A Semidefinite Programming Relaxation under False Data Injection Attacks against Power Grid AC State Estimation

Ming Jin[1], Javad Lavaei[2], Karl Johansson[3]

***Abstract*— The integration of sensing and information technology renders the power grid susceptible to cyber-attacks. To understand how vulnerable the state estimator is, we study its behavior under the worst attacks possible. A general false data injection attack (FDIA) based on the AC model is formulated, where the attacker manipulates sensor measurements to mislead the system operator to make decisions based on a falsified state. To stage such an attack, the optimization problem incorporates constraints of limited resources (allowing only a limited number of measurements to be altered), and stealth operation (ensuring the cyber hack cannot be identified by the bad data detection algorithm). Due to the nonlinear AC power flow model and combinatorial selection of compromised sensors, the problem is nonconvex and cannot be solved in polynomial time; however, it is shown that convexification of the original problem based on a semidefinite programming (SDP) relaxation and a sparsity penalty is able to recover a near-optimal solution. This represents the first study to solve the AC-based FDIA. Simulations on a 30-bus system illustrate that the proposed attack requires only sparse sensor manipulation and remains stealthy from the residual-based bad data detection mechanism. In light of the analysis, this study raises new challenges on grid defense mechanism and attack detection strategy.**

I. Introduction

The convergence of ubiquitous sensing and information technology enables enhanced efficiency and agility of the modern grid [1], [2]. Managed by supervisory control and data acquisition (SCADA) systems, a wealth of data on transmission and distribution power flows are collected and used to facilitate power system state estimation (SE) [3], [4] and demand response [2], [5]. The growing reliance on data communication raises concerns about cyber-security that is heightened in the aftermath of severe cyber-attacks [6], [7]. In smart grid where information is sent via remote terminal units (RTUs), it is imperative to guard against improper information modification to ensure data integrity [3], [8].

In power grid vulnerability analysis, one critical class of threat is false data injection attack (FDIA) [7], [9], which attempts to stealthily modify data to introduce error into grid SE (Fig. 1). To stage an FDIA, the attacker needs to compromise power measurements by hacking the communication between RTUs and SCADA systems. Pioneered by Liu et al. [9] and following the works [10]–[15], a stealth FDIA is possible to evade bad data detection (BDD) by the control center, with potential consequences of load shedding [14], economic loss [7], [16], and even blackouts [17].

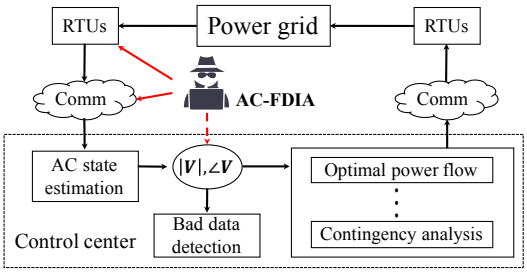

Fig. 1. Illustration of AC-based FDIA, where the attacker takes control over the RTUs or communication channel to inject false data in order to influence the grid state estimates.

While previous works on FDIA and countermeasures primarily focus on a simplified power flow model, i.e., DC model [9]–[15], [18], [19], an FDIA based on a more accurate AC model is within the realm of possibility [20]. 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], [21]. 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 sparse attacks is feasible and perhaps more detrimental than an DC-based FDIA. Once constructed, this new class of attack could be hard to detect by existing methods.

Motivated by the theoretical challenges of continuous nonconvexity and discrete nonlinearity posed by AC-based FDIA, we propose a novel convexification framework using semidefinite programming (SDP), and prove conditions on exact solution recovery and objective value bounds, which broadens the perspectives on power system security and vulnerability analysis. By investigating the least-effort strategy from the attacker's perspective, this study provides a realistic metric on the grid security, based on the number of individual sensors required to thwart an FDIA. The results also motivate protection mechanisms for AC-based SE, such as the redesign of BDD [22]. The main contributions of this work are as follows:

- Formulation of AC-based FDIA and a convexification framework using SDP and sparsity penalty;

[1]Ming Jin is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720, USA. Email: jinming@berkeley.edu

[2]Javad Lavaei is with the Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA 94720, USA. Email: lavaei@berkeley.edu

[3]Karl 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, AFOSR YIP Award, NSF CAREER Award 1351279 and NSF EECS Award 1406865

- Analysis of a condition for a near-global attack, and establishment of objective value bounds;
- Simulation study on a 30-bus system to illustrate that the planned attack is sparse and stealthy.

The rest of the paper is organized as follows. Previous works are surveyed in Section II. Section III provides preliminaries on power system modeling and SE. A general framework of AC-based FDIA is proposed in Section IV, which is convexified and analyzed in Section V. Experimental results are discussed in section VI. Conclusions are drawn in Section VII.

II. Related work

Previous works on power system vulnerability analysis have addressed potential adversarial FDIA strategies [9], [11], [14], [21], [23], negative impacts [14], [17], and possible defense mechanisms [10], [11]. From a practitioner's point of view, there are mainly two categories, based on either DC or AC models [7], [16]. For DC-FDIA, an unobservability condition was derived and the attack was numerically shown to be sparse [9], [11], [14]. Distributed DC-FDIA with partial knowledge about the topology was considered in [8], [15]. The vulnerability was quantified by the minimum number of sensors needed to compromise in order to stage stealth FDIA [10], [11], [13]. This can be formulated as a minimum cardinality problem, where different algorithms were proposed for efficient computation [18], [19]. As for the attack impact, FDIA was studied on the electric market [12] and load redistribution [14], causing significant financial losses.

Only a few works have been published on AC-based FDIA, due to the recognized complexity of nonlinear systems [3], [21]. The paper [23] has 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 [21] has 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.

