
Article:
Sun, K., Esnaola, I. orcid.org/0000-0001-5597-1718, Perlaza, S.M. et al. (1 more author) (2020) Stealth attacks on the smart grid. IEEE Transactions on Smart Grid, 11 (2). pp. 1276-1285. ISSN 1949-3053

https://doi.org/10.1109/TSG.2019.2935353

Stealth Attacks on the Smart Grid

Ke Sun, Iñaki Esnaola, Samir M. Perlaza, and H. Vincent Poor

***Abstract*—Random attacks that jointly minimize the amount of information acquired by the operator about the state of the grid and the probability of attack detection are presented. The attacks minimize the information acquired by the operator by minimizing the mutual information between the observations and the state variables describing the grid. Simultaneously, the attacker aims to minimize the probability of attack detection by minimizing the Kullback-Leibler (KL) divergence between the distribution when the attack is present and the distribution under normal operation. The resulting cost function is the weighted sum of the mutual information and the KL divergence mentioned above. The trade-off between the probability of attack detection and the reduction of mutual information is governed by the weighting parameter on the KL divergence term in the cost function. The probability of attack detection is evaluated as a function of the weighting parameter. A sufficient condition on the weighting parameter is given for achieving an arbitrarily small probability of attack detection. The attack performance is numerically assessed on the IEEE 14-Bus, 30-Bus, and 118-Bus test systems.**

***Index Terms*—Stealth, data injection attacks, information-theoretic security, mutual information, probability of detection**

I. Introduction

THE smart grid relies on the effective integration of the power grid and advanced communication and sensing infrastructure. Consistency between the physical layer of the power grid and the energy management system (EMS) in the cyber layer facilitates an economic and reliable operation of the power system. The 2003 North American outage caused by an alarm system failure [1] and the 2015 Ukraine power failure caused by the BlackEnergy virus incident [2] emphasize the need for cybersecurity mechanisms for the power system. However, the cybersecurity threats to which the smart grid is exposed are not well understood yet, and therefore, practical security solutions need to come forth as a multidisciplinary effort combining technologies such as cryptography, machine learning, and information-theoretic security [3].

Data injection attacks (DIAs) have emerged as a major source of concern and exemplify the type of cybersecurity threats that specifically target power systems [4]. DIAs manipulate the state estimation process in the EMS by altering the measurements of the state variables without triggering the bad data detection mechanism put in place by the operator. In [4], it is shown that attacks that lie in the column space of the Jacobian measurement matrix are undetectable by testing the residual. To decrease the number of sensors that need to be compromised by the attacker while remaining undetectable, the ℓ_0 norm of the attack vector is used as minimization objective yielding sparse attack in [5], [6], [7] and [8]. The case in which sparse attacks are constructed in a distributed setting with multiple attackers is discussed in [9] and [10]. Estimation of the operating point is studied in [11] using the power flow or power injection information in an inference problem formulation. Attack detection methods that incorporate the statistical structure of the state variables are presented for the centralized case in [12] and for the decentralized case in [13]. Similarly, other instances of prior knowledge use include using load forecasts in [14], [15], and overload mitigation based on corrective dispatch in [16].

The complex nature of the power system leads naturally to a stochastic modelling of the state variables describing the grid. For instance, the state variables of low voltage distribution systems are well described as following a multivariate Gaussian distribution [17]. DIAs within a Bayesian framework with minimum mean square error estimation are studied in [18] for the centralized case and in [19] for the distributed case. However, the fundamental limits governing the performance of attacks in the smart grid are not well understood yet.

Information-theoretic tools are well suited to analyze power system by leveraging the stochastic description of the state variables. A sensor placement strategy that accounts for the amount of information acquired by the sensing infrastructure is studied in [20]. Information-theoretic privacy guarantees for smart meter users are proposed in [21], [22], [23] for memoryless stochastic processes and in [24] for general random processes. In [25], stealth Gaussian DIA constructions are studied in terms of information measures that quantify the information loss and the probability of attack detection induced by the attack. Therein, the proposed cost function gives the same weight to the information loss and the probability of detection which results in the effective secrecy framework proposed by [26] in the context of *stealth* communications. Stealth DIA constructions are also studied in [5], [27] for the case in which the detection is based on the residual and in a Bayesian hypothesis testing framework in [28]. The approaches in [5] and [27] consider the minimum cost of compromising the meters and the communication substation, respectively. On the other hand, [28] focuses on the delay between the time of attacker launching the attack and the time of operator detecting the attack.

In this paper, the stealth attacks in [25] are generalized by introducing a weight parameter to the objective describing the probability of detection, which allows the attacker to construct attacks with arbitrarily low probability of detection.

K. Sun and I. Esnaola 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: ke.sun@sheffield.ac.uk, esnaola@sheffield.ac.uk).

S. M. Perlaza is with the Institut National de Recherche en Informatique et Automatique (INRIA), Lyon, France, and also with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (email: samir.perlaza@inria.fr).

H. V. Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

Operating under the assumption that the state variables are described by a multivariate Gaussian distribution [18], [19], we characterize the optimal Gaussian generalized stealth attacks. Since the performance of the attacks depends on the weighting parameter governing the probability of detection, we provide a sufficient condition on the weighting parameter that achieves a desired probability of attack detection. To this end, we characterize the probability of attack detection via an upper bound which leverages a concentration inequality in [29].

The rest of the paper is organized as follows: In Section II, a Bayesian framework with linearized dynamics for DIA is introduced. The generalized stealth attack construction and performance analysis are presented in Section III. Section IV provides the probability of detection of the generalized stealth attack, and the concentration inequality is used to derive the upper bound for probability of detection. Section V evaluates the proposed attack performance on IEEE test systems. The paper ends with conclusions in Section VI.

II. System Model

A. Bayesian Framework with Linearized Dynamics

The measurement model for state estimation with linearized dynamics is given by

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

where $Y^m \in \mathbb{R}^m$ is a vector of random variables describing the measurements; $X^n \in \mathbb{R}^n$ is a vector of random variables describing the state variables; $\mathbf{H} \in \mathbb{R}^{m\times n}$ is the linearized Jacobian measurement matrix which is determined by the power network topology and the admittances of the branches; and $Z^m \in \mathbb{R}^m$ is the additive white Gaussian noise (AWGN) with distribution $\mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_m)$ that is introduced by the sensors as a result of the thermal noise, c.f. [30] and [31]. In the remaining of the paper, we assume that the vector of the state variables follows a multivariate Gaussian distribution given by

$$X^n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{XX}), \tag{2}$$

where $\mathbf{\Sigma}_{XX} \in \mathcal{S}_+^n$ is the covariance matrix of the distribution of the state variables and $\mathcal{S}_+^n$ denotes the set of positive semidefinite matrices of size $n\times n$. As a result of the linearized dynamic in (1), the vector of measurements also follows a multivariate Gaussian distribution denoted by

$$Y^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{YY}), \tag{3}$$

where $\mathbf{\Sigma}_{YY} = \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}} + \sigma^2\mathbf{I}_m$ is the covariance matrix of the distribution of the vector of measurements.

