Stealth Data Injection Attacks with Sparsity Constraints

Xiuzhen Ye, Iñaki Esnaola, Samir M. Perlaza, Robert F Harrison

Stealth Data Injection Attacks with Sparsity Constraints

Xiuzhen Ye, Iñaki Esnaola, Samir M. Perlaza, and Robert F. Harrison

***Abstract*—Sparse stealth attack constructions that minimize the mutual information between the state variables and the observations are proposed. The attack construction is formulated as the design of a multivariate Gaussian distribution that aims to minimize the mutual information while limiting the Kullback-Leibler divergence between the distribution of the observations under attack and the distribution of the observations without attack. The sparsity constraint is incorporated as a support constraint of the attack distribution. Two heuristic greedy algorithms for the attack construction are proposed. The first algorithm assumes that the attack vector consists of independent entries, and therefore, requires no communication between different attacked locations. The second algorithm considers correlation between the attack vector entries and achieves a better disruption to stealth tradeoff at the cost of requiring communication between different locations. We numerically evaluate the performance of the proposed attack constructions on IEEE test systems and show that it is feasible to construct stealth attacks that generate significant disruption with a low number of compromised sensors.**

I. Introduction

Monitoring and controlling processes that are supported by supervisory control and data acquisition (SCADA) systems facilitate an economic and reliable operation of the power system [1]. The integration between the physical layer of the power system and the cyber layer enables efficient, scalable, and secure operation of the system [2]. While advanced communication systems that acquire and transmit observations to a state estimator provide reliable and low-latency state information [3], this cyber layer also exposes the system to malicious attacks. One of the main cybersecurity threats faced by modern power systems are data injection attacks (DIAs), which were first introduced in [4]. DIAs alter the state estimate of the system by compromising the system observations and altering the data without triggering the data detection mechanisms set by the system operator. A large body of literature studies the case in which attack detection is performed by a residual test [5] under the assumption that state estimation is deterministic both in centralized and decentralized scenarios [6], [7], [8], [9]. In this setting, attack construction that requires access to a small set of observations yields l_0-norm minimization problems, which are in general hard to solve. In [10], it is shown that the operator can secure a small fraction of observations to make undetectable attack constructions significantly harder.

The unprecedented data acquisition capabilities that are now available to cyberphysical systems promote the efficient operation of the smart grid but also increase the threat posed by DIAs because accurate stochastic models of the system can be generated. This problem is cast in a Bayesian framework in [11]. In this Bayesian paradigm, the attack detection can be formulated as the likelihood ratio test [12] or alternatively machine learning methods [13] can be employed to learn the geometry of the data generated by the system. Data analytics are increasingly important in the operation of modern power systems and they are central to the advanced estimation, control, and management of the smart grid [14]. For this reason, it is essential to study attack constructions in fundamental terms to understand the impact over a wide range of data analysis paradigms.

Stealth data injection attacks within Bayesian framework were first introduced in [15] and then generalized in [16]. In this research, the attack construction uses information theoretic measures, i.e. mutual information and Kullback-Leibler (KL) divergence, to characterize the fundamental limits of the attack. In [11] [15] [16] [17], the state variables are assumed to follow a Gaussian distribution. From a practical point of view, the adoption of Gaussian random vectors as the data injection attack vectors is validated by real data [18] [19]. However, both the stealth attacks con-

This research was supported in part by the European Commission through the H2020-MSCA-RISE-2019 program under grant 872172 and in part by the China Scholarship Council.

X. Ye, I. Esnaola, and R. F. Harrison are with the Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield S1 3JD, UK. I. Esnaola is also with the Department of Electrical Engineering, Princeton University, Princeton NJ 08544, USA. (email: xye15@sheffield.ac.uk, esnaola@sheffield.ac.uk, r.f.harrison@sheffield.ac.uk).

S. M. Perlaza is with INRIA at Sophia Antipolis, France; also with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA; and also with the GAATI Laboratory of the University of French Polynesia, Faaa, Tahiti. (email: samir.perlaza@inria.fr).

structed in [15] and [16] require that the attacker tampers with all the observations in the system, which is not feasible in most scenarios. Information theoretic attack constructions that incorporate sparsity constraints are first proposed in [17]. However, the construction of attack vectors that effectively exploits the correlation between attack variables is still an open problem that requires novel approaches. In this paper, we present novel sparse stealth attack constructions that leverage the coordination between different attacked observations to attain a better attack disruption to stealth tradeoff.

The rest of the paper is organized as follows: In Section II, we introduce a Bayesian framework with linearized dynamics for DIAs. Stealth attacks incorporating sparsity constraints are presented in Section III. Independent sparse stealth attacks and correlated sparse stealth attacks are presented in Section IV and Section V, respectively. In Section VI, we evaluate the performance of the proposed attack constructions for both independent and correlated scenarios on IEEE test systems. The paper closes with conclusions in Section VII.

Notation: We denote the number of state variables on a given IEEE test system by n and the number of the observations by m. The set of positive semidefinite matrices of size $n \times n$ is denoted by S_+^n. The n-dimensional identity matrix is denoted as $\mathbf{I}_n$. The elementary vector $\mathbf{e}_i \in \mathbb{R}^n$ is a vector of zeros with a one in the i-th entry. Random variables are denoted by capital letters and their realizations by the corresponding lower case, e.g. x is a realization of the random variable X. Vectors of n random variables are denoted by a superscript, e.g. $X^n = (X_1, \ldots, X_n)^\mathsf{T}$ with corresponding realizations denoted by $\mathbf{x}$. Given an n-dimensional vector $\boldsymbol{\mu} \in \mathbb{R}^n$ and a matrix $\boldsymbol{\Sigma} \in S_+^n$, we denote by $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ the multivariate Gaussian distribution of dimension n with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The mutual information between random variables X and Y is denoted by $I(X;Y)$ and the Kullback-Leibler (KL) divergence between the distributions P and Q is denoted by $D(P\|Q)$.

