{4.40}$$

where the union is over all subsets $\mathcal{K} \subseteq \{1, 2, \ldots, m\}$ with $\mathrm{card}\,(\mathcal{K}) = k \leq m$. Note that it holds that $\tilde{\mathcal{S}}_k \subseteq \mathcal{S}_k$.

Corollary 12.1. *Under the independence assumption, the optimization problem in* (4.38) *boils down to the following optimization problem:*

$$\min_{\mathbf{\Sigma}_{AA} \in \tilde{\mathcal{S}}_k} J(\mathbf{\Sigma}_{AA}), \tag{4.41}$$

where the optimization domain $\tilde{\mathcal{S}}_k$ *is in* (4.40) *and the cost function* $J : \tilde{\mathcal{S}}_k \to \mathbb{R}_+$ *is in* (4.35)*.*

The optimization problem in (4.41) is hard to solve due to the combinatorial character of identifying the support of the sparse random attack vector. For that reason, we first tackle the case in which the attacker only comprises one measurement, that is, $k = 1$ in Section 4.4.1 and propose two different greedy constructions that sequentially update the set $\text{supp}(A^m)$, with A^m in (4.6), and determines the corresponding entry in the diagonal of the matrix $\mathbf{\Sigma}_{AA}$ in (4.22) in Section 4.4.2 and Section 4.4.3, respectively.

4.4.1 Optimal Single Measurement Attack Construction

Despite having narrowed it down to Gaussian distributions, the optimization problem in (4.41) is still challenging due to its combinatorial character. For that reason, this section tackles the case in which the attacker only comprises one measurement, that is, $k = 1$. The rationale for this is that it is expected to leverage the insight developed for the single sensor case in the construction of the general k-sparse case. The following theorem provides the optimal solution for the case in which the attacker corrupts a single measurement.

Theorem 13. *The solution to the sparse stealth attack construction problem in* (4.41) *for the case* $k = 1$ *is given by*

$$\bar{\mathbf{\Sigma}}_{AA} = v\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}, \tag{4.42}$$

where

$$i = \underset{j\in\{1,2,\ldots,m\}}{\arg\min} \left\{ \left(\mathbf{\Sigma}_{YY}^{-1}\right)_{jj} \right\}, \tag{4.43a}$$

$$v = -\frac{\sigma^2}{2} + \frac{1}{2}\left(\sigma^4 - \frac{4(\underline{w}\sigma^2 - 1)}{\lambda \underline{w}^2}\right)^{\frac{1}{2}}, \tag{4.43b}$$

with $\underline{w} \triangleq (\mathbf{\Sigma}_{YY}^{-1})_{ii}$.

Proof. The proof of Theorem 13 is presented in Appendix C. □

Remark 4. *Knowledge of the second order moments of the measurements* $\mathbf{\Sigma}_{YY}$ *and the variance of the AWGN introduced by the sensors* σ^2 *suffice to construct the optimal single measurement attack.*

4.4.2 Greedy Constructions with Jacobian Updated

The extension to the k-sparse case of the solution proposed in Section 4.4.1 does not get around the combinatorial optimization in (4.41). For that reason, in the following a greedy construction that leverages the insight distilled in the $k = 1$ case to select the set of k attacked measurements is proposed. The construction is based on a classical greedy procedure that sequentially selects a measurement to attack by minimizing the cost in terms of the decision at each step.

Let us denote by $\mathcal{K}$ the set of measurement indices that are attacked, that is, $\mathcal{K} \triangleq \text{supp}(A^m)$. The greedy algorithm operates by sequentially updating the elements in $\mathcal{K}$ by adding a new index in each step until k indices are selected. For that reason, the resulting

entries of the attack vector are independent, and therefore, the covariance matrix of the attack vector obtained via the proposed greedy approach is in the set in (4.40). The proposed greedy construction is described in Algorithm 1. Let $\mathbf{H}_j$ be the Jacobian matrix at iteration j, with $j \in \{1, 2, \ldots, k\}$. The algorithm updates the Jacobian matrix at iteration j by removing all the rows that correspond to the compromised measurements in $\mathcal{K}_{j-1}$ from the original Jacobian matrix $\mathbf{H}$ in (4.1). In the other words, the Jacobian matrix $\mathbf{H}_j$ is the observation matrix formed by the measurements in $\mathcal{K}_{j-1}^c$ that have not been compromised. The resulting Jacobian matrix is denoted as $\mathbf{H}_{\mathcal{K}_{j-1}^c} \in \mathbb{R}^{\text{card}\left(\mathcal{K}_{j-1}^c\right)\times n}$ in step 3 in the Algorithm. Therefore, the complexity of Algorithm 1 is $m!$.

Algorithm 1 k-sparse stealth attack construction with Jacobian updated

Input: the Jacobian matrix $\mathbf{H} \in \mathbb{R}^{m\times n}$ in (4.1),
the variance of the noise $\sigma^2 \in \mathbb{R}_+$ in (4.2),
the covariance matrix $\mathbf{\Sigma}_{XX} \in \mathcal{S}_+^n$ in (4.3),
the weighting parameter $\lambda \geq 1$ in (4.34);
and the sparse constraint $k \leq m$.

Output: the covariance matrix of the attack vector $\bar{\mathbf{\Sigma}}_{AA}$ in (4.22), and the set of indices of attacked measurements $\mathcal{K}$.

1: Set $\mathcal{K}_0 = \{\emptyset\}$
2: **for** $j = 1$ to k **do**
3: Set $\mathbf{H}_j = \mathbf{H}_{\mathcal{K}_{j-1}^c}$
4: Compute $\mathbf{W}_j = \left(\mathbf{H}_j \mathbf{\Sigma}_{XX} \mathbf{H}_j^\mathsf{T} + \sigma^2 \mathbf{I}_{\text{card}\left(\mathcal{K}_{j-1}^c\right)}\right)^{-1}$
5: Set $\alpha_j = \arg\min_i \{(\mathbf{W}_j)_{ii}\}$,
6: Set $\underline{w}_j \triangleq (\mathbf{W}_j)_{\alpha_j\alpha_j}$
7: Set $v_{\alpha_j} = -\frac{\sigma^2}{2} + \frac{1}{2}\left(\sigma^4 - \frac{4(\underline{w}_j\sigma^2-1)}{\lambda \underline{w}_j^2}\right)^{\frac{1}{2}}$
8: Set $\mathcal{K}_j = \mathcal{K}_{j-1} \cup \{\alpha_j\}$
9: **end for**
10: Set $\mathcal{K} = \mathcal{K}_k$
11: Set $\bar{\mathbf{\Sigma}}_{AA} = \sum_{i\in\mathcal{K}} v_i \mathbf{e}_i \mathbf{e}_i^\mathsf{T}$

Remark 5. *The k-sparse stealth attack construction in Algorithm 1 requires the knowledge of the second order moments of the state variables $\mathbf{\Sigma}_{XX}$, the Jacobian matrix $\mathbf{H}$ and the variance of the AWGN introduced by the sensors σ^2.*

4.4.3 Greedy Constructions with Optimal Single-Step Sequential Procedure

The proposed construction hinges on the idea that approaching the sensor selection problem in a sequential fashion resembles the single sensor selection problem discussed in Section 4.4.1. This enables us to leverage the single sensor selection construction to analytically characterize the cost difference induced by the addition of a new element to the set $\text{supp}(A^m)$, with A^m in (4.6), and determines the corresponding entry in the diagonal of the matrix $\mathbf{\Sigma}_{AA}$ in (4.22).

More specifically, given the sparsity constraint in (4.37), for some $k \leqslant m$, the construction can be divided into k epochs. At each epoch a new element is added to $\text{supp}(A^m)$. At epoch i, let $\mathbf{\Sigma}_i \in S_+^m$ be the covariance matrix of the attack vector. Let the set $\mathcal{K}_i$ be the set of indices corresponding to the entries of the vector $\text{diag}(\mathbf{\Sigma}_i)$ that are nonzero, that is,

$$\mathcal{K}_i = \{j \in \{1, 2, \ldots, m\} : (\mathbf{\Sigma}_i)_{jj} > 0\}. \tag{4.44}$$

For all $i \in \{1, 2, \ldots, k\}$, it is imposed that $\mathcal{K}_i \subseteq \{1, 2, \ldots, m\}$ and $\text{card}\,(\mathcal{K}_i) = i$. This implies that $\mathcal{K}_1 \subset \mathcal{K}_2 \subset \ldots \subset \mathcal{K}_k \subset \{1, 2, \ldots, m\}$. Hence, the following holds

$$\mathbf{\Sigma}_i = \mathbf{\Sigma}_{i-1} + v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}, \tag{4.45}$$

where, for $i = 1$, $\mathbf{\Sigma}_0$ is a matrix of zeros, the integer $j \in \{1, 2, \ldots, m\} \setminus \mathcal{K}_{i-1}$ is the index of the new entry at epoch i, and $v \in \mathbb{R}_+$ is the value of such entry that corresponds to the variance of the attack variable added to A^m. For ease of presentation, this section denotes the set of indices available to the attacker to choose at epoch i, that is, the entries of the vector $\text{diag}(\mathbf{\Sigma}_{i-1})$ that are zero, as

$$\mathcal{K}_{i-1}^{\mathsf{c}} \triangleq \{1, 2, \ldots, m\} \setminus \mathcal{K}_{i-1}. \tag{4.46}$$

Our proposition to choose both $j \in \mathcal{K}_{i-1}^{\mathsf{c}}$ and $v \in \mathbb{R}_+$ at epoch i as described in (4.45) is based on the following optimization problem

$$\min_{(j,v)\in\mathcal{K}_{i-1}^{\mathsf{c}}\times\mathbb{R}_+} J(\mathbf{\Sigma}_{i-1} + v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}). \tag{4.47}$$

The following lemma sheds light on the solution to the problem (4.47).

Lemma 14. *Let $\mathbf{\Sigma}_i \in S_+^m$ and $\mathbf{\Sigma}_{i-1} \in S_+^m$ be two matrices in epoch i and epoch $i-1$ that satisfy $\mathbf{\Sigma}_i = \mathbf{\Sigma}_{i-1} + v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}$ with $\mathcal{K}_{i-1}$ in (4.44), $j \in \mathcal{K}_{i-1}^c$ and $v \in \mathbb{R}_+$. The cost function J in (4.35) satisfies that*

$$J(\mathbf{\Sigma}_i) = J(\mathbf{\Sigma}_{i-1}) + f(\mathbf{\Sigma}_{i-1}, v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}), \tag{4.48}$$

where the function $f : \mathbb{R}^{m\times m} \times \mathbb{R}^{m\times m} \to \mathbb{R}$ is such that

$$\begin{aligned} f(\mathbf{\Sigma}_{i-1}, v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}) \triangleq &(1-\lambda)\log\left|\mathbf{I}_m + (\mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_{i-1})^{-1} v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right| \\ &- \log\left|\mathbf{I}_m + \left(\sigma^2\mathbf{I}_m + \mathbf{\Sigma}_{i-1}\right)^{-1} v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right| + \lambda\text{tr}\left(\mathbf{\Sigma}_{YY}^{-1} v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right), \end{aligned} \tag{4.49}$$

where $\lambda \geq 1$ is introduced in (4.34), the matrix $\mathbf{\Sigma}_{YY}$ is in (4.5).

Proof. The proof of Lemma 14 is presented in Appendix D. □

The relevance of Lemma 14 is that it enables the selection of both $j \in \mathcal{K}_{i-1}^{\mathsf{c}}$ and $v \in \mathbb{R}_+$ at epoch i based on a simpler optimization problem than that in (4.47). Indeed, the selection problem results in

$$\min_{(j,v)\in\mathcal{K}_{i-1}^{\mathsf{c}}\times\mathbb{R}_+} f(\mathbf{\Sigma}_{i-1}, v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}), \tag{4.50}$$

where the function f is defined in (4.49).

The following lemma characterizes the convexity of the cost function in (4.50) with respect to v.

Proposition 7. *Let $\lambda \geq 1$. Then the optimization problem in* (4.50) *is convex with respect to v.*

Proof. The proof of Proposition 7 is presented in Appendix E. □

The following theorem provides the solution to the optimization problem in (4.50).

Theorem 15. *Let k satisfy $0 < k \leqslant m$, and for all $i \in \{1, 2, \ldots, k\}$, denote by $(j^\star, v^\star) \in \mathcal{K}_{i-1}^{\mathsf{c}} \times \mathbb{R}_+$ the solution to the optimization problem in* (4.47)*. Then, the following holds*

$$j^\star = \underset{j \in \mathcal{K}_{i-1}^{\mathsf{c}}}{\arg\min}\, J(\mathbf{\Sigma}_{i-1} + v_j \mathbf{e}_j \mathbf{e}_j^{\mathsf{T}}) \quad \textit{and} \tag{4.51}$$

$$v^\star = v_{j^\star}, \tag{4.52}$$

where, for all $j \in \mathcal{K}_{i-1}^{\mathsf{c}}$,

$$v_{j^*} = \left(\frac{\beta_j - \alpha_j + \beta_j \alpha_j \sigma^2}{2\beta_j \alpha_j} \right) \left(\sqrt{1 - \frac{4\beta_j \alpha_j \left(\beta_j \sigma^2 - \alpha_j \sigma^2 - \dfrac{\alpha_j \sigma^2 + 1}{\lambda} \right)}{\left(\beta_j - \alpha_j + \beta_j \alpha_j \sigma^2 \right)^2}} - 1 \right), \tag{4.53}$$

with

$$\alpha_j \stackrel{\Delta}{=} \mathrm{tr}\left(\left(\mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_{i-1} \right)^{-1} \mathbf{e}_{j^*} \mathbf{e}_{j^*}^{\mathsf{T}} \right), \tag{4.54}$$

$$\beta_j \stackrel{\Delta}{=} \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1} \mathbf{e}_{j^*} \mathbf{e}_{j^*}^{\mathsf{T}} \right), \tag{4.55}$$

and the real $\sigma > 0$ in (4.53) *is introduced in* (4.2)*.*

Proof. The proof of Theorem 15 is presented in Appendix F. □

The proposed greedy construction with optimal sequential procedure is described in Algorithm 2.

Remark 6. *The k-sparse stealth attack construction in Algorithm 2 requires the knowledge of the second order moments of the measurements $\mathbf{\Sigma}_{YY}$ and the variance of the AWGN introduced by the sensors σ^2. This implies that the second order statistics of the measurements and variance of the noise introduced by the sensors suffice to construct the attacks.*

4.5 Numerical Results

This section numerically evaluates the performance of the proposed attack construction algorithms on a direct current (DC) state estimation setting for the IEEE 9-bus, IEEE 14-bus and IEEE 30-bus test systems [112]. The voltage magnitudes are set to 1.0 per unit, which implies that the state estimation is based on the measurements of active power flow injections to all the buses and the active power flow between physically connected buses as described in Section 2.1.2. The Jacobian matrix $\mathbf{H}$ is determined by the reactance of the branches and the topology of the corresponding systems. The MATPOWER [113] is

Algorithm 2 k-sparse independent attack construction

Input: the variance of the noise $\sigma^2 \in \mathbb{R}_+$ in (4.2),
the covariance matrix $\mathbf{\Sigma}_{YY} \in \mathcal{S}_+^m$ in (4.5),
the weighting parameter $\lambda \geq 1$ in (4.34);
and the sparse constraint $k \leq m$ in (4.44).

Output: the covariance matrix of the attack vector $\bar{\mathbf{\Sigma}}_{AA}$ in (4.22),
and the set of indices of attacked measurements $\mathcal{K}$.

Set $\mathcal{K}_0 = \{\emptyset\}$
Set $\mathbf{\Sigma}_0 = \mathbf{0}$
for $j = 1$ to k **do**
  for $\ell \in \mathcal{K}_{i-1}^{\mathsf{c}}$ **do**
    Compute v_ℓ in (4.53)
  end for
  Compute $j^\star$ in (4.51)
  Compute $v^\star$ in (4.52)
  Set $\mathcal{K}_j = \mathcal{K}_{j-1} \cup \{j^\star\}$
  Set $\mathbf{\Sigma}_j = \sum_{i \in \mathcal{K}_j} v_i \mathbf{e}_i \mathbf{e}_i^{\mathsf{T}}$
end for
$\bar{\mathbf{\Sigma}}_{AA} = \sum_{i \in \mathcal{K}_k} v_i \mathbf{e}_i \mathbf{e}_i^{\mathsf{T}}$

adopted to generate $\mathbf{H}$ for each test system. The attack constructions in Theorem 13 and Theorem 15 show that the variance of the random attacks is a function of the covariance matrix of the measurements. To obtain the covariance matrix of measurements, this section first captures the statistical dependence between the state variables by a Toeplitz model for the covariance matrix $\mathbf{\Sigma}_{XX} \in \mathcal{S}_+^n$ that arises in a wide range of practical settings, such as autoregressive stationary processes [13, 17, 114]. Specifically, the correlation between state variables X_i and X_j is modelled with the exponential decay parameter $\rho \in \mathbb{R}_+$ that defines the entries of the covariance matrix of the state variables as $(\mathbf{\Sigma}_{XX})_{ij} = \rho^{\text{abs}(i-j)}$, with $(i, j) \in \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}$. That is

$$\mathbf{\Sigma}_{XX} = \begin{pmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{n-2} & \rho^{n-1} \\ \rho & 1 & \rho & \ldots & \rho^{n-3} & \rho^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \ldots & 1 & \rho \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \ldots & \rho & 1 \end{pmatrix}. \tag{4.56}$$

From (4.56), the parameter ρ characterizes the correlation strength between state variables X_i and X_j, with $(i, j) \in \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}$, that is, the correlation strength between the phase angle of the voltage in the buses. In this setting, the performance of the proposed sparse stealth attack is not only a function of the attack constructions but also the correlation parameter ρ, the noise variance σ^2, and the topology of the system described by $\mathbf{H}$. In the simulations, the observation model noise regime is set to be the signal to noise ratio (SNR) defined as

$$\text{SNR} \triangleq 10 \log_{10} \left(\frac{\text{tr}(\mathbf{H} \mathbf{\Sigma}_{\text{XX}} \mathbf{H}^{\mathsf{T}})}{m \sigma^2} \right). \tag{4.57}$$

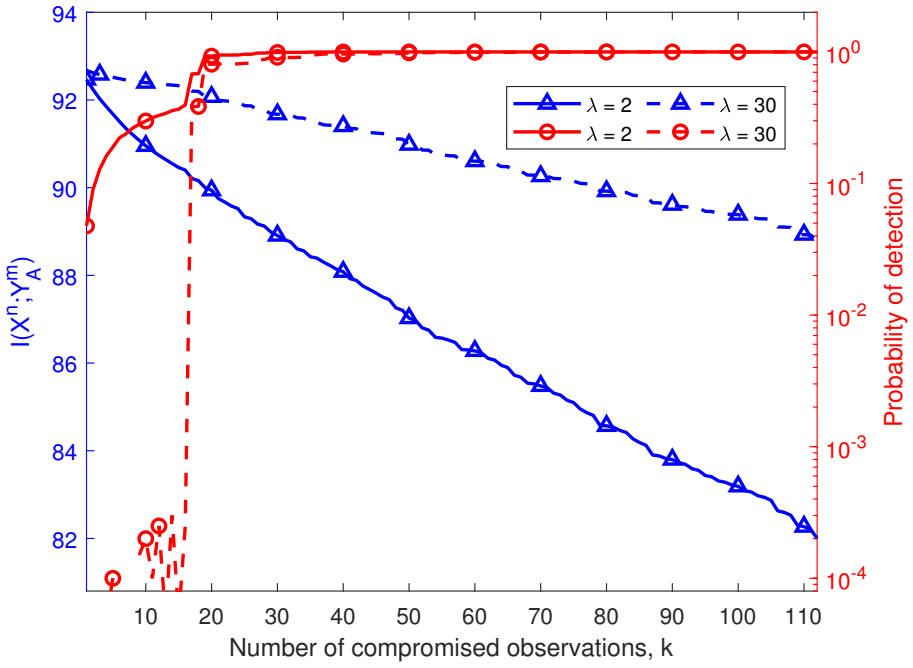

Figure 4.1: Performance of the sparse attack in terms of mutual information, probability of detection for different values of λ when SNR $= 30$ dB, $\rho = 0.1, \tau = 2$ on the IEEE 30-bus test system.

The results in this section are obtained by averaging 2×10^4 realizations of the measurements as described in (4.6).

4.5.1 Performance of Attack Construction with Jacobian Updated

The probability of detection in this section is obtained by LRT in (4.14), that is,

$$L(\bar{\mathbf{y}}) = \frac{f_{Y_A^m}(\bar{\mathbf{y}})}{f_{Y^m}(\bar{\mathbf{y}})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau, \tag{4.58}$$

where the functions $f_{Y_A^m}$ and f_{Y^m} are the pdf of $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{Y_A Y_A})$ obtained from Algorithm 1 and the pdf of $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{YY})$ in (4.4), respectively, and τ is the decision threshold set by the system operator to meet the false alarm constraint. The results are obtained by averaging 2×10^4 realizations of the measurements as described in (4.6).

Fig. 4.1 depicts the mutual information and the probability of detection that the attack constructed by Algorithm 1 induces for different values of the number of compromised measurements and the weighting parameter λ. As expected, the mutual information decreases monotonically, approximately linearly with the number of compromised measurements, while the probability of detection increases. Interestingly, the probability of detection exhibits an abrupt increase that suggests a *threshold effect* when a critical number of compromised measurements is reached. The weighting parameter λ governs the minimum achievable probability of detection, e.g. a probability of detection of 10^{-2} is not attainable when $\lambda = 2$. Indeed,

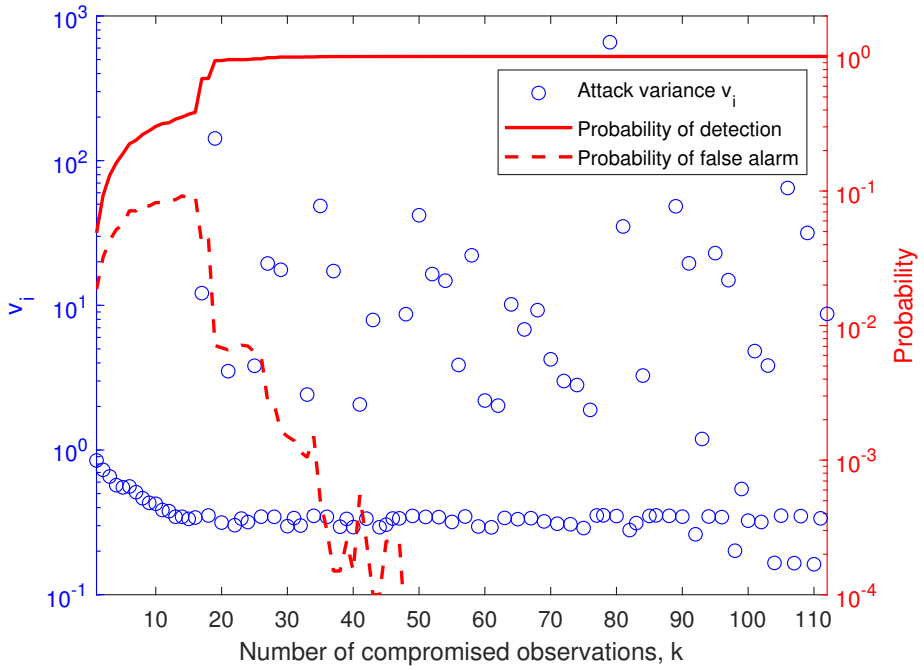

Figure 4.2: Variance of the attack vector entries, probability of detection, and probability of false alarm of the sparse attack when $\lambda = 2$, SNR $= 30$ dB, $\rho = 0.1, \tau = 2$ on the IEEE 30-bus test system.

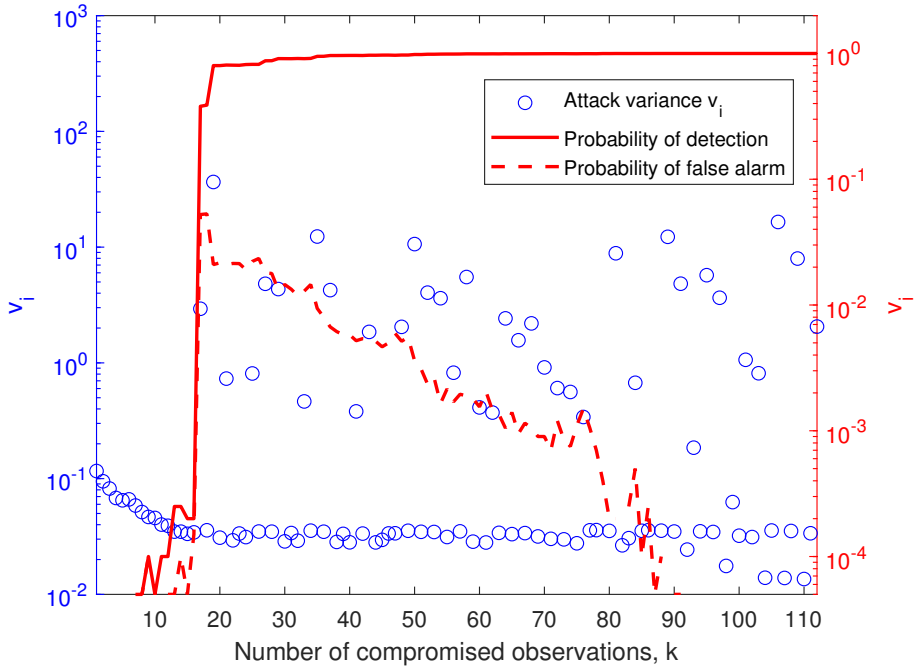

Figure 4.3: Variance of the attack vector entries, probability of detection, and probability of false alarm of the sparse attack when $\lambda = 30$, SNR $= 30$ dB, $\rho = 0.1, \tau = 2$ on the IEEE 30-bus test system.

increasing the value of λ to 30 yields a smaller probability of detection for small values of k but the threshold effect takes place for the same number of compromised measurements, for both values of λ. This suggests that the topology of the system governs the position of the threshold.

The variance of the random attack variables used to compromise each sensor, the probability of detection, and the probability of false alarm as a function of the number of compromised measurements are illustrated in Fig. 4.2 and Fig. 4.3 for $\lambda = 2$ and $\lambda = 30$, respectively. As shown in Theorem 13, λ is a scaling factor on the variances of the attack vector, and therefore, the values of the variance for the case $\lambda = 2$ are simply scaled in the case $\lambda = 30$. There are two distinguishable attack regimes depending on the variance of the attack vector entries. Algorithm 1 does not yield a monotonically decreasing profile of variances. Instead the variance of the entries selected by the algorithm switches between small and large values as the number of compromised measurements increases. This suggests, that certain measurement entries are significantly more sensitive to additive attacks than others and the existence of more vulnerable sensors that are determined by the topology of the system, as shown in (4.43b). For both cases, the probability of false alarm exhibits a non-monotonic behaviour with the number of compromised measurements, and interestingly, the change in monotonicity coincides with the threshold.

4.5.2 Performance of Attack Construction with Optimal Sequential Procedure

The simulation of the linearized AC power flow model is carried out to verify the performance of the proposed attacks. Let $\mathbf{x}_0$ be the state variables of the nominal operation point when the system is operating under optimal power flow. The MATPOWER [113] is adopted to obtain the optimal power flow where the nominal operation point lies on. The corresponding Jacobian matrix is

$$\mathbf{H}_0 = \frac{\partial}{\partial \mathbf{x}} \mathbf{h}(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_0},$$

where $\mathbf{h}(\mathbf{x}) \in \mathbb{R}^m$ denotes the vector of random variables induced by the nonlinear relation between the state variables and the measurements and $\mathbf{H}_0$ is the corresponding Jacobian matrix in linearized AC model when system is operating under optimal power flow with the vector of state variables $\mathbf{x}_0$.

Performance in terms of information theoretic cost

Let $\mathbf{\Sigma}_k^{\text{ind}}$ be the output of the k-sparse attack construction of Algorithm 2. This section evaluates the attack performance in terms of the sparsity penalty defined as

$$\eta \triangleq \frac{J(\mathbf{\Sigma}_k^{\text{ind}}) - J(\mathbf{\Sigma}_m^{\text{ind}})}{J(\mathbf{\Sigma}_m^{\text{ind}})}, \tag{4.59}$$

where J is the cost defined in (4.35). Note that $J(\mathbf{\Sigma}_m^{\text{ind}})$ denotes the cost induced by the construction when all the sensors are attacked. In that sense, this metric captures the performance loss of the attack when only k measurements are attacked.

Fig. 4.4 depicts the performance of the independent sparse stealth attack constructions obtained with Algorithm 2 in DC model on different IEEE test systems as a function of the proportion of compromised sensors, that is k/m, for correlation parameter $\rho = 0.9$ and $\lambda = 8$. As expected, the sparsity penalty decreases monotonically with the proportion of compromised sensors. In the independent sparse attack case, the sparsity penalty does not change significantly in terms of the proportion of compromised sensors. Note that the exponential decrease slope is approximately constant, which indicates that the advantage of adding more sensors to the attack construction decreases exponentially at an approximately constant rate. Remarkably, this exponential decrease is observed for all system sizes and SNR regimes. Interestingly, the size of the network does not determine the performance the attack. Specifically, the IEEE 14-bus system is the most vulnerable to attacks. This suggests that the topology of the system fundamentally changes the performance of the attack.

Fig. 4.5 depicts the performance of the independent sparse stealth attack construction obtained with Algorithm 2 in linearized AC model on different IEEE test systems with the same setting as in Fig. 4.4. As expected, the sparsity penalty decreases monotonically with the proportion of compromised measurements.

Performance in terms of the tradeoff between mutual information and KL divergence

Fig. 4.6 and Fig. 4.7 depict the multiobjective performance of the Algorithm 2 attack construction in DC model in terms of the tradeoff between mutual information and KL divergence for different values of the proportion of compromised sensors when SNR = 30 dB and $\rho = 0.9$ on the IEEE 9-bus and the IEEE 14-bus systems, respectively. As expected, larger values of the parameter λ yield smaller values of KL divergence, that is, the probability of detection is prioritized in the construction over the mutual information decrease for all the scenarios. Moreover, smaller values of k yield smaller reductions of the mutual information, which indicates that remaining stealthy in a sparse setting necessarily implies reducing the amount of disruption of the attack. On the other hand, larger values of k enable the attacker to more effectively tradeoff disruption for stealth.

Fig. 4.8 and Fig. 4.9 depict the multiobjective performance of the Algorithm 2 attack constructions in linearized AC model in terms of the tradeoff between mutual information and KL divergence for different values of the proportion of compromised measurements when SNR = 30 dB and $\rho = 0.9$ on the IEEE 9-bus and the IEEE 14-bus systems, respectively.

Performance in terms of state estimate degradation and probability of detection

Fig. 4.10 depicts the deviation of the LS estimate caused by one realization of the independent attack constructions via Algorithm 2 on the IEEE 9-bus test system with different values of k when $\lambda = 2$, SNR = 30 dB and $\rho = 0.9$. Note that the magnitude of the state estimate is in per unit where the base quantity is the actual value of state variables accordingly. The LS state estimate denoted by black dots serves as a benchmark for the state estimate after attacks. The LS estimate for all the state variables under independent attacks derivates from the LS estimate without attacks in both small and large values of k. The attack constructions successfully deviate the LS estimates for all state variables, albeit with different deviation

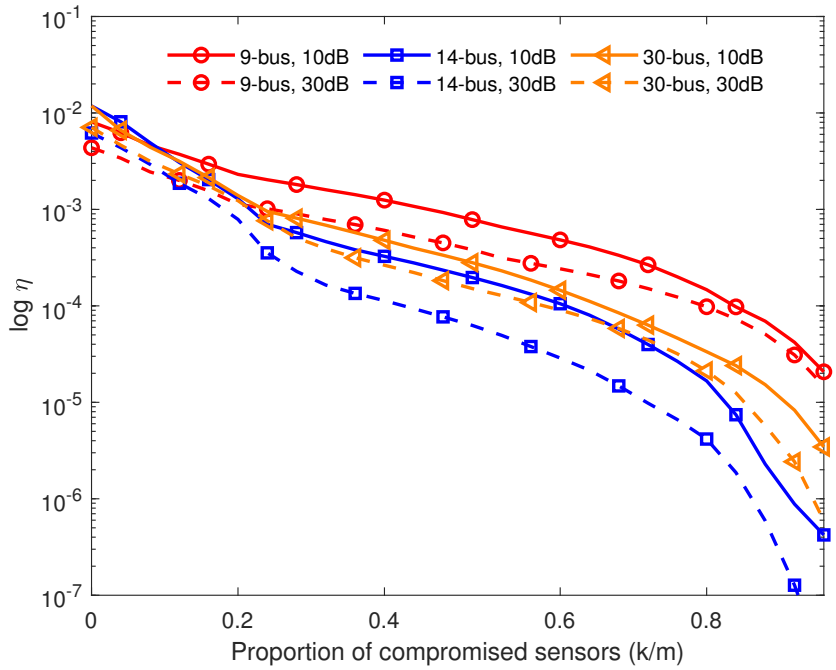

Figure 4.4: Performance of independent attack constructions in DC model on different IEEE test systems with $\rho = 0.9$ and $\lambda = 8$.

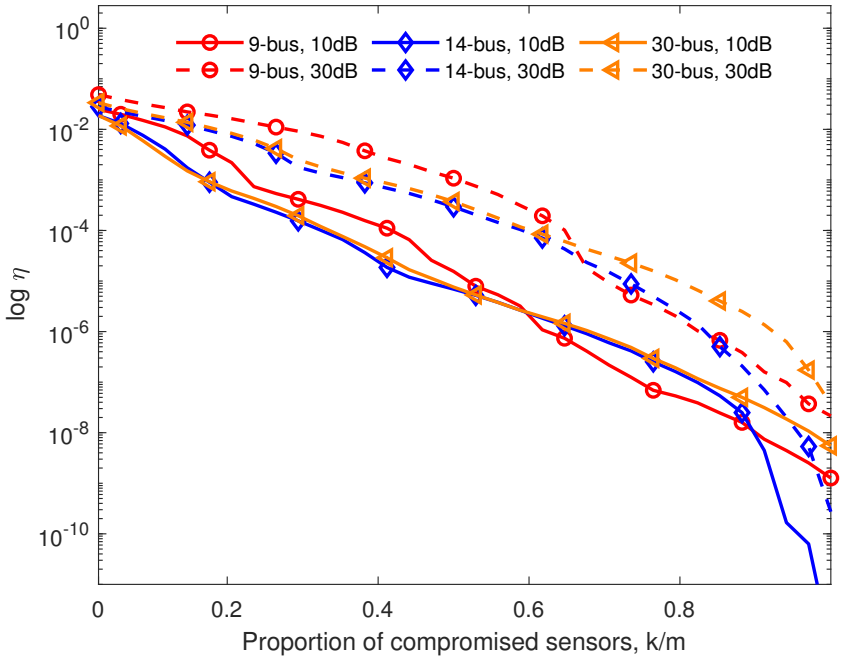

Figure 4.5: Performance of independent attack constructions in linearized AC model on different IEEE test systems with $\rho = 0.9$ and $\lambda = 8$.

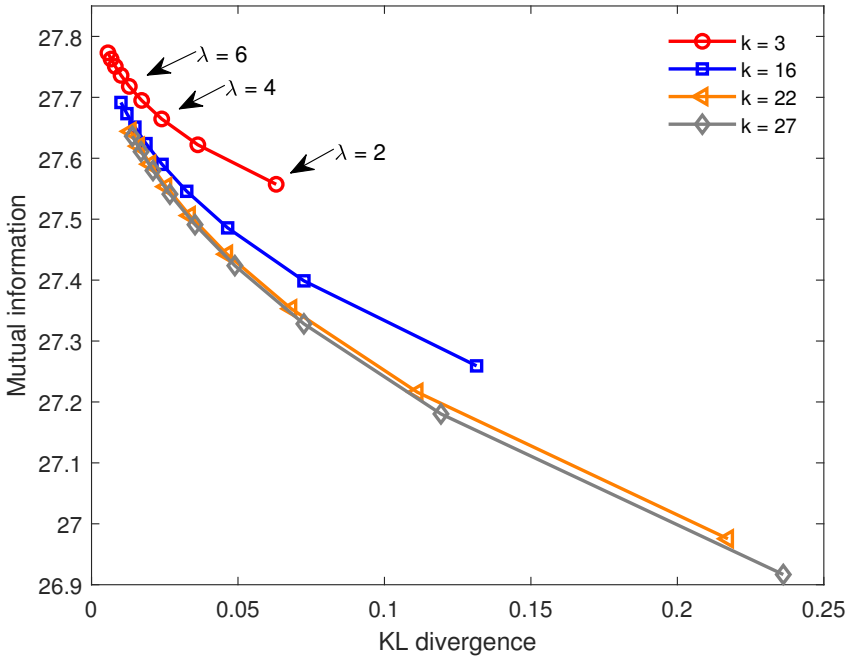

Figure 4.6: Performance of independent sparse attack construction in DC model in terms of mutual information and KL divergence for different values of λ on the IEEE 9-bus test system with SNR = 30 dB and $\rho = 0.9$.

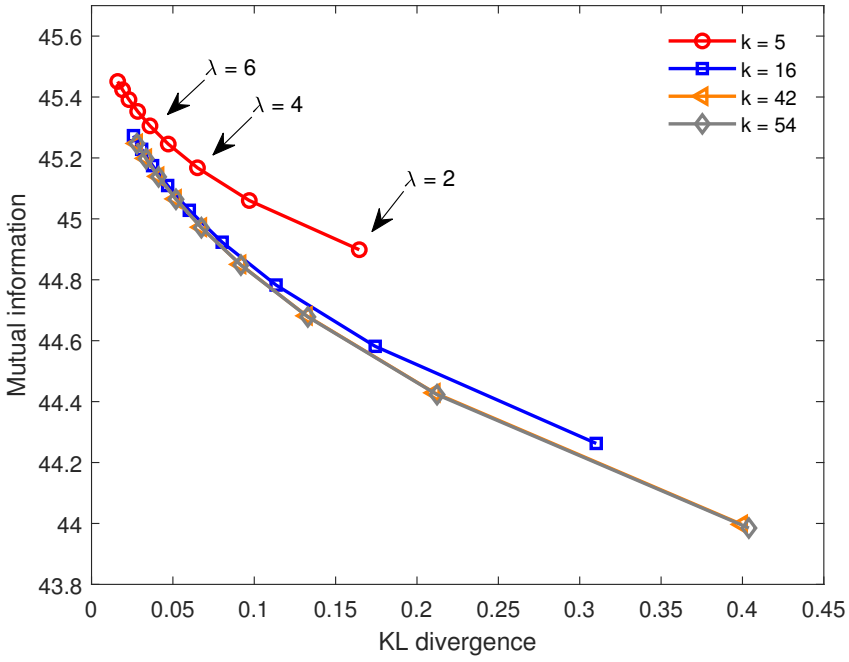

Figure 4.7: Performance of independent sparse attack construction in DC model in terms of mutual information and KL divergence for different values of λ on the IEEE 14-bus test system with SNR = 30 dB and $\rho = 0.9$.

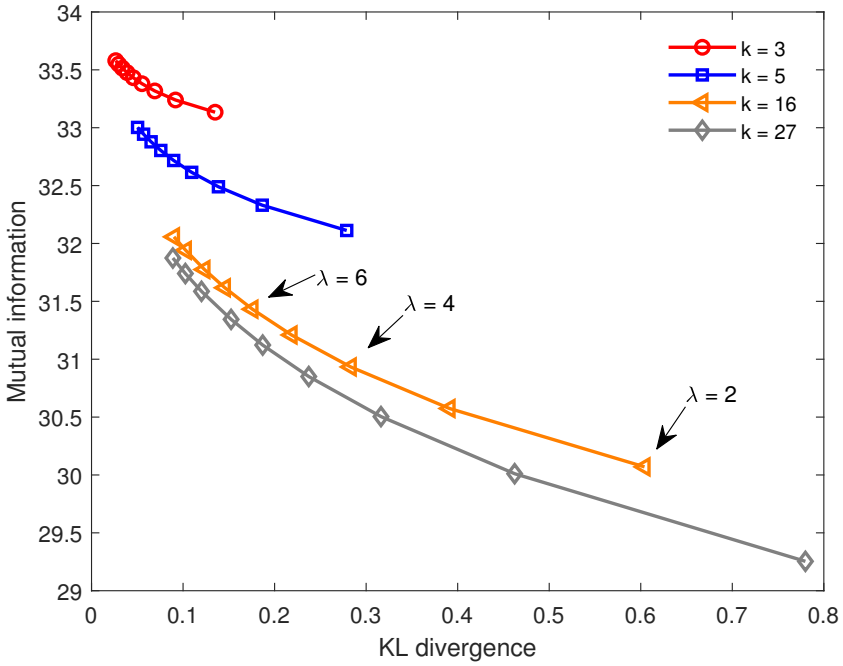

Figure 4.8: Performance of independent sparse attack construction in linearized AC model in terms of mutual information and KL divergence for different values of λ on the IEEE 9-bus test system with SNR = 30 dB and $\rho = 0.9$.

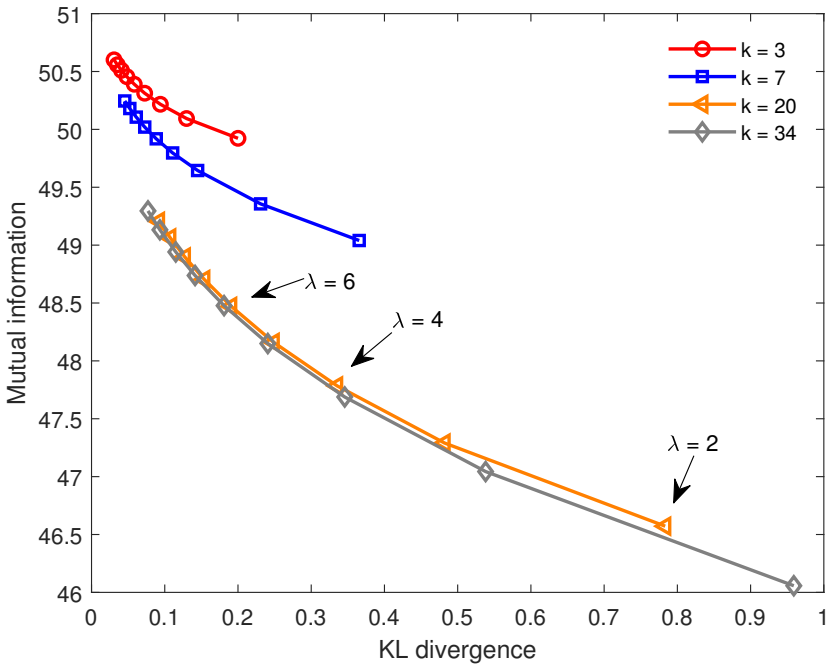

Figure 4.9: Performance of independent sparse attack construction in linearized AC model in terms of mutual information and KL divergence for different values of λ on the IEEE 14-bus test system with SNR = 30 dB and $\rho = 0.9$.