Data injection attacks corrupt the measurements available to the operator by adding an attack vector to the measurements. The resulting vector of compromised measurements is given by

$$Y_A^m = \mathbf{H}X^n + Z^m + A^m, \tag{4}$$

where $A^m \in \mathbb{R}^m$ is the attack vector and $Y_A^m \in \mathbb{R}^m$ is the vector containing the compromised measurements [4]. Given the stochastic nature of the state variables, it is reasonable for the attacker to pursue a stochastic attack construction strategy. In the following, an attack vector which is independent of the state variables is constructed under the assumption that the attack vector follows a multivariate Gaussian distribution denoted by

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

where $\mathbf{\Sigma}_{AA} \in \mathcal{S}_+^m$ is the covariance matrix of the attack distribution. The rationale for choosing a Gaussian distribution for the attack vector follows from the fact that for the measurement model in (4) the additive attack distribution that minimizes the mutual information between the vector of state variables and the compromised measurements is Gaussian [32]. Because of the Gaussianity of the attack distribution, the vector of compromised measurements is distributed as

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

where $\mathbf{\Sigma}_{Y_AY_A} = \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}} + \sigma^2\mathbf{I}_m + \mathbf{\Sigma}_{AA}$ is the covariance matrix of the distribution of the compromised measurements.

It is worth noting that the independence of the attack vector with respect to the state variables implies that constructing the attack vector does not require access to the realizations of the state variables. In fact, knowledge of the second order moments of the state variables and the variance of the AWGN introduced by the measurement process suffices to construct the attack. This assumption significantly reduces the difficulty of the attack construction.

The operator of the power system makes use of the acquired measurements to detect the attack. The detection problem is cast as a hypothesis testing problem with hypotheses

$$\mathcal{H}_0 : Y^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{YY}), \quad \text{versus} \tag{7}$$

$$\mathcal{H}_1 : Y^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{Y_AY_A}). \tag{8}$$

The null hypothesis $\mathcal{H}_0$ describes the case in which the power system is not compromised, while the alternative hypothesis $\mathcal{H}_1$ describes the case in which the power system is under attack.

Two types of error are considered in hypothesis testing problems, Type I error is the probability of accepting $\mathcal{H}_1$ when $\mathcal{H}_0$ is the ground truth, i.e. a false alarm or false positive; and Type II error is the probability of accepting $\mathcal{H}_0$ when $\mathcal{H}_1$ is the ground truth, i.e. a true negative. The Neyman-Pearson lemma [33, Proposition II.D.1] states that for a fixed probability of Type I error, the likelihood ratio test (LRT) achieves the minimum Type II error when compared with any other test with an equal or smaller Type I error. Consequently, the LRT is chosen to decide between $\mathcal{H}_0$ and $\mathcal{H}_1$ based on the available measurements. The LRT between $\mathcal{H}_0$ and $\mathcal{H}_1$ takes following form:

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

where $\mathbf{y} \in \mathbb{R}^m$ is a realization of the vector of random variables modelling the measurements; $f_{Y_A^m}$ and f_{Y^m} denote the probability density functions (p.d.f.'s) of Y_A^m and Y^m, respectively; and τ is the decision threshold set by the operator to meet a given false alarm constraint.

B. Information-Theoretic Setting

The mutual information between two random variables is a measure of the amount of information that each random variable contains about the other random variable. Consequently, the amount of information that the vector of measurements contains about the vector of state variables is determined by the mutual information between the vector of state variables and the vector of measurements. Information measures have previously been used to quantify the amount of information acquired by different monitoring systems in a smart grid context. For instance, in [20] mutual information is used to quantify the amount of information obtained by phasor measurement units from the grid. Similarly, mutual information is used to quantify the amount of information leaked by smart meters in [21] and [22].

The Kullback-Leibler (KL) divergence between two probability distributions is a measure of the statistical difference between the distributions. As such, it is a practical measure to quantify the deviation of the measurement statistics with respect to the statistics under normal operating conditions. For instance, in [13] it is used to test abnormal behaviors on the grid. For the hypothesis testing problem in (9), a small value of the KL divergence between $P_{Y_A^m}$ and P_{Y^m} implies that on average the attack is unlikely to be detected by the LRT set by the attacker for a fixed value of τ.

The purpose of the attacker is to disrupt the normal state estimation procedure by minimizing the information that the operator acquires about the state variables, while guaranteeing that the probability of attack detection is small enough, and therefore, remain concealed in the system.

An information-theoretic framework for the attack construction is adopted in this paper. To minimize the information that the operator acquires about the state variables from the measurements, the attacker minimizes the mutual information between the vector of state variables and the vector of compromised measurements. Specifically, the attacker aims to minimize $I(X^n; Y_A^m)$. The rationale for choosing mutual information to measure the evidence acquired by the measurements stems from the fundamental character of information-theoretic measures. In particular, mutual information describes the amount of information two random variables share, and therefore, it establishes in quantitative terms how much evidence the measurements contain. For that reason, it is natural for the attacker to attempt to minimize the mutual information with the aim of disrupting the monitoring process of the network operator.

On the other hand, the probability of attack detection is determined by the detection threshold τ set by the operator and the distribution induced by the attack on the vector of compromised measurements. An analytical expression of the probability of attack detection can be described in closed-form as a function of the distributions describing the measurements under both hypotheses. However, the expression is involved in general and it is not straightforward to incorporate it into an analytical formulation of the attack construction. For that reason, we instead consider the asymptotic performance of the LRT to evaluate the detection performance of the operator. The Chernoff-Stein lemma [34, Theorem 11.7.3] characterizes the asymptotic exponent of the probability of detection when the number of observations of measurement vectors grows to infinity. In our setting, the Chernoff-Stein lemma states that for any LRT and $\epsilon \in (0, 1/2)$, it holds that

$$\lim_{k\to\infty} \frac{1}{k} \log \beta_k^\epsilon = -D(P_{Y_A^m} \| P_{Y^m}), \tag{10}$$

where $D(\cdot\|\cdot)$ is the KL divergence, β_k^ϵ is the minimum Type II error such that the Type I error α satisfies $\alpha < \epsilon$, and k is the number of m-dimensional measurement vectors that are available for the LRT. Therefore, for the attacker, minimizing the asymptotic detection probability is equivalent to minimizing $D(P_{Y_A^m} \| P_{Y^m})$, where $P_{Y_A^m}$ and P_{Y^m} denote the probability distributions of Y_A^m and Y^m, respectively.

III. Information-Theoretic Attack

A. Generalized Stealth Attacks

When these two information-theoretic objectives are considered by the attacker, [25] proposes an stealthy attack construction that combines the two objectives in one cost function, i.e.,

$$I(X^n; Y_A^m) + D(P_{Y_A^m} \| P_{Y^m}) = D(P_{X^n Y_A^m} \| P_{X^n} P_{Y^m}), \tag{11}$$

where $P_{X^n Y_A^m}$ is the joint distribution of X^n and Y_A^m. The resulting optimization problem to construct the attack is given by

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

Therein, it is shown that (12) is a convex optimization problem and the covariance matrix of the optimal Gaussian attack is $\mathbf{\Sigma}_{AA} = \mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}$. However, numerical simulations on IEEE test system show that the attack construction proposed above yields large values of probability of detection in practical settings.