II. System model

A. Observation Model and Attack Setting

The operation state of a power system is described by a vector $\mathbf{x} \in \mathbb{R}^n$ containing the voltages and phases at all the generation and load buses. The state vector $\mathbf{x}$ is observed through the acquisition function $F : \mathbb{R}^n \to \mathbb{R}^m$. When a linearized observation model is considered for state estimation, it yields an observation model of the form

$$Y^m = \mathbf{H}\mathbf{x} + Z^m, \tag{1}$$

where $\mathbf{H} \in \mathbb{R}^{m \times n}$ is the Jacobian of the function F at a given operating point and is determined by the system components and the topology of the network. The vector Y^m containing the observations is corrupted by additive white Gaussian noise introduced by the sensors, c.f., [2] and [3]. Such noise is modelled by the vector Z^m in (1), which follows a multivariate Gaussian distribution. That is,

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

where σ^2 is the noise variance.

In a Bayesian estimation framework, the state variables are described by a random vector X^n with a given distribution. In this study, the random vector X^n is assumed to follow a multivariate Gaussian distribution with a null mean vector and covariance matrix

$$\boldsymbol{\Sigma}_{XX} \in S_+^n. \tag{3}$$

Hence, the vector of observations Y^m in (1) follows a multivariate Gaussian distribution with null mean vector and a covariance matrix $\boldsymbol{\Sigma}_{YY}$ satisfying that

$$\boldsymbol{\Sigma}_{YY} \triangleq \mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T} + \sigma^2 \mathbf{I}_m. \tag{4}$$

The resulting observations are corrupted by a malicious attack vector $A^m \sim P_{A^m}$, where P_{A^m} is the distribution of the random vector A^m. In the following, P_{A^m} is assumed to be a multivariate Gaussian distribution that satisfies

$$A^m \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{AA}), \tag{5}$$

where $\mathbf{0} = (0, 0, \ldots, 0)$ and $\boldsymbol{\Sigma}_{AA} \in S_+^m$ are the mean vector and the covariance matrix of the random vector A^m.

The choice in (5) is justified by the fact that a multivariate Gaussian distribution minimizes the mutual information between the state variables and the compromised observations under the assumption that the covariance matrix $\boldsymbol{\Sigma}_{YY}$ is fixed [20]. Consequently, the compromised observations denoted by Y_A^m are given by

$$Y_A^m = \mathbf{H}X^n + Z^m + A^m, \tag{6}$$

where Y_A^m follows a multivariate Gaussian distribution given by

$$Y_A^m \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{Y_A Y_A}) \tag{7}$$

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

B. Attack Detection

As a part of a security strategy, the operator implements an attack detection procedure prior to performing

state estimation. Detection is cast as a hypothesis testing problem given by

$$\mathcal{H}_0 : \text{There is no attack,} \tag{8a}$$

$$\mathcal{H}_1 : \text{Observations are compromised.} \tag{8b}$$

At time step $i \in \mathbb{N}$, the system operator acquires a vector of observations $\bar{Y}_i^m$ and decides whether the vector of observations $\bar{Y}_i^m$ is produced following a no attack scenario as described in (1) or is the result of the attack as described in (6). In our setting, the hypothesis test can be recast in terms of the probability density functions induced by the state variables, the system noise, and the attack onto the observations $\bar{Y}^m$. Hence, the hypotheses in (8) become

$$\mathcal{H}_0 : \bar{Y}^m \sim P_{Y^m}, \tag{9a}$$

$$\mathcal{H}_1 : \bar{Y}^m \sim P_{Y_A^m}. \tag{9b}$$

A test to determine what distribution generates the observation data is a deterministic test $T : \mathbb{R}^m \to \{0, 1\}$. Given an observation vector $\bar{\mathbf{y}}$, let $T(\bar{\mathbf{y}}) = 0$ denote the case in which the test decides $\mathcal{H}_0$ upon the observation of $\bar{\mathbf{y}}$; and $T(\bar{\mathbf{y}}) = 1$ the case in which the test decides $\mathcal{H}_1$. The performance of the test is assessed in terms of the Type-I error, denoted by $\alpha \triangleq \mathbb{P}\left[T\left(\bar{Y}^m\right) = 1\right]$, with $\bar{Y}^m \sim P_{Y^m}$; and the Type-II error, denoted by $\beta \triangleq \mathbb{P}\left[T\left(\bar{Y}^m\right) = 0\right]$, with $\bar{Y}^m \sim P_{Y_A^m}$. Given the requirement that the Type-I error satisfies $\alpha \leq \alpha'$, with $\alpha' \in [0, 1]$, the likelihood ratio test (LRT) is optimal in the sense that it induces the smallest Type-II error β [21]. In this setting, the LRT is given by

$$T(\bar{\mathbf{y}}) = \mathbb{1}_{\{L(\bar{\mathbf{y}}) \gtrless \tau\}}, \tag{10}$$

with $L(\bar{\mathbf{y}})$ the likelihood ratio, i.e.,

$$L(\bar{\mathbf{y}}) = \frac{f_{Y_A^m}(\bar{\mathbf{y}})}{f_{Y^m}(\bar{\mathbf{y}})}, \tag{11}$$

where the functions $f_{Y_A^m}$ and f_{Y^m} are respectively the probability density function (pdf) of Y_A^m in (6) and the pdf of Y^m in (1); and $\tau \in \mathbb{R}_+$ in (10) is the decision threshold. Note that changing the value of τ is equivalent to changing the tradeoff between Type-I and Type-II errors.

III. Sparse Stealth Attacks

A. Information Theoretic Metric

The aim of the attacker is twofold. First, it aims to inflict a data integrity attack that disrupts all processes that use the observations of the system; and second, to guarantee a stealthy attack. Hence, instead of assuming a particular state estimation procedure, we adopt the methodology in [16] to construct stealth attacks that minimize the amount of information acquired by the observations about the state variables. In doing so, the attacker targets a universal utility metric consisting in a weighted sum of two terms: (a) the mutual information between the state variables and the observations; and (b) the KL divergence between the probability distribution functions of the observations with and without attack. By minimizing this metric, the attacker guarantees a stealthy attack that impinges upon any procedure using the observations.

The KL divergence term guarantees a stealthy attack in the sense that its minimization leads to minimizing the absolute difference between Type-I and Type-II probability of errors, i.e., $|\alpha - \beta|$, for a given mutual information target [21].