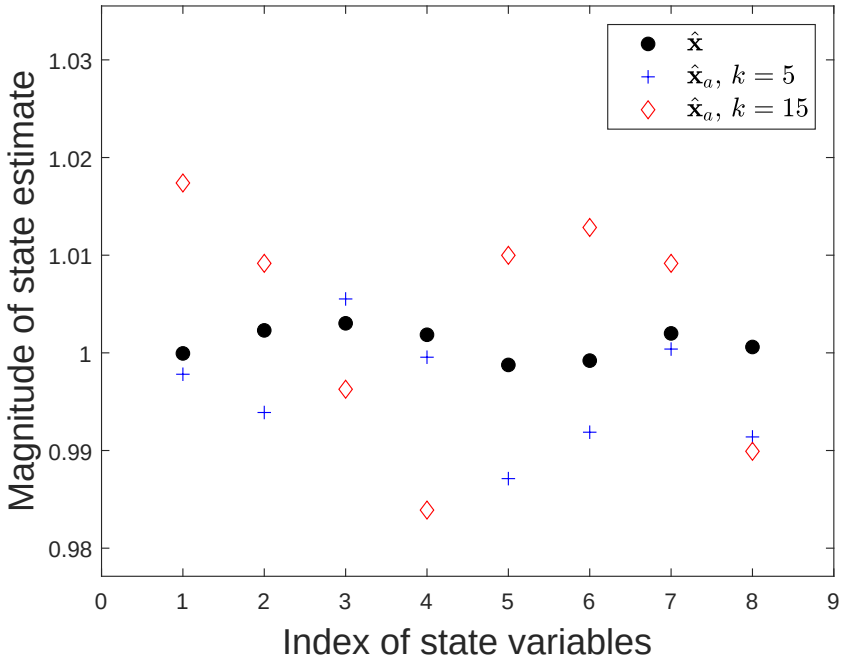

Figure 4.10: Performance of independent attack construction in terms of state estimate with and without attacks on the IEEE 9-bus test system with one realization when $\lambda = 2$, $\rho = 0.9$, SNR $= 30$ dB.

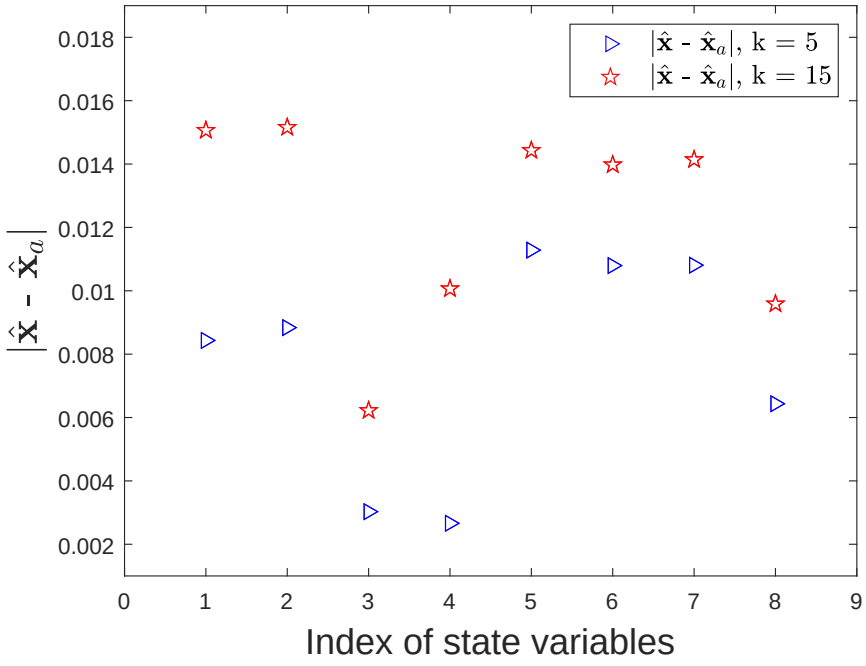

Figure 4.11: Performance of independent attack construction in terms of the average absolute deviation of the state estimate on the IEEE 9-bus test system with 20000 realizations when $\lambda = 2$, $\rho = 0.9$, SNR $= 30$ dB.

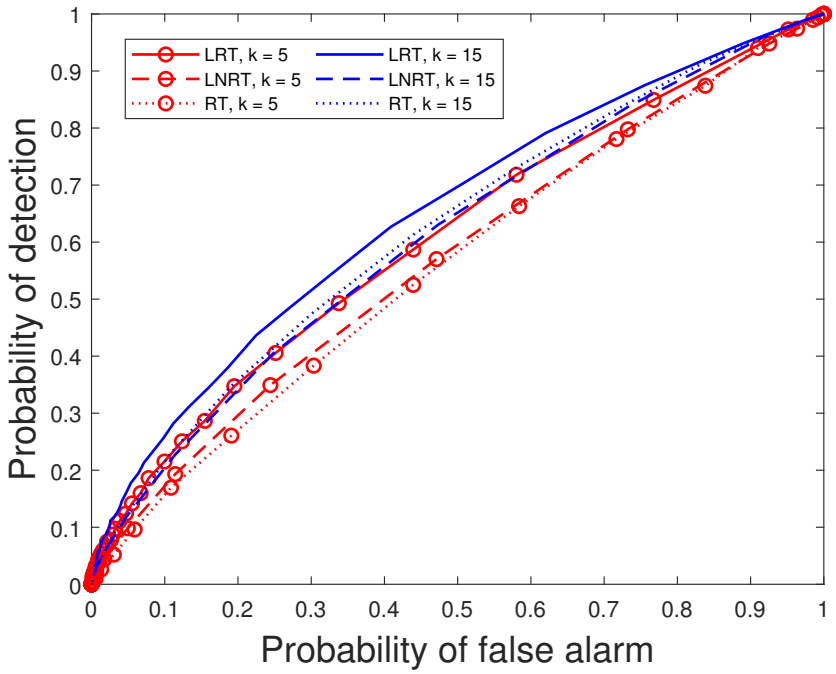

Figure 4.12: Performance of independent attack construction in terms of probability of detection and probability of false alarm in LRT, LNRT and RT on the IEEE 9-bus test system when $\lambda = 2$, $\rho = 0.9$, SNR $= 30$ dB.

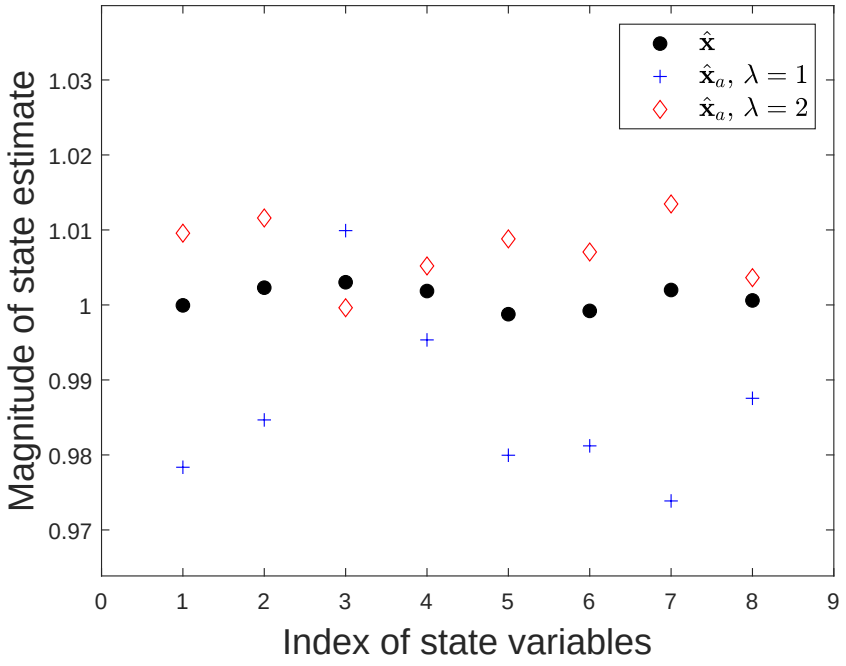

Figure 4.13: Performance of independent attack construction in terms of state estimate with and without attacks on the IEEE 9-bus test system with one realization when $k = 15$, $\rho = 0.9$, SNR $= 30$ dB.

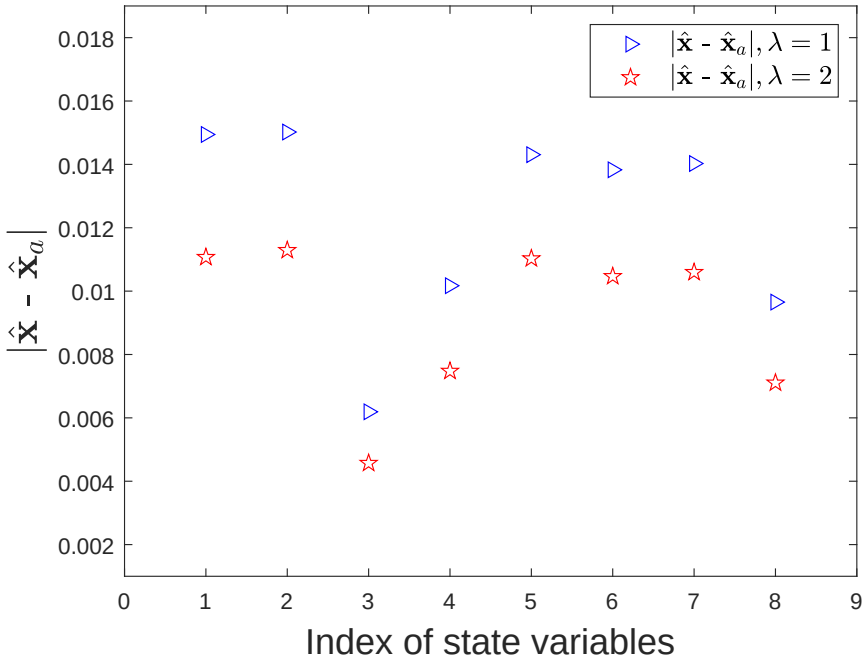

Figure 4.14: Performance of independent attack construction in terms of the average absolute deviation of the state estimate on the IEEE 9-bus test system with 20000 realizations when $k = 15$, $\rho = 0.9$, SNR $= 30$ dB.

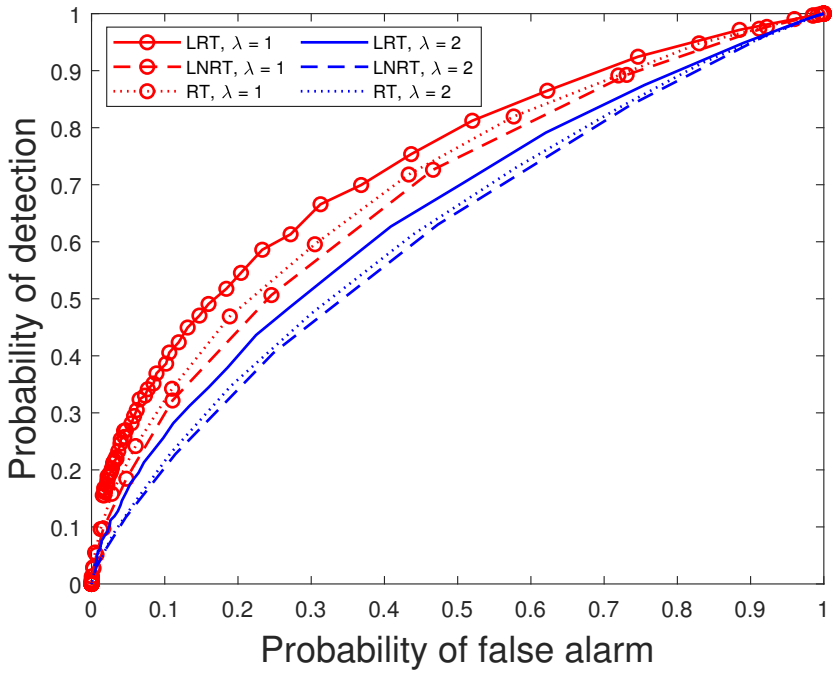

Figure 4.15: Performance of independent attack construction in terms of probability of detection and probability of false alarm in LRT, LNRT and RT on the IEEE 9-bus test system when $k = 15$, $\rho = 0.9$, SNR $= 30$ dB.

strengths. In the worst case, for example the first state variable, the deviation is around 0.02 per unit. Fig. 4.11 depicts the absolute value of the deviation of the LS estimate caused by averaging 2×10^4 realizations of the attack constructions via Algorithm 2 on the IEEE 9-bus test system with different values of k when $\lambda = 2$, SNR = 30 dB and $\rho = 0.9$. As expected, the attack constructions with larger k yield larger deviation of estimate in all the state variables. Surprisingly, state variables $i, i \in \{1, 2, 5, 6, 7\}$ deviates more than state variable $j, j \in \{3, 4, 8\}$ both when $k = 5$ and $k = 15$. This indicates that some state variables are more sensitive to the random attacks.

The tradeoff between probability of detection and probability of false alarm for the attack construction via Algorithm 2 on the IEEE 9-bus test system with LRT, largest normalized residual test (LNRT) and residual test (RT) is depicted in Fig. 4.12 for different k when $\lambda = 2$, $\rho = 0.9$, SNR = 30 dB. As expected, the LRT outperforms other detection method and smaller value of k yields smaller probability of detection in every detection method.

Fig. 4.13 depicts the deviation of the LS estimate caused by one realization of the independent attack constructions via Algorithm 2 on the IEEE 9-bus test system with different values of λ when $k = 15$, SNR = 30 dB and $\rho = 0.9$. The LS estimate under independent attacks for all the state variables deviates from the LS estimate without attacks in both small and large values of λ. The attack constructions successfully deviate the LS estimates for all state variables, albeit with different deviation strengths. Fig. 4.14 depicts the absolute value of the deviation of the LS estimate caused by averaging 2×10^4 realizations of the attack construction via Algorithm 2 on the IEEE 9-bus test system with different values of λ when $k = 15$, SNR = 30 dB and $\rho = 0.9$. As expected, the attack construction with smaller λ yields larger deviation of estimate in all the state variables at the cost of stealthiness. Fig. 4.15 depicts the tradeoff between probability of detection and probability of false alarm for the attack construction via Algorithm 2 on the IEEE 9-bus test system with LRT, LNRT and RT for different λ when $k = 15$, $\rho = 0.9$, SNR = 30 dB. As expected that smaller values of λ yields larger probability of detection in all detection methods.

4.6 Summary

This chapter proposed a novel independent stealth attack construction with sparsity constraints. The proposed attack constructions minimize the mutual information between the state variables and the measurements obtained by the operator while minimizing the probability of detection. To that end, this chapter proposed a cost function that combines the mutual information and the KL divergence that is amenable to sparse attack constructions. This chapter has theoretically characterized the optimal single measurement attack case by proving that the resulting cost function is convex and obtaining the optimal attack construction for this case. This chapter distils the insight obtained from the single measurement attack case to propose a sparse attack construction via the greedy algorithm described in Algorithm 1 that overcomes the combinatorial challenge by sequentially removing the attacked sensors and update the Jacobian matrix of the system. This chapter has numerically assessed the performance of the proposed attack on the IEEE 30-bus system and observed that the probability of detection exhibits a threshold effect when a critical number of measurements are compromised.

In addition to the greedy attack constructions in Algorithm 1, the insight obtained from minimizing the cost difference induced by incorporating an additional sensor to the attack is distilled to construct heuristic greedy constructions presented in Algorithm 2. This chapter shows that the greedy step results in a convex optimization problem which can be solved efficiently and yields a low complexity attack update rule. This chapter has numerically evaluated the attack performance in several IEEE test systems and shown that it is feasible to implement disruptive attacks that have access to small number of measurements. Furthermore, it is observed that the topology and the SNR regime govern the performance of the attacks.

Chapter 5

Correlated Sparse Stealth Attacks

In this chapter, the stealth sparse attacks introduced in Chapter 4 are generalized by dropping the assumption of independence and considering correlation between attack vector entries. This chapter tackles the challenge of the combinatorial character of identifying the support of the sparse attack vector by incorporating an additional sensor that yields a sequential sensor selection problem. The convexity of the resulting optimization problem is proved and the insight obtained from incorporating an additional sensor has been distilled to propose an heuristic greedy algorithm. This chapter has numerically assessed the performance of the proposed correlated attack constructions on the IEEE test systems and observed that the performance of the correlated attacks outperform independent attacks at the expense of requiring coordination, i.e., communication, between different attack locations.

5.1 Correlated Sparse Stealth Attack

5.1.1 Sparse Stealth Attack Formulation

The observation model with linearized dynamics given in (4.1) is presented here again for convenience. Recall that

$$Y^m = \mathbf{H}X^n + Z^m, \tag{5.1}$$

where the matrix $\mathbf{H} \in \mathbb{R}^{m\times n}$ is the Jacobian matrix defined in (2.9) at a given operating point and is determined by the system components and the topology of the network, the vector $X^n \in \mathbb{R}^n$ is the vector of random state variables such that $X^n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{XX})$, the vector of random variables $Y^m \in \mathbb{R}^m$ is the vector of measurements that are corrupted by AWGN that is described by the random noise vector $Z^m \in \mathbb{R}^m$ such that

$$Z^m \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_m). \tag{5.2}$$

Therefore, it follows that

$$Y^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{YY}), \tag{5.3}$$

where $\mathbf{\Sigma}_{YY} = \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T} + \sigma^2\mathbf{I}_m$.

The measurements Y^m are corrupted by a random attack vector $A^m \sim P_{A^m}$. This yields the observation model under attacks given by

$$Y_A^m = \mathbf{H}X^n + Z^m + A^m. \tag{5.4}$$

The sparsity constraint is modelled as

$$\text{card}\left(\text{supp}(A^m)\right) = k, \tag{5.5}$$

where $k \leq m$. In view of this, the domain of the k-sparse attack construction is the set of distributions over the attack vector that has at most $k \leq m$ nonzero entries given by (4.31), that is,

$$\tilde{\mathcal{P}}_k \triangleq \left\{P_{A^m} : \text{card}\left(\text{supp}(A^m)\right) = k\right\}. \tag{5.6}$$

The objectives of the attackers are disruption and detection that are captured by mutual information and KL divergence as discussed in Section 4.2. Therefore, the k-sparse attack construction problem that jointly minimizes the mutual information and the KL divergence is cast as an optimization problem given by

$$\min_{\tilde{\mathcal{P}}_k} I(X^n; Y_A^m) + \lambda D(P_{Y_A^m} \| P_{Y^m}) \tag{5.7}$$

5.1.2 Gaussian Sparse Attack Construction

From Lemma 11, the optimal Gaussian attack construction is with a null mean vector, that is

$$A^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{AA}), \tag{5.8}$$

where $\mathbf{\Sigma}_{AA} \in \mathcal{S}_+^m$. From (5.4) and (5.8), the compromised measurements denoted by Y_A^m satisfies that

$$Y_A^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{Y_AY_A}), \tag{5.9}$$

where $\mathbf{\Sigma}_{Y_AY_A} = \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T} + \sigma^2\mathbf{I}_m + \mathbf{\Sigma}_{AA}$. In this section, the assumption of independence in (4.39) is dropped. Instead, the following case is considered:

$$P_{A^m} \neq \prod_{i=1}^{m} P_{A_i}. \tag{5.10}$$

This case boils down to the attack construction given in (4.38), that is

$$\min_{\mathbf{\Sigma}_{AA} \in \mathcal{S}_k} J(\mathbf{\Sigma}_{AA}), \tag{5.11}$$

where $\mathcal{S}_k$ and the cost function J are defined in (4.37) and (4.35), respectively. Recall that

$$\mathcal{S}_k = \left\{\mathbf{S} \in S_+^m : \|\text{diag}(\mathbf{S})\|_0 = k\right\}, \tag{5.12}$$

$$J(\mathbf{\Sigma}_{AA}) = (1-\lambda)\log|\mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_{AA}| - \log|\sigma^2\mathbf{I}_m + \mathbf{\Sigma}_{AA}| + \lambda\text{tr}(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{AA}), \tag{5.13}$$

where $k \leq m$ and the weighting parameter $\lambda \geq 1$ is introduced in (5.7), the real $\sigma \in \mathbb{R}$ is in (5.2) and the matrix $\mathbf{\Sigma}_{YY}$ is in (5.3). The optimization domain given by (5.12) is the set of covariance matrices with k nonzero entries in the diagonal. In next section, the correlation between the attack vector entries that is captured by the nonzero entries of the covariance matrix $\mathbf{\Sigma}_{AA}$ is incorporated.

5.2 Correlated Sparse Stealth Attacks

5.2.1 Correlation Structure

The optimization in (5.11) is carried over the set of covariance matrices with nonzero off-diagonal entries that account for the correlation between different attack entries. As a result, in addition to the set in (5.12) that incorporates the general sparsity constraints, the following set that takes into account the correlation between the attack vector entries is defined:

$$\check{\mathcal{S}}_k \triangleq \bigcup_{\mathcal{K}} \left\{ \mathbf{S} \in S_+^m : i \in \mathcal{K}, (\mathbf{S})_{ii} > 0 \text{ and } (i,j) \in \mathcal{K} \times \mathcal{K}, (\mathbf{S})_{ij} \neq 0 \right\}, \tag{5.14}$$

where $(\mathbf{S})_{ij}$ denotes the entry of the matrix $\mathbf{S}$ in the i-th row and j-th column, the union is over all subsets $\mathcal{K} \subseteq \{1, 2, \ldots, m\}$ with $\text{card}\,(\mathcal{K}) = k \leq m$. Note that it holds that $\check{\mathcal{S}}_k \subseteq \mathcal{S}_k$. Consequently, the optimization problem resulting from the construction of correlated k-sparse attacks is

$$\min_{\mathbf{\Sigma}_{AA} \in \check{\mathcal{S}}_k} J(\mathbf{\Sigma}_{AA}), \tag{5.15}$$

The optimization problem in (5.15) is hard to solve due to the combinatorial character of identifying the support of the sparse random attack vector. To circumvent this problem, this section proposes a structure that firstly sequentially updates the set of measurements being attacked and determines the corresponding entry in the diagonal of the matrix $\mathbf{\Sigma}_{AA}$. Then the nonzero off-diagonal entries that accounts for the correlation between the new index and the attacked measurements introduced are determined.

Specifically, given the sparsity constraint in the optimization domain in (5.15), for all $k < m$, the construction can be divided into k epochs. At each epoch a new element is added to $\text{diag}(\mathbf{\Sigma}_{AA})$. At epoch i, let $\mathbf{\Sigma}_i \in S_+^m$ be the covariance matrix of the vector attack. Let the set $\mathcal{K}_i$ be the set of indices corresponding to the entries of the vector $\text{diag}(\mathbf{\Sigma}_i)$ that are nonzero, that is,

$$\mathcal{K}_i = \{ j \in \{1, 2, \ldots, m\} : (\mathbf{\Sigma}_i)_{jj} > 0 \}. \tag{5.16}$$

In operational terms, the set $\mathcal{K}_i$ is the set of attacked measurements. For all $i \in \{1, 2, \ldots, k\}$, it is imposed that $\mathcal{K}_i \subseteq \{1, 2, \ldots, m\}$ and $\text{card}\,(\mathcal{K}_i) = i$. This implies that $\mathcal{K}_1 \subset \mathcal{K}_2 \subset \ldots \subset \mathcal{K}_k \subset \{1, 2, \ldots, m\}$. Hence,

$$\mathbf{\Sigma}_i = \mathbf{\Sigma}_{i-1} + \mathbf{\Delta}_i, \tag{5.17}$$

where $\mathbf{\Delta}_i \in \mathcal{D}_i$ with

$$\mathcal{D}_i \triangleq \bigcup_{\mathbf{s} \in \mathbb{R}^m} \left\{ \mathbf{D} \in \mathbb{R}^{m \times m} : \mathbf{D} = \mathbf{s}^{\mathsf{T}} \otimes \mathbf{e}_i + \mathbf{s} \otimes \mathbf{e}_i^{\mathsf{T}}, i \in \mathcal{K}_{i-1}^c \right\}. \tag{5.18}$$

Note that the vector $\mathbf{s}$ determines the second order moments describing the covariance between attacked measurements As in the independent attack construction, characterizing the difference enables to formulate the optimization problem that yields the minimum cost increase introduced by a new index in the attack support. Let $\mathbf{\Sigma}_{i-1} \in \mathcal{S}_{i-1}$ be the covariance matrix of the attack vector over $i-1$ measurements. Then the sensor selection problem at

step i is given by the optimization problem:

$$\begin{aligned} \min_{j,\boldsymbol{\Delta}} \; & J\left(\boldsymbol{\Sigma}_{i-1}+\boldsymbol{\Delta}\right) \\ \text{s.t.} \quad & j \in \mathcal{K}_{i-1}^{\mathsf{c}}, \\ & \boldsymbol{\Delta} \in \mathcal{D}_j, \\ & \boldsymbol{\Sigma}_{i-1}+\boldsymbol{\Delta} \in S_+^m. \end{aligned} \tag{5.19}$$

The following lemma sheds light on the solution to the problem in (5.19).

Lemma 16. *Let $\boldsymbol{\Sigma}_i \in S_+^m$ and $\boldsymbol{\Sigma}_{i-1} \in S_+^m$ be two matrices in epoch i and $i-1$, respectively, that satisfy $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_{i-1}+\boldsymbol{\Delta}$ with $\mathcal{K}_{i-1}$ in (5.16), $j \in \mathcal{K}_{i-1}^c$ and $\boldsymbol{\Delta} \in \mathcal{D}_i$. The cost function J in (5.13) satisfies that*

$$J(\boldsymbol{\Sigma}_i) = J(\boldsymbol{\Sigma}_{i-1}) + f_{\text{cor}}(\boldsymbol{\Sigma}_{i-1}, \boldsymbol{\Delta}), \tag{5.20}$$

where the function $f : \mathbb{R}^{m\times m} \times \mathbb{R}^{m\times m} \to \mathbb{R}$ is such that

$$\begin{aligned} f_{\text{cor}}(\boldsymbol{\Sigma}_{i-1}, \boldsymbol{\Delta}) \overset{\Delta}{=} & (1-\lambda)\log\left|\mathbf{I}_m + \left(\boldsymbol{\Sigma}_{YY}+\boldsymbol{\Sigma}_{i-1}\right)^{-1}\boldsymbol{\Delta}\right| - \log\left|\mathbf{I}_m + \left(\sigma^2\mathbf{I}_m+\boldsymbol{\Sigma}_{i-1}\right)^{-1}\boldsymbol{\Delta}\right| \\ & + \lambda \text{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Delta}\right), \end{aligned} \tag{5.21}$$

where $\lambda \geq 1$ is introduced in (5.7) and the matrix $\boldsymbol{\Sigma}_{YY}$ is defined in (5.3).

Proof. The proof consists in showing that the difference between $J(\boldsymbol{\Sigma}_i)$ and $J(\boldsymbol{\Sigma}_{i-1})$ yields

$$\begin{aligned} & J(\boldsymbol{\Sigma}_i) - J(\boldsymbol{\Sigma}_{i-1}) \\ = & (1-\lambda)\log\left|\mathbf{I}_m + \left(\boldsymbol{\Sigma}_{YY}+\boldsymbol{\Sigma}_{i-1}\right)^{-1}\boldsymbol{\Delta}\right| - \log\left|\mathbf{I}_m + \left(\sigma^2\mathbf{I}_m+\boldsymbol{\Sigma}_{i-1}\right)^{-1}\boldsymbol{\Delta}\right| + \lambda\text{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Delta}\right). \end{aligned} \tag{5.22}$$

This completes the proof. □

The relevance of Lemma 16 is that it enables the selection of both $j \in \mathcal{K}_{i-1}^{\mathsf{c}}$ and $\boldsymbol{\Delta} \in \mathcal{D}_j$ at epoch i based on a simpler optimization problem than that in (5.19). Indeed, from Lemma 16, the selection problem in (5.19) is equivalent to

$$\min_{(j,v)\in\mathcal{K}_{i-1}^{\mathsf{c}}\times\mathbb{R}_+} f_{\text{cor}}(\boldsymbol{\Sigma}_{i-1}, \boldsymbol{\Delta}), \tag{5.23}$$

where the function f_{cor} is defined in (5.21).

In the following, it is shown that when the choice of the index selected for attacks in an epoch is fixed, the optimization in (5.19) is convex in the matrix $\boldsymbol{\Delta}$.

Theorem 17. *Let $\boldsymbol{\Sigma}_{i-1} \in \breve{\mathcal{S}}_{i-1}$ and $j \in \mathcal{K}_{i-1}^{\mathsf{c}}$. Then the optimization problem given by*

$$\begin{aligned} \min_{\boldsymbol{\Delta}} \; & J\left(\boldsymbol{\Sigma}_{i-1}+\boldsymbol{\Delta}\right) \\ \text{s.t.} \quad & \boldsymbol{\Delta} \in \mathcal{D}_j, \\ & \boldsymbol{\Sigma}_{i-1}+\boldsymbol{\Delta} \in S_+^m, \end{aligned} \tag{5.24}$$

is convex in $\boldsymbol{\Delta}$.

Proof. From Lemma 16, the following holds for the optimization problem in (5.24).

$$J(\mathbf{\Sigma}_{i-1}+\mathbf{\Delta}) = J(\mathbf{\Sigma}_{i-1}) + f_{\text{cor}}(\mathbf{\Sigma}_{i-1},\mathbf{\Delta}), \tag{5.25}$$

where the function f_{cor} is in (5.21). Hence, the optimization problem in (5.24) is equivalent to

$$\begin{aligned} \min_{\mathbf{\Delta}} \; & f_{\text{cor}}(\mathbf{\Sigma}_{i-1},\mathbf{\Delta}) \\ \text{s.t.} \; & \mathbf{\Delta} \in \mathcal{D}_j, \\ & \mathbf{\Sigma}_{i-1}+\mathbf{\Delta} \in S_+^m, \end{aligned} \tag{5.26}$$

that is,

$$\begin{aligned} \min_{\mathbf{\Delta}} \; & (1-\lambda)\log\left|\mathbf{I}_m + \left(\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{i-1}\right)^{-1}\mathbf{\Delta}\right| - \log\left|\mathbf{I}_m + \left(\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_{i-1}\right)^{-1}\mathbf{\Delta}\right| \\ & +\lambda \text{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Delta}\right) \\ \text{s.t.} \; & \mathbf{\Delta} \in \mathcal{D}_j, \\ & \mathbf{\Sigma}_{i-1}+\mathbf{\Delta} \in S_+^m, \end{aligned} \tag{5.27}$$

which is equivalent to

$$\begin{aligned} \min_{\mathbf{\Delta}} \; & (1-\lambda)\log\left|\mathbf{I}_m + \left(\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{i-1}\right)^{-1}\mathbf{\Delta}\right| - \log\left|\mathbf{I}_m + \left(\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_{i-1}\right)^{-1}\mathbf{\Delta}\right| \\ & +\lambda \text{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Delta}\right) + (1-\lambda)\log\left|\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{i-1}\right| - \log\left|\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_{i-1}\right| \\ \text{s.t.} \; & \mathbf{\Delta} \in \mathcal{D}_j, \\ & \mathbf{\Sigma}_{i-1}+\mathbf{\Delta} \in S_+^m. \end{aligned} \tag{5.28}$$

$$\begin{aligned} = \min_{\mathbf{\Delta}} \; & (1-\lambda)\log\left|\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{i-1}+\mathbf{\Delta}\right| - \log\left|\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_{i-1}+\mathbf{\Delta}\right| + \lambda \text{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Delta}\right) \\ \text{s.t.} \; & \mathbf{\Delta} \in \mathcal{D}_j, \\ & \mathbf{\Sigma}_{i-1}+\mathbf{\Delta} \in S_+^m. \end{aligned} \tag{5.29}$$

Noting that in (5.29), the sets $\mathcal{D}_j$ are convex for all $j \in \mathcal{K}_{i-1}^{\mathsf{c}}$, that the logarithm terms are convex [115] for $\lambda \geq 1$, and that the trace term is linear, yields that the optimization problem in (5.29) is convex. Therefore, the optimization problem (5.24) is convex in $\mathbf{\Delta}$. This completes the proof. □

5.2.2 Greedy Construction

The proposed greedy construction for correlated attack case is described in Algorithm 3. Note that the matrix obtained in the optimization problem in Theorem 17 is constrained by projecting the sum of the update and the previous covariance matrix in the positive semidefinite cone to guarantee that the resulting covariance matrix is indeed positive semidefinite. This is reflected in the last step of Algorithm 3 where the resulting matrix construction is projected by minimizing the Frobenius distance to the positive semidefinite cone. The computation complexity of Algorithm 3 comes from computing the matrix $\mathbf{\Delta}$ that describes the variance of the new compromised measurement and the covariances with the compromised ones. As provided in Theorem 17, the optimization problem is convex. Hence, the algorithm converges and the computation complexity is $O(m(m-k))$.

Algorithm 3 k-sparse correlated attack construction

Input: $\mathbf{H}$ in (5.1);
σ^2 in (5.1);
$\mathbf{\Sigma}_{XX}$ in (5.1);
k in (5.6) and λ in (5.7).
Output: $\mathbf{\Sigma}_{AA}$ in (5.15).

Set $\mathcal{K}_0 = \{\emptyset\}$
Set $\mathbf{\Sigma}_0 = \mathbf{0}$
for $j = 1$ to k **do**
  for $\ell \in \mathcal{K}_{j-1}^{\mathsf{c}}$ **do**
    Compute $\mathbf{\Delta}_\ell = \arg\min_{\mathbf{\Delta} \in \mathcal{D}_\ell} J(\mathbf{\Sigma}_{j-1} + \mathbf{\Delta})$
  end for
  Compute $j^\star = \arg\min_{\ell \in \mathcal{A}_{j-1}^{\mathsf{c}}} J(\mathbf{\Sigma}_{j-1} + \mathbf{\Delta}_\ell)$
  Set $\mathcal{K}_j = \mathcal{K}_{j-1} \cup \{j^\star\}$
  Set $\mathbf{\Sigma}_j = \mathbf{\Sigma}_{j-1} + \mathbf{\Delta}_{j^\star}$
end for
Compute $\mathbf{\Sigma}_{AA} = \arg\min_{\mathbf{S} \in \mathcal{S}_+^m} \|\mathbf{\Sigma}_k - \mathbf{S}\|_{\mathsf{F}}$.

5.3 Numerical Results

5.3.1 Performance in terms of Information Theoretic Cost

Let $\mathbf{\Sigma}_k^{\text{cor}}$ be the output of the k-sparse attack construction of Algorithm 3. This section evaluates the attack performance in terms of the sparsity penalty defined as

$$\eta \triangleq \frac{J(\mathbf{\Sigma}_k^{\text{cor}}) - J(\mathbf{\Sigma}_m^{\text{cor}})}{J(\mathbf{\Sigma}_m^{\text{cor}})}, \tag{5.30}$$

where J is in (5.13). Note that $J(\mathbf{\Sigma}_m^{\text{cor}})$ denotes the cost induced by the construction when all the sensors are attacked. In that sense, this metric captures the performance loss of the attack when only k sensors are attacked.

Fig. 5.1 depicts the performance of the correlated sparse stealth attack construction obtained with Algorithm 3 in different IEEE test systems as a function of the proportion of compromised sensors, that is, k/m, for $\rho = 0.9$ and $\lambda = 8$. As expected, the sparsity penalty decreases monotonically with the proportion of compromised sensors. The sparsity penalty decreases exponentially in the number of compromised sensors. Note that the exponential decrease slope is approximately constant, which indicates that the advantage of adding more sensors to the attack construction decreases exponentially at an approximately constant rate. Remarkably, this exponential decrease is observed for all system sizes and SNR regimes.

It is worth noting that for most systems, operating with larger SNR yields a lower mutual information for the same KL divergence. However, in Fig. 5.1 for the IEEE 30-bus test system the 10 dB and 30 dB performance curves cross, which indicates that the lower SNR regime benefits the attacker when the number of comprised sensors grows. Interestingly, the size of the network does not determine the performance the attack. From Fig. 5.1, 14-bus system is the most vulnerable system to the attacks. This statement only holds for high SNR

regime. In lower SNR regime, 30-bus system is the most vulnerable system. This suggests that the topology of the network fundamentally changes the performance of the attack but the specific mechanisms are left for future study. Fig. 5.2, depict the performance of the correlated attacks from Algorithm 3 in terms of the sparsity penalty in linearized AC model.

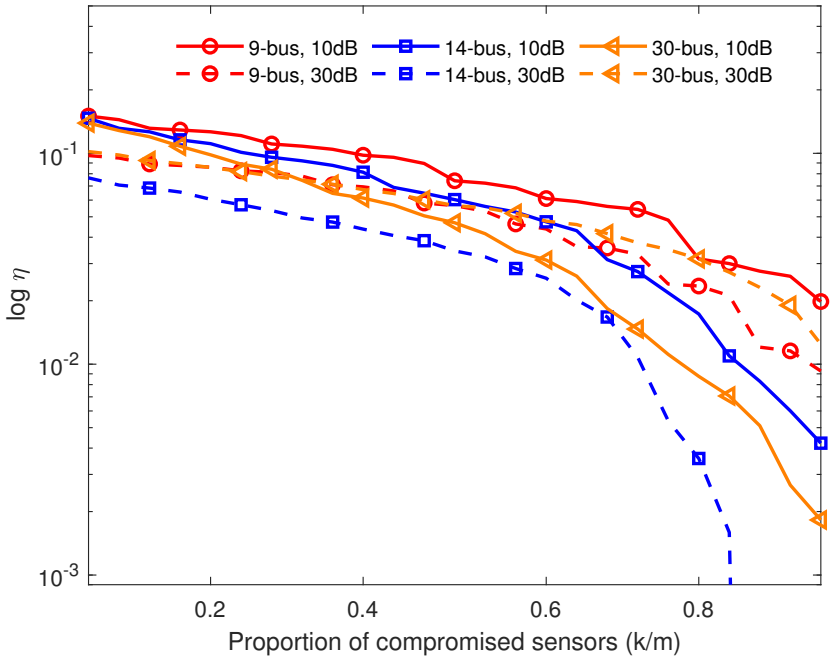

Figure 5.1: Performance of correlated attack constructions in DC model on different IEEE test systems with $\rho = 0.9$ and $\lambda = 8$.

5.3.2 Performance in terms of the Tradeoff between Mutual Information and KL Divergence

Fig. 5.3 and Fig. 5.4 depict the multiobjective performance of the correlated attack construction via Algorithm 3 in terms of the tradeoff between mutual information and KL divergence for different values of the proportion of compromised sensors when SNR $= 30$ dB and $\rho = 0.9$. As expected, larger values of the parameter λ yield smaller values of KL divergence, i.e. the probability of detection is prioritized in the construction over the mutual information decrease for all the scenarios. Moreover, smaller values of k yield smaller reductions of the mutual information, which indicates that remaining stealthy in a sparse setting necessarily implies reducing the disruption induced by the attacks in the state variables. On the other hand, larger values of k enable the attacker to more effectively tradeoff disruption for stealth, which reinforces the previous observation regarding the value of coordination between attack variables to achieve stealth.

Fig. 5.5 and Fig. 5.6 depict the multiobjective performance of the correlated attack case in linearized AC model via Algorithm 3 for the IEEE 9-bus and the IEEE 14-bus systems,

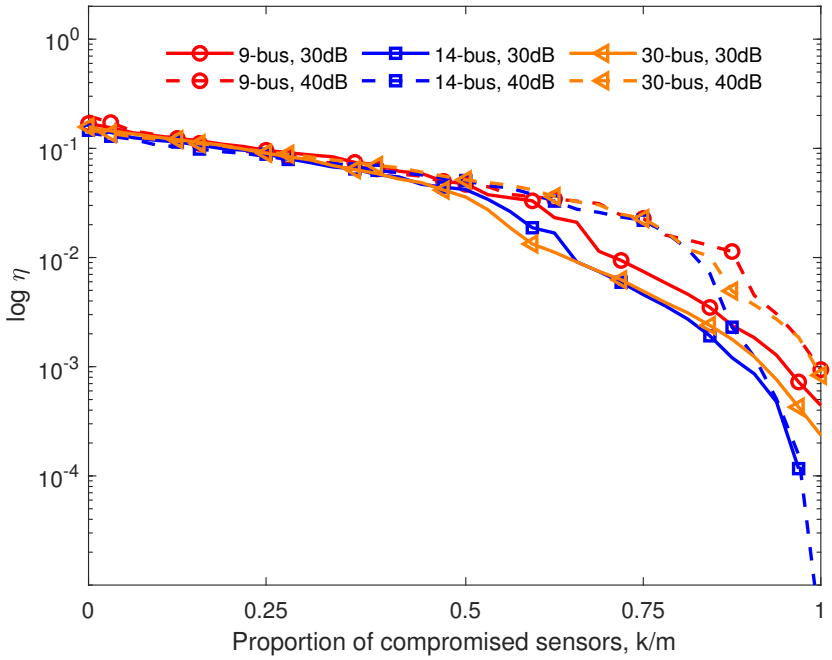

Figure 5.2: Performance of correlated attack constructions in linearized AC model on different IEEE test systems with $\rho = 0.9$ and $\lambda = 8$.

respectively, with SNR = 30 dB and $\rho = 0.9$.

5.3.3 Performance in terms of Disruption to State Estimation

Fig. 5.7 depicts the deviation of the LS estimate caused by one realization of the correlated attack constructions via Algorithm 3 on the IEEE 9-bus test system with different values of k when $\lambda = 2$, SNR = 30 dB and $\rho = 0.9$. With correlated attacks, both small and large values of k deviate the LS estimate for all state variables. The attack construction results in deviation of the LS estimate for all state variables to different extent. In the worst case, the deviation can cause 2 percent deviation. Fig. 5.8 depicts the absolute value of the deviation of the LS estimate caused by averaging 2×10^4 realizations of the attack construction via Algorithm 3 on the IEEE 9-bus test system with different values of k when $\lambda = 2$, SNR = 30 dB and $\rho = 0.9$. As expected, the attack construction with larger k yields larger deviation of the estimates for all the state variables.

The tradeoff between probability of detection and probability of false alarm for the attack construction via Algorithm 3 on the IEEE 9-bus test system with LRT, LNRT and RT is depicted in Fig. 5.9 for different k when $\lambda = 2$, $\rho = 0.9$, SNR = 30 dB. It is expected that LRT outperforms other detection method and smaller value of k yields smaller probability of detection in all every detection methods.