Differentiated from prior literature, this study is the *first of its kind* to solve a general AC-based FDIA *exactly*, with theoretical guarantees of sparsity and unobservability (Theorem 2). The presented method has both practical and theoretical implications on solving real-world nonlinear and nonconvex problems beyond AC-based FDIA.

III. Preliminaries

A. Power system modeling

Consider an electric grid with the graph $\mathcal{G} = \{\mathcal{N}, \mathcal{L}\}$, where $\mathcal{N} \coloneqq [n_b]$ and $\mathcal{L} \coloneqq [n_l]$ represent its sets of buses and branches (we use $[x]$ to indicate the discrete set $\{1, 2, ..., x\}$). Denote the admittance of each branch as y_{st} for every $(s,t) : l \in \mathcal{L}$. 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 here. 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 (see [24], Ch. 3). Throughout this paper, we use $\mathbf{v}^*$ to indicate conjugate transpose of a vector $\mathbf{v}$ and use $\mathbf{v}^\top$ to show its transpose. We also use $\mathbb{H}^n$ to denote the set of $n \times n$ Hermitian matrix.

The state is described by $\mathbf{v} = \left[v_1, ..., v_{n_b}\right]^\top \in \mathbb{C}^{n_b}$, where $v_k \in \mathbb{C}$ is the complex voltage at bus $k \in \mathcal{N}$ with magnitude $|v_k|$ and phase $\angle v_k$. To find the underlying state of the system, a set of measurements $\mathbf{m} \in \mathbb{R}^{n_m}$ can be obtained:

$$\mathbf{m} = \mathbf{f}(\mathbf{v}) + \mathbf{e} \tag{1}$$

where $\mathbf{f} : \mathbb{C}^{n_b} \mapsto \mathbb{R}^{n_m}$ is the measurement mapping and $\mathbf{e}$ denotes random noise [20]. Based on simplifying assumptions, the DC formulation corresponds to measurements that depend linearly on voltage phases, i.e., $\mathbf{f}(\mathbf{v}) = \mathbf{H} \times \angle\mathbf{v}$ for a measurement matrix $\mathbf{H}$ [20]. In the AC formulation, the power flow and voltage magnitude measurement functions are nonlinear, but can be written in a quadratic form as

$$f_i(\mathbf{v}) = \text{trace}\,(\mathbf{M}_i \mathbf{v}\mathbf{v}^*), \qquad \forall\ i \in [n_m] \tag{2}$$

where $\mathbf{M}_i \in \mathbb{R}^{n_b \times n_b}$ depends on line admittances [1], [4]. For instance, the voltage magnitude at bus k follows the formula $|v_k|^2 = \text{trace}\,(\mathbf{E}_k \mathbf{v}\mathbf{v}^*)$ and the real power flows on a branch l connecting buses s and t are given as

$$p_{l,f} = \text{trace}\left(\mathbf{Y}_{pf}^{(l)} \mathbf{v}\mathbf{v}^*\right), \;\; p_{l,t} = \text{trace}\left(\mathbf{Y}_{pt}^{(l)} \mathbf{v}\mathbf{v}^*\right)$$

where

$$\begin{aligned}
\mathbf{E}_k &\coloneqq \mathbf{e}_k \mathbf{e}_k^\top, \\
\mathbf{Y}_{pf}^{(l)} &\coloneqq \frac{1}{2}\left(\mathbf{Y}_f^* \mathbf{d}_l \mathbf{e}_s^\top + \mathbf{e}_s \mathbf{d}_l^\top \mathbf{Y}_f\right), \\
\mathbf{Y}_{pt}^{(l)} &\coloneqq \frac{1}{2}\left(\mathbf{Y}_f^* \mathbf{d}_l \mathbf{e}_t^\top + \mathbf{e}_t \mathbf{d}_l^\top \mathbf{Y}_f\right),
\end{aligned}$$

and $\{\mathbf{e}_1, ..., \mathbf{e}_{n_b}\}$ and $\{\mathbf{d}_1, ..., \mathbf{d}_{n_l}\}$ are the sets of canonical vectors in $\mathbb{R}^{n_b}$ and $\mathbb{R}^{n_l}$, respectively.

B. State estimation

Based on noisy measurements, SE is used to monitor the system operating conditions. In general, the method finds a state that "almost matches" the observations by minimizing some distance function:

$$\min_{\hat{\mathbf{v}} \in \mathcal{V}} \|\mathbf{m} - \mathbf{f}(\hat{\mathbf{v}})\| \tag{3}$$

where $\mathcal{V}$ is a feasible set, and $\|\cdot\|$ can be any matrix norm [4], [20].

For DC models with a 2-norm objective function, the estimated state has an analytic form corresponding to a least-square solution [20]. However, due to the nonlinearity of the AC model, AC-SE is often solved by the Gauss-Newton algorithm in practice [20]. A convexification framework using SDP is proposed recently, which can recover the true state exactly under mild conditions [4]. While existing SE methods are effective under noisy measurements, the solution can be misleading under cyber-attacks, as is discussed next.

IV. General framework of AC-based FDIA

FDIA is one type of cyber-attack, which compromises SE estimator by injecting false data, namely $\mathbf{x}^a \in \mathbb{R}^{n_m}$, to n_m grid sensors [7], [8], i.e.,

$$\mathbf{m} = \mathbf{f}(\mathbf{v}) + \mathbf{e} + \mathbf{x}^a, \tag{4}$$

where $\mathbf{f}(\mathbf{v}) \in \mathbb{R}^{n_m}$ is the noiseless measurement function in (2), and $\mathbf{e} \in \mathbb{R}^{n_m}$ is random noise. The false data is maliciously injected to lead system operators to believe in an operating state, namely $\tilde{\mathbf{v}}$, other than the true state $\mathbf{v}$. As an illustrative example (Fig. 2), the operator will be "tricked" if the attacker manages to tamper certain power flow measurements to generate a fake state for the system.