Within this framework, stealth attacks are constructed as random vectors whose probability distribution functions are the solution to the following optimization problem:

$$\min_{P_{A^m}} I(X^n; Y_A^m) + \lambda D(P_{Y_A^m} \| P_{Y^m}), \tag{12}$$

where the optimization domain is the set of all possible m-dimensional Gaussian probability distributions; and $\lambda \geq 1$ is a weighting parameter that determines the tradeoff between the attack disruption and probability of attack detection.

The solution to the optimization in (12) is a multivariate Gaussian distribution for the attack vector. It is shown in [16] that the optimal Gaussian attack is given by $\bar{P}_{A^m} \sim \mathcal{N}(\mathbf{0}, \bar{\mathbf{\Sigma}})$ where

$$\bar{\mathbf{\Sigma}} = \lambda^{-1/2} \mathbf{H} \mathbf{\Sigma}_{XX} \mathbf{H}^{\mathsf{T}}. \tag{13}$$

Note that (13) yields a stealth attack vector that is not sparse, indeed all the components of the attack realizations are nonzero with probability one, i.e. $\mathbb{P}\left[|\text{supp}(A^m)| = m\right] = 1$, where we define the support of the attack vector A^m as

$$\text{supp}(A^m) \triangleq \{i : \mathbb{P}\left[A_i = 0\right] = 0\}. \tag{14}$$

B. Sparse Stealth Attack Formulation

The attack implementation requires access to the sensing infrastructure of the industrial control system (ICS) operating the power system. Data injection attacks usually exploit the vulnerabilities existing in the field zone by comprising remote terminal units or local secondary level control systems, or alternatively, by getting access to the SCADA system coordinating the control zone of the ICS. For that reason, attack constructions that are required to intrude the least amount of monitoring and

data acquisition infrastructure are particularly interesting. In view of this, we study sparse attacks that require access to a limited number of sensors, i.e. we pose the attack construction problem with sparsity constraints by setting the domain as the set of distributions over the attack vector that put non-zero mass on at most $k \leq m$ attack vector components.

In our formulation, this is reflected by an additional optimization constraint of the form $|\text{supp}(A^m)| = k$, for some given $k \leqslant m$. Hence, the attacker chooses the distribution over the set of multivariate Gaussian distributions given by

$$\mathcal{P}_k \triangleq \left\{ P_{A^m} \sim \mathcal{N}(\mathbf{0}, \bar{\mathbf{\Sigma}}) : |\text{supp}(A^m)| = k \right\}. \quad (15)$$

The resulting k-sparse stealth attack construction is therefore posed as the optimization problem:

$$\min_{P_{A^m} \in \mathcal{P}_k} I(X^n; Y_A^m) + \lambda D(P_{Y_A^m} \| P_{Y^m}). \quad (16)$$

The optimization domain including the sparcity constraint in (15) implies an additional difficulty in the construction of stealth attacks with respect to the construction proposed in [16]. This additional difficulty lies on the combinatorial problem arising from the selection of at most k out of m dimensions of the vector attack to form the support of A^m. To tackle this difficulty, we exploit the structure that the Gaussian attack embeds into the sparse attack problem formulation to propose novel attack construction algorithms with verifiable performance guarantees.

C. Gaussian Sparse Stealth Attack Construction

The probability distribution function of a random vector is determined by two parameters, i.e., the mean vector and the covariance matrix. Hence, writing the objective function of the optimization problems in (12) and (16) in terms of the mean vector and covariance matrix of the attack random vector A^m leads to observing that it is equal to the following expression, up to a constant additive term,

$$\begin{aligned} J(\mathbf{\Sigma}_{AA}) \triangleq & (1-\lambda) \log |\mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_{AA}| \\ & - \log |\sigma^2 \mathbf{I}_m + \mathbf{\Sigma}_{AA}| + \lambda \text{tr}(\mathbf{\Sigma}_{YY}^{-1} \mathbf{\Sigma}_{AA}), \end{aligned} \quad (17)$$

where $\lambda \geq 1$ is introduced in (12); and the matrix $\mathbf{\Sigma}_{YY}$ is defined by (4).

Hence, the optimization problem in (12) is equivalent to the following optimization problem:

$$\min_{\mathbf{\Sigma}_{AA} \in S_+^m} J(\mathbf{\Sigma}_{AA}). \quad (18)$$

In order to write the optimization domain of the problem in (16) in terms of the mean vector and covariance matrix of the attack random vector, it suffices to observe that the sparsity constraint in (15) translates into a constraint on the number of nonzero entries in the diagonal of the covariance matrix of the attack vector. More specifically, the optimization domain becomes:

$$\mathcal{S}_k \triangleq \left\{ \mathbf{S} \in S_+^m : \|\text{diag}(\mathbf{S})\|_0 = k \right\}, \quad (19)$$

where $\text{diag}(\mathbf{S})$ denotes the vector formed by the diagonal entries of $\mathbf{S}$. Solving (18) within the optimization domain specified by (19) re-casts the equivalent k-sparse stealth attack construction problem in (16) as:

$$\min_{\mathbf{\Sigma}_{AA} \in \mathcal{S}_k} J(\mathbf{\Sigma}_{AA}). \quad (20)$$

IV. Independent Sparse Stealth Attacks

We first tackle the case in which the attack vector entries are independent. More specifically, the focus is on product probability measures of the form

$$P_{A^m} = \prod_{i=1}^{m} P_{A_i}, \quad (21)$$

where, for all $i \in \{1, 2, \ldots, m\}$, the probability density function of the measure P_{A_i} is Gaussian with zero measure and variance σ_i^2.