Fig. 5.10 depicts the deviation of the LS estimate caused by one realization of the correlated attack constructions via Algorithm 3 on the IEEE 9-bus test system with different values of λ when $k = 15$, SNR = 30 dB and $\rho = 0.9$. With correlated attacks, both small

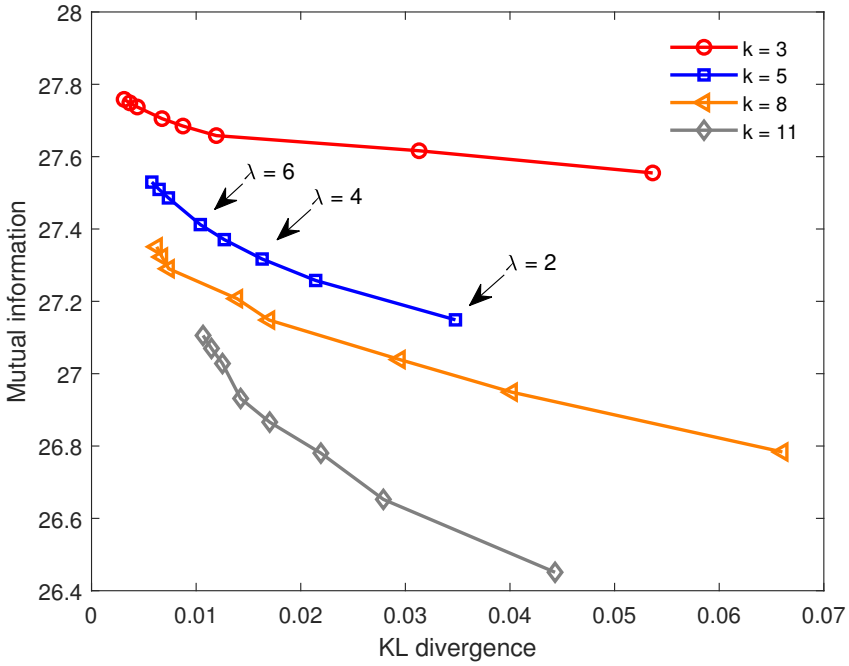

Figure 5.3: Performance of correlated sparse attack construction in DC model in terms of mutual information and KL divergence for different values of λ on the IEEE 9-bus system with SNR = 30 dB and $\rho = 0.9$.

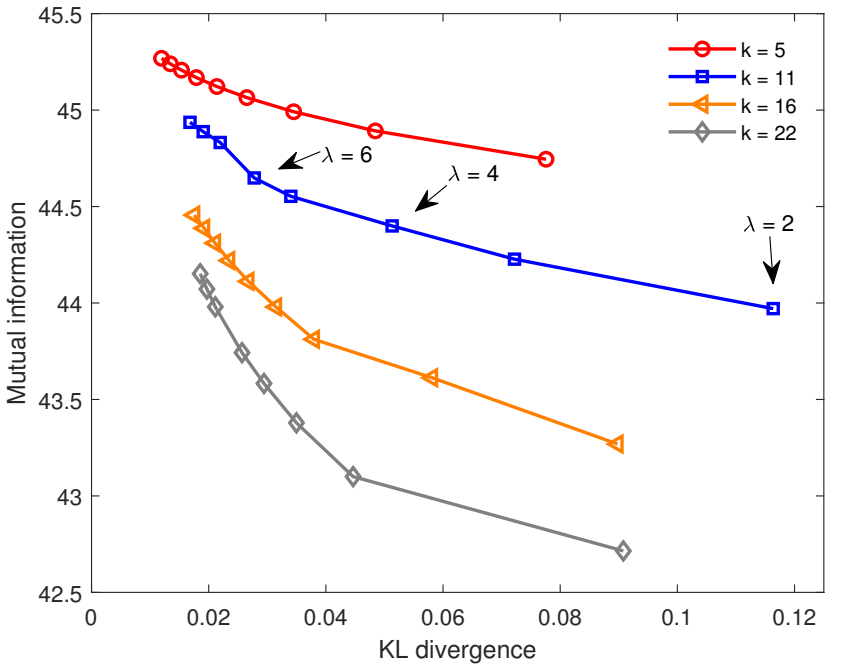

Figure 5.4: Performance of correlated sparse attack construction in DC model in terms of mutual information and KL divergence for different values of λ on the IEEE 14-bus system with SNR = 30 dB and $\rho = 0.9$.

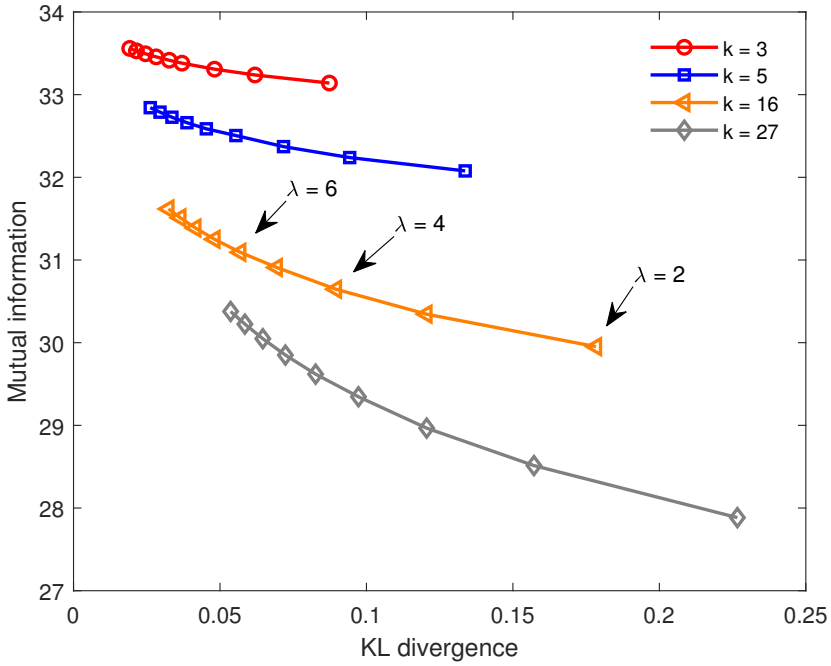

Figure 5.5: Performance of correlated sparse attack construction in linearized AC model in terms of mutual information and KL divergence for different values of λ on the IEEE 9-bus system with SNR = 30 dB and $\rho = 0.9$.

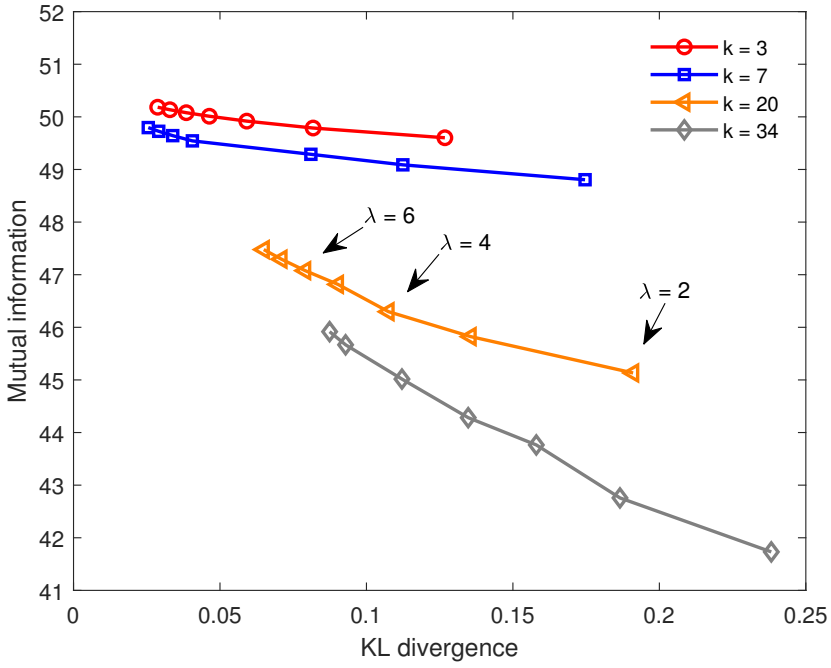

Figure 5.6: Performance of correlated sparse attack construction in linearized AC model in terms of mutual information and KL divergence for different values of λ on the IEEE 14-bus system with SNR = 30 dB and $\rho = 0.9$.

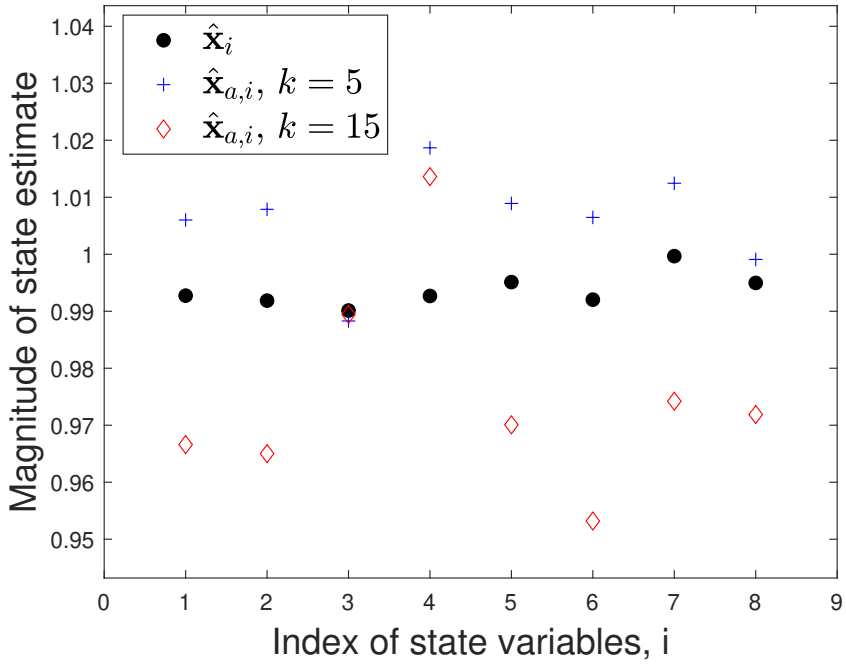

Figure 5.7: Performance of correlated attack construction in terms of WLS state estimate with and without attacks on the IEEE 9-bus test system with one realization when $\lambda = 2$.

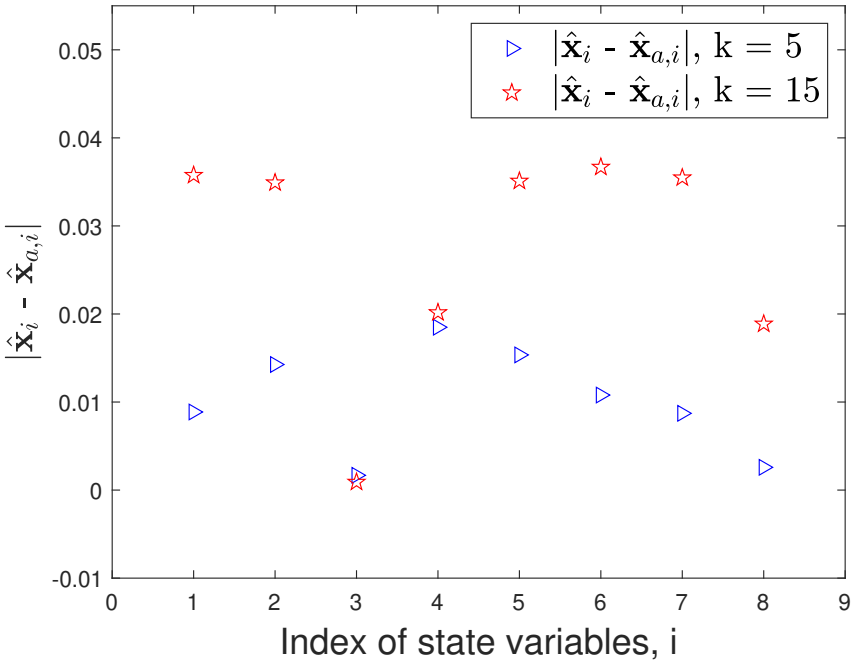

Figure 5.8: Performance of correlated attack construction in terms of the average absolute deviation of WLS state estimate on the IEEE 9-bus test system with 2×10^4 realizations when $\lambda = 2$.

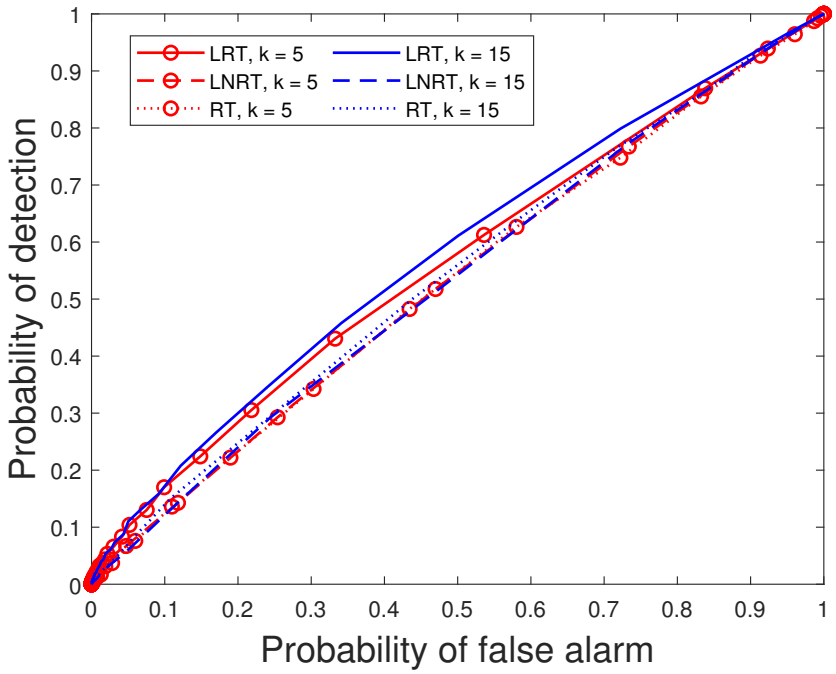

Figure 5.9: Performance of correlated attack construction in terms of probability of detection and probability of false alarm in LRT, LNRT and RT on the IEEE 9-bus test system when $\lambda = 2$.

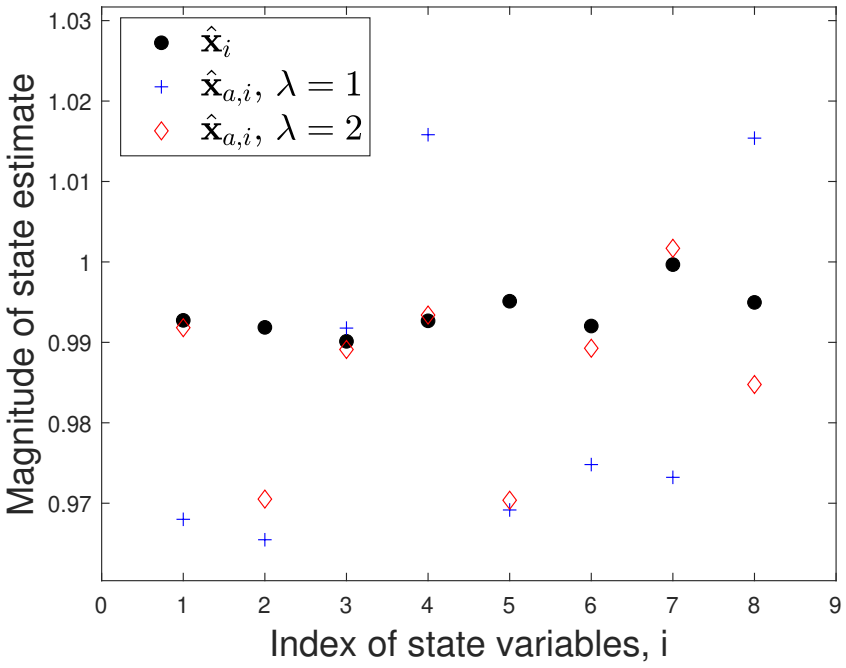

Figure 5.10: Performance of correlated attack construction in terms of WLS state estimate with and without attacks on the IEEE 9-bus test system with one realization when $k = 15$.

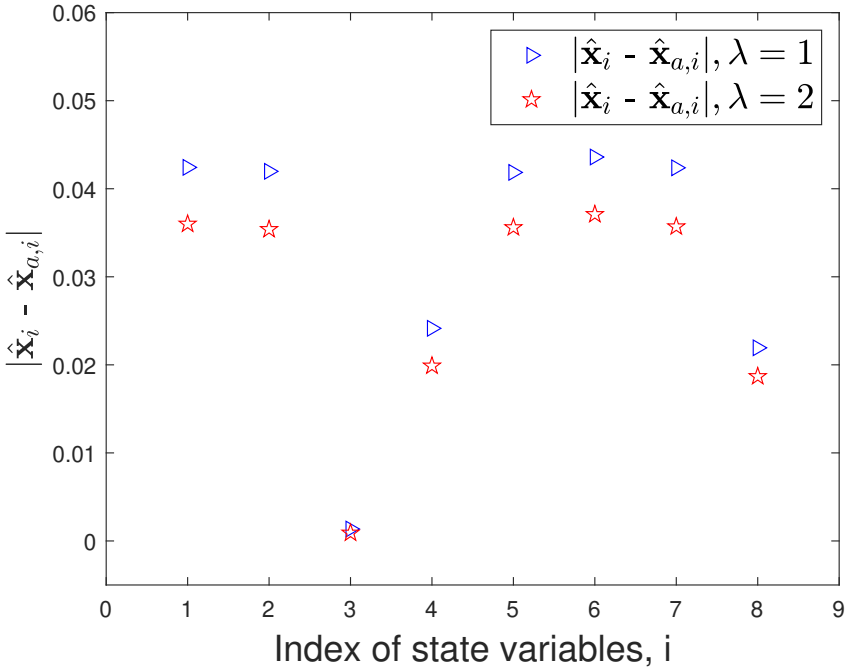

Figure 5.11: Performance of correlated attack construction in terms of the average absolute deviation of WLS state estimate on the IEEE 9-bus test system with 2×10^4 realizations when $k = 15$.

and large values of λ yield deviation of the LS estimate for all state variables. The attacks induce a deviation of the LS estimate for all state variables with differences in the extent of the disruption depending on the state variable index. Fig. 5.11 depicts the absolute value of the deviation of the LS estimate caused by averaging 2×10^4 realizations of the attack constructions via Algorithm 3 on the IEEE 9-bus test system with different values of λ when $k = 15$, SNR $= 30$ dB and $\rho = 0.9$. As expected, the attack construction with smaller λ yields larger deviation of estimate in all the state variables at the cost of stealthiness. Fig.5.12 depicts the tradeoff between probability of detection and probability of false alarm for the attack construction via Algorithm 3 on the IEEE 9-bus test system with LRT, LNRT and RT for different λ when $k = 15$, $\rho = 0.9$, SNR $= 30$ dB. As expected smaller value of λ yields larger probability of detection in all detection methods.

5.3.4 Performance in terms of Mutual Information and Probability of Attack Detection

Fig. 5.13 and Fig. 5.14 depict the performance of the attack construction for different values of λ and sparse constraint k with SNR $= 30$ dB, $\rho = 0.9$ and $\tau = 2$ for the IEEE 9-bus and the IEEE 14-bus test systems, respectively. As expected, larger values of the parameter λ yield smaller values of the probability of attack detection while increasing the mutual information between the vector of state variables and the vector of observations in the systems. Note that the probability of attack detection decreases approximately linearly with respect to

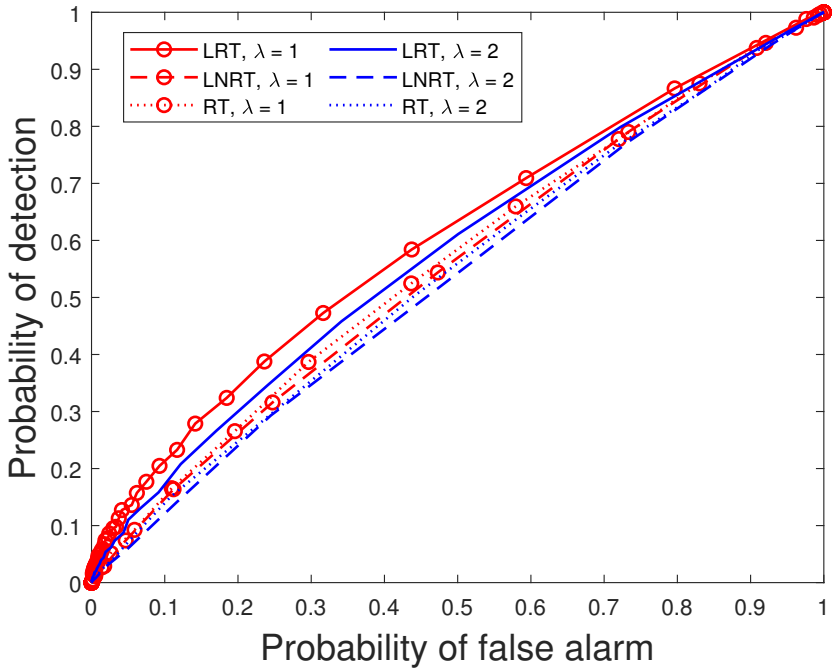

Figure 5.12: Performance of correlated attack construction in terms of probability of detection and probability of false alarm in LRT, LNRT and RT on the IEEE 9-bus test system when $k = 15$.

$\log \lambda$ for small values of λ. Simultaneously for this range of λ, mutual information increases approximately linearly with respect to $\log \lambda$. For moderate values of λ, it is observed that there is a significant decrease in the probability of detection with respect to $\log \lambda$ with a smaller rate of increase in mutual information. The comparison between independent and correlated attack constructions, shows that for the same sparsity constraint, the correlated attack construction successfully exploits the coordination between different locations to yield a smaller probability of detection and a smaller mutual information.

In order to numerically evaluate the performance correlated attacks via Algorithm 3 in comparison with independent attacks construction via Algorithm 2 in Chapter 4, Fig. 5.15 depicts the average absolute deviation of the state estimate of independent attacks and correlated attacks when $\lambda = 1$ and $k = 15$ on the IEEE 9-bus test system. The deviation for correlated attacks is in general larger than for independent attacks, which implies larger disruption in state estimate. Fig. 5.16 depicts the corresponding probability of detection and probability of false alarm in independent attacks and correlated attacks when $\lambda = 1$ and $k = 15$ on the IEEE 9-bus test system. For the same probability of false alarm, correlated attacks achieves a lower probability of detection in comparison with independent attacks. Overall, correlated attacks yield larger disruption but obtain lower probability of detection. Therefore, correlated attacks outperform independent attacks at the expense of coordination between different locations.

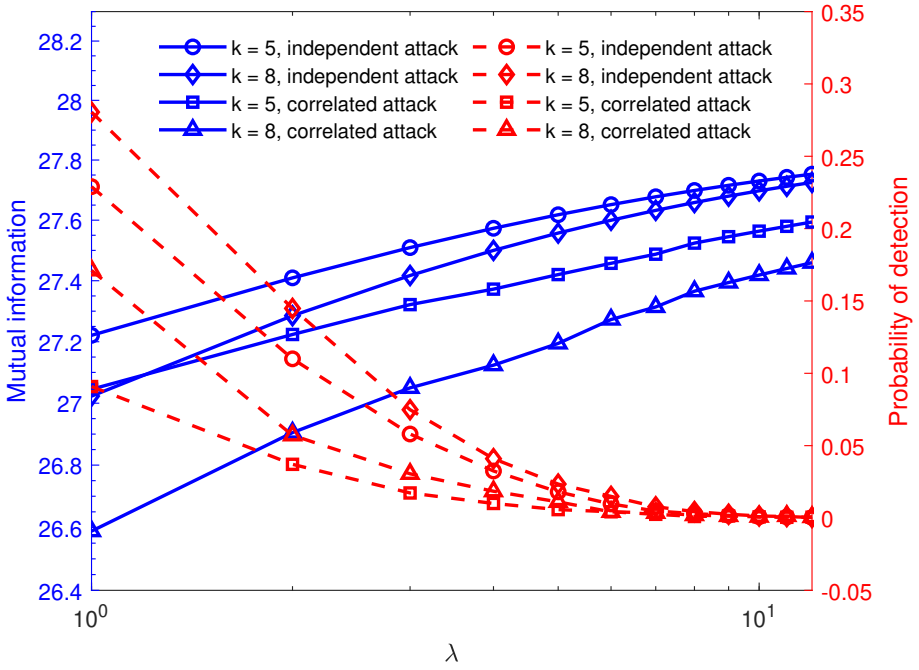

Figure 5.13: Performance of attack constructions on the IEEE 9-bus test system with $\rho = 0.9$, SNR $= 30$ dB and $\tau = 2$.

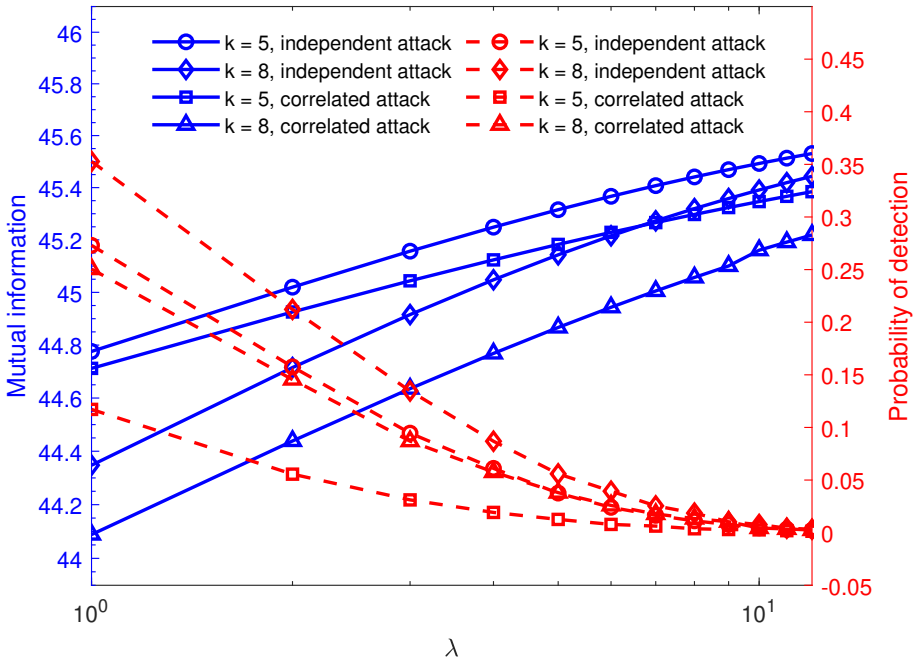

Figure 5.14: Performance of attack constructions on the IEEE 14-bus test system with $\rho = 0.9$, SNR $= 30$ dB and $\tau = 2$.

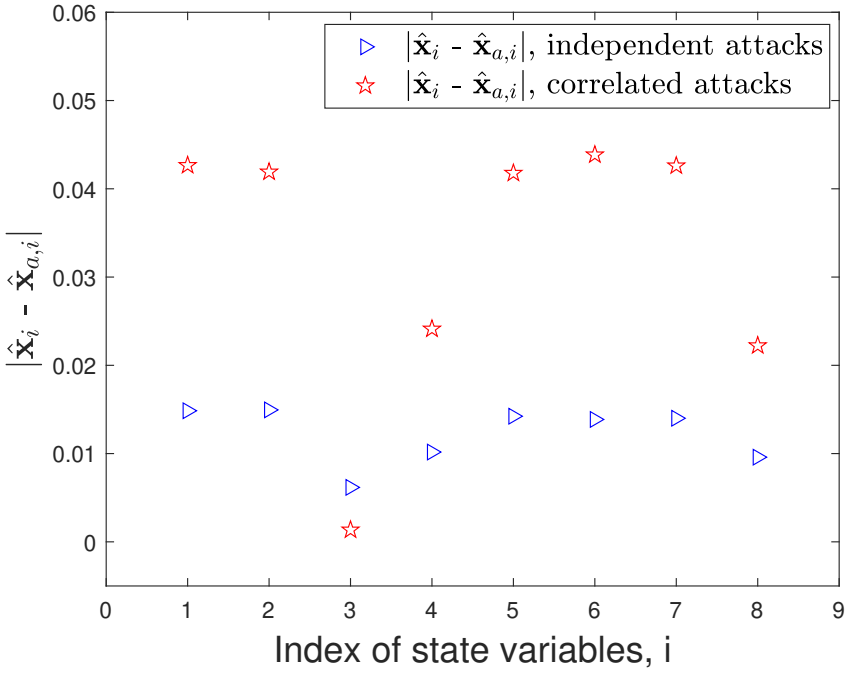

Figure 5.15: Performance of independent attacks and correlated attacks in terms of the average absolute deviation on WLS estimate on the IEEE 9-bus test system with 2×10^4 realizations when $\lambda = 1$ and $k = 15$.

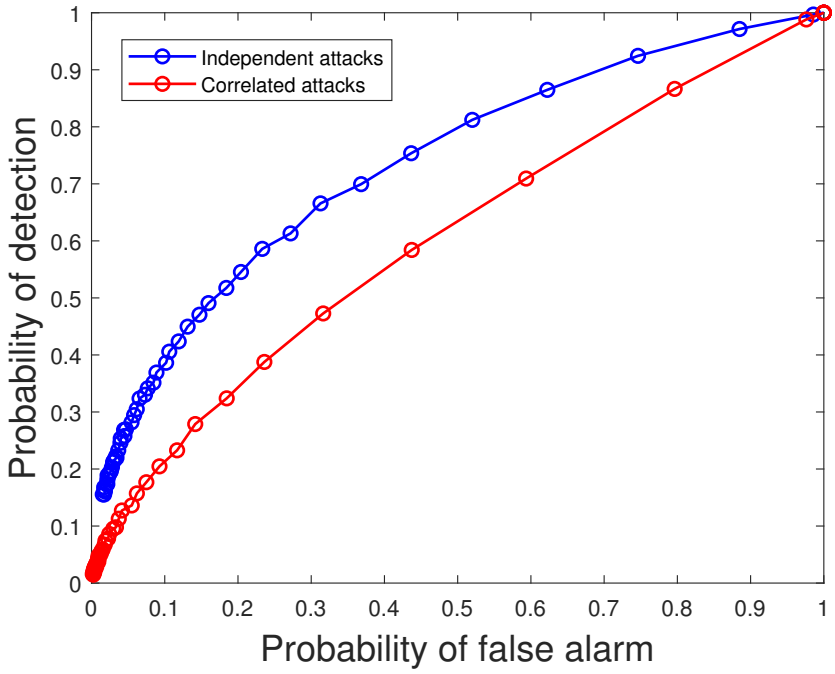

Figure 5.16: Performance of independent attacks and correlated attacks in terms of probability of detection and probability of false alarm on the IEEE 9-bus test system when $\lambda = 1$ and $k = 15$.

5.4 Summary

This chapter has proposed novel stealth attack construction with sparsity constraints. The insight obtained from the problem of incorporating an additional sensor to the attack has been distilled to construct heuristic greedy constructions for the correlated attack cases. It is shown that the greedy step results in a convex optimization problem which can be solved efficiently and yields a low complexity attack update rule. The simulation has numerically evaluated the attack performance in several IEEE test systems and shown that it is feasible to implement disruptive attacks that have access to a small number of measurements. The increase of KL divergence and decrease of mutual information quantitatively with different sparsity constraints are depicted in Fig. 5.3 to Fig. 5.6 in DC and AC systems. The disruption to state estimation quantitatively and the probability of detection under different detection methods are presented in Fig. 5.7 to Fig. 5.16. Furthermore, it is observed that the topology and the SNR regime govern the performance of the attack and numerically characterized the dependence.

Chapter 6

Measurement Vulnerability Analysis

This chapter proposed a fundamental metric to assess the vulnerability of measurements on the smart grid to data integrity attacks. The new metric, coined vulnerability index (VuIx), leverages information theoretic measures to assess the attack effect on the fundamental limits of the disruption and detection tradeoff. As in the previous chapters, an information theoretic framework is adopted to characterize the fundamental information loss induced by data integrity attacks. The result of computing the VuIx of the measurements in the system yields an ordering of the measurements vulnerability based on the level of exposure to data integrity attacks. This new framework is used to assess the vulnerability of the measurements of the IEEE test systems and it is observed that power injection measurements are overwhelmingly more vulnerable to data integrity attacks than power flow measurements. A detailed numerical evaluation of the VuIx metric for the IEEE test systems is provided.

6.1 Information Theoretic Attacks Modelling

The observation model with linearized dynamics within Bayesian framework is given in definition 7, that is

$$Y^m = \mathbf{H}X^n + Z^m, \tag{6.1}$$

where the Jacobian matrix $\mathbf{H} \in \mathbb{R}^{m\times n}$ is defined in (2.9), the vector $X^n \in \mathbb{R}^n$ is the vector of random state variables such that

$$X^n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{XX}), \tag{6.2}$$

the vector $Y^m \in \mathbb{R}^m$ is the vector of measurements that are corrupted by AWGN that is described by the random noise vector $Z^m \in \mathbb{R}^m$ such that

$$Z^m \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_m). \tag{6.3}$$

Therefore, it follows that

$$Y^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{YY}), \tag{6.4}$$

where

$$\mathbf{\Sigma}_{YY} = \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T} + \sigma^2\mathbf{I}_m. \tag{6.5}$$

The measurements Y^m are corrupted by a random attack vector $A^m \sim P_{A^m}$. This yields the observation model under attacks as follows

$$Y_A^m = \mathbf{H}X^n + Z^m + A^m. \tag{6.6}$$

From Lemma 11, the optimal Gaussian attack construction is with a null mean vector, that is

$$A^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{AA}), \tag{6.7}$$

where $\mathbf{\Sigma}_{AA} \in \mathcal{S}_+^m$. From (6.6) and (6.7), the compromised measurements denoted by Y_A^m satisfies that

$$Y_A^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{Y_AY_A}), \tag{6.8}$$

where $\mathbf{\Sigma}_{Y_AY_A} = \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T} + \sigma^2\mathbf{I}_m + \mathbf{\Sigma}_{AA}$.

6.2 Attack Structure with Sequential Sensor Selection

To assess the impact of the attacks to different measurements, the entries of the attack vector with independency are modelled as

$$P_{A^m} = \prod_{i=1}^{m} P_{A_i}, \tag{6.9}$$

where for all $i \in \{1, 2, \ldots, m\}$, the distribution P_{A_i} is Gaussian with zero mean and variance $v \in \mathbb{R}_+$. Consider that k sensors have been compromised with $k \in \{1, 2, \ldots, m-1\}$ and let the covariance matrix of the corresponding attack vector A^m in (6.6) be

$$\mathbf{\Sigma} \in \mathcal{S}_k, \tag{6.10}$$

where $\mathcal{S}_k$ is the set of m-dimensional positive semidefinite matrix with k nonzero entries in the diagonal, that is,

$$\mathcal{S}_k \triangleq \{\mathbf{S} \in \mathcal{S}_+^m : \|\mathrm{diag}(\mathbf{S})\|_0 = k\}. \tag{6.11}$$

Let the set of sensors that have not been compromised be

$$\mathcal{K}_o = \{i \in \{1, 2, \ldots, m\} : (\mathbf{\Sigma})_{ii} = 0\}. \tag{6.12}$$

The sequential sensor selection imposes the following structure in the covariance matrix of the attack vector:

$$\mathbf{\Sigma}_{AA} = \mathbf{\Sigma} + v\mathbf{e}_i\mathbf{e}_i^\mathsf{T}, \tag{6.13}$$

where $i \in \mathcal{K}_o$ and $v \in \mathbb{R}_+$. The cost function $f : \mathcal{S}_k \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathcal{K}_o \to \mathbb{R}_+$ defined by adding (4.18) and (4.21) is as follows:

$$f(\mathbf{\Sigma}, \lambda, v, i) \triangleq I(X^n; Y_A^m) + \lambda D(P_{Y_A^m} \| P_{Y^m}) \tag{6.14}$$

$$= \frac{1}{2}(1-\lambda)\log\left|\mathbf{\Sigma}_{YY} + \mathbf{\Sigma} + v\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right| - \frac{1}{2}\log\left|\mathbf{\Sigma} + v\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \sigma^2\mathbf{I}_m\right| \tag{6.15}$$

$$+ \frac{1}{2}\lambda\left(\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\left(\mathbf{\Sigma} + v\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)\right) + \log|\mathbf{\Sigma}_{YY}|\right).$$

6.3 Information Theoretic Vulnerability of A Measurement

This section proposes a notion of vulnerability that is linked to the information theoretic cost function proposed in [17] to characterize the disruption and detection tradeoff incurred by the attacks. Taking the state of the system with k compromised measurements as the baseline, this section quantifies the vulnerability of each measurement in terms of the cost decrease induced by attacking a sensor i with $i \in \mathcal{K}_o$. In the following, the vulnerability of a measurement is defined.

Definition 10. *The function $\Delta : \mathcal{S}_+^m \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathcal{K}_o \to \mathbb{R}_+$, where $\mathcal{K}_o$ is in* (6.12)*, defines the vulnerability of measurement i in the following form:*

$$\Delta(\mathbf{\Sigma}, \lambda, v, i) \stackrel{\Delta}{=} f(\mathbf{\Sigma}, \lambda, v, i) - f(\mathbf{\Sigma}, \lambda, 0, i), \tag{6.16}$$

where the function f is defined in (6.14)*.*

Note that the attacker aims to minimize (6.14) by choosing an index i and a variance v, and therefore, the definition above implies that given that k sensors in $\{1, 2, \ldots, m\} \setminus \mathcal{K}_o$ are already attacked in the system, the most vulnerable measurement is obtained by solving the following optimization problem

$$\min_{i \in \mathcal{K}_o} \Delta(\mathbf{\Sigma}, \lambda, v, i), \tag{6.17}$$

where $\mathcal{K}_o$ is defined in (6.12).

6.3.1 Vulnerability Analysis of Uncompromised Systems

This section first consideres the case in which no sensors are under attack, that is, $k = 0$, and the attacker selects a single sensor and corrupts the corresponding measurement with a given budget $v \leq v_0$. The vulnerability of measurement i is quantified in terms of $\Delta(\mathbf{\Sigma}, \lambda, v, i)$.

For the uncompromised system case, the optimization problem in (6.17) can be solved in closed form expression. The following theorem provides the solution.

Theorem 18. *The solution to the problem in* (6.17)*, with $\mathcal{K}_o = \{1, 2, \ldots, m\}$, is*

$$i = \underset{j \in \{1,2,\ldots,m\}}{\arg\min} \left\{ \left(\mathbf{\Sigma}_{YY}^{-1} \right)_{jj} \right\}, \tag{6.18}$$

where $\mathbf{\Sigma}_{YY}$ is in (6.5)*.*

Proof. Note that in (6.15), $f(\mathbf{0}, \lambda, 0, i)$ is a constant with respect to i. Hence, optimization problem in (6.17) is equivalent to

$$\min_{i \in \{1,2,\ldots,m\}} f(\mathbf{0}, \lambda, v, i). \tag{6.19}$$

Let $\lambda \in \mathbb{R}_+$ and $v \in \mathbb{R}_+$. The resulting problem in (6.19) is equivalent to the following optimization problem:

$$\min_{i \in \{1,2,\ldots,m\}} (1-\lambda) \log \left(1 + v \operatorname{tr} \left(\mathbf{\Sigma}_{YY}^{-1} \mathbf{e}_i \mathbf{e}_i^{\mathsf{T}} \right) \right) + \lambda v \operatorname{tr} \left(\mathbf{\Sigma}_{YY}^{-1} \mathbf{e}_i \mathbf{e}_i^{\mathsf{T}} \right). \tag{6.20}$$

The proof concludes by noting that the cost function in (6.20) is monotonically increasing with respect to $\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right)$. □

From Theorem 18, it follows that the identification of the most vulnerable measurement is independent of λ in (6.14) and the variance v. That is, it exclusively depends on the system topology denoted by $\mathbf{\Sigma}_{YY}$ in (6.5). This result coincides with Theorem 13 in the sense that in the attack construction for $k = 1$ in (4.41), the most vulnerable measurement is in (4.43a), which is independent of the value of λ.

The following corollary characterizes the vulnerability ordering in uncompromised systems.

Corollary 18.1. *Let the parameters $\mathbf{\Sigma} = \boldsymbol{0}$, $v \in \mathbb{R}_+$, $\lambda \in \mathbb{R}_+$ and $i \in \{1, 2, \ldots, m\}$. The measurements vulnerability ascending ordering for an uncompromised system is given by the ascending ordering of* $\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right)$.

6.3.2 Information Theoretic Vulnerability Index (VuIx)

The vulnerability analysis of uncompromised systems in Section 6.3.1 is constrained to $k = 0$. To generalize the vulnerability analysis to compromised systems when $k > 0$, in the following, a novel metric, coined *vulnerability index*, for all $i \in \mathcal{K}_o$ is proposed.

Definition 11. *For $k \in \{1, 2, \ldots, m-1\}$ and $\mathcal{S}_k$ in (6.11), consider the parameters $\mathbf{\Sigma} \in \mathcal{S}_k$, $v \in \mathbb{R}_+$, $\lambda \in \mathbb{R}_+$. Consider also the set $\{(i, \Delta) : i \in \mathcal{K}_o\}$, with $\mathcal{K}_o$ in (6.12) and $\Delta_i \triangleq \Delta(\mathbf{\Sigma}, \lambda, v, i)$. Let the vulnerability ranking $\mathbf{r} = (r_1, r_2, \ldots, r_{\mathrm{card}(\mathcal{K}_o)})$ be such that for all $i \in \{1, 2, \ldots, \mathrm{card}\,(\mathcal{K}_o)\}$, $r_i \in \mathcal{K}_o$ and moreover,*

$$\Delta_{r_1} \leq \Delta_{r_2} \leq \ldots \leq \Delta_{r_{\mathrm{card}(\mathcal{K}_o)}}. \tag{6.21}$$

The vulnerability index (VuIx) of measurement $r_j \in \mathcal{K}_o$ is j, that is, $\mathrm{VuIx}(r_j) = j$.

Note that the measurement with the smallest VuIx is the most vulnerable measurement that corresponds the solution to the optimization problem in (6.17). The proposed VuIx for $i \in \mathcal{K}_o$ is obtained from Algorithm 4.

6.4 Numerical Results

This section numerically evaluates the VuIx of the measurements on a DC model for the IEEE test systems [112]. The voltage magnitudes are set to 1.0 per unit, that is, the measurements of the systems are active power flow between the buses that are physically connected and active power injection to all the buses. The Jacobian matrix $\mathbf{H}$ in (6.1) determined by the topology of the system and the physical parameter of the branches is generated by MATPOWER [113]. A Toeplitz model is adopted for the covariance matrix $\mathbf{\Sigma}_{XX}$ defined in (6.2). In this setting, the VuIx of the measurements is also a function of the correlation parameter ρ, the noise variance σ^2, and the Jacobian matrix $\mathbf{H}$. The noise regime in the observation model is characterized by SNR defined in (4.57). As discussed in Theorem 18,

Algorithm 4 Computation of Vulnerability Index (VuIx)

Input: $\mathbf{H}$ in (6.1);
$\mathbf{\Sigma}_{XX}$ in (6.2);
σ^2 in (6.3);
$\mathbf{\Sigma} \in \mathcal{S}_k$ in (6.10);
$\lambda \in \mathbb{R}_+$ and $v \in \mathbb{R}_+$.
Output: the VuIx for all $i \in \mathcal{K}_o$.
1: Set $\mathcal{K}_o$ in (6.12)
2: **for** $i \in \mathcal{K}_o$ **do**
3: Compute $\Delta(\mathbf{\Sigma}, \lambda, v, i)$ in (6.16)
4: **end for**
5: Sort $\Delta(\mathbf{\Sigma}, \lambda, v, i)$ in ascending order
6: Set $\mathbf{r} = (r_1, r_2, \ldots, r_{\mathrm{card}(\mathcal{K}_o)})$
7: Set the VuIx of measurement $r_j \in \mathcal{K}_o$ as j.

the solution to the optimization problem is unique and analytical. The algorithm converges and the computation complexity of Algorithm 4 is $O(m)$.

