Information Theoretic Data Injection Attacks with Sparsity Constraints

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

Information Theoretic Data Injection Attacks with Sparsity Constraints

Xiuzhen Ye*, Iñaki Esnaola*†, Samir M. Perlaza §†, and Robert F. Harrison*
*Dept. of Automatic Control and Systems Engineering, University of Sheffield, Sheffield S1 3JD, UK
†Dept. of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA
§INRIA, Centre de Recherche de Sophia Antipolis - Méditerranée, France

***Abstract*—Information theoretic sparse attacks that minimize simultaneously the information obtained by the operator and the probability of detection are studied in a Bayesian state estimation setting. The attack construction is formulated as an optimization problem that aims to minimize the mutual information between the state variables and the observations while guaranteeing the stealth of the attack. Stealth is described in terms of the Kullback-Leibler (KL) divergence between the distributions of the observations under attack and without attack. To overcome the difficulty posed by the combinatorial nature of a sparse attack construction, the attack case in which only one sensor is compromised is analytically solved first. The insight generated in this case is then used to propose a greedy algorithm that constructs random sparse attacks. The performance of the proposed attack is evaluated in the IEEE 30 Bus Test Case.**

I. Introduction

State estimation enables efficient, scalable, and secure operation of power systems. This is in part thanks to monitoring and control processes that are supported by Supervisory Control and Data Acquisition (SCADA) systems and more recently by advanced communication systems that acquire and transmit observations to a state estimator [1]. This cyber layer exposes the system to malicious attacks that exploit the vulnerabilities of the sensing and communication infrastructure. One of the main threats faced by modern power systems are data injection attacks (DIAs) [2] that alter the state estimate of the operator by compromising the system observations. A large body of literature is concerned with the case in which attack detection is performed by a residual test [3] under the assumption that state estimation is deterministic. In this setting, constructing DIAs that require access to a small set of observations yields optimization problems with sparsity constraints, which are often difficult to solve. In [4], it is shown that the operator can secure a small fraction of observations to make the attack construction significantly harder. This problem has been studied extensively in the literature in both centralized and decentralized scenarios [5], [6], [7], [8].

The unprecedented data acquisition capabilities in the smart grid elevate the threat of attack precisely because accurate stochastic models can be generated for the system. In view of this, attack constructions that exploit this prior knowledge can be posed within a Bayesian framework [9]. In this setting, the attack detection problem is no longer cast as a residual test. Instead, detection strategies consider the likelihood ratio test [10] or alternatively machine learning methods [11]. The operator produces a stochastic model of the system based on the observations generated by the monitoring system. Moreover, data analytics on the system depend on the reliability of the observations that are used with a variety of estimation, statistical and machine learning tools that provide the operator with different insight. In view of this, it is essential to assess attacks in fundamental terms to understand the impact over a wide range of estimation and data analysis paradigms.

Information theoretic attacks are first introduced in [12] and then generalized in [13]. In this approach, attack disruption is measured in terms of two information measures: (a) the mutual information between the state variables and the observations under attack; and (b) the probability of detection, which is governed by the Kullback-Leibler (KL) divergence. The advantage of using these information measures is that the attack disrupts a wide range of estimation, statistical and machine learning methods that are available to the operator. Given that the attack vector corrupts the observations in an additive fashion, mutual information minimization yields a Gaussian attack construction that has the maximum entropy, i.e. maximum uncertainty, among all the distributions with fixed variance [14]. From a practical point of view, the assumption is validated given the data shared by Electricity North West Limited [15]. In this case, mutual information decreases monotonically with

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.

the variance of the attack vector entries [16] and the converse holds for the probability of attack detection. The information theoretic attacks in [13] require that the attacker tampers with all the observations used by the operator [17]. Hence, incorporating sparsity constraints with information theoretic attacks is still an open problem that requires novel approaches. In this paper, we present a novel information theoretic sparse attack construction based on a greedy observation selection mechanism.

