This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2852774, IEEE Transactions on Automatic Control

Power Grid AC-based State Estimation: Vulnerability Analysis Against Cyber Attacks

Ming Jin[1], Javad Lavaei[1], and Karl Henrik Johansson[2]

***Abstract*—To ensure grid efficiency and reliability, power system operators continuously monitor the operational characteristics of the grid through a critical process called state estimation (SE), which performs the task by filtering and fusing various measurements collected from grid sensors. This study analyzes the vulnerability of the key operation module, namely AC-based SE, against potential cyber attacks on data integrity, also known as false data injection attack (FDIA). A general form of FDIA can be formulated as an optimization problem, whose objective is to find a stealthy and sparse data injection vector on the sensor measurements with the aim of making the state estimate spurious and misleading. Due to the nonlinear AC measurement model and the cardinality constraint, the problem includes both continuous and discrete nonlinearities. To solve the FDIA problem efficiently, we propose a novel convexification framework based on semidefinite programming (SDP). By analyzing a globally optimal SDP solution, we delineate the "attackable region" for any given set of measurement types and grid topology, where the spurious state can be falsified by FDIA. Furthermore, we prove that the attack is stealthy and sparse, and derive performance bounds. Simulation results on various IEEE test cases indicate the efficacy of the proposed convexification approach. From the grid protection point of view, the results of this study can be used to design a security metric for the current practice against cyber attacks, redesign the bad data detection scheme, and inform proposals of grid hardening. From a theoretical point of view, the proposed framework can be used for other nonconvex problems in power systems and beyond.**

***Index Terms*—State estimation, nonconvex optimization, convexification, semidefinite programming, false data injection attack, cyber attack, power system, resilience, security**

I. Introduction

THE convergence of automation and information technology has enhanced reliability, efficiency, and agility of the modern grid [1]. Managed by supervisory control and data acquisition (SCADA) systems, a wealth of sensor data from transmission and distribution infrastructures are collected and filtered in order to facilitate a key procedure known as power system state estimation (SE), which is conducted on a regular basis (e.g., every few minutes), as shown in Fig. 1 [2]–[5]. The outcome presents system operators with essential information about the real-time operating status to improve situational awareness, make economic decisions, and take contingency actions in response to potential threats that could endanger the grid reliability [6].

In smart grid where information is sent via remote terminal units (RTUs), maintaining the security of the communication network is imperative to guard against system intrusion and ensure operational integrity [3], [8], [9]. However, traditional approaches such as security software, firewalls, and "air gaps", i.e., no connection between systems, are recognized as inadequate in the face of growing likelihood of breaches and cyber threat, such as the 2015 cyber attack on Ukraine's electricity infrastructure [10]. In a recent report from the National Academies of Sciences, Engineering, and Medicine, titled "Enhancing the resilience of the nation's electricity system", the committee concluded that the United States' electric grid is vulnerable to a range of threats, among which terrorism and cyber attacks are most severe and could potentially cause long-term and widespread blackouts [6]. A process called "envisioning process" is recommended to improve the cyber security and resilience, which stresses the importance of "anticipating myriad ways in which the system might be disrupted and the many social, economic, and other consequences of such disruptions".

The objective of this study is to analyze power grid vulnerability against cyber attack – more specifically, one critical class of threat known as false data injection attack (FDIA), which attempts to stealthily modify data to introduce error into grid SE (Fig. 1) [11], [12]. To stage an FDIA, the attacker needs to compromise power measurements by hacking the communication with SCADA. Previous works [11], [13]–[18] have demonstrated that a stealth FDIA is possible to evade bad data detection (BDD) by the control center, and can cause potential damages of load shedding [17], economic loss [12], [19], and even blackouts [20]. While these works have primarily studied a simplified power flow model, i.e., DC model [11], [13]–[18], [21], [22], an FDIA based on a more accurate AC model is within the realm of possibility [2], [23], [24]. In a system where measurements are nonlinear functions of the state parameters, it is usually not easy to construct a state that evades BDD. Indeed, DC-based FDIA can be easily detected by AC-based BDD [8], [25]. On the other hand, the nonlinearity of equality power-flow constraints also makes the co-existence of multiple states and spurious solutions possible, which is a fundamental reason why an AC-based FDIA with

[1] Ming Jin and Javad Lavaei are with the Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA 94720, USA. Email: {jinming,lavaei}@berkeley.edu

[2] Karl H. Johansson is with the School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden. Email: kallej@kth.se

*This work was supported by DARPA YFA, ONR YIP Award, ARO Award, AFOSR YIP Award, and NSF CAREER Award. Johansson was partially supported by the Knut and Alice Wallenberg Foundation, the Swedish Strategic Research Foundation, and the Swedish Research Council. Parts of this work have appeared in the conference paper "Ming Jin, Javad Lavaei, and Karl Henrik Johansson, A semidefinite programming relaxation under false data injection attacks against power grid AC state estimation, IEEE Allerton Conference on Communication, Control, and Computing, 2017".

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2852774, IEEE Transactions on Automatic Control

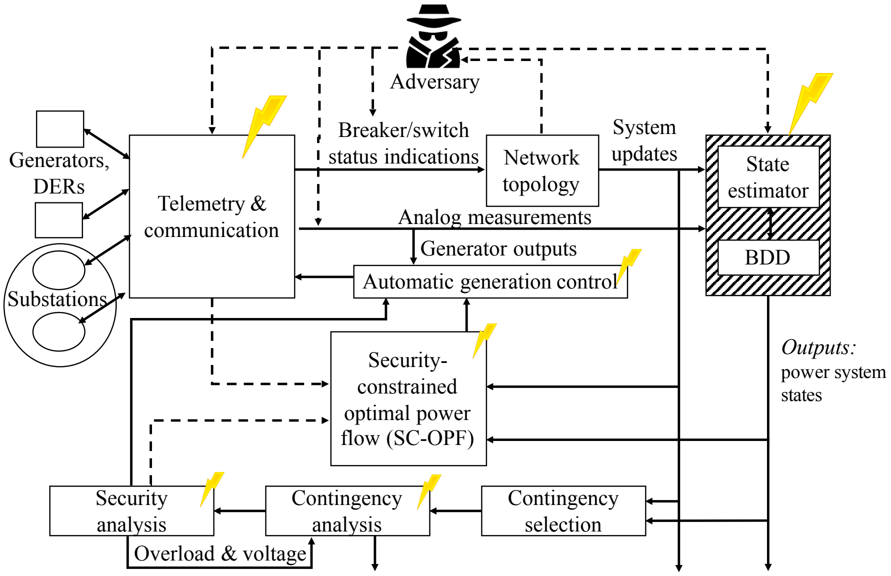


Fig. 1: Illustration of power system operation and its vulnerability to cyber attack (adapted from [7]). With unfettered access to the communication network and grid information system through cyber intrusion, an adversary would be able to stage an attack on the system without any physical sabotage, by simply injecting false data to the state estimator to impact the decision making for the system.

sparse attacks is feasible and perhaps more detrimental than an DC-based FDIA. Once constructed, this new class of attacks could be hard to detect by existing methods. Thus, it is vital to understand its mechanism and devise protection/detection methods to thwart such attacks.

A. Related Work

Potential adversarial FDIA strategies have been addressed in previous works on power system vulnerability analysis [11], [14], [17], [25], [26]. The negative impacts and possible defense mechanisms have also been studied [13], [14], [17], [20], [23]. From a practitioner's point of view, there are mainly two categories, based on either DC or AC models [12], [19]. For DC-FDIA, an unobservability condition was derived and the attack was numerically shown to be sparse [11], [14], [17]. Distributed DC-FDIA with partial knowledge about the topology was considered in [18] and [8]. The vulnerability was quantified by the minimum number of sensors needed to compromise in order to stage stealth FDIA [13], [14], [16]. This can be formulated as a minimum cardinality problem, where different algorithms have been proposed for efficient computation [21], [22]. As for the attack impact, FDIA has been studied on the electric market [15] and load redistribution [17] to show significant financial losses.

Only a few works have been published on AC-based FDIA, due to the recognized complexity of nonlinear systems [3], [25]. The paper [26] introduced a graph-based algorithm to identify a set of compromised sensors that suffices to construct an unobservable attack; however, this only offers an *upper bound* on the cardinality, rather than resource-constrained sparsity. The work [25] studied AC-based FDIA based on linearization around the target state under the assumption that SE is obtained by a specific algorithm, which could be too stringent in practice. Joint cyber and physical attacks on power grids has been studied in [27], which has been extended to a simplified AC power flow model [28] and to the false data injection scenarios [29]. Robust AC-based state estimation has been investigated to guard against bad data and adversarial injections [4], [23], [30]–[32] (see, e.g., [24] for a review on this subject); however, there are no guarantees to detect stealthy injections. As validated in the experiments of Section IV, attacks planned by solving the AC-based FDIA can evade BDD even when robust SE methods are employed.

Differentiated from prior literature, this study is the *first of its kind* to solve a general FDIA for the AC-based SE, with theoretical guarantees of sparsity and stealthiness.

B. Contributions

This study focuses on the problem of FDIA under an AC model, which can be formulated as an optimization problem with a quadratic objective function subject to quadratic equality constraints and a cardinality condition. This problem can be written as

$$\begin{aligned} \min_{\tilde{\mathbf{v}}\in\mathbb{C}^{n_b},\mathbf{b}\in\mathbb{R}^{n_m}} \quad & h(\tilde{\mathbf{v}}) \\ \text{s. t.} \quad & \mathbf{f}(\tilde{\mathbf{v}}) = \mathbf{m} + \mathbf{b} \qquad \text{(NC-FDIA)} \\ & \|\mathbf{b}\|_0 \le c \end{aligned}$$

where the variable $\tilde{\mathbf{v}} \in \mathbb{C}^{n_b}$ is a complex-valued vector of dimension n_b, $\mathbf{b} \in \mathbb{R}^{n_m}$ is a real-valued vector of dimension n_m, $h(\tilde{\mathbf{v}})$ is a quadratic function of $\tilde{\mathbf{v}}$, $\mathbf{f}(\tilde{\mathbf{v}}) \in \mathbb{R}^{n_m}$ is a real-valued vector function with entries being quadratic in $\tilde{\mathbf{v}}$, and $\|\mathbf{b}\|_0$ is the cardinality of $\mathbf{b}$ that is upper limited by a positive integer c. With respect to FDIA, $\tilde{\mathbf{v}}$ is the spurious state, $\mathbf{m}$ includes sensor measurements, $\mathbf{b}$ is the sparsity-constrained attack vector, $h(\cdot)$ is the FDIA objective, and $\mathbf{f}(\cdot)$ is the AC-model measurement function. Due to the quadratic equality constraints as well as the cardinality constraint, (NC-FDIA) is nonconvex. By investigating the least-effort strategy from the attacker's perspective, this study provides a realistic metric for the grid security based on the number of individual sensors required to thwart an FDIA. This broadens the perspectives on power system security and vulnerability analysis. The results also motivate protection mechanisms for AC-based SE, such as the redesign of BDD [33]. The main contributions of this work are as follows:

- Formulation of a novel convexification framework based on SDP to solve the AC-based FDIA problem (NC-FDIA) for a near-globally optimal strategy;
- Analysis of the outcome of the SDP framework from the perspectives of the attackable region, attack stealthiness, and performance bounds;
- Simulation study on an array of power systems to illustrate that the planned attack is sparse and stealthy.

We also note that the presented method has both practical and theoretical implications on solving real-world nonlinear and nonconvex problems that can be formulated in the abstract form (NC-FDIA). This work extends the conference paper [34] to a more general scenario that includes both power branch and nodal measurements, in addition to a new analysis of the attackable region, attack stealthiness and performance bounds.

C. Organization

The rest of the paper is organized as follows. We will first introduce the notations used throughout the paper in the following section. Sec. II provides an overview of the vulnerability issue of AC-based SE. More specifically, we will introduce the power system modeling, AC-based SE methods, and a general framework of FDIA for AC-based SE. Since the presented framework is nonconvex, a convexification framework based on SDP is proposed in Sec. III. We will analyze the (globally) optimal solution of this SDP in terms of the "attackable region" (Sec. III-B) and performance bounds (Sec. III-C). Experimental results on several IEEE bus systems are discussed in Sec. IV. Conclusions are drawn in Sec. V.

D. Notations

Set notations. We use $\mathbb{R}$ and $\mathbb{C}$ as the sets of real and complex numbers, and $\mathbb{S}^n$ and $\mathbb{H}^n$ to represent the spaces of $n \times n$ real symmetric matrices and $n \times n$ complex Hermitian matrices, respectively. A set of indices $\{1, 2, ..., k\}$ is denoted by $[k]$. The set cardinality Card$(\cdot)$ is the number of elements in a set. The support of a vector $\mathbf{x}$, denoted as supp$\,(\mathbf{x})$, is the set of indices of the nonzero entries of $\mathbf{x}$. For a set $\mathcal{S} \subset \mathbb{R}^n$, we use $\mathcal{S}^c = \mathbb{R}^{n_m} \setminus \mathcal{S}$ to denote its complement. The notation int Γ is used to represent the interior of the set Γ.

Matrix notations. Vectors are shown by bold letters, and matrices are shown by bold and capital letters. The symbols $\mathbf{0}_n$, $\mathbf{1}_n$, $\mathbf{0}_{m\times n}$, $\mathbf{I}_{n\times n}$ denote the $n \times 1$ zero vector, $n \times 1$ one vector, $m \times n$ zero matrix, and $n \times n$ identity matrix, respectively. Let $[\mathbf{x}]_i$ denote the i-th element of vector $\mathbf{x}$. For an $m \times n$ matrix $\mathbf{W}$, let $\mathbf{W}[\mathcal{X}, \mathcal{Y}]$ denote the submatrix of $\mathbf{W}$ whose rows are chosen from $\mathcal{X} \in [m]$ and whose columns are chosen from $\mathcal{Y} \in [n]$. The notation $\mathbf{W} \succeq 0$ indicates that $\mathbf{W}$ is Hermitian and positive semidefinite (PSD), and $\mathbf{W} \succ 0$ indicates that $\mathbf{W}$ is Hermitian and positive definite.

Operator notations. The symbols $(\cdot)^\top$ and $(\cdot)^*$ represent the transpose and conjugate transpose operators. We use $\Re(\cdot)$, $\Im(\cdot)$, trace$\,(\cdot)$, and $\det(\cdot)$ to denote the real part, imaginary part, trace, and determinant of a scalar/matrix. The dot product is represented by $\mathbf{x}_1 \cdot \mathbf{x}_2 = \mathbf{x}_1^\top \mathbf{x}_2$, for $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$. The imaginary unit is denoted as $\mathbb{i}$. The notations $\angle x$ and $|x|$ indicate the angle and magnitude of a complex scalar; moreover, $\angle \mathbf{x}$ and $|\mathbf{x}|$ are defined based on the angles and magnitudes of all entries of the vector $\mathbf{x}$. For a convex function $g(\mathbf{x})$, we use $\partial g(\mathbf{x})$ to denote its subgradient. The notations $\|\mathbf{x}\|_0$, $\|\mathbf{x}\|_1$, $\|\mathbf{x}\|_2$ and $\|\mathbf{x}\|_\infty$ show the cardinality, 1-norm, 2-form and ∞-norm of $\mathbf{x}$.

II. Vulnerability of AC-based state estimation

A. Power system modeling

We model the electric grid as a graph $\mathcal{G} := \{\mathcal{N}, \mathcal{L}\}$, where $\mathcal{N} := [n_b]$ and $\mathcal{L} := [n_l]$ represent its set of buses and branches. Denote the admittance of each branch $l \in \mathcal{L}$ that connects bus s and bus t as y_{st}. The mathematical framework of this work applies to more detailed models with shunt elements and transformers; but to streamline the presentation, these are not considered in the theoretical analysis of this paper. The grid topology is encoded in the bus admittance matrix $\mathbf{Y} \in \mathbb{C}^{n_b \times n_b}$, as well as the *from* and *to* branch admittance matrices $\mathbf{Y}_f \in \mathbb{C}^{n_l \times n_b}$ and $\mathbf{Y}_t \in \mathbb{C}^{n_l \times n_b}$, respectively. To illustrate these definitions, consider the simple 2-bus system given in Fig. 2. The bus admittance matrix can be written as

$$\mathbf{Y} = \begin{bmatrix} y_k + y_{kk'} & -y_{kk'} \\ -y_{kk'} & y_{k'} + y_{kk'} \end{bmatrix},$$

where $y_{kk'}$ is the admittance of the branch that connects bus k to bus k', and y_k (resp. y'_k) accounts for the admittance of the load as well as the admittance-to-ground at bus k (resp. bus k'). The branch admittance matrices are given by:

$$\mathbf{Y}_f = \begin{bmatrix} y_k + y_{kk'} & -y_{kk'} \end{bmatrix}, \mathbf{Y}_t = \begin{bmatrix} -y_{kk'} & y_{k'} + y_{kk'} \end{bmatrix}.$$

A general procedure for the construction of $\mathbf{Y}$, $\mathbf{Y}_f$ and $\mathbf{Y}_t$ can be found in [35, Ch. 3].