For all $\lambda \in \mathbb{R}_+$ and $v \in \mathbb{R}_+$, this section generates a realization of k attacked indices $\mathcal{K}_a \subseteq \{1, 2, \ldots, m\}$ that is uniformly sampled from the set of attack k-tuples given by

$$\tilde{\mathcal{K}} = \{\mathcal{A} \subseteq \{1, 2, \ldots, m\} : \mathrm{card}\,(\mathcal{A}) = k\}\,. \tag{6.22}$$

Then, a random covariance matrix is constructed to describe the existing attacks on the system as

$$\tilde{\mathbf{\Sigma}} = \sum_{i \in \mathcal{K}_a} \mathbf{e}_i \mathbf{e}_i^{\mathsf{T}}, \tag{6.23}$$

with $\mathcal{K}_a \in \tilde{\mathcal{K}}$. In the numerical simulation, the vulnerability of measurement i is obtained by computing

$$\Delta(\tilde{\mathbf{\Sigma}}, \lambda, 1, i), \tag{6.24}$$

where $i \in \mathcal{K}_o$ is in (6.12) and Δ is defined in (6.16).

6.4.1 Assessment of Vulnerability Index (VuIx)

Fig. 6.1 and Fig. 6.2 depict the mean and variance of the VuIx obtained from Algorithm 4 for all the measurements with SNR = 10 dB, $\lambda = 2$ and $\rho = 0.1$ on the IEEE 9-bus system when $k = 1$ and $k = 2$, respectively. It is observed that power injection measurements yield higher priority vulnerability indices, which indicates that power injection measurements are more vulnerable to data integrity attacks. Most power injection measurements correspond to higher ranked vulnerability indices but there are instances of power flow measurements with a higher ranked VuIx than that of some power injection measurements. Interestingly, the power injection measurements with lower vulnerability indices correspond to the buses that are isolated in the system, that is, the buses with a lower number of connections. On the other hand, the power flow measurements with higher ranked vulnerability indices correspond to the branches with higher admittance. The VuIx for $k = 0$ obtained in Corollary 18.1 is

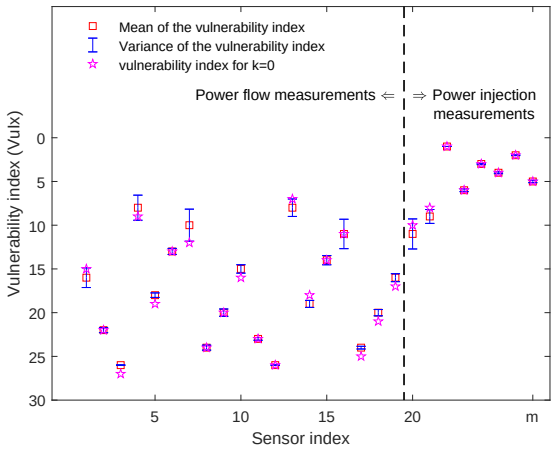

Figure 6.1: Vulnerability index (VuIx) when $k = 1$, SNR $= 10$ dB, $\lambda = 2$ and $\rho = 0.1$ on the IEEE 9-bus system.

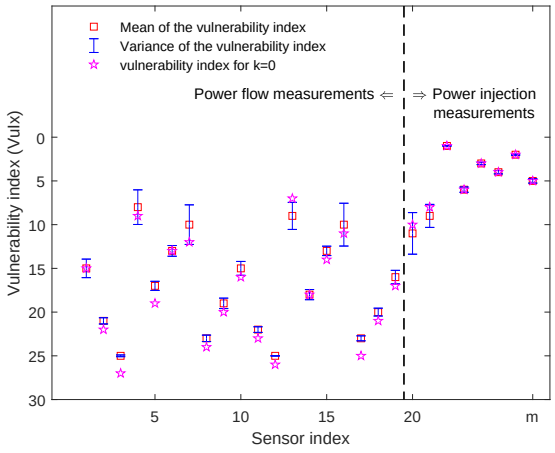

Figure 6.2: Vulnerability index (VuIx) when $k = 2$, SNR $= 10$ dB, $\lambda = 2$ and $\rho = 0.1$ on the IEEE 9-bus system.

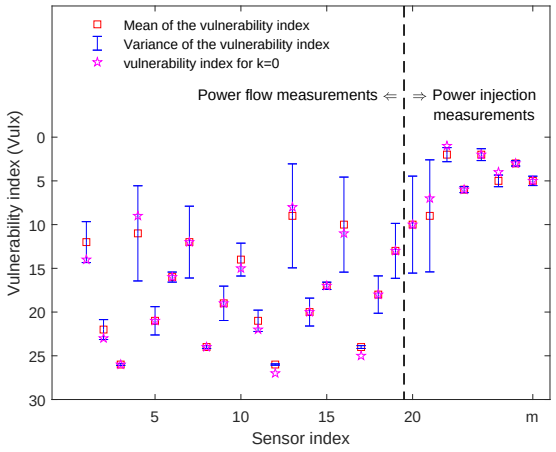

Figure 6.3: Vulnerability index (VuIx) when $k = 1$, SNR $= 30$ dB, $\lambda = 2$ and $\rho = 0.1$ on the IEEE 9-bus system.

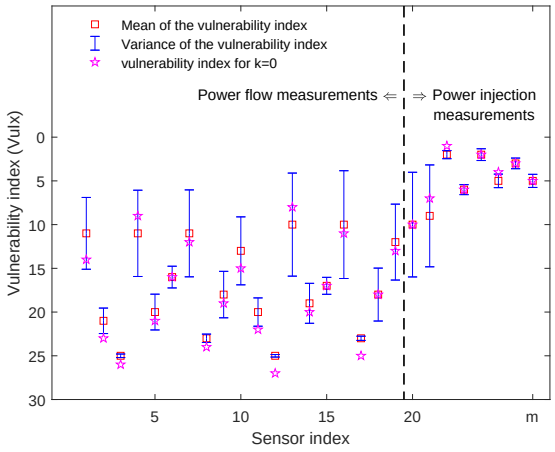

Figure 6.4: Vulnerability index (VuIx) when $k = 2$, SNR $= 30$ dB, $\lambda = 2$ and $\rho = 0.1$ on the IEEE 9-bus system.

depicted for the purpose of serving as a reference to assess the deviation when $k > 0$. Interestingly, the VuIx of most measurements does not change significantly for different values of k, which suggests that the VuIx is insensitive to the state of the system.

Fig. 6.3 and Fig. 6.4 depict the mean and variance of the VuIx from Algorithm 4 for all the measurements with SNR $= 30$ dB, $\lambda = 2$ and $\rho = 0.1$ on the IEEE 9-bus system when $k = 1$ and $k = 2$, respectively. Interestingly, the mean of the VuIx for most of the measurements does not deviate significantly from the case when $k = 0$. Instead, most of the variances deviate significantly in comparison with the cases in Fig. 6.1 and Fig. 6.2 with

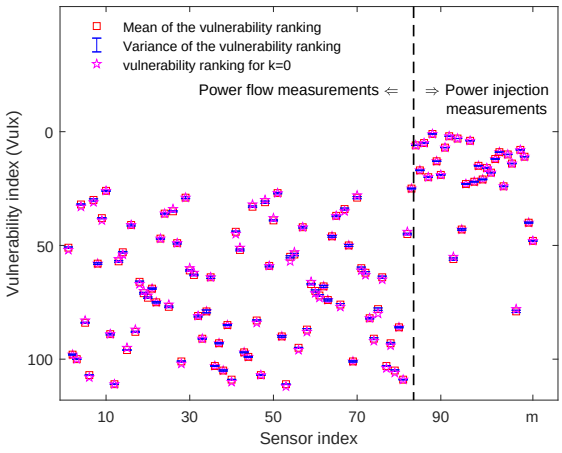

Figure 6.5: Vulnerability index (VuIx) when $k = 1$, SNR $= 10$ dB, $\lambda = 2$ and $\rho = 0.1$ on the IEEE 30-bus system.

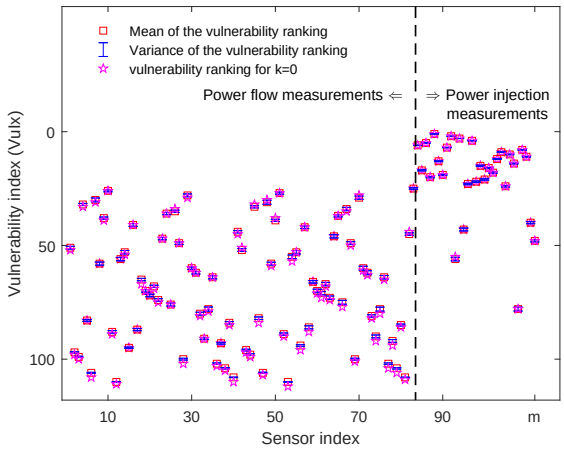

Figure 6.6: Vulnerability index (VuIx) when $k = 2$, SNR $= 10$ dB, $\lambda = 2$ and $\rho = 0.1$ on the IEEE 30-bus system.

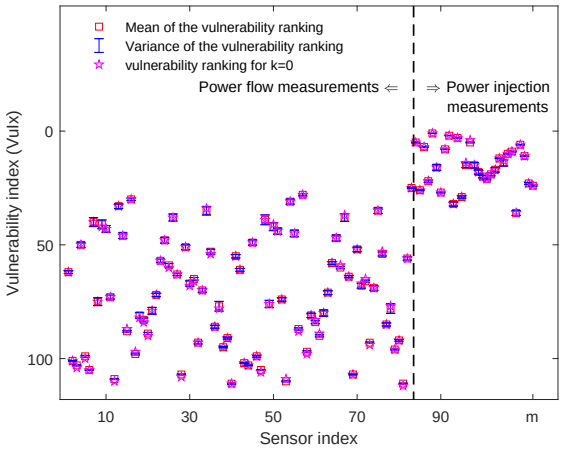

Figure 6.7: Vulnerability index (VuIx) when $k = 1$, SNR $= 30$ dB, $\lambda = 2$ and $\rho = 0.1$ on the IEEE 30-bus system.

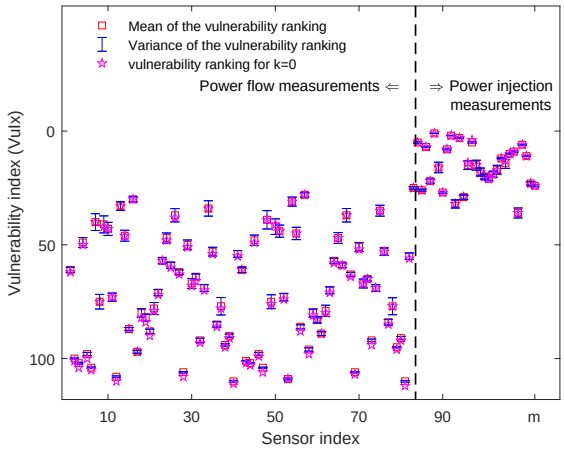

Figure 6.8: Vulnerability index (VuIx) when $k = 2$, SNR $= 30$ dB, $\lambda = 2$ and $\rho = 0.1$ on the IEEE 30-bus system.

SNR $= 10$ dB. Fig. 6.5 and Fig. 6.6 depict the results on the IEEE 30-bus systems with the same setting as in Fig. 6.1 and Fig. 6.2, respectively. Fig. 6.7 and Fig. 6.8 depict the results on the IEEE 30-bus systems with the same setting as in Fig. 6.3 and Fig. 6.4, respectively. Surprisingly, the mean of the VuIx in larger systems coincides with that obtained for the case $k = 0$, which shows that the VuIx is a robust security metric for large systems. Interestingly, the power injection measurements corresponding to the least connected buses decrease in the VuIx when SNR $= 10$ dB.

6.4.2 Comparative Vulnerability Assessment of Power Flow and Power Injection Measurements

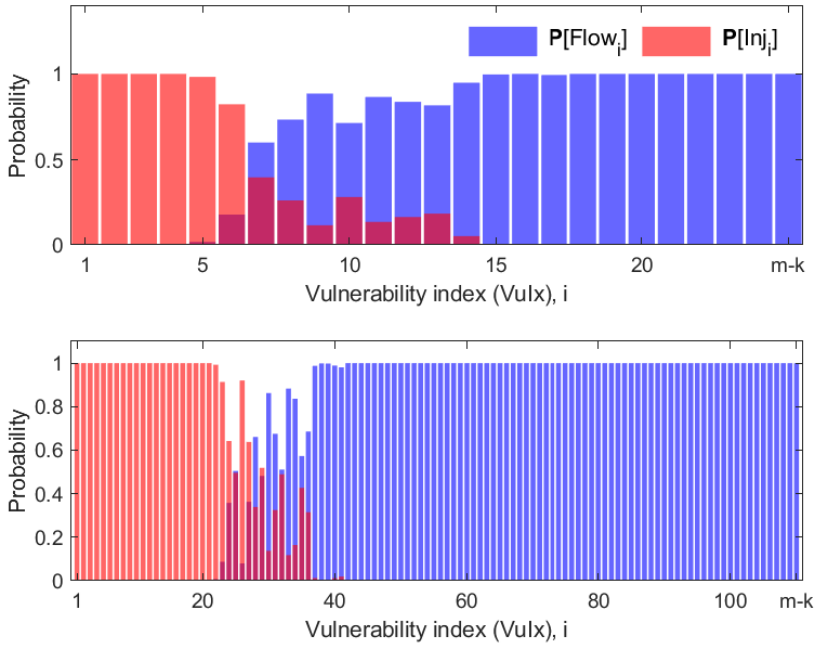

Figure 6.9: Probability of Vulnerability index (VuIx) corresponds to power injection measurements and power flow measurements when $\lambda = 2$, $k = 2$, SNR $= 30$ dB and $\rho = 0.1$ on the IEEE 9-bus and 30-bus systems.

In Section 6.4.1, it is established that power injection measurements and power flow measurements are qualitatively different in terms of the VuIx. To provide a quantitative description of this difference, Fig. 6.9 depicts the probability of a given VuIx $i \in \{1, 2, \ldots, m - \text{card}\,(\mathcal{K}_a)\}$ being taken by a power injection measurement or by a power flow measurement for the IEEE 9-bus and 30-bus systems when $\lambda = 2$, $k = 2$, SNR $= 30$ dB and $\rho = 0.1$. Specifically, Fig. 6.9 depicts the probability of the following events:

$$\mathsf{Flow}_i : \text{VuIx } i \text{ corresponds to a power flow measurement,}$$
$$\mathsf{Inj}_i : \text{VuIx } i \text{ corresponds to a power injection measurement.}$$

It is observed that in both systems, small VuIx values are more likely to correspond to power injection measurements than to power flow measurements, that is, $\mathbb{P}[\mathsf{Inj}_i] > \mathbb{P}[\mathsf{Flow}_i]$ for small values of i. Conversely, it holds that $\mathbb{P}[\mathsf{Inj}_i] < \mathbb{P}[\mathsf{Flow}_i]$ for large values of i. In fact, small VuIx values corresponding to power injection measurements is with probability one, which shows that the most vulnerable measurements in the system are always power injection measurements. Similarly, larger VuIx values corresponding to power flow measurements is with probability one, which indicates that the least vulnerable measurements are always power flow measurements. Interestingly, there is a clear demarcation for each system

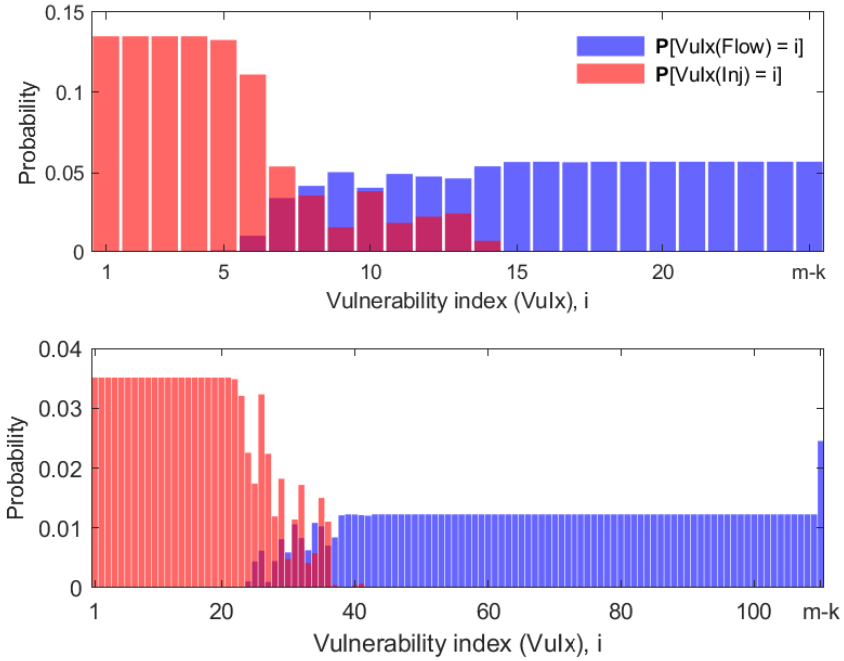

Figure 6.10: Distributions of Vulnerability index (VuIx) for power injection measurements and power flow measurements when $\lambda = 2$, $k = 2$, SNR $= 30$ dB and $\rho = 0.1$ on the IEEE 9-bus and 30-bus systems.

for which $\mathbb{P}[\mathsf{Inj}_i]$ and $\mathbb{P}[\mathsf{Flow}_i]$ change rapidly with the VuIx value, which suggests a phase transition type phenomenon for measurement vulnerability.

Fig. 6.10 depicts the probability mass function (pmf) of the VuIx $i \in \{1, 2, \ldots, m - \text{card}\,(\mathcal{K}_a)\}$ for power injection measurements or power flow measurements on the IEEE 9-bus and 30-bus systems when $\lambda = 2$, $k = 2$, SNR $= 30$ dB and $\rho = 0.1$. Specifically, in Fig. 6.10, $\mathbb{P}[\mathsf{VuIx(Inj)}]$ and $\mathbb{P}[\mathsf{VuIx(Flow)}]$ depict the pmf of the VuIx for power injection measurements and power flow measurements, respectively, on the IEEE 9-bus and 30-bus systems when $\lambda = 2$, $k = 2$, SNR $= 30$ dB and $\rho = 0.1$. The probability mass for power injection measurements concentrates on the vulnerability indices with higher priority. Whereas, the probability mass for power flow measurements concentrates on the low ranked vulnerability indices. The pmf with high and low vulnerability indices are evenly distributed. Interestingly, in 30-bus system, the probability of lowest ranked VuIx for power flow measurements experiences a sharp increase.

6.4.3 Comparative Vulnerability Assessment of Selected Power Flow and Power Injection Measurements

In Section 6.4.2, the VuIx of power injection measurements and power flow measurements are quantitatively assessed. To provide a quantitative description of typical power injection measurements and power flow measurements, this section presents the pmf of selected power injection measurements or power flow measurements. Specifically, Fig. 6.11 depicts the

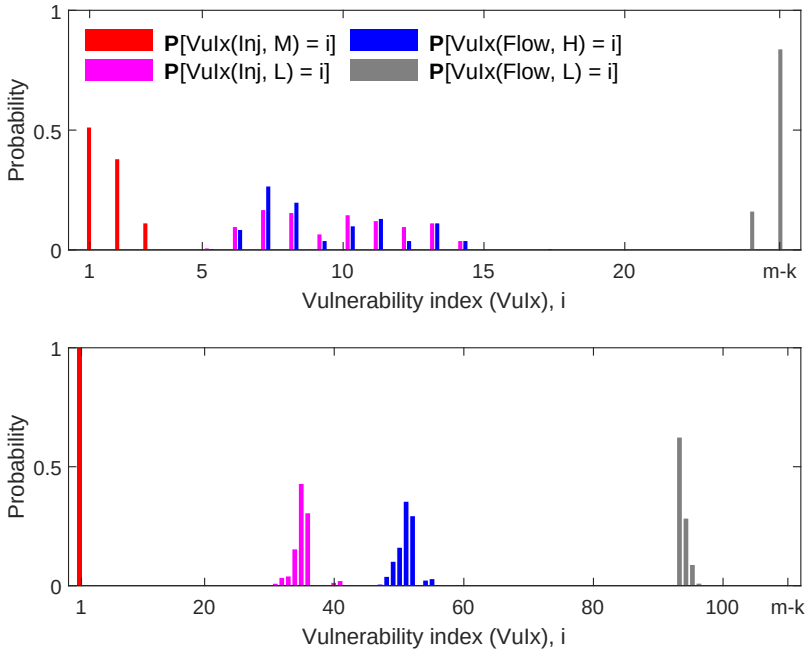

Figure 6.11: The pdf of Vulnerability index (VuIx) for power injection measurements and power flow measurements when $\lambda = 2$, $k = 2$, SNR $= 30$ dB and $\rho = 0.1$ on the IEEE 9-bus and 30-bus systems.

distributions of the following events:

$$\mathsf{VuIX}(\mathsf{Inj}, \mathsf{M}) = \mathsf{i} : \text{VuIx } i \text{ corresponds to the most connected power injection measurement,}$$
$$\mathsf{VuIX}(\mathsf{Inj}, \mathsf{L}) = \mathsf{i} : \text{VuIx } i \text{ corresponds to the least connected power injection measurement,}$$
$$\mathsf{VuIX}(\mathsf{Flow}, \mathsf{H}) = \mathsf{i} : \text{VuIx } i \text{ corresponds to power flow measurement with the highest admittance,}$$
$$\mathsf{VuIX}(\mathsf{Flow}, \mathsf{L}) = \mathsf{i} : \text{VuIx } i \text{ corresponds to power flow measurement with the lowest admittance.}$$

It is observed that in both systems, there is a significant difference in the pmf for these four selected measurements. The most connected power injection measurements have high probability mass in VuIx for small values of i. Specifically, in the 30-bus system, the VuIx of the most connected power injection measurement is 1 is with probability 1, that is $\mathbb{P}[\mathsf{VuIx}(\mathsf{Inj}, \mathsf{M}) = 1] = 1$, which indicates the most connected power injection measurement is always the most vulnerable measurement in the system. Conversely, the power flow measurements with the lowest admittance have high probability mass in VuIx for small values of i. Specifically, in 9-bus system, for large values of i, it holds that $\mathbb{P}[\mathsf{VuIx}(\mathsf{Flow}, \mathsf{L}) = \mathsf{i}]$ is large, which indicates that the power flow measurement with lowest admittance is likely to be the most not vulnerable measurement in the system. This coincides with the results on optimal single measurement attack case in Theorem 13 in Chapter 4. For the least connected power injection measurement and the power flow measurement with the highest admittance, the pmf takes nonzero values for medium i. Interestingly, in 9-bus system, for medium i, it holds that $\mathbb{P}[\mathsf{VuIx}(\mathsf{Inj}, \mathsf{L}) = \mathsf{i}] \neq 0$ and $\mathbb{P}[\mathsf{VuIx}(\mathsf{Flow}, \mathsf{M}) = \mathsf{i}] \neq 0$, which indicates these two selected

measurements can have the same VuIx. However, in 30-bus system, the least connected power injection measurement always has smaller VuIx than the power flow measurement with highest admittance.

6.5 Summary

This chapter has designed a novel security metric referred to as VuIx that characterizes vulnerability of power system measurements to data integrity attacks from a fundamental perspective. This is achieved by embedding information theoretic measures into the metric definition. The resulting VuIx framework evaluates the vulnerability of all the measurements in the systems and enables the operator to identify those that are more exposed to data integrity threats. The simulations have tested the framework for the IEEE test systems and concluded that power injection measurements are more vulnerable to data integrity attacks than power flow measurements.

Chapter 7

Decentralized Stealth Attacks

This chapter presents the main results on decentralized stealth attacks with coordination. Specifically, the interaction between the attackers is modelled as a game in a normal form. Considering the information theoretic metrics, the cost functions for the attackers to launch a random attack cooperatively is with different objectives inspired by mutual information and KL divergence both globally and locally. It is proved that the games are potential games with corresponding potential functions. The uniqueness and achievability of the Nash Equilibriums (NEs) in the games are obtained. The best response in each game is characterized. This chapter also proposed best response dynamics to evaluate the performance of the decentralized stealth attacks and achieve the NEs in the games.

7.1 System Model

In a decentralized system, the observation model is as described in Section 3.6, that is,

$$Y^m \sim P_{Y^m} = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{YY}), \tag{7.1}$$

with

$$\mathbf{\Sigma}_{YY} \triangleq \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}} + \sigma^2\mathbf{I}_m. \tag{7.2}$$

Note that for all $i \in \{1, 2, \ldots, m\}$, the probability distribution of the i-th entry of the random vector Y^m is denoted by P_{Y_i}, that is,

$$Y_i \sim P_{Y_i} = \mathcal{N}(0, \mathbf{e}_i^{\mathsf{T}}\mathbf{\Sigma}_{YY}\mathbf{e}_i). \tag{7.3}$$

The attacker manipulates the measurements in (6.1) at physically protected locations that yields additive FDIAs [2]. In (7.1), given the stochastic nature of the measurements, the attacker pursues a random attack construction strategy. A random malicious attack, denoted by a random vector A^m, compromises the vector of measurements, which yields the vector of compromised measurements as follows:

$$Y_A^m = \mathbf{H}X^n + Z^m + A^m, \tag{7.4}$$

where $A^m \sim P_{A^m}$ and P_{A^m} is the distribution of the random attack vector A^m. In this study, P_{A^m} is assumed to be a multivariate Gaussian distribution that satisfies

$$A^m \sim P_{A^m} = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{AA}), \tag{7.5}$$

where $\mathbf{\Sigma}_{AA} \in S_+^m$ is a covariance matrix.

7.1.1 Decentralized Data Injection Attacks

In [17], the stealth attacks is studied where there is a unique attacker referred to as *centralized attacks*. In a decentralized system, DIAs are constructed by several attackers that have access to the measurements referred to as *decentralized attacks* [13]. In this scenario, the attackers decide the attack vector $A \in \mathbb{R}^m$ cooperatively. The aim of one attacker is to autonomously decide its attack vector to maximize the damage to the system, e.g. distortion to the state estimate, while staying undetected. All the attackers have the same interests, which reveals a cooperative manner among the attackers in decentralized attacks.

Therefore, the random attack A^m in (7.5) in decentralized attacks is modelled with the independence of the entries of attack vector, that is,

$$A^m \triangleq (A_1, A_2, \ldots, A_m)^\mathsf{T}, \tag{7.6}$$

such that

$$P_{A^m} = \prod_{i=1}^{m} P_{A_i} \tag{7.7}$$

where, for all $i \in \{1, 2, \ldots, m\}$, the probability density function of P_{A_i} is Gaussian with zero mean and variance $v_i \in [0, +\infty)$, that is,

$$A_i \sim \mathcal{N}(0, v_i). \tag{7.8}$$

The independence between the random attacks in (7.7) implies that decentralized attacks does not require the communication between different attack locations. That being the case, the attack construction by attackers in different locations is much more practical and particularly interesting. Note that with the independence, the covariance matrix in (7.5) is

$$\mathbf{\Sigma}_{AA} = \sum_{i=1}^{m} v_i \mathbf{e}_i \mathbf{e}_i^\mathsf{T}. \tag{7.9}$$

Consequently, the vector of compromised measurements Y_A^m under the random attacks follows a multivariate Gaussian distribution $P_{Y_A^m}$, that is,

$$Y_A^m \sim P_{Y_A^m} = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{Y_A Y_A}) \tag{7.10}$$

with

$$\mathbf{\Sigma}_{Y_A Y_A} \triangleq \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T} + \sigma^2 \mathbf{I}_m + \mathbf{\Sigma}_{AA}. \tag{7.11}$$

From (7.3) and (7.8), for all $i \in \{1, 2, \ldots, m\}$, the i-th entry of the random vector Y_A^m denoted by $Y_{A,i}$ is such that

$$Y_{A_i} \sim P_{Y_{A,i}} = \mathcal{N}(0, \mathbf{e}_i^\mathsf{T}\mathbf{\Sigma}_{YY}\mathbf{e}_i + v_i), \tag{7.12}$$

where $\mathbf{\Sigma}_{YY}$ is defined in (7.2) and v_i is introduced in (7.8).

The independence assumption of the entries of random attack vector in (7.7) allows the attackers to launch attacks independently. Let $\mathcal{M} \triangleq \{1, 2, \ldots, m\}$ be the set of measurement indices in the system and

$$\mathcal{K} \triangleq \{1, 2, \ldots, m\} \tag{7.13}$$

be the set of attack indices. Assume that measurement i, with $i \in \mathcal{M}$, is the only measurement that attacker i can compromise. Let $A_i^m \in \mathbb{R}^m$ be the random attack vector produced by attacker i and $\mathcal{A}_i$ be the set of random attack vectors that can be injected into the system by attacker i, with $i \in \mathcal{K}$, that is,

$$\mathcal{A}_i = \{A_i^m \in \mathbb{R}^m : (A_i^m)_j = 0 \text{ for all } j \neq i\}. \tag{7.14}$$

Hence, from (7.6) and (7.14), for all $i \in \mathcal{K}$, the following holds

$$A_i^m = A_i \otimes \mathbf{e}_i, \tag{7.15}$$

where $A_i \otimes \mathbf{e}_i$ is the Kronecker product of A_i and $\mathbf{e}_i$. Let the Minkowski sum of $\mathcal{A}_i$ and $\mathcal{A}_j$ be denoted by $\mathcal{A}_i \oplus \mathcal{A}_j$. For all $A^m \in \mathcal{A}_i \oplus \mathcal{A}_j$, there exists a pair of random vectors $(A_i^m, A_j^m) \in \mathcal{A}_i \times \mathcal{A}_j$ such that $A^m = A_i^m + A_j^m$. Let the set of all possible random attack vectors be

$$\mathcal{A} \triangleq \mathcal{A}_1 \oplus \mathcal{A}_2 \oplus \ldots \oplus \mathcal{A}_m, \tag{7.16}$$

and the set of complementary random attack vectors with respect to the attacker i be

$$\mathcal{A}_{-i} \triangleq \mathcal{A}_1 \oplus \mathcal{A}_2 \oplus \ldots \mathcal{A}_{i-1} \oplus \mathcal{A}_{i+1} \ldots \oplus \mathcal{A}_{m-1} \oplus \mathcal{A}_m. \tag{7.17}$$

Denote the random attack vector constructed by the attacker i by $A_i^m \in \mathcal{A}_i$. Hence, the resulting random attack vector is $A^m \in \mathbb{R}^m$ in (7.4) and satisfies

$$A^m = \sum_{i \in \mathcal{K}} A_i^m \in \mathcal{A}. \tag{7.18}$$

Denote also the complementary random attack vector of A_i^m as follows:

$$A_{-i}^m \triangleq \sum_{j \in \mathcal{K} \setminus \{i\}} A_j^m \in \mathcal{A}_{-i}. \tag{7.19}$$

Given the actions by all the other attackers A_{-i}^m, the aim of attacker i is to corrupt the measurements by injecting its random attack vector $A_i^m \in \mathcal{A}_i$ to compromise the data integrity while guaranteeing a low probability of attack detection. For modelling this behaviour, attacker i, with $i \in \mathcal{K}$, adopts the cost function ϕ_i: $\mathbb{R}^m \to \mathbb{R}$ to determine whether a random attack vector $A_i^m \in \mathcal{A}_i$ is more beneficial than another attack vector $B_i^m \in \mathcal{A}_i$. In this context, the attack vector A_i^m is preferred to B_i^m if $\phi_i(A_i^m + A_{-i}^m) < \phi_i(B_i^m + A_{-i}^m)$.

7.2 Information Theoretic Metrics

The aim of attacker i is to corrupt the measurements by injecting a random attack vector $A_i^m \in \mathcal{A}_i$ that maximizes the disruption to state estimate while stays undetectable. For modelling the disruption and the stealthness of the attacks, information theoretic metrics both globally and locally are adopted.

7.2.1 Disruption Metrics

Precisely, the disruption is captured by the mutual information. The following proposition provides the global mutual information.

Proposition 8. *The mutual information between the random vector of state variables $X^n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{XX})$ and the random vector of compromised measurements $Y_A^m \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{Y_AY_A})$ is [17]*

$$I(X^n; Y_A^m) = \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_{XX}||\boldsymbol{\Sigma}_{Y_AY_A}|}{|\boldsymbol{\Sigma}|}, \tag{7.20}$$

where $\boldsymbol{\Sigma}_{XX}$ is in (4.3)*, the matrix $\boldsymbol{\Sigma}_{Y_AY_A}$ is in* (4.10) *and $\boldsymbol{\Sigma}$ is the covariance matrix of the joint distribution of (X^n, Y_A^m) in* (4.17)*.*

The following proposition provides the local mutual information between the state variables X^n and the i-th compromised measurement $Y_{A,i} \triangleq Y_i + A_i$ in (7.12).

Proposition 9. *The mutual information between the vector of random state variables $X^n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{XX})$ and the random variable $Y_{A,i}$ is*

$$I(X^n; Y_{A,i}) = \frac{1}{2}\log\left(1 + \frac{\mathrm{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)}{\sigma^2 + v_i}\right). \tag{7.21}$$

Proof. The proof of Proposition 9 is presented in Appendix G. □

7.2.2 Detection Metrics

As a part of security strategies, the system operator implements an attack detection procedure prior to performing state estimation. The detection is cast as a hypothesis test with hypotheses

$$\mathcal{H}_0\text{: There is no attack,} \tag{7.22a}$$
$$\mathcal{H}_1\text{: Measurements are compromised.} \tag{7.22b}$$

Global Detection

In global detection, the operator decides if the vector of measurements is produced under attacks. This is described in Section 4.1.2. The global detection is captured by the global KL divergence in the following proposition.

Proposition 10. *The KL divergence between the probability distribution functions of the measurements with attacks and without attacks, i.e., $P_{Y_A^m}$ and P_{Y^m}, respectively, is [17]*

$$D(P_{Y_A^m} \| P_{Y^m}) = \frac{1}{2}\left(\log\frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{Y_AY_A}|} - m + \mathrm{tr}(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{Y_AY_A})\right), \tag{7.23}$$

where $\boldsymbol{\Sigma}_{YY}$ is in (4.4) *and the matrix $\boldsymbol{\Sigma}_{Y_AY_A}$ is in* (4.10)*.*

Local Detection

Similarly, in local detection, the operator decides if measurement i, with $i \in \mathcal{M}$, is produced under attacks. Specifically, the operator acquires measurement i denoted by $\bar{Y}_i$ and decides whether it is produced under attack. The hypothesis test problem becomes

$$\mathcal{H}_0\colon \bar{Y}_i \sim \mathcal{N}(0, \mathbf{e}_i^\mathsf{T} \mathbf{\Sigma}_{YY} \mathbf{e}_i), \tag{7.24a}$$
$$\mathcal{H}_1\colon \bar{Y}_i \sim \mathcal{N}(0, \mathbf{e}_i^\mathsf{T} \mathbf{\Sigma}_{YY} \mathbf{e}_i + v_i). \tag{7.24b}$$

A deterministic test $T_{\text{ID}} : \mathbb{R} \to \{0, 1\}$ is adopted to determine which distribution generates the measurements. Given measurement i denoted by $\bar{y}_i$, let $T_{\text{ID}}(\bar{y}_i) = j$ denote the case in which the test decides for $\mathcal{H}_j$ upon $\bar{y}_i$, with $j \in \{0, 1\}$. Therefore, the deterministic test T_{ID} is

$$\mathcal{H}_0\colon T_{\text{ID}}(\bar{y}_i) = 0, \tag{7.25a}$$
$$\mathcal{H}_1\colon T_{\text{ID}}(\bar{y}_i) = 1. \tag{7.25b}$$

In this setting, the LRT is given by

$$T_{\text{ID}}(\bar{y}_i) = \mathbb{1}_{\{L_{\text{ID}}(\bar{y}_i) \gtrless \tau\}}, \tag{7.26}$$

with $\tau \in \mathbb{R}_+$ the decision threshold and the likelihood ratio $L_{\text{ID}}(\bar{y}_i)$ given by

$$L_{\text{ID}}(\bar{y}_i) = \frac{f_{Y_{A,i}}(\bar{y}_i)}{f_{Y_{A,i}}(\bar{y}_i)}, \tag{7.27}$$

where the functions $f_{Y_{A,i}}$ and f_{Y_i} are the probability density function of $Y_{A,i}$ in (7.12) and Y_i in (7.3), respectively.

The following proposition provides the local KL divergence.

Proposition 11. *The KL divergence between two one-dimensional Gaussian distributions $P_{Y_{A,i}}$ and P_{Y_i} is given by*

$$D(P_{Y_{A,i}} \| P_{Y_i}) = \frac{1}{2} \left(\frac{v_i}{\text{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) + \sigma^2} + \log \frac{\text{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) + \sigma^2}{\text{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) + \sigma^2 + v_i} \right). \tag{7.28}$$

Proof. The proof of Proposition 11 is presented in Appendix H. □

7.3 Game Formulation

The cost for the attacker i to launch a random attack not only depends on its own attack vector A_i^m but also on the random attacks A_{-i}^m of all the other attackers. This is implied from the overall resulting random attack vector in (7.18). Particularly, in Gaussian attack construction, given the action profile $\mathbf{v} = (v_1, v_2, \ldots, v_m)^\mathsf{T}$, the cost function of the attacker i does not only depend on its own variance v_i that is added in the i-th sensor in the network,

but also the variance by all the other attackers $v_j, j \neq i$. Therefore, the interaction of all the attackers in the network can be described by a game in normal form

$$\mathcal{G}_p = (\mathcal{K}, \{\mathcal{V}_i\}_{i\in\mathcal{K}}, \{\phi_i^p\}_{i\in\mathcal{K}}), \tag{7.29}$$

where $p \in \mathbb{Z}_+$ yields different aims of the attackers, i.e., different games, when considering global metrics in (7.20) and (7.23) and local metrics in (7.21) and (7.28) in Proposition 9 and Proposition 11, respectively. The set of attacker $\mathcal{K}$ is the set of all the players. The set of all possible action of player i is denoted by $\mathcal{V}_i$ and ϕ_i^p is the cost function when the attacker chooses different information theoretic metrics.

7.3.1 Game Objectives

Global Disruption and Global Detection

When the attacker i considers global mutual information in (7.20), i.e., mutual information between the vector of measurements and the vector of state variables, and global KL divergence in (7.23), i.e., the KL divergence between the distributions of the vector of measurements with and without attacks. This yields the attack construction for attacker i as the following optimization problem:

$$\min_{v_i\in\mathbb{R}_+} I(X^n; Y_A^m) + \lambda D(P_{Y_A^m} \| P_{Y^m}). \tag{7.30}$$

Local Disruption and Global Detection

When the attacker i considers local mutual information in (7.21) in Proposition 9, i.e., mutual information between the random vector of state variables and the i-th measurement with attacks in the system, and global KL divergence in (7.23), i.e., the KL divergence between the distributions of the random vector of measurements with and without attacks. This yields the attack construction for attacker i as the following optimization problem:

$$\min_{v_i\in\mathbb{R}_+} I(X^n; Y_{A,i}) + \lambda D(P_{Y_A^m} \| P_{Y^m}). \tag{7.31}$$

Global Disruption and Local Detection

When the attacker considers global mutual information in (7.20), i.e., mutual information between the random vector of measurements and the random vector of state variables, and local KL divergence in (7.28) in Proposition 11, that is, the KL divergence between the distributions of the i-th measurement with attacks and the i-th measurement without attacks. This yields the attack construction for attacker i as the following optimization problem:

$$\min_{v_i\in\mathbb{R}_+} I(X^n; Y_A^m) + \lambda D(P_{Y_{A,i}} \| P_{Y_i}). \tag{7.32}$$

7.3.2 Costs of the Games

Global Disruption and Global Detection

When considering global mutual information and global KL divergence in (7.30), the game formulation in (7.29) is particularized as

$$\mathcal{G}_1 = \left(\mathcal{K}, \{\mathcal{V}_i\}_{i\in\mathcal{K}}, \{\phi_i^1\}_{i\in\mathcal{K}}\right), \tag{7.33}$$

where $\mathcal{K}$ is the set of players, $\mathcal{V}_i$ is the set of all possible actions by the i-th player, ϕ_i^1 is the utility function of the attacker i in $\mathcal{G}_1$.

The following proposition provides the analytical expression of the cost function for attacker i.

Proposition 12. *Given the action profile* $\mathbf{v} = (v_1, v_2, \ldots, v_m)^\mathsf{T}$*, the cost function for attacker* i *in* $\mathcal{G}_1$ *is* $\phi_i^1 : \mathbb{R}_+^m \to \mathbb{R}_+$ *such that*

$$\phi_i^1(v_1, v_2, \ldots, v_m) \tag{7.34}$$

$$\triangleq I(X^n; Y_A^m) + \lambda D(P_{Y_A^m} \| P_{Y^m}) \tag{7.35}$$

$$= \frac{1}{2}(1-\lambda)\log\left|\mathbf{\Sigma}_{YY} + v_i\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right| - \frac{1}{2}\log\left|\sigma^2\boldsymbol{I}_m + v_i\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right|$$
$$+ \frac{1}{2}\lambda\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\left(v_i\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right)\right). \tag{7.36}$$

Proof. Note that $\mathbf{\Sigma}_{AA} = \sum_{i=1}^m v_i\mathbf{e}_i\mathbf{e}_i^\mathsf{T}$. The proof follows by taking (7.20) and (7.23) into the cost in (7.30). □

The following lemma proposes an equivalent expression for the optimization problem in (7.30).

Lemma 19. *For all* $i \in \{1, 2, \ldots, m\}$*, the optimization problem in* (7.30) *is equivalent to*

$$\min_{v\in\mathbb{R}_+}(1-\lambda)\log\left|\mathbf{\Sigma}_{YY} + v\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right| - \log(\sigma^2+v) + \lambda v\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) \tag{7.37}$$

Proof. The proof of Lemma 19 is presented in Appendix J. □

The following proposition characterizes the convexity of the cost function ϕ_i^1 in (7.34).

Proposition 13. *Let* $\lambda \geq 1$*. Then the cost function of attacker* i ϕ_i^1*:* $\mathbb{R}_+^m \to \mathbb{R}_+$ *in* (7.34) *is convex in* v*.*

Proof. Note that in (7.37), it holds that $\mathbf{\Sigma}_{YY} + \sum_{i=1}^m v_i\mathbf{e}_i\mathbf{e}_i^\mathsf{T} \in \mathbf{S}_{++}^m$. Therefore, the term $\log\left|\mathbf{\Sigma}_{YY} + \sum_{i=1}^m v_i\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right|$ is concave [115]. It yields that $(1-\lambda)\log\left|\mathbf{\Sigma}_{YY} + \sum_{i=1}^m v_i\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right|$ is convex. The trace is a linear operation. It follows that the optimization problem in (7.37) is convex. Therefore, from Lemma 19, the optimization problem in (7.30) convex in v. From Proposition 12, the cost function ϕ_i^1 is convex in v. □

Local Disruption and Global Detection

When considering local mutual information and global KL divergence in (7.31), the game formulation in (7.29) is particularized as

$$\mathcal{G}_2 = \left(\mathcal{K}, \{\mathcal{V}_i\}_{i\in\mathcal{K}}, \{\phi_i^2\}_{i\in\mathcal{K}}\right), \tag{7.38}$$

where $\mathcal{K}$ is the set of players, $\mathcal{V}_i$ is the set of all possible actions by the i-th player, ϕ_i^2 is the utility function of the attacker i in $\mathcal{G}_2$.