A brief description of notation follows. Consider matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, then $(\mathbf{A})_{ij}$ denotes the entry in row i and column j. We denote by $\mathbf{A}_{\mathcal{I}}$ the matrix formed with the rows of $\mathbf{A} \in \mathbb{R}^{m\times n}$ given by the indices in $\mathcal{I} \subseteq \{1,\dots,m\}$ in increasing order. We denote the complement of set $\mathcal{I}$ by $\mathcal{I}^{\mathsf{c}}$. The elementary vector $\mathbf{e}_i$ is a vector of zeros with a one in the entry i. Random variables are denoted by capital letters and their realizations by the correponding 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. $X^n = (X_1,\dots,X_n)$ with corresponding realizations denoted by $\mathbf{x}$. The set of positive semidefinite matrices of size $n\times n$ is denoted by S_+^n.

II. System model

A. Power system state estimation

In a power system the state vector $\mathbf{x} \in \mathbb{R}^n$ containing the voltages and angles at all the generation and load buses describes the operation state of the system. State vector $\mathbf{x}$ is observed by the acquisition function $F : \mathbb{R}^n \to \mathbb{R}^m$. A linearized observation model is considered for state estimation, yielding the observation model

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

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 containing observations Y^m is corrupted by additive white Gaussian noise introduced by the sensors [1], [18]. The noise vector Z^m follows a multivariate Gaussian distribution $Z^m \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_m)$, where σ^2 is the noise variance.

In a Bayesian estimation framework, the state variables are described by a vector of random variables X^n with a given distribution. As the Gaussian distribution has the maximum entropy among all distributions with the same variance, we assume X^n follows a multivariable Gaussian distribution with zero mean and covariance matrix $\mathbf{\Sigma}_{XX} \in S_+^n$. From (1), it follows that the vector of observations is zero mean and with covariance matrix

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

The resulting observations are corrupted by the malicious attack vector

$$A^m \sim P_{A^m}, \tag{3}$$

where P_{A^m} is the distribution of the random attack vector A^m. Since the Gaussian distribution minimizes the mutual information between the state variables and the compromised observations with a fixed covariance matrix [16], we adopt a Gaussian random attack framework given by

$$A^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{AA}), \tag{4}$$

where $\mathbf{\Sigma}_{AA}$ is the covariance matrix of attack vector A^m. Consequently, the compromised observations denoted by Y_A^m are given by

$$Y_A^m = \mathbf{H}X^n + Z^m + A^m, \tag{5}$$

where Y_A^m follows a multivariate Gaussian distribution given by

$$Y_A^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{Y_AY_A}) \tag{6}$$

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

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:

$$\mathcal{H}_0 : \text{There is no attack,} \tag{7}$$

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

In this setting, the optimal test is the likelihood ratio test (LRT) [19] given by

$$L(\mathbf{y}) = \frac{f_{Y_A^m}(\mathbf{y})}{f_{Y^m}(\mathbf{y})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau, \tag{9}$$

where $\mathbf{y}$ is the realization of the observations to be tested for attack; $f_{Y_A^m}(\mathbf{y})$ is the probability density function (pdf) of Y_A^m in (5), $f_{Y^m}(\mathbf{y})$ is the pdf of Y^m in (1), and $\tau \in \mathbb{R}_+$ in (9) is the decision threshold. The performance of the test is assessed in terms of the Type I error, defined as $\alpha \triangleq \mathbb{P}\left[L(\bar{Y}^m) \geq \tau\right]$ with $\bar{Y}^m \sim P_{Y^m}$, and the Type II error, denoted by $\beta \triangleq \mathbb{P}\left[L(\bar{Y}^m) < \tau\right]$ with $\bar{Y}^m \sim P_{Y_A^m}$. Note that the LRT is optimal, and therefore, changing the value of τ is equivalent to changing the tradeoff between Type I and Type II errors.