The power system state is described by the bus voltage vector $\mathbf{v} = \begin{bmatrix} v_1, ..., v_{n_b} \end{bmatrix}^\top \in \mathbb{C}^{n_b}$, where $v_k \in \mathbb{C}$ is the complex voltage at bus $k \in \mathcal{N}$ with magnitude $|v_k|$ and phase $\angle v_k$. Given the complex nodal vector, the nodal current injection can be written as $\mathbf{i} = \mathbf{Y}\mathbf{v}$, and the branch currents at the from and to ends of all branches are given by $\mathbf{i}_f = \mathbf{Y}_f\mathbf{v}$ and $\mathbf{i}_t = \mathbf{Y}_t\mathbf{v}$, respectively. Define $\{\mathbf{e}_1, ..., \mathbf{e}_{n_b}\}$ and $\{\mathbf{d}_1, ..., \mathbf{d}_{n_l}\}$ as the sets of canonical vectors in $\mathbb{R}^{n_b}$ and $\mathbb{R}^{n_l}$, respectively. We can derive various types of power and voltage measurements as follows (see Fig. 2 for an illustration):

- *Voltage magnitude.* The voltage magnitude square at bus k is given by $|v_k|^2 = \text{trace}\,(\mathbf{E}_k\mathbf{v}\mathbf{v}^*)$, where $\mathbf{E}_k := \mathbf{e}_k\mathbf{e}_k^\top$.

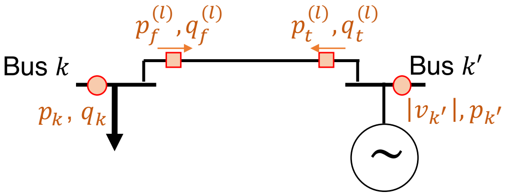


Fig. 2: Illustration of a simple 2-bus system with $n_b = 2$ and $n_l = 1$. Bus k is connected to a load with measurements of real and reactive nodal power injections p_k and q_k. Bus k' is connected to a generator with measurements of real nodal power injection $p_{k'}$ and bus voltage magnitude $|v_{k'}|$. The branch power flows over the line l are also measured.

- *Nodal power injection.* The power injection at bus node k consists of real and reactive powers, $p_k + \mathbb{i}q_k$, where:
$$p_k = \Re\left(i_k^* v_k\right) = \text{trace}\left(\tfrac{1}{2}\left(\mathbf{Y}^*\mathbf{E}_k + \mathbf{E}_k\mathbf{Y}\right)\mathbf{v}\mathbf{v}^*\right)$$
$$q_k = \Im\left(i_k^* v_k\right) = \text{trace}\left(\tfrac{1}{2\mathbb{i}}\left(\mathbf{Y}^*\mathbf{E}_k - \mathbf{E}_k\mathbf{Y}\right)\mathbf{v}\mathbf{v}^*\right).$$
- *Branch power flows.* Given a line $l \in \mathcal{L}$ from node s to node t, the real and reactive power flows in both directions are given by:
$$p_f^{(l)} = \Re\left([\mathbf{i}_f]_l^* v_s\right) = \text{trace}\left(\tfrac{1}{2}\left(\mathbf{Y}_f^*\mathbf{d}_l\mathbf{e}_s^\top + \mathbf{e}_s\mathbf{d}_l^\top\mathbf{Y}_f\right)\mathbf{v}\mathbf{v}^*\right)$$
$$p_t^{(l)} = \Re\left([\mathbf{i}_f]_l^* v_t\right) = \text{trace}\left(\tfrac{1}{2}\left(\mathbf{Y}_f^*\mathbf{d}_l\mathbf{e}_t^\top + \mathbf{e}_t\mathbf{d}_l^\top\mathbf{Y}_f\right)\mathbf{v}\mathbf{v}^*\right)$$
$$q_f^{(l)} = \Im\left([\mathbf{i}_f]_l^* v_s\right) = \text{trace}\left(\tfrac{1}{2\mathbb{i}}\left(\mathbf{Y}_f^*\mathbf{d}_l\mathbf{e}_s^\top - \mathbf{e}_s\mathbf{d}_l^\top\mathbf{Y}_f\right)\mathbf{v}\mathbf{v}^*\right)$$
$$q_t^{(l)} = \Im\left([\mathbf{i}_f]_l^* v_t\right) = \text{trace}\left(\tfrac{1}{2\mathbb{i}}\left(\mathbf{Y}_f^*\mathbf{d}_l\mathbf{e}_t^\top - \mathbf{e}_t\mathbf{d}_l^\top\mathbf{Y}_f\right)\mathbf{v}\mathbf{v}^*\right).$$

Thus, each customary measurement in power systems that belongs to one of the above *measurement types* can be written as:

$$f_i(\mathbf{v}) = \text{trace}\left(\mathbf{M}_i\mathbf{v}\mathbf{v}^*\right), \tag{1}$$

where $\mathbf{M}_i \in \mathbb{H}^{n_b}$ is the Hermitian measurement matrix for the i-th noiseless measurement (it is straightforward to include linear PMU measurements in our analysis as well).

B. AC-based state estimation

The SE problem aims at finding the unknown operating point of a power network, namely $\mathbf{v}$, based on a given set of measurements. During the operation, a set of measurements $\mathbf{m} \in \mathbb{R}^{n_m}$ are acquired:

$$\mathbf{m} = \mathbf{f}(\mathbf{v}) + \mathbf{e} + \mathbf{b}, \tag{2}$$

where $\mathbf{f} : \mathbb{C}^{n_b} \mapsto \mathbb{R}^{n_m}$ is the measurement function whose scalar elements are designated in (1), $\mathbf{e} \in \mathbb{R}^{n_m}$ denotes random noise, and $\mathbf{b} \in \mathbb{R}^{n_m}$ is the bad data error that accounts for sensor failure or adversarial injection [4], [11], [23]. In the case of no bad data error, a common strategy for solving SE is to form a nonlinear weighted least squares problem:

$$\min_{\hat{\mathbf{v}} \in \mathcal{V}} \sum_{i=1}^{n_m} w_i(m_i - f_i(\hat{\mathbf{v}}))^2, \tag{3}$$

where $\mathcal{V}$ is the region of potential operating points, w_i is the inverse variance of sensor i, and $f_i(\hat{\mathbf{v}})$ is given in (1).

In the case that the sensor measurements are not corrupted by bad data and noise, i.e., $\mathbf{b} = \mathbf{e} = \mathbf{0}$, we describe a condition under which a state is "observable" based on the measurement types (matrices) $\mathcal{M} = \{\mathbf{M}_1, ..., \mathbf{M}_{n_m}\}$ [36]–[38]. First, we introduce some notations. Let $\mathcal{O}$ denote the set of all buses except the slack bus. The complex vector $\mathbf{v} \in \mathcal{C}^{n_b}$ can be represented by its real-valued counterpart:

$$\underline{\mathbf{v}} = \left[\Re\left(\mathbf{v}[\mathcal{N}]^\top\right) \quad \Im\left(\mathbf{v}[\mathcal{O}]^\top\right)\right]^\top \in \mathbb{R}^{2n_b-1}.$$

Accordingly, any $n \times n$ Hermitian matrix $\mathbf{M}$ can be characterized by a $(2n-1) \times (2n-1)$ real skew-symmetric matrix:

$$\underline{\mathbf{M}} = \begin{bmatrix} \Re\left(\mathbf{M}[\mathcal{N},\mathcal{N}]\right) & -\Im\left(\mathbf{M}[\mathcal{N},\mathcal{O}]\right) \\ \Im\left(\mathbf{M}[\mathcal{O},\mathcal{N}]\right) & \Re\left(\mathbf{M}[\mathcal{O},\mathcal{O}]\right) \end{bmatrix} \in \mathbb{R}^{(2n-1)\times(2n-1)}.$$

Based on (1) and the above notations, the vector-valued function $\mathbf{f}(\mathbf{v})$ maps the state to a set of noiseless measurements:

$$\mathbf{f}(\mathbf{v}) = \begin{bmatrix} \mathbf{v}^*\mathbf{M}_1\mathbf{v} \\ \vdots \\ \mathbf{v}^*\mathbf{M}_{n_m}\mathbf{v} \end{bmatrix} = \begin{bmatrix} \underline{\mathbf{v}}^\top\underline{\mathbf{M}}_1\underline{\mathbf{v}} \\ \vdots \\ \underline{\mathbf{v}}^\top\underline{\mathbf{M}}_{n_m}\underline{\mathbf{v}} \end{bmatrix} \in \mathbb{R}^{n_m}, \tag{4}$$

whose Jacobian matrix is given by:

$$\mathbf{J}(\mathbf{v}) = \left[(\underline{\mathbf{M}}_1 + \underline{\mathbf{M}}_1^\top)\underline{\mathbf{v}} \quad \cdots \quad (\underline{\mathbf{M}}_{n_m} + \underline{\mathbf{M}}_{n_m}^\top)\underline{\mathbf{v}}\right]^\top. \tag{5}$$

Motivated by the inverse function theorem, which states that the inverse of the function $\mathbf{f}(\mathbf{v})$ exists locally if $\mathbf{J}(\mathbf{v})$ has full column rank, an "observability" definition is introduced below.

Definition 1 (Observability). *A state $\mathbf{v} \in \mathbb{C}^{n_b}$ is observable from a set of measurement types $\mathcal{M}$ if the Jacobian $\mathbf{J}(\mathbf{v})$ has full column rank. For a given set of measurement types $\mathcal{M}$, the observable set $\mathcal{V}(\mathcal{M})$ is the set of all observable states.*

As implied by the observability property and the Kantorovich theorem, if the state $\mathbf{v}$ is observable, then we can find it using the Gauss-Newton method by starting from any initial point sufficiently close to $\mathbf{v}$. More generally, the SE problem (3) may be solved using first-order methods or SDP-based convexification techniques with theoretical guarantees on global optimality in the case where the number of measurements is sufficiently large or some prior information about the solution is available [4], [37]–[39].

As captured by the bad data vector $\mathbf{b}$, the sensor measurements might be corrupted by aberrant data. The common practice is to employ a BDD based on statistical hypothesis testing [3]. Under the null hypothesis that no bad injection exists, namely $b_i = 0$, the residual $(m_i - f_i(\hat{\mathbf{v}}))^2$ should follow the chi-squared distribution, where $\hat{\mathbf{v}}$ is the estimated state and the random error e_i is assumed to be normally distributed. A threshold value is set based on confidence levels to detect large residuals, whose corresponding data are discarded and a new iteration of SE starts. Robust SE approaches such as the least-absolute-value and the least-trimmed-squares based estimators have also been investigated in the literature [4], [30]–[32]. For a least-trimmed-squares analysis, the objective in (3) is replaced by the sum of squared residuals over a subset of data points [30]. For least-absolute-value estimators, the squared errors in (3) are replaced by the absolute errors [4], [31], [32]. These methods are able to sift out randomly occurring bad data; however, it can be ineffective to guard against systematically fabricated bad data, a type of cyber attack known as FDIA.

C. FDIA framework

FDIA is a cyber attack on the data analytic process, where a malicious agent intentionally injects false data $\mathbf{b} \in \mathbb{R}^{n_m}$ into the n_m grid sensors to make system operators believe in an operating state, namely $\tilde{\mathbf{v}}$, other than the true state $\mathbf{v}$ [8], [12]. As an illustrative example (Fig. 3), the operator would be "tricked" if the attacker manages to tamper with certain power flow measurements to generate a fake state estimate of the system.

FDIA differs from randomly occurring bad data in its stealth operation to evade BDD. Existing works have investigated stealth conditions for FDIA on DC-based SE [11], [14]. The following definition of "stealth" is provided to include cases of both DC- and AC-based models.

Definition 2 (Stealth)**.** *An attack* $\mathbf{b}$ *is stealthy under state* $\mathbf{v}$ *if, in the absence of the measurement noise* $\mathbf{e}$*, there exists a nonzero vector* $\mathbf{c}$ *such that* $\mathbf{f}(\mathbf{v}) + \mathbf{b} = \mathbf{f}(\mathbf{v} + \mathbf{c})$.

A definition of observable attack has also been introduced in [23] using set notations. Our definition of stealthy attack is equivalent to an attack that is not observable by the definition given in [23]. The following lemma provides a sufficient condition for AC-based attacks to remain stealthy.

Lemma 1 (Sufficient condition for stealth attack)**.** *An attack* $\mathbf{b}$ *is stealthy if there exists a nonzero vector* $\mathbf{c}$ *such that* $\mathbf{M}_i\mathbf{c} = \mathbf{0}$ *for every* $i \in [n_m]$ *that is not in the support of* $\mathbf{b}$.[1]

Proof. Since $f_i(\mathbf{v}) = \text{trace}\,(\mathbf{M}_i\mathbf{v}\mathbf{v}^*)$, we have

$$f_i(\mathbf{v} + \mathbf{c}) = \text{trace}\,(\mathbf{M}_i(\mathbf{v} + \mathbf{c})(\mathbf{v} + \mathbf{c})^*) = f_i(\mathbf{v}),$$

for every $i \in [n_m]$ that is not in the support of $\mathbf{b}$. □

Lemma 1 implies that an attack is unobservable if the state deviation $\mathbf{c}$ lies in the *null space* of the measurement matrices of those sensors the attacker does not tamper with. This is applicable to the situation discussed in [26] for a single bus attack. To better understand this, consider a vector $\mathbf{c}$ that has zeros everywhere except at location j. Since the j-th column of $\mathbf{M}_i$, denoted as $[\mathbf{M}_i]_{:j}$, is zero unless $\mathbf{M}_i$ corresponds to the measurement of a branch that connects to bus j, this delineates a "superset" of sensors needed to hack to guarantee a stealth attack.

An upper bound on the minimum number of compromised sensors can be derived for a multi-bus attack; however, the sufficient condition could be too stringent because the attacker only needs to satisfy $b_i = \text{trace}\,(\mathbf{M}_i\mathbf{c}\mathbf{c}^*) + \text{trace}\,(\mathbf{M}_i\mathbf{c}\mathbf{v}^*) + \text{trace}\,(\mathbf{M}_i\mathbf{v}\mathbf{c}^*) = 0$ for all $i \notin \text{supp}\,(\mathbf{b})$ to remain stealthy. For instance, consider the system in Fig. 3. Since the bus states are all under attack, the upper bound on the minimum number of sensors to infiltrate is 40, or all the measurements, according to [26] and Lemma 1. But due to the "clever" design, FDIA is conducted successfully by tampering with only 18 sensors, which is a sparser subset of the upper bound. It is also worthwhile to note that one can think of a strategy that offsets the phases of bus voltages at bus 2, 3, 5 and 6 by a constant. This will keep the real power flows the same as before and only change the reactive flows. However, even with this ad hoc strategy, the number of sensors to tamper with is 19. This indicates the efficiency of the demonstrated strategy. However, to find such an attack vector, a general strategy can be formulated as an optimization problem (NC-FDIA) to maximize sabotage with limited resources and to evade detection, where $\mathbf{f}(\cdot)$ is the AC-model measurement function (1), $\tilde{\mathbf{v}}$ is the spurious state, $h(\cdot)$ is an optimization criterion to be specified later, and c is a constant number. The constraints amount to the unobservability condition (Definition 2) and the sparsity requirement. The following assumption is made throughout the analysis on the adversary's capability to acquire grid topology and measurement data:

Assumption 1. *The attacker can form a strategy after accessing the grid topology and the measurement vector* $\mathbf{m}$.

The above assumption depicts a powerful adversary and a completely adversarial scenario. Using the full set of measurements, the attacker can perform SE to estimate the true state $\mathbf{v}$, and tailor the attack to be stealthy. However, if this assumption is violated, the attacker risks being detected by the BDD [8]. The analysis provided in this paper is based on Assumption 1 because it helps understand the behavior of the system under the worst attack possible (using the full knowledge of the system) and simplifies the mathematical treatment. In addition, by solving the non-convex AC-based SE (3), one may end up with spurious solutions that might not be easily distinguishable from the spurious stationary point caused by FDIA. However, in practice most of the states have phases close to 0 and magnitudes close to 1 per unit, and hence may be recovered with high accuracy. For the common case where there is a large number of redundant measurements, the spurious state is often very close to the true state [39]. In this regard, spurious states in normal conditions can be much less harmful than spurious states caused by FDIA because the latter modifies the globally optimal solution of SE.

Several objectives are possible for the attacker to fulfill various malicious goals, such as:

- *Target state attack:* $h(\tilde{\mathbf{v}}) = \|\tilde{\mathbf{v}} - \mathbf{v}_{tg}\|_2^2$, which intentionally misguides the operator towards $\mathbf{v}_{tg}$;
- *Voltage collapse attack:* $h(\tilde{\mathbf{v}}) = \|\tilde{\mathbf{v}}\|_2^2$, which deceives the operator to believe in low voltages;
- *State deviation attack:* $h(\tilde{\mathbf{v}}) = -\|\tilde{\mathbf{v}} - \mathbf{v}\|_2^2$, which yields the estimated state $\tilde{\mathbf{v}}$ to be maximally different from the true state $\mathbf{v}$.

An FDIA attack can be formed by solving (NC-FDIA) with one of the above objectives; however, the problem is challenging due to: 1) a possibly nonconvex objective function, e.g., concave for the state deviation attack, 2) nonlinear equalities, and 3) cardinality constraints. The next section develops an efficient strategy to deal with these issues.