The following proposition provides the analytical expression of the cost function for attacker i.

Proposition 14. *Given the action profile* $\mathbf{v} = (v_1, v_2, \ldots, v_m)^\mathsf{T}$*, the cost function for attacker* i *in* $\mathcal{G}_2$ *is* $\phi_i^2 : \mathbb{R}_+^m \to \mathbb{R}_+$ *such that*

$$\phi_i^2(v_1, v_2, \ldots, v_m) \tag{7.39}$$

$$\stackrel{\Delta}{=} I(X^n; Y_{A,i}) + \lambda D(P_{Y_A^m} \| P_{Y^m}) \tag{7.40}$$

$$= \frac{1}{2}\log\left(1 + \frac{\mathrm{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)}{\sigma^2 + v_i}\right) \tag{7.41}$$

$$+ \frac{1}{2}\lambda\left(\log\frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{YY} + v_i\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}|} + \mathrm{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\left(v_i\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right)\right)\right).$$

Proof. Note that $\boldsymbol{\Sigma}_{AA} = \sum_{i=1}^m v_i\mathbf{e}_i\mathbf{e}_i^\mathsf{T}$. The proof follows by taking (7.21) and (7.23) into the cost in (7.31). □

The following lemma proposes an equivalent expression for the optimization problem in (7.31).

Lemma 20. *For all* $i \in \{1, 2, \ldots, m\}$*, the optimization problem in* (7.31) *is equivalent to*

$$\min_{v\in\mathbb{R}_+} \log\left(1 + \frac{\mathrm{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)}{\sigma^2 + v}\right) - \lambda\log\left|\boldsymbol{\Sigma}_{YY} + v\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right| + \lambda v\mathrm{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) \tag{7.42}$$

Proof. The proof of Lemma 20 is presented in Appendix K. □

The following proposition characterizes the convexity of the cost function ϕ_i^2 in (7.39).

Proposition 15. *Let* $\lambda \geq 0$*. Then the cost function of attacker* i ϕ_i^2*:* $\mathbb{R}_+^m \to \mathbb{R}$ *in* (7.39) *is convex in* v*.*

Proof. Note that in (7.42), it holds that $\boldsymbol{\Sigma}_{YY} + v\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T} \in \mathbf{S}_{++}^m$. Therefore, the logarithm term $\log\left|\boldsymbol{\Sigma}_{YY} + v\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right|$ is concave [115]. The term

$$\log\left(1 + \frac{\mathrm{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)}{\sigma^2 + v}\right) \tag{7.43}$$

is convex in v and the trace is a linear operation. This completes the proof. □

Global Disruption and Local Detection

When considering global mutual information and local KL divergence in (7.32), the game formulation in (7.29) is particularized as

$$\mathcal{G}_3 = \left(\mathcal{K}, \{\mathcal{V}_i\}_{i\in\mathcal{K}}, \{\phi_i^3\}_{i\in\mathcal{K}}\right), \tag{7.44}$$

where $\mathcal{K}$ is the set of players, $\mathcal{V}_i$ is the set of all possible actions by the i-th player, ϕ_i^3 is the utility function of the attacker i in $\mathcal{G}_3$.

The following proposition provides the analytical expression of the cost function for attacker i.

Proposition 16. *Given the action profile $\mathbf{v} = (v_1, v_2, \ldots, v_m)^\mathsf{T}$, the cost function for attacker i in $\mathcal{G}_3$ is $\phi_i^3 : \mathbb{R}_+^m \to \mathbb{R}_+$ such that*

$$\begin{aligned}
&\phi_i^3(v_1, v_2, \ldots, v_m) && (7.45)\\
\overset{\Delta}{=}&I(X^n; Y_A^m) + \lambda D(P_{Y_{A,i}} \| P_{Y_i}) && (7.46)\\
=&\frac{1}{2}\log\frac{|\mathbf{\Sigma}_{YY} + v_i\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}|}{|\sigma^2\boldsymbol{I}_m + v_i\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}|} && (7.47)\\
&+\frac{1}{2}\left(\frac{v_i}{\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) + \sigma^2} + \log\frac{\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) + \sigma^2}{\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) + \sigma^2 + v_i}\right).
\end{aligned}$$

Proof. Noth that $\mathbf{\Sigma}_{AA} = \sum_{i=1}^m v_i\mathbf{e}_i\mathbf{e}_i^\mathsf{T}$. The proof follows by taking (7.20) and (7.28) into the cost in (7.32). □

The following lemma proposes an equivalent expression for the optimization problem in (7.32).

Lemma 21. *For all $i \in \{1, 2, \ldots, m\}$, the optimization problem in (7.32) is equivalent to*

$$\begin{aligned}
\min_{v\in\mathbb{R}_+} &\log\left|\frac{1}{\sigma^2+v}\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\sum_{j\in\mathcal{K}\setminus\{i\}}\frac{1}{\sigma^2+v_j}\mathbf{e}_j\mathbf{e}_j^\mathsf{T} + \boldsymbol{I}_m\right|\\
&+\frac{\lambda}{\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) + \sigma^2}v - \lambda\log\left(\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) + \sigma^2 + v\right)
\end{aligned} \tag{7.48}$$

Proof. The proof of Lemma 21 is presented in Appendix L. □

The following proposition characterizes the convexity of the cost function ϕ_i^3 in (7.45).

Proposition 17. *Let $\lambda \geq 0$. Then the cost function of attacker i $\phi_i^3 : \mathbb{R}_+^m \to \mathbb{R}_+$ is convex in v.*

Proof. The proof of Proposition 17 is presented in Appendix I. □

7.3.3 Best Response

Each attacker is a player in the game $\mathcal{G}_p, p \in \{1,2,3\}$ and it is identified by an index from the set $\mathcal{K}$. The action that player i adopts is the variance of the random variable A_i added to measurement i of the system, i.e., v_i. The underlying assumption in the following of this section is that, given a sequence of actions by all the other players, player i adopts an action such that the cost of launching the attack $\phi_i^p(v_1, v_2, ..., v_m)$ is minimized. That is,

$$v_i \in \mathrm{BR}_i^p(v_1, v_2, ..., v_{i-1}, v_{i+1}, ..., v_{m-1}, v_m), \tag{7.49}$$

where the correspondence BR_i^p: $\mathbb{R}_+^{m-1} \to \mathbb{R}_+$ is the best response correspondence, i.e.,

$$\mathrm{BR}_i^p(v_1, v_2, ..., v_{i-1}, v_{i+1}, ..., v_{m-1}, v_m) = \underset{v_i \in \mathbb{R}_+}{\arg\min}\, \phi_i^p(v_1, v_2, ..., v_m). \tag{7.50}$$

Best Response in $\mathcal{G}_1$

In $\mathcal{G}_1$, The best response v_i^* of player i in the game $\mathcal{G}_1$ in (7.33) is

$$v_i^* \in \mathrm{BR}_i^1(v_1, v_2, ..., v_{i-1}, v_{i+1}, ..., v_{m-1}, v_m) \tag{7.51}$$

$$= \underset{v \in \mathbb{R}_+}{\arg\min}\, I(X^n; Y_A^m) + \lambda D(P_{Y_A^m} \| P_{Y^m}) \tag{7.52}$$

$$= \underset{v \in \mathbb{R}_+}{\arg\min} (1-\lambda) \log \left| \mathbf{\Sigma}_{YY} + v\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \sum_{j \in \mathcal{K}\setminus\{i\}} v_j \mathbf{e}_j \mathbf{e}_j^\mathsf{T} \right| - \log(\sigma^2 + v) + \lambda v \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1} \mathbf{e}_i \mathbf{e}_i^\mathsf{T}\right), \tag{7.53}$$

where (7.53) follows from Lemma 19.

The following theorem provides the analytical solution for the best response of player i in $\mathcal{G}_1$.

Theorem 22. *Given the action profile $\mathbf{v} = (v_1, v_2, \ldots, v_m)^\mathsf{T}$, for all $i \in \{1, 2, \ldots, m\}$, the best response for the player i in $\mathcal{G}_1$ is*

$$v_i^* = \frac{-(\beta_i + \alpha_i\sigma^2\beta_i - \alpha_i) + \sqrt{(\beta_i + \alpha_i\sigma^2\beta_i - \alpha_i)^2 - 4\beta_i\alpha_i(\beta_i\sigma^2 - \alpha_i\sigma^2 + \frac{\alpha_i\sigma^2 - 1}{\lambda})}}{2\beta_i\alpha_i}, \tag{7.54}$$

where $\alpha_i = \mathrm{tr}\left(\left(\mathbf{\Sigma}_{YY} + \sum_{j \in \mathcal{K}\setminus\{i\}} v_j \mathbf{e}_j \mathbf{e}_j^\mathsf{T}\right)^{-1} \mathbf{e}_i \mathbf{e}_i^\mathsf{T}\right), \beta_i = \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1} \mathbf{e}_i \mathbf{e}_i^\mathsf{T}\right)$.

Proof. The proof of Theorem 22 is presented in Appendix M. □

Best Response in $\mathcal{G}_2$

In $\mathcal{G}_2$, the best response v_i^* of player i of the game $\mathcal{G}_2$ in (7.38) is

$$\begin{aligned} & v_i^* \in \mathrm{BR}_i^2\left(v_1, v_2, ..., v_{i-1}, v_{i+1}, ..., v_{m-1}, v_m\right) && (7.55)\\ =&\arg\min_{v\in\mathbb{R}_+} I(X^n; Y_{A,i}) + \lambda D(P_{Y_A^m} \| P_{Y^m}) && (7.56)\\ =&\arg\min_{v\in\mathbb{R}_+} \log\left(1+\frac{\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)}{\sigma^2+v}\right) - \lambda\log\left|\mathbf{\Sigma}_{YY} + v\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right| \\ & +\lambda v\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right), && (7.57)\end{aligned}$$

where (7.57) follows from Lemma 20.

The following theorem provides the analytical solution for the best response of player i in $\mathcal{G}_2$.

Theorem 23. *Given the action profile $\mathbf{v} = (v_1, v_2, \ldots, v_m)^\mathsf{T}$, for all $i \in \{1, 2, \ldots, m\}$, the best response for player i in $\mathcal{G}_2$ is v such that*

$$\frac{-\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)}{\left(\sigma^2+v\right)\left(\sigma^2+v+\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)\right)} - \lambda\frac{\alpha_i}{1+v\alpha_i} + \lambda\beta_i = 0, \tag{7.58}$$

where $\alpha_i = \mathrm{tr}\left(\left(\mathbf{\Sigma}_{YY} + \sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right)^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)$ and $\beta_i = \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)$.

Proof. The proof of Theorem 23 is presented in Appendix N. □

Best Response in $\mathcal{G}_3$

In $\mathcal{G}_3$, The best response v_i^* of the i-th player of the game in (7.44) is

$$\begin{aligned} & v_i^* \in \mathrm{BR}_i^3\left(v_1, v_2, ..., v_{i-1}, v_{i+1}, ..., v_{m-1}, v_m\right) \\ =&\arg\min_{v\in\mathbb{R}_+} I(X^n; Y_A^m) + \lambda D(P_{Y_{A,i}} \| P_{Y_i}) \\ =&\arg\min_{v\in\mathbb{R}_+} \log\left|\frac{1}{\sigma^2+v}\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\sum_{j\in\mathcal{K}\setminus\{i\}}\frac{1}{\sigma^2+v_j}\mathbf{e}_j\mathbf{e}_j^\mathsf{T} + \mathbf{I}_m\right| && (7.59)\\ & + \frac{\lambda}{\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)+\sigma^2}v - \lambda\log\left(\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)+\sigma^2+v\right).\end{aligned}$$

The following theorem provides the analytical solution for the best response of player i in $\mathcal{G}_3$.

Theorem 24. *Given the action profile $\mathbf{v} = (v_1, v_2, \ldots, v_m)^\mathsf{T}$, for all $i \in \{1, 2, \ldots, m\}$, the best response for player i in $\mathcal{G}_3$ is v such that*

$$-\frac{\gamma_i}{\left(\sigma^2+v\right)\left(\sigma^2+v+\alpha_i\right)} + \lambda\frac{v}{\left(\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)+\sigma^2+v\right)\left(\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)+\sigma^2\right)}, \tag{7.60}$$

where $\gamma_i \triangleq \mathrm{tr}\left(\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\sum_{j\in\mathcal{K}\setminus\{i\}}\frac{1}{\sigma^2+v_j}\mathbf{e}_j\mathbf{e}_j^\mathsf{T} + \mathbf{I}_m\right)^{-1}\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)$.

Proof. The proof of Theorem 24 is presented in Appendix O. □

7.4 Potential Games

7.4.1 Potential functions

Potential function of $\mathcal{G}_1$

The following proposition highlights an important property of the game $\mathcal{G}_1$ in (7.33).

Proposition 18. *The game $\mathcal{G}_1$ in (7.33) is a potential game.*

Proof. Let us define a function P_1: $\mathbb{R}_+^m \to \mathbb{R}$:

$$P_1(v_1, v_2, ..., v_m) \tag{7.61}$$
$$\triangleq I(X^n; Y_A^m) + \lambda D(Y_A^m \| Y^m) \tag{7.62}$$
$$=(1-\lambda)\log\left|\mathbf{\Sigma}_{YY} + \sum_{i=1}^m v_i \mathbf{e}_i \mathbf{e}_i^\mathsf{T}\right| - \sum_{i=1}^m \log(\sigma^2 + v_i) + \lambda \sum_{i=1}^m v_i \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1} \mathbf{e}_i \mathbf{e}_i^\mathsf{T}\right) \tag{7.63}$$

where (7.63) holds from plugging (7.35) and (7.36) into (7.62). For every attacker $i, i \in \{1, 2, ..., m\}$, and for every $(v_1, v_2, ..., v_{i-1}, v_{i+1}, ..., v_m)^\mathsf{T} \in \mathbb{R}_+^{m-1}$, it satisfies that for all $v_i \in \mathbb{R}_+, x_i \in \mathbb{R}_+$,

$$\phi_i^1(v_1, ..., v_{i-1}, v_i, v_{i+1}, ..., v_m) - \phi_i^1(v_1, ..., v_{i-1}, x_i, v_{i+1}, ..., v_m) < 0 \tag{7.64}$$

holds if and only if

$$P_1(v_1, ..., v_{i-1}, v_i, v_{i+1}, ..., v_m) - P_1(v_1, ..., v_{i-1}, x_i, v_{i+1}, ..., v_m) < 0 \tag{7.65}$$

holds. Therefore, $\mathcal{G}_1$ is an potential game. This completes the proof. □

The following lemma presents the potential function of the game $\mathcal{G}_1$ in (7.33).

Lemma 25. *The potential function of $\mathcal{G}_1$ in (7.33) is given by*

$$P_1(v_1, v_2, ..., v_m) \tag{7.66}$$
$$=(1-\lambda)\log\left|\mathbf{\Sigma}_{YY} + v_i \mathbf{e}_i \mathbf{e}_i^\mathsf{T} + \sum_{j \in \mathcal{K} \setminus \{i\}} v_j \mathbf{e}_j \mathbf{e}_j^\mathsf{T}\right| - \sum_{i=1}^m \log(\sigma^2 + v_i) + \lambda \sum_{i=1}^m v_i \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1} \mathbf{e}_i \mathbf{e}_i^\mathsf{T}\right)$$

Proof. For every attacker $i \in \{1, 2, ..., m\}$, and for every $(v_1, v_2, ..., v_{i-1}, v_{i+1}, ..., v_m)^\mathsf{T} \in \mathbb{R}_+^{m-1}$, for all $v_i \in \mathbb{R}_+, x_i \in \mathbb{R}_+$, the following holds

$$\begin{aligned} &\phi_i^1(v_1, ..., v_{i-1}, v_i, v_{i+1}, ..., v_m) - \phi_i^1(v_1, ..., v_{i-1}, x_i, v_{i+1}, ..., v_m) \\ =&P_1(v_1, ..., v_{i-1}, v_i, v_{i+1}, ..., v_m) - P_1(v_1, ..., v_{i-1}, x_i, v_{i+1}, ..., v_m). \end{aligned} \tag{7.67}$$

Therefore, P_1 is a potential function of game $\mathcal{G}_1$. This completes the proof. □

Potential function of $\mathcal{G}_2$

The following proposition highlights an important property of the game $\mathcal{G}_2$ in (7.38).

Proposition 19. *The game $\mathcal{G}_2$ in (7.38) is a potential game.*

Proof. Let us define a function P_2: $\mathbb{R}_+^m \to \mathbb{R}$:

$$\begin{aligned} P_2(v_1, v_2, ..., v_m) \triangleq & \lambda D(Y_A^m \| Y^m) + \sum_{i=1}^{m} I(X^n; Y_{A,i}) && (7.68) \\ = & \frac{1}{2}\lambda \left(\log \frac{|\mathbf{\Sigma}_{YY}|}{|\mathbf{\Sigma}_{YY} + \sum_{i=1}^m v_i \mathbf{e}_i \mathbf{e}_i^\mathsf{T}|} + \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1} \left(\sum_{i=1}^{m} v_i \mathbf{e}_i \mathbf{e}_i^\mathsf{T} \right) \right) \right) \\ & + \frac{1}{2} \sum_{i=1}^{m} \log \left(1 + \frac{\mathrm{tr}\left(\mathbf{H} \mathbf{\Sigma}_{XX} \mathbf{H}^\mathsf{T} \mathbf{e}_i \mathbf{e}_i^\mathsf{T} \right)}{\sigma^2 + v_i} \right), && (7.69) \end{aligned}$$

where (7.69) holds from plugging (7.40) and (7.41) into (7.68). For every attacker $i, i \in \{1, 2, ..., m\}$, and for every $(v_1, v_2, ..., v_{i-1}, v_{i+1}, ..., v_m)^\mathsf{T} \in \mathbb{R}_+^{m-1}$, for all $v_i \in \mathbb{R}_+, x_i \in \mathbb{R}_+$, it satisfies that

$$\phi_i^2(v_1, ..., v_{i-1}, v_i, v_{i+1}, ..., v_m) - \phi_i^2(v_1, ..., v_{i-1}, x_i, v_{i+1}, ..., v_m) < 0 \tag{7.70}$$

holds if and only if

$$P_2(v_1, ..., v_{i-1}, v_i, v_{i+1}, ..., v_m) - P_2(v_1, ..., v_{i-1}, x_i, v_{i+1}, ..., v_m) < 0 \tag{7.71}$$

holds. Therefore, $\mathcal{G}_2$ is an potential game. This completes the proof. □

The following lemma presents the potential function of the game $\mathcal{G}_2$ in (7.38).

Lemma 26. *The potential function of $\mathcal{G}_2$ in (7.38) is given by*

$$\begin{aligned} P_2(v_1, v_2, ..., v_m) = & \frac{1}{2}\lambda \left(\log \frac{|\mathbf{\Sigma}_{YY}|}{|\mathbf{\Sigma}_{YY} + \sum_{i=1}^m v_i \mathbf{e}_i \mathbf{e}_i^\mathsf{T}|} + \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1} \left(\sum_{i=1}^{m} v_i \mathbf{e}_i \mathbf{e}_i^\mathsf{T} \right) \right) \right) \\ & + \frac{1}{2} \sum_{i=1}^{m} \log \left(1 + \frac{\mathrm{tr}\left(\mathbf{H} \mathbf{\Sigma}_{XX} \mathbf{H}^\mathsf{T} \mathbf{e}_i \mathbf{e}_i^\mathsf{T} \right)}{\sigma^2 + v_i} \right). && (7.72) \end{aligned}$$

Proof. For every attacker $i \in \{1, 2, ..., m\}$, and for every $(v_1, v_2, ..., v_{i-1}, v_{i+1}, ..., v_m)^\mathsf{T} \in \mathbb{R}_+^{m-1}$, for all $v_i \in \mathbb{R}_+, x_i \in \mathbb{R}_+$, the following holds

$$\begin{aligned} & \phi_i^2(v_1, ..., v_{i-1}, v_i, v_{i+1}, ..., v_m) - \phi_i^2(v_1, ..., v_{i-1}, x_i, v_{i+1}, ..., v_m) && (7.73) \\ = & P_2(v_1, ..., v_{i-1}, v_i, v_{i+1}, ..., v_m) - P_2(v_1, ..., v_{i-1}, x_i, v_{i+1}, ..., v_m). \end{aligned}$$

Therefore, P_2 is a potential function of game $\mathcal{G}_2$. This completes the proof. □

Potential function of $\mathcal{G}_3$

The following proposition highlights an important property of the game $\mathcal{G}_3$ in (7.44).

Proposition 20. *The game $\mathcal{G}_3$ in (7.44) is a potential game.*

Proof. Let us define a function P_3: $\mathbb{R}_+^m \to \mathbb{R}$:

$$
\begin{aligned}
&P_3(v_1, v_2, ..., v_m) && (7.74)\\
\triangleq &I(X^n; Y_A^m) + \lambda \sum_{i=1}^{m} D(Y_{A,i} \| Y_i) && (7.75)\\
= &\frac{1}{2}\log \frac{|\mathbf{\Sigma}_{YY} + \sum_{i=1}^m v_i \mathbf{e}_i \mathbf{e}_i^\mathsf{T}|}{|\sigma^2 \mathbf{I}_m + \sum_{i=1}^m v_i \mathbf{e}_i \mathbf{e}_i^\mathsf{T}|} && (7.76)\\
&+ \frac{1}{2}\lambda \sum_{i=1}^{m} \left(\frac{v_i}{\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) + \sigma^2} + \log \frac{\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) + \sigma^2}{\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) + \sigma^2 + v_i} \right),
\end{aligned}
$$

where (7.76) holds from plugging (7.46) and (7.47) into (7.75). For all attacker i, $i \in \{1, 2, ..., m\}$ and for all $(v_1, v_2, ..., v_{i-1}, v_{i+1}, ..., v_m)^\mathsf{T} \in \mathbb{R}_+^{m-1}$, it satisfies that for all $v_i \in \mathbb{R}_+, x_i \in \mathbb{R}_+$,

$$
\phi_i^3(v_1, ..., v_{i-1}, v_i, v_{i+1}, ..., v_m) - \phi_i^3(v_1, ..., v_{i-1}, x_i, v_{i+1}, ..., v_m) > 0 \tag{7.77}
$$

holds if and only if

$$
P_3(v_1, ..., v_{i-1}, v_i, v_{i+1}, ..., v_m) - P_3(v_1, ..., v_{i-1}, x_i, v_{i+1}, ..., v_m) > 0 \tag{7.78}
$$

holds. Therefore, $\mathcal{G}_3$ is an potential game. This completes the proof. □

The following lemma presents the potential function of the game $\mathcal{G}_3$ in (7.44).

Lemma 27. *The potential function of $\mathcal{G}_3$ in (7.44) is given by*

$$
\begin{aligned}
P_3(v_1, v_2, ..., v_m) = &\frac{1}{2}\log \frac{|\mathbf{\Sigma}_{YY} + \sum_{i=1}^m v_i \mathbf{e}_i \mathbf{e}_i^\mathsf{T}|}{|\sigma^2 \boldsymbol{I}_m + \sum_{i=1}^m v_i \mathbf{e}_i \mathbf{e}_i^\mathsf{T}|} \qquad (7.79)\\
&+ \frac{1}{2}\lambda \sum_{i=1}^{m} \left(\frac{v_i}{\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) + \sigma^2} + \log \frac{\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) + \sigma^2}{\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) + \sigma^2 + v_i} \right).
\end{aligned}
$$

Proof. From all attacker $i, i \in \{1, 2, ..., m\}$ and for all $(v_1, v_2, ..., v_{i-1}, v_{i+1}, ..., v_m)^\mathsf{T} \in \mathbb{R}_+^{m-1}$, for all $v_i \in \mathbb{R}_+, x_i \in \mathbb{R}_+$, the following holds

$$
\begin{aligned}
&\phi_i^3(v_1, ..., v_{i-1}, v_i, v_{i+1}, ..., v_m) - \phi_i^3(v_1, ..., v_{i-1}, x_i, v_{i+1}, ..., v_m)\\
=&P_3(v_1, ..., v_{i-1}, v_i, v_{i+1}, ..., v_m) - P_3(v_1, ..., v_{i-1}, x_i, v_{i+1}, ..., v_m). \qquad (7.80)
\end{aligned}
$$

Therefore, P_3 is a potential function of game $\mathcal{G}_3$. This completes the proof. □

7.4.2 Nash Equilibriums (NEs)

From the best response described in the last section, a game solution that is particularly relevant for this analysis is the Nash Equilibrium.

Definition 12. *The action profile formed by all the attackers* $(v_1, v_2, ..., v_m)^\mathsf{T}$ *is an NE of the game* $\mathcal{G}_p$ *if and only if it is a solution of the fix point equation*

$$\sum_{i=1}^{m} v_i \mathbf{e}_i \mathbf{e}_i^\mathsf{T} = \mathrm{BR}^p \left(v_1, v_2, ...v_m\right), \tag{7.81}$$

with BR^p*:* $\mathbb{R}_+^m \to \mathbb{R}_+^m$ *being the global best response correspondence, i.e.,*

$$\begin{aligned} &\mathrm{BR}^p \left(v_1, v_2, ...v_m\right) \\ =&\mathrm{BR}_1^p \left(v_2, v_3, ..., v_m\right) \mathbf{e}_1 \mathbf{e}_1^\mathsf{T} + \mathrm{BR}_2^p \left(v_1, v_3, ..., v_m\right) \mathbf{e}_2 \mathbf{e}_2^\mathsf{T} + \cdots + \mathrm{BR}_m^p \left(v_1, v_2, ..., v_{m-1}\right) \mathbf{e}_m \mathbf{e}_m^\mathsf{T}. \end{aligned} \tag{7.82}$$

Essentially, at an NE, attackers achieve the minimal cost given the actions adopted by all the other attackers. This implies that an NE is an operating point where any deviation from the action at NE does not lead to a smaller cost.

7.4.3 Existence of the NE

The following proposition highlights an important property of the game $\mathcal{G}_1$ in (7.33).

Proposition 21. *The game* $\mathcal{G}_1$ *in* (7.33) *possesses only one NE.*

Proof. Note that P_1 in (7.66) is continuous over the set of all possible actions $v_i \in \mathbb{R}_+, i \in \{1, 2, \ldots, m\}$ and $\mathbb{R}_+$ is a convex set, therefore, there always exists a minimum of the potential function P_1 in $\mathbb{R}_+$. From Proposition 13, the potential function is convex, i.e., there is only one minimum of the potential function. From Lemma 4.3 in [103], it follows that such a minimum corresponds to an NE. Therefore, the game possesses only one NE. □

The following proposition highlights an important property of the game $\mathcal{G}_2$ in (7.38).

Proposition 22. *The game* $\mathcal{G}_2$ *in* (7.38) *possesses only one NE.*

Proof. Note that P_2 in (7.72) is continuous over the set of all possible actions $v_i \in \mathbb{R}_+, i \in \{1, 2, \ldots, m\}$ and $\mathbb{R}_+$ is a convex set, therefore, there always exists a minimum of the potential function P_2 in $\mathbb{R}_+$. From Proposition 15, the potential function is convex, i.e., there is only one minimum of the potential function. From Lemma 4.3 in [103], it follows that such a minimum corresponds to an NE. Therefore, the game possesses only one NE. □

The following proposition highlights an important property of the game $\mathcal{G}_3$ in (7.44).

Proposition 23. *The game* $\mathcal{G}_3$ *in* (7.44) *possesses only one NE.*

Proof. Note that P_3 in (7.79) is continuous over the set of all possible actions $v_i \in \mathbb{R}_+, i \in \{1, 2, \ldots, m\}$ and $\mathbb{R}_+$ is a convex set, therefore, there always exists a minimum of the potential function P_3 in $\mathbb{R}_+$. From Proposition 17, the potential function is convex, i.e., there is only one minimum of the potential function. From Lemma 4.3 in [103], it follows that such a minimum corresponds to an NE. Therefore, the game possesses only one NE. □

7.4.4 Achievability of the NE

The attackers are said to play a sequential best response dynamic (BRD) if the attackers can sequentially decide their own variance from their sets of best responses following a round-robin (increasing) order. Let us denote the choice of attacker i during round $t \in \mathbb{N}$ and assume that attackers are able to observa all the other attackers' decision. Under this assumption, the BRD is defined as follows.

Definition 13. *(Best Response Dynamics). The players of the game $\mathcal{G}_p$ are said to play a best response dynamics if there exists an round-robin order of the elements of K in which at each round $t \in \mathbb{N}$, the following holds*

$$v_i^*(t) = \mathrm{BR}_i\left(v_1^*(t), v_2^*(t), ..., v_{i-1}^*(t), v_{i+1}^*(t-1), ..., v_{m-1}^*(t-1), v_m^*(t-1)\right). \tag{7.83}$$

From the properties of potential games in [103, Lemma 4.2], the following Lemma follows.

Lemma 28. *(Achievability of NE attacks). Any BRD in the game $\mathcal{G}_p$ converges to a Gaussian attack construction that is the only NE in this game.*

The relevance of Lemma 28 is that it establishes that if attackers can play the game for at least a round-robin, they are always able to attack the network that minimizes the potential.

The proposed BRD in game $\mathcal{G}_1$, $\mathcal{G}_2$ and $\mathcal{G}_3$ are described in Algorithm 5, Algorithm 6 and Algorithm 7, respectively. As discussed in Theorem 22, Theorem 23 and Theorem 24, the optimal solution is unique and the achievability is discussed in Section 7.4.4. The computation complexity of all three Algorithms is $O(mt_{\max})$.

Algorithm 5 Best Response Dynamics for $\mathcal{G}_1$

Input: the observation matrix $\mathbf{H}$, the covariance matrix of the state variables $\mathbf{\Sigma}_{XX}$, the variance of the noise σ^2, and the weighting parameter λ.

Output: the action profile in the NE $(v_1, v_2, \ldots, v_m)^\mathsf{T}$.

Initialize the actions by all the players $v_i(t_0), i \in \{1, 2, ..., m\}$;

for $0 < t < t_{\max}$ **do**,

for $1 \leq i \leq m$ **do**,

Get $v_i^*(t)$ in Theorem 22, that is,

$$\frac{-(\beta_i + \alpha_i\sigma^2\beta_i - \alpha_i) + \sqrt{(\beta_i + \alpha_i\sigma^2\beta_i - \alpha_i)^2 - 4\beta_i\alpha_i(\beta_i\sigma^2 - \alpha_i\sigma^2 + \frac{\alpha_i\sigma^2 - 1}{\lambda})}}{2\beta_i\alpha_i},$$

where $\alpha_i = \mathrm{tr}\left(\left(\mathbf{\Sigma}_{YY} + \sum_{j\in\mathcal{K}, j<i} v_j(t)\mathbf{e}_j\mathbf{e}_j^\mathsf{T} + \sum_{j\in\mathcal{K}, j>i} v_j(t-1)\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right)^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)$ and $\beta_i = \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)$.

end for

$t = t + 1$

end for

Algorithm 6 Best Response Dynamics for $\mathcal{G}_2$

Input: the observation matrix $\mathbf{H}$, the covariance matrix of the state variables $\mathbf{\Sigma}_{XX}$, the variance of the noise σ^2, and the weighting parameter λ.

Output: the action profile in the NE $(v_1, v_2, \ldots, v_m)^\mathsf{T}$.

Initialize the actions by all the players $v_i(t_0), i \in \{1, 2, ..., m\}$;

for $0 < t < t_{\max}$ **do**,

for $1 \leq i \leq m$ **do**,

Get $v_i(t)$ in Theorem 23 such that

$$\frac{-\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)}{\left(\sigma^2 + v_i(t)\right)\left(\sigma^2 + v_i(t) + \mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)\right)} - \lambda\frac{\alpha_i}{1 + v_i(t)\alpha_i} + \lambda\beta_i = 0,$$

where $\alpha_i = \mathrm{tr}\left(\left(\mathbf{\Sigma}_{YY} + \sum_{j\in\mathcal{K}, j<i} v_j(t)\mathbf{e}_j\mathbf{e}_j^\mathsf{T} + \sum_{j\in\mathcal{K}, j>i} v_j(t-1)\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right)^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)$ and $\beta_i = \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)$.

end for

$t = t + 1$

end for

Algorithm 7 Best Response Dynamics for $\mathcal{G}_3$

Input: the observation matrix $\mathbf{H}$, the covariance matrix of the state variables $\mathbf{\Sigma}_{XX}$, the variance of the noise σ^2, and the weighting parameter λ.

Output: the action profile in the NE $(v_1, v_2, \ldots, v_m)^\mathsf{T}$.

Initialize the actions by all the players $v_i(t_0), i \in \{1, 2, ..., m\}$;

for $0 < t < t_{\max}$ **do**,

for $1 \leq i \leq m$ **do**,

Get $v_i(t)$ in Theorem 24 such that

$$-\frac{\gamma_i}{\left(\sigma^2 + v\right)\left(\sigma^2 + v + \alpha_i\right)} + \lambda\frac{v}{\left(\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) + \sigma^2 + v\right)\left(\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) + \sigma^2\right)},$$

with

$$\gamma_i = \mathrm{tr}\left(\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\sum_{j\in\mathcal{K}, j<i}\frac{1}{\sigma^2 + v_j(t)}\mathbf{e}_j\mathbf{e}_j^\mathsf{T} + \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\sum_{j\in\mathcal{K}, j>i}\frac{1}{\sigma^2 + v_j(t-1)}\mathbf{e}_j\mathbf{e}_j^\mathsf{T} + \mathbf{I}_m\right)^{-1}\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right).$$

end for

$t = t + 1$

end for

7.5 Numerical Results

The performance of the decentralized attack constructions in this section are verified on IEEE test systems. It is assumed that all the attackers can observe the actions taken by the other attacks. There are m attackers and m measurements in the system and each attacker has access to one unique measurement.

7.5.1 Game Convergence in terms of Potential Function

Fig. 7.1 depicts convergence of the potential function of $\mathcal{G}_1$ given by (7.66) in terms of round robin for different λ when $\rho = 0.9$, SNR = 30 dB in the IEEE 9-bus test system. The NEs with different λ are numerically evaluated and presented by red squares. In $\mathcal{G}_1$, the potential function in (7.66) is the same as the cost function of attacker i, with $i \in \{1, 2, \ldots, m\}$, in (7.34). From $t = 0$ to $t = 1$, all the attackers inject attacks following the best response in Theorem 22, that is, the variance of the random attack $v_i = 0$, $i \in \{1, 2, \ldots, m\}$ by attacker i would be modified to $v_i > 0$. The potential function from $t = 0$ to $t = 1$ decreases monotonically, which implies that all attackers benefit from the attacks launched by the other attackers in this round robin. Note that after all the attacker have injected an attack, that is, $t = 1$, and potential function is convex shown in Proposition 13. Hence, from $t = 1$, each attacker adjust its attacks that yields a decrease in the potential function until the NE in this game is achieved.

Similarly, Fig. 7.2 and Fig. 7.3 depict the potential function of $\mathcal{G}_2$ given by (7.72) and the potential function of $\mathcal{G}_3$ given by (7.79), respectively, in terms of round robin for different λ when $\rho = 0.9$, SNR = 30 dB in the IEEE 9-bus test system. Following the same round robin process as in $\mathcal{G}_1$, the attackers in $\mathcal{G}_2$ and $\mathcal{G}_3$ start to inject the attacks in the first round robin from $t = 0$ to $t = 1$ that yields the decrease in potential functions monotonically. The attackers adjust the variance of the attacks after $t = 1$ based on the best response in Theorem 23 and Theorem 24, respectively. With different objectives in $\mathcal{G}_1$, $\mathcal{G}_2$ and $\mathcal{G}_3$, the potential functions decrease differently but the monotonicity is guaranteed in all three games. Note that with larger λ, the value of the potential functions are larger in all three games. This implies that mutual information has a larger impact on the potential functions as λ is the weighting parameter between mutual information and KL divergence. The next section shows the game convergence in terms of the tradeoff between mutual information and KL divergence.

7.5.2 The Tradeoff between Mutual Information and KL Divergence

Fig. 7.4 depicts the tradeoff between mutual information and KL divergence in $\mathcal{G}_1$ for different λ when $\rho = 0.9$, SNR = 30 dB on the IEEE 9-bus system. As expected, from $t = 0$ to $t = 1$, the attackers start to compromise the measurements, which yields the decrease in mutual information and increase in KL divergence. From $t = 1$ to $t = 2$, the attackers adjust the variance of its attacks based on the best response in Theorem 22. Interestingly, at $t = 1$, the attackers achieve a smaller mutual information and larger KL divergence in comparison with in the following round robin. Hence, from $t = 1$ to $t = 2$, the attackers modify the

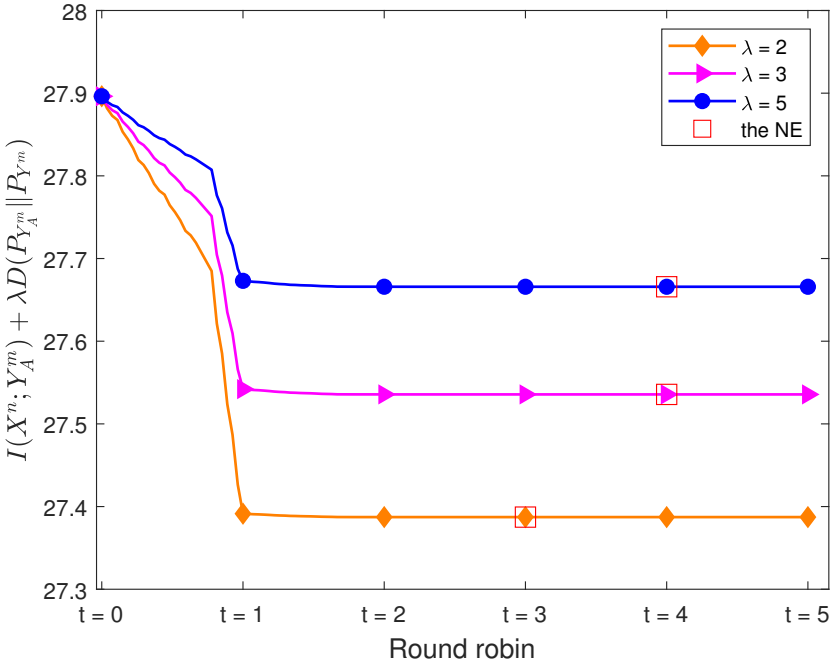

Figure 7.1: The convergence in $\mathcal{G}_1$ in terms of the potential function P_1 on the IEEE 9-bus test system when $\rho = 0.9$, SNR $= 30$ dB and the NEs with different λ.

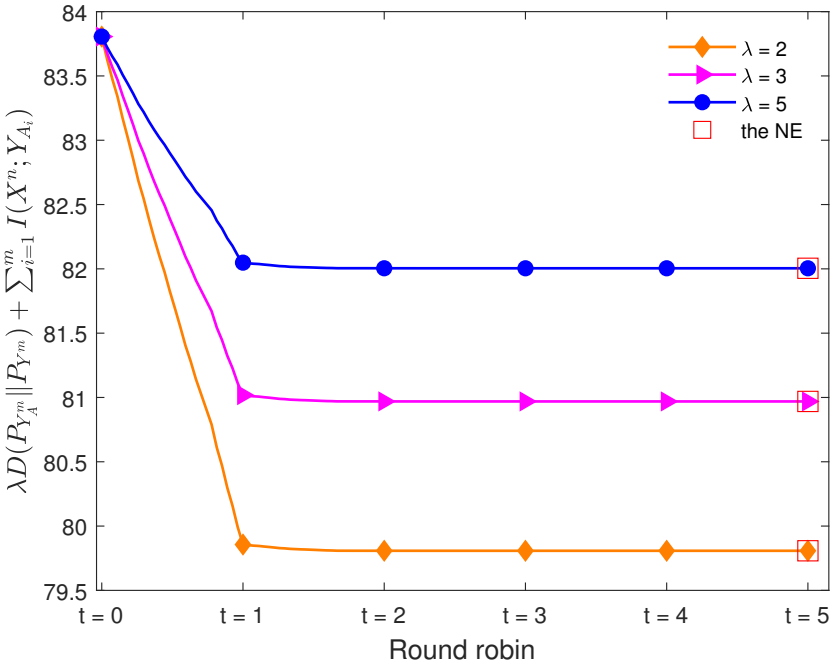

Figure 7.2: The convergence in $\mathcal{G}_2$ in terms of the potential function P_2 on the IEEE 9-bus test system when $\rho = 0.9$, SNR $= 30$ dB and the NEs with different λ.

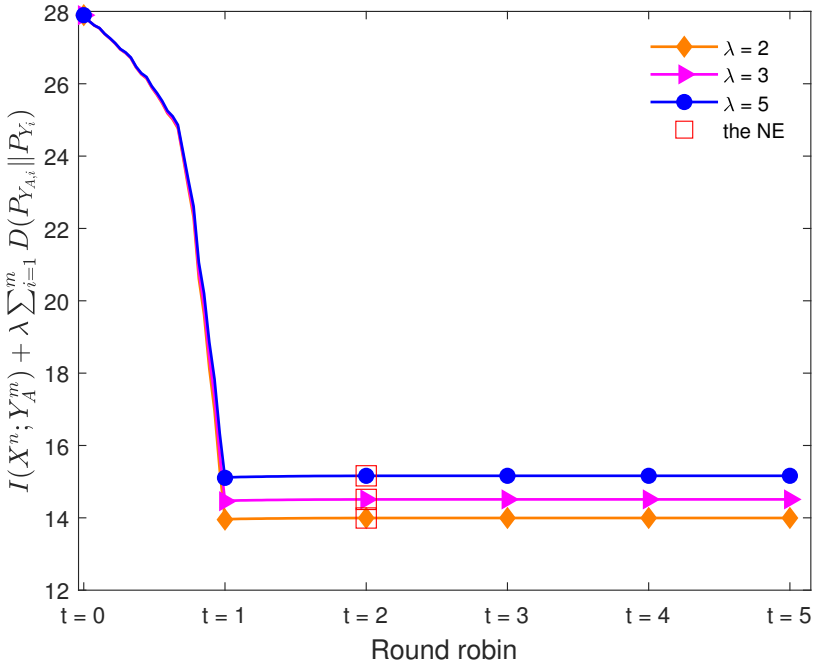

Figure 7.3: The convergence in $\mathcal{G}_3$ in terms of the potential function P_3 on the IEEE 9-bus test system when $\rho = 0.9$, SNR $= 30$ dB and the NEs with different λ.