III. Sparse information theoretic attacks

A. Information theoretic setting

The attack construction in [13] incorporates a detection constraint based on the KL divergence between the distributions $P_{Y_A^m}$ in (5) and P_{Y^m} in (1) which results in the construction of *stealth attacks*. Specifically, the construction is given by the solution to the following optimization problem:

$$\min_{P_{A^m}} I(X^n; Y_A^m) + \lambda D(P_{Y_A^m} \| P_{Y^m}), \tag{10}$$

where $I(X;Y)$ is the mutual information between random variables X and Y, $D(P\|Q)$ denotes the KL divergence between distributions P and Q, and $\lambda \geq 1$ is the weighting parameter that determines the tradeoff between attack disruption and probability of detection. Note that the optimization in (10) searches for the distribution of the attack vector of random variables over the set of Gaussian multivariate distributions of m dimensions, or equivalently, it chooses the optimal covariance matrix for the distribution of the attack. It is shown in [13] that the optimal Gaussian attack is given by $\bar{P}_{A^m} = \mathcal{N}(\mathbf{0}, \bar{\mathbf{\Sigma}})$ where

$$\bar{\mathbf{\Sigma}} = \lambda^{-1} \mathbf{H} \mathbf{\Sigma}_{XX} \mathbf{H}^{\mathsf{T}}. \tag{11}$$

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

$$\text{supp}(A^m) \triangleq \left\{ i : \mathbb{P}\left[A_i = 0\right] = 0 \right\}. \tag{12}$$

B. Sparse attack formulation

Given that the operator is likely to have access control policies in place [20], an attack construction that requires access to all the observations is costly and unrealistic for the attacker in most scenarios. For that reason, in the following we study stealth attack constructions that require access to a limited number of sensors. In particular, we pose the optimization problem with sparsity constraints by considering distributions over the attack vector that put non-zero mass on at most $k \leq m$ attack vector components. Thus, we include the additional requirement that $|\text{supp}(A^m)| = k$ in the attack construction. In view of this, the attacker chooses the distribution of the attack vector over the set of multivariate Gaussian distributions given by

$$\mathcal{P}_k \triangleq \left\{ P_{A^m} : |\text{supp}(A^m)| = k \right\}. \tag{13}$$

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

$$\min_{P_{A^m} \in \mathcal{P}_k} I(X^n; Y_A^m) + \lambda D(P_{Y_A^m} \| P_{Y^m}). \tag{14}$$

Solving this problem is hard in general owing to the combinatorial nature of the attack vector support selection. For that reason, in Section IV we tackle the problem by proposing a greedy attack construction algorithm that results in k-sparse attack vectors.

C. Gaussian sparse attack construction

In the following, we particularize the attack construction in (10) by considering Gaussian distributed state variables, i.e. $X^n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{XX})$, and assuming that the attack vector follows the Gaussian distribution given in (4). In this setting, the optimization problem in (10) is equivalent [13] to the following optimization problem:

$$\begin{aligned} \min_{\mathbf{\Sigma}_{AA} \in S_+^m} & (1-\lambda) \log |\mathbf{I}_m + \mathbf{W}\mathbf{\Sigma}_{AA}| \\ & - \log |\sigma^2 \mathbf{I}_m + \mathbf{\Sigma}_{AA}| + \lambda \text{tr}(\mathbf{W}\mathbf{\Sigma}_{AA})), \end{aligned} \tag{15}$$

where $\mathbf{W} \triangleq \mathbf{\Sigma}_{YY}^{-1}$. In order to incorporate sparsity constraints in (15), the minimization domain is constrained to the set of covariance matrices that induce k-sparse supports over the attack vectors, i.e., the set given by

$$\mathcal{S}_k \triangleq \left\{ \mathbf{S} \in S_+^m : \|\text{diag}(\mathbf{S})\|_0 = k \right\}, \tag{16}$$

where $\text{diag}(\mathbf{S})$ denotes the vector formed by the diagonal entries of $\mathbf{S}$. Solving (15) within the optimization domain specified by (16) re-casts the equivalent k-sparse stealth attack construction problem in (14) as follows:

$$\begin{aligned} \min_{\mathbf{\Sigma}_{AA} \in \mathcal{S}_k} & (1-\lambda) \log |\mathbf{I}_m + \mathbf{W}\mathbf{\Sigma}_{AA}| \\ & - \log |\sigma^2 \mathbf{I}_m + \mathbf{\Sigma}_{AA}| + \lambda \text{tr}(\mathbf{W}\mathbf{\Sigma}_{AA})). \end{aligned} \tag{17}$$