[1] The support of $\mathbf{b}$ is the set of all indices of the measurements that have been accessed and modified by the attacker.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2852774, IEEE Transactions on Automatic Control

III. SDP CONVEXIFICATION OF THE FDIA PROBLEM

Since the original attack problem (NC-FDIA) is nonconvex and difficult to tackle, we propose a convexification method based on SDP, which can be solved efficiently. Based on this framework, an "attackable region" of system states is characterized, where a strategy is guaranteed to exist and can be found efficiently. To streamline the presentation, we focus the analysis on the case of "target state attack", where $h(\tilde{\mathbf{v}}) = \|\tilde{\mathbf{v}} - \mathbf{v}_{tg}\|_2^2$ with $\mathbf{v}_{tg}$ chosen by the adversary *a priori*. The results hold for many other objective functions as well.

A. SDP convexification

By introducing an auxiliary variable $\mathbf{W} \in \mathbb{H}^{n_b}$ and the associated function $\bar{h}(\tilde{\mathbf{v}}, \mathbf{W}) = \operatorname{trace}(\mathbf{W}) - \tilde{\mathbf{v}}^* \mathbf{v}_{tg} - \mathbf{v}_{tg}^* \tilde{\mathbf{v}}$, (NC-FDIA) can be reformulated as:

$$
\begin{aligned}
\min_{\substack{\tilde{\mathbf{v}} \in \mathbb{C}^{n_b}, \mathbf{b} \in \mathbb{R}^{n_m},\\ \mathbf{W} \in \mathbb{H}^{n_b}}} \quad & \bar{h}(\tilde{\mathbf{v}}, \mathbf{W}) \\
\text{s. t.} \quad & \operatorname{trace}(\mathbf{M}_i \mathbf{W}) = m_i + b_i, \quad \forall i \in [n_m] \\
& \|\mathbf{b}\|_0 \leq c \\
& \mathbf{W} = \tilde{\mathbf{v}} \tilde{\mathbf{v}}^*
\end{aligned}
\tag{NC-FDIA-r}
$$

Note that this reformulation can be applied to the state deviation attack to convexify the objective function. A cardinality-included SDP relaxation of the above nonconvex problem can be obtained by replacing $\mathbf{W} = \tilde{\mathbf{v}}\tilde{\mathbf{v}}^*$ with a general PSD constraint:

$$
\begin{aligned}
\min_{\substack{\tilde{\mathbf{v}} \in \mathbb{C}^{n_b}, \mathbf{b} \in \mathbb{R}^{n_m},\\ \mathbf{W} \in \mathbb{H}^{n_b}}} \quad & \bar{h}(\tilde{\mathbf{v}}, \mathbf{W}) \\
\text{s. t.} \quad & \operatorname{trace}(\mathbf{M}_i \mathbf{W}) = m_i + b_i, \quad \forall i \in [n_m] \\
& \|\mathbf{b}\|_0 \leq c \\
& \begin{bmatrix} 1 & \tilde{\mathbf{v}}^* \\ \tilde{\mathbf{v}} & \mathbf{W} \end{bmatrix} \succeq 0
\end{aligned}
\tag{NC-FDIA-c}
$$

To study the relationship between the nonconvex problem (NC-FDIA-r) and its cardinality-included relaxation (NC-FDIA-c), we define an augmented matrix:

$$
\hat{\boldsymbol{Z}} = \begin{bmatrix} 1 & \hat{\mathbf{v}}^* \\ \hat{\mathbf{v}} & \hat{\mathbf{W}} \end{bmatrix}, \tag{6}
$$

where $(\hat{\mathbf{v}}, \hat{\mathbf{W}})$ is a solution of (NC-FDIA-c). It is straightforward to verify that if $\operatorname{rank}(\hat{\boldsymbol{Z}})$ is equal to 1, then we must have $\hat{\mathbf{W}} = \hat{\mathbf{v}}\hat{\mathbf{v}}^*$. Thus, $(\hat{\mathbf{v}}, \hat{\mathbf{W}})$ is feasible for (NC-FDIA-r) and consequently optimal since the objective value of (NC-FDIA-c) is a lower bound for (NC-FDIA-r). In fact, by exploring the special features of the problem, we can derive a milder condition to guarantee the equivalence. This will be elaborated next.

Assumption 2a. *Given a solution* $(\hat{\mathbf{v}}, \hat{\mathbf{W}}, \hat{\mathbf{b}})$ *of* (NC-FDIA-c), $\hat{\mathbf{v}}$ *and* $\mathbf{v}_{tg}$ *point along the same "general direction" in the sense that:*

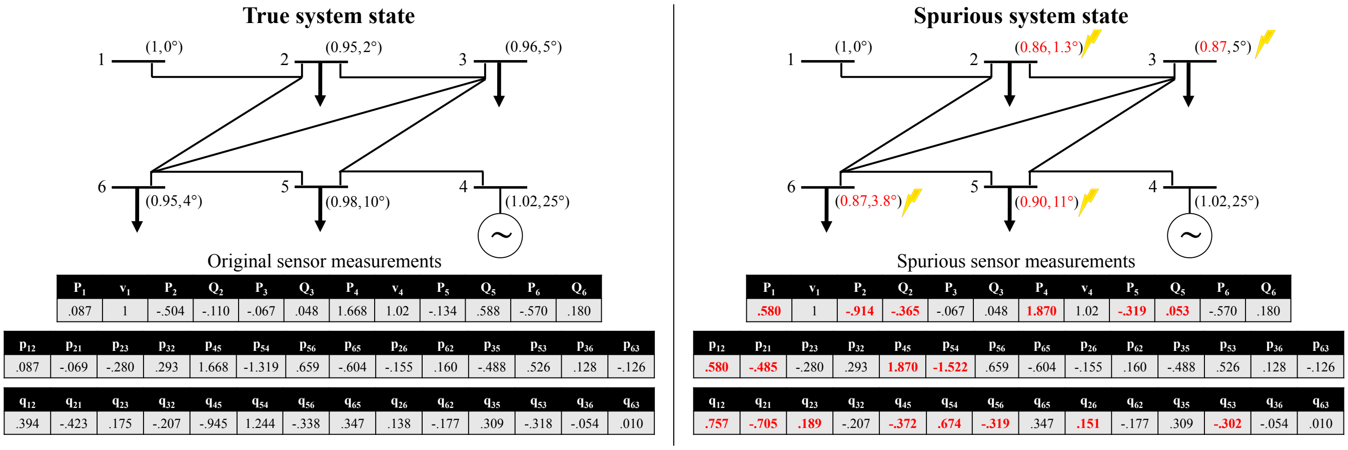
$$
\hat{\mathbf{v}}^* \mathbf{v}_{tg} + \mathbf{v}_{tg}^* \hat{\mathbf{v}} > 0. \tag{7}
$$

Note that the objective function of (NC-FDIA-c) helps with the satisfaction of Assumption 2a, since the objective aims at making $\hat{\mathbf{v}}$ and $\mathbf{v}_{tg}$ be as close as possible to each other.

Theorem 1. *The relaxation* (NC-FDIA-c) *recovers a solution of the nonconvex problem* (NC-FDIA) *and finds an optimal attack if it has a solution* $(\hat{\mathbf{v}}, \hat{\mathbf{W}}, \hat{\mathbf{b}})$ *satisfying Assumption 2a such that* $\operatorname{rank}(\hat{\mathbf{W}}) = 1$.

Proof. See Appendix A. □

Theorem 1 ensures that if $\operatorname{rank}(\hat{\mathbf{W}}) = 1$, then $\operatorname{rank}(\hat{\boldsymbol{Z}}) = 1$ (even though it could theoretically be 2), in which case (NC-FDIA-c) is able to find an optimal attack. Nevertheless, the optimal solution of (NC-FDIA-c) is not guaranteed to be rank-1 (i.e., the solution $\hat{\mathbf{W}}$ cannot be written in the form of $\mathbf{u}\mathbf{u}^*$ for any $\mathbf{u} \in \mathbb{C}^{n_b}$), and in addition the cardinality



Original sensor measurements

P_1	v_1	P_2	Q_2	P_3	Q_3	P_4	v_4	P_5	Q_5	P_6	Q_6
.087	1	-.504	-.110	-.067	.048	1.668	1.02	-.134	.588	-.570	.180

p_{12}	p_{21}	p_{23}	p_{32}	p_{45}	p_{54}	p_{56}	p_{65}	p_{26}	p_{62}	p_{35}	p_{53}	p_{36}	p_{63}
.087	-.069	-.280	.293	1.668	-1.319	.659	-.604	-.155	.160	-.488	.526	.128	-.126

q_{12}	q_{21}	q_{23}	q_{32}	q_{45}	q_{54}	q_{56}	q_{65}	q_{26}	q_{62}	q_{35}	q_{53}	q_{36}	q_{63}
.394	-.423	.175	-.207	-.945	1.244	-.338	.347	.138	-.177	.309	-.318	-.054	.010

Spurious sensor measurements

P_1	v_1	P_2	Q_2	P_3	Q_3	P_4	v_4	P_5	Q_5	P_6	Q_6
.580	1	**-.914**	**-.365**	-.067	.048	**1.870**	1.02	**-.319**	**.053**	-.570	.180

p_{12}	p_{21}	p_{23}	p_{32}	p_{45}	p_{54}	p_{56}	p_{65}	p_{26}	p_{62}	p_{35}	p_{53}	p_{36}	p_{63}
.580	**-.485**	-.280	.293	**1.870**	**-1.522**	.659	-.604	-.155	.160	-.488	.526	.128	-.126

q_{12}	q_{21}	q_{23}	q_{32}	q_{45}	q_{54}	q_{56}	q_{65}	q_{26}	q_{62}	q_{35}	q_{53}	q_{36}	q_{63}
.757	**-.705**	**.189**	-.207	**-.372**	**.674**	**-.319**	.347	**.151**	-.177	.309	**-.302**	-.054	.010

Fig. 3: An example of a 6-bus system, where the nodal voltage magnitudes and power injections as well as branch power flows are measured (p.u.). The attacker injects false data (red) to influence the bus state estimates (shown on the right side of each bus). The per unit bases for power and voltage are 100MW and 240kV, respectively. The line admittance values are identical to $1 + 1\mathrm{i}$. The FDIA injection is solved by SDP-FDIA, with parameters shown in Table I. Note that p_{ij} and q_{ij} show the active and reactive power flows over the line (i, j).

constraint $\|\mathbf{b}\|_0 \leq c$ in this optimization problem is intractable. We introduce a series of techniques to deal with each issue.

To enforce (NC-FDIA-c) to possess a rank-1 solution, we aim at penalizing the rank of its solution via a convex term. Low-rank optimization has been considered in problems such as spectral estimation [40], system identification [41], and compressed sensing [42]. A common approach is to employ the nuclear norm penalty trace $(\mathbf{W})$ [42]. However, this penalty is not appropriate for power systems, since it penalizes the voltage magnitude at each bus and may yield impractical results. Instead, a more general penalty term in the form of trace $(\mathbf{M}_0\mathbf{W})$ will be used in this paper:

$$\begin{aligned} \min_{\substack{\tilde{\mathbf{v}}\in\mathbb{C}^{n_b},\mathbf{b}\in\mathbb{R}^{n_m},\\ \mathbf{W}\in\mathbb{H}^{n_b}}} \quad & \bar{h}(\tilde{\mathbf{v}},\mathbf{W}) + \text{trace}\,(\mathbf{M}_0\mathbf{W}) \\ \text{s. t.} \quad & \text{trace}\,(\mathbf{M}_i\mathbf{W}) = m_i + b_i, \;\; \forall i \in [n_m] \\ & \|\mathbf{b}\|_0 \leq c \\ & \begin{bmatrix} 1 & \tilde{\mathbf{v}}^* \\ \tilde{\mathbf{v}} & \mathbf{W} \end{bmatrix} \succeq 0, \end{aligned} \tag{NC-FDIA-p}$$

where $\mathbf{M}_0$ is to be designed. Similar to Lasso [43], we can replace the cardinality constraint in the above problem with an l_1-norm penalty added to the objective function to induce sparsity, which leads to the convex program:

$$\begin{aligned} \min_{\substack{\tilde{\mathbf{v}}\in\mathbb{C}^{n_b},\mathbf{b}\in\mathbb{R}^{n_m},\\ \mathbf{W}\in\mathbb{H}^{n_b}}} \quad & \bar{h}(\tilde{\mathbf{v}},\mathbf{W}) + \text{trace}\,(\mathbf{M}_0\mathbf{W}) + \alpha\|\mathbf{b}\|_1 \\ \text{s. t.} \quad & \text{trace}\,(\mathbf{M}_i\mathbf{W}) = m_i + b_i, \; \forall i \in [n_m] \\ & \begin{bmatrix} 1 & \tilde{\mathbf{v}}^* \\ \tilde{\mathbf{v}} & \mathbf{W} \end{bmatrix} \succeq 0 \end{aligned} \tag{SDP-FDIA}$$

where α is a constant regularization parameter. After this convexification, (SDP-FDIA) is thus an SDP (after reformulating the l_1-norm term in a linear way), which can be solved efficiently using standard numerical solvers (e.g., SeDuMi and SDPT3) [44]. On the other hand, we recognize that by including penalty terms for rank and sparsity, we inevitably introduce bias to the optimization problem. Thus, the result obtained by (SDP-FDIA) should be described as "near-optimal", in comparison to a global minimum of NC-FDIA. This is an artifact that arises from the computational complexity of the problem, and can be only remedied by a careful selection of the penalty coefficients.

Assumption 2b. *Given a solution* $(\hat{\mathbf{v}},\hat{\mathbf{W}},\hat{\mathbf{b}})$ *of* (SDP-FDIA), $\hat{\mathbf{v}}$ *and* $\mathbf{v}_{tg}$ *have the same general direction in the sense of* (7).

Similar to Assumption 2a, this assumption is mild since the objective function of the target state attack strives to make $\hat{\mathbf{v}}$ and $\mathbf{v}_{tg}$ as close as possible to each other. We will make this assumption throughout the analysis unless otherwise specified.

Lemma 2 (Stealth attack)**.** *Let* $(\hat{\mathbf{v}},\hat{\mathbf{W}},\hat{\mathbf{b}})$ *be a solution of* (SDP-FDIA) *satisfying Assumption 2b. The attack* $\hat{\mathbf{b}}$ *is stealthy if rank*$(\hat{\mathbf{W}}) = 1$.

Proof. See Appendix A. □

B. Attackable region

In this section, we first introduce and characterize the set of voltages that the attacker can achieve by solving (SDP-FDIA) for the malicious data injection. Then, we analyze the sabotage scale under the studied FDIA. Throughout this section, let $(\hat{\mathbf{v}},\hat{\mathbf{W}},\hat{\mathbf{b}})$ denote an optimal solution of (SDP-FDIA). Given any stealth attack $\mathbf{b}$, we define an optimization problem based on (SDP-FDIA) to minimize over $(\mathbf{v},\mathbf{W})$ with a fixed $\mathbf{b}$, and denote its optimal objective value as $g(\mathbf{b})$:

$$\begin{aligned} g(\mathbf{b}) = \min_{\substack{\tilde{\mathbf{v}}\in\mathbb{C}^{n_b},\\ \mathbf{W}\in\mathbb{H}^{n_b}}} \quad & \bar{h}(\tilde{\mathbf{v}},\mathbf{W}) + \text{trace}\,(\mathbf{M}_0\mathbf{W}) \\ \text{s. t.} \quad & \text{trace}\,(\mathbf{M}_i\mathbf{W}) = m_i + b_i, \;\; \forall i \in [n_m] \\ & \begin{bmatrix} 1 & \tilde{\mathbf{v}}^* \\ \tilde{\mathbf{v}} & \mathbf{W} \end{bmatrix} \succeq 0 \end{aligned} \tag{FDIA-SE}$$

In the following, we will use $g(\mathbf{b})$ as a proxy for the sabotage scale.[2] Now, we define an "attackable" state below.

Definition 3 (Attackable state)**.** *A state* $\mathbf{v}_{at}$ *is attackable if* $(\mathbf{v}_{at},\mathbf{W} = \mathbf{v}_{at}\mathbf{v}_{at}^*)$ *is the unique and optimal solution of* (FDIA-SE) *for some stealth attack vector* $\mathbf{b} \in \mathbb{R}^m$.

Definition 4 (Attackable region)**.** *The attackable region* $\mathcal{A}(\mathcal{M},\rho)$ *for a given set of measurement types* $\mathcal{M}$ *is the set of states* $\mathbf{v}_{at}$ *that is attackable for some* $\mathbf{M}_0$ *with bounded norm* $\|\mathbf{M}_0\|_2 \leq \rho$.

In other words, for any state $\mathbf{v}_{at} \in \mathcal{A}(\mathcal{M},\rho)$ in the attackable region, there exists a stealth attack $\mathbf{b}$ such that $(\mathbf{v}_{at},\mathbf{W} = \mathbf{v}_{at}\mathbf{v}_{at}^*,\mathbf{b})$ is a feasible solution of (SDP-FDIA) and that $(\mathbf{v}_{at},\mathbf{W} = \mathbf{v}_{at}\mathbf{v}_{at}^*)$ is optimal if we fix the attack $\mathbf{b}$.