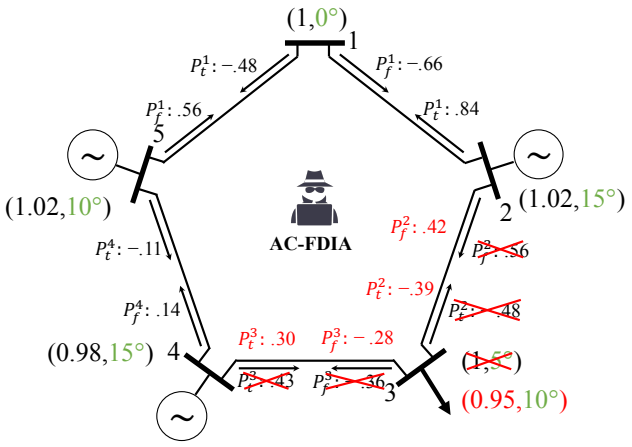

Fig. 2. Toy example of a 5-bus system, where the bus voltage magnitudes and branch real power are measured (per unit, or p.u.). The attacker injects false data (red) to influence the bus phase estimates (green).

A. Stealth attack

The sabotage cannot be detected by common BDD methods, e.g., hypothesis tests based on residuals $(m_i - f_i(\hat{\mathbf{v}}))^2$ [3]. This gives rise to the following definition of "unobservability", which generalizes previous requirements for DC [9], [11] to be applicable to AC models.

Definition 1 (Unobservability): An attack $\mathbf{x}^a$ is unobservable under state $\mathbf{v}$ if, in the absence of measurement noise, there exists a nonzero vector $\mathbf{c}$ such that $\mathbf{f}(\mathbf{v})+\mathbf{x}^a = \mathbf{f}(\mathbf{v}+\mathbf{c})$.

Analogous to the DC-based unobservability condition in [11], the following lemma provides a sufficient condition for AC-based attacks.

Lemma 1 (Sufficient condition for unobservability in AC): An attack $\mathbf{x}^a$ is unobservable 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 $\mathbf{x}^a$, i.e., $\text{supp}\,(\mathbf{x}^a)$.[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 $\text{supp}\,(\mathbf{x}^a)$, which indicates that $\mathbf{x}^a$ is unobservable. ■

[1]The support of a vector $\mathbf{x}^a$, denoted as $\text{supp}\,(\mathbf{x}^a)$, is the set of indices of the nonzero entries of $\mathbf{x}^a$.

Remark 1: 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 [23] for a single bus attack. To 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 is too stringent because the attacker only needs to satisfy $x_i^a = \text{trace}\,(\mathbf{M}_i\mathbf{c}\mathbf{c}^*) + \text{trace}\,(\mathbf{M}_i\mathbf{c}\mathbf{v}^*) + \text{trace}\,(\mathbf{M}_i\mathbf{v}\mathbf{c}^*) = 0$ for all $i \notin \text{supp}\,(\mathbf{x}^a)$ to remain stealthy. Finding a feasible vector $\mathbf{c}$ requires solving a quadratic constrained program, which is NP-hard in general.

B. Optimal attack

A general strategy for an attacker is to formulate FDIA as an optimization problem to maximize sabotage with limited resources and to evade detection:

$$\begin{aligned} \min_{\tilde{\mathbf{v}}\in\mathbb{C}^{n_b},\mathbf{x}^a\in\mathbb{R}^{n_m}} \quad & h(\tilde{\mathbf{v}}) \\ \text{s. t.} \quad & \mathbf{f}(\tilde{\mathbf{v}}) = \mathbf{m} + \mathbf{x}^a \\ & \|\mathbf{x}^a\|_0 \le b \end{aligned} \tag{P0}$$

where $h(\cdot)$ is an optimization criterion to be specified later, $\tilde{\mathbf{v}}$ is the state that seems correct to the system operator (albeit erroneously), and the constraints amount to the unobservability condition and the sparsity requirement for a given number b (note that $\|\cdot\|_0$ is the cardinality operator).

Assumption 1: The attacker has access to the grid topology and the measurement vector $\mathbf{m}$.

Assumption 1 is recognized as necessary for stealth attack against AC-SE [16]. Using the full set of measurements, the attacker can perform AC-SE to estimate the true state $\mathbf{v}$ to form a proxy. If Assumption 1 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).

The attacker can choose $h(\tilde{\mathbf{v}})$ in different ways 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 $\mathbf{v}$.

The program (P0) is challenging due to three reasons: 1) a possibly nonconvex objective function, e.g., concave for the state deviation attack, 2) nonlinear equalities, and 3) cardinality constraints. In what follows, we will develop a novel convexification framework using SDP to efficiently address the above issues.

V. Convexification and exact recovery

In this section, we first derive an SDP relaxation of problem (P0), and show that it is *exact* if the solution has rank one. We then introduce penalty terms for rank and sparsity, and prove that the solution is guaranteed to be rank-1 and sparse with established performance bounds under mild conditions. To streamline the presentation, we focus the analysis of this paper on the case 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 relaxation and sparsity penalty

Define the function $\bar{h}(\tilde{\mathbf{v}}, \mathbf{W}) = \text{trace}\,(\mathbf{W}) - \tilde{\mathbf{v}}^* \mathbf{v}_{tg} - \mathbf{v}_{tg}^* \tilde{\mathbf{v}}$, where $\mathbf{W} \in \mathbb{H}^{n_b}$. Then, (P0) can be reformulated as:

$$\begin{aligned} \min_{\substack{\tilde{\mathbf{v}} \in \mathbb{C}^{n_b}, \mathbf{x}^a \in \mathbb{R}^{n_m}, \\ \mathbf{W} \in \mathbb{H}^{n_b}}} \quad & \bar{h}(\tilde{\mathbf{v}}, \mathbf{W}) \\ \text{s. t.} \quad & \text{trace}\,(\mathbf{M}_i \mathbf{W}) = m_i + x_i^a, \ \ \forall i \in [n_m] \\ & \|\mathbf{x}^a\|_0 \le b \\ & \mathbf{W} = \tilde{\mathbf{v}} \tilde{\mathbf{v}}^* \end{aligned} \tag{P0'}$$