D. Optimal single observation attack case

Despite having narrowed it down to Gaussian distributions, the above optimization problem is still challenging and combinatorial in nature. For that reason, we first tackle the case in which the attacker only comprises one sensor, i.e. $k = 1$. The rationale for this is that we use the insight developed for the single sensor case in the construction of the general k-sparse case. The following theorem provides the optimal solution for the case in which the attacker corrupts a single sensor.

Theorem 1. *The solution to the sparse stealth attack construction problem in (17) for the case $k = 1$ is given by*

$$\bar{\mathbf{\Sigma}}_{AA} = \bar{\sigma}^2 \mathbf{e}_\alpha \mathbf{e}_\alpha^{\mathsf{T}}, \tag{18}$$

where

$$\alpha = \arg\min_i \{(\mathbf{W})_{ii}\}, \tag{19}$$

$$\bar{\sigma}^2 = -\frac{\sigma^2}{2} + \frac{1}{2}\left(\sigma^4 - \frac{4(\underline{w}\sigma^2 - 1)}{\lambda \underline{w}^2}\right)^{\frac{1}{2}}, \tag{20}$$

with $\underline{w} \triangleq (\mathbf{W})_{\alpha\alpha}$.

Proof: We start by noting that for $k = 1$ the set of attack covariance matrices is given by

$$\mathcal{S}_1 \triangleq \bigcup_{i=1,\ldots,m} \left\{\mathbf{S} \in S_+^m : \mathbf{S} = \sigma_i^2 \mathbf{e}_i \mathbf{e}_i^\mathsf{T} \text{with } \sigma_i \in \mathbb{R}_+\right\}. \tag{21}$$

The covariance matrices in set $\mathcal{S}_1$ comprise matrices with a single nonzero element in the diagonal. The non-zero entry i denotes the index of the sensor that is attacked. Let $i \in \{1, 2, ..., m\}$ be the index of the non-zero entry of the covariance matrix $\bar{\mathbf{\Sigma}}_{AA}$. The non-zero entry denoted by σ_i^2 is the variance of the random variable used to attack observation i.

Let $\lambda > 1$ and restrict the optimization domain in (17) to $\mathcal{S}_1$. Thus, the resulting optimization problem is equivalent to:

$$\min_{\bar{\sigma}>0} \min_i \log \frac{(1 + (\mathbf{W})_{ii}\bar{\sigma}^2)^{1-\lambda}}{(\sigma^2 + \bar{\sigma}^2)} + \lambda(\mathbf{W})_{ii}\bar{\sigma}^2. \tag{22}$$

We proceed by solving the inner part of the optimization problem above. Consider the cost given by

$$f((\mathbf{W})_{ii}) \triangleq \log \frac{(1 + (\mathbf{W})_{ii}\bar{\sigma}^2)^{1-\lambda}}{(\sigma^2 + \bar{\sigma}^2)} + \lambda(\mathbf{W})_{ii}\bar{\sigma}^2, \tag{23}$$

which can be rewritten as

$$f(t) = (1 - \lambda)\log t - \log(\sigma^2 + \bar{\sigma}^2) + \lambda t - \lambda, \tag{24}$$

where $t = 1 + (\mathbf{W})_{ii}\bar{\sigma}^2$. It follows that (24) is convex with respect to t because λt is a linear term and $(1 - \lambda)\log t$ is convex in t for $\lambda > 1$. Therefore, $f((\mathbf{W})_{ii})$ is convex with respect to $(\mathbf{W})_{ii}$ and the minimum is attained for $(\mathbf{W})_{ii} = -\frac{1}{\lambda\bar{\sigma}^2}$. Since $(\mathbf{W})_{ii} > 0$ the inner minimization in (22) is equivalent to selecting the index i that minimizes $(\mathbf{W})_{ii}$. The definition of α in (19) and $\underline{w}$ in (20) follow from this observation.