The size of $\mathcal{A}(\mathcal{M},\rho)$ also depends on ρ; more specifically, we have $\mathcal{A}(\mathcal{M},\rho_1) \subseteq \mathcal{A}(\mathcal{M},\rho_2)$ for $\rho_1 \leq \rho_2$. This follows from the definition that, for every $\mathbf{v} \in \mathcal{A}(\mathcal{M},\rho_1)$, there exists a stealth attack vector $\mathbf{b}$ and a penalty matrix $\mathbf{M}_0$ with $\|\mathbf{M}_0\| \leq \rho_1$ such that $(\mathbf{v},\mathbf{W} = \mathbf{v}\mathbf{v}^*)$ is the unique and optimal solution of (FDIA-SE); therefore, since $\|\mathbf{M}_0\| \leq \rho_1 \leq \rho_2$, we also have $\mathbf{v} \in \mathcal{A}(\mathcal{M},\rho_2)$. We will now characterize the attackable region.

Theorem 2. *If* $\mathcal{A}(\mathcal{M},\rho)$ *is non-empty for some* $\rho > 0$*, the intersection of the attackable region and the observable set, i.e.,* $\mathcal{A}(\mathcal{M},\rho) \cap \mathcal{V}(\mathcal{M})$*, is an open set.*

Proof. See Appendix B. □

For some special cases, we can have a more explicit characterization of the attackable region, as explained later.

Theorem 3. *Consider the "target state attack" with* $\bar{h}(\tilde{\mathbf{v}},\mathbf{W}) = trace\,(\mathbf{W}) - \tilde{\mathbf{v}}^*\mathbf{v}_{tg} - \mathbf{v}_{tg}^*\tilde{\mathbf{v}}$*, where* $\mathbf{v}_{tg} \in \mathcal{V}(\mathcal{M})$ *is chosen to be observable. Then,* $\mathbf{v}_{tg} \in \mathcal{A}(\mathcal{M},\rho)$ *for some* $\rho > 0$*, i.e.,* $\mathbf{v}_{tg}$ *is attackable.*

Proof. See Appendix B. □

Note that the proof of Theorem 3 allows computing ρ explicitly. Since we consider the case when $\hat{\mathbf{v}} = \mathbf{v}_{tg}$, Assump-

[2] For an optimal solution of (SDP-FDIA), the term trace $\left(\mathbf{M}_0\hat{\mathbf{W}}\right)$ can be bounded within limited ranges; as a result, $g(\mathbf{b})$ acts as a "proxy" for $\bar{h}(\hat{\mathbf{v}},\hat{\mathbf{W}})$.

tion 2b is satisfied automatically (i.e., $\mathbf{v}_{tg}^*\mathbf{v}_{tg} + \mathbf{v}_{tg}\mathbf{v}_{tg}^* > 0$). Define a set of voltages $\mathcal{R}(\mathbf{Y}) \subset \mathbb{C}^{n_b}$ such that $\mathbf{v} \in \mathcal{R}(\mathbf{Y})$ if and only if, for each line $l \in \mathcal{L}$ that connect nodes s and t, we have:

$$-\pi \leq \angle v_s - \angle v_t - \angle y_{st} \leq 0 \tag{8a}$$

$$0 \leq \angle v_s - \angle v_t + \angle y_{st} \leq \pi \tag{8b}$$

where y_{st} is the branch admittance between buses s and t. Since real-world transmission systems feature low resistance-to-reactance ratios, the angle of each line admittance y_{st} is close to $-\pi/2$ [2], and thus a realistic vector $\mathbf{v}$ would belong to $\mathcal{R}(\mathbf{Y})$ under normal conditions where the voltage phase difference along each line is relatively small. The following result gives an explicit form for a region that is attackable, in the case where the set of measurement types includes only the branch power flows and nodal voltage magnitudes, but not the nodal bus injections. Henceforth, we will refer to this case as the "special case" (compared to the "general case" where nodal bus injections can also be included in the measurements).

Theorem 4. *Let $\mathcal{V}(\mathcal{M}) \subset \mathbb{C}^{n_b}$ denote the set of observable states for a given set of measurement types $\mathcal{M}$ including the branch power flows and nodal voltage magnitudes, but not the nodal bus injections. Then, if Assumption 2b holds, we have $\mathcal{V}(\mathcal{M}) \cap \mathcal{R}(\mathbf{Y}) \subseteq \mathcal{A}(\mathcal{M}, \rho)$ for some $\rho > 0$.*

Proof. See Appendix D. □

The attackable region is an important concept that characterizes the outcome of solving (SDP-FDIA), meaning that if a state is in the attackable region, then it is a candidate attack strategy as well as the unique solution of (FDIA-SE) for some stealth attack. However, this does not imply that no stealth attack exists for a state $\tilde{\mathbf{v}}$ that is not in the attackable region; in fact, we can always construct a stealth data injection $\mathbf{b} = \mathbf{f}(\tilde{\mathbf{v}}) - \mathbf{v}$, where $\mathbf{v}$ is the true state. For example, if the measurement set $\mathcal{M}$ is so small that a part of the grid remains unobservable (see Definition 1), then (FDIA-SE) does not have a unique solution for any stealth attack $\mathbf{b}$. In that case, the attack-targeted state $\tilde{\mathbf{v}}$ does not belong to $\mathcal{A}(\mathcal{M}, \rho)$. In light of Theorem 2, if a state $\mathbf{v}_{at}$ is attackable, then any state in its small neighborhood is also attackable. Since we do not know the outcome of (SDP-FDIA) a priori, it is helpful to design a particular rank penalty matrix $\mathbf{M}_0$; indeed, as shown in Theorem 3, this can guarantee that a desired observable state is attackable. Further, Theorem 4 indicates that any observable state is attackable over a set of branch power flow measurements. In fact, we will give an explicit formula for $\mathbf{M}_0$ in this case (see the proof of Theorem 4 in Appendix D) such that the solution to (SDP-FDIA) is unique and in the form of $(\hat{\mathbf{v}}, \hat{\mathbf{W}} = \hat{\mathbf{v}}\hat{\mathbf{v}}^*, \hat{\mathbf{b}})$.

C. Performance bounds for (SDP-FDIA)

The main objective of this section is to compare the solution of (SDP-FDIA) to an "oracle attack" to be defined later, and provide guarantees for stealthy solutions (Lemma 2). First, we focus on the properties of the sabotage scale $g(\mathbf{b})$ defined in (FDIA-SE).

Lemma 3. *$g(\mathbf{b})$ is convex and sub-differentiable.*

Proof. See Appendix C. □

To proceed with the paper, we consider an "oracle attack" that is able to solve (NC-FDIA-p).

Definition 5 (Oracle attack)**.** *The oracle attack $\mathbf{b}^\star \in \mathbb{R}^{n_m}$ is a global minimum of the nonconvex program* (NC-FDIA-p)*. Define $\mathcal{B} \subseteq \mathbb{R}^{n_m}$ as the set of all vectors in $\mathbb{R}^{n_m}$ with the same support as $\mathbf{b}^\star$.*

Let $\mathbf{\Delta}_\mathcal{B} = \arg\min_{\mathbf{\Delta}_t \in \mathcal{B}} \|\mathbf{\Delta} - \mathbf{\Delta}_t\|_2^2$ be the projection of a vector $\mathbf{\Delta}$ onto the set $\mathcal{B}$. The deviation of the solution of (SDP-FDIA) from the oracle, namely $\hat{\mathbf{\Delta}} = \hat{\mathbf{b}} - \mathbf{b}^\star$, belongs to a cone.

Lemma 4. *[45] For every $\alpha \geq 2\|\partial g(\mathbf{b}^\star)\|_\infty$, the error $\hat{\mathbf{\Delta}} = \hat{\mathbf{b}} - \mathbf{b}^\star$ belongs to the cone $C(\mathcal{B}, \mathcal{B}^c; \mathbf{b}^\star) = \{\mathbf{\Delta} \in \mathbb{R}^{n_m} | \|\mathbf{\Delta}_{\mathcal{B}^c}\|_1 \leq 3\|\mathbf{\Delta}_\mathcal{B}\|_1\}$.*

For a general set of measurements that might include an arbitrary set of voltage magnitudes, nodal injections, and branch power flows as discussed in Sec. II-A, the following theorem provides performance bounds and a condition for stealthy attack using (SDP-FDIA).

Theorem 5. *Consider* (SDP-FDIA) *for a "target state attack" with $\bar{h}(\tilde{\mathbf{v}}, \mathbf{W}) = \text{trace}(\mathbf{W}) - \tilde{\mathbf{v}}^*\mathbf{v}_{tg} - \mathbf{v}_{tg}^*\tilde{\mathbf{v}}$, where $\mathbf{v}_{tg} \in \mathcal{V}(\mathcal{M})$ is chosen to be observable. Let $(\hat{\mathbf{v}}, \hat{\mathbf{W}}, \hat{\mathbf{b}})$ denote an optimal solution of* (SDP-FDIA) *for an arbitrary α greater than or equal to $2\|\partial g(\mathbf{b}^\star)\|_\infty$. The difference between the sabotage scale of the solved attack and the oracle attack satisfies the inequalities:*

$$-2\alpha\|\hat{\mathbf{\Delta}}_\mathcal{B}\|_1 \leq g(\hat{\mathbf{b}}) - g(\mathbf{b}^\star) \leq \alpha\Big(\|\hat{\mathbf{\Delta}}_\mathcal{B}\|_1 - \|\hat{\mathbf{\Delta}}_{\mathcal{B}^c}\|_1\Big),$$

where $\hat{\mathbf{\Delta}} = \hat{\mathbf{b}} - \mathbf{b}^\star$ is the difference with the oracle $\mathbf{b}^\star$.

Proof. See Appendix D. □

According to Theorem 5, there is a trade-off between attack sparsity and outcome in the sense that a tighter bound can be achieved with more entries outside the oracle sparse set $\mathcal{B}$. However, this also means that the attacker needs to tamper with more sensors. Moreover, the matrix $\mathbf{M}_0$ in (SDP-FDIA) can be constructed systematically using the Gram-Schmidt process (as detailed in the proof of Theorem 3 in Appendix B).

D. Discussions of theoretical results

To analyze the vulnerability of AC-based SE against potential FDIA, Sec. II-C formulates the adversarial problem as an optimization that aims at finding a sparse and stealthy attack vector to maliciously lead SE to make wrong determinations. However, the formulated NC-FDIA is highly nonconvex and thus computationally challenging to solve, conforming to the common belief that such an attack is difficult to be carried out without modifying a large number of measurements. Perhaps, the most surprising result of this section is that a near-globally optimal attack may be computed efficiently by solving the SDP relaxation (SDP-FDIA). Throughout the analysis, the *rank-1 condition* on the auxiliary matrix $\hat{\mathbf{W}} \in \mathbf{H}^{n_b}$ is needed to ensure that $\hat{\mathbf{W}} = \hat{\mathbf{v}}\hat{\mathbf{v}}^*$ for the problem (NC-FDIA-r) leading to a stealthy attack, where the auxiliary variable $\hat{\mathbf{W}}$ emerges in the reformulation of (NC-FDIA) (see Lemma 2). The key technique to induce a rank-1 solution without imposing the rank condition is the introduction of the penalty term trace $\left(\mathbf{M}_0\hat{\mathbf{W}}\right)$, where $\mathbf{M}_0$ can be chosen based on the target state $\mathbf{v}_{tg}$.

Further, we have characterized the attackable region of the AC-based FDIA, which is defined as the set of those states that are the unique and optimal solution of (FDIA-SE) for some stealth attack $\mathbf{b}$. Specifically, we have shown that if a state $\mathbf{v}$ is attackable, then all observable states that are close to $\mathbf{v}$ are also attackable, since the intersection of the attackable region and the set of observable states is an open set (Theorem 2). Regarding the target state attack with the target state $\mathbf{v}_{tg}$, the result of Theorem 3 suggests that $\mathbf{v}_{tg}$ is attackable as long as $\mathbf{v}_{tg}$ is observable for the given set of measurements. In a special case where the measurements only include branch power flows and nodal voltage magnitudes, any observable state satisfying conditions (8a) and (8b) is attackable (Theorem 4). In addition, to quantify the sub-optimality of the attack solved by (SDP-FDIA), the difference between the SDP solution and the oracle solution of (NC-FDIA) is bounded in Theorem 5. The bounds depend on how many measurements the adversary needs to modify (i.e., the sparsity of the attack).

The above-mentioned theoretical analysis is based on some key assumptions. Assumption 1 depicts a worst-case scenario where the adversary can access information about the grid topology and measurement values, though the ability to modify sensor values can be limited in terms of the number and locations of the sensors to manipulate. Because of the sparsity-inducing l_1-penalty on the attack $\mathbf{b}$, the spurious state $\hat{\mathbf{v}}$ obtained from (SDP-FDIA) can be different from the target state $\mathbf{v}_{tg}$. In this regard, Assumption 2b requires $\hat{\mathbf{v}}$ to be close to $\mathbf{v}_{tg}$, which is valid in the experiments and is a mild condition because the optimization minimizes the distance term $\|\hat{\mathbf{v}} - \mathbf{v}_{tg}\|_2^2$. The main message of the developed theoretical results is that even though a globally optimal solution of AC-based FDIA is hard to obtain, a stealthy and sparse attack that is near-globally optimal may be found efficiently. This is validated experimentally in the next section. Last but not least, NC-FDIA is a static problem that considers a snapshot of the measurements for state estimation. This is mainly because a static state estimation problem is often solved in practice (partly because voltage phases could change abruptly from one time period to another). In the case where the operator solves a dynamic state estimation, the attacker can plan a strategy accordingly by manipulating the time-stamped data. This can be performed by solving a larger problem that considers multiple time steps, with a penalty for the smoothness of spurious state changes. While such an attack has been studied for linear time-invariant systems [46], the results of this paper could be adopted to study nonlinear attacks for dynamical systems. This extension is left as future work.

IV. Experiments

This section numerically studies the vulnerability of power system AC-based SE under FDIA. More specifically, the objective is to validate whether the solution of (SDP-FDIA) is sparse and stealthy.

We first study the 30-bus system provided in MATPOWER [35] (Fig. 4). The states of this system are randomly initialized with voltage magnitudes uniformly distributed over $[0.98, 1.02]$ and voltage angles uniformly distributed over $[-15°, 15°]$. We consider a comprehensive measurement portfolio, which includes nodal voltage magnitudes, power injections, and branch real/reactive power flows. To streamline the presentation, we will focus on the target state attack, i.e., $h(\tilde{\mathbf{v}}) = \|\tilde{\mathbf{v}} - \mathbf{v}_{tg}\|_2^2$, where the entries of the target $\mathbf{v}_{tg}$ have been deliberately chosen to have low magnitudes (around 0.9), and phases identical to their counterparts in the true state. This would often trigger misguided contingency response, in an attempt to recover from the voltage sag [47]. Throughout the experiments, we assume that the sensor noise is Gaussian distributed with zero mean and a standard deviation of 1% of the measurement value.

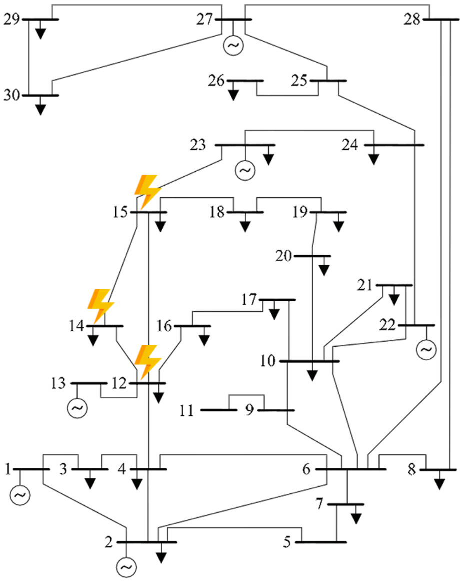
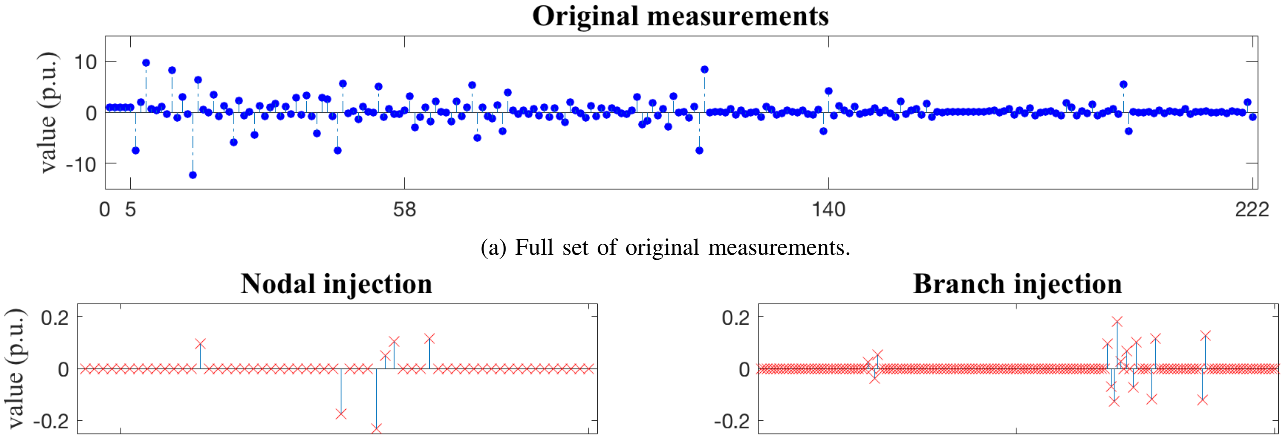


Fig. 4: The IEEE 30-bus test case [35].