To address the issue of high probability of detection, in the following we propose an attack construction strategy that tunes the probability of detection with a parameter that weights the detection term in the cost function. The resulting optimization problem is given by

$$\min_{A^m} \; I(X^n; Y_A^m) + \lambda D(P_{Y_A^m} \| P_{Y^m}), \tag{13}$$

where $\lambda \geq 1$ governs the weight given to each objective in the cost function. It is interesting to note that for the case in which $\lambda = 1$ the proposed cost function boils down to the effective secrecy proposed in [26] and the attack construction in (13) coincides with that in [25]. For $\lambda > 1$, the attacker adopts a conservative approach and prioritizes remaining undetected over minimizing the amount of information acquired by the operator. By increasing the value of λ the attacker decreases the probability of detection at the expense of increasing the amount of information acquired by the operator via the measurements. The case for $\lambda < 1$ requires a different treatment and is left as future work.

B. Optimal Attack Construction

The attack construction in (13) is formulated in a general setting. The following propositions particularize the KL divergence and mutual information to our multivariate Gaussian setting.

Proposition 1. [34] *The KL divergence between m-dimensional multivariate Gaussian distributions $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{Y_A Y_A})$ and $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{YY})$ is given by*

$$D(P_{Y_A^m} \| P_{Y^m}) = \frac{1}{2}\left(\log \frac{|\mathbf{\Sigma}_{YY}|}{|\mathbf{\Sigma}_{Y_A Y_A}|} - m + \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{Y_A Y_A}\right)\right). \quad (14)$$

Proposition 2. [34] *The mutual information between the vectors of random variables $X^n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{XX})$ and $Y_A^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{Y_A Y_A})$ is given by*

$$I(X^n; Y_A^m) = \frac{1}{2}\log \frac{|\mathbf{\Sigma}_{XX}||\mathbf{\Sigma}_{Y_A Y_A}|}{|\mathbf{\Sigma}|}, \quad (15)$$

where $\mathbf{\Sigma}$ is the covariance matrix of the joint distribution of (X^n, Y_A^m).

Substituting (14) and (15) in (13) we can now pose the Gaussian attack construction as the following optimization problem:

$$\min_{\mathbf{\Sigma}_{AA} \in \mathcal{S}_+^m} \; -(\lambda - 1)\log|\mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_{AA}| - \log|\mathbf{\Sigma}_{AA} + \sigma^2 \mathbf{I}_m| + \lambda \mathrm{tr}(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{AA}). \quad (16)$$

We now proceed to solve the optimization problem above. First, note that the optimization domain $\mathcal{S}_+^m$ is a convex set. The following proposition characterizes the convexity of the cost function.

Proposition 3. *Let $\lambda \geq 1$. Then the cost function in the optimization problem in (16) is convex.*

Proof. Note that the term $-\log|\mathbf{\Sigma}_{AA} + \sigma^2 \mathbf{I}_m|$ is a convex function on $\mathbf{\Sigma}_{AA} \in \mathcal{S}_+^m$ [35]. Additionally, $-(\lambda - 1)\log|\mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_{AA}|$ is a convex function on $\mathbf{\Sigma}_{AA} \in \mathcal{S}_+^m$ when $\lambda \geq 1$. Since the trace operator is a linear operator and the sum of convex functions is convex, it follows that the cost function in (16) is convex on $\mathbf{\Sigma}_{AA} \in \mathcal{S}_+^m$. □

Theorem 1. *Let $\lambda \geq 1$. Then the solution to the optimization problem in (16) is*

$$\mathbf{\Sigma}_{AA}^{\star} = \frac{1}{\lambda}\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}. \quad (17)$$

Proof. Denote the cost function in (16) by $f(\mathbf{\Sigma}_{AA})$. Taking the derivative of the cost function with respect to $\mathbf{\Sigma}_{AA}$ yields

$$\frac{\partial f(\mathbf{\Sigma}_{AA})}{\partial \mathbf{\Sigma}_{AA}} = -2(\lambda - 1)(\mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_{AA})^{-1} - 2(\mathbf{\Sigma}_{AA} + \sigma^2 \mathbf{I}_m)^{-1} + 2\lambda \mathbf{\Sigma}_{YY}^{-1} + (\lambda - 1)\mathrm{diag}\left((\mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_{AA})^{-1}\right) + \mathrm{diag}\left((\mathbf{\Sigma}_{AA} + \sigma^2 \mathbf{I}_m)^{-1}\right) - \lambda \mathrm{diag}(\mathbf{\Sigma}_{YY}^{-1}). \quad (18)$$

Note that the only critical point is $\mathbf{\Sigma}_{AA}^{\star} = \frac{1}{\lambda}\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}$. Theorem 1 follows immediately from combining this result with Proposition 3. □

Corollary 1. *The mutual information between the vector of state variables and the vector of compromised measurements induced by the optimal attack construction is given by*

$$I(X^n; Y_A^n) = \frac{1}{2}\log\left|\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}\left(\sigma^2 \mathbf{I}_m + \frac{1}{\lambda}\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^{\mathrm{T}}\right)^{-1} + \mathbf{I}_m\right|. \quad (19)$$

Theorem 1 shows that the generalized stealth attacks share the same structure of the stealth attacks in [25] up to a scaling factor determined by λ. The solution in Theorem 1 holds for the case in which $\lambda \geq 1$, and therefore, lacks full generality. However, the case in which $\lambda < 1$ yields unreasonably high probability of detection [25] which indicates that the proposed attack construction is indeed of practical interest in a wide range of state estimation settings. Furthermore the optimization problem in (16) results in a non-convex problem when $\lambda < 1$ and the solution obtained above no longer holds. For this reason the case with $\lambda < 1$ is left as a future research question.

Changing the value of λ yields different solutions on the Pareto front of the optimization problem in (16) as we show in the numerical results in Section V-B. For any $\lambda \geq 1$, Theorem 1 guarantees that the generalized stealth attack is the only Parento efficient solution, i.e. the attack construction that minimizes the mutual information subject to the probability of detection constraint being satisfied. By increasing the value of λ the attacker places more importance on the probability of detection than on the mutual information which results in a more conservative attack that disrupts less but is more difficult to detect.

Theorem 1 also shows that the resulting attack construction is remarkably simple to implement provided that the information about the system is available to the attacker. Indeed, the attacker only requires access to the linearized Jacobian measurement matrix $\mathbf{H}$ and the second order statistics of the state variables, but the variance of the noise introduced by the sensors is not necessary. To obtain the Jacobian, a malicious attacker needs to know the topology of the grid, the admittances of the branches, and the operation point of the system. The second order statistics of the state variables on the other hand, can be estimated using historical data. In [25] it is shown that the attack construction with a sample covariance matrix of the state variables obtained with historical data is asymptotically optimal when the size of the training data grows to infinity.

Corollary 1 shows that the mutual information increases monotonically with λ and that it asymptotically converges to $I(X^n; Y^m)$, i.e. the case in which there is no attack. While the evaluation of the mutual information as shown in Corollary 1 is straightforward, the computation of the associated probability of detection yields involved expressions that do not provide much insight. For that reason, the probability of detection of optimal attacks is treated in the following section.