The assumption of independence relaxes the correlation requirements between the components of the attack vector. As a result, the set of covariance matrices given by (19), with $k \leqslant m$, that arises from considering Gaussian attacks is the set

$$\tilde{\mathcal{S}}_k \triangleq \bigcup_{\mathcal{K}} \left\{ \mathbf{S} \in S_+^m : \mathbf{S} = \sum_{i \in \mathcal{K}} v_i \mathbf{e}_i \mathbf{e}_i^{\mathsf{T}} \text{ with } v_i \in \mathbb{R}_+ \right\}, \quad (22)$$

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

Under the independence assumption adopted in this section, the optimization problem in (18) boils down to the following problem:

$$\min_{\mathbf{\Sigma}_{AA} \in \tilde{\mathcal{S}}_k} J(\mathbf{\Sigma}_{AA}), \quad (23)$$

which is hard to solve due to the combinatorial character of identifying the support of the sparse random attack vector. To circumvent this problem, we propose a greedy construction that sequentially updates the set $\text{supp}(A^m)$ in (14) and determines the corresponding entry in the diagonal of the matrix $\mathbf{\Sigma}_{AA}$ in (5).

A. Greedy Independent Attack Construction

The proposed construction hinges on the idea that approaching the sensor selection problem in a sequential fashion resembles the single sensor selection problem discussed in [17]. This enables us to leverage the single sensor selection construction to analytically characterize the cost difference induced by the addition of a new element to the set $\text{supp}(A^m)$ in (14).

More specifically, given the sparsity constraint in (19), for some $k \leqslant m$, the construction can be divided into k epochs. At each epoch a new element is added to $\text{supp}(A^m)$. At epoch i, let $\mathbf{\Sigma}_i \in S_+^m$ be the covariance matrix of the vector attack under construction. Let the set $\mathcal{A}_i$ be the set of indices corresponding to the entries of the vector $\text{diag}(\mathbf{\Sigma}_i)$ that are different from zero. That is,

$$\mathcal{A}_i = \{j \in \{1, 2, \ldots, m\} : \mathbf{e}_j^\mathsf{T} \mathbf{\Sigma}_i \mathbf{e}_j > 0\}. \tag{24}$$

For all $i \in \{1, 2, \ldots, k\}$, it is imposed that $\mathcal{A}_i \subseteq \{1, 2, \ldots, m\}$ and $|\mathcal{A}_i| = i$. This implies that $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \ldots \subset \mathcal{A}_k \subset \{1, 2, \ldots, m\}$. Hence,

$$\mathbf{\Sigma}_i = \mathbf{\Sigma}_{i-1} + v \mathbf{e}_j \mathbf{e}_j^\mathsf{T}, \tag{25}$$

where $\mathbf{\Sigma}_0$ is a matrix of zeros; the integer $j \in \{1, 2, \ldots, m\} \setminus \mathcal{A}_{i-1}$ is the index of the new entry at epoch i; and $v > 0$ is the value of such entry. For ease of presentation we denote the set of indices available to the attacker to choose at epoch i, i.e. the entries of the vector $\text{diag}(\mathbf{\Sigma}_{i-1})$ that are zero, as

$$\mathcal{A}_{i-1}^\mathsf{c} \triangleq \{1, 2, \ldots, m\} \setminus \mathcal{A}_{i-1}. \tag{26}$$

Our proposition to choose both $j \in \mathcal{A}_{i-1}^\mathsf{c}$ and $\theta > 0$ at epoch i as described in (25) is based on the following optimization problem

$$\min_{(j,v) \in \mathcal{A}_{i-1}^\mathsf{c} \times \mathbb{R}_+} J(\mathbf{\Sigma}_{i-1} + v \mathbf{e}_j \mathbf{e}_j^\mathsf{T}). \tag{27}$$

The following lemma sheds light on the solution to the problem (27).

Lemma 1. *Let $\mathbf{\Sigma}_1 \in S_+^m$ and $\mathbf{\Sigma}_2 \in S_+^m$ be two matrices that satisfy $\mathbf{\Sigma}_2 = \mathbf{\Sigma}_1 + \mathbf{\Delta}$, with $\mathbf{\Delta} \in \mathbb{R}^{m \times m}$. Then, the cost function J in (17) satisfies that*

$$J(\mathbf{\Sigma}_2) = J(\mathbf{\Sigma}_1) + f(\mathbf{\Sigma}_1, \mathbf{\Delta}), \tag{28}$$

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

$$\begin{aligned} f(\mathbf{\Sigma}_1, \mathbf{\Delta}) = & (1 - \lambda) \log \left| \mathbf{I}_m + \left(\mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_1 \right)^{-1} \mathbf{\Delta} \right| \\ & - \log \left| \mathbf{I}_m + \left(\sigma^2 \mathbf{I}_m + \mathbf{\Sigma}_1 \right)^{-1} \mathbf{\Delta} \right| \\ & + \lambda \text{tr} \left(\mathbf{\Sigma}_{YY}^{-1} \mathbf{\Delta} \right), \end{aligned} \tag{29}$$

the $\lambda \geq 1$ is introduced in (12); and the matrix $\mathbf{\Sigma}_{YY}$ is defined by (4).

Proof. The proof consists in showing that the difference between $J(\mathbf{\Sigma}_2)$ and $J(\mathbf{\Sigma}_1)$ yields

$$\begin{aligned} J(\mathbf{\Sigma}_2) - J(\mathbf{\Sigma}_1) = & (1 - \lambda) \log \left| \mathbf{I}_m + \left(\mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_1 \right)^{-1} \mathbf{\Delta} \right| \\ & - \log \left| \mathbf{I}_m + \left(\sigma^2 \mathbf{I}_m + \mathbf{\Sigma}_1 \right)^{-1} \mathbf{\Delta} \right| \\ & + \lambda \text{tr} \left(\mathbf{\Sigma}_{YY}^{-1} \mathbf{\Delta} \right), \end{aligned} \tag{30}$$

which completes the proof. □

The relevance of Lemma 1 is that it enables the selection of both $j \in \mathcal{A}_{i-1}^\mathsf{c}$ and $v > 0$ at epoch i based on a simpler optimization problem than that in (27). Indeed, the selection problem results in

$$\min_{(j,v) \in \mathcal{A}_{i-1}^\mathsf{c} \times \mathbb{R}_+} f(\mathbf{\Sigma}_{i-1}, v \mathbf{e}_j \mathbf{e}_j^\mathsf{T}), \tag{31}$$

where the function f is defined in (29). Theorem 2 provides the solution to the optimization problem in (31).