An FDIA injection is obtained in Fig. 5 by solving (SDP-FDIA) with parameters listed in Table I. There are 222 measurements in total, which are organized in Fig. 5a by voltage magnitudes (indices 1–5), nodal real and reactive power injections (indices 5–58), branch real power flows (indices 58–140), and branch reactive power flows (indices 140–222). The FDIA injections for nodal measurements and branch measurements are also shown in Fig. 5. It can be observed that the injection values are relatively sparse, especially for real power flows over branches (indices 1–82 in Fig. 5c). This is due to the fact that they depend mainly on the phase differences between buses, but the target voltages have identical phases as the true state. The geographic locations of the attacked sensors include the locations of buses under attack (buses 12, 14 and 15) and the locations of the adjacent power lines, as confined within the superset used to calculate the upper bound [26]. In addition, the spurious measurements against the original values are depicted in Fig. 6. Given the



(a) Full set of original measurements.

(b) FDIA on nodal power and voltage measurements.

(c) FDIA on branch power flow measurements.

Fig. 5: There are 222 measurements in total, which are organized in Figure (a) by voltage magnitudes (indices 1–5), nodal real and reactive power injections (indices 5–58), branch real power flows (indices 58–140), and branch reactive power flows (indices 140–222). The FDIA injections for nodal measurements are shown in Figure (b), where indices 1–5 and 5–58 correspond to voltage magnitudes and bus injections, respectively. The FDIA injections for branch measurements are provided in Figure (c), where indices 1–82 and 82–164 correspond to real power flows and reactive power flows, respectively.

TABLE I: Simulation experiments, lists of the regularization parameters α and ϵ, the rank of $\hat{\boldsymbol{Z}}$, and the cardinality of $\hat{\mathbf{b}}$, as well as the upper bound given by [26].

system	α	ϵ	rank($\hat{\boldsymbol{Z}}$)	Card($\hat{\mathbf{b}}$)	upper bound	buses attacked*	pass BDD
6-bus†	.4	1/6	1	18	40	[2,3,5,6]	Yes
14-bus	.2	1/14	1	16	46	[2,3,4]	Yes
30-bus	1.16	1/30	1	21	54	[12,14,15]	Yes
39-bus	1.82	1/39	1	18	36	[26,28,29]	Yes
57-bus	0.5	1/57	1	30	92	[6,7,8]	Yes
118-bus	0.55	1/118	1	36	272	[77,78,80]	Yes

* The attacked bus numbers are identical to the MATPOWER description.
† The 6-bus system is described in Fig. 3.

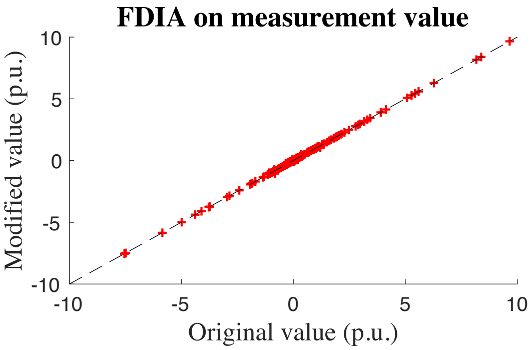


Fig. 6: This plot shows the spurious values against the original values for all the measurements. The identity relation $y = x$ is illustrated by the dotted line. It can be observed that, given the presence of innate sensor noise, the spurious values are almost identical to the original measurements.

presence of innate sensor noise, it is difficult to identify the attack on the raw measurement values by observation. In other words, the attack is "hidden" among the sensor noises.

Assume that the FDIA visualized in Fig. 5 is successfully implemented by the adversary on the set of measurements, and then the system operator solves the SE problem using either the Gauss-Newton algorithm implemented in MATPOWER (note that the attack is SE-algorithm-agnostic) or the robust SE algorithm based on least-absolute value reported in [4]. The spurious states obtained by both the Gauss-Newton algorithm and the least-absolute-value algorithm are very close to the solved target state $\tilde{\mathbf{v}}$ in (SDP-FDIA), since the construction of the attack in (SDP-FDIA) is SE-algorithm-agnostic. Fig. 7 shows the voltage magnitudes and phases of the solution of the least-absolute-value algorithm against the true states. Even though the system operates in a normal state with magnitudes in the prescribed interval $[0.98, 1.02]$, FDIA "tricks" the operator to believe in a potential voltage sag where some of the voltage magnitudes are outside of the above interval (green area in Fig. 7). Consequently, the operator may take harmful contingency actions. It is worthwhile to note that since the phases of the designed target states $\mathbf{v}_{tg}$ are identical to those of the true states by design, the spurious states estimated by the operator change insignificantly in phases, as shown in the right plot of Fig. 7.

To examine the effect of the regularization parameter α on the solution sparsity, we have run ten independent experiments with random sensor noise values and plotted the cardinality of $\hat{\mathbf{b}}$ with respect to α, as shown in Fig. 8. While the absence of $\|\cdot\|_1$ penalty (i.e., $\alpha = 0$) results in a dense solution, as α increases, the attack $\hat{\mathbf{x}}^a$ becomes significantly sparser compared to the upper bound provided by [26]. However, as α continuously increase, since the attack becomes sparser, its effect on SE reduces. This fact is reflected in the performance

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2852774, IEEE Transactions on Automatic Control

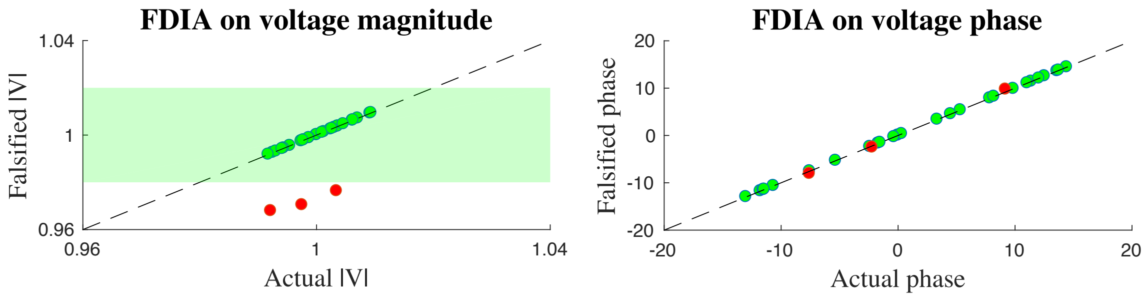


Fig. 7: These plots depict the voltage magnitude (left) and voltage phases (right) of the spurious state estimation against the true state, where the estimation is given by robust SE based on least absolute values. In both plots, the dotted line indicates the $y = x$ relationship. For the magnitude plot, the green region specifies the normal operating interval $[0.98, 1.02]$. Observe that some spurious voltage magnitudes fall out of this prescribed operating region, while all of the spurious states have almost the same phases as their counterparts in the true states, due to the specifications by the FDIA target voltage vector.

bounds in Theorem 5.

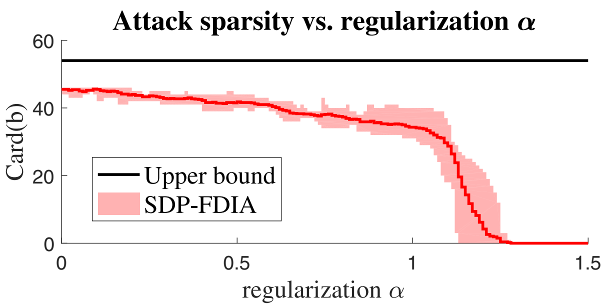


Fig. 8: This plot shows the cardinality of the solution $\hat{\mathbf{b}}$ with respect to α. The upper bound is derived according to [26]. Ten independent experiments were performed to obtain the mean (red line) and min/max (shaded region).

As for the choice of $\mathbf{M}_0$, we set $\mathbf{M}_0 = -\mathbf{I} + \epsilon \mathbf{v}_{tg}\mathbf{v}_{tg}^* + \boldsymbol{L}_0$, for a matrix $\boldsymbol{L}_0 \in \mathbb{R}^{n_b \times n_b}$ that satisfies the following properties: 1) $\boldsymbol{L}_0 \succeq 0$, 2) 0 is a simple eigenvalue of $\boldsymbol{L}_0$, 3) the vector $\mathbf{v}_{tg}$ belongs to the null space of $\boldsymbol{L}_0$ (outlined in the proof of Theorem 3). The matrix $\boldsymbol{L}_0$ is obtained via the standard Gram-Schmidt procedure by starting with the target $\mathbf{v}_{tg}$. For the choice of ϵ, the proof of Theorem 3 (Appendix B) provides a guideline to use the equation $\epsilon = \frac{1}{\mathbf{v}_{tg}^* \hat{\mathbf{v}}}$; while $\hat{\mathbf{v}}$ cannot be known *a priori*, it is desirable to be close to $\mathbf{v}_{tg}^*$. Therefore, for the 30-bus system, a value of ϵ that leads to a rank-1 solution is close to 1/30≈0.033. In addition, the algorithm has been tested on several other power systems, with parameters listed in Table I. We employ the robust SE based on least absolute values, and use the residual errors for BDD. According to the results, the constructed FDIA attack can always evade BDD detection with ϵ close to $1/n_b$. Indeed, the measurement residuals are all on the order of 0.001, which are much lower than the BDD detection threshold. As for the sparsity, we have found that the cardinality Card($\hat{\mathbf{b}}$) is lower than the upper bound by [26] at the obtained scale of attack.

As the analysis shows, by having access to the sensor measurements, the adversary can solve (SDP-FDIA) to obtain a sparse attack vector. To thwart FDIA, a set of security sensors may need to be placed at locations under potential attack as indicated by $\hat{\mathbf{b}}$ of (SDP-FDIA). For any power system, the cardinality of a potential FDIA stealth attack can be used to indicate the vulnerability of the system against potential cyber threat [13].

V. Conclusion

This study analyzes the vulnerability of power system AC-based state estimation against a critical class of cyber attacks known as false data injection attack. Since constructing an FDIA against AC-based state estimation requires solving a highly nonconvex problem, it is often believed that such attacks could be easily detected. However, this study shows that a near-globally optimal stealth attack can be found efficiently for a general scenario through a novel convexification framework based on SDP, where the measurement set could include nodal voltage magnitudes, real and reactive power injections at buses, and power flows over branches. This study further analyzes the "attackable region" and derives performance bounds for a given set of measurement types and grid topology, where an attacker can plan an attack in polynomial time with limited resources.

For protection purposes, the results can be used to understand the mechanism of FDIA on AC-based SE in order to design new BDD procedures. In addition, the outcome of anticipating such an attack can be used to evaluate the security of a given system. Above all, the proposed convexification method and its associated theoretical analysis can be applied to other nonconvex problems in power systems and beyond where the solution requires sparsity and rank conditions. This paper provides a detailed analysis on the design of a rank penalty function as well as bounds on the sparsity of the optimal solution.

References

[1] M. L. Tuballa and M. L. Abundo, "A review of the development of smart grid technologies," *Renewable and Sustainable Energy Reviews*, vol. 59, pp. 710–725, 2016.

[2] A. Abur and A. G. Exposito, *Power system state estimation: theory and implementation*. CRC press, 2004.

[3] A. Teixeira, K. C. Sou, H. Sandberg, and K. H. Johansson, "Secure control systems: A quantitative risk management approach," *IEEE Control Systems*, vol. 35, no. 1, pp. 24–45, 2015.

[4] Y. Zhang, R. Madani, and J. Lavaei, "Conic relaxations for power system state estimation with line measurements," *IEEE Transactions on Control of Network Systems*, vol. PP, no. 99, pp. 1–1, 2017.

[5] M. Jin, H. Feng, and J. Lavaei, "Multiplier-based observer design for large-scale lipschitz systems," 2018. [Online]. Available: http://www.ieor.berkeley.edu/∼lavaei/Observer_2018_1.pdf

[6] National Academies of Sciences, Engineering, and Medicine, *Enhancing the Resilience of the Nation's Electricity System*. Washington, DC: The National Academies Press, 2017.

[7] J. McCalley, "Lecture notes EE 553: Steady-state analysis - Power system operation and control," http://home.engineering.iastate.edu/∼jdm/ee553/SE1.pdf, accessed: 2017-9-21.

[8] W. Wang and Z. Lu, "Cyber security in the smart grid: Survey and challenges," *Computer Networks*, vol. 57, no. 5, pp. 1344–1371, 2013.

[9] D. Wei, Y. Lu, M. Jafari, P. M. Skare, and K. Rohde, "Protecting smart grid automation systems against cyberattacks," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 782–795, 2011.

[10] N. Beach-Westmoreland, J. Styczynski, and S. Stables, "When the lights went out – a comprehensive review of the 2015 attacks on Ukrainian critical infrastructure," 2016. [Online]. Available: http://www.infragardmd.org/announcements/whenthelightswentout

[11] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security*, vol. 14, no. 1, p. 13, 2011.

[12] G. Liang, J. Zhao, F. Luo, S. Weller, and Z. Y. Dong, "A review of false data injection attacks against modern power systems," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1630–1638, 2017.

[13] 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*, 2010.

[14] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures," in *IEEE International Conference on Smart Grid Communications*, 2010, pp. 220–225.

[15] L. Xie, Y. Mo, and B. Sinopoli, "False data injection attacks in electricity markets," in *IEEE International Conference on Smart Grid Communications*, 2010, pp. 226–231.

[16] G. Dan and H. Sandberg, "Stealth attacks and protection schemes for state estimators in power systems," in *IEEE International Conference on Smart Grid Communications*. IEEE, 2010, pp. 214–219.

[17] Y. Yuan, Z. Li, and K. Ren, "Modeling load redistribution attacks in power systems," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 382–390, 2011.

[18] F. Pasqualetti, R. Carli, and F. Bullo, "A distributed method for state estimation and false data detection in power networks," in *IEEE International Conference on Smart Grid Communications*, 2011, pp. 469–474.

[19] J. Wang, L. C. Hui, S. Yiu, E. K. Wang, and J. Fang, "A survey on cyber attacks against nonlinear state estimation in power systems of ubiquitous cities," *Pervasive and Mobile Computing*, vol. 39, pp. 52 – 64, 2017.

[20] J. Liang, O. Kosut, and L. Sankar, "Cyber attacks on ac state estimation: Unobservability and physical consequences," in *IEEE PES General Meeting— Conference & Exposition*, 2014, pp. 1–5.

[21] K. C. Sou, H. Sandberg, and K. H. Johansson, "On the exact solution to a smart grid cyber-security analysis problem," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 856–865, 2013.

[22] J. M. Hendrickx, K. H. Johansson, R. M. Jungers, H. Sandberg, and K. C. Sou, "Efficient computations of a security index for false data attacks in power networks," *IEEE Transactions on Automatic Control*, vol. 59, no. 12, pp. 3194–3208, 2014.

[23] H. Zhu and G. B. Giannakis, "Robust power system state estimation for the nonlinear AC flow model," in *Proc. of the IEEE North American Power Symposium*, 2012, pp. 1–6.

[24] V. Kekatos, G. Wang, H. Zhu, and G. B. Giannakis, "PSSE redux: Convex relaxation, decentralized, robust, and dynamic approaches," *arXiv preprint arXiv:1708.03981*, 2017.

[25] M. A. Rahman and H. Mohsenian-Rad, "False data injection attacks against nonlinear state estimation in smart power grids," in *IEEE Power and Energy Society General Meeting*, 2013, pp. 1–5.

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

[27] S. Soltan, M. Yannakakis, and G. Zussman, "Power grid state estimation following a joint cyber and physical attack," *IEEE Transactions on Control of Network Systems*, vol. PP, no. 99, pp. 1–1, 2016.

[28] S. Soltan and G. Zussman, "EXPOSE the line failures following a cyber-physical attack on the power grid," *arXiv preprint arXiv:1709.07399*, 2017.

[29] S. Soltan, M. Yannakakis, and G. Zussman, "REACT to cyber attacks on power grids," *arXiv preprint arXiv:1709.06934*, 2017.

[30] L. Mili, M. G. Cheniae, and P. J. Rousseeuw, "Robust state estimation of electric power systems," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 41, no. 5, pp. 349–358, 1994.

[31] W. W. Kotiuga and M. Vidyasagar, "Bad data rejection properties of weughted least absolute value techniques applied to static state estimation," *IEEE Transactions on Power Apparatus and Systems*, no. 4, pp. 844–853, 1982.

[32] M. K. Celik and A. Abur, "A robust WLAV state estimator using transformations," *IEEE Transactions on Power Systems*, vol. 7, no. 1, pp. 106–113, 1992.