We now proceed to solve the outer optimization. In this case, the cost is given by

$$g(r) = (1 - \lambda)\log(1 + \underline{w}r) - \log(\sigma^2 + r) + \lambda\underline{w}r, \tag{25}$$

where $r \triangleq \bar{\sigma}^2$. Noticing that the above function has a single minimizer given by

$$r = -\frac{\sigma^2}{2} + \frac{1}{2}\left(\sigma^4 - \frac{4(\underline{w}\sigma^2 - 1)}{\lambda \underline{w}^2}\right)^{\frac{1}{2}} \tag{26}$$

completes the proof. ■

Algorithm 1 k-sparse stealth attack construction

Input: the observation matrix $\mathbf{H}$; the covariance matrix of the state variables $\mathbf{\Sigma}_{XX}$; the variance of the noise σ^2; and the weighting parameter λ; number of nonzero attack vector components k.

Output: the covariance matrix of the attack vector $\bar{\mathbf{\Sigma}}_{AA}$; and the set of indices of attacked sensors $\mathcal{A}$.

Set $\mathcal{A}_0 = \{\emptyset\}$
for $j = 1$ to k **do**
  Set $\mathbf{H}_j = \mathbf{H}_{\mathcal{A}_{j-1}^c}$
  Compute $\mathbf{W}_j = \left(\mathbf{H}_j \mathbf{\Sigma}_{XX} \mathbf{H}_j^\mathsf{T} + \sigma^2 \mathbf{I}_{|\mathcal{A}_{j-1}^c|}\right)^{-1}$
  Set $\alpha_j = \arg\min_i \{(\mathbf{W}_j)_{ii}\}$,
  Set $\underline{w}_j \triangleq (\mathbf{W}_j)_{\alpha_j\alpha_j}$
  Set $\bar{\sigma}_j^2 = -\frac{\sigma^2}{2} + \frac{1}{2}\left(\sigma^4 - \frac{4(\underline{w}_j\sigma^2 - 1)}{\lambda \underline{w}_j^2}\right)^{\frac{1}{2}}$
  Set $\mathcal{A}_j = \mathcal{A}_{j-1} \cup \{\alpha_j\}$
end for
Set $\mathcal{A} = \mathcal{A}_k$
Set $\bar{\mathbf{\Sigma}}_{AA} = \sum_{i\in\mathcal{A}} \bar{\sigma}_i^2 \mathbf{e}_i \mathbf{e}_i^\mathsf{T}$

IV. Greedy construction of sparse attacks

The extension to the k-sparse case of the solution proposed in Section III-D does not get around the combinatorial optimization in (17). For that reason, in the following we propose a greedy construction that leverages the insight distilled in the $k = 1$ case to select the set of k attacked sensors. The construction is based on a classical greedy procedure that sequentially selects an observation to attack by maximizing the performance in terms of the decision at each step. Let us denote by $\mathcal{A}$ the set of observation indices that are attacked, i.e. $\mathcal{A} \triangleq \text{supp}(A^m)$. The greedy algorithm operates by sequentially updating the entries in $\mathcal{A}$ by adding a new index in each step until k indices are selected. For that reason, the resulting entries of the attack vector are independent, and therefore, the covariance matrix of the attack vector obtained via the proposed greedy approach belongs to the set

$$\tilde{\mathcal{S}}_k \triangleq \bigcup_{\mathcal{K}} \left\{\mathbf{S} \in S_+^m : \mathbf{S} = \sum_{i\in\mathcal{K}} \sigma_i^2 \mathbf{e}_i \mathbf{e}_i^\mathsf{T} \text{with } \sigma_i \in \mathbb{R}_+\right\}, \tag{27}$$

where the union is over all subsets $\mathcal{K} \subseteq \{1, 2, \ldots, m\}$ with $|\mathcal{K}| = k \leq m$. The proposed greedy construction is described in Algorithm 1.