Theorem 2. *Let k satisfy $0 < k \leqslant m$, and for all $i \in \{1, 2, \ldots, k\}$, denote by $(j^\star, v^\star) \in \mathcal{A}_{i-1}^\mathsf{c} \times \mathbb{R}_+$ the solution to the optimization problem in (27). Then, the following holds*

$$j^\star = \arg\min_{j \in \mathcal{A}_{i-1}^\mathsf{c}} J(\mathbf{\Sigma}_{i-1} + v_j \mathbf{e}_j \mathbf{e}_j^\mathsf{T}) \quad \textit{and} \tag{32}$$

$$v^\star = v_{j^\star}, \tag{33}$$

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

$$\begin{aligned} v_{j^*} = & \left(\frac{\beta_j - \alpha_j + \beta_j \alpha_j \sigma^2}{2 \beta_j \alpha_j} \right) \\ & \cdot \left(\sqrt{1 - \frac{4 \beta_j \alpha_j \left(\beta_j \sigma^2 - \alpha_j \sigma^2 - \dfrac{\alpha_j \sigma^2 + 1}{\lambda} \right)}{\left(\beta_j - \alpha_j + \beta_j \alpha_j \sigma^2 \right)^2}} - 1 \right) \end{aligned} \tag{34}$$

with

$$\alpha_j \triangleq \text{tr} \left(\left(\mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_{i-1} \right)^{-1} \mathbf{e}_{j^*} \mathbf{e}_{j^*}^\mathsf{T} \right), \tag{35}$$

$$\beta_j \triangleq \text{tr} \left(\mathbf{\Sigma}_{YY}^{-1} \mathbf{e}_{j^*} \mathbf{e}_{j^*}^\mathsf{T} \right), \tag{36}$$

and the real $\sigma > 0$ in (34) is introduced in (2).

Proof. It follows from Lemma 1 that the optimization problem in (31) is equivalent to

$$\begin{aligned} \min_{(j,v) \in \mathcal{A}_{i-1}^\mathsf{c} \times \mathbb{R}_+} & (1 - \lambda) \log \left| \mathbf{I}_m + \left(\mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_{i-1} \right)^{-1} v \mathbf{e}_j \mathbf{e}_j^\mathsf{T} \right| \\ & - \log \left| \mathbf{I}_m + \left(\sigma^2 \mathbf{I}_m + \mathbf{\Sigma}_{i-1} \right)^{-1} v \mathbf{e}_j \mathbf{e}_j^\mathsf{T} \right| \\ & + \lambda \text{tr} \left(\mathbf{\Sigma}_{YY}^{-1} v \mathbf{e}_j \mathbf{e}_j^\mathsf{T} \right). \end{aligned} \tag{37}$$

After some algebraic manipulation it follows that

$$\min_{(j,v)\in\mathcal{A}_{i-1}^{\mathsf{c}}\times\mathbb{R}_+}(1-\lambda)\log(1+\alpha_j v)-\log(1+\frac{v}{\sigma^2})+\lambda\beta_j v, \tag{38}$$

which is convex for $\lambda \geq 1$. The only solution of the minimization problem in (38) is obtained by letting the derivative to zero, which yields

$$\beta_j\alpha_j v^2+(\beta_j-\alpha_j+\beta_j\alpha_j\sigma^2)v+\beta_j\sigma^2-\alpha_j\sigma^2-\frac{\alpha_j\sigma^2+1}{\lambda}=0 \tag{39}$$

Note that (39) is quadratic with two solutions. The result follows by choosing the solution such that $v \in \mathbb{R}_+$. This completes the proof. □

The proposed greedy construction is described in Algorithm 1.

Algorithm 1 k-sparse independent attack construction

Input: $\mathbf{H}$ in (1);
σ^2 in (2);
$\mathbf{\Sigma}_{XX}$ in (3);
λ in (17); and
k in (24).
Output: $\mathbf{\Sigma}_{AA}$ in (5).

Set $\mathcal{A}_0 = \{\emptyset\}$
Set $\mathbf{\Sigma}_0 = \mathbf{0}$
for $j = 1$ to k **do**
  for $\ell \in \mathcal{A}_{i-1}^{\mathsf{c}}$ **do**
    Compute v_ℓ in (34)
  end for
  Compute $j^\star$ in (32)
  Compute $v^\star$ in (33)
  Set $\mathcal{A}_j = \mathcal{A}_{j-1} \cup \{j^\star\}$
  Set $\mathbf{\Sigma}_j = \sum_{i\in\mathcal{A}_j} v_i \mathbf{e}_i \mathbf{e}_i^{\mathsf{T}}$
end for
$\mathbf{\Sigma}_{AA} = \sum_{i\in\mathcal{A}_k} v_i \mathbf{e}_i \mathbf{e}_i^{\mathsf{T}}$

V. Correlated Sparse Stealth Attacks

A. Correlation Structure

In this section, the assumption of independence in (21) is dropped. This case boils down to the attack construction given in (20), i.e. the optimization is carried over the set of covariance matrices with non-zero off-diagonal entries that account for the correlation between different attack entries. In this case the addition of a new index to the set of k attacked observations introduces off-diagonal entries in the difference between covariance matrices described in Lemma 1. More precisely, the difference introduced by selecting the index i is given by $\Delta_i \in \mathcal{D}_i$ with

$$\mathcal{D}_i = \bigcup_{\mathbf{s}\in\mathbb{R}^m} \left\{\mathbf{D}\in\mathbb{R}^{m\times m} : \mathbf{D} = \mathbf{s}^{\mathsf{T}}\otimes\mathbf{e}_i + \mathbf{s}\otimes\mathbf{e}_i^{\mathsf{T}}\right\}. \tag{40}$$

Note that the vector $\mathbf{s}$ determines the second order moments describing the covariance between attacked observations. As in the independent case, characterizing the difference enables to formulate the optimization problem that yields the minimum cost increase introduced by a new index in the attack support. Let $\mathcal{A}_{k-1}$ denote set of indices of attacked observations and $\mathbf{\Sigma}_{i-1} \in \mathcal{S}_{i-1}$ the covariance matrix of the attack vector over those $i-1$ observations. Then the sensor selection problem at step i is given by the optimization problem:

$$\begin{aligned} \min_{j,\mathbf{\Delta}}\ & J\left(\mathbf{\Sigma}_{i-1}+\mathbf{\Delta}\right) \\ \text{s.t.}\ & j\in\mathcal{A}_{i-1}^{\mathsf{c}}, \\ & \mathbf{\Delta}\in\mathcal{D}_j, \\ & \mathbf{\Sigma}_{i-1}+\mathbf{\Delta}\in S_+^m. \end{aligned} \tag{41}$$

In the following we show that when the choice of the next index selected for attacks is fixed, the optimization in (41) is convex in the matrix difference.

