

Stealth Data Injection Attacks with Sparsity Constraints

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

▶ **To cite this version:**

Xiuzhen Ye, Iñaki Esnaola, Samir M. Perlaza, Robert F Harrison. Stealth Data Injection Attacks with Sparsity Constraints. [Research Report] RR-9481, Institut National de Recherche en Informatique et en Automatique (INRIA). 2022. hal-03781671v2

HAL Id: hal-03781671
https://hal.science/hal-03781671v2

Submitted on 28 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

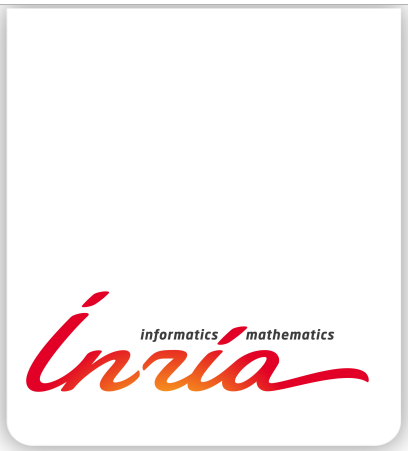


Stealth Data Injection Attacks with Sparsity Constraints

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

RESEARCH REPORT
N° 9481
September 2022
Project-Team NEO

ISSN 0249-6399 ISRN INRIA/RR--9481--FR+ENG



Stealth Data Injection Attacks with Sparsity Constraints

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

Project-Team NEO

Research Report n° 9481 — September 2022 — 37 pages

Abstract: In this report, sparse stealth data-injection attacks 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 aiming to minimize the mutual information subject to a constraint to the Kullback-Leibler divergence between the distribution of the observations under attack and 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. The second algorithm considers correlations between the attack vector entries, which results in larger disruption and smaller probability of detection. A performance analysis of the proposed attack constructions on IEEE test systems is presented. Using a numerical example, it is shown that it is feasible to construct stealth attacks that generate significant disruption with a low number of compromised sensors.

Key-words: Data Injection Attacks, Cyber-security, Sparse Constraints, Smart Grid, Information Theory.

Xiuzhen Ye, Iñaki Esnaola and Robert F. Harrison are with the Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, UK. Samir M. Perlaza is with INRIA, Sophia Antipolis, France. (samir.perlaza@inria.fr); the Department of Electrical and Computer Engineering, Princeton University, USA; and the Mathematics Laboratory (GAATI), University of the French Polynesia, Tahiti, French Polynesia. Iñaki Esnaola is also with the Department of Electrical and Computer Engineering, Princeton University, USA.

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.

RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Résumé : Dans ce rapport, des constructions d'attaques furtives ciblant un sous-ensemble des capteurs qui minimisent l'information mutuelle entre les variables d'état et les observations sont proposées. La construction de l'attaque est formulée comme la conception d'une distribution gaussienne multivariée visant à minimiser l'information mutuelle tout en limitant la divergence de Kullback-Leibler entre la distribution des observations sous attaque et la distribution des observations sans attaque. La contrainte pour modéliser le fait que l'attaque cible uniquement un sous-ensemble des capteurs est incorporée en tant que contrainte sur le support de la distribution de probabilité de l'attaque. Deux algorithmes heuristiques gloutons pour la construction des attaques sont proposés. Le premier algorithme suppose que le vecteur d'attaque se compose d'entrées indépendantes et, par conséquent, ne nécessite aucune communication entre les différents emplacements attaqués. Le deuxième algorithme prend en compte les corrélations entre les entrées du vecteur d'attaque, ce qui entraîne une perturbation plus importante et une probabilité de détection plus faible. Une analyse des performances des constructions d'attaque proposées sur les systèmes de test IEEE est présentée. À l'aide d'un exemple numérique, il est démontré qu'il est possible de construire des attaques furtives qui génèrent des perturbations importantes avec un faible nombre de capteurs compromis.

Mots-clés : Attaques par injection de données, cybersécurité, contraintes parcimonieuses, réseau intelligent, théorie de l'information

Contents

1 Introduction **4**

2 System model **6**
2.1 Observation Model and Attack Setting 6
2.2 Attack Detection 7

3 Sparse Stealth Attacks **8**
3.1 Information Theoretic Metric 8
3.2 Sparse Stealth Attack Formulation 10
3.3 Gaussian Sparse Stealth Attack Construction 11

4 Independent Sparse Stealth Attacks **12**
4.1 Independent Structure 12
4.2 Greedy Independent Attack Construction 12

5 Correlated Sparse Stealth Attacks **15**
5.1 Correlation Structure 15
5.2 Greedy Correlated Attack Construction 15

6 Numerical Results **16**
6.1 Performance in terms of information theoretic cost 17
6.2 Performance in terms of the tradeoff between mutual information and KL divergence 20
6.3 Performance in terms of mutual information and probability of attack detection 27

7 Conclusions **28**

Appendices **29**

A Proof of Proposition 3.1.1 **29**

B Proof of Proposition 3.1.2 **31**

C Proof of Lemma 4.1 **32**

D Proof of Theorem 4.1 **32**

E Proof of Theorem 5.1 **34**

1 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 obtained from different estimation methods by compromising the system observations without triggering bad data detection mechanisms set by the system operator [5]. A large body of literature studies the case in which attack detection is performed by a residual test [6] under the assumption that state estimation is deterministic both in centralized and decentralized scenarios [7–10]. 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 [11], 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 [12]. In this Bayesian paradigm, the attack detection can be formulated as the likelihood ratio test [13] or alternatively machine learning methods [14] 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 [15]. 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 [16] and then generalized in [17]. In this research, the attack construction uses information theoretic measures, that is, Mutual Information and Kullback-Leibler (KL) divergence, to characterize the fundamental limits of the attack [18]. In [12, 16, 17, 19], 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 [20, 21]. However, both the stealth attacks constructed in [16] and [17] 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 [19] and studied in [22]. 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 2, we introduce a Bayesian framework with linearized dynamics for DIAs. Stealth attacks incorporating sparsity constraints are presented in Section 3. Independent sparse stealth attacks and correlated sparse stealth attacks are presented in Section 4 and Section 5, respectively. In Section 6, 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 7.