V. Numerical results

In this section, we present the simulation results on a direct current (DC) state estimation setting for the IEEE 30 Bus Test Case [21]. The voltage magnitudes are set

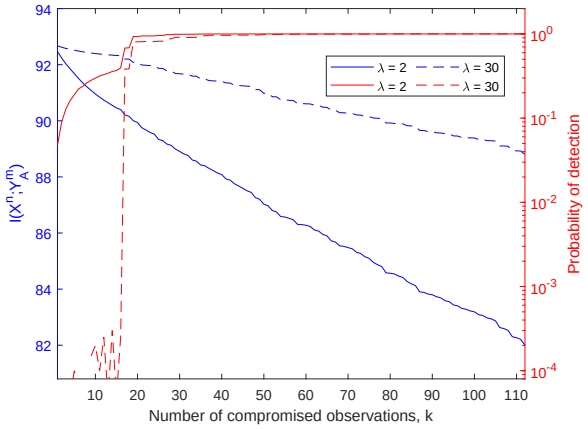

Figure 1: Performance of the sparse attack in terms of mutual information, probability of detection for different values of λ when SNR $= 30$dB, $\rho = 0.1, \tau = 2$ on the IEEE 30 Bus Test Case.

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 reactances of the branches and the topology of the system. MATPOWER [22] is adopted to generate $\mathbf{H}$. To capture the statistical dependence between the state variables we adopt a Toeplitz model for the covariance matrix $\mathbf{\Sigma}_{XX}$ that arises in a wide range of practical settings, such as autoregressive stationary processes. Specifically, we model the correlation between state variables X_i and X_j with the exponential decay parameter ρ that results in $(\mathbf{\Sigma}_{XX})_{ij} = \rho^{|i-j|}$ with $i, j = 1, 2, \ldots, n$.

In this setting, the performance of the proposed sparse stealth attack is a function of the correlation parameter ρ, noise variance σ^2, and the topology of the system as described by $\mathbf{H}$. We define the signal to noise ratio (SNR) as

$$\text{SNR} \triangleq 10 \log_{10} \left(\frac{\text{tr}(\mathbf{H}\mathbf{\Sigma}_{\text{XX}}\mathbf{H}^{\text{T}})}{m\sigma^2} \right). \tag{28}$$

The results in this section are obtained by averaging 2×10^4 realizations of the observations as described in (5). Fig. 1 depicts the mutual information and the probability of detection that the attack constructed via Algorithm 1 induces for different values of the number of compromised observations and the weighting parameter λ. As expected, the mutual information decreases monotonically, approximately linearly with the number of compromised observations, while the probability of detection increases monotonically. Interestingly, the probability of detection exhibits an abrupt increase that suggests a *threshold effect* when a critical number of compromised observations is reached. The weighting parameter λ governs the minimum achievable probability of detection, e.g. a probability of detection of 10^{-2} is not attainable when $\lambda = 2$. Indeed, increasing the value of λ to 30 yields a smaller probability of detection for small values of k but the threshold effect takes place for the same number of compromised observations, for both values of λ. This suggests that the topology of the system governs the position of the threshold.

The variance of the random variables used to attack each sensor, the probability of detection, and the probability of false alarm as a function of the number of compromised observations are illustrated in Fig. 2 and Fig. 3 for $\lambda = 2$ and $\lambda = 30$, respectively. As shown in Theorem 1, λ is a scaling factor on the variances of the attack vector, and therefore, the values of the variance for the case $\lambda = 2$ are simply scaled in the case $\lambda = 30$. There are two distinguishable attack regimes depending on the variance of the attack vector entries. Algorithm 1 does not yield a monotonically decreasing profile of variances. Instead the variance of the entries selected by the algorithm switches between small and large values as the number of compromised observations increases. This suggests, that certain entries are significantly more sensitive to additive attack than others and the existence of more vulnerable sensors that are determined by the topology of the system, as shown in (20). For both cases, the probability of false alarm exhibits non-monotonic behavior with the number of compromised observations, and interestingly, the change in monotonicity coincides with the threshold.