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 positive semi-definite (PSD) constraint:

$$\begin{aligned} \min_{\substack{\tilde{\mathbf{v}} \in \mathbb{C}^{n_b}, \mathbf{x}^a \in \mathbb{R}^{n_m}, \\ \mathbf{W} \in \mathbb{H}^{n_b}}} \quad & \bar{h}(\tilde{\mathbf{v}}, \mathbf{W}) \\ \text{s. t.} \quad & \text{trace}\,(\mathbf{M}_i \mathbf{W}) = m_i + x_i^a, \ \ \forall i \in [n_m] \\ & \|\mathbf{x}^a\|_0 \le b \\ & \begin{bmatrix} 1 & \tilde{\mathbf{v}}^* \\ \tilde{\mathbf{v}} & \mathbf{W} \end{bmatrix} \succeq 0 \end{aligned} \tag{P1}$$

In addition, we make an assumption about its solution:

Assumption 2: Given a solution $(\hat{\mathbf{v}}, \hat{\mathbf{W}}, \hat{\mathbf{x}}^a)$ of (P1), assume that $\hat{\mathbf{v}}$ is close to $\mathbf{v}_{tg}$ in the sense that:

$$\hat{\mathbf{v}}^* \mathbf{v}_{tg} + \mathbf{v}_{tg}^* \hat{\mathbf{v}} > 0. \tag{5}$$

Note that the objective function of (P1) helps with the satisfaction of Assumption 2 because it aims at making $\hat{\mathbf{v}}$ and $\mathbf{v}_{tg}$ be as closely as possible to each other. The following theorem describes a condition for the equivalence of the nonconvex problem (P0') and its cardinality-included convex relaxation (P1).

Theorem 1: The SDP relaxation (P1) recovers a solution of (P0') and finds an optimal attack if it has a solution $(\hat{\mathbf{v}}, \hat{\mathbf{W}}, \hat{\mathbf{x}}^a)$ satisfying Assumption 2 such that $\text{rank}(\hat{\mathbf{W}}) = 1$.

Proof: See Appendix A. ∎

Remark 2: Define:

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

Theorem 1 ensures that if $\text{rank}(\hat{\mathbf{W}}) = 1$, then $\text{rank}(\hat{\boldsymbol{Z}})$ is equal to 1 (even thought it could theoretically be 2), in which case (P1) is able to find an optimal attack. There are still two challenges: 1) an optimal solution of (P1) is not guaranteed to be rank-1, and 2) the cardinality constraint $\|\mathbf{x}^a\|_0 \le b$ is intractable.

To enforce (P1) to possess a rank-1 solution, we aim at penalizing the rank of its solution via a convex term. The literature of compressed sensing suggests using the nuclear norm penalty $\text{trace}\,(\mathbf{W})$ [25]. However, this penalty is not appropriate for power systems, since it would penalize the voltage magnitude at each bus and may yield impractical results. Instead, a more general penalty term in the form of $\text{trace}\,(\mathbf{M}_0 \mathbf{W})$ will be used as follows:

$$\begin{aligned} \min_{\substack{\tilde{\mathbf{v}} \in \mathbb{C}^{n_b}, \mathbf{x}^a \in \mathbb{R}^{n_m}, \\ \mathbf{W} \in \mathbb{H}^{n_b}}} \quad & \bar{h}(\tilde{\mathbf{v}}, \mathbf{W}) + \text{trace}\,(\mathbf{M}_0 \mathbf{W}) \\ \text{s. t.} \quad & \text{trace}\,(\mathbf{M}_i \mathbf{W}) = m_i + x_i^a, \ \ \forall i \in [n_m] \\ & \|\mathbf{x}^a\|_0 \le b \\ & \begin{bmatrix} 1 & \tilde{\mathbf{v}}^* \\ \tilde{\mathbf{v}} & \mathbf{W} \end{bmatrix} \succeq 0, \end{aligned} \tag{P2}$$

where the constant matrix $\mathbf{M}_0$ is to be designed. Similar to Lasso [26], we can replace the cardinality constraint in the above problem with an l_1-norm penalty added to the objective function to induce sparsity, resulting in the following program:

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

where α is a constant regularization parameter. After this convexification, (FDIA-SDP) is thus an SDP (after reformulating the norm term in a linear way), which can be solved efficiently using standard numerical methods (e.g., SeDuMi, SDPT3) [27]. We analyze its solution next, and derive a rank-1 condition for the design of a near-global attack, as well as performance bounds.

B. Exact recovery and performance bounds

Throughout this section, let $(\hat{\mathbf{v}}, \hat{\mathbf{W}}, \hat{\mathbf{x}}^a)$ denote an optimal solution of (FDIA-SDP). In light of Theorem 1, it is desirable to have $\text{rank}(\hat{\mathbf{W}}) = 1$. Given any attack $\mathbf{x}^a$, define $g(\mathbf{x}^a)$ as the optimal objective value of (FDIA-SDP) without the l_1 penalty:

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

Note that $g(\mathbf{x}^a)$ can be considered as a proxy for the sabotage scale.[2] In the following, we study some properties of $g(\mathbf{x}^a)$.

Lemma 2: $g(\mathbf{x}^a)$ is convex and sub-differentiable.