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

[34] M. Jin, J. Lavaei, and K. Johansson, "A semidefinite programming relaxation under false data injection attacks against power grid ac state estimation," in *55th Annual Allerton Conference on Communication, Control, and Computing*, 2017, pp. 236–243.

[35] 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 Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, 2011.

[36] F. F. Wu and A. Monticelli, "Network observability: theory," *IEEE Transactions on Power Apparatus and Systems*, no. 5, pp. 1042–1048, 1985.

[37] H. Zhu and G. B. Giannakis, "Estimating the state of AC power systems using semidefinite programming," in *Proc. of the IEEE North American Power Symposium*, 2011, pp. 1–7.

[38] R. Madani, J. Lavaei, and R. Baldick, "Convexification of power flow equations for power systems in presence of noisy measurements," 2016. [Online]. Available: http://www.ieor.berkeley.edu/∼lavaei/SE_J_2016.pdf

[39] R. Y. Zhang, J. Lavaei, and R. Baldick, "Spurious local minima in power system state estimation," 2018. [Online]. Available: http://www.ieor.berkeley.edu/∼lavaei/SE_2018_1.pdf

[40] M. Zorzi and A. Chiuso, "Sparse plus low rank network identification: A nonparametric approach," *Automatica*, vol. 76, pp. 355–366, 2017.

[41] M. Zorzi and R. Sepulchre, "Ar identification of latent-variable graphical models," *IEEE Transactions on Automatic Control*, vol. 61, no. 9, pp. 2327–2340, 2016.

[42] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[43] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[44] H. Wolkowicz, R. Saigal, and L. Vandenberghe, *Handbook of semidefinite programming: theory, algorithms, and applications*. Springer Science & Business Media, 2012, vol. 27.

[45] S. N. Negahban, P. Ravikumar, M. J. Wainwright, and B. Yu, "A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers," *Statistical Science*, vol. 27, no. 4, pp. 538–557, 2012.

[46] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli, "False data injection attacks against state estimation in wireless sensor networks," in *Proc. of the IEEE Conference on Decision and Control*, 2010, pp. 5967–5972.

[47] S. Burke and E. Schneider, "Enemy number one for the electric grid: mother nature," *SAIS Review of International Affairs*, vol. 35, no. 1, pp. 73–86, 2015.

[48] R. Madani, J. Lavaei, and R. Baldick, "Convexification of power flow problem over arbitrary networks," in *IEEE Conference on Decision and Control*, 2015, pp. 1–8.

[49] D. P. Bertsekas, *Nonlinear programming*. Athena scientific Belmont, 1999.

[50] R. T. Rockafellar, *Convex analysis*. Princeton University Press, 2015.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2018.2852774, IEEE Transactions on Automatic Control

Appendix A
Proof of Theorem 1 and Lemma 2

A. *Proof of Theorem 1*

First, we prove that the equation $\text{rank}(\hat{\mathbf{W}}) = 1$ implies that $\hat{\mathbf{W}} = a^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*$, for some a such that $|a| \geq 1$. Since $\begin{bmatrix} 1 & \hat{\mathbf{v}}^* \\ \hat{\mathbf{v}} & \hat{\mathbf{W}} \end{bmatrix} \succeq 0$, by Schur complement, we have $\hat{\mathbf{W}} \succeq 0$, and $\hat{\mathbf{W}} - \hat{\mathbf{v}}\hat{\mathbf{v}}^* \succeq 0$. Due to $\text{rank}(\hat{\mathbf{W}}) = 1$, we can express $\hat{\mathbf{W}} = \mathbf{w}\mathbf{w}^*$. Since $\mathbf{w}\mathbf{w}^* - \hat{\mathbf{v}}\hat{\mathbf{v}}^* \succeq 0$, one can write $\mathbf{w} = a\hat{\mathbf{v}}$, where $|a| \geq 1$ (otherwise, there exists a vector $\boldsymbol{\nu} \in \mathbb{C}^{n_b}$ such that $\boldsymbol{\nu}^*\mathbf{w} = 0$, but $\boldsymbol{\nu}^*\hat{\mathbf{v}} \neq 0$ and $\boldsymbol{\nu}^*\left(\mathbf{w}\mathbf{w}^* - \hat{\mathbf{v}}\hat{\mathbf{v}}^*\right)\boldsymbol{\nu} = -|\boldsymbol{\nu}^*\hat{\mathbf{v}}|^2 < 0$, which violates the PSD condition).

Now, we show by contradiction that the equation $\hat{\mathbf{W}} = \hat{\mathbf{v}}\hat{\mathbf{v}}^*$ holds at optimality. Assume that $(\hat{\mathbf{v}}, \hat{\mathbf{W}} = \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*, \hat{\mathbf{b}})$ is an optimal solution of (NC-FDIA-c) and that $\hat{a} > 1$ (the case $\hat{a} < -1$ is similar). It is obvious that $(\hat{a}\hat{\mathbf{v}}, \hat{\mathbf{W}} = \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*, \hat{\mathbf{b}})$ is also feasible. This gives rise to the relation:

$$\begin{aligned} \bar{h}(\hat{\mathbf{v}}, \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*) &= \text{trace}\left(\hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*\right) - (\tilde{\mathbf{v}}^*\mathbf{v}_{tg} + \mathbf{v}_{tg}^*\tilde{\mathbf{v}}) \\ &> \text{trace}\left(\hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*\right) - \hat{a}(\tilde{\mathbf{v}}^*\mathbf{v}_{tg} + \mathbf{v}_{tg}^*\tilde{\mathbf{v}}) \\ &= \bar{h}(\hat{a}\hat{\mathbf{v}}, \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*), \end{aligned}$$

where the inequality follows from Assumption 2a. This contradicts the optimality of $(\hat{\mathbf{v}}, \hat{\mathbf{W}} = \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*, \hat{\mathbf{b}})$. Therefore, we must have $\hat{a} = 1$, implying that $\hat{\mathbf{W}} = \hat{\mathbf{v}}\hat{\mathbf{v}}^*$.

Recall that (NC-FDIA-c) provides a lower bound for (NC-FDIA-r), which is a reformulation of (NC-FDIA). Therefore, since $(\hat{\mathbf{v}}, \hat{\mathbf{W}} = \hat{\mathbf{v}}\hat{\mathbf{v}}^*, \hat{\mathbf{b}})$ is feasible for (NC-FDIA-r), it is optimal for (NC-FDIA).

B. *Proof of Lemma 2*

Let $(\hat{\mathbf{v}}, \hat{\mathbf{W}}, \hat{\mathbf{b}})$ denote an optimal solution of (SDP-FDIA). If $\text{rank}(\hat{\mathbf{W}}) = 1$, then using a similar reasoning as in the proof for Theorem 1, we have $\hat{\mathbf{W}} = a^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*$ for every $|a| \geq 1$ due to the PSD constraint. Now, we show by contradiction that the relation $\hat{\mathbf{W}} = \hat{\mathbf{v}}\hat{\mathbf{v}}^*$ holds at optimality. Let $(\hat{\mathbf{v}}, \hat{\mathbf{W}} = \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*, \hat{\mathbf{b}})$ be an optimal solution of (SDP-FDIA), and $\hat{a} > 1$ (the case $\hat{a} < -1$ is similar). It is obvious that $(\hat{a}\hat{\mathbf{v}}, \hat{\mathbf{W}} = \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*, \hat{\mathbf{b}})$ is also feasible. For a fixed $\hat{\mathbf{b}}$, this gives rise to the relation:

$$\begin{aligned} &\bar{h}(\hat{\mathbf{v}}, \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*) + \hat{a}^2\text{trace}(\mathbf{M}_0\hat{\mathbf{v}}\hat{\mathbf{v}}^*) \\ &= \text{trace}\left(\hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*\right) - (\tilde{\mathbf{v}}^*\mathbf{v}_{tg} + \mathbf{v}_{tg}^*\tilde{\mathbf{v}}) + \hat{a}^2\text{trace}(\mathbf{M}_0\hat{\mathbf{v}}\hat{\mathbf{v}}^*) \\ &> \text{trace}\left(\hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*\right) - \hat{a}(\tilde{\mathbf{v}}^*\mathbf{v}_{tg} + \mathbf{v}_{tg}^*\tilde{\mathbf{v}}) + \hat{a}^2\text{trace}(\mathbf{M}_0\hat{\mathbf{v}}\hat{\mathbf{v}}^*) \\ &= \bar{h}(\hat{a}\hat{\mathbf{v}}, \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*) + \hat{a}^2\text{trace}(\mathbf{M}_0\hat{\mathbf{v}}\hat{\mathbf{v}}^*) \end{aligned}$$

where the inequality follows from Assumption 2b. This contradicts the optimality of $(\hat{\mathbf{v}}, \hat{\mathbf{W}} = \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*, \hat{\mathbf{b}})$. Therefore, we must have $\hat{a} = 1$, implying that $\hat{\mathbf{W}} = \hat{\mathbf{v}}\hat{\mathbf{v}}^*$. Moreover, since

$$\begin{aligned} f_i(\hat{\mathbf{v}}) &= \text{trace}\left(\mathbf{M}_i\hat{\mathbf{v}}\hat{\mathbf{v}}^*\right) = \text{trace}\left(\mathbf{M}_i\hat{\mathbf{W}}\right) \\ &= m_i + \hat{b}_i = f_i(\mathbf{v}) + \hat{b}_i,\ \forall i \in [n_m], \end{aligned}$$

the stealth condition is satisfied, implying that $\hat{\mathbf{b}}$ is stealthy.

Appendix B
Proof of Theorems 2 and 3

In the case of $\bar{h}(\tilde{\mathbf{v}}, \mathbf{W}) = \text{trace}\left(\mathbf{W}\right) - \tilde{\mathbf{v}}^*\mathbf{v}_{tg} - \mathbf{v}_{tg}^*\tilde{\mathbf{v}}$, the dual of (FDIA-SE) can be written as

$$\begin{aligned} \min_{\boldsymbol{\xi}\in\mathbb{R}^{n_m}, q_0\in\mathbb{R}} \quad & \boldsymbol{\xi}\cdot(\mathbf{m}+\mathbf{b}) \\ \text{s. t.} \quad & \begin{bmatrix} q_0 & -\mathbf{v}_{tg}^* \\ -\mathbf{v}_{tg} & \mathbf{I} + \mathbf{M}_0 + \sum_i \xi_i\mathbf{M}_i \end{bmatrix} \succeq 0, \end{aligned}$$

where $\boldsymbol{\xi}$ is the vector of dual variables. The complementary slackness condition is given by:

$$\begin{aligned} &\begin{bmatrix} q_0 & -\mathbf{v}_{tg}^* \\ -\mathbf{v}_{tg} & \mathbf{I} + \mathbf{M}_0 + \sum_i \xi_i\mathbf{M}_i \end{bmatrix} \begin{bmatrix} 1 & \tilde{\mathbf{v}}^* \\ \tilde{\mathbf{v}} & \mathbf{W} \end{bmatrix} \\ &= \begin{bmatrix} q_0 - \mathbf{v}_{tg}^*\tilde{\mathbf{v}} & q_0\tilde{\mathbf{v}}^* - \mathbf{v}_{tg}^*\mathbf{W} \\ -\mathbf{v}_{tg} + \mathbf{Q}_0\tilde{\mathbf{v}} & -\mathbf{v}_{tg}\tilde{\mathbf{v}}^* + \mathbf{Q}_0\mathbf{W} \end{bmatrix} = \mathbf{0}. \end{aligned} \tag{9}$$

Let $(\hat{\mathbf{v}}, \hat{\mathbf{W}})$ be an optimal solution of (FDIA-SE) and $(\hat{q}_0, \hat{\boldsymbol{\xi}})$ be a dual optimal solution. It follows from the above equation that $\hat{q}_0 = \mathbf{v}_{tg}^*\hat{\mathbf{v}}$. By defining

$$\mathbf{Q}_0 = \mathbf{I} + \mathbf{M}_0 + \sum_i \xi_i\mathbf{M}_i, \tag{10a}$$

$$\boldsymbol{L}_0 = -\frac{1}{\hat{q}_0}\mathbf{v}_{tg}\mathbf{v}_{tg}^* + \mathbf{I} + \mathbf{M}_0, \tag{10b}$$

$$\mathbf{H}(\boldsymbol{\xi}) = \boldsymbol{L}_0 + \sum_i \xi_i\mathbf{M}_i, \tag{10c}$$

and using the Schur complement, the dual problem can be reformulated as

$$\begin{aligned} \min_{\boldsymbol{\xi}\in\mathbb{R}^{n_m}} \quad & \boldsymbol{\xi}\cdot(\mathbf{m}+\mathbf{b}) \\ \text{s. t.} \quad & \mathbf{H}(\boldsymbol{\xi}) = \boldsymbol{L}_0 + \sum_i \xi_i\mathbf{M}_i \succeq 0. \end{aligned} \tag{FDIA-SE-d}$$

The following lemma proves strong duality between (FDIA-SE) and its dual formulation.

Lemma 5. *Suppose that there exists a vector* $\mathbf{v} \in \mathcal{V}(\mathcal{M})$ *that is feasible for* (FDIA-SE). *Then, strong duality holds between* (FDIA-SE) *and its dual formulation* (FDIA-SE-d).

Proof. To prove the lemma, it suffices to find a strictly feasible point for the dual problem. Since there exists a vector $\mathbf{v} \in \mathcal{V}(\mathcal{M})$ that is feasible for (FDIA-SE), we have $\mathbf{J}(\mathbf{v})\underline{\mathbf{v}} \neq \mathbf{0}$ due to the full rank property of $\mathbf{J}(\mathbf{v})$. Therefore, there exists an index $i \in [n_m]$ such that $\mathbf{v}^*\mathbf{M}_i\mathbf{v} \neq 0$. Let $\{\mathbf{d}_1, ..., \mathbf{d}_{n_m}\}$ denote the standard basis vectors in $\mathbb{R}^{n_m}$. Then, we can select $\hat{\boldsymbol{\xi}} = \boldsymbol{\xi} + \delta \times \mathbf{d}_i$ for any feasible dual vector $\boldsymbol{\xi}$, where $\delta \in \mathbb{R}$ is a nonzero number with an arbitrarily small absolute value such that $\delta \times \mathbf{v}^*\mathbf{M}_i\mathbf{v} > 0$. Therefore, one can write:

$$\mathbf{H}(\hat{\boldsymbol{\xi}}) = \boldsymbol{L}_0 + \sum_i \hat{\xi}_i\mathbf{M}_i = \mathbf{H}(\boldsymbol{\xi}) + c\mathbf{M}_i \succ 0 \tag{11}$$

if c is sufficiently small. Hence, $\hat{\boldsymbol{\xi}}$ is a strictly feasible dual point and, by Slater's condition, strong duality holds. □

Definition 6. *Define* $\Omega(\boldsymbol{L}_0, \mathbf{v})$ *as a set of dual variables such that*

$$\mathbf{J}(\mathbf{v})^\top\boldsymbol{\xi} = -2\underline{\boldsymbol{L}_0}\underline{\mathbf{v}}, \tag{12}$$

for every $\boldsymbol{\xi} \in \Omega(\boldsymbol{L}_0, \mathbf{v})$, where $\mathbf{J}(\mathbf{v}) \in \mathbb{R}^{n_m \times (2n_b - 1)}$ is the Jacobian matrix in (5).

Since $\mathbf{H}(\boldsymbol{\xi}) = \boldsymbol{L}_0 + \sum_i \xi_i \mathbf{M}_i$, we have

$$\underline{\mathbf{H}(\boldsymbol{\xi})}\,\underline{\mathbf{v}} = \underline{\boldsymbol{L}_0}\,\underline{\mathbf{v}} + \sum_i \xi_i \underline{\mathbf{M}_i}\,\underline{\mathbf{v}} = \underline{\boldsymbol{L}_0}\,\underline{\mathbf{v}} + \tfrac{1}{2}\mathbf{J}(\mathbf{v})^\top \boldsymbol{\xi} = \mathbf{0},$$

for all $\boldsymbol{\xi} \in \Omega(\boldsymbol{L}_0, \mathbf{v})$, which indicates that $\mathbf{v}$ lies in the null space of $\mathbf{H}(\boldsymbol{\xi}) \in \mathbb{S}^{n_b}$ for every $\boldsymbol{\xi} \in \Omega(\boldsymbol{L}_0, \mathbf{v})$.