VI. Conclusion

We have proposed an information theoretic sparse attack construction within a smart grid Bayesian state estimation framework. The proposed attack construction minimizes the mutual information between the state variables in the smart grid and the observations obtained by the operator while minimizing the probability of detection. To that end, we have proposed a cost function that combines the mutual information and the KL divergence that is amenable to sparse attack constructions. We have theoretically characterized the single observation attack case by proving that the resulting cost function is convex and obtaining the optimal attack construction for this case. We distill the insight obtained from the single observation case to propose a sparse attack construction via a greedy algorithm that overcomes the combinatorial challenge posed by the observation selection problem.

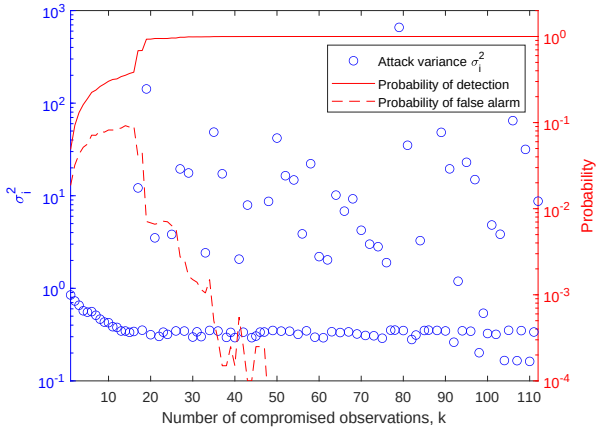

Figure 2: Variance of the attack vector entries, probability of detection, and probability of false alarm of the sparse attack when $\lambda = 2, \text{SNR} = 30$ dB, $\rho = 0.1, \tau = 2$ on the IEEE 30 Bus Test Case.

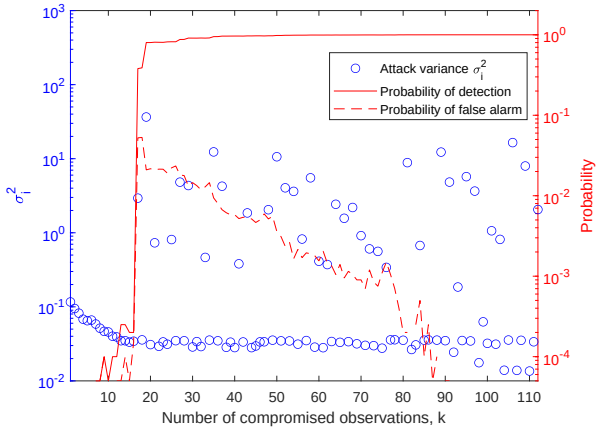

Figure 3: Variance of the attack vector entries, probability of detection, and probability of false alarm of the sparse attack when $\lambda = 30, \text{SNR} = 30$ dB, $\rho = 0.1, \tau = 2$ on the IEEE 30 Bus Test Case.

We have numerically assessed the performance of the proposed attack in the IEEE 30 Bus Test Case and observed that the probability of detection exhibits a threshold effect when a critical number of observations are compromised.

REFERENCES

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

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

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

[14] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 1999.

[15] C. Genes, I. Esnaola, S. M. Perlaza, L. F. Ochoa, and D. Coca, "Robust recovery of missing data in electricity distribution systems," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 4057–4067, Jun. 2018.

[16] 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.

[17] K. Sun, I. Esnaola, A. M. Tulino, and H. V. Poor, "Learning requirements for stealth attacks," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Brighton, UK, Apr. 2019, pp. 8102–8106.

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

[19] 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.

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

[21] U. of Washington, "Power systems test case archive," 1999. [Online]. Available: https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/radiometric

[22] 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.