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

Proof: See Appendix B. ∎

Define $\partial g(\mathbf{x}^a)$ as the subgradient of $g(\mathbf{x}^a)$. To proceed with the paper, we consider an "oracle attack" that is able to solve (P2).

Definition 2 (Oracle attack): The oracle attack $\mathbf{x}^{a,\star} \in \mathbb{R}^{n_m}$ is a solution of the nonconvex program (P2). Define $\mathcal{B} \subseteq \mathbb{R}^{n_m}$ as the set of all vectors in $\mathbb{R}^{n_m}$ with the same support as $\mathbf{x}^{a,\star}$, and define $\mathcal{B}^c$ as the complement $\mathbb{R}^{n_m} \setminus \mathcal{B}$. 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 (FDIA-SDP)'s solution from the oracle, namely $\hat{\mathbf{\Delta}} = \hat{\mathbf{x}}^a - \mathbf{x}^{a,\star}$, belongs to a cone.

Lemma 3: For the pair $(\mathcal{B}, \mathcal{B}^c)$ and $\alpha \geq 2\|\partial g(\mathbf{x}^{a,\star})\|_\infty$, the error $\hat{\mathbf{\Delta}} = \hat{\mathbf{x}}^a - \mathbf{x}^{a,\star}$ belongs to the cone $C(\mathcal{B}, \mathcal{B}^c; \mathbf{x}^{a,\star}) = \{\mathbf{\Delta} \in \mathbb{R}^{n_m} | \|\mathbf{\Delta}_{\mathcal{B}^c}\|_1 \leq 3\|\mathbf{\Delta}_{\mathcal{B}}\|_1\}$.

Proof: See Appendix C. ∎

To ensure that the optimal solution of (FDIA-SDP), namely $\hat{\mathbf{W}}$, is rank-1, the following assumption is made on the attack state:

Assumption 3: The attack state $\hat{\mathbf{v}}$ as the solution of (FDIA-SDP) satisfies the following phase conditions, for all $(s,t) : l \in \mathcal{L}$:

$$-\pi \leq \angle \hat{v}_s - \angle \hat{v}_t - \angle y_{st} \leq 0$$
$$0 \leq \angle \hat{v}_s - \angle \hat{v}_t + \angle y_{st} \leq \pi$$

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 the line admittance y_{st} is close to $-\pi/2$ [20], and thus Assumption 3 would be satisfied under normal conditions where the voltage phase difference along each line is relatively small.

Assumption 4: For AC-SE, we assume the availability of voltage magnitude measurements at every bus, and active power measurements for both *from* and *to* ends of a branch (see Section III-A).

Assumption 5: Let Λ be the set of all regularization parameters α for which the optimal solution $\hat{\mathbf{x}}^a$ of (FDIA-SDP) corresponds to a feasible state, i.e., there exists $\mathbf{v} \in \mathbb{C}^{n_b}$ such that

$$\mathbf{f}(\mathbf{v}) = \mathbf{m} + \hat{\mathbf{x}}^a.$$

Assume that Λ is not empty and includes at least one value of α that corresponds to a non-zero injection $\hat{\mathbf{x}}^a$.

The following theorem provides performance bounds and a condition for rank-1 recovery under Assumptions 4 and 5.

Theorem 2: Let $\mathbf{M}_0$ in (FDIA-SDP) be given by:

$$\mathbf{M}_0 = -\mathbf{I} + \epsilon \mathbf{v}_{tg}\mathbf{v}_{tg}^* + \sum_{l\in\mathcal{L}} \tilde{\mathbf{M}}_{pf}^{(l)} + \sum_{l\in\mathcal{L}} \tilde{\mathbf{M}}_{pt}^{(l)} \quad (7)$$

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 $(s,t) \in \{1,...,n_b\} \times \{1,...,n_b\}$, assume that the (s,t) elements 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:

$$-\pi \leq \angle y_{st} - \angle \tilde{M}_{pf,st}^{(l)} \leq 0 \quad (8)$$
$$\pi \leq \angle y_{st} + \angle \tilde{M}_{pt,st}^{(l)} \leq 2\pi. \quad (9)$$

For every $\alpha \geq 2\|\partial g(\mathbf{x}^{a,\star})\|_\infty$ and some ϵ, the optimal solution $(\hat{\mathbf{v}}, \hat{\mathbf{W}}, \hat{\mathbf{x}}^a)$ of (FDIA-SDP) satisfies the equation:

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

where $\hat{\mathbf{\Delta}}$ is equal to $\hat{\mathbf{x}}^a - \mathbf{x}^{a,\star}$, i.e., the difference between $\hat{\mathbf{x}}^a$ and the oracle $\mathbf{x}^{a,\star}$. In addition, the attack $\hat{\mathbf{x}}^a$ is unobservable for every $\alpha \in \Lambda$.

Proof: See Appendix C. ∎

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.

As for the choice of $\mathbf{M}_0$, to conform with (8) and (9), instead of the term $\sum_{l\in\mathcal{L}} \tilde{\mathbf{M}}_{pf}^{(l)} + \sum_{l\in\mathcal{L}} \tilde{\mathbf{M}}_{pt}^{(l)}$ in (7), we can use the negative of the susceptance matrix (the imaginary part of $\mathbf{Y}$) with the exception that its diagonal entries are all zero. The value of ϵ can be chosen based on $\frac{1}{n_b}$, as discussed in the proof (see Appendix C).

In what follows, we will conduct some experiments to verify the above theoretical results and compare them with the existing literature.

VI. Experiments

For the experiment, we study a 30-bus system (shown in Fig. 3) provided in MATPOWER [24], whose states are randomly initialized with magnitudes close to 1 and small phases. The measurements include branch real power flows and bus magnitudes.