Theorem 3. *Let $\mathbf{\Sigma}_{i-1} \in \mathcal{S}_{i-1}$ and $j \in \mathcal{A}_{i-1}^{\mathsf{c}}$, then the optimization problem given by*

$$\begin{aligned} \min_{\mathbf{\Delta}}\ & J\left(\mathbf{\Sigma}_{i-1}+\mathbf{\Delta}\right) \\ \text{s.t.}\ & \mathbf{\Delta}\in\mathcal{D}_j, \\ & \mathbf{\Sigma}_{i-1}+\mathbf{\Delta}\in S_+^m, \end{aligned} \tag{42}$$

is a convex optimization problem.

Proof. It follows from Lemma 1 and some algebraic manipulation that the optimization problem in (42) is equivalent to

$$\begin{aligned} \min_{\mathbf{\Delta}}\ & (1-\lambda)\log\left|\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{i-1}+\mathbf{\Delta}\right| \\ & -\log\left|\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_{i-1}+\mathbf{\Delta}\right|+\lambda\text{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Delta}_i\right) \\ \text{s.t.}\ & \mathbf{\Delta}\in\mathcal{D}_j, \\ & \mathbf{\Sigma}_{i-1}+\mathbf{\Delta}\in S_+^m. \end{aligned} \tag{43}$$

Noting that the sets $\mathcal{D}_j$ are convex for all $j \in \mathcal{A}_{i-1}^{\mathsf{c}}$, that the logarithm terms are convex [22] for $\lambda \geq 1$, and that the trace term is linear, yields that the optimization problem in (42) is convex in $\mathbf{\Delta}$. This completes the proof. □

The proposed greedy construction for independent attack case is described in Algorithm 2. Note that the matrix obtained in the optimization problem in Theorem 3 is constrained by projecting the sum of the update and

the previous covariance matrix in the positive semidefinite cone to guarantee that the resulting covariance matrix is indeed positive semidefinite. This is reflected in the last step of Algorithm 2 where the resulting matrix construction is projected by minimizing the Frobenius distance to the positive semidefinite cone.

Algorithm 2 k-sparse correlated attack construction

Input: $\mathbf{H}$ in (1);
σ^2 in (2);
$\mathbf{\Sigma}_{XX}$ in (3);
λ in (17); and
k in (24).
Output: $\mathbf{\Sigma}_{AA}$ in (5).

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

VI. Numerical Results

In this section, we numerically evaluate the performance of the proposed attack construction algorithms on a direct current (DC) state estimation setting for the IEEE 9-Bus, IEEE 14-Bus and IEEE 30-Bus test systems [23]. The voltage magnitudes are set to 1.0 per unit, which implies that the state estimation is based on the observations of active power flow injections to all the buses and the active power flow between physically connected buses. The Jacobian matrix $\mathbf{H}$ is determined by the reactance of the branches and the topology of the corresponding systems. We use MATPOWER [24] to generate $\mathbf{H}$ for each test system. The statistical dependence between the state variables is captured by a Toeplitz model for the covariance matrix $\mathbf{\Sigma}_{XX} \in \mathcal{S}_+^n$ that arises in a wide range of practical settings, such as autoregressive stationary processes [12], [16], [25]. Specifically, we model the correlation between state variables X_i and X_j with the exponential decay parameter $\rho \in \mathbb{R}_+$ that defines the entries of the covariance matrix of the state variables as $(\mathbf{\Sigma}_{XX})_{ij} = \rho^{|i-j|}$ with $(i, j) \in \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}$.

In this setting, the performance of the proposed sparse stealth attack is not only a function of the attack constructions but also the correlation parameter ρ, the noise variance σ^2, and the topology of the system described by $\mathbf{H}$. In the simulations, we set the observation model noise regime in terms of the signal to noise ratio (SNR) defined as

$$\text{SNR} \triangleq 10 \log_{10} \left(\frac{\text{tr}(\mathbf{H}\mathbf{\Sigma}_{\text{XX}}\mathbf{H}^{\mathsf{T}})}{m\sigma^2} \right). \tag{44}$$

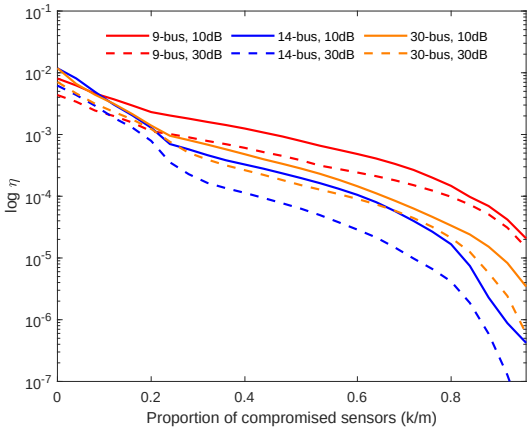

Fig. 1: Performance of independent attack constructions on different IEEE test systems with $\rho = 0.9$ and $\lambda = 8$.

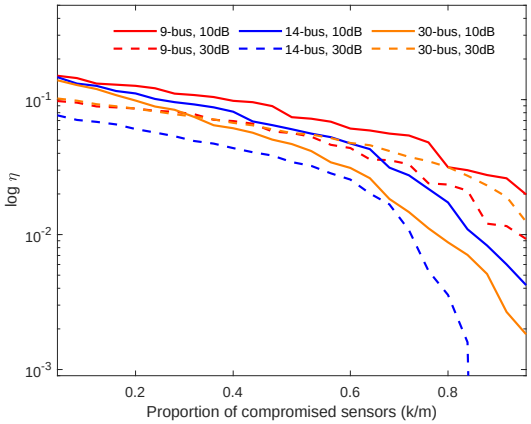

Fig. 2: Performance of correlated attack constructions on different IEEE test systems with $\rho = 0.9$ and $\lambda = 8$.