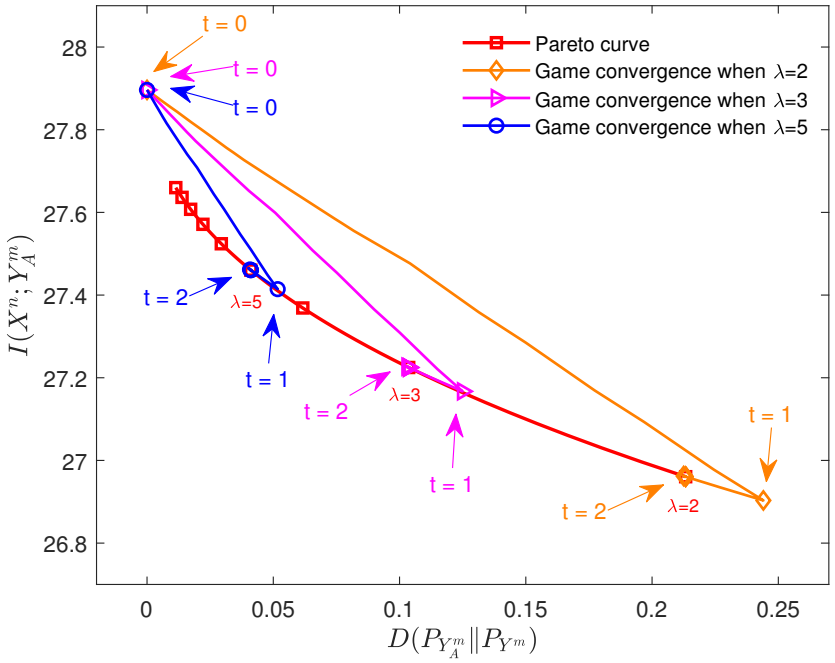

Figure 7.4: The tradeoff between mutual information and KL divergence in $\mathcal{G}_1$ on the IEEE 9-bus test system when $\rho = 0.9$, SNR $= 30$ dB.

attacks to constrain KL divergence. This is because in the first round robin from $t = 0$ to $t = 1$, the attack decisions were made when there were limited attackers that attacked the system. Therefore, the decisions from best response in Theorem 22 are more aggressive but still within the constraints of KL divergence. However, from $t = 1$ to $t = 2$, all the attackers have launched attacks, the result from best response in Theorem 22 limits the variance of the attacks which yields a decrease in KL divergence. After several round robins, the best response dynamics lead to the NE that is on the Pareto curve.

Fig. 7.5 and Fig. 7.6 depict the tradeoff between mutual information and KL divergence in $\mathcal{G}_2$ and $\mathcal{G}_3$ for different λ when $\rho = 0.9$, SNR $= 30$ dB on the IEEE 9-bus system, respectively. It is observed the same phenomenon as in $\mathcal{G}_1$ that in the end of the first round robin when $t = 1$, the attackers achieved lower mutual information at the expense of a large KL divergence. Hence, from $t = 1$ to $t = 2$ the best response dynamics yield to a decrease in KL divergence. Note that in $\mathcal{G}_1$ and $\mathcal{G}_2$, the convergence from $t = 0$ to $t = 1$ is linear that indicates the decrease of mutual information is at the expense of KL divergence linearly. Interestingly, in $\mathcal{G}_3$, it is observed that there is a significant decrease in mutual information in the end of the first round robin from $t = 0$ to $t = 1$. It is worth noting that the significant decrease comes from the attacking power injection measurements. This coincides with the results in chapter 6 that the power injection measurements are more vulnerable to data integrity attacks than power flow measurements [27].

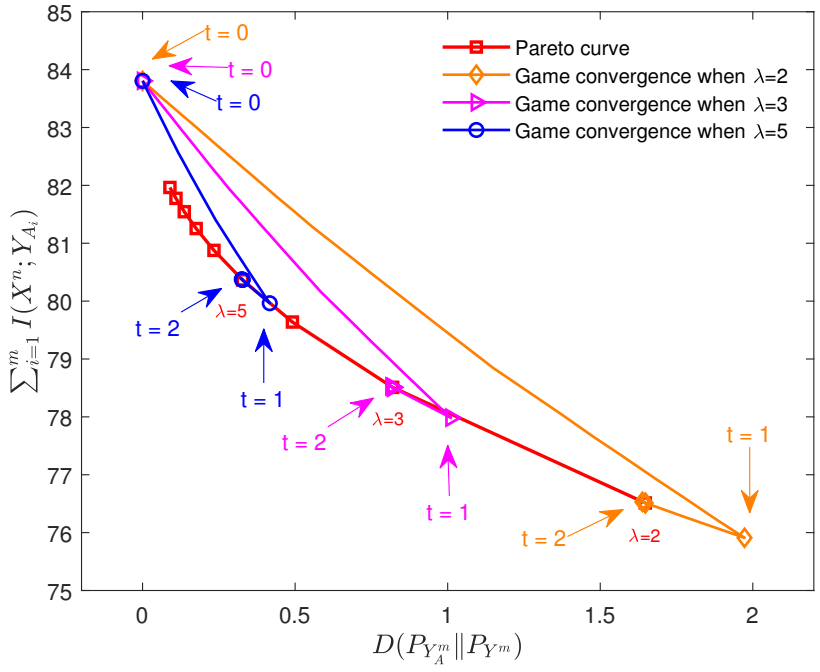

Figure 7.5: The tradeoff between mutual information and KL divergence in $\mathcal{G}_2$ on the IEEE 9-bus test system when $\rho = 0.9$, SNR $= 30$ dB.

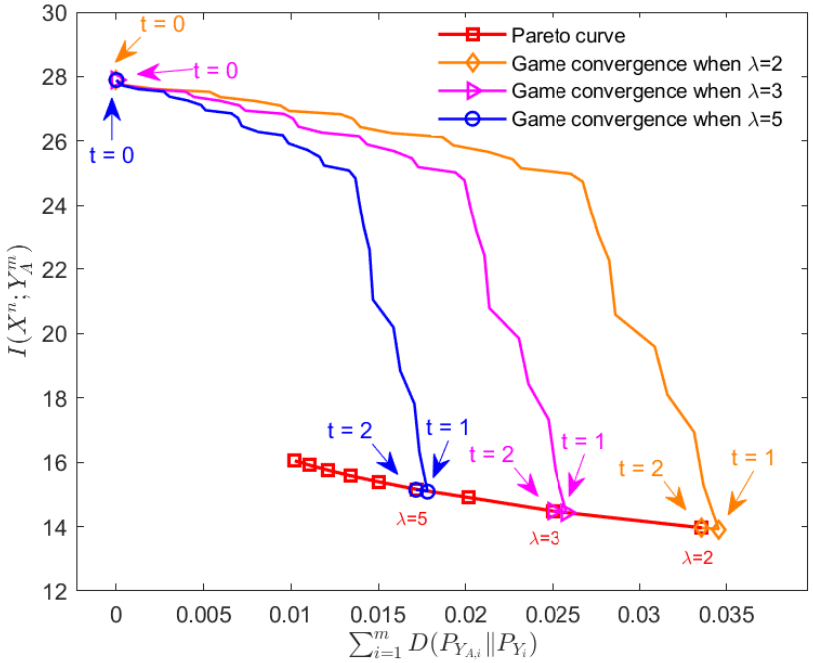

Figure 7.6: The tradeoff between mutual information and KL divergence in $\mathcal{G}_3$ on the IEEE 9-bus test system when $\rho = 0.9$, SNR $= 30$ dB.

7.6 Summary

This chapter has proposed a novel decentralized stealth attack constructions where game theoretic techniques are adopted in this framework. The objectives of the attack constructions are the disruption and detection both globally and locally that are measured by information theoretic metrics. The interaction between the attackers is utilized to formulate games in attack constructions that are motived by different information metrics. It is proved that the games are potential games as well as the existence and convexity of the potential functions. This chapter also characterizes the best response for the attackers and best response dynamics are proposed to achieve the NEs of the games, accordingly. The simulations have numerically evaluated the performance of the decentralized attacks on the IEEE test systems and shown the interaction between the attackers converge to the NEs on the Pareto curve.

Chapter 8

Decentralized Sparse Stealth Attacks

This chapter presents the main results on decentralized stealth attacks with sparsity constraints. The decentralized model and the game analysis are described in Section 3.6. Specifically, the sets of measurements that each attacker has access to form a partition of the set of measurements in the system. Every attacker minimizes the information theoretic cost of launching a random attack to one of the measurements that it has access to in a coordinated fashion. The decentralized sparse attacks with partition is modelled in a game form, which yields a potential game. characterized the potential function in this game and obtained the uniqueness and achievability of the Nash Equilibrium in this game. The best response dynamics is proposed to achieve the unique NE in the game. The performance of the decentralized sparse attack constructions is evaluated on IEEE test systems. It is observed that the game achieves better performance with smaller λ and larger k.

8.1 System Model

Let $\mathcal{M} \triangleq \{1, 2, \ldots, m\}$ be the set of measurements on the power system and the set

$$\mathcal{K} \triangleq \{1, 2, \ldots, k\}, \tag{8.1}$$

where $k < m$ be the set of attackers that can perform a random data injection attacks to the system.

For all $j \in \mathcal{K}$, the set of sensors that the attacker j has access to is $\mathcal{M}_j, \mathcal{M}_j \subseteq \mathcal{M}$. The sets $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_k$ form a partition of $\mathcal{M}$, that is,

$$\mathcal{M} = \bigcup_{j=1}^{k} \mathcal{M}_j, \tag{8.2}$$

and for all $(i, j) \in \mathcal{K}$, with $i \neq j$,

$$\mathcal{M}_i \cap \mathcal{M}_j = \emptyset. \tag{8.3}$$

This chapter assumes that each attacker only compromises one sensor. For all $j \in \mathcal{K}$, the index of the measurement that the attacker j compromises is denoted by s_j. That is, for all

$j \in \mathcal{K}$, $s_j \in \mathcal{M}_j$ with $\mathcal{M}_j$ in (8.2). The data injection attacks of the attacker j is a random variable that follows a Gaussian distribution with zero mean, that is,

$$A_{s_j} \sim \mathcal{N}(0, v_j), \tag{8.4}$$

with $v_j \in \mathbb{R}_+$. In this setting, the action of the attacker j is to choose an index $s_j \in \mathcal{M}_j$ in (8.2) and v_j in (8.4). The action of the attacker j is given by

$$(s_j, v_j) \in \mathcal{A}_j, \tag{8.5}$$

where

$$\mathcal{A}_j \triangleq \mathcal{M}_j \times \mathbb{R}_+. \tag{8.6}$$

The tuple

$$\Big((s_1, v_1), (s_2, v_2), ..., (s_k, v_k) \Big) \in \mathcal{A}_1 \times \mathcal{A}_2 \times ... \times \mathcal{A}_k \tag{8.7}$$

is referred to as an *action profile* where for all $j \in \mathcal{K}$, $s_j \in \mathcal{M}_j$ and $v_j \in \mathbb{R}_+$.

Note that the action profile forms the covariance matrix of the random attack vector denoted by $\mathbf{\Sigma}_{AA}$ in (4.8), that is,

$$\mathbf{\Sigma}_{AA} = \sum_{j \in \mathcal{K}} v_j \mathbf{e}_{s_j} \mathbf{e}_{s_j}^{\mathsf{T}}. \tag{8.8}$$

8.2 Game Formulation

The cost for the attacker j to launch an attack does not only depend on its own sensor index s_j and variance v_j that is added in the corresponding sensor, but also the indices and variances by all the other attackers. Therefore, the interaction of all attackers in the system is described by a game in normal form

$$\mathcal{G} = (\mathcal{K}, \{\mathcal{A}_j\}_{j \in \mathcal{K}}, \{\psi_j\}_{j \in \mathcal{K}}), \tag{8.9}$$

where the set $\mathcal{K}$ is in (8.1), the set $\mathcal{A}_j$ is in (8.6), for all $j \in \mathcal{K}$ the function ψ_j is the cost function for attacker j that is defined in the following.

Definition 14. *Given the action profile* $\Big((s_1, v_1), (s_2, v_2), ..., (s_k, v_k) \Big) \in \mathcal{A}_1 \times \mathcal{A}_2 \times ... \times \mathcal{A}_k$, *the cost function of the attacker* j *denoted by* $\psi_j : \mathcal{A}_1 \times \mathcal{A}_2 \times ... \times \mathcal{A}_k \to \mathbb{R}_+$ *defined by adding* (4.18) *and* (4.21) *is*

$$\psi_j \Big((s_1, v_1), (s_2, v_2), ..., (s_k, v_k) \Big) \tag{8.10}$$

$$\triangleq I(X^n; Y_A^m) + \lambda D(P_{Y_A^m} \| P_{Y^m}) \tag{8.11}$$

$$= \frac{1}{2}(1-\lambda) \log \left| \mathbf{\Sigma}_{YY} + \sum_{j \in \mathcal{K}} v_j \mathbf{e}_{s_j} \mathbf{e}_{s_j}^{\mathsf{T}} \right| - \frac{1}{2} \log \left| \sum_{j \in \mathcal{K}} v_j \mathbf{e}_{s_j} \mathbf{e}_{s_j}^{\mathsf{T}} + \sigma^2 \boldsymbol{I}_m \right|$$
$$+ \frac{1}{2} \lambda \left(\mathrm{tr} \left(\mathbf{\Sigma}_{YY}^{-1} \left(\sum_{j \in \mathcal{K}} v_j \mathbf{e}_{s_j} \mathbf{e}_{s_j}^{\mathsf{T}} \right) \right) + \log |\mathbf{\Sigma}_{YY}| \right). \tag{8.12}$$

8.2.1 Best Response

The cost of the attacks $\psi_j\Big((s_1,v_1),(s_2,v_2),...,(s_k,v_k) \Big)$ for attacker j not only depends on its own action (s_j,v_j), but also on the actions by all the other attackers. For all $j \in \mathcal{K}$, given the actions by all the other attackers $\Big((s_1,v_1),(s_2,v_2),...,(s_k,v_k) \Big) \setminus (s_j,v_j)$, the best response for attacher j is given by

$$\left(s_j^*, v_j^*\right) \in \mathrm{BR}_j\left((s_1,v_1),(s_2,v_2),...,(s_{j-1},v_{j-1}),(s_{j+1},v_{j+1}),\ldots,(s_{k-1},v_{k-1}),(s_k,v_k)\right), \tag{8.13}$$

where the correspondence $\mathrm{BR}_j : \mathcal{A}_1 \times \mathcal{A}_2, \ldots \mathcal{A}_{j-1} \times \mathcal{A}_{j+1} \ldots \mathcal{A}_{k-1} \times \mathcal{A}_k \to \mathbb{R}_+$ is the best response correspondence, that is,

$$\mathrm{BR}_j\left((s_1,v_1),(s_2,v_2),...,(s_{j-1},v_{j-1}),(s_{j+1},v_{j+1}),\ldots,(s_{k-1},v_{k-1}),(s_k,v_k)\right) \tag{8.14}$$

$$= \underset{(s_j,v_j)\in\mathcal{A}_j}{\arg\min}\ \psi_j\Big((s_1,v_1),(s_2,v_2),...,(s_k,v_k) \Big) \tag{8.15}$$

The following lemma presents the equivalent optimization problem to minimize the cost for attacker j in (8.10).

Lemma 29. *For attacker $j, j \in \mathcal{K}$, minimizing the cost function ψ_j in* (8.10) *is equivalent to*

$$\underset{(s_j,v_j)\in\mathcal{A}_j}{\arg\min}\,(1-\lambda)\log\left|\mathbf{\Sigma}_{YY}+\sum_{j\in\mathcal{K}} v_j\mathbf{e}_{s_j}\mathbf{e}_{s_j}^{\mathsf{T}}\right| - \log\left|\sum_{j\in\mathcal{K}} v_j\mathbf{e}_{s_j}\mathbf{e}_{s_j}^{\mathsf{T}}+\sigma^2\boldsymbol{I}_m\right| + \lambda v_j \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_{s_j}\mathbf{e}_{s_j}^{\mathsf{T}}\right). \tag{8.16}$$

Proof. The proof follows by removing the constants that is not a function of (s_j, v_j). □

The following proposition provides the convexity of the cost function in (8.16).

Proposition 24. *The cost function in* (8.16) *is convex.*

Proof. From [115, Sec. 3.1.5] and the fact that $\mathbf{\Sigma}_{YY}+\sum_{j\in\mathcal{K}} v_j\mathbf{e}_{s_j}\mathbf{e}_{s_j}^{\mathsf{T}} \in \mathbf{S}_{++}^m$ and $\sum_{j\in\mathcal{K}} v_j\mathbf{e}_{s_j}\mathbf{e}_{s_j}^{\mathsf{T}}+\sigma^2\mathbf{I}_m \in \mathbf{S}_{++}^m$, the terms $\log\left|\mathbf{\Sigma}_{YY}+\sum_{j\in\mathcal{K}} v_j\mathbf{e}_{s_j}\mathbf{e}_{s_j}^{\mathsf{T}}\right|$ and $\log\left|\sum_{j\in\mathcal{K}} v_j\mathbf{e}_{s_j}\mathbf{e}_{s_j}^{\mathsf{T}}+\sigma^2\mathbf{I}_m\right|$ are concave. Therefore, the terms

$$(1-\lambda)\log\left|\mathbf{\Sigma}_{YY}+\sum_{j\in\mathcal{K}} v_j\mathbf{e}_{s_j}\mathbf{e}_{s_j}^{\mathsf{T}}\right|, \tag{8.17}$$

and $-\log\left|\sum_{j\in\mathcal{K}} v_j\mathbf{e}_{s_j}\mathbf{e}_{s_j}^{\mathsf{T}}+\sigma^2\mathbf{I}_m\right|$ are convex. Given that the trace is a linear operator, the term $\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_{s_j}\mathbf{e}_{s_j}^{\mathsf{T}}\right)$ is linear with respect to v_j. From [115, Sec. 3.2.1], the non-negative weighted sums of convex functions is convex, that is, the cost function in (8.16) is convex. This completes the proof. □

The following theorem proposes the analytical expression of the best response.

Theorem 30. *Let $\lambda \geq 1$. In the game $\mathcal{G} = (\mathcal{K}, \{\mathcal{A}_j\}_{j\in\mathcal{K}}, \{\psi_j\}_{j\in\mathcal{K}})$ in (8.9), for all $j \in \mathcal{K}$ and $s_j \in \mathcal{M}_j$, given a action profile by all the attackers except attacker j, that is,*

$$\begin{aligned}&\Big((i_1, v_1), (i_2, v_2), ..., (i_{j-1}, v_{j-1}), (i_{j+1}, v_{j+1}), ..., (s_{k-1}, v_{k-1}), (s_k, v_k) \Big) \\ &\in \mathcal{A}_1 \times \mathcal{A}_2 \times ... \times \mathcal{A}_{j-1} \times \mathcal{A}_{j+1} \times ... \times \mathcal{A}_{k-1} \times \mathcal{A}_k,\end{aligned} \tag{8.18}$$

the following holds

$$s_j^* = \arg\min_{q\in\mathcal{M}_j} (1-\lambda)\log\left|\mathbf{\Sigma}_{YY} + v_q^*\mathbf{e}_q\mathbf{e}_q^\mathsf{T} + \sum_{p\in\mathcal{K}\setminus\{j\}} v_p\mathbf{e}_{s_p}\mathbf{e}_{s_p}^\mathsf{T}\right| - \log\left(\sigma^2 + v_q^*\right) + \lambda v_q^* \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_q\mathbf{e}_q^\mathsf{T}\right) \tag{8.19}$$

$$v_q^* = \frac{-(\beta + \alpha\sigma^2\beta - \alpha) + \sqrt{(\beta + \alpha\sigma^2\beta - \alpha)^2 - 4\beta\alpha(\beta\sigma^2 - \alpha\sigma^2 + \frac{\alpha\sigma^2 - 1}{\lambda})}}{2\beta\alpha}, \tag{8.20}$$

where $\alpha \triangleq \mathrm{tr}\left(\left(\mathbf{\Sigma}_{YY} + \sum_{p\in\mathcal{K}\setminus\{j\}} v_p\mathbf{e}_{s_p}\mathbf{e}_{s_p}^\mathsf{T}\right)^{-1}\mathbf{e}_q\mathbf{e}_q^\mathsf{T}\right), \beta \triangleq \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_q\mathbf{e}_q^\mathsf{T}\right)$, $q \in \mathcal{M}_j$.

Proof. From Lemma 29, given the action by all the other players in (8.18), for all $j \in \mathcal{K}$, the best response is given by

$$(s_j, v_j) = \arg\min_{(i,v)\in\mathcal{A}_j} (1-\lambda)\log\left|\mathbf{\Sigma}_{YY} + v\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \sum_{p\in\mathcal{K}\setminus\{j\}} v_p\mathbf{e}_{s_p}\mathbf{e}_{s_p}^\mathsf{T}\right| - \log\left(\sigma^2 + v\right) + \lambda v \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right). \tag{8.21}$$

The optimization problem in (8.21) is brokendown as follows:

$$\min_{i\in\mathcal{M}_j}\min_{v\in\mathbb{R}_+} (1-\lambda)\log\left|\mathbf{\Sigma}_{YY} + v\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \sum_{p\in\mathcal{K}\setminus\{j\}} v_p\mathbf{e}_{s_p}\mathbf{e}_{s_p}^\mathsf{T}\right| - \log\left(\sigma^2 + v\right) + \lambda v \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right). \tag{8.22}$$

The inner optimization problem is convex. Therefore, the best response of v given a fixed index i is

$$v_i^* = \frac{-(\beta + \alpha\sigma^2\beta - \alpha) + \sqrt{(\beta + \alpha\sigma^2\beta - \alpha)^2 - 4\beta\alpha(\beta\sigma^2 - \alpha\sigma^2 + \frac{\alpha\sigma^2 - 1}{\lambda})}}{2\beta\alpha}, \tag{8.23}$$

where $\alpha \triangleq \mathrm{tr}\left(\left(\mathbf{\Sigma}_{YY} + \sum_{p\in\mathcal{K}\setminus\{j\}} v_p\mathbf{e}_{s_p}\mathbf{e}_{s_p}^\mathsf{T}\right)^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right), \beta \triangleq \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)$. Then, the outer optimization problem in (8.22) is equivalent to the following:

$$s_j^* = \arg\min_{i\in\mathcal{M}_j} (1-\lambda)\log\left|\mathbf{\Sigma}_{YY} + v_i^*\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \sum_{p\in\mathcal{K}\setminus\{j\}} v_p\mathbf{e}_{s_p}\mathbf{e}_{s_p}^\mathsf{T}\right| - \log\left(\sigma^2 + v_i^*\right) + \lambda v_i^* \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right), \tag{8.24}$$

where v_i^* is given by (8.23). i^* is obtained by searching the index over $\mathcal{M}_j$, which completes the proof. □

8.2.2 Potential Game

The following proposition highlights an important property of the game $\mathcal{G}$ in (8.9).

Proposition 25. *The game $\mathcal{G}$ in* (8.9) *is an ordinal potential game.*

Proof. Let us define a function $P : \mathcal{A}_1 \times \mathcal{A}_2 \times ... \times \mathcal{A}_k \to \mathbb{R}_+$:

$$P\Big((s_1, v_1), (s_2, v_2), ..., (s_k, v_k) \Big) \tag{8.25}$$

$$= \frac{1}{2}(1-\lambda)\log \left| \mathbf{\Sigma}_{YY} + \sum_{j \in \mathcal{K}} v_j \mathbf{e}_{s_j} \mathbf{e}_{s_j}^{\mathsf{T}} \right| - \frac{1}{2}\log \left| \sum_{j \in \mathcal{K}} v_j \mathbf{e}_{s_j} \mathbf{e}_{s_j}^{\mathsf{T}} + \sigma^2 \mathbf{I}_m \right|$$

$$+ \frac{1}{2}\lambda \left(\mathrm{tr} \left(\mathbf{\Sigma}_{YY}^{-1} \left(\sum_{j \in \mathcal{K}} v_j \mathbf{e}_{s_j} \mathbf{e}_{s_j}^{\mathsf{T}} \right) \right) + \log |\mathbf{\Sigma}_{YY}| \right). \tag{8.26}$$

For every attacker j, $j \in \mathcal{K}$, and for every action profile by all the attackers except attacker j, that is,

$$\Big((i_1, v_1), (i_2, v_2), ..., (i_{j-1}, v_{j-1}), (i_{j+1}, v_{j+1}), ..., (s_{k-1}, v_{k-1}), (s_k, v_k) \Big) \tag{8.27}$$
$$\in \mathcal{A}_1 \times \mathcal{A}_2 \times ... \times \mathcal{A}_{j-1} \times \mathcal{A}_{j+1} \times ... \times \mathcal{A}_{k-1} \times \mathcal{A}_k,$$

for all $(s_j, v_j) \in \mathcal{A}_j$ and $(x, y) \in \mathcal{A}_j$, the following inequation holds

$$\psi_j \Big((s_1, v_1), ..., (s_{j-1}, v_{j-1}), (s_j, v_j), (s_{j+1}, v_{j+1}), ..., (s_k, v_k) \Big) \tag{8.28}$$
$$- \psi_j \Big((s_1, v_1), ..., (s_{j-1}, v_{j-1}), (x, y), (s_{j+1}, v_{j+1}), ..., (s_k, v_k) \Big) < 0$$

if and only if

$$P\Big((s_1, v_1), ..., (s_{j-1}, v_{j-1}), (s_j, v_j), (s_{j+1}, v_{j+1}), ..., (s_k, v_k) \Big) \tag{8.29}$$
$$- P\Big((s_1, v_1), ..., (s_{j-1}, v_{j-1}), (x, y), (s_{j+1}, v_{j+1}), ..., (s_k, v_k) \Big) < 0$$

holds, where ψ_j is the cost function of the attacker j defined in (8.10). This completes the proof. □

The following lemma presents the potential function of the game $\mathcal{G}$ in (8.9).

Proposition 26. *The potential function of* $\mathcal{G}$ *in* (8.9) *is given by*

$$P\Big(\,(s_1,v_1),(s_2,v_2),...,(s_k,v_k)\,\Big) \tag{8.30}$$

$$\begin{aligned}=&\frac{1}{2}(1-\lambda)\log\left|\boldsymbol{\Sigma}_{YY}+\sum_{j\in\mathcal{K}}v_j\mathbf{e}_{s_j}\mathbf{e}_{s_j}^{\mathsf{T}}\right|-\frac{1}{2}\log\left|\sum_{j\in\mathcal{K}}v_j\mathbf{e}_{s_j}\mathbf{e}_{s_j}^{\mathsf{T}}+\sigma^2\boldsymbol{I}_m\right|\\&+\frac{1}{2}\lambda\left(\mathrm{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\left(\sum_{j\in\mathcal{K}}v_j\mathbf{e}_{s_j}\mathbf{e}_{s_j}^{\mathsf{T}}\right)\right)+\log|\boldsymbol{\Sigma}_{YY}|\right),\end{aligned} \tag{8.31}$$

where $\Big(\,(s_1,v_1),(s_2,v_2),...,(s_k,v_k)\,\Big)\in\mathcal{A}_1\times\mathcal{A}_2\times...\times\mathcal{A}_k$ *in* (8.7).

Proof. For every attacker j, $j\in\mathcal{K}$, and for every action profile by all the attackers except attacker j, that is,

$$\begin{aligned}&\Big(\,(i_1,v_1),(i_2,v_2),...,(i_{j-1},v_{j-1}),(i_{j+1},v_{j+1}),...,(s_{k-1},v_{k-1}),(s_k,v_k)\,\Big)\\&\in\mathcal{A}_1\times\mathcal{A}_2\times...\times\mathcal{A}_{j-1}\times\mathcal{A}_{j+1}\times...\times\mathcal{A}_{k-1}\times\mathcal{A}_k,\end{aligned} \tag{8.32}$$

for all $(s_j,v_j)\in\mathcal{A}_j$ and $(x,y)\in\mathcal{A}_j$, the following holds

$$\begin{aligned}&\psi_j\Big(\,(s_1,v_1),...,(s_{j-1},v_{j-1}),(s_j,v_j),(s_{j+1},v_{j+1}),...,(s_k,v_k)\,\Big)\\&-\psi_j\Big(\,(s_1,v_1),...,(s_{j-1},v_{j-1}),(x,y),(s_{j+1},v_{j+1}),...,(s_k,v_k)\,\Big)\end{aligned} \tag{8.33}$$

$$\begin{aligned}=&P\Big(\,(s_1,v_1),...,(s_{j-1},v_{j-1}),(s_j,v_j),(s_{j+1},v_{j+1}),...,(s_k,v_k)\,\Big)\\&-P\Big(\,(s_1,v_1),...,(s_{j-1},v_{j-1}),(x,y),(s_{j+1},v_{j+1}),...,(s_k,v_k)\,\Big),\end{aligned} \tag{8.34}$$

where ψ_j is the cost function of the attacker j defined in (8.10). This completes the proof. □

8.2.3 Existence and Achievability of the NE

The following lemma provides the existence of Nash Equilibrium.

Lemma 31. *The game* $\mathcal{G}$ *in* (7.33) *has always at least one NE in pure strategies.*

Proof. The Lemma holds from Propostion 25, that is, every finite ordinal potential game possesses a pure-strategy equilibrium [103, Cor. 2.2]. □

The attacker are said to play a sequential best response dynamic (BRD) if the attackers can sequentially decide their own variance from their sets of best responses following a round-robin (increasing) order. Let us denote the choice of attacker j during round $t\in\mathbb{N}$ and assume that attackers are able to observe all the other attackers' decision. Under this assumption, the BRD is defined as follows.

Definition 15. *(Best Response Dynamics). The players of the game $\mathcal{G}$ are said to play a best response dynamics if there exists an round-robin order of the elements of K in which at each round $t \in \mathbb{N}$, the following holds*

$$\left(s_j^*(t), v_j^*(t)\right) = \mathrm{BR}_j\left(\left(s_1^*(t), v_1^*(t)\right), \left(s_2^*(t), v_2^*(t)\right), ..., \left(s_{j-1}^*(t), v_{j-1}^*(t)\right), \left(s_{j+1}^*(t-1), v_{j+1}^*(t-1)\right), \ldots, \left(s_{k-1}^*(t-1), v_{k-1}^*(t-1)\right), \left(s_k^*(t-1), v_k^*(t-1)\right)\right) \tag{8.35}$$

From the properties of potential games in [103, Lemma 4.2], the following lemma follows.

Lemma 32. *(Achievability of NE attacks). Any BRD in the game $\mathcal{G}$ converges to a Gaussian attack construction that is the NE.*

The proposed BRD in game $\mathcal{G}$ to achieve the NE is described in Algorithm 8. As in Theorem 30, the optimization problem is brokendown as a convex problem. The algorithm converges quickly and the computation complexity is $O(t_{\max}k)$.

Algorithm 8 Best Response Dynamics for $\mathcal{G}$

Input: the observation matrix $\mathbf{H}$, the covariance matrix of the state variables $\mathbf{\Sigma}_{XX}$, the variance of the noise σ^2, and the weighting parameter λ.

Output: the action profile in the NE $(v_1, v_2, \ldots, v_m)^\mathsf{T}$.

Initialize the actions by all the players $v_i(t_0), i \in \{1, 2, ..., k\}$;

for $0 < t < t_{\max}$ **do**,

for $1 \le i \le k$ **do**,

Get the sensor selection s_j^* and variance v_q^* in Theorem 30, that is,

$$s_j^* = \arg\min_{q \in \mathcal{M}_j} (1-\lambda)\log\left|\mathbf{\Sigma}_{YY} + v_q^* \mathbf{e}_q \mathbf{e}_q^\mathsf{T} + \sum_{p \in \mathcal{K}\setminus\{j\}} v_p \mathbf{e}_{s_p} \mathbf{e}_{s_p}^\mathsf{T}\right| - \log\left(\sigma^2 + v_q^*\right) + \lambda v_q^* \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1} \mathbf{e}_q \mathbf{e}_q^\mathsf{T}\right),$$

$$v_q^* = \frac{-(\beta + \alpha\sigma^2\beta - \alpha) + \sqrt{(\beta + \alpha\sigma^2\beta - \alpha)^2 - 4\beta\alpha(\beta\sigma^2 - \alpha\sigma^2 + \frac{\alpha\sigma^2 - 1}{\lambda})}}{2\beta\alpha},$$

where $\alpha \triangleq \mathrm{tr}\left(\left(\mathbf{\Sigma}_{YY} + \sum_{p \in \mathcal{K}\setminus\{j\}} v_p \mathbf{e}_{s_p} \mathbf{e}_{s_p}^\mathsf{T}\right)^{-1} \mathbf{e}_q \mathbf{e}_q^\mathsf{T}\right), \beta \triangleq \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1} \mathbf{e}_q \mathbf{e}_q^\mathsf{T}\right), q \in \mathcal{M}_j$.

end for

$t = t + 1$

end for

8.3 Numerical Results

The numerical results of decentralized attack constructions with sparsity constraints in this section are obtained where the sets of the measurements that attackers have access to form

a proper partition of the set of the measurements in the system. Each attacker compromises one measurement that is accessible. The simulations assume that all the attackers can see the actions taken by the other attacks.

8.3.1 Game Convergence with Different Weighting Parameter λ

Fig. 8.1 to Fig. 8.4 depict the performance of decentralized attack constructions by Algorithm 8 with different λ when $k = 2$, $\rho = 0.9$ and SNR = 30 dB on the IEEE 9-bus test system. With $k = 2$, there are two attackers in this game. The simulations in this section assume attacker 1 has access to the set of power flow measurement with indices from 1 to 22 and attacker 2 has access to the set of power injection measurement with indices from 23 to m. These two sets form a proper partition of the set of the measurements on the system.

Specifically, Fig. 8.1 depicts the convergence of the potential function of $\mathcal{G}$ given by (8.31) in terms of round robin with different λ when $k = 2$, $\rho = 0.9$ and SNR = 30 dB on the IEEE 9-bus test system. From $t = 0$ to $t = 1$, attackers launched the attacks obtained from the best response in Theorem 30, that is, attackers determined the variance of the attacks for all the measurements from (8.20) and obtained the best sensor selection according to (8.19). Hence, from $t = 0$ to $t = 1$, the potential function P decreases monotonically, which implies that all attackers benefit from the attacks launched by the other attackers. Note that after all the attackers have injected an attack, from $t = 1$ to $t = 2$, each attacker modifies its attacks based on the best response in Theorem 30 after observing the attacks by the other attackers. The best response yields a decrease in the value of the potential function until the NE in this game is achieved. Note that in Fig. 8.1, the game with smaller λ achieve a lower value of potential function. This implies emphasizing on the mutual information benefit the potential function of the game.

Fig. 8.2 depicts the tradeoff between mutual information and KL divergence in terms of round robin with different λ when $k = 2$, $\rho = 0.9$ and SNR = 30 dB on the IEEE 9-bus test system. Similarly, the game starts when $t = 0$ where there are no attacks to the system. Hence, the KL divergence is 0. As expected, from $t = 0$ to $t = 1$, the attackers inject attacks to the system which yields a decrease in mutual information and increase in KL divergence and the game with larger λ results in smaller KL divergence and larger mutual information. However, with larger λ, attackers reduce KL divergence at a higher expense of increasing mutual information. Therefore, the game with smaller λ achieves a lower cost overall. This coincides with the value of potential functions in NEs in Fig. 8.1.

Fig. 8.3 and Fig. 8.4 depict the sensor selection and the corresponding variances in terms of round robin when $k = 2$, $\rho = 0.9$ and SNR = 30 dB on the IEEE 9-bus test system with $\lambda = 2$ and $\lambda = 5$, respectively. Attacker 1 has access to the sensors with indices 1 to 22 and attacker 2 has access to the sensors with indices 23 to m. Interestingly, in both Fig. 8.3 and Fig. 8.4, attacker 1 selects different sensors in different round robin and chooses one sensor as the game convergences while the variances are similar. However, in both Fig. 8.3 and Fig. 8.4, attacker 2 chooses the same sensor from the first round robin until the game convergences even though the variances in different round robin are slightly different. From the topology of IEEE 9-bus system, the sensor selected by attacker 2 has the most physical connections with the other buses. This coincides with the sensor vulnerability analysis in Chapter 6 where the measurement that has more connection to the others is the most vulnerable one.

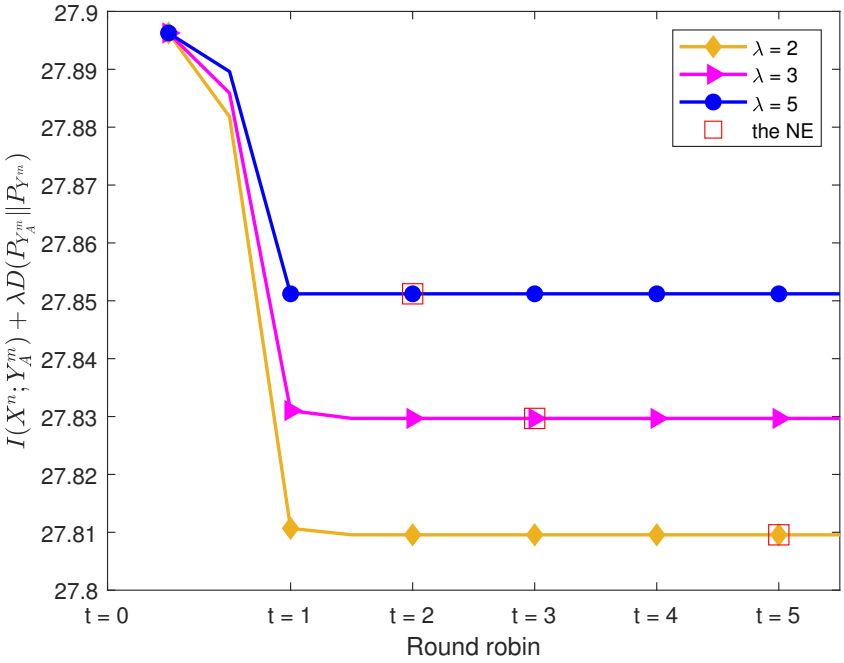

Figure 8.1: The convergence of the potential function P in $\mathcal{G}$ with different λ when $\rho = 0.9$, SNR $= 30$ dB, $k = 2$ on the IEEE 9-bus test system.

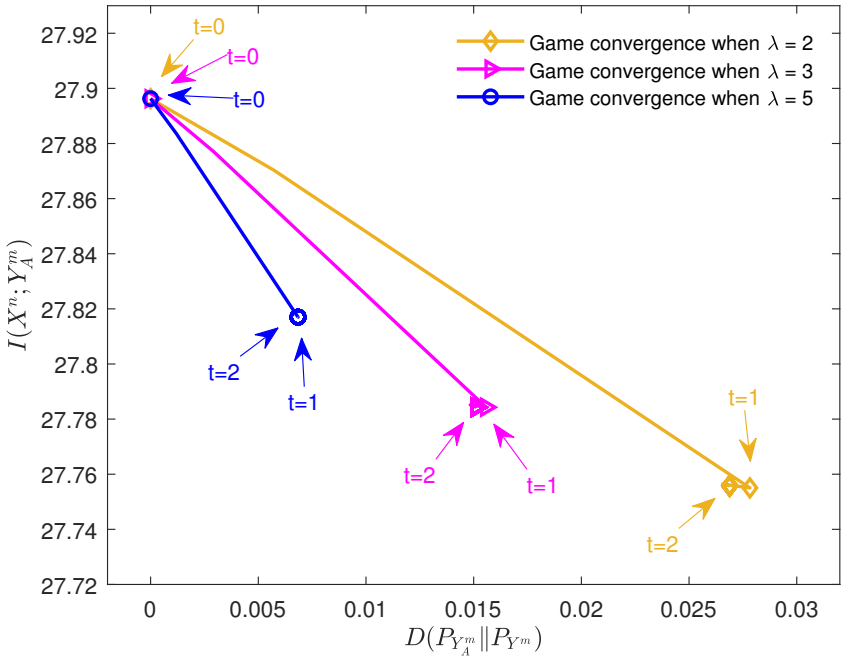

Figure 8.2: The tradeoff between mutual information and KL divergence in $\mathcal{G}$ with different λ when $\rho = 0.9$, SNR $= 30$ dB and $k = 2$ on the IEEE 9-bus test system.

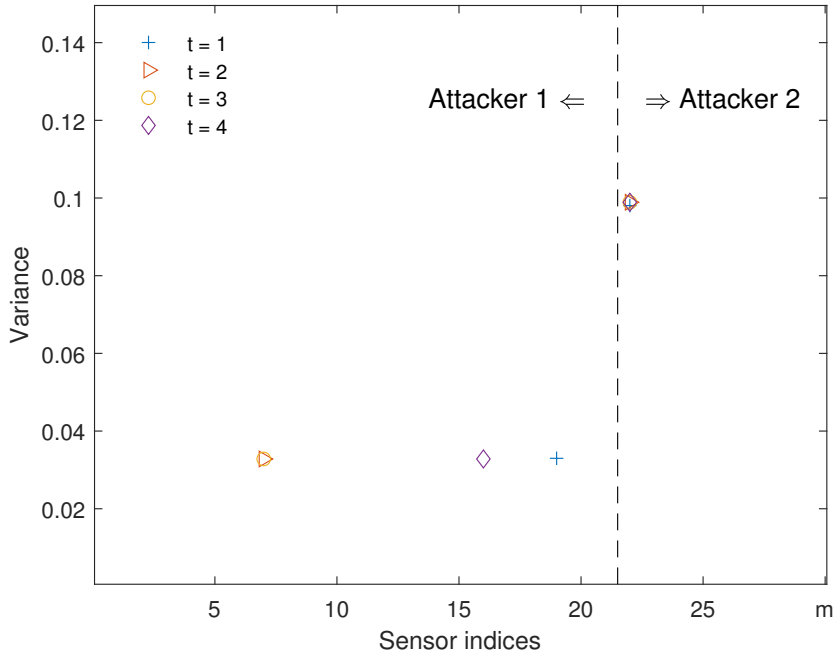

Figure 8.3: The sensor selection and the corresponding variances in different the round robin when $\rho = 0.9$, SNR $= 30$ dB, $\lambda = 2$ and $k = 2$ on the IEEE 9-bus test system.

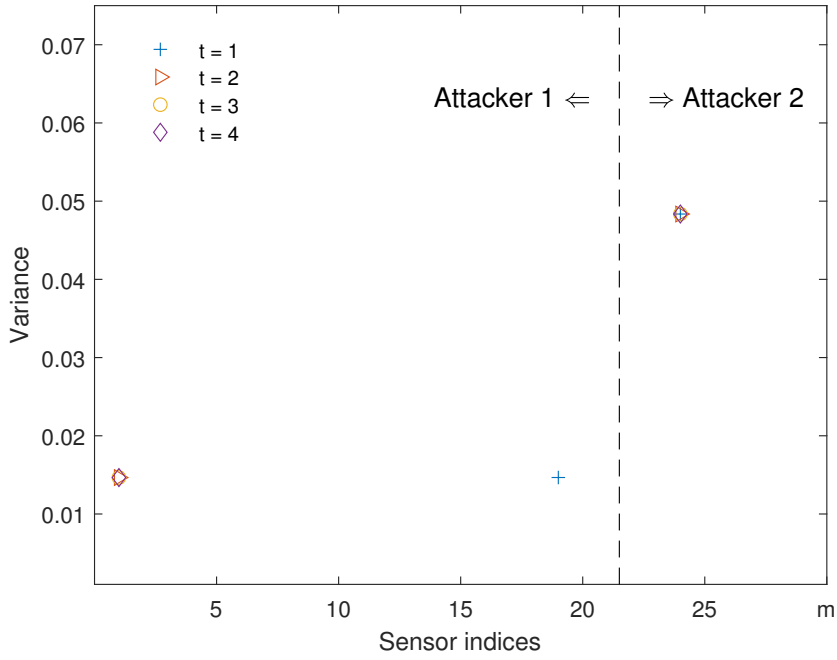

Figure 8.4: The sensor selection and the corresponding variances in different round robin when $\rho = 0.9$, SNR $= 30$ dB, $\lambda = 5$ and $k = 2$ on the IEEE 9-bus test system.