IV. Probability of Detection of Generalized Stealth Attacks

The asymptotic probability of detection of the generalized stealth attacks characterized in Section III-B is governed by

the KL divergence as described in (10). However in the non-asymptotic case, determining the probability of detection is difficult, and therefore, choosing a value of λ that provides the desired probability of detection is a challenging task. In this section we first provide a closed-form expression of the probability of detection by direct evaluation and show that the expression does not provide any practical insight over the choice of λ that achieves the desired detection performance. That being the case, we then provide an upper bound on the probability of detection, which, in turn, provides a lower bound on the value of λ that achieves the desired probability of detection.

A. Direct Evaluation of the Probability of Detection

Detection based on the LRT with threshold τ yields a probability of detection given by

$$\mathsf{P}_\mathsf{D} \triangleq \mathbb{E}\left[\mathbb{1}_{\left\{L(Y_A^m)\geq\tau\right\}}\right], \tag{20}$$

where $\mathbb{1}_{\{\cdot\}}$ is the indicator function. The following proposition particularizes the above expression to the optimal attack construction described in Section III-B.

Lemma 1. *The probability of detection of the LRT in (9) for the attack construction in (17) is given by*

$$\mathsf{P}_\mathsf{D}(\lambda) \triangleq \mathbb{P}\left[(U^p)^\mathrm{T}\boldsymbol{\Delta}U^p \geq \lambda\left(2\log\tau+\log\left|\mathbf{I}_p+\lambda^{-1}\boldsymbol{\Delta}\right|\right)\right], \tag{21}$$

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

Proof. The probability of detection of the stealth attack is,

$$\mathsf{P}_\mathsf{D}(\lambda) = \int_{\mathcal{S}} \mathrm{d}P_{Y_A^m} \tag{22}$$

$$= \frac{1}{(2\pi)^{\frac{m}{2}}\left|\boldsymbol{\Sigma}_{Y_AY_A}\right|^{\frac{1}{2}}}\int_{\mathcal{S}} \exp\left\{-\frac{1}{2}\mathbf{y}^\mathrm{T}\boldsymbol{\Sigma}_{Y_AY_A}^{-1}\mathbf{y}\right\}\mathrm{d}\mathbf{y}, \tag{23}$$

where

$$\mathcal{S} = \{\mathbf{y}\in\mathbb{R}^m : L(\mathbf{y}) \geq \tau\}. \tag{24}$$

Algebraic manipulation yields the following equivalent description of the integration domain:

$$\mathcal{S} = \left\{\mathbf{y}\in\mathbb{R}^m\colon \mathbf{y}^\mathrm{T}\boldsymbol{\Delta}_0\mathbf{y} \geq 2\log\tau+\log\left|\mathbf{I}_m+\boldsymbol{\Sigma}_{AA}\boldsymbol{\Sigma}_{YY}^{-1}\right|\right\}, \tag{25}$$

with $\boldsymbol{\Delta}_0 \triangleq \boldsymbol{\Sigma}_{YY}^{-1} - \boldsymbol{\Sigma}_{Y_AY_A}^{-1}$. Let $\boldsymbol{\Sigma}_{YY} = \mathbf{U}_{YY}\boldsymbol{\Lambda}_{YY}\mathbf{U}_{YY}^\mathrm{T}$ where $\boldsymbol{\Lambda}_{YY} \in \mathbb{R}^{m\times m}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{\Sigma}_{YY}$ in descending order and $\mathbf{U}_{YY} \in \mathbb{R}^{m\times m}$ is a unitary matrix whose columns are the eigenvectors of $\boldsymbol{\Sigma}_{YY}$ ordered matching the order of the eigenvalues. Applying the change of variable $\mathbf{y}_1 \triangleq \mathbf{U}_{YY}\mathbf{y}$ in (23) results in

$$\mathsf{P}_\mathsf{D}(\lambda) = \frac{1}{(2\pi)^{\frac{m}{2}}\left|\boldsymbol{\Sigma}_{Y_AY_A}\right|^{\frac{1}{2}}}\int_{\mathcal{S}_1} \exp\left\{-\frac{1}{2}\mathbf{y}_1^\mathrm{T}\boldsymbol{\Lambda}_{Y_AY_A}^{-1}\mathbf{y}_1\right\}\mathrm{d}\mathbf{y}_1, \tag{26}$$

where $\boldsymbol{\Lambda}_{Y_AY_A} \in \mathbb{R}^{m\times m}$ denotes the diagonal matrix containing the eigenvalues of $\boldsymbol{\Sigma}_{Y_AY_A}$ in descending order. Noticing that $\boldsymbol{\Sigma}_{AA}$ and $\boldsymbol{\Sigma}_{Y_AY_A}$ are also diagonalized by $\mathbf{U}_{YY}$, the integration domain $\mathcal{S}_1$ is given by

$$\mathcal{S}_1 = \left\{\mathbf{y}_1\in\mathbb{R}^m\colon \mathbf{y}_1^\mathrm{T}\boldsymbol{\Delta}_1\mathbf{y}_1 \geq 2\log\tau+\log\left|\mathbf{I}_m+\boldsymbol{\Lambda}_{AA}\boldsymbol{\Lambda}_{YY}^{-1}\right|\right\}, \tag{27}$$

where $\boldsymbol{\Delta}_1 \triangleq \boldsymbol{\Lambda}_{YY}^{-1} - \boldsymbol{\Lambda}_{Y_AY_A}^{-1}$ and $\boldsymbol{\Lambda}_{AA}$ denotes the diagonal matrix containing the eigenvalues of $\boldsymbol{\Sigma}_{AA}$ in descending order. Further applying the change of variable $\mathbf{y}_2 \triangleq \boldsymbol{\Lambda}_{Y_AY_A}^{-\frac{1}{2}}\mathbf{y}_1$ in (26) results in

$$\mathsf{P}_\mathsf{D}(\lambda) = \frac{1}{\sqrt{(2\pi)^m}}\int_{\mathcal{S}_2} \exp\{-\frac{1}{2}\mathbf{y}_2^\mathrm{T}\mathbf{y}_2\}\mathrm{d}\mathbf{y}_2, \tag{28}$$

with the transformed integration domain given by

$$\mathcal{S}_2 = \left\{\mathbf{y}_2\in\mathbb{R}^m\colon \mathbf{y}_2^\mathrm{T}\boldsymbol{\Delta}_2\mathbf{y}_2 \geq 2\log\tau+\log\left|\mathbf{I}_m+\boldsymbol{\Delta}_2\right|\right\}, \tag{29}$$

with

$$\boldsymbol{\Delta}_2 \triangleq \boldsymbol{\Lambda}_{AA}\boldsymbol{\Lambda}_{YY}^{-1}. \tag{30}$$

Setting $\boldsymbol{\Delta} \triangleq \lambda\boldsymbol{\Delta}_2$ and noticing that rank$(\boldsymbol{\Delta})$ = rank$(\mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathrm{T})$ concludes the proof. □

Lemma 1 shows that the probability of detection is equivalent to the probability that a weighted sum of independent χ^2 random variables exceeds a certain threshold. In our setting, the threshold is determined by the trade-off parameter λ. Notice that the left-hand term $(U^p)^\mathrm{T}\boldsymbol{\Delta}U^p$ in (21) is a weighted sum of independent χ^2 distributed random variables with one degree of freedom where the weights are determined by the diagonal entries of $\boldsymbol{\Delta}$ which depend on the second order statistics of the state variables, the Jacobian measurement matrix, and the variance of the noise; i.e. the attacker has no control over this term. The right-hand side contains in addition λ and τ, and therefore, the probability of attack detection is described as a function of the parameter λ.