A. Performance in terms of information theoretic cost

Let $\mathbf{\Sigma}_i^k$ be the output of the k-sparse attack construction of Algorithm i. We evaluate the attack performance in terms of the sparsity penalty defined as

$$\eta \triangleq \frac{J(\mathbf{\Sigma}_i^k) - J(\mathbf{\Sigma}_i^m)}{J(\mathbf{\Sigma}_i^m)}, \tag{45}$$

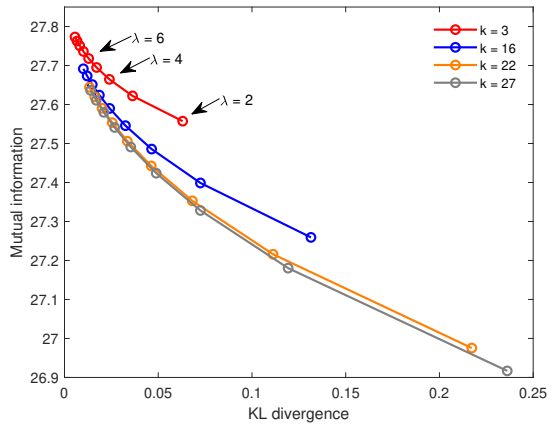

Fig. 3: Performance of independent sparse attack construction in terms of mutual information and KL divergence for different values of λ on the IEEE 9-bus system with SNR = 30 dB and $\rho = 0.9$.

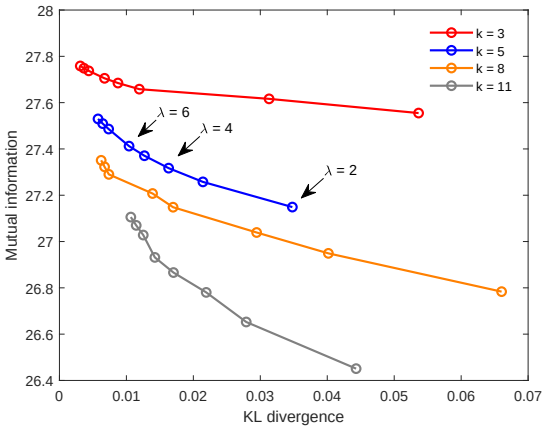

Fig. 5: Performance of correlated sparse attack construction in terms of mutual information and KL divergence for different values of λ on the IEEE 9-bus system with SNR = 30 dB and $\rho = 0.9$.

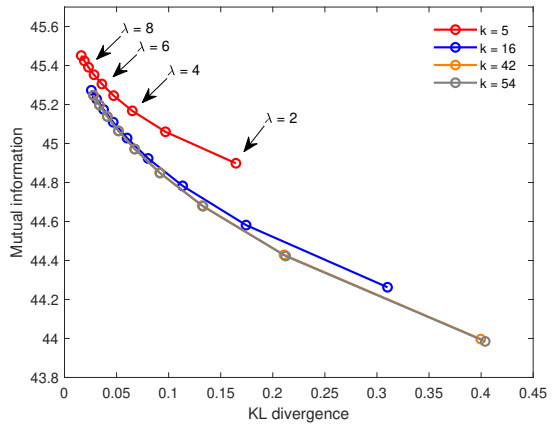

Fig. 4: Performance of independent sparse attack construction in terms of mutual information and KL divergence for different values of λ on the IEEE 14-bus system with SNR = 30 dB and $\rho = 0.9$.

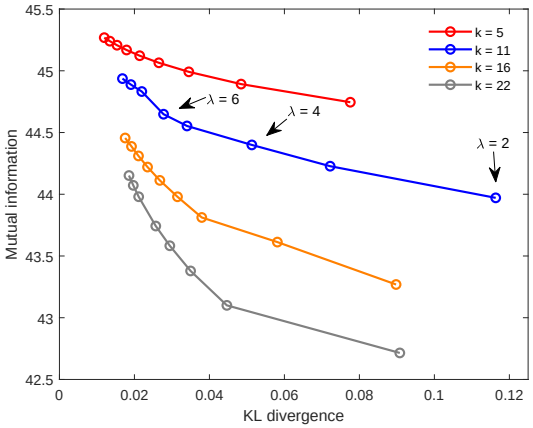

Fig. 6: Performance of correlated sparse attack construction in terms of mutual information and KL divergence for different values of λ on the IEEE 14-bus system with SNR = 30 dB and $\rho = 0.9$.

where $J(\cdot)$ is the cost defined in (17). Note that $J(\mathbf{\Sigma}_i^m)$ denotes the cost induced by the construction when all the sensors are attacked. In that sense, this metric captures the performance loss of the attack when only k sensors are attacked. Fig. 1 depicts the performance of the independent sparse stealth attack construction obtained with Algorithm 1 in different IEEE test systems as a function of the proportion of compromised sensors, i.e. k/m, for correlation parameter $\rho = 0.9$ and $\lambda = 8$. Similarly, Fig. 2 depicts the performance of the correlated sparse stealth attack construction from Algorithm 2 in the same setting as in Fig. 1. As expected, in both cases the sparsity penalty decreases monotonically with the proportion of compromised sensors. In the independent sparse attack case, the sparsity penalty does not change significantly in terms of the proportion of compromised sensors while in the Algorithm 2 construction case the sparsity penalty decreases exponentially in the number of compromised sensors. Note that the exponential decrease slope is approximately constant, which indicates that the advantage of adding more sensors to the attack construction decreases exponentially at an approximately constant rate. Remarkably, this exponential decrease is observed for all system sizes and SNR regimes.