8.3.2 Game Convergence with Sparsity Constraints k

Fig. 8.5 depicts the convergence of the potential function P in $\mathcal{G}$ with different sparsity constraints k when $\rho = 0.9$, SNR $= 30$ dB, $\lambda = 2$ on the IEEE 9-bus test system. The number of the attackers in the game is constrained by k. The sets of measurements that the attackers have access to form a proper partition of the set of the measurements in the system. Note that there are different number of attackers in Fig. 8.5. One marker denotes the start of the next round robin or an end of the previous round robin. As expected, when there are more attackers, i.e., larger k achieves a lower value of potential function. In fact, given there are more attackers and one attacker compromises one measurement, it is expected the game with larger k achieves better performance in the NE. In other words, the relaxation of sparsity constraints leads to a better performance.

Fig. 8.6 depicts tradeoff between mutual information and KL divergence in $\mathcal{G}$ with different sparsity constraints k when $\rho = 0.9$, SNR $= 30$ dB, $\lambda = 2$ on the IEEE 9-bus test system. Larger sparsity constraints k indicate a better achievable performance of the attacks. As expected, the game with $k = 8$ achieves much lower mutual information at the expense of a small increase in KL divergence. Interestingly, in all the cases with different k, in the end of the first round robin when $t = 1$, the attacks obtain smaller mutual information and larger KL divergence in comparison with in the NE. Hence, from $t = 1$ to $t = 2$, the attackers modify the attacks to constrain KL divergence. This is because in the first round robin from $t = 0$ to $t = 1$, the attack decisions were made when there were limited attackers that attacked the system. Therefore, the decisions from best response in Theorem 30 are more aggressive. However, from $t = 1$ to $t = 2$, all the attackers have launched attacks, the result from best response in Theorem 30 limits the variance of the attacks which yields a decrease in KL divergence.

8.4 Summary

This chapter has proposed a novel decentralized stealth attack constructions with coordination and sparsity constraints. The sets of measurements that different attackers have access to form a proper partition of the set of the measurements in the systems. The objectives of the attack constructions are the disruption and detection that are measured by mutual information and KL divergence. The interaction between the attackers is modelled in a game framework. It is proved the game is a potential game as well as the existence of the potential function. The best response for the attackers is characterized and best response dynamics are proposed to achieve the NE of the game. The simulations have numerically evaluated the performance of the decentralized attacks with coordination and sparsity constraints on the IEEE test systems and shown the game achieves better performance with smaller λ and larger k in the NE.

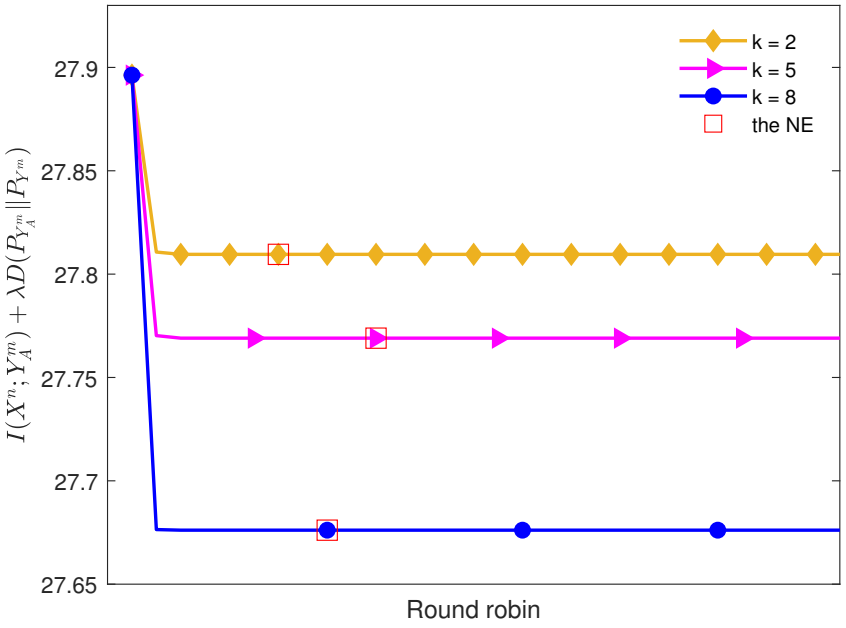

Figure 8.5: The convergence of the potential function P in $\mathcal{G}$ with different sparsity constraints k when $\rho = 0.9$, SNR $= 30$ dB, $\lambda = 2$ on the IEEE 9-bus test system.

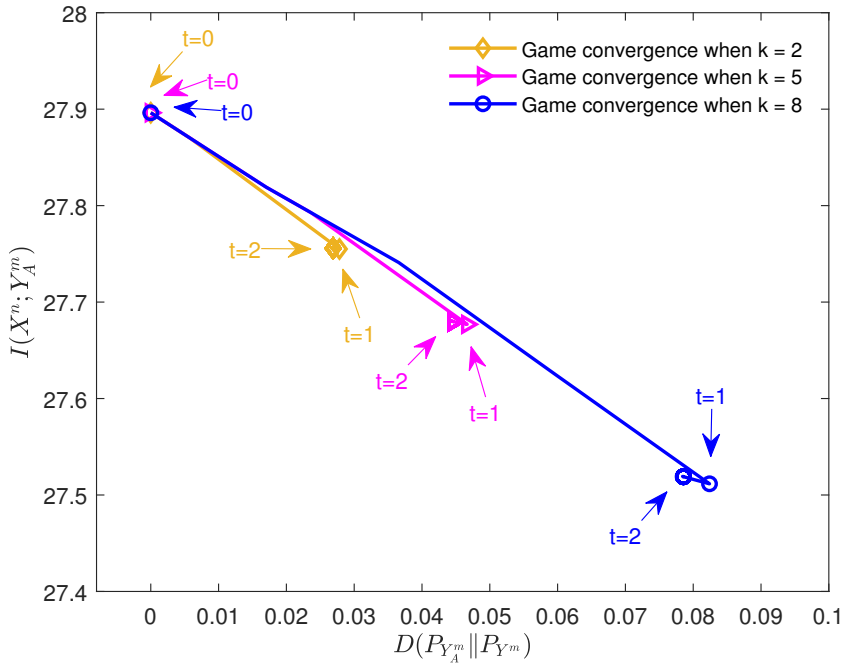

Figure 8.6: The tradeoff between mutual information and KL divergence in $\mathcal{G}$ with different sparsity constraints k when $\rho = 0.9$, SNR $= 30$ dB, $\lambda = 2$ on the IEEE 9-bus test system.

Chapter 9

Conclusions and Future Work

9.1 Conclusions

In this thesis, the information theoretic data injection attack (DIA) constructions with sparsity and coordination constraints are studied. Information theoretic metrics are considered as the objectives of the attack constructions which poses the attack constructions. The attack constructions aim at the disruption to the state estimate and the probability of detection. The disruption and detection are captured by mutual information and KL divergence, respectively.

Particularly, in Chapter 4 and Chapter 5, the information theoretic attacks are proposed with sparsity constraints. A novel stealth attack constructions with sparsity constraints are proposed. The proposed attack constructions minimize the mutual information between the state variables and the compromised measurements obtained by the operator while minimizing KL divergence between the distributions of measurements under attacks and without attacks. The optimal single measurement attack case is analytically characterized. To overcome the combinatorial challenge of identifying the measurements to be attacked, the insight on optimal single measurement attack case is distilled to construct greedy algorithms are proposed to minimize the additional cost of compromising one more measurement and sequentially update the set of compromised measurements. In Chapter 4, the attacks on different measurements are assumed to be independent. In Chapter 5, the correlation between different random attacks are considered. It is shown that the greedy step results in a convex optimization problems which can be solved efficiently and yields a low complexity attack update rule. The performance of the proposed independent attacks and correlated attacks are numerically assessed on the IEEE test systems. A better attack performance is achieved in correlated attack constructions at the expense of the communication between different attack locations.

In attack constructions with sparsity constraints in Chapter 4 and Chapter 5, it is observed from Fig. 4.1 to Fig. 4.3 that the probability of detection exhibits a threshold effect when a critical number of measurements are compromised. This observation leads to the analysis on the measurement vulnerability in Chapter 6. In this research, a novel security metric that referred to as VuIx is designed. The VuIx characterizes vulnerability of power system measurements to data integrity attacks from a fundamental perspective. This is

achieved by embedding information theoretic measures into the metric definition. The VuIx serves as an metric to assess the measurement vulnerability and gives the insight on the measurements that are more exposed to data integrity attacks. The simulations have tested the framework for the IEEE test systems and concluded that power injection measurements are more vulnerable to data integrity attacks than power flow measurements.

In Chapter 7, it is assumed that there are multiple attackers that construct the attacks in a coordinated fashion in a decentralized system. In decentralized systems, central decision maker does not exist. There are multiple attackers aim to disrupt the state estimate while limiting the probability of detection. From the disruption perspective, as an individual, one attacker is interested in the disruption to the state estimate that results from its own attack or the disruption that results from overall attacks. This research proposes the mutual information between the state variables and the measurement that one attacker has access to as the metric to capture the disruption that results from one attacker's own attack. This is referred to as local mutual information. The mutual information between the state variables and the measurements in the system is referred to as global mutual information which describes the disruption results from overall attacks. From the detection perspective, as an individual, one attacker is interested in the probability of detection under joint detection on the vector of measurements and local detection on the one measurement it has access to. Hence, the global KL divergence between the distributions of the measurements under attacks and without attacks is proposed to capture the probability of joint detection. Similarly, local KL divergence between the distributions of one measurement under attacks and without attacks is proposed to capture the probability of local detection. This research has developed these metrics into objectives in different attack constructions as games in Chapter 7. The games are proved to be potential games with potential functions accordingly. It is proved the convexity of the potential functions followed by the existence and the uniqueness of the Nash Equilibrium in each game. Best response has been analytically characterized and best response dynamics are proposed to achieve the unique Nash Equilibrium in each game, accordingly.

This thesis has also proposed a decentralized attacks with sparsity constraints where the set of measurements that the attackers have access to form a proper partition. The objectives of the decentralized attack constructions with sparsity constraints are the disruption and detection that are captured by the global mutual information and global KL divergence as in Chapter 7. The decentralized attacks with sparsity constraints are developed to form a game in Chapter 8. It is shown that the game is a potential game and the corresponding potential function is characterized. This chapter has also proved the existence and achievability of the Nash Equilibrium. The simulations in Chapter 7 have numerically evaluated the performance of the decentralized attacks and the performance of the decentralized attacks with sparsity constraints on the IEEE test systems. It is shown that the coordination in decentralized attacks can be developed to games with different objectives and the games converge to the NEs.

9.2 Future Work

9.2.1 Sensitivity Analysis of the Measurement Vulnerability

In Chapter 4 and Chapter 5, it is observed that some measurements are more vulnerable to others. The attackers in sparse attack constructions are more likely to attack power injection measurements which can be numerically verified by the vulnerability analysis in Chapter 6. Meanwhile, note that the optimal single measurement attack construction in Theorem 13 suggests that the measurements that have more connections to other buses are more likely to be compromised. Generally, power injection measurements have more connection to the other buses than power flow measurements. However, these are numerical observations. Further sensitivity analysis is needed to give the insights on the the measurement vulnerability.

9.2.2 Analysis on the Topology

In attack constructions with sparsity constraints, sparsity penalty suggests that the topology of the system fundamentally changes the performance of the attack but the specific mechanisms are left for future study. For example, in Fig. 4.1, there is a threshold over which the probability of detection increases significantly. The threshold effect happens for different cases with both small and large values of λ that is a weighting parameter of the disruption and detection. This suggests that the topology of the system governs the position of the threshold. It is also observed that in Fig. 4.11 and Fig. 5.8, some state estimate deviate more than the others both in independent attacks and correlated attacks. This suggests that different state variables suffer from DIAs differently and the topology of the system has an impact on the deviation of the state estimate under attacks.

9.2.3 Decentralized DIAs with Sparsity Constraints

This thesis has proposed the decentralized DIAs with sparsity constraints that the set of measurements that the attackers have access to form a proper partition of the set of the measurements on the systems. However, this thesis does not explored the general sparsity constraints on the decentralized attack constructions. The general sparsity constraints do not assume the set of measurements that the attackers have access to form a proper partition which may result in the collision in sensor selection when different attackers choose the same measurement to attack.

Appendix

A Proof of Proposition 5

Proof. Let $W^{n+m} \triangleq (X^n, Y_A^m)$. It follows that $W^{n+m} \sim \mathcal{N}(\boldsymbol{\mu}_W, \boldsymbol{\Sigma})$ such that

$$\boldsymbol{\mu}_W \triangleq \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\mu}_A \end{pmatrix}, \tag{1}$$

$$\boldsymbol{\Sigma} \triangleq \begin{pmatrix} \boldsymbol{\Sigma}_{XX} & \boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T} \\ \mathbf{H}\boldsymbol{\Sigma}_{XX} & \mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T} + \sigma^2\mathbf{I}_m + \boldsymbol{\Sigma}_{AA} \end{pmatrix}. \tag{2}$$

Note that

$$I(X^n; Y_A^m) \triangleq \mathbb{E}_{X^n, Y_A^m}\left[\log \frac{f_{X^n, Y_A^m}}{f_{X^n} f_{Y_A^m}}\right] \tag{3}$$

$$= \mathbb{E}_{W^{n+m}}\left[\log \frac{f_{W^{n+m}}}{f_{X^n} f_{Y_A^m}}\right], \tag{4}$$

where the functions f_{X^n, Y_A^m}, f_{X^n} and $f_{Y_A^m}$ in (3) are the pdf of (X^n, Y_A^m), X^n and Y_A^m, respectively, the function $f_{W^{n+m}}$ in (4) is the pdf of W^{n+m} and $f_{X^n, Y_A^m} = f_{W^{n+m}}$. It follows

that

$$
\begin{aligned}
& I(X^n; Y_A^m) && (5)\\
=& \mathbb{E}_{W^{n+m}}\left[\log \frac{\exp(-\frac{1}{2}(W^{n+m}-\boldsymbol{\mu}_W)^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(W^{n+m}-\boldsymbol{\mu}_W))}{(2\pi)^{\frac{n+m}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\right] && (6)\\
& -\mathbb{E}_{X^n}\left[\log \frac{\exp(-\frac{1}{2}(X^n)^{\mathsf{T}}\boldsymbol{\Sigma}_{XX}^{-1}X^n)}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}_{XX}|^{\frac{1}{2}}}\right]\\
& -\mathbb{E}_{Y_A^m}\left[\log \frac{\exp(-\frac{1}{2}(Y_A^m-\boldsymbol{\mu}_A)^{\mathsf{T}}\boldsymbol{\Sigma}_{Y_AY_A}^{-1}(Y_A^m-\boldsymbol{\mu}_A))}{(2\pi)^{\frac{m}{2}}|\boldsymbol{\Sigma}_{Y_AY_A}|^{\frac{1}{2}}}\right]\\
=& -\frac{1}{2}\mathbb{E}_{W^{n+m}}\left[\left(W^{n+m}-\boldsymbol{\mu}_W\right)^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\left(W^{n+m}-\boldsymbol{\mu}_W\right)\right]-\frac{1}{2}\log(2\pi)^{n+m}|\boldsymbol{\Sigma}| && (7)\\
& +\frac{1}{2}\mathbb{E}_{X^n}\left[(X^n)^{\mathsf{T}}\boldsymbol{\Sigma}_{XX}^{-1}X^n\right]+\frac{1}{2}\log(2\pi)^n|\boldsymbol{\Sigma}_{XX}|\\
& +\frac{1}{2}\mathbb{E}_{Y_A^m}\left[(Y_A^m-\boldsymbol{\mu}_A)^{\mathsf{T}}\boldsymbol{\Sigma}_{Y_AY_A}^{-1}(Y_A^m-\boldsymbol{\mu}_A)\right]+\frac{1}{2}\log(2\pi)^m|\boldsymbol{\Sigma}_{Y_AY_A}|\\
=& -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbb{E}_{W^{n+m}}\left[\left(W^{n+m}-\boldsymbol{\mu}_W\right)\left(W^{n+m}-\boldsymbol{\mu}_W\right)^{\mathsf{T}}\right]\right)+\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}_{XX}^{-1}\mathbb{E}_{X^n}\left[X^n(X^n)^{\mathsf{T}}\right]\right) && (8)\\
& +\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}_{Y_AY_A}^{-1}\mathbb{E}_{Y_A^m}\left[\left(Y_A^m-\boldsymbol{\mu}_A\right)\left(Y_A^m-\boldsymbol{\mu}_A\right)^{\mathsf{T}}\right]\right)+\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_{XX}||\boldsymbol{\Sigma}_{Y_AY_A}|}{|\boldsymbol{\Sigma}|}\\
=& -\frac{1}{2}(n+m)+\frac{1}{2}n+\frac{1}{2}m+\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_{XX}||\boldsymbol{\Sigma}_{Y_AY_A}|}{|\boldsymbol{\Sigma}|} && (9)\\
=& \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_{XX}||\boldsymbol{\Sigma}_{Y_AY_A}|}{|\boldsymbol{\Sigma}|}, && (10)
\end{aligned}
$$

where (6) follows from taking the expression of the probability density functions of multivariate Gaussian distributions, that is, f_{X^n}, $f_{Y_A^m}$ and $f_{W^{n+m}}$ into (4), (7) follows from taking constants out of the expectation, (8) follows from the fact that

$$
\begin{aligned}
& \mathbb{E}_{W^{n+m}}\left[\left(W^{n+m}-\boldsymbol{\mu}_W\right)^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\left(W^{n+m}-\boldsymbol{\mu}_W\right)\right] && (11)\\
=& \mathbb{E}_{W^{n+m}}\left[\mathrm{tr}\left(\left(W^{n+m}-\boldsymbol{\mu}_W\right)^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\left(W^{n+m}-\boldsymbol{\mu}_W\right)\right)\right]\\
=& \mathbb{E}_{W^{n+m}}\left[\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\left(W^{n+m}-\boldsymbol{\mu}_W\right)\left(W^{n+m}-\boldsymbol{\mu}_W\right)^{\mathsf{T}}\right)\right]\\
=& \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbb{E}_{W^{n+m}}\left[\left(W^{n+m}-\boldsymbol{\mu}_W\right)\left(W^{n+m}-\boldsymbol{\mu}_W\right)^{\mathsf{T}}\right]\right),
\end{aligned}
$$

$$
\mathbb{E}_{X^n}\left[(X^n)^{\mathsf{T}}\boldsymbol{\Sigma}_{XX}^{-1}X^n\right]=\mathrm{tr}\left(\boldsymbol{\Sigma}_{XX}^{-1}\mathbb{E}_{X^n}\left[X^n(X^n)^{\mathsf{T}}\right]\right), \quad (12)
$$

$$
\mathbb{E}_{Y_A^m}\left[(Y_A^m-\boldsymbol{\mu}_A)^{\mathsf{T}}\boldsymbol{\Sigma}_{Y_AY_A}^{-1}(Y_A^m-\boldsymbol{\mu}_A)\right]=\mathrm{tr}\left(\boldsymbol{\Sigma}_{Y_AY_A}^{-1}\mathbb{E}_{Y_A^m}\left[(Y_A^m-\boldsymbol{\mu}_A)(Y_A^m-\boldsymbol{\mu}_A)^{\mathsf{T}}\right]\right), \quad (13)
$$

and (9) follows from the fact that

$$\begin{aligned}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbb{E}_{W^{n+m}}\left[\left(W^{n+m}-\boldsymbol{\mu}_W\right)\left(W^{n+m}-\boldsymbol{\mu}_W\right)^{\mathsf{T}}\right]\right)=&\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\right)=n+m, &(14)\\ \operatorname{tr}\left(\boldsymbol{\Sigma}_{XX}^{-1}\mathbb{E}_{X^n}\left[X^n\left(X^n\right)^{\mathsf{T}}\right]\right)=&\operatorname{tr}\left(\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XX}\right)=n, &(15)\\ \operatorname{tr}\left(\boldsymbol{\Sigma}_{Y_AY_A}^{-1}\mathbb{E}_{Y_A^m}\left[\left(Y_A^m-\boldsymbol{\mu}_A\right)\left(Y_A^m-\boldsymbol{\mu}_A\right)^{\mathsf{T}}\right]\right)=&\operatorname{tr}\left(\boldsymbol{\Sigma}_{Y_AY_A}^{-1}\boldsymbol{\Sigma}_{Y_AY_A}\right)=m. &(16)\end{aligned}$$

This completes the proof. □

B Proof of Proposition 6

Proof. Note that

$$D(P_{Y_A^m}\|P_{Y^m})\triangleq\mathbb{E}_{Y_A^m}\left[\log\frac{f_{Y_A^m}}{f_{Y^m}}\right], \tag{17}$$

where the functions $f_{Y_A^m}$ and f_{Y^m} are the pdf of Y_A^m and Y^m, respectively. It follows that

$$\begin{aligned}&D(P_{Y_A^m}\|P_{Y^m}) &(18)\\ =&\mathbb{E}_{Y_A^m}\left[\log\frac{\exp(-\frac{1}{2}(Y_A^m-\boldsymbol{\mu}_A)^{\mathsf{T}}\boldsymbol{\Sigma}_{Y_AY_A}^{-1}(Y_A^m-\boldsymbol{\mu}_A))}{(2\pi)^{\frac{m}{2}}|\boldsymbol{\Sigma}_{Y_AY_A}|^{\frac{1}{2}}}-\log\frac{\exp(-\frac{1}{2}(Y_A^m)^{\mathsf{T}}\boldsymbol{\Sigma}_{YY}^{-1}Y_A^m)}{(2\pi)^{\frac{m}{2}}|\boldsymbol{\Sigma}_{YY}|^{\frac{1}{2}}}\right] &(19)\\ =&\frac{1}{2}\mathbb{E}_{Y_A^m}\left[\log\frac{\exp(-(Y_A^m-\boldsymbol{\mu}_A)^{\mathsf{T}}\boldsymbol{\Sigma}_{Y_AY_A}^{-1}(Y_A^m-\boldsymbol{\mu}_A))}{|\boldsymbol{\Sigma}_{Y_AY_A}|}-\log\frac{\exp(-(Y_A^m)^{\mathsf{T}}\boldsymbol{\Sigma}_{YY}^{-1}Y_A^m)}{|\boldsymbol{\Sigma}_{YY}|}\right] &(20)\\ =&\frac{1}{2}\mathbb{E}_{Y_A^m}\left[-(Y_A^m-\boldsymbol{\mu}_A)^{\mathsf{T}}\boldsymbol{\Sigma}_{Y_AY_A}^{-1}(Y_A^m-\boldsymbol{\mu}_A)+(Y_A^m)^{\mathsf{T}}\boldsymbol{\Sigma}_{YY}^{-1}Y_A^m\right]+\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{Y_AY_A}|} &(21)\\ =&-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_{Y_AY_A}^{-1}\mathbb{E}_{Y_A^m}\left[\left(Y_A^m-\boldsymbol{\mu}_A\right)\left(Y_A^m-\boldsymbol{\mu}_A\right)^{\mathsf{T}}\right]\right)+\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\mathbb{E}_{Y_A^m}\left[Y_A^m(Y_A^m)^{\mathsf{T}}\right]\right) &(22)\\ &+\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{Y_AY_A}|}\\ =&-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_{Y_AY_A}^{-1}\boldsymbol{\Sigma}_{Y_AY_A}\right)+\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}(\boldsymbol{\Sigma}_{Y_AY_A}+\boldsymbol{\mu}_A\boldsymbol{\mu}_A^{\mathsf{T}})\right)+\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{Y_AY_A}|} &(23)\\ =&-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_{Y_AY_A}^{-1}\boldsymbol{\Sigma}_{Y_AY_A}\right)+\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{Y_AY_A}+\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\mu}_A\boldsymbol{\mu}_A^{\mathsf{T}}\right)+\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{Y_AY_A}|} &(24)\\ =&\frac{1}{2}\left(\log\frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{Y_AY_A}|}-m+\operatorname{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{Y_AY_A}\right)+\operatorname{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\mu}_A\boldsymbol{\mu}_A^{\mathsf{T}}\right)\right) &(25)\end{aligned}$$

where (19) follows from taking the definition of KL divergence into (18), (21) follows from taking constants out of the expectation, (23) follows from (13) and the fact that

$$\mathbb{E}_{Y_A^m}\left[(Y_A^m)^{\mathsf{T}}\boldsymbol{\Sigma}_{YY}^{-1}Y_A^m\right]=\operatorname{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\mathbb{E}_{Y_A^m}\left[(Y_A^m)^{\mathsf{T}}Y_A^m\right]\right), \tag{26}$$

and

$$\begin{aligned}\mathbb{E}_{Y_A^m}\left[Y_A^m(Y_A^m)^{\mathsf{T}}\right]=&\mathbb{E}_{Y_A^m}\left[\left(Y_A^m-\boldsymbol{\mu}_A\right)\left(Y_A^m-\boldsymbol{\mu}_A\right)^{\mathsf{T}}\right]+\boldsymbol{\mu}_A\boldsymbol{\mu}_A^{\mathsf{T}} &(27)\\ =&\boldsymbol{\Sigma}_{Y_AY_A}+\boldsymbol{\mu}_A\boldsymbol{\mu}_A^{\mathsf{T}}.\end{aligned}$$

This completes the proof. □

C Proof of Theorem 13

Proof. The proof starts by noting that for $k = 1$ the set of attack covariance matrices $\tilde{\mathcal{S}}_k$ is

$$\tilde{\mathcal{S}}_1 \triangleq \bigcup_{i=1,\dots,m} \left\{ \mathbf{S} \in S_+^m : \mathbf{S} = v_i \mathbf{e}_i \mathbf{e}_i^\mathsf{T} \text{ with } v_i \in \mathbb{R}_+ \right\}. \tag{28}$$

The covariance matrices in set $\tilde{\mathcal{S}}_1$ are matrices with a single positive real element in the diagonal. The non-zero entry i denotes the index of the measurement that is compromised. Let $i \in \{1, 2, ..., m\}$ be the index of the non-zero entry of the covariance matrix $\bar{\mathbf{\Sigma}}_{AA}$. The non-zero entry denoted by v_i is the variance of the random variable used to attack the measurement i.

Let $\lambda \geq 1$, $\mathbf{W} = \mathbf{\Sigma}_{YY}^{-1}$ and restrict the optimization domain in (4.41) to $\tilde{\mathcal{S}}_1$. Thus, the resulting optimization problem is equivalent to:

$$\min_{v \in \mathbb{R}_+} \min_{i \in \{1,2,\dots,m\}} \log \frac{(1 + (\mathbf{W})_{ii} v)^{1-\lambda}}{(\sigma^2 + v)} + \lambda (\mathbf{W})_{ii} v. \tag{29}$$

The proof proceeds by solving the inner part of the optimization problem above. Consider the cost given by

$$f((\mathbf{W})_{ii}) \triangleq \log \frac{(1 + (\mathbf{W})_{ii} v)^{1-\lambda}}{\sigma^2 + v} + \lambda (\mathbf{W})_{ii} v, \tag{30}$$

which can be rewritten as

$$f(t) = (1 - \lambda) \log t - \log(\sigma^2 + v) + \lambda t - \lambda, \tag{31}$$

where $t = 1 + (\mathbf{W})_{ii} v$. It follows that (31) is convex with respect to t because λt is a linear term and $(1 - \lambda) \log t$ is convex in t for $\lambda \geq 1$. Therefore, $f((\mathbf{W})_{ii})$ is convex with respect to $(\mathbf{W})_{ii}$ and the minimum is obtained for $(\mathbf{W})_{ii} = -\frac{1}{\lambda v}$. Since $(\mathbf{W})_{ii} > 0$ the inner minimization in (29) is equivalent to selecting the index i that minimizes $(\mathbf{W})_{ii}$. The definition of i in (4.43a) and $\underline{w}$ in (4.43b) follow from this observation.

The proof now proceeds to solve the outer optimization. In this case, the cost is given by

$$g(v) = (1 - \lambda) \log(1 + \underline{w} v) - \log(\sigma^2 + v) + \lambda \underline{w} v, \tag{32}$$

where $r \triangleq v$. Noticing that the above function has a single minimizer given by

$$v = -\frac{\sigma^2}{2} + \frac{1}{2} \left(\sigma^4 - \frac{4(\underline{w}\sigma^2 - 1)}{\lambda \underline{w}^2} \right)^{\frac{1}{2}}. \tag{33}$$

This completes the proof. □

D Proof of Lemma 14

Proof. The proof follows by showing that the difference between $J(\mathbf{\Sigma}_i)$ and $J(\mathbf{\Sigma}_{i-1})$, that is,

$$
\begin{aligned}
& J(\mathbf{\Sigma}_i) - J(\mathbf{\Sigma}_{i-1}) && (34)\\
=&(1-\lambda)\log|\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_i| - \log|\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_i| + \lambda\mathrm{tr}(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_i) \\
&-\left((1-\lambda)\log|\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{i-1}| - \log|\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_{i-1}| + \lambda\mathrm{tr}(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{i-1})\right) && (35)\\
=&(1-\lambda)\log\frac{\left|\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{i-1}+v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right|}{|\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{i-1}|} - \log\frac{\left|\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_{i-1}+v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right|}{|\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_{i-1}|} && (36)\\
&+\lambda\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\left(\mathbf{\Sigma}_i-\mathbf{\Sigma}_{i-1}\right)\right) \\
=&(1-\lambda)\log\left|\mathbf{I}_m+\left(\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{i-1}\right)^{-1}v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right| - \log\left|\mathbf{I}_m+\left(\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_{i-1}\right)^{-1}v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right| && (37)\\
&+\lambda\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right),
\end{aligned}
$$

where (35) follows from taking $\mathbf{\Sigma}_{i-1}$ and $\mathbf{\Sigma}_i$ into (3.49), (36) follows from replacing $\mathbf{\Sigma}_i$ with $\mathbf{\Sigma}_{i-1}+v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}$ and (37) follows from eliminating $|\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{i-1}|$ and $|\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_{i-1}|$. This completes the proof. □

E Proof of Proposition 7

Proof. From Lemma 14, the optimization problem in (4.50) is

$$
\begin{aligned}
&\min_{(j,v)\in\mathcal{K}_{i-1}^c\times\mathbb{R}_+}(1-\lambda)\log\left|\mathbf{I}_m+\left(\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{i-1}\right)^{-1}v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right| - \log\left|\mathbf{I}_m+\left(\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_{i-1}\right)^{-1}v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right| \\
&\qquad +\lambda v\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right) && (38)\\
=&\min_{(j,v)\in\mathcal{K}_{i-1}^c\times\mathbb{R}_+}(1-\lambda)\log(1+\alpha_j v) - \log(1+\frac{v}{\sigma^2}) + \lambda\beta_j v, && (39)
\end{aligned}
$$

where $\alpha_j \triangleq \mathrm{tr}\left(\left(\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{i-1}\right)^{-1}\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right) > 0$ and $\beta_j \triangleq \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right)$. Note that

$$
\log\left|\mathbf{I}_m+\left(\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{i-1}\right)^{-1}v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right| = \log(1+\alpha_j v) \tag{40}
$$

follows from the fact that the matrix $\left(\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{i-1}\right)^{-1}v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}$ is a matrix with only one nonzero column in position j and

$$
\log\left|\mathbf{I}_m+\left(\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_{i-1}\right)^{-1}v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right| = \log(1+\frac{v}{\sigma^2}) \tag{41}
$$

follows from the fact that $j\in\mathcal{K}_{i-1}^c$. This completes the proof. □

F Proof of Theorem 15

Proof. It follows from Lemma 14 the optimization problem in (4.47) is equivalent to the optimization problem in (4.50) which is convex for $\lambda \geq 1$ from Proposition 7. After some

algebraic manipulation, it follows that the optimization problem in (4.50) is

$$\min_{(j,v)\in\mathcal{K}_{i-1}^{\mathrm{c}}\times\mathbb{R}_+} (1-\lambda)\log\left|\mathbf{I}_m+\left(\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{i-1}\right)^{-1}v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right|-\log\left|\mathbf{I}_m+\left(\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_{i-1}\right)^{-1}v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right| \\ +\lambda v\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right) \tag{42}$$

$$= \min_{(j,v)\in\mathcal{K}_{i-1}^{\mathrm{c}}\times\mathbb{R}_+} (1-\lambda)\log(1+\alpha_j v)-\log(1+\frac{v}{\sigma^2})+\lambda\beta_j v, \tag{43}$$

where $\alpha_j \triangleq \mathrm{tr}\left(\left(\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{i-1}\right)^{-1}\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right) > 0$ and $\beta_j \triangleq \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right)$. Therefore, the proof proceeds to characterize the first derivative of the cost in (43), that is,

$$\beta_j\alpha_j v^2+(\beta_j-\alpha_j+\beta_j\alpha_j\sigma^2)v+\beta_j\sigma^2-\alpha_j\sigma^2-\frac{\alpha_j\sigma^2+1}{\lambda}=0 \tag{44}$$

Note that (44) is quadratic with two solutions. The result follows from choosing the solution such that $v\in\mathbb{R}_+$. This completes the proof. □

G Proof of Proposition 9

Proof. Note that $Y_i \sim \mathcal{N}\left(0, \mathbf{e}_i^{\mathsf{T}}\mathbf{\Sigma}_{YY}\mathbf{e}_i\right)$. Consequently, it follows that

$$Y_{A,i} \sim \mathcal{N}\left(0, \mathbf{e}_i^{\mathsf{T}}\mathbf{\Sigma}_{YY}\mathbf{e}_i + v_i\right). \tag{45}$$

The mutual information between random vector of state variables $X^n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{XX})$ and the i-th random measurement $Y_{A,i}$ is

$$I(X^n; Y_{A,i}) \triangleq \mathbb{E}_{X^n,Y_{A,i}}\left[\log\frac{f_{X^n,Y_{A,i}}}{f_{X^n}f_{Y_{A,i}}}\right], \tag{46}$$

where f_{X^n}, $f_{Y_{A,i}}$ and $f_{X^n,Y_{A,i}}$ are the probability density functions of the random variables X^n, Y_{A_i} and (X^n, Y_{A_i}), respectively. Let us denote the random vector $W^{n+1} \triangleq (X^n, Y_{A,i}) \in \mathbb{R}^{n+1}$ and f_W be the probability density function of W such that $f_W = f_{X^n,Y_{A,i}}$. The random vector W^{n+1} follows a joint multivariate Gaussian distribution given by

$$W^{n+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}), \tag{47}$$

where the block covariance matrix has the following structure:

$$\mathbf{\Sigma} \triangleq \begin{bmatrix} \mathbf{\Sigma}_{XX} & \mathbf{\Sigma}_{XX}\mathbf{h}_i^{\mathsf{T}} \\ \mathbf{h}_i\mathbf{\Sigma}_{XX} & \mathbf{e}_i^{\mathsf{T}}\mathbf{\Sigma}_{YY}\mathbf{e}_i + v_i \end{bmatrix}, \tag{48}$$

where $\mathbf{h}_i$ is the i-th row of $\mathbf{H}$. Hence, the following holds

$$
\begin{aligned}
I(X^n;Y_{A,i}) =& \mathbb{E}_W\left[\log\frac{f_W}{f_{X^n}f_{Y_{A,i}}}\right], && (49)\\
=& \mathbb{E}_W\left[\log f_W\right] - \mathbb{E}_W\left[\log f_{X^n}\right] - \mathbb{E}_W\left[\log f_{Y_{A,i}}\right], && (50)\\
=& \mathbb{E}_W\left[\log\frac{\exp\left(-\frac{1}{2}W^{\mathsf{T}}\mathbf{\Sigma}^{-1}W\right)}{(2\pi)^{\frac{n+1}{2}}|\mathbf{\Sigma}|^{\frac{1}{2}}}\right] - \mathbb{E}_X\left[\log\frac{\exp\left(-\frac{1}{2}X^{\mathsf{T}}\mathbf{\Sigma}_{XX}^{-1}X\right)}{(2\pi)^{\frac{n}{2}}|\mathbf{\Sigma}_{XX}|^{\frac{1}{2}}}\right] && (51)\\
& -\mathbb{E}_{Y_{A,i}}\left[\log\frac{\exp\left(-\frac{1}{2}\frac{y_a^2}{\mathbf{e}_i^{\mathsf{T}}\mathbf{\Sigma}_{YY}\mathbf{e}_i+v_i}\right)}{(2\pi)^{\frac{1}{2}}(\mathbf{e}_i^{\mathsf{T}}\mathbf{\Sigma}_{YY}\mathbf{e}_i+v_i)^{\frac{1}{2}}}\right]\\
=& \frac{1}{2}\mathbb{E}_W\left[-W^{\mathsf{T}}\mathbf{\Sigma}^{-1}W-\log|\mathbf{\Sigma}|\right] + \frac{1}{2}\mathbb{E}_X\left[X^{\mathsf{T}}\mathbf{\Sigma}_{XX}^{-1}X+\log|\mathbf{\Sigma}_{XX}|\right] && (52)\\
& +\frac{1}{2}\mathbb{E}_{Y_{A,i}}\left[\frac{y_a^2}{\mathbf{e}_i^{\mathsf{T}}\mathbf{\Sigma}_{YY}\mathbf{e}_i+v_i}+\log(\mathbf{e}_i^{\mathsf{T}}\mathbf{\Sigma}_{YY}\mathbf{e}_i+v_i)\right]\\
=& \frac{1}{2}\mathbb{E}_W\left[-\mathrm{tr}(\mathbf{\Sigma}^{-1}WW^{\mathsf{T}})\right] + \frac{1}{2}\mathbb{E}_X\left[\mathrm{tr}(\mathbf{\Sigma}_{XX}^{-1}XX^{\mathsf{T}})\right] && (53)\\
& +\frac{1}{2\left(\mathbf{e}_i^{\mathsf{T}}\mathbf{\Sigma}_{YY}\mathbf{e}_i+v_i\right)}\mathbb{E}_{Y_{A,i}}\left[y_a^2\right] + \frac{1}{2}\log\frac{|\mathbf{\Sigma}_{XX}|(\mathbf{e}_i^{\mathsf{T}}\mathbf{\Sigma}_{YY}\mathbf{e}_i+v_i)}{|\mathbf{\Sigma}|}\\
=& \frac{1}{2}\left(-(n+1)+n+1\right) + \frac{1}{2}\log\frac{|\mathbf{\Sigma}_{XX}|(\mathbf{e}_i^{\mathsf{T}}\mathbf{\Sigma}_{YY}\mathbf{e}_i+v_i)}{|\mathbf{\Sigma}|} && (54)\\
=& \frac{1}{2}\log\frac{|\mathbf{\Sigma}_{XX}|(\mathbf{e}_i^{\mathsf{T}}\mathbf{\Sigma}_{YY}\mathbf{e}_i+v_i)}{|\mathbf{\Sigma}|} && (55)
\end{aligned}
$$

where (51) holds from taking the probability density functions of W^{n+1}, $Y_{A,i}$ and X^n to (50); (53) follows from taking the constants out of expectation. The following holds

$$
\begin{aligned}
|\mathbf{\Sigma}| =& |\mathbf{\Sigma}_{XX}||\mathbf{e}_i^{\mathsf{T}}\mathbf{\Sigma}_{YY}\mathbf{e}_i+v_i-\mathbf{h}_i\mathbf{\Sigma}_{XX}\mathbf{\Sigma}_{XX}^{-1}\mathbf{\Sigma}_{XX}\mathbf{h}_i^{\mathsf{T}}| && (56)\\
=& |\mathbf{\Sigma}_{XX}||\mathbf{e}_i^{\mathsf{T}}\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i+\sigma^2+v_i-\mathbf{h}_i\mathbf{\Sigma}_{XX}\mathbf{h}_i^{\mathsf{T}}| && (57)\\
=& |\mathbf{\Sigma}_{XX}|(\sigma^2+v_i), && (58)
\end{aligned}
$$

where (56) holds from [116, 14.17(a)]; (57) follows from $\mathbf{\Sigma}_{YY}=\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}+\sigma^2\mathbf{I}_m$; (58) follows from $\mathbf{h}_i=\mathbf{e}_i^{\mathsf{T}}\mathbf{H}$ and $\mathbf{h}_i^{\mathsf{T}}=\mathbf{H}^{\mathsf{T}}\mathbf{e}_i$. Therefore, from (55), it yields that

$$
\begin{aligned}
I(X^n;Y_i+A_i) =& \frac{1}{2}\log\frac{|\mathbf{\Sigma}_{XX}|(\mathbf{e}_i^{\mathsf{T}}\mathbf{\Sigma}_{YY}\mathbf{e}_i+v_i)}{|\mathbf{\Sigma}|}\\
=& \frac{1}{2}\log\frac{|\mathbf{\Sigma}_{XX}|(\mathbf{e}_i^{\mathsf{T}}\mathbf{\Sigma}_{YY}\mathbf{e}_i+v_i)}{|\mathbf{\Sigma}_{XX}|(\sigma^2+v_i)}\\
=& \frac{1}{2}\log\frac{\mathbf{e}_i^{\mathsf{T}}\mathbf{\Sigma}_{YY}\mathbf{e}_i+v_i}{\sigma^2+v_i}\\
=& \frac{1}{2}\log\left(1+\frac{\mathbf{e}_i^{\mathsf{T}}\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i}{\sigma^2+v_i}\right)
\end{aligned} \qquad (59)
$$

This completes the proof. □

H Proof of Proposition 11

The proof of Proposition 11 is obtained by applying the definition of KL divergence between two one dimensional Gaussian distributions.

Proof. Let $f_{P_{Y_{A,i}}}$ and $f_{P_{Y_i}}$ denote the probability density function of $P_{Y_{A,i}}$ and P_{Y_i}, respectively. Note that

$$Y_i \sim \mathcal{N}\left(0, \operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2\right), \tag{60}$$

and

$$Y_{A,i} \sim \mathcal{N}\left(0, \operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2 + v_i\right). \tag{61}$$

The KL divergence between $P_{Y_{A,i}}$ and P_{Y_i} is given by

$$D(P_{Y_{A,i}} \| P_{Y_i}) \tag{62}$$

$$\stackrel{\Delta}{=} \mathbb{E}_{P_{Y_{A,i}}}\left[\log \frac{f_{Y_{A,i}}}{f_{Y_i}}\right] \tag{63}$$

$$= \mathbb{E}_{P_{Y_{A,i}}}\left[\log \frac{\frac{1}{\sqrt{\operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2 + v_i}\sqrt{2\pi}} \exp\left[\frac{-x^2}{2(\operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2 + v_i)}\right]}{\frac{1}{\sqrt{\operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2}\sqrt{2\pi}} \exp\left[\frac{-x^2}{2\operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2}\right]}\right] \tag{64}$$

$$= \frac{1}{2}\mathbb{E}_{P_{Y_{A,i}}}\left[\frac{-x^2}{\operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2 + v_i} - \frac{-x^2}{\operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2}\right] \tag{65}$$

$$+ \frac{1}{2}\log \frac{\operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2}{\operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2 + v_i}$$

$$= \frac{1}{2}\frac{v_i}{\left(\operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2\right)\left(\left(\operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2 + v_i\right)\right)}\mathbb{E}_{P_{Y_{A,i}}}\left[x^2\right] \tag{66}$$

$$+ \frac{1}{2}\log \frac{\operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2}{\operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2 + v_i}$$

$$= \frac{1}{2}\frac{v_i}{\left(\operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2\right)\left(\operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2 + v_i\right)}(\operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2 + v_i)$$

$$+ \frac{1}{2}\log \frac{\operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2}{\operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2 + v_i} \tag{67}$$

$$= \frac{1}{2}\left(\frac{v_i}{\operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2} + \log \frac{\operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2}{\operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2 + v_i}\right), \tag{68}$$

where (64) follows from taking the density function of $P_{Y_{A,i}}$ and P_{Y_i}; (67) follows from the fact that the expectation of the random variable x^2 such that $x \sim \mathcal{N}(0, \operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2 + v_i)$ is $\operatorname{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) + \sigma^2 + v_i$. This completes the proof. □

I Proof of Proposition 17

The proof of Proposition 17 is obtained by characterizing the three terms in (7.48) are convex in v_i. Specifically, the first derivative of the first term is negative and the second derivative is positive, which yields the convexity of the first term.