Lemma 6. *For every $\mathbf{v} \in \mathcal{V}(\mathcal{M})$ and $n_m \geq 2n_b - 1$, there is a vector $\boldsymbol{\xi} \in \mathbb{R}^{n_m}$ such that* (12) *is satisfied. Therefore, $\Omega(\boldsymbol{L}_0, \mathbf{v})$ is nonempty for every observable state vector $\mathbf{v}$.*

Proof. Since $\mathbf{v} \in \mathcal{V}(\mathcal{M})$ is observable, $\mathbf{J}(\mathbf{v})$ has full column rank. This implies that, for every $\boldsymbol{L}_0$, as long as the number of rows of $\mathbf{J}(\mathbf{v})$, namely n_m, is greater than or equal to the number of columns, namely $2n_b - 1$, there is a vector $\boldsymbol{\xi}$ satisfying (12). □

A. Proof of Theorem 2

Define $\kappa(\mathbf{H}(\boldsymbol{\xi}))$ as the sum of the two smallest eigenvalues of the Hermitian matrix $\mathbf{H}(\boldsymbol{\xi}) \in \mathbb{S}^{n_b}$. It can be shown that the intersection of the attackable region and observable set, i.e., $\mathcal{A}(\mathcal{M}, \rho) \cap \mathcal{V}(\mathcal{M})$, can be represented as

$$\{\mathbf{v} \in \mathcal{V}(\mathcal{M}) | \kappa(\mathbf{H}(\boldsymbol{\xi})) > 0, \boldsymbol{\xi} \in \Omega(\boldsymbol{L}_0, \mathbf{v})\}.$$

Now, consider a vector $\mathbf{v}$ in $\{\mathbf{v} \in \mathcal{V}(\mathcal{M}) | \kappa(\mathbf{H}(\boldsymbol{\xi})) > 0, \boldsymbol{\xi} \in \Omega(\boldsymbol{L}_0, \mathbf{v})\}$, and let δ denote the second smallest eigenvalue of $\mathbf{H}(\mathbf{M}, \boldsymbol{\xi})$. Due to the continuity of the mapping from a state $\mathbf{v}$ to a set $\Omega(\boldsymbol{L}_0, \mathbf{v})$, there exists a neighborhood $\mathcal{T} \in \mathbb{C}^{n_b}$ such that there exists a $\boldsymbol{\xi}_t \in \Omega(\boldsymbol{L}_0, \mathbf{v}_t)$ with the following property:

$$\|\mathbf{H}(\boldsymbol{\xi}) - \mathbf{H}(\boldsymbol{\xi}_t)\|_F < \sqrt{\delta} \tag{13}$$

for every $\mathbf{v}_t \in \mathcal{V}(\mathcal{M}) \cap \mathcal{T}$ (note that $\|.\|_F$ represents the Frobenius norm). Using an eigenvalue perturbation argument (Lemma 5 in [48]), it can be concluded that $\mathbf{H}(\boldsymbol{\xi}_t) \succeq 0$ and $\text{rank}(\mathbf{H}(\boldsymbol{\xi}_t)) = n_b - 1$, which imply that $\kappa(\mathbf{H}(\boldsymbol{\xi}_t)) > 0$ and $\mathbf{v}_t \in \{\mathbf{v} \in \mathcal{V}(\mathcal{M}) | \kappa(\mathbf{H}(\boldsymbol{\xi})) > 0, \boldsymbol{\xi} \in \Omega(\boldsymbol{L}_0, \mathbf{v})\}$. Hence, $\mathcal{A}(\mathcal{M}, \rho) \cap \mathcal{V}(\mathcal{M})$ is an open set.

B. Proof of Theorem 3

Let $\mathbf{M}_0$ be chosen as $\mathbf{M}_0 = -\mathbf{I} + \epsilon \mathbf{v}_{tg}\mathbf{v}_{tg}^* + \boldsymbol{L}_0$, for some $\epsilon > 0$ and a matrix $\boldsymbol{L}_0 \in \mathbb{R}^{n_b \times n_b}$ satisfying the following properties: 1) $\boldsymbol{L}_0 \succeq 0$, 2) 0 is a simple eigenvalue of $\boldsymbol{L}_0$, 3) the vector $\mathbf{v}_{tg}$ belongs to the null space of $\boldsymbol{L}_0$. By construction, $\text{rank}(\boldsymbol{L}_0) = n_b - 1$. Let $\rho = \|\mathbf{M}_0\|_2$ defined above. Note that $\boldsymbol{\xi} = \mathbf{0}$ is a feasible dual point since $\mathbf{H}(\mathbf{0}) = \boldsymbol{L}_0 \succeq 0$. It can be shown that the pair of primal solution $(\tilde{\mathbf{v}} = \mathbf{v}_{tg}, \mathbf{W} = \mathbf{v}_{tg}\mathbf{v}_{tg}^*)$ is optimal for (FDIA-SE) since it satisfies the KKT conditions (i.e., the primal-dual feasibility and complementary slackness (9)).

It is desirable to show that the optimal solution $(\tilde{\mathbf{v}} = \mathbf{v}_{tg}, \mathbf{W} = \mathbf{v}_{tg}\mathbf{v}_{tg}^*)$ is unique. Due to the complementary slackness condition (9), it follows that $\mathbf{H}(\mathbf{0})\mathbf{W} = \boldsymbol{L}_0\mathbf{W} = \mathbf{0}$. We have shown that $\boldsymbol{\xi} = \mathbf{0}$ is a feasible dual solution, and it is optimal since the sufficient optimality conditions are satisfied (see e.g., [49, Chap. 5]). The rank-1 condition for $\mathbf{W}$ follows from the equation $\text{rank}(\mathbf{H}(\mathbf{0})) = \text{rank}(\boldsymbol{L}_0) = n_b - 1$ (since this together with $\mathbf{H}(\mathbf{0})\mathbf{W} = \mathbf{0}$ implies that $\mathbf{W}$ lies in the null space of $\mathbf{H}(\hat{\boldsymbol{\xi}})$, which is at most rank 1). Moreover, because of the equation $\mathbf{H}(\mathbf{0})\mathbf{v}_{tg} = \boldsymbol{L}_0\mathbf{v}_{tg} = \mathbf{0}$, the matrix $\mathbf{W}$ must be in the form of $\xi\mathbf{v}_{tg}\mathbf{v}_{tg}^*$, where ξ is a nonzero constant. Using the result of Lemma 2, we can establish that $\xi = 1$, and that $(\tilde{\mathbf{v}} = \mathbf{v}_{tg}, \mathbf{W} = \mathbf{v}_{tg}\mathbf{v}_{tg}^*)$ is the unique optimal solution. By Definition 3, this means that $\mathbf{v}_{tg} \in \mathcal{A}(\mathcal{M}, \rho)$ is attackable.

Appendix C
Proof of Lemma 3

For any two attacks $\mathbf{b}_1$ and $\mathbf{b}_2$, let the optimal states be denoted as $(\hat{\mathbf{v}}^{(1)}, \hat{\mathbf{W}}^{(1)})$ and $(\hat{\mathbf{v}}^{(2)}, \hat{\mathbf{W}}^{(2)})$. For every number $\lambda \in [0, 1]$, the point $(\lambda\hat{\mathbf{v}} + (1 - \lambda)\hat{\mathbf{v}}^{(2)}, \lambda\hat{\mathbf{W}} + (1 - \lambda)\hat{\mathbf{W}}^{(2)})$ is a feasible solution for the attack $\lambda\mathbf{b}_1 + (1 - \lambda)\mathbf{b}_2$:

$$g(\lambda\mathbf{b}_1 + (1 - \lambda)\mathbf{b}_2) \leq \lambda g(\mathbf{b}_1) + (1 - \lambda) g(\mathbf{b}_2),$$

which proves the convexity. In what follows, in addition to proving the continuity of $g(\mathbf{b})$, we will derive a bound on the subgradient of $g(\mathbf{b})$, which is used in Theorem 5. The method is an extension of [44] to the primal formulation. In particular, our analysis is a type of parametric programming, which characterizes the change of the solution with respect to small perturbations of the parameters (see [44], Ch. 4). Consider a disturbance γ to the vector $\mathbf{b} \in \mathbb{R}^{n_m}$ in (FDIA-SE) along the direction $\underline{\mathbf{b}}$. The primal problem changes as

$$\begin{aligned} \min_{\tilde{\mathbf{v}}, \mathbf{W}} \quad & \bar{h}(\tilde{\mathbf{v}}, \mathbf{W}) + \text{trace}\,(\mathbf{M}_0\mathbf{W}) \\ \text{s. t.} \quad & \text{trace}\,(\mathbf{M}_i\mathbf{W}) = m_i + b_i + \gamma\underline{b}_i \\ & \begin{bmatrix} 1 & \tilde{\mathbf{v}}^* \\ \tilde{\mathbf{v}} & \mathbf{W} \end{bmatrix} \succeq 0 \end{aligned} \tag{P$_\gamma$}$$

and its dual formulation is given by:

$$\begin{aligned} \min_{\boldsymbol{\xi}, q_0} \quad & \boldsymbol{\xi} \cdot (\mathbf{m} + \mathbf{b} + \gamma\underline{\mathbf{b}}) \\ \text{s. t.} \quad & \begin{bmatrix} q_0 & -\mathbf{v}_{tg}^* \\ -\mathbf{v}_{tg} & \mathbf{I} + \mathbf{M}_0 + \sum_i \xi_i\mathbf{M}_i \end{bmatrix} \succeq 0 \end{aligned} \tag{D$_\gamma$}$$

Let Γ be the set of all vectors γ for which (D$_\gamma$) has a bounded solution and is strictly feasible. Assume that $0 \in \Gamma$. It is straightforward to verify that Γ is a closed (and possibly unbounded) interval. Due to duality, (P$_\gamma$) is feasible and has a bounded solution for every $\gamma \in \Gamma$, and the duality gap is zero.

Let $\mathcal{F}_\gamma$ denote the feasible set of (D$_\gamma$), $\mathbf{b}(\gamma) = \mathbf{m} + \mathbf{b} + \gamma\underline{\mathbf{b}}$, and $\boldsymbol{\xi}(\gamma) \in \{\boldsymbol{\xi}_\gamma : \boldsymbol{\xi}_\gamma = \arg\min\{\boldsymbol{\xi}_\gamma \cdot \mathbf{b}(\gamma), \boldsymbol{\xi}_\gamma \in \mathcal{F}_\gamma\}$. Moreover, let $\phi(\gamma; \mathbf{b}, \underline{\mathbf{b}}) = \boldsymbol{\xi}(\gamma) \cdot \mathbf{b}(\gamma)$ be the optimal value function. Obviously, we have $\phi(0; \mathbf{b}, \underline{\mathbf{b}}) = g(\mathbf{b})$ by the Slater's condition, and $\phi(\gamma; \mathbf{b}, \underline{\mathbf{b}})$ is *concave* in γ. We will use the shorthand notation $\phi(\gamma)$ henceforth.

Next, we derive the subdifferential of $\phi(\gamma)$, which is equivalent to $\partial g(\mathbf{b})$ when $\gamma = 0$ and $\underline{\mathbf{b}}$ is one of the canonical basis in $\mathbb{R}^{n_m}$. For any $\gamma \in \text{int}\,\Gamma$, choose $d\gamma$ small enough such that the point $\boldsymbol{\xi}(\gamma + d\gamma)$ lies in a compact set. Let $\boldsymbol{\xi}^+(\gamma)$ and $\boldsymbol{\xi}^-(\gamma)$ denote the limit as $d\gamma \to +0$ and -0, respectively.

Lemma 7. *The equations*

$$\lim_{d\gamma \to +0} \frac{\mathbf{b}(\gamma) \cdot (\boldsymbol{\xi}(\gamma + d\gamma) - \boldsymbol{\xi}^+(\gamma))}{d\gamma} = 0$$

$$\lim_{d\gamma\to -0} \frac{\mathbf{b}(\gamma)\cdot(\boldsymbol{\xi}(\gamma+d\gamma)-\boldsymbol{\xi}^{-}(\gamma))}{d\gamma} = 0$$

hold for every $\gamma \in$ int Γ.

Proof. It is straightforward to verify that $\boldsymbol{\xi}^{+}(\gamma)$ is an optimal solution of (D_γ). Assume that

$$\lim_{d\gamma\to +0} \frac{\mathbf{b}(\gamma)\cdot(\boldsymbol{\xi}(\gamma+d\gamma)-\boldsymbol{\xi}^{+}(\gamma))}{d\gamma} \geq \epsilon > 0.$$

There exists a sequence $\{d\gamma_k\} \to +0$ such that

$$\begin{aligned}&\mathbf{b}(\gamma+d\gamma_k)\cdot\boldsymbol{\xi}(\gamma+d\gamma_k)\\ &\geq \mathbf{b}(\gamma+d\gamma_k)\cdot\boldsymbol{\xi}^{+}(\gamma)+\epsilon d\gamma_k+d\gamma_k\underline{\mathbf{b}}\cdot(\boldsymbol{\xi}(\gamma+d\gamma_k)-\boldsymbol{\xi}^{+}(\gamma))o(d\gamma_k)\\ &> \mathbf{b}(\gamma+d\gamma_k)\cdot\boldsymbol{\xi}^{+}(\gamma)\end{aligned}$$

if $d\gamma_k$ is sufficiently small. This contradicts the optimality of $\boldsymbol{\xi}(\gamma+d\gamma_k)$ for ($\mathrm{D}_{\gamma+d\gamma_k}$). Similarly, assume that

$$\lim_{d\gamma\to +0} \frac{\mathbf{b}(\gamma)\cdot(\boldsymbol{\xi}(\gamma+d\gamma)-\boldsymbol{\xi}^{+}(\gamma))}{d\gamma} \leq \epsilon < 0$$

Then, there exists $\{d\gamma_k\} \to +0$ such that

$$\begin{aligned}\mathbf{b}(\gamma)\cdot\boldsymbol{\xi}(\gamma+d\gamma_k) &\leq \mathbf{b}(\gamma)\cdot\boldsymbol{\xi}^{+}(\gamma)+\epsilon d\gamma_k+o(d\gamma_k)\\ &< \mathbf{b}(\gamma)\cdot\boldsymbol{\xi}^{+}(\gamma),\end{aligned}$$

which contradicts that $\boldsymbol{\xi}^{+}(\gamma)$ is optimal for (D_γ). A similar argument can be made in the case where $d\gamma \to -0$. □

We now derive the directional derivative of $\phi(\gamma)$.

Lemma 8. *The equations*

$$\lim_{d\gamma\to +0} \frac{\phi(\gamma+d\gamma)-\phi(\gamma)}{d\gamma} = \boldsymbol{\xi}^{+}(\gamma)\cdot\underline{\mathbf{b}}$$

$$\lim_{d\gamma\to -0} \frac{\phi(\gamma+d\gamma)-\phi(\gamma)}{d\gamma} = \boldsymbol{\xi}^{-}(\gamma)\cdot\underline{\mathbf{b}}$$

holds for every $\gamma \in$ int Γ.

Proof. Since $\mathbf{b}(\gamma)\cdot\boldsymbol{\xi}(\gamma) = \mathbf{b}(\gamma)\cdot\boldsymbol{\xi}^{+}(\gamma) = \mathbf{b}(\gamma)\cdot\boldsymbol{\xi}^{-}(\gamma)$, it follows from Lemma 7 that

$$\begin{aligned}&\lim_{d\gamma\to +0} \frac{\phi(\gamma+d\gamma)-\phi(\gamma)}{d\gamma}\\ &= \lim_{d\gamma\to +0} \frac{\boldsymbol{\xi}(\gamma+d\gamma)\cdot\mathbf{b}(\gamma+d\gamma)-\boldsymbol{\xi}(\gamma)\cdot\mathbf{b}(\gamma)}{d\gamma}\\ &= \lim_{d\gamma\to +0} \boldsymbol{\xi}(\gamma+d\gamma)\cdot\underline{\mathbf{b}} + \frac{\mathbf{b}(\gamma)\cdot(\boldsymbol{\xi}(\gamma+d\gamma)-\boldsymbol{\xi}(\gamma))}{d\gamma}\\ &= \boldsymbol{\xi}^{+}(\gamma)\cdot\underline{\mathbf{b}}\end{aligned}$$

The proof for the case $d\gamma \to -0$ is similar. □

Notice that $\phi(\gamma)$ is continuously differentiable at γ if and only if $\underline{\mathbf{b}}\cdot\boldsymbol{\xi}^{+}(\gamma) = \underline{\mathbf{b}}\cdot\boldsymbol{\xi}^{-}(\gamma)$, which occurs either when (D_γ) has a unique solution or any feasible direction of the optimal face is orthogonal to $\underline{\mathbf{b}}$. To wrap up this section, we state the following lemma to bound the subdifferential $\partial g(\mathbf{b})$.