It is worth noting that for most systems, operating with larger SNR yields a lower mutual information for the same KL divergence. However, in Fig. 2 for the IEEE 30-bus test system the 10 dB and 30 dB performance curves cross, which indicates that the lower SNR regime benefits the attacker when the number of comprised sensors grows. Interestingly, the size of the network does not determine the performance the attack. For the Algorithm 1 construction, the IEEE 14-bus system is the most vulnerable to attacks, while for the Algorithm 2 construction the statement only holds for high SNR regime. This suggests that the topology of the network fundamentally changes the performance of the attack but the specific mechanisms are left for future study.

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

Fig. 3 and Fig. 4 depict the multiobjective performance of the Algorithm 1 attack construction in terms of the tradeoff between mutual information and KL divergence for different values of the proportion of compromised sensors when SNR = 30 dB and $\rho = 0.9$. Similarly, Fig. 5 and Fig. 6 depict the same setting for the Algorithm 2 attack construction. As expected, larger values of the parameter λ yield smaller values of KL divergence, i.e. the probability of detection is prioritized in the construction over the mutual information decrease for all the scenarios. Moreover, smaller values of k yield smaller reductions of the mutual information, which indicates that remaining stealthy in a sparse setting necessarily implies reducing the amount of disruption of the attack. On the other hand, larger values of k enable the attacker to more effectively tradeoff disruption for stealth. This effect is particularly marked in the correlated attack construction case, which reinforces the previous observation regarding the value of coordination between attack variables to achieve stealth.

VII. Conclusion

We have proposed novel stealth attack construction with sparsity constraints. The insight obtained from the problem of incorporating an additional sensor to the attack has been distilled to construct heuristic greedy constructions for both the independent and the correlated attack cases. We show that for both cases, the greedy step results in a convex optimization problem which can be solved efficiently and yields a low complexity attack update rule. We have numerically evaluated the attack performance in several IEEE test systems and shown that it is feasible to implement disruptive attacks that have access to small number of observations. Furthermore, we have observed that the topology and the SNR regime govern the performance of the attack and numerically characterized the dependence.

References

[1] E. J. Colbert and A. Kott, *Cyber-security of SCADA and other industrial control systems*. Springer, 2016.

[2] J. J. Grainger and W. D. Stevenson, *Power system analysis*. McGraw-Hill, 1994.

[3] A. Abur and A. G. Exposito, *Power system state estimation: Theory and implementation*. CRC press, Mar. 2004.

[4] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Trans. Info. Syst. Sec*, vol. 14, no. 1, pp. 1–33, May 2011.

[5] O. Vuković, K. C. Sou, G. Dán, and H. Sandberg, "Network-layer protection schemes against stealth attacks on state estimators in power systems," in *Proc. IEEE Int. Conf. on Smart Grid Comm.*, Brussels, Belgium, Oct. 2011, pp. 184–189.

[6] A. Tajer, S. Kar, H. V. Poor, and S. Cui, "Distributed joint cyber attack detection and state recovery in smart grids," in *Proc. IEEE Int. Conf. on Smart Grid Comm.*, Brussels, Belgium, Oct. 2011, pp. 202–207.

[7] S. Cui, Z. Han, S. Kar, T. T. Kim, H. V. Poor, and A. Tajer, "Coordinated data-injection attack and detection in the smart grid: A detailed look at enriching detection solutions," *IEEE Signal Process. Mag*, vol. 29, no. 5, pp. 106–115, Aug. 2012.

[8] M. Ozay, I. Esnaola, F. T. Y. Vural, S. R. Kulkarni, and H. V. Poor, "Sparse attack construction and state estimation in the smart grid: Centralized and distributed models," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1306–1318, Jul. 2013.

[9] I. Esnaola, S. M. Perlaza, and H. V. Poor, "Equilibria in data injection attacks," in *Proc. IEEE Global Conference on Signal and Information Processing*, Atlanta, GA, USA, Dec. 2014, pp. 779–783.

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

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

[12] I. Esnaola, S. M. Perlaza, H. V. Poor, and O. Kosut, "Maximum distortion attacks in electricity grids," *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 2007–2015, Jul. 2016.

[13] M. Ozay, I. Esnaola, F. T. Yarman Vural, S. R. Kulkarni, and H. V. Poor, "Machine learning methods for attack detection in the smart grid," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 8, pp. 1773–1786, Aug. 2016.

[14] A. Tajer, S. M. Perlaza, and H. V. Poor, *Advanced Data Analytics for Power Systems*. Cambridge University Press, 2021.

[15] K. Sun, I. Esnaola, S. M. Perlaza, and H. V. Poor, "Information-theoretic attacks in the smart grid," in *Proc. IEEE Int. Conf. on Smart Grid Comm.*, Dresden, Germany, Oct. 2017, pp. 455–460.

[16] ——, "Stealth attacks on the smart grid," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1276–1285, Aug. 2019.

[17] X. Ye, I. Esnaola, S. M. Perlaza, and F. H. Robert, "Information theoretic data injection attacks with sparsity constraints," in *Proc. IEEE Int. Conf. on Smart Grid Comm.*, Tempe, AZ, USA, Oct. 2020, pp. 1–6.

[18] C. Genes, I. Esnaola, S. M. Perlaza, L. F. Ochoa, and D. Coca, "Recovering missing data via matrix completion in electricity distribution systems," in *Proc. Int. Workshop on Signal Processing Advances in Wireless Communications*, Edinburgh, United Kingdom, Jul. 2016, pp. 1–6.

[19] ——, "Robust recovery of missing data in electricity distribution systems," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 4057–4067, Jun. 2018.

[20] I. Shomorony and A. S. Avestimehr, "Worst-case additive noise in wireless networks," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3833–3847, Jun. 2013.
[21] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Philosophical Trans. of the Royal Society of London*, vol. 231, pp. 289–337, Feb. 1933.
[22] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
[23] U. of Washington, "Power systems test case archive," 1999. [Online]. Available: https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/radiometric
[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 Trans. Power Syst*, vol. 26, no. 1, pp. 12–19, Feb. 2010.
[25] I. Esnaola, A. M. Tulino, and J. Garcia-Frias, "Linear analog coding of correlated multivariate Gaussian sources," *IEEE Trans. on Commun.*, vol. 61, no. 8, pp. 3438–3447, Aug. 2013.