Unfortunately, no closed-form expression is available for the distribution of a positively weighted sum of independent χ^2 random variables with one degree of freedom [36]. Usually, some moment matching approximation approaches, such as the Lindsay–Pilla–Basak (LPB) method [37], are utilized to solve this problem but the resulting expressions are complex and the relation of the probability of detection with λ is difficult to describe analytically following this course of action.

In the following an upper bound on the probability of attack detection is derived. The upper bound is then used to provide a simple lower bound on the value λ that achieves the desired probability of detection.

B. Upper Bound on the Probability of Detection

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

Theorem 2. *Let $\tau > 1$ be the decision threshold of the LRT in (9). Given $t > 0$ it holds that for all $\lambda \geq \max(\lambda^\star(t), 1)$ the probability of attack detection satisfies*

$$\mathsf{P}_\mathsf{D}(\lambda) \leq e^{-t}, \tag{31}$$

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

$$2\lambda\log\tau - \frac{1}{2\lambda}\mathrm{tr}(\boldsymbol{\Delta}^2) - 2\sqrt{\mathrm{tr}(\boldsymbol{\Delta}^2)t} - 2\|\boldsymbol{\Delta}\|_\infty t = 0, \tag{32}$$

and $\|\cdot\|_\infty$ is the infinity norm.

Proof. We start with the result of Lemma 1 which gives

$$\mathsf{P}_\mathsf{D}(\lambda)=\mathbb{P}\Big[(U^p)^\mathrm{T}\mathbf{\Delta}U^p\geq\lambda\left(2\log\tau+\log\left|\mathbf{I}_p+\lambda^{-1}\mathbf{\Delta}\right|\right)\Big]. \quad (33)$$

We now proceed to expand the term $\log\left|\mathbf{I}_p+\lambda^{-1}\mathbf{\Delta}\right|$ using a Taylor series expansion resulting in

$$\begin{aligned}&\log\left|\mathbf{I}_p+\lambda^{-1}\mathbf{\Delta}\right|\\&=\sum_{i=1}^{p}\log\left(1+\lambda^{-1}(\mathbf{\Delta})_{i,i}\right) \quad (34)\\&=\sum_{i=1}^{p}\left(\sum_{j=1}^{\infty}\left(\frac{\left(\lambda^{-1}(\mathbf{\Delta})_{i,i}\right)^{2j-1}}{2j-1}-\frac{\left(\lambda^{-1}(\mathbf{\Delta})_{i,i}\right)^{2j}}{2j}\right)\right). \quad (35)\end{aligned}$$

Since $(\mathbf{\Delta})_{i,i}\leq 1$ for $i=1,\ldots,p$, and $\lambda\geq 1$, then

$$\frac{\left(\lambda^{-1}(\mathbf{\Delta})_{i,i}\right)^{2j-1}}{2j-1}-\frac{\left(\lambda^{-1}(\mathbf{\Delta})_{i,i}\right)^{2j}}{2j}\geq 0,\ \text{for } j\in\mathbb{Z}^+. \quad (36)$$

Thus, (35) is lower bounded by the second order Taylor expansion, i.e.,

$$\begin{aligned}\log|\mathbf{I}_p+\mathbf{\Delta}|&\geq\sum_{i=1}^{p}\left(\lambda^{-1}(\mathbf{\Delta})_{i,i}-\frac{\left(\lambda^{-1}(\mathbf{\Delta})_{i,i}\right)^2}{2}\right) \quad (37)\\&=\frac{1}{\lambda}\mathrm{tr}(\mathbf{\Delta})-\frac{1}{2\lambda^2}\mathrm{tr}(\mathbf{\Delta}^2). \quad (38)\end{aligned}$$

Substituting (38) in (33) yields

$$\mathsf{P}_\mathsf{D}(\lambda)\leq\mathbb{P}\left[(U^p)^\mathrm{T}\mathbf{\Delta}U^p\geq\mathrm{tr}(\mathbf{\Delta})+2\lambda\log\tau-\frac{1}{2\lambda}\mathrm{tr}(\mathbf{\Delta}^2)\right]. \quad (39)$$

Note that $\mathbb{E}\left[(U^p)^\mathrm{T}\mathbf{\Delta}U^p\right]=\mathrm{tr}(\mathbf{\Delta})$, and therefore, evaluating the probability in (39) is equivalent to evaluating the probability of $(U^p)^\mathrm{T}\mathbf{\Delta}U^p$ deviating $2\lambda\log\tau-\frac{1}{2\lambda}\mathrm{tr}(\mathbf{\Delta}^2)$ from the mean. In view of this and the results in [29] and [38], the right-hand side in (39) is upper bounded by

$$\begin{aligned}\mathsf{P}_\mathsf{D}(\lambda)&\leq\mathbb{P}\left[(U^p)^\mathrm{T}\mathbf{\Delta}U^p\geq\mathrm{tr}(\mathbf{\Delta})+2\sqrt{\mathrm{tr}(\mathbf{\Delta}^2)t}+2||\mathbf{\Delta}||_\infty t\right] \quad (40)\\&\leq e^{-t}, \quad (41)\end{aligned}$$

for $t>0$ satisfying

$$2\lambda\log\tau-\frac{1}{2\lambda}\mathrm{tr}(\mathbf{\Delta}^2)\geq 2\sqrt{\mathrm{tr}(\mathbf{\Delta}^2)t}+2||\mathbf{\Delta}||_\infty t. \quad (42)$$

The expression in (42) is satisfied with equality for two values of λ, one is strictly negative and the other one is strictly positive denoted by $\lambda^*(t)$, when $\tau>1$. The result follows by noticing that the left-hand term of (42) increases monotonically for $\lambda>0$ and choosing $\lambda\geq\max\left(\lambda^\star(t),1\right)$. This concludes the proof. □

It is interesting to note that for large values of λ the probability of detection decreases exponentially fast with λ. We will later show in the numerical results that the regime in which the exponentially fast decrease kicks in does not align with the saturation of the mutual information loss induced by the attack.

V. Numerical Simulation

In this section, we present simulations to evaluate the performance of the proposed attack strategy in practical state estimation settings. In particular, the IEEE 14-Bus, 30-Bus and 118-Bus test systems are considered in the simulation. In state estimation with linearized dynamics, the Jacobian measurement matrix is determined by the operation point. We assume a DC state estimation scenario [30], [31], and thus, we set the resistances of the branches to 0 and the bus voltage magnitude to 1.0 per unit. Note that in this setting it is sufficient to specify the network topology, the branch reactances, real power flow, and the power injection values to fully characterize the system. Specifically, we use the IEEE test system framework provided by MATPOWER [39]. We choose the bus voltage angle to be the state variables, and use the power injection and the power flows in both directions as the measurements.