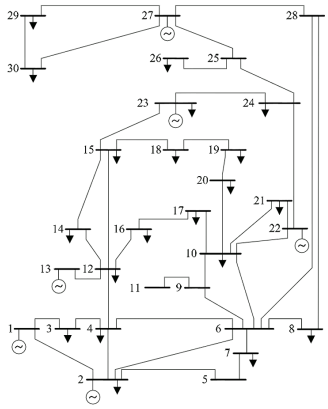

Fig. 3. The IEEE 30-bus test case [24].

Consider the target state attack problem with the measure $h(\tilde{\mathbf{v}}) = \|\tilde{\mathbf{v}} - \mathbf{v}_{tg}\|_s^2$, where the target $\mathbf{v}_{tg}$ is identical to the true state $\mathbf{v}$ except for a random subset of buses whose voltage magnitudes are deliberately chosen to be low (around 0.95). This would often trigger misguided contingency response, in an attempt to recover from a possible voltage sag [6].

An example of the (FDIA-SDP) solution is shown in Fig. 4, which depicts the true state magnitudes, $|v_i|$'s against the falsified state magnitudes $|\tilde{v}_i|$'s obtained by the system operator using the exact convexification technique described in [4] or any other global optimization techniques (in fact,

the attack is SE-algorithm-agnostic). Even though the system operates in a normal state, FDIA "tricks" the operator to assume a potential voltage sag. The operator may then take falsified harmful contingency actions. The FDIA is triggered by tampering with a small set of sensors (see the sparsity pattern in Fig. 5), and the modified measurements are hardly distinguishable from the original values (Fig. 6).

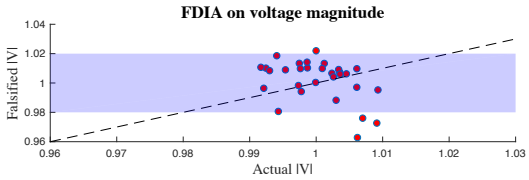

Fig. 4. The FDIA effect on SE voltage magnitudes, where the dotted line indicates identity. Several buses have magnitudes outside the normal operation region (shaded), from the system operator's perspective.

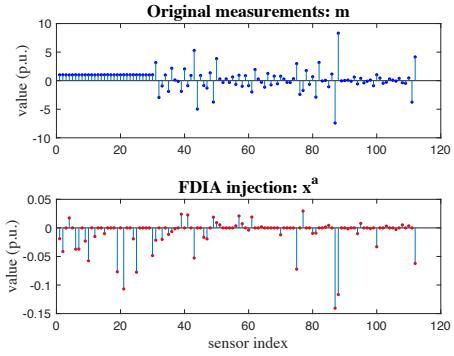

Fig. 5. An instance of the original grid measurements (top) and injected false data (bottom).

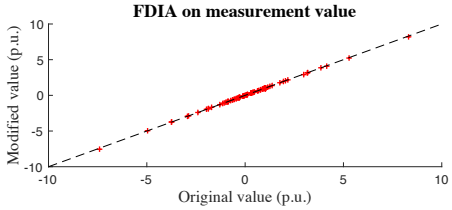

Fig. 6. Comparison of the modified and original values for all sensors, where the dotted line indicates identity. The influence of FDIA on actual sensor measurements is hard to detect.

The solution's sparsity and rank are then examined with respect to the regularization parameters. While the absence of the $\|\cdot\|_1$ 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 in [23] (Fig. 7). On the other hand, the rank of $\hat{\mathbf{W}}$ increases for large values of α, since $\mathbf{m} + \hat{\mathbf{x}}^a$ does not correspond to a valid state.

The choice of $\mathbf{M}_0$ follows (7), where the part $\sum_{l\in\mathcal{L}} \tilde{\mathbf{M}}_{pf}^{(l)} + \sum_{l\in\mathcal{L}} \tilde{\mathbf{M}}_{pt}^{(l)}$ is substituted by the susceptance matrix with zero diagonal entries. As for the choice of ϵ, Theorem 2 provides a guideline to use the equation $\epsilon = \frac{1}{\mathbf{v}_{tg}^* \hat{\mathbf{v}}}$ (see Appendix C for the proof); 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, the value of ϵ that corresponds to a rank-1 solution is close to .033, as corroborated in Fig. 8. The

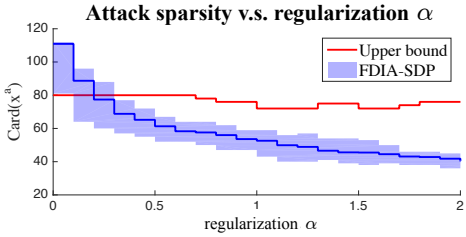

Fig. 7. Influence of the regularization parameter α on the cardiality of the solution of (FDIA-SDP). The upper bound is derived according to [23]. Ten independent experiments were performed to obtain the mean (blue line), and min/max (shaded region). We used 0.001 as the threshold for computing the cardinality of $\mathbf{x}^a$.

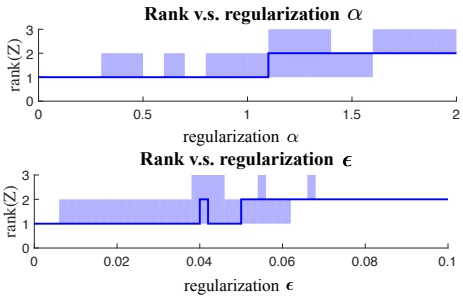

Fig. 8. Effects of the regularization parameters on the rank of $\hat{\boldsymbol{Z}}$ defined in (6). The median (blue line) and the min/max ranks (shaded) are shown.

rank-1 condition also guarantees that $\hat{\mathbf{v}}$ is close to $\mathbf{v}_{tg}$, as measured by $\|\hat{\mathbf{v}} - \mathbf{v}_{tg}\|_2^2$ (Fig. 9).