Proof. Let us define $\mathbf{A} \triangleq \mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}} \sum_{j\in\mathcal{K}\setminus\{i\}} \frac{1}{\sigma^2+v_j}\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}} + \mathbf{I}_m$ and $\alpha_i \triangleq \mathrm{tr}\left(\mathbf{A}^{-1}\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right)$. The derivative of the first term with respect of v_i in (7.48) is

$$\frac{\partial}{\partial v_i}\log\left|\frac{1}{\sigma^2+v_i}\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}} + \mathbf{A}\right| \tag{69}$$

$$=-\frac{1}{(\sigma^2+v_i)^2}\mathrm{tr}\left(\left(\frac{1}{\sigma^2+v_i}\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}} + \mathbf{A}\right)^{-1}\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) \tag{70}$$

$$=-\frac{1}{(\sigma^2+v_i)^2} \tag{71}$$

$$\cdot\mathrm{tr}\left(\left(\mathbf{A}^{-1}-\frac{\frac{1}{\sigma^2+v_i}}{1+\frac{1}{\sigma^2+v_i}\mathbf{e}_i^{\mathsf{T}}\mathbf{A}^{-1}\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i}\mathbf{A}^{-1}\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\mathbf{A}^{-1}\right)\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) \tag{72}$$

$$=-\frac{1}{(\sigma^2+v_i)^2}\mathrm{tr}\left(\mathbf{A}^{-1}\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) \tag{73}$$

$$+\frac{1}{(\sigma^2+v_i)^2}\frac{\frac{1}{\sigma^2+v_i}}{1+\frac{1}{\sigma^2+v_i}\mathbf{e}_i^{\mathsf{T}}\mathbf{A}^{-1}\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i}\mathrm{tr}\left(\mathbf{A}^{-1}\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\mathbf{A}^{-1}\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) \tag{74}$$

$$=-\frac{1}{(\sigma^2+v_i)^2}\alpha_i + \frac{1}{(\sigma^2+v_i)^2}\frac{\frac{1}{\sigma^2+v_i}}{1+\frac{1}{\sigma^2+v_i}\alpha_i}\alpha_i^2 \tag{75}$$

$$=-\frac{\alpha_i}{(\sigma^2+v_i)\left(\sigma^2+v_i+\alpha_i\right)}, \tag{76}$$

where (70) follows from taking the derivative of $\log\left|\frac{1}{\sigma^2+v_i}\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}} + \mathbf{A}\right|$ with respect of v_i [116, Statement 17.18(a)] and taking $-\frac{1}{(\sigma^2+v_i)^2}$ out of the trace, (72) follows from Sherman-Morrison Formula in [116, 15.2(b)]. Note that $\alpha_i > 0$ and $v_i > 0$. It follows that $-\frac{\alpha_i}{(\sigma^2+v_i)\left(\sigma^2+v_i+\alpha_i\right)} < 0$. The proof now proceeds to exam the second derivative of $\log\left|\frac{1}{\sigma^2+v_i}\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}} + \mathbf{A}\right|$, that is,

$$\frac{\partial}{\partial v_i}\left(-\frac{\alpha_i}{(\sigma^2+v_i)\left(\sigma^2+v_i+\alpha_i\right)}\right) = \frac{\alpha_i\left(2(\sigma^2+v_i)+\alpha_i\right)}{(\sigma^2+v_i)^2\left(\sigma^2+v_i+\alpha_i\right)^2}. \tag{77}$$

Note that $\alpha_i > 0$ and $v_i > 0$. It follows that $\dfrac{\alpha_i \left(2(\sigma^2 + v_i) + \alpha_i\right)}{(\sigma^2 + v_i)^2 \left(\sigma^2 + v_i + \alpha_i\right)^2} > 0$, which yields that the first term $\log \left| \dfrac{1}{\sigma^2 + v_i} \mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}} + \mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}} \sum_{j \in \mathcal{K}\setminus\{i\}} \dfrac{1}{\sigma^2 + v_j}\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}} + \mathbf{I}_m \right|$ in (7.48) is convex. The second term is linear in v_i and the third term is convex in v_i. Therefore, the utility function of the i-th attacker in (7.48) is convex. This completes the proof. □

J Proof of Lemma 19

Proof. From (7.30) and Proposition 12, it follows that

$$\min_{v\in\mathbb{R}_+} I(X^n;Y_A^m)+\lambda D(P_{Y_A^m}\|P_{Y^m}) \tag{78}$$

$$=\min_{v\in\mathbb{R}_+} \phi_i^1(v_1,\ldots,v_{i-1},v,v_{i+1},\ldots,v_m) \tag{79}$$

$$=\frac{1}{2}\log\frac{|\mathbf{\Sigma}_{XX}|\left|\mathbf{\Sigma}_{YY}+v_i\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right|}{|\mathbf{\Sigma}|} \tag{80}$$

$$+\frac{1}{2}\lambda\left(\log\frac{|\mathbf{\Sigma}_{YY}|}{|\mathbf{\Sigma}_{YY}+v_i\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}|}-m\right)$$

$$+\frac{1}{2}\lambda\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\left(\mathbf{\Sigma}_{YY}+v_i\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right)\right)$$

$$=\frac{1-\lambda}{2}\log\left|\mathbf{\Sigma}_{YY}+v_i\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right|-\frac{1}{2}\log\left|\sigma^2\mathbf{I}_m+v_i\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right|$$

$$+v_i\frac{\lambda}{2}\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)+c \tag{81}$$

$$=\min_{v\in\mathbb{R}_+}\frac{1}{2}(1-\lambda)\log\left|\mathbf{\Sigma}_{YY}+v\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right|-\frac{1}{2}\log\left|\sigma^2\mathbf{I}_m+v\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right|$$

$$+\frac{1}{2}\lambda\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\left(v\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right)\right) \tag{82}$$

$$=\min_{v\in\mathbb{R}_+}\frac{1-\lambda}{2}\log\left|\mathbf{\Sigma}_{YY}+v\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right|-\frac{1}{2}\log(\sigma^2+v)-\frac{1}{2}\sum_{j\in\mathcal{K}\setminus\{i\}}\log(\sigma^2+v_j)$$

$$+v\frac{\lambda}{2}\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)+\frac{\lambda}{2}\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\left(\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right)\right) \tag{83}$$

$$\Leftrightarrow\min_{v\in\mathbb{R}_+}\frac{1-\lambda}{2}\log\left|\mathbf{\Sigma}_{YY}+v\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right|-\frac{1}{2}\log(\sigma^2+v)+v\frac{\lambda}{2}\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)+c \tag{84}$$

$$\Leftrightarrow\min_{v}(1-\lambda)\log\left|\mathbf{\Sigma}_{YY}+v\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right|-\log(\sigma^2+v)+v\lambda\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) \tag{85}$$

where (82) follows from Proposition 12, (83) follows from noting that the matrix $\sigma^2\mathbf{I}_m+v_i\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}$ is a diagonal matrix and therefore $|\sigma^2\mathbf{I}_m+v_i\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}|=\prod_{i=1}^m(\sigma^2+v_i)$, (84) follows from combining all the terms that are not a function of v and denote it as c. This completes proof. □

K Proof of Lemma 20

Proof. From (7.31) and Proposition 14, it follows that

$$\min_{v\in\mathbb{R}_+} I(X^n;Y_{A,i}) + \lambda D(P_{Y_A^m}\|P_{Y^m}) \tag{86}$$

$$= \min_{v\in\mathbb{R}_+} \frac{1}{2}\log\left(1+\frac{\mathrm{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)}{\sigma^2+v}\right) + \frac{1}{2}\lambda\left(\log\frac{|\boldsymbol{\Sigma}_{YY}|}{\left|\boldsymbol{\Sigma}_{YY}+v\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right|}\right)$$
$$+\frac{1}{2}\lambda\mathrm{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\left(v\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right)\right) \tag{87}$$

$$= \min_{v\in\mathbb{R}_+} \frac{1}{2}\log\left(1+\frac{\mathrm{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)}{\sigma^2+v}\right) - \frac{1}{2}\lambda\log\left|\boldsymbol{\Sigma}_{YY}+v\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right| \tag{88}$$
$$+\frac{1}{2}\lambda v\mathrm{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)+c$$

$$\Leftrightarrow \min_{v\in\mathbb{R}_+} \log\left(1+\frac{\mathrm{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)}{\sigma^2+v}\right) - \lambda\log\left|\boldsymbol{\Sigma}_{YY}+v\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right| \tag{89}$$
$$+\lambda v\mathrm{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)$$

where (87) follows from Proposition 14, (88) follows from combining the constant term that is not a function of v and denote it as c. □

L Proof of Lemma 21

Proof. From (7.32) and Proposition 16, it follows that

$$\min_{v\in\mathbb{R}_+} I(X^n;Y_A^m)+\lambda D(P_{Y_{A,i}}\|P_{Y_i}) \tag{90}$$

$$= \min_{v\in\mathbb{R}_+} \frac{1}{2}\log\frac{\left|\mathbf{\Sigma}_{YY}+v\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right|}{\left|\sigma^2\mathbf{I}_m+v\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right|} \tag{91}$$

$$+\frac{1}{2}\lambda\left(\frac{v}{\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)+\sigma^2}+\log\frac{\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)+\sigma^2}{\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)+\sigma^2+v}\right)$$

$$= \min_{v\in\mathbb{R}_+} \frac{1}{2}\log\left|\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\left(\sigma^2\mathbf{I}_m+v\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right)^{-1}+\mathbf{I}_m\right| \tag{92}$$

$$+\frac{\lambda}{2\left(\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)+\sigma^2\right)}v-\frac{1}{2}\lambda\log\left(\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)+\sigma^2+v\right)+c$$

$$= \min_{v\in\mathbb{R}_+} \frac{1}{2}\log\left|\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\left(\frac{1}{\sigma^2+v}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}}\frac{1}{\sigma^2+v_j}\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right)+\mathbf{I}_m\right| \tag{93}$$

$$+\frac{\lambda}{2\left(\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)+\sigma^2\right)}v-\frac{1}{2}\lambda\log\left(\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)+\sigma^2+v\right)+c$$

$$= \min_{v\in\mathbb{R}_+} \frac{1}{2}\log\left|\frac{1}{\sigma^2+v}\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\sum_{j\in\mathcal{K}\setminus\{i\}}\frac{1}{\sigma^2+v_j}\mathbf{e}_j\mathbf{e}_j^\mathsf{T}+\mathbf{I}_m\right| \tag{94}$$

$$+\frac{\lambda}{2\left(\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)+\sigma^2\right)}v-\frac{1}{2}\lambda\log\left(\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)+\sigma^2+v\right)+c$$

$$\Leftrightarrow \min_{v\in\mathbb{R}_+} \log\left|\frac{1}{\sigma^2+v}\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\sum_{j\in\mathcal{K}\setminus\{i\}}\frac{1}{\sigma^2+v_j}\mathbf{e}_j\mathbf{e}_j^\mathsf{T}+\mathbf{I}_m\right| \tag{95}$$

$$+\frac{\lambda}{\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)+\sigma^2}v-\lambda\log\left(\mathrm{tr}\left(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)+\sigma^2+v\right),$$

where (92) follows from eliminating the term $\left|\sigma^2\mathbf{I}_m+v\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right|$ in the term $\log\frac{\left|\mathbf{\Sigma}_{YY}+v\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right|}{\left|\sigma^2\mathbf{I}_m+v\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right|}$ and denoting the term that is not a function of v as c. This completes the proof. □

M Proof of Theorem 22

The proof of Theorem 22 is obtained by noting that the utility function of $\mathcal{G}_1$ is convex.

Proof. According to Proposition 13, the utility function of $\mathcal{G}_1$ is convex. Therefore, the solution for the minimization problem in (7.37) is obtained in the first critical point. That

is

$$\begin{aligned}\frac{\partial \phi_i^1}{\partial v} =& \frac{\partial}{\partial v}(1-\lambda)\log\left|\mathbf{\Sigma}_{YY}+v\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right|-\log(\sigma^2+v)+\lambda v\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) \\ =&(1-\lambda)\mathrm{tr}\left(\left(\mathbf{\Sigma}_{YY}+v\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right)^{-1}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right)-\frac{1}{\sigma^2+v}+\lambda\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right).\end{aligned} \tag{96}$$

Consider the first term $(1-\lambda)\mathrm{tr}\left(\left(\mathbf{\Sigma}_{YY}+v\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right)^{-1}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right)$ in (96). Let $\mathbf{A}\triangleq\mathbf{\Sigma}_{YY}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}$ and $\alpha_i\triangleq\mathrm{tr}\left(\left(\mathbf{\Sigma}_{YY}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right)^{-1}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right)$. It follows that

$$(1-\lambda)\mathrm{tr}\left(\left(\mathbf{\Sigma}_{YY}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}+v\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right)^{-1}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) \tag{97}$$

$$=(1-\lambda)\mathrm{tr}\left(\left(\mathbf{A}+v\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right)^{-1}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) \tag{98}$$

$$=(1-\lambda)\mathrm{tr}\left(\left(\mathbf{A}^{-1}-\frac{v}{1+v\mathbf{e}_i^{\mathsf{T}}\mathbf{A}^{-1}\mathbf{e}_i}\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\mathbf{A}^{-1}\right)\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) \tag{99}$$

$$=(1-\lambda)\mathrm{tr}\left(\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right)-(1-\lambda)\frac{v}{1+v\mathbf{e}_i^{\mathsf{T}}\mathbf{A}^{-1}\mathbf{e}_i}\mathrm{tr}\left(\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) \tag{100}$$

$$=(1-\lambda)\frac{\mathrm{tr}\left(\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right)}{1+v\mathrm{tr}\left(\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right)} \tag{101}$$

$$=(1-\lambda)\frac{\alpha_i}{1+v\alpha_i}. \tag{102}$$

Hence, the first derivative of the utility in (96) is

$$\begin{aligned}&\frac{\partial \phi_i^1}{\partial v}(1-\lambda)\log\left|\mathbf{\Sigma}_{YY}+v\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}+\sum_{j\in\mathcal{K}\setminus\{i\}}v_j\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right|-\log(\sigma^2+v)+\lambda v\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right) \\ =&(1-\lambda)\frac{\alpha_i}{1+v\alpha_i}-\frac{1}{\sigma^2+v}+\lambda\beta_i,\end{aligned} \tag{103}$$

where $\beta_i\triangleq\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^{\mathsf{T}}\right)$. Let the first derivative in (96) equals to 0. The following holds

$$\beta_i\alpha_i v^2+(\beta_i+\alpha_i\sigma^2\beta_i-\alpha_i)v+\beta_i\sigma^2-\alpha_i\sigma^2+\frac{\alpha_i\sigma^2-1}{\lambda}=0 \tag{104}$$

Note that (104) is a quadratic form with two solutions as follows:

$$v_1=\frac{-(\beta_i+\alpha_i\sigma^2\beta_i-\alpha_i)+\sqrt{(\beta_i+\alpha_i\sigma^2\beta_i-\alpha_i)^2-4\beta_i\alpha_i(\beta_i\sigma^2-\alpha_i\sigma^2+\frac{\alpha_i\sigma^2-1}{\lambda})}}{2\beta_i\alpha_i} \tag{105}$$

$$v_2=\frac{-(\beta_i+\alpha_i\sigma^2\beta_i-\alpha_i)-\sqrt{(\beta_i+\alpha_i\sigma^2\beta_i-\alpha_i)^2-4\beta_i\alpha_i(\beta_i\sigma^2-\alpha_i\sigma^2+\frac{\alpha_i\sigma^2-1}{\lambda})}}{2\beta_i\alpha_i} \tag{106}$$

Note that the utility function is convex in v and $v \in \mathbb{R}_+$. The only critical point is positive, i.e. the solution in (105). Therefore, the best response for the i-th player i is

$$v^* = \frac{-(\beta_i + \alpha_i\sigma^2\beta_i - \alpha_i) + \sqrt{(\beta_i + \alpha_i\sigma^2\beta_i - \alpha_i)^2 - 4\beta_i\alpha_i(\beta_i\sigma^2 - \alpha_i\sigma^2 + \frac{\alpha_i\sigma^2 - 1}{\lambda})}}{2\beta_i\alpha_i} \quad (107)$$

This competes the proof. □

N Proof of Theorem 23

The proof of Theorem 23 is obtained by noting that the utility function of $\mathcal{G}_2$ is convex.

Proof. According to Proposition 15, the utility function of $\mathcal{G}_2$ is convex. Therefore, the solution for the optimization problem in (7.55) is obtained in the first critical point. That is

$$\begin{aligned}
&\frac{\partial}{\partial v}\phi_i^2 \\
=&\frac{\partial}{\partial v}\log\left(1 + \frac{\mathrm{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)}{\sigma^2 + v}\right) - \lambda\log\left|\boldsymbol{\Sigma}_{YY} + v\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right| + \lambda v\mathrm{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) \\
=&\frac{-\mathrm{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)}{(\sigma^2 + v)\left(\sigma^2 + v + \mathrm{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)\right)} - \lambda\mathrm{tr}\left(\left(\boldsymbol{\Sigma}_{YY} + v\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right)^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) \\
&+ \lambda\mathrm{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right).
\end{aligned}$$

Consider the second term $\mathrm{tr}\left(\left(\boldsymbol{\Sigma}_{YY} + v\mathbf{e}_i\mathbf{e}_i^\mathsf{T} + \sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right)^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)$. From (98) to (102), it follows that

$$\mathrm{tr}\left(\left(\boldsymbol{\Sigma}_{YY} + \sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T} + v\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right) = \frac{\alpha_i}{1 + v\alpha_i}, \quad (108)$$

where $\alpha_i = \mathrm{tr}\left(\left(\boldsymbol{\Sigma}_{YY} + \sum_{j\in\mathcal{K}\setminus\{i\}} v_j\mathbf{e}_j\mathbf{e}_j^\mathsf{T}\right)^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)$. Let $\beta_i \triangleq \mathrm{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)$. Therefore, the first critical point is obtained from the following equation:

$$\frac{-\mathrm{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)}{(\sigma^2 + v)\left(\sigma^2 + v + \mathrm{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)\right)} - \lambda\frac{\alpha_i}{1 + v\alpha_i} + \lambda\beta_i = 0. \quad (109)$$

This competes the proof. □

O Proof of Theorem 24

Proof. From Proposition 17, the utility function of $\mathcal{G}_3$ in (7.48) is convex. The best response is obtained by taking the derivative of the utility function. Let $h_i \triangleq \text{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)$ and $\gamma_i \triangleq \text{tr}\left(\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\sum_{j\in\mathcal{K}\setminus\{i\}}\frac{1}{\sigma^2+v_j}\mathbf{e}_j\mathbf{e}_j^\mathsf{T}+\mathbf{I}_m\right)^{-1}\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)$. It follows that

$$\begin{aligned}
&\frac{\partial\phi_i^3}{\partial v}\log\left|\frac{1}{\sigma^2+v}\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}+\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\sum_{j\in\mathcal{K}\setminus\{i\}}\frac{1}{\sigma^2+v_j}\mathbf{e}_j\mathbf{e}_j^\mathsf{T}+\mathbf{I}_m\right| \\
&\quad+\frac{\lambda}{\text{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)+\sigma^2}v-\lambda\log\left(\text{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)+\sigma^2+v\right) \qquad (110)\\
=&-\frac{\gamma_i}{(\sigma^2+v)(\sigma^2+v+\alpha_i)}+\frac{\lambda}{h_i+\sigma^2}-\frac{\lambda}{h_i+\sigma^2+v}\\
=&-\frac{\gamma_i}{(\sigma^2+v)(\sigma^2+v+\alpha_i)}+\lambda\frac{v}{(h_i+\sigma^2+v)(h_i+\sigma^2)}
\end{aligned}$$

The only solution of $\frac{\partial\phi_i^3}{\partial v}=0$ such that $v_i\in\mathbb{R}_+$ is given by

$$v^*=-\frac{\alpha_i+2\sigma^2}{3}-\frac{2^{1/3}(-3\alpha_i\sigma_{y_i}\lambda-\alpha_i^2\lambda^2-\alpha_i\sigma^2\lambda^2-\sigma^4\lambda^2)}{3\lambda\xi}+\frac{\xi}{32^{1/3}\lambda}, \qquad (111)$$

where

$$\begin{aligned}
\alpha_i \triangleq&\text{tr}\left(\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\sum_{j\in\mathcal{K}\setminus\{i\}}\frac{1}{\sigma^2+v_j}\mathbf{e}_j\mathbf{e}_j^\mathsf{T}+\mathbf{I}_m\right)^{-1}\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)\\
\sigma_{y_i} \triangleq&\text{tr}\left(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T}\mathbf{e}_i\mathbf{e}_i^\mathsf{T}\right)\\
\xi \triangleq&\left(\theta+\sqrt{4(-3\alpha_i\sigma_{y_i}\lambda-\alpha_i^2\lambda^2-\alpha_i\sigma^2\lambda^2-\sigma^4\lambda^2)^3+\theta^2}\right)^{1/3}\\
\theta \triangleq&-9\alpha_i^2\sigma_{y_i}\lambda^2-18\alpha_i\sigma^2\sigma_{y_i}\lambda^2+27\alpha_i\sigma_{y_i}^2\lambda^2-2\alpha_i^3\lambda^3-3\alpha_i^2\sigma^2\lambda^3+3\alpha_i\sigma^4\lambda^3+2\sigma^8\lambda^3.
\end{aligned}$$

This completes the proof. □

Bibliography

[1] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. ACM Conf. on Comput. and Commun. Security*, Chicago, IL, USA, Nov. 2009, pp. 21–32.

[2] ——, "False data injection attacks against state estimation in electric power grids," *ACM Trans. Info. Syst. Sec*, vol. 14, no. 1, pp. 1–33, May 2011.

[3] A. Bretas, N. Bretas, J. B. London Jr, and B. Carvalho, *Cyber-physical power systems state estimation*. Elsevier, 2021.

[4] A. Abur and A. G. Exposito, *Power system state estimation: Theory and implementation*. CRC press, Mar. 2004.

[5] J. J. Grainger and W. D. Stevenson, *Power system analysis*. McGraw-Hill, 1994.

[6] O. Vuković, K. C. Sou, G. Dán, and H. Sandberg, "Network-layer protection schemes against stealth attacks on state estimators in power systems," in *Proc. IEEE Int. Conf. on Smart Grid Comm.*, Brussels, Belgium, Oct. 2011, pp. 184–189.

[7] A. Tajer, S. Kar, H. V. Poor, and S. Cui, "Distributed joint cyber attack detection and state recovery in smart grids," in *Proc. IEEE Int. Conf. on Smart Grid Comm.*, Brussels, Belgium, Oct. 2011, pp. 202–207.

[8] S. Cui, Z. Han, S. Kar, T. T. Kim, H. V. Poor, and A. Tajer, "Coordinated data-injection attack and detection in the smart grid: A detailed look at enriching detection solutions," *IEEE Signal Process. Mag*, vol. 29, no. 5, pp. 106–115, Aug. 2012.

[9] M. Ozay, I. Esnaola, F. T. Y. Vural, S. R. Kulkarni, and H. V. Poor, "Sparse attack construction and state estimation in the smart grid: Centralized and distributed models," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1306–1318, Jul. 2013.

[10] I. Esnaola, S. M. Perlaza, and H. V. Poor, "Equilibria in data injection attacks," in *Proc. IEEE Global Conference on Signal and Information Processing*, Atlanta, GA, USA, Dec. 2014, pp. 779–783.

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

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

[13] I. Esnaola, S. M. Perlaza, H. V. Poor, and O. Kosut, "Maximum distortion attacks in electricity grids," *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 2007–2015, Jul. 2016.

[14] M. Ozay, I. Esnaola, F. T. Yarman Vural, S. R. Kulkarni, and H. V. Poor, "Machine learning methods for attack detection in the smart grid," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 8, pp. 1773–1786, Aug. 2016.

[15] A. Tajer, S. M. Perlaza, and H. V. Poor, *Advanced Data Analytics for Power Systems*. Cambridge University Press, 2021.

[16] K. Sun, I. Esnaola, S. M. Perlaza, and H. V. Poor, "Information-theoretic attacks in the smart grid," in *Proc. IEEE Int. Conf. on Smart Grid Comm.*, Dresden, Germany, Oct. 2017, pp. 455–460.

[17] ——, "Stealth attacks on the smart grid," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1276–1285, Aug. 2019.

[18] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[19] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 1999.

[20] X. Ye, I. Esnaola, S. M. Perlaza, and R. F. Harrison, "Information theoretic data injection attacks with sparsity constraints," in *Proc. IEEE Int. Conf. on Smart Grid Comm.*, Phoenix, USA, Nov. 2020, pp. 1–6.

[21] C. Genes, I. Esnaola, S. M. Perlaza, L. F. Ochoa, and D. Coca, "Recovering missing data via matrix completion in electricity distribution systems," in *IEEE Workshop on Signal Processing Advances in Wireless Communications*, Edinburgh, UK, Jul. 2016, pp. 1–6.

[22] ——, "Robust recovery of missing data in electricity distribution systems," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 4057–4067, Jun. 2018.

[23] M. J. Osborne and A. Rubinstein, *A course in game theory*. MIT press, 1994.

[24] W. Saad, Z. Han, H. V. Poor, and T. Basar, "Game-theoretic methods for the smart grid: An overview of microgrid systems, demand-side management, and smart grid communications," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 86–105, 2012.

[25] I. Esnaola, S. M. Perlaza, and H. V. Poor, "Equilibria in data injection attacks," in *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2014, pp. 779–783.

[26] X. Ye, I. Esnaola, S. M. Perlaza, and R. F. Harrison, "Stealth data injection attacks with sparsity constraints," *IEEE Trans. Smart Grid*, pp. 1–1, 2023.

[27] ——, "An information theoretic vulnerability metric for data integrity attacks on smart grids," *arXiv preprint arXiv:2211.02538*, 2022.

[28] H. O. Hartley, "The modified gauss-newton method for the fitting of non-linear regression functions by least squares," *Technometrics*, vol. 3, no. 2, pp. 269–280, 1961.

[29] Y. Wang, "Gauss-newton method," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 4, no. 4, pp. 415–420, 2012.

[30] A. S. Dobakhshari, V. Terzija, and S. Azizi, "Normalized deleted residual test for identifying interacting bad data in power system state estimation," *IEEE Trans. on Power Systems*, 2022.

[31] I. Esnaola, S. M. Perlaza, H. V. Poor, and O. Kosut, "Maximum distortion attacks in electricity grids," *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 2007–2015, Jul. 2016.

[32] R. Zhang and P. Venkitasubramaniam, "Stealthy control signal attacks in linear quadratic gaussian control systems: Detectability reward tradeoff," *IEEE Trans. on Inf. Forensics and Security*, vol. 12, no. 7, pp. 1555–1570, Feb. 2017.

[33] N. Forti, G. Battistelli, L. Chisci, and B. Sinopoli, "A bayesian approach to joint attack detection and resilient state estimation," in *2016 IEEE 55th Conf. on Decision and Control*. IEEE, Dec. 2016, pp. 1192–1198.

[34] H. V. Poor, *An introduction to signal detection and estimation*. New York, NY, USA: Springer, 1994.

[35] A. G. Tsikalakis and N. D. Hatziargyriou, "Centralized control for optimizing microgrids operation," in *2011 IEEE power and energy society general meeting*. IEEE, 2011, pp. 1–8.

[36] P. Basak, S. Chowdhury, S. H. nee Dey, and S. Chowdhury, "A literature review on integration of distributed energy resources in the perspective of control, protection and stability of microgrid," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 8, pp. 5545–5556, 2012.

[37] H. Lund and W. Kempton, "Integration of renewable energy into the transport and electricity sectors through v2g," *Energy policy*, vol. 36, no. 9, pp. 3578–3587, 2008.

[38] Z. Wang, B. Chen, J. Wang *et al.*, "Decentralized energy management system for networked microgrids in grid-connected and islanded modes," *IEEE Trans. on Smart Grid*, vol. 7, no. 2, pp. 1097–1105, 2015.

[39] T. Yokoyama and T. Nagata, "Comparison of centralized and decentralized systems in power system restoration," *Electrical Engineering in Japan*, vol. 189, no. 2, pp. 26–33, 2014.

[40] G. Dan and H. Sandberg, "Stealth attacks and protection schemes for state estimators in power systems," in *Proc. IEEE Int. Conf. on Smart Grid Comm.*, Gaithersburg, MD, Oct. 2010, pp. 214–219.

[41] S. Bi and Y. J. Zhang, "Defending mechanisms against false-data injection attacks in the power system state estimation," in *IEEE GLOBECOM Workshops*, Houston, USA, Dec. 2011, pp. 1162–1167.

[42] ——, "Graphical methods for defense against false-data injection attacks on power system state estimation," *IEEE Transactions on Smart Grid*, vol. 5, no. 3, pp. 1216–1227, 2014.

[43] K. C. Sou, H. Sandberg, and K. H. Johansson, "On the exact solution to a smart grid cyber-security analysis problem," *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 856–865, Mar. 2013.

[44] M. Ozay, I. Esnaola, F. T. Y. Vural, S. R. Kulkarni, and H. V. Poor, "Sparse attack construction and state estimation in the smart grid: Centralized and distributed models," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1306–1318, Jul. 2013.

[45] K. Sun, I. Esnaola, A. M. Tulino, and H. V. Poor, "Learning requirements for stealth attacks," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Brighton, UK, Apr. 2019, pp. 8102–8106.

[46] M. A. Rahman and H. Mohsenian-Rad, "False data injection attacks with incomplete information against smart power grids," in *2012 IEEE Global Communications Conference (GLOBECOM)*. Anaheim, CA, USA: IEEE, Dec. 2012, pp. 3153–3158.

[47] X. Liu and Z. Li, "Local load redistribution attacks in power systems with incomplete network information," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1665–1676, 2014.

[48] ——, "False data attacks against ac state estimation with incomplete network information," *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2239–2248, 2016.

[49] X. Liu, Z. Bao, D. Lu, and Z. Li, "Modeling of local false data injection attacks with reduced network information," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1686–1696, 2015.

[50] J. Kim, L. Tong, and R. J. Thomas, "Subspace methods for data attack on state estimation: A data driven approach," *IEEE Trans. on Signal Processing*, vol. 63, no. 5, pp. 1102–1114, 2014.

[51] A. Tajer, "False data injection attacks in electricity markets by limited adversaries: Stochastic robustness," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 128–138, 2017.

[52] J. Kim and L. Tong, "On topology attack of a smart grid: Undetectable attacks and countermeasures," *IEEE J. Sel. Areas in Commun.*, vol. 31, no. 7, pp. 1294–1305, 2013.

[53] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, "A framework for cyber-topology attacks: Line-switching and new attack scenarios," *IEEE Trans. on Smart Grid*, vol. 10, no. 2, pp. 1704–1712, 2017.

[54] D.-H. Choi and L. Xie, "Economic impact assessment of topology data attacks with virtual bids," *IEEE Trans. on Smart Grid*, vol. 9, no. 2, pp. 512–520, 2016.

[55] ——, "Impact analysis of locational marginal price subject to power system topology errors," in *2013 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. IEEE, 2013, pp. 55–60.

[56] M. A. Rahman and H. Mohsenian-Rad, "False data injection attacks against nonlinear state estimation in smart power grids," in *2013 IEEE Power & Energy Society General Meeting*. IEEE, Jul. 2013, pp. 1–5.

[57] A. Tajer, "False data injection attacks in electricity markets by limited adversaries: Stochastic robustness," *IEEE Trans. on Smart Grid*, vol. 10, no. 1, pp. 128–138, 2017.

[58] 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, 2011.

[59] L. Jia, J. Kim, R. J. Thomas, and L. Tong, "Impact of data quality on real-time locational marginal price," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 627–636, 2013.

[60] M. A. Rahman, E. Al-Shaer, and R. Kavasseri, "Impact analysis of topology poisoning attacks on economic operation of the smart power grid," in *2014 IEEE 34th International Conference on Distributed Computing Systems*. IEEE, 2014, pp. 649–659.

[61] Y. Yuan, Z. Li, and K. Ren, "Modeling load redistribution attacks in power systems," *IEEE Trans. on Smart Grid*, vol. 2, no. 2, pp. 382–390, 2011.

[62] ——, "Quantitative analysis of load redistribution attacks in power systems," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 9, pp. 1731–1738, 2012.

[63] J. Lin, W. Yu, X. Yang, G. Xu, and W. Zhao, "On false data injection attacks against distributed energy routing in smart grid," in *2012 IEEE/ACM Third International Conference on Cyber-Physical Systems*. IEEE, 2012, pp. 183–192.

[64] R. Deng, G. Xiao, and R. Lu, "Defending against false data injection attacks on power system state estimation," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 1, pp. 198–207, 2015.

[65] J. Chen and A. Abur, "Placement of pmus to enable bad data detection in state estimation," *IEEE Trans. on Power Systems*, vol. 21, no. 4, pp. 1608–1615, 2006.

[66] C. Pei, Y. Xiao, W. Liang, and X. Han, "Pmu placement protection against coordinated false data injection attacks in smart grid," *IEEE Trans. on Industry Applications*, vol. 56, no. 4, pp. 4381–4393, 2020.

[67] Q. Yang, L. Jiang, W. Hao, B. Zhou, P. Yang, and Z. Lv, "Pmu placement in electric transmission networks for reliable state estimation against false data injection attacks," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1978–1986, 2017.

[68] M. Talebi, C. Li, and Z. Qu, "Enhanced protection against false data injection by dynamically changing information structure of microgrids," in *2012 IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM)*. IEEE, 2012, pp. 393–396.

[69] Y. Huang, H. Li, K. A. Campbell, and Z. Han, "Defending false data injection attack on smart grid network using adaptive cusum test," in *2011 45th Annual Conference on Information Sciences and Systems*. IEEE, 2011, pp. 1–6.

[70] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, "Detecting false data injection attacks on power grid by sparse optimization," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 612–621, 2014.

[71] S. Li, Y. Yılmaz, and X. Wang, "Quickest detection of false data injection attack in wide-area smart grids," *IEEE Trans. on Smart Grid*, vol. 6, no. 6, pp. 2725–2735, 2014.

[72] G. Chaojun, P. Jirutitijaroen, and M. Motani, "Detecting false data injection attacks in ac state estimation," *IEEE Trans. on Smart Grid*, vol. 6, no. 5, pp. 2476–2483, 2015.

[73] X. Liu, P. Zhu, Y. Zhang, and K. Chen, "A collaborative intrusion detection mechanism against false data injection attack in advanced metering infrastructure," *IEEE Trans. on Smart Grid*, vol. 6, no. 5, pp. 2435–2443, 2015.

[74] E. J. Colbert and A. Kott, *Cyber-security of SCADA and other industrial control systems*. Springer, 2016.

[75] Q. Yang, J. Yang, W. Yu, D. An, N. Zhang, and W. Zhao, "On false data-injection attacks against power system state estimation: Modeling and countermeasures," *IEEE Trans. on Parallel Distrib. Syst.*, vol. 25, no. 3, pp. 717–729, Mar. 2013.

[76] S. Arora and B. Barak, *Computational complexity: a modern approach*. Cambridge University Press, 2009.

[77] J. C. Lagarias and A. M. Odlyzko, "Solving low-density subset sum problems," *Journal of the ACM (JACM)*, vol. 32, no. 1, pp. 229–246, 1985.

[78] E. L. Lawler, "The traveling salesman problem: a guided tour of combinatorial optimization," *Wiley-Interscience Series in Discrete Mathematics*, 1985.

[79] D. S. Hochba, "Approximation algorithms for np-hard problems," *ACM Sigact News*, vol. 28, no. 2, pp. 40–52, 1997.

[80] Z. Hussain and J. Shawe-Taylor, "Theory of matching pursuit," *Advances in neural information processing systems*, vol. 21, 2008.

[81] T. Zhang, "Sparse recovery with orthogonal matching pursuit under rip," *IEEE trans. on information theory*, vol. 57, no. 9, pp. 6215–6221, 2011.

[82] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[83] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proceedings of 27th Asilomar conf. on signals, syst. and computers*, Pacific Grove, USA, Nov. 1993, pp. 40–44.

[84] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE trans. on inf. theory*, vol. 51, no. 12, pp. 4203–4215, Nov. 2005.

[85] E. J. Candes *et al.*, "The restricted isometry property and its implications for compressed sensing," *Comptes rendus mathematique*, vol. 346, no. 9-10, pp. 589–592, Feb. 2008.

[86] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, Nov. 2010.

[87] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *Journal of Fourier analysis and applications*, vol. 14, no. 5-6, pp. 877–905, Oct. 2008.

[88] A. Zymnis, S. Boyd, and E. Candes, "Compressed sensing with quantized measurements," *IEEE Signal Process. Lett.*, vol. 17, no. 2, pp. 149–152, Oct. 2009.

[89] H. Sandberg, A. Teixeira, and K. H. Johansson, "On security indices for state estimators in power networks," in *First Workshop on Secure Control Systems*, Stockholm, Sweden, Apr. 2010, pp. 8102–8106.

[90] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on the smart grid state estimation: Attack strategies and countermeasures," in *Proc. IEEE Int. Conf. on Smart Grid Comm.*, Gaithersburg, MD, Oct. 2010, pp. 220–225.

[91] M. Prais and A. Bose, "A topology processor that tracks network modifications," *IEEE trans. on Power Systems*, vol. 3, no. 3, pp. 992–998, 1988.

[92] L. Jia, J. Kim, R. J. Thomas, and L. Tong, "Impact of data quality on real-time locational marginal price," *IEEE Trans. on Power Systems*, vol. 29, no. 2, pp. 627–636, 2013.

[93] X. Liu, Z. Li, X. Liu, and Z. Li, "Masking transmission line outages via false data injection attacks," *IEEE Trans. on Information Forensics and Security*, vol. 11, no. 7, pp. 1592–1602, 2016.

[94] Z. Li, M. Shahidehpour, A. Alabdulwahab, and A. Abusorrah, "Analyzing locally coordinated cyber-physical attacks for undetectable line outages," *IEEE Trans. on Smart Grid*, vol. 9, no. 1, pp. 35–47, 2016.

[95] J. Zhang and L. Sankar, "Physical system consequences of unobservable state-and-topology cyber-physical attacks," *IEEE Trans. on Smart Grid*, vol. 7, no. 4, 2016.

[96] Z. Li, M. Shahidehpour, A. Alabdulwahab, and A. Abusorrah, "Bilevel model for analyzing coordinated cyber-physical attacks on power systems," *IEEE Trans. on Smart Grid*, vol. 7, no. 5, pp. 2260–2272, 2015.

[97] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 1999.

[98] J. Hou and G. Kramer, "Effective secrecy: Reliability, confusion and stealth," in *2014 IEEE International Symposium on Information Theory*, Honolulu, USA, Aug. 2014, pp. 601–605.

[99] Q. Li, T. Cui, Y. Weng, R. Negi, F. Franchetti, and M. D. Ilic, "An information-theoretic approach to PMU placement in electric power systems," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 446–456, Dec. 2012.

[100] M. Arrieta and I. Esnaola, "Smart meter privacy via the trapdoor channel," in *IEEE Int. Conf. on Smart Grid Comm.* Dresden, Germany: IEEE, Apr. 2018, pp. 277–282.

[101] B. Qin, J. Huang, Q. Wang, X. Luo, B. Liang, and W. Shi, "Cecoin: A decentralized PKI mitigating MitM attacks," *Future Generation Computer Systems*, vol. 107, pp. 805–815, 2020.

[102] L. Zhou, X. Xiong, J. Ernstberger, S. Chaliasos, Z. Wang, Y. Wang, K. Qin, R. Wattenhofer, D. Song, and A. Gervais, "Sok: Decentralized finance (DeFi) attacks," *Cryptology ePrint Archive*, 2022.

[103] D. Monderer and L. S. Shapley, "Potential games," *Games and economic behavior*, vol. 14, no. 1, pp. 124–143, 1996.

[104] P. Chen, S. Liu, B. Chen, and L. Yu, "Multi-agent reinforcement learning for decentralized resilient secondary control of energy storage systems against dos attacks," *IEEE Trans. Smart Grid*, vol. 13, no. 3, pp. 1739–1750, 2022.

[105] H. H. Alhelou, M. E. H. Golshan, and N. D. Hatziargyriou, "A decentralized functional observer based optimal LFC considering unknown inputs, uncertainties, and cyber-attacks," *IEEE Trans. Power Systems*, vol. 34, no. 6, pp. 4408–4417, 2019.

[106] B. Alangot, D. Reijsbergen, S. Venugopalan, P. Szalachowski, and K. S. Yeo, "Decentralized and lightweight approach to detect eclipse attacks on proof of work blockchains," *IEEE Trans. on Network and Service Management*, vol. 18, no. 2, pp. 1659–1672, 2021.

[107] L. Zha, E. Tian, X. Xie, Z. Gu, and J. Cao, "Decentralized event-triggered H_∞ control for neural networks subject to cyber-attacks," *Information Sciences*, vol. 457, pp. 141–155, 2018.

[108] D. Cai, X. He, Z. Yu, L. Wang, G. Xie, and Q. Ai, "3D power-map for smart grids—an integration of high-dimensional analysis and visualization," in *International Conference on Renewable Power Generation (RPG 2015)*. IET, 2015, pp. 1–5.

[109] I. Shomorony and A. S. Avestimehr, "Worst-case additive noise in wireless networks," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3833–3847, Jun. 2013.

[110] P. Lévy, "Propriétés asymptotiques des sommes de variables aléatoires enchaînées," *J. Math. Pures Appl.*, vol. 14, pp. 109–128, 1935.

[111] H. Cramér, "Über eine Eigenschaft der normalen Verteilungsfunktion," *Math. Z.*, vol. 41, pp. 405–414, 1936.

[112] U. of Washington, "Power systems test case archive," 1999. [Online]. Available: https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/radiometric

[113] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Trans. Power Syst*, vol. 26, no. 1, pp. 12–19, Feb. 2010.

[114] I. Esnaola, A. M. Tulino, and J. Garcia-Frias, "Linear analog coding of correlated multivariate Gaussian sources," *IEEE Trans. on Commun.*, vol. 61, no. 8, pp. 3438–3447, Aug. 2013.

[115] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[116] G. A. Seber, *A matrix handbook for statisticians*. John Wiley & Sons, 2008, vol. 15.