As stated in Section IV-A, there is no closed-form expression for the distribution of a positively weighted sum of independent χ^2 random variables, which is required to calculate the probability of detection of the generalized stealth attacks as shown in Lemma 1. For that reason, we use the LPB method and the MOMENTCHI2 package [40] to numerically evaluate the probability of attack detection.

The simulation settings are the same as in [25]. The covariance matrix of the state variables is assumed to be a Toeplitz matrix with exponential decay parameter ρ, where the exponential decay parameter ρ determines the correlation strength between different entries of the state variable vector. The performance of the generalized stealth attack is a function of weight given to the detection term in the attack construction cost function, i.e. λ, the correlation strength between state variables, i.e. ρ, and the Signal-to-Noise Ratio (SNR) of the power system which is defined as

$$\mathrm{SNR}\stackrel{\Delta}{=}10\log_{10}\left(\frac{\mathrm{tr}(\mathbf{H}\mathbf{\Sigma}_{XX}\mathbf{H}^\mathrm{T})}{m\sigma^2}\right). \quad (43)$$

A. Generalized Stealth Attack Performance

Fig. 1 and Fig. 2 depict the performance of the optimal attack construction given in (17) for different values of ρ with SNR $=$ 10 dB and SNR $=$ 20 dB, respectively, when $\lambda=2$ and $\tau=2$. Interestingly, the performance of the attack construction does not change monotonically with the correlation strength, which suggests that the correlation among the state variables does not necessarily provide an advantage to the attacker. Admittedly, for a small and moderate values of ρ, the performance of the attack does not change significantly with ρ for both objectives. This effect is more noticeable in the high SNR scenario. However, for large values of ρ the performance of the attack improves significantly in terms of both mutual information and probability of detection. Moreover, the advantage provided by large values of ρ is more significant for the 118-Bus system than for the 30-Bus system,

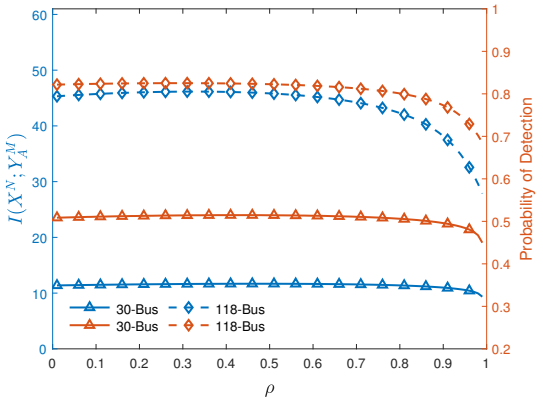

Fig. 1. Performance of the generalized stealth attack in terms of mutual information and probability of detection for different values of ρ when $\lambda = 2$, $\tau = 2$, and SNR $= 10$ dB.

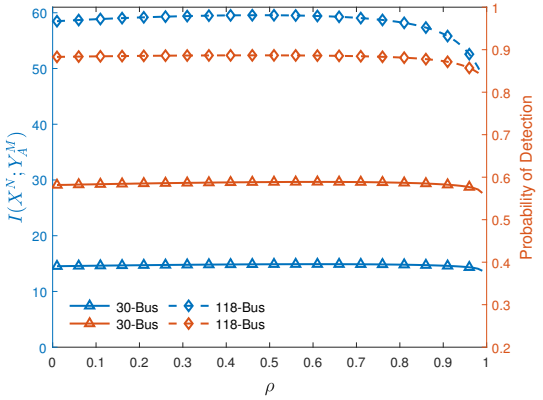

Fig. 2. Performance of the generalized stealth attack in terms of mutual information and probability of detection for different values of ρ when $\lambda = 2$, $\tau = 2$, and SNR $= 20$ dB.

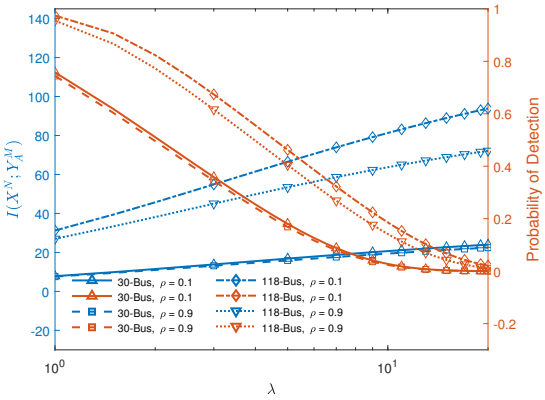

Fig. 3. Performance of the generalized stealth attack in terms of mutual information and probability of detection for different values of λ and system size when $\rho = 0.1$, $\rho = 0.9$, SNR $=$ 10 dB and $\tau = 2$.

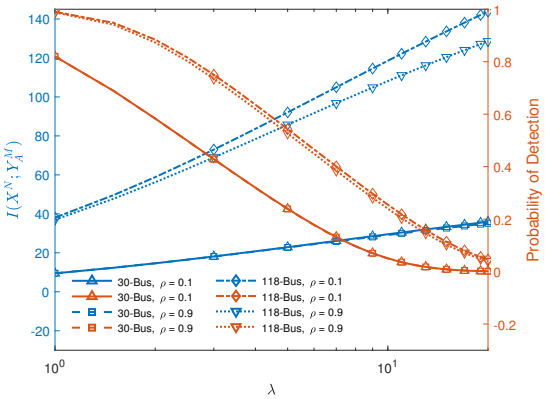

Fig. 4. Performance of the generalized stealth attack in terms of mutual information and probability of detection for different values of λ and system size when $\rho = 0.1$, $\rho = 0.9$, SNR $=$ 20 dB and $\tau = 2$.

which indicates that correlation between the state variables is easier to exploit for the attacker in large systems.

Fig. 3 and Fig. 4 depict the performance of the optimal attack construction for different values of λ and ρ with SNR $= 10$ dB and SNR $= 20$ dB, respectively, when $\tau = 2$. As expected, larger values of the parameter λ yield smaller values of the probability of attack detection while increasing the mutual information between the state variables vector and the compromised measurement vector. We observe that the probability of detection decreases approximately linearly with respect to $\log \lambda$ for moderate values of λ. On the other hand, Theorem 2 states that for large values of λ the probability of detection decreases exponentially fast to zero. However, for the range of values of λ in which the decrease of probability of detection is approximately linear with respect to $\log \lambda$, there is no significant reduction on the rate of growth of mutual information. In view of this, the attacker needs to choose the value of λ carefully as the convergence of the mutual information to the asymptote $I(X^N;Y^M)$ is slower than that of the probability of detection to zero.