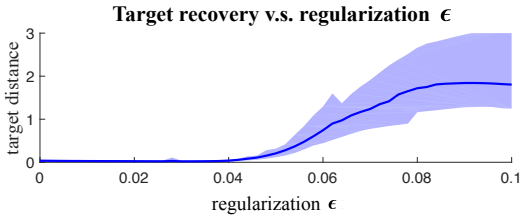

Fig. 9. Success of FDIA measured by the distance between the operator's recovered SE $\tilde{\mathbf{v}}$ and target $\mathbf{v}_{tg}$, or $\|\tilde{\mathbf{v}} - \mathbf{v}_{tg}\|_2^2$, as ϵ varies. Both the mean (blue line) and the min/max regions (shaded) are shown.

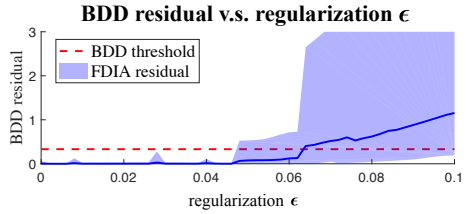

Fig. 10. Residuals $\|\mathbf{m} + \mathbf{x}^a - \mathbf{f}(\tilde{v})\|_2^2$ obtained for BDD. The detection threshold represents the variance of the measurement noise [3]. Both the mean (blue line) and the min/max regions (shaded) are shown.

The rank-1 (FDIA-SDP) solution will evade any residual-based BDD (Fig. 10). To thwart FDIA, one can place a set of security sensors at locations under potential attack as indicated by $\mathbf{x}^a$ of (FDIA-SDP).

VII. Conclusion

In the study, we first formulated a general AC-based FDIA problem with stealth and sparsity constraints. To address the

problem's nonconvexity and nonlinearity, a novel framework using SDP and l_1 penalty was proposed in (FDIA-SDP). A condition on exact recovery was proved (Theorem 2), providing a *first analytical result* on the NP-hard AC-based FDIA problem. From the perspective of power grid security, (FDIA-SDP) identifies a small subset of grid sensors that could enable staging a stealth attack. This information is essential for designing a security index based on AC modeling. As future work, it is important to investigate protection and BDD strategies against AC-based FDIA.

REFERENCES

[1] J. Lavaei and S. H. Low, "Zero duality gap in optimal power flow problem," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 92–107, 2012.

[2] M. Jin, W. Feng, P. Liu, C. Marnay, and C. Spanos, "Mod-dr: Microgrid optimal dispatch with demand response," *Applied Energy*, vol. 187, pp. 758–776, 2017.

[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, R. Jia, and C. Spanos, "Virtual occupancy sensing: Using smart meters to indicate your presence," *IEEE Transactions on Mobile Computing*, vol. 99, p. 1, 2017.

[6] 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.

[7] 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*, 2016.

[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] 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 (TISSEC)*, vol. 14, no. 1, p. 13, 2011.

[10] H. Sandberg, A. Teixeira, and K. H. Johansson, "On security indices for state estimators in power networks," in *First Workshop on Secure Control Systems (SCS), Stockholm, 2010*, 2010.

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

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

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

[14] 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.

[15] 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 (SmartGridComm)*, 2011, pp. 469–474.

[16] 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*, 2017.

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

[18] 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.

[19] 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.

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

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

[22] 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.

[23] 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.

[24] 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.

[25] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

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

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

[28] R. T. Rockafellar, *Convex analysis*. Princeton university press, 2015.

[29] 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.

APPENDIX

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 $(\hat{\mathbf{v}}, \hat{\mathbf{W}} = \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*, \hat{\mathbf{x}}^a)$ is an optimal solution of (P1) and that $\hat{a} > 1$ (the case $\hat{a} < -1$ is similar). It is obvious that $(\hat{a}\hat{\mathbf{v}}, \hat{\mathbf{W}} = \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*, \hat{\mathbf{x}}^a)$ 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 2. This contradicts the optimality of $(\hat{\mathbf{v}}, \hat{\mathbf{W}} = \hat{a}^2\hat{\mathbf{v}}\hat{\mathbf{v}}^*, \hat{\mathbf{x}}^a)$. Therefore, we must have $\hat{a} = 1$, implying that $\hat{\mathbf{W}} = \hat{\mathbf{v}}\hat{\mathbf{v}}^*$.

Recall that (P1) provides a lower bound for (P0'), which is a reformulation of (P0). Therefore, since $(\hat{\mathbf{v}}, \hat{\mathbf{W}} = \hat{\mathbf{v}}\hat{\mathbf{v}}^*, \hat{\mathbf{x}}^a)$ is feasible for (P0'), it is optimal for (P0).

B. Poof of Lemma 2

For any two attacks $\mathbf{x}^{a1}$ and $\mathbf{x}^{a2}$, let the optimal states be denoted as $(\hat{\mathbf{v}}^{(1)}, \hat{\mathbf{W}}^{(1)})$ and $(\hat{\mathbf{v}}^{(2)}, \hat{\mathbf{W}}^{(2)})$. Then, for every

$\lambda \in [0,1]$, 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 $\lambda\mathbf{x}^{a1} + (1-\lambda)\mathbf{x}^{a2}$:

$$g(\lambda\mathbf{x}^{a1} + (1-\lambda)\mathbf{x}^{a2}) \leq \lambda g(\mathbf{x}^{a1}) + (1-\lambda)g(\mathbf{x}^{a2}),$$

which proves the convexity. The subdifferentiability follows from [28].