Lemma 9. *Let $[\boldsymbol{\xi}^{+}(0)]_i$ and $[\boldsymbol{\xi}^{-}(0)]_i$ denote the i-th entry of $\boldsymbol{\xi}(d\gamma)$ as $d\gamma \to +0$ and -0 along the direction of the i-th canonical basis in $\mathbb{R}^{n_m}$. For every attack* $\mathbf{b}$*, assume that* $0 \in \Gamma$*. The subdifferential of $g(\mathbf{b})$ is bounded element-wise as*

$$[\boldsymbol{\xi}^{+}(0)]_i \leq [\partial g(\mathbf{b})]_i \leq [\boldsymbol{\xi}^{-}(0)]_i, \quad \forall i \in [n_m]$$

Proof. The proof follows from the strong duality between (P_γ) and (D_γ) at $\gamma = 0$, the concavity of $\phi(\gamma)$, and Theorem 24.1 in [50] on the monotonicity of subdifferential. □

Appendix D
Proofs of Theorems 4 and 5

A. Proof of Theorem 4

For every $\hat{\mathbf{v}} \in \mathcal{V}(\mathcal{M}) \cap \mathcal{R}(\mathbf{Y})$, we show that by choosing $\mathbf{b} = \boldsymbol{f}(\hat{\mathbf{v}}) - \mathbf{m}$ where $f_i(\hat{\mathbf{v}})$ is given in (1), the unique optimal solution of (FDIA-SE) is given by $(\hat{\mathbf{v}}, \hat{\mathbf{W}} = \hat{\mathbf{v}}\hat{\mathbf{v}}^*)$, hence $\hat{\mathbf{v}} \in \mathcal{A}(\mathcal{M}, \rho)$ is attackable for some ρ defined below. We adopt the argument made in [4]. Let $\mathbf{M}_0$ in (SDP-FDIA) be given by the formula:

$$\mathbf{M}_0 = -\mathbf{I} + \epsilon\mathbf{v}_{tg}\mathbf{v}_{tg}^* + \sum_{l\in\mathcal{L}} \tilde{\mathbf{M}}_{pf}^{(l)} + \sum_{l\in\mathcal{L}} \tilde{\mathbf{M}}_{pt}^{(l)}, \qquad (14)$$

where $\epsilon > 0$ is a constant parameter, and $\tilde{\mathbf{M}}_{pf}^{(l)}$ and $\tilde{\mathbf{M}}_{pt}^{(l)}$ are arbitrary matrices in $\mathbb{H}^{n_b}$. For every $(s,t) \in [n_b]\times[n_b]$, assume that the (s,t) entries of $\tilde{\mathbf{M}}_{pf}^{(l)}$ and $\tilde{\mathbf{M}}_{pt}^{(l)}$ are equal to zero if $(s,t) \notin \mathcal{L}$ and otherwise satisfy the following inequalities:

$$-\pi \leq \angle y_{st} - \angle\tilde{M}_{pf,st}^{(l)} \leq 0 \qquad (15a)$$

$$\pi \leq \angle y_{st} + \angle\tilde{M}_{pt,st}^{(l)} \leq 2\pi. \qquad (15b)$$

Choose $\rho = \mathbf{M}_0$ defined in (14) accordingly.

Let $\boldsymbol{\xi} \in \mathbb{R}^{n_m}$ and $\mathbf{Q} = \begin{bmatrix} q_0 & \mathbf{q}^* \\ \mathbf{q} & \mathbf{Q}_0 \end{bmatrix} \in \mathbb{H}^{n_b+1}$ be the dual variables. By the KKT conditions for optimality, we have: a) the stationarity conditions: $\mathbf{q} = -\mathbf{v}_{tg}$ and $\mathbf{Q}_0 = \mathbf{I} + \mathbf{M}_0 + \sum_i \xi_i\mathbf{M}_i$, b) the dual feasibility condition: $\mathbf{Q} \succeq 0$, and c) the complementary slackness condition: $\mathbf{Q}\begin{bmatrix} 1 & \mathbf{v}^* \\ \mathbf{v} & \mathbf{W} \end{bmatrix} = \mathbf{0}$. Let $\mathbf{H}(\boldsymbol{\xi}) = -\frac{1}{q_0}\mathbf{v}_{tg}\mathbf{v}_{tg}^* + \mathbf{Q}_0$ and $q_0 = \mathbf{v}_{tg}^*\mathbf{v}$. Based on a) and c), we have $\mathbf{H}(\boldsymbol{\xi})\mathbf{W} = \mathbf{0}$. Due to b) and Schur complement, it is required that $\mathbf{H}(\boldsymbol{\xi}) \succeq 0$.

By Slater's condition, strong duality holds if one can construct a strictly feasible dual solution $\hat{\boldsymbol{\xi}}$, which is optimal if KKT conditions are satisfied. The rank-1 condition for $\mathbf{W}$ follows if we can further show that rank$(\mathbf{H}(\hat{\boldsymbol{\xi}})) = n_b - 1$ (since together with $\mathbf{H}(\hat{\boldsymbol{\xi}})\mathbf{W} = \mathbf{0}$, it implies that $\mathbf{W}$ lies in the null space of $\mathbf{H}(\hat{\boldsymbol{\xi}})$, which is at most rank 1).

For the three types of measurements considered in this paper, the measurement matrices are: 1) $\mathbf{M}_i = \mathbf{E}_i$ for every $i \in \mathcal{N}$ (associated with voltage magnitudes), 2) $\mathbf{M}_{i+n_b} = \mathbf{Y}_{pf}^{(l)}$ for every $i \in \mathcal{L}$ (associated with real power flow *from* the bus), and 3) $\mathbf{M}_{i+n_b+n_l} = \mathbf{Y}_{pt}^{(l)}$ for every $i \in \mathcal{L}$ (associated with real power flow *to* the bus). By denoting $\hat{\boldsymbol{\xi}} = \sum_{l\in\mathcal{L}} \hat{\boldsymbol{\xi}}_{pf}^{(l)} + \sum_{l\in\mathcal{L}} \hat{\boldsymbol{\xi}}_{pt}^{(l)}$, we can write

$$\mathbf{H}(\hat{\boldsymbol{\xi}}) = \sum_{l\in\mathcal{L}} \mathbf{H}_{pf}^{(l)}(\hat{\boldsymbol{\xi}}_{pf}^{(l)}) + \sum_{l\in\mathcal{L}} \mathbf{H}_{pt}^{(l)}(\hat{\boldsymbol{\xi}}_{pt}^{(l)}),$$

where

$$\mathbf{H}_{pf}^{(l)}(\hat{\boldsymbol{\xi}}_{pf}^{(l)}) = \tilde{\mathbf{M}}_{pf}^{(l)} + \hat{\xi}_{pf,s}^{(l)}\mathbf{E}_s + \hat{\xi}_{pf,t}^{(l)}\mathbf{E}_t + \hat{\xi}_{pf,l+n_b}^{(l)}\mathbf{Y}_{pf}^{(l)}$$

$$\mathbf{H}_{pt}^{(l)}(\hat{\boldsymbol{\xi}}_{pt}^{(l)}) = \tilde{\mathbf{M}}_{pt}^{(l)} + \hat{\xi}_{pt,s}^{(l)}\mathbf{E}_s + \hat{\xi}_{pt,t}^{(l)}\mathbf{E}_t + \hat{\xi}_{pt,l+n_l+n_b}^{(l)}\mathbf{Y}_{pt}^{(l)}$$

and $\sum_{l\in\mathcal{L}}\tilde{\mathbf{M}}_{pf}^{(l)}+\sum_{l\in\mathcal{L}}\tilde{\mathbf{M}}_{pt}^{(l)}=\mathbf{I}+\mathbf{M}_0-\frac{1}{q_0}\mathbf{v}_{tg}\mathbf{v}_{tg}^*$. Define $\hat{\boldsymbol{\xi}}_{pf}^{(l)}$ in such a way that

$$\hat{\xi}_{pf,l+n_b}^{(l)}=-\frac{2\Im\left(\hat{v}_s\hat{v}_t^*\tilde{M}_{pf,st}^{(l)*}\right)}{\Im\left(\hat{v}_s\hat{v}_t^*y_{st}^*\right)},\hat{\xi}_{pf,t}^{(l)}=\frac{|\hat{v}_s|^2\Im\left(\tilde{M}_{pf,st}^{(l)*}y_{st}\right)}{\Im\left(\hat{v}_s\hat{v}_t^*y_{st}^*\right)}$$

$$\hat{\xi}_{pf,s}^{(l)}=\frac{|\hat{v}_t|^2}{|\hat{v}_s|^2}\hat{\xi}_{pf,t}^{(l)}+\Re(y_{st})\hat{\xi}_{pf,l+n_b}^{(l)} \tag{16}$$

and $\hat{\boldsymbol{\xi}}_{pt}^{(l)}$ such that

$$\hat{\xi}_{pt,l+n_b+n_l}^{(l)}=-\frac{2\Im\left(\hat{v}_s\hat{v}_t^*\tilde{M}_{pt,st}^{(l)*}\right)}{\Im\left(\hat{v}_s\hat{v}_t^*y_{st}\right)},\hat{\xi}_{pt,t}^{(l)}=-\frac{|v_s|^2\Im\left(\tilde{M}_{pt,st}^{(l)}y_{st}\right)}{\Im\left(\hat{v}_s\hat{v}_t^*y_{st}\right)}$$

$$\hat{\xi}_{pt,s}^{(l)}=\frac{|\hat{v}_t|^2}{|\hat{v}_s|^2}\hat{\xi}_{pt,t}^{(l)}+\Re(y_{st})\hat{\xi}_{pt,l+n_b+n_l}^{(l)} \tag{17}$$

where $\hat{\mathbf{v}}$ is an optimal solution of the primal problem (FDIA-SE). It can be verified that $\mathbf{H}_{pf}^{(l)}\hat{\mathbf{v}}=\mathbf{0}$, $\mathbf{H}_{pt}^{(l)}\hat{\mathbf{v}}=\mathbf{0}$, $\mathbf{H}_{pf}^{(l)}\succeq 0$ and $\mathbf{H}_{pt}^{(l)}\succeq 0$, as long as:

$$-\pi\le\angle\hat{v}_s-\angle\hat{v}_t-\angle y_{st}\le 0 \tag{18a}$$

$$0\le\angle\hat{v}_s-\angle\hat{v}_t+\angle y_{st}\le\pi \tag{18b}$$

$$-\pi\le\angle y_{st}-\angle\tilde{M}_{pf,st}^{(l)}\le 0 \tag{18c}$$

$$\pi\le\angle y_{st}+\angle\tilde{M}_{pt,st}^{(l)}\le 2\pi. \tag{18d}$$

The inequalities (18a) and (18b) are satisfied since $\hat{\mathbf{v}}\in\mathcal{R}(\mathbf{Y})$. The inequalities (18c) and (18d) require that $\tilde{M}_{pf,st}^{(l)}$ and $\tilde{M}_{pt,st}^{(l)}$ to lie in the second or third quadrants of the complex plane, which is satisfied by the design in (15a) and (15b).

Our next goal is to show that $\text{rank}(\mathbf{H}(\hat{\boldsymbol{\xi}}))=n_b-1$, or equivalently, $\dim(\text{null}(\mathbf{H}(\hat{\boldsymbol{\xi}})))=1$. For every $\mathbf{x}\in\text{null}(\mathbf{H}(\hat{\boldsymbol{\xi}}))$, since $\mathbf{H}_{pf}^{(l)}\succeq 0$ and $\mathbf{H}_{pt}^{(l)}\succeq 0$, we have $\mathbf{H}_{pf}^{(l)}\mathbf{x}=\mathbf{H}_{pt}^{(l)}\mathbf{x}=\mathbf{0}$. By the construction of (16) and (17), for every line l with the endpoints s and t, it holds that $\frac{x_s}{\hat{v}_s}=\frac{x_t}{\hat{v}_t}$. This reasoning can be applied to another line $l':(t,a)$ to obtain $\frac{x_t}{\hat{v}_t}=\frac{x_a}{\hat{v}_a}$. By repeating the argument over a connected spanning graph of the network, one can obtain:

$$\frac{x_s}{\hat{v}_s}=\frac{x_t}{\hat{v}_t}=\frac{x_a}{\hat{v}_a}=\cdots=c \tag{19}$$

which indicates that $\mathbf{x}=\gamma\hat{\mathbf{v}}$. As a result, $\dim(\text{null}(\mathbf{H}(\hat{\boldsymbol{\xi}})))=1$ and $\text{rank}(\mathbf{H}(\hat{\boldsymbol{\xi}}))=n_b-1$. By the complementary slackness condition, it can be concluded that $\text{rank}(\hat{\mathbf{W}})=1$. By Lemma 2, we have $\hat{\mathbf{W}}=\hat{\mathbf{v}}\hat{\mathbf{v}}^*$. We also know that $\mathbf{b}$ is stealthy since $\text{trace}\left(\hat{\mathbf{v}}^*\mathbf{M}_i\hat{\mathbf{v}}\right)=m_i+b_i,\ \forall i\in[n_m]$ by choice.

B. Proof of Theorem 5

In what follows, we will derive performance bounds for $\hat{\mathbf{x}}$ compared to $\mathbf{b}^\star$. By the definition of $g(\mathbf{b})$ in (FDIA-SE), we can rewrite (SDP-FDIA) only in terms of $\mathbf{b}$ as

$$\max_{\mathbf{b}}\quad g(\mathbf{b})+\alpha\|\mathbf{b}\|_1 \tag{P4}$$

Define $r(\boldsymbol{\Delta})=g(\mathbf{b}^\star+\boldsymbol{\Delta})-g(\mathbf{b}^\star)+\alpha(\|\mathbf{b}^\star+\boldsymbol{\Delta}\|_1-\|\mathbf{b}^\star\|_1)$ and $\hat{\boldsymbol{\Delta}}=\hat{\mathbf{b}}-\mathbf{b}^\star$. The separability of the l_1-norm yields that

$$\begin{aligned}\|\mathbf{b}^\star+\hat{\boldsymbol{\Delta}}\|_1&\ge\|\mathbf{b}_{\mathcal{B}}^\star+\hat{\boldsymbol{\Delta}}_{\mathcal{B}^c}\|_1-\|\mathbf{b}_{\mathcal{B}^c}^\star+\hat{\boldsymbol{\Delta}}_{\mathcal{B}}\|_1\\&=\|\mathbf{b}_{\mathcal{B}}^\star\|_1+\|\hat{\boldsymbol{\Delta}}_{\mathcal{B}^c}\|_1-\|\hat{\boldsymbol{\Delta}}_{\mathcal{B}}\|_1\\&=\|\mathbf{b}^\star\|_1+\|\hat{\boldsymbol{\Delta}}_{\mathcal{B}^c}\|_1-\|\hat{\boldsymbol{\Delta}}_{\mathcal{B}}\|_1.\end{aligned}$$

Together with $r(\hat{\boldsymbol{\Delta}})\le 0$ that results from the optimality of $\hat{\mathbf{b}}$, we have proved the upper bound. For the lower bound, one can write:

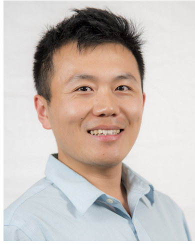
$$g(\hat{\mathbf{b}})-g(\mathbf{b}^\star)\ge\langle\partial g(\mathbf{b}^\star),\hat{\boldsymbol{\Delta}}\rangle\ge-|\langle\partial g(\mathbf{b}^\star),\hat{\boldsymbol{\Delta}}\rangle| \tag{20a}$$

$$\ge-\|\partial g(\mathbf{b}^\star)\|_\infty\|\hat{\boldsymbol{\Delta}}\|_1 \tag{20b}$$

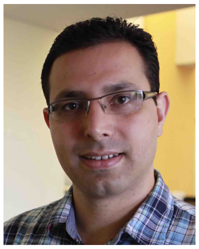
$$\ge-\frac{\alpha}{2}\left(\|\hat{\boldsymbol{\Delta}}_{\mathcal{B}}\|_1+\|\hat{\boldsymbol{\Delta}}_{\mathcal{B}^c}\|_1\right) \tag{20c}$$

$$\ge-2\alpha\|\hat{\boldsymbol{\Delta}}_{\mathcal{B}}\|_1 \tag{20d}$$

where (20a) is due to the convexity of $g(\mathbf{b})$ (Lemma 3), (20b) is by Hölder's inequality, (20c) is due to the assumption of α, and (20d) is due to Lemma 4 (see Lemma 1 and [45]).



Ming Jin is a postdoctoral researcher in the Department of Industrial Engineering and Operations Research at University of California, Berkeley. His current research interests include nonlinear optimization, data-efficient analytics, learning and control with applications to energy cyber-physical systems. He was the recipient of the Siebel scholarship, Best Paper Award at the International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, Best Paper Award at the International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, Electronic and Computer Engineering Department Scholarship, School of Engineering Scholarship, and University Scholarship at the Hong Kong University of Science and Technology.



Javad Lavaei is an Assistant Professor in the Department of Industrial Engineering and Operations Research at University of California, Berkeley. He obtained his Ph.D. degree in Control and Dynamical Systems from California Institute of Technology in 2011. He has won multiple awards, including NSF CAREER Award, Office of Naval Research Young Investigator Award, AFOSR Young Faculty Award, DARPA Young Faculty Award, Donald P. Eckman Award, INFORMS Optimization Society Prize for Young Researchers, INFORMS ENRE Energy Best Publication Award, and SIAM Control and Systems Theory Prize. He is an associate editor of the IEEE Transactions on Smart Grid and of the IEEE Control Systems Letters, and serves on the conference editorial boards of the IEEE Control Systems Society and European Control Association.

Karl Henrik Johansson is Director of the Stockholm Strategic Research Area ICT The Next Generation and Professor at the School of Electrical Engineering, KTH Royal Institute of Technology. He received MSc and PhD degrees in Electrical Engineering from Lund University. He has held visiting positions at UC Berkeley, Caltech, NTU, HKUST Institute of Advanced Studies, and NTNU. His research interests are in networked control systems, cyber-physical systems, and applications in transportation, energy, and automation. He is a member of the IEEE Control Systems Society Board of Governors and the European Control Association Council. He has received several best paper awards and other distinctions, including a ten-year Wallenberg Scholar Grant, a Senior Researcher Position with the Swedish Research Council, the Future Research Leader Award from the Swedish Foundation for Strategic Research, and the triennial Young Author Prize from IFAC. He is member of the Royal Swedish Academy of Engineering Sciences, Fellow of the IEEE, and IEEE Distinguished Lecturer.