The comparison between the 30-Bus and 118-Bus systems shows that for the smaller size system the probability of detection decreases faster to zero while the rate of growth of mutual information is smaller than that on the larger system. This suggests that the choice of λ is particularly critical in large size systems as smaller size systems exhibit a more robust attack performance for different values of λ. The effect of the correlation between the state variables is significantly more noticeable for the 118-bus system. While there is a performance gain for the 30-bus system in terms of both mutual information and probability of detection due to the high correlation between the state variables, the improvement is more noteworthy for the 118-bus case. Remarkably, the difference in terms of mutual information between the case in which $\rho = 0.1$ and $\rho = 0.9$ increases as λ increases which indicates that the cost in terms of mutual information of reducing the probability of detection is large in the small values of correlation.

The performance of the upper bound given by Theorem 2 on the probability of detection for different values of λ and ρ when $\tau = 2$ and SNR $= 10$ dB is shown in Fig. 5. Similarly, Fig. 6 depicts the upper bound with the same parameters but with SNR $= 20$ dB. As shown by Theorem 2 the bound decreases exponentially fast for large values of

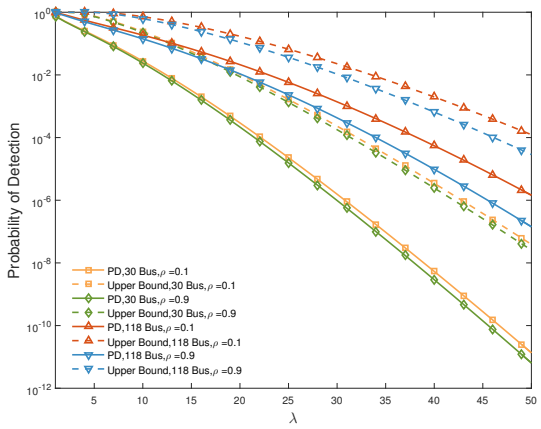

Fig. 5. Upper bound on probability of detection given in Theorem 2 for different values of λ when $\rho = 0.1$ or 0.9, SNR $= 10$ dB, and $\tau = 2$.

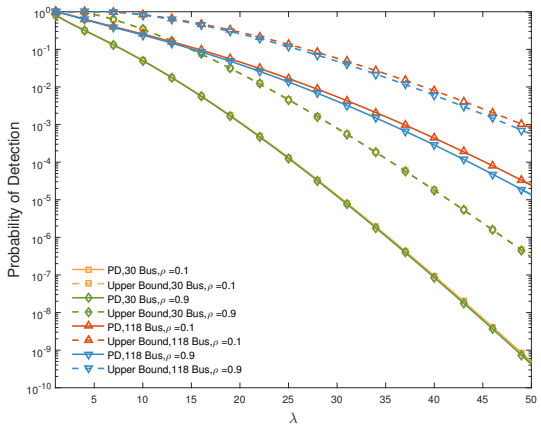

Fig. 6. Upper bound on probability of detection given in Theorem 2 for different values of λ when $\rho = 0.1$ or 0.9, SNR $= 20$ dB, and $\tau = 2$.

λ. Still, there is a significant gap to the probability of attack detection evaluated numerically. This is partially due to the fact that our bound is based on the concentration inequality in [29] which introduces a gap of more than an order of magnitude. Interestingly, the gap decreases when the value of ρ increases although the change is not significant. More importantly, the bound is tighter for lower values of SNR for both 30-bus and 118-bus systems.

B. Performance and Sensitivity under AC State Estimation

In the AC state estimation case the iterative estimation methods require a nominal operation point that is updated for each iteration. When the attacker has the perfect information about the operation point in each iteration, i.e. perfect information about Jacobian matrix $\mathbf{H}$ in each iteration, the resulting mutual information and probability of detection follow from Corollary 1 and Lemma 1 directly. *In the following, we study the impact of imperfect nominal operation point information on the attack performance.* In particular the generalized stealth attacks are constructed as $A_0^m \sim \mathcal{N}(\mathbf{0}, \frac{1}{\lambda}\mathbf{H}_0\mathbf{\Sigma}_{XX}\mathbf{H}_0^{\mathrm{T}})$, where $\mathbf{H}_0$ is the Jacobian matrix at the nominal operation point $\mathbf{x}_0$ and is given

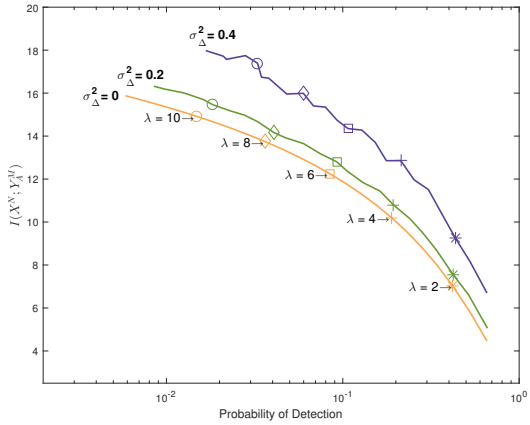

Fig. 7. Performance of generalized stealth attack in terms of mutual information and probability of detection for different values of σ_Δ^2 and λ on IEEE 14-Bus system when $\rho = 0.1$, $\tau = 2$, and SNR $= 20$ dB. The marker represents the same value of λ is used in the attack construction.

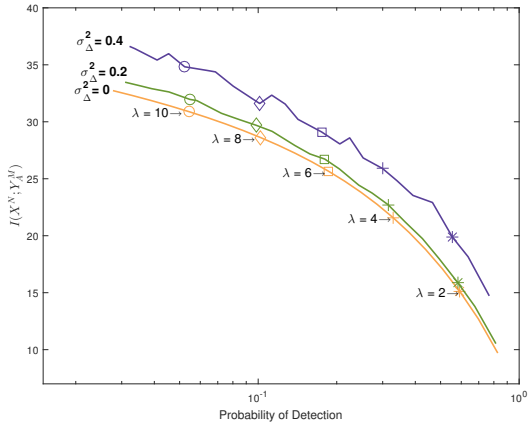

Fig. 8. Performance of generalized stealth attack in terms of mutual information and probability of detection for different values of σ_Δ^2 and λ on IEEE 30-Bus system when $\rho = 0.1$, $\tau = 2$, and SNR $= 20$ dB. The marker represents the same value of λ is used in the attack construction.

by

$$\mathbf{H}_0 = \frac{\partial}{\partial(X^n)} H(X^n)|_{X^n=\mathbf{x}_0}, \tag{44}$$

with $H(X^n) \in \mathbb{R}^m$ denoting the vector of random variables induced by the nonlinear relation between the state variables and the measurements. To model the imperfect knowledge of the nominal point, the nominal linearization point is perturbed with random variable $\Delta \sim \mathcal{N}(\mathbf{0}, \sigma_\Delta^2\mathbf{I})$ resulting in the Jacobian matrix $\mathbf{H}$ given by

$$\mathbf{H} = \frac{\partial}{\partial(X^n)} H(X^n)|_{X^n=\mathbf{x}_0+\Delta}. \tag{45}$$

Note that the introduction of this random perturbation gives us a way to control the strength of the perturbation, i.e. the uncertainty over the nominal linearization point, and as a result we study the sensitivity of the attacks under AC state estimation by changing the variance σ_Δ^2 in the simulations.