C. Proof of Theorem 2

First, we show that $\text{rank}(\hat{\mathbf{W}}) = 1$, $\hat{\mathbf{W}} = \hat{\mathbf{v}}\hat{\mathbf{v}}^*$, and $\hat{\mathbf{x}}^a$ is unobservable. The rank-1 condition is derived by solving (FDIA-SE) with the injection $\hat{\mathbf{x}}^a$ fixed at its optimal value. 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 $\text{rank}(\mathbf{H}(\hat{\boldsymbol{\xi}})) = n_b - 1$ (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

$$\hat{\xi}_{pf,l+n_b}^{(l)} = -\frac{2\Im\left(\hat{v}_s\hat{v}_t^*\tilde{M}_{pf,st}^{(l)*}\right)}{\Im\left(\hat{v}_s\hat{v}_t^*y_{st}^*\right)}, \hat{\xi}_{pf,t}^{(l)} = \frac{|\hat{v}_s|^2\Im\left(\tilde{M}_{pf,st}^{(l)*}y_{st}\right)}{\Im\left(\hat{v}_s\hat{v}_t^*y_{st}^*\right)}$$
$$\hat{\xi}_{pf,s}^{(l)} = \frac{|\hat{v}_t|^2}{|\hat{v}_s|^2}\hat{\xi}_{pf,t}^{(l)} + \Re(y_{st})\hat{\xi}_{pf,l+n_b}^{(l)} \tag{10}$$

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

$$\hat{\xi}_{pt,l+n_b+n_l}^{(l)} = -\frac{2\Im\left(\hat{v}_s\hat{v}_t^*\tilde{M}_{pt,st}^{(l)*}\right)}{\Im\left(\hat{v}_s\hat{v}_t^*y_{st}\right)}, \hat{\xi}_{pt,t}^{(l)} = -\frac{|v_s|^2\Im\left(\tilde{M}_{pt,st}^{(l)}y_{st}\right)}{\Im\left(\hat{v}_s\hat{v}_t^*y_{st}\right)}$$
$$\hat{\xi}_{pt,s}^{(l)} = \frac{|\hat{v}_t|^2}{|\hat{v}_s|^2}\hat{\xi}_{pt,t}^{(l)} + \Re(y_{st})\hat{\xi}_{pt,l+n_b+n_l}^{(l)} \tag{11}$$

where $\hat{\mathbf{v}}$ is the optimal solution of the primal (FDIA-SE) after fixing $\hat{\mathbf{x}}^a$ at its optimal value. 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:

$$-\pi \leq \angle\hat{v}_s - \angle\hat{v}_t - \angle y_{st} \leq 0 \tag{12}$$
$$0 \leq \angle\hat{v}_s - \angle\hat{v}_t + \angle y_{st} \leq \pi \tag{13}$$
$$-\pi \leq \angle y_{st} - \angle\tilde{M}_{pf,st}^{(l)} \leq 0 \tag{14}$$
$$\pi \leq \angle y_{st} + \angle\tilde{M}_{pt,st}^{(l)} \leq 2\pi. \tag{15}$$

The inequalities (12) and (13) are satisfied by Assumption 3, which is often valid for real-world power systems. The inequalities (14) and (15) 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.

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 $\boldsymbol{\nu} \in \text{null}(\mathbf{H}(\hat{\boldsymbol{\xi}}))$, since $\mathbf{H}_{pf}^{(l)} \succeq 0$ and $\mathbf{H}_{pt}^{(l)} \succeq 0$, we have $\mathbf{H}_{pf}^{(l)}\boldsymbol{\nu} = \mathbf{H}_{pt}^{(l)}\boldsymbol{\nu} = \mathbf{0}$. By the construction of (10) and (11), for every line $l : (s,t)$, it holds that $\frac{\nu_s}{\hat{v}_s} = \frac{\nu_t}{\hat{v}_t}$. This reasoning can be applied to another line $l' : (t,a)$ to obtain $\frac{\nu_t}{\hat{v}_t} = \frac{\nu_a}{\hat{v}_a}$. By repeating the argument over a connected spanning graph of the network, one can obtain:

$$\frac{\nu_s}{\hat{v}_s} = \frac{\nu_t}{\hat{v}_t} = \frac{\nu_a}{\hat{v}_a} = \cdots = c \tag{16}$$

which indicates that $\boldsymbol{\nu} = \gamma\hat{\mathbf{v}}$. 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 $\text{rank}(\hat{\mathbf{W}}) = 1$. Based on a proof similar to Theorem 1, we have $\hat{\mathbf{W}} = \hat{\mathbf{v}}\hat{\mathbf{v}}^*$. The unobservability of $\hat{\mathbf{x}}^a$ follows immediately from the constraint $\text{trace}\,(\hat{\mathbf{v}}^*\mathbf{M}_i\hat{\mathbf{v}}) = m_i + \hat{x}_i^a,\ \forall i \in [n_m]$.

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

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

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

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

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

$$g(\hat{\mathbf{x}}^a) - g(\mathbf{x}^{a,\star}) \geq \langle\partial g(\mathbf{x}^{a,\star}), \hat{\boldsymbol{\Delta}}\rangle \geq -|\langle\partial g(\mathbf{x}^{a,*}), \hat{\boldsymbol{\Delta}}\rangle| \tag{17}$$
$$\geq -\|\partial g(\mathbf{x}^{a,*})\|_\infty\|\hat{\boldsymbol{\Delta}}\|_1 \tag{18}$$
$$\geq -\frac{\alpha}{2}\left(\|\hat{\boldsymbol{\Delta}}_{\mathcal{B}}\|_1 + \|\hat{\boldsymbol{\Delta}}_{\mathcal{B}^c}\|_1\right) \tag{19}$$
$$\geq -2\alpha\|\hat{\boldsymbol{\Delta}}_{\mathcal{B}}\|_1 \tag{20}$$

where (17) is due to the convexity of $g(\mathbf{x}^a)$ (Lemma 2), (18) is by Hölder's inequality, (19) is due to the assumption of α, and (20) is due to Lemma 3 (see [29], Lemma 1).