Fig. 7 depicts the performance of the generalized stealth attacks in terms of the mutual information and the probability of detection for different values of σ_Δ^2 and λ on the IEEE 14-Bus system when $\rho = 0.1$, $\tau = 2$, and SNR $= 20$ dB. Similarly Fig. 8 shows the performance of the attacks under the same setting on IEEE 30-Bus system. We generate 200 realizations of Δ per point and for each realization of Δ we evaluate 2000 realizations of the state variables. The curve corresponding to the case when $\sigma_\Delta^2 = 0$ describes the performance of the attacks with perfect knowledge of the nominal operation point. As expected, when there is less accurate knowledge about the nominal operation point, i.e. σ_Δ^2 increases, the performance of the attack A_0^m decreases. Interestingly the performance decrease translates in a larger value of mutual information for all cases. However, the change in probability of detection is not as significant, to the extent that in some cases the probability of detection decreases. Note that for all cases, overall the attack performance decreases when perfect operation point is not available. Interestingly, the stealth of the attacks is more robust for the IEEE 30-Bus system than for the IEEE 14-Bus system, which suggests that the attacker is better positioned to cope with system uncertainty for larger networks.

VI. Conclusions

We have proposed a novel data injection attacks based on information-theoretic performance measures. Specifically, we have posed the attack construction problem as an optimization problem in which the cost function combines the mutual information and the probability of attack detection. The proposed cost function allows to obtain an arbitrarily small probability of attack detection via a parameter that weights the effect of the mutual information and the probability of detection. The resulting random attack construction has been analyzed in terms of the information loss and the probability of attack detection that it induces on the system. We have characterized the probability of attack detection by obtaining an easy to compute upper bound. The upper bound has been used to provide a practical attack construction guideline by determining the cost function that achieves a given probability of attack detection.

References

[1] U.S.-Canada Power System Outage Task Force, *Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations*, 2004.

[2] D. Alderson and R. Di Pietro, "Operational technology: Are you vulnerable?," *Governance Directions*, vol. 68, no. 6, pp. 339–343, Jul. 2016.

[3] Royal Society, *Progress and Research in Cybersecurity: Supporting a Resilient and Trustworthy System for the UK*, 2016.

[4] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. ACM Conf. on Comput. and Commun. Security*, Chicago, IL, USA, Nov. 2009, pp. 21–32.

[5] G. Dán and H. Sandberg, "Stealth attacks and protection schemes for state estimators in power systems," in *Proc. IEEE Int. Conf. on Smart Grid Comm.*, Gaithersburg, MD, USA, Oct. 2010, pp. 214–219.

[6] H. Sandberg, A. Teixeira, and K. H. Johansson, "On security indices for state estimators in power networks," in *Proc. 1st Workshop on Secure Control Syst.*, Stockholm, Sweden, Apr. 2010.

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

[8] 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.

[9] 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.

[10] 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.

[11] R. Deng, P. Zhuang, and H. Liang, "False data injection attacks against state estimation in power distribution systems," *IEEE Trans. Smart Grid (Early Access)*, 2018.

[12] Y. Hao, M. Wang, and J. H. Chow, "Likelihood analysis of cyber data attacks to power systems with Markov decision processes," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3191–3202, Jul. 2018.

[13] R. Moslemi, A. Mesbahi, and J. M. Velni, "A fast, decentralized covariance selection-based approach to detect cyber attacks in smart grids," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4930–4941, Sep. 2018.

[14] A. Ashok, M. Govindarasu, and V. Ajjarapu, "Online detection of stealthy false data injection attacks in power system state estimation," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 1636–1646, May 2018.

[15] S. Sridhar and M. Govindarasu, "Model-based attack detection and mitigation for automatic generation control," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 580–591, Mar. 2014.

[16] L. Che, X. Liu, and Z. Li, "Mitigating false data attacks induced overloads using a corrective dispatch scheme," *IEEE Trans. Smart Grid (Early Access)*, 2018.

[17] 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. IEEE Workshop on Signal Process. Advances in Wireless Commun.*, Edinburgh, UK, Jul. 2016, pp. 1–6.

[18] 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.

[19] 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.

[20] Q. Li, T. Cui, Y. Weng, R. Negi, F. Franchetti, and M. D. Ilić, "An information-theoretic approach to PMU placement in electric power systems," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 446–456, Mar. 2013.

[21] D. Varodayan and A. Khisti, "Smart meter privacy using a rechargeable battery: Minimizing the rate of information leakage," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process.*, Prague, Czech Republic, May 2011, pp. 1932–1935.

[22] L. Sankar, S.R. Rajagopalan, S. Mohajer, and H.V. Poor, "Smart meter privacy: A theoretical framework," *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 837–846, Jun. 2013.

[23] O. Tan, D. Gündüz, and H. V. Poor, "Increasing smart meter privacy through energy harvesting and storage devices," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1331–1341, Jul. 2013.

[24] M. Arrieta and I. Esnaola, "Smart meter privacy via the trapdoor channel," in *Proc. IEEE Int. Conf. on Smart Grid Comm.*, Dresden, Germany, Oct. 2017, pp. 227–282.

[25] 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.

[26] J. Hou and G. Kramer, "Effective secrecy: Reliability, confusion and stealth," in *Proc. IEEE Int. Symp. on Inf. Theory*, Honolulu, HI, USA, Jun. 2014, pp. 601–605.

[27] 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.

[28] Y. Huang, M. Esmalifalak, H. Nguyen, R. Zheng, Z. Han, H. Li, and L. Song, "Bad data injection in smart grid: Attack and defense mechanisms," *IEEE Commun. Mag.*, vol. 51, no. 1, pp. 27–33, Jan. 2013.

[29] B. Laurent and P. Massart, "Adaptive estimation of a quadratic functional by model selection," *Ann Statist*, vol. 28, no. 5, pp. 1302–1338, 2000.

[30] A. Abur and A. G. Expósito, *Power System State Estimation: Theory and Implementation*, CRC Press, Mar. 2004.

[31] J. J. Grainger and W. D. Stevenson, *Power System Analysis*, McGraw-Hill, 1994.

[32] 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.
[33] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer, New York, 1994.
[34] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Nov. 2012.
[35] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Mar. 2004.
[36] D. A. Bodenham and N. M. Adams, "A comparison of efficient approximations for a weighted sum of chi-squared random variables," *Stat Comput*, vol. 26, no. 4, pp. 917–928, Jul. 2016.
[37] Bruce G. Lindsay, Ramani S. Pilla, and Prasanta Basak, "Moment-based approximations of distributions using mixtures: Theory and applications," *Ann. Inst. Stat. Math.*, vol. 52, no. 2, pp. 215–230, Jun. 2000.
[38] D. Hsu, S.M. Kakade, and T. Zhang, "A tail inequality for quadratic forms of subgaussian random vectors," *Electron. Commun. in Probab.*, vol. 17, no. 52, pp. 1–6, 2012.
[39] 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. 2011.
[40] D. Bodenham, *Momentchi2: Moment-Matching Methods for Weighted Sums of Chi-Squared Random Variables*. (2016) [Online]. Available: https://cran.r-project.org/web/packages/momentchi2/index.html.