The main contributions of this paper follow: (1) A novel stealth attack construction with sparsity constraints in Bayesian framework is proposed where the sparse attack is constructed as random attacks. (2) Information measures are firstly used to construct sparse attacks. Precisely, the attack construction jointly minimizes mutual information and KL divergence. (3) We tackle the challenge of the combinatorial character of identifying the support of the sparse attack vector by incorporating an additional sensor that yields a sequential sensor selection problem. (4) Both independent attacks and correlated attacks are considered. In the first case, the random attack requires no communication between locations because its entries are independent. On the other hand, there is correlation between entries in the second case which leads to a better attack performance at the expense of communication. The convexity of the resulting optimization problems in both cases are provided and the insight obtained from incorporating an additional sensor has been distilled to propose heuristic greedy algorithms, accordingly.

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 $D(P\|Q)$.

2 System model

2.1 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 entries 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

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

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

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

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

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

where $\boldsymbol{\mu}_A \in \mathbb{R}^m$ and $\mathbf{\Sigma}_{AA} \in S_+^m$ are the mean vector and the covariance matrix of the random vector A^m.

The choice in (5) is justified by the fact that when $Z^m + A^m$ in (6) follows a Gaussian distribution, the mutual information between the state variables X^n and the compromised observations Y_A^m, denoted by $I(X^n; Y_A^m)$, is minimized [23]. Hence, from the Lévy-Cramér decomposition theorem [24] [25], it holds that for the sum $Z^m + A^m$ to be Gaussian, given that Z^m satisfies (2), then, A^m must be Gaussian. This choice is further discussed in Section 3.1.

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

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

with

$$\boldsymbol{\Sigma}_{Y_A Y_A} \triangleq \mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T} + \sigma^2\mathbf{I}_m + \boldsymbol{\Sigma}_{AA}. \tag{8}$$

2.2 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

$$\mathcal{H}_0 : \text{There is no attack,} \tag{9a}$$
$$\mathcal{H}_1 : \text{Observations are compromised.} \tag{9b}$$

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 (9) become

$$\mathcal{H}_0 : \bar{Y}^m \sim P_{Y^m}, \tag{10a}$$
$$\mathcal{H}_1 : \bar{Y}^m \sim P_{Y_A^m}. \tag{10b}$$

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 [26]. In this setting, the LRT is given by

$$T(\bar{\mathbf{y}}) = \mathbb{1}_{\{L(\bar{\mathbf{y}}) \gtrless \tau\}}, \tag{11}$$

with $L(\bar{\mathbf{y}})$ is the likelihood ratio, that is,

$$L(\bar{\mathbf{y}}) = \frac{f_{Y_A^m}(\bar{\mathbf{y}})}{f_{Y^m}(\bar{\mathbf{y}})}, \tag{12}$$

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 (11) is the decision threshold. Note that changing the value of τ is equivalent to change the tradeoff between Type-I and Type-II errors.

3 Sparse Stealth Attacks

3.1 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 [17] 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 [27]: (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 following proposition presents the analytical expression of mutual information with X^n in (3) and Y_A^m in (6).

Proposition 3.1.1. *The mutual information between the random variable* $X^n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{XX})$ *and* $Y_A^m \sim \mathcal{N}(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{Y_AY_A})$ *is*

$$I(X^n; Y_A^m) = \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_{XX}||\boldsymbol{\Sigma}_{Y_AY_A}|}{|\boldsymbol{\Sigma}|}, \tag{13}$$

where the matrix $\boldsymbol{\Sigma}_{XX}$ *is in* (3)*, the matrix* $\boldsymbol{\Sigma}_{Y_AY_A}$ *is in* (8) *and the matrix* $\boldsymbol{\Sigma}$ *is the covariance matrix of the joint distribution of* X^n *and* Y_A^m*, that is,* $(X^n; Y_A^m) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ *with*

$$\boldsymbol{\Sigma} \triangleq \begin{pmatrix} \boldsymbol{\Sigma}_{XX} & \boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}} \\ \mathbf{H}\boldsymbol{\Sigma}_{XX} & \mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^{\mathsf{T}} + \sigma^2 \boldsymbol{I}_m + \boldsymbol{\Sigma}_{AA} \end{pmatrix}. \tag{14}$$

Proof. The proof is presented in Appendix A. □

Corollary 3.1. *The mutual information between the vector of random variables* $X^n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{XX})$ *and* $Y_{A^m} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{Y_AY_A})$

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

where the matrix $\boldsymbol{\Sigma}_{XX}$ *is in* (3)*, the matrix* $\boldsymbol{\Sigma}_{Y_AY_A}$ *is in* (8) *and the matrix* $\boldsymbol{\Sigma}$ *is in* (14)*.*

The KL divergence term guarantees a stealthy attack in the sense that its minimization leads to minimizing the absolute difference between the probability of false alarm and the probability of attack detection, that is, $|\alpha - (1 - \beta)|$ [26, 28]. The following proposition presents the analytical expression of KL divergence.

Proposition 3.1.2. *The KL divergence between* $Y_A^m \sim \mathcal{N}(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{Y_AY_A})$ *and* $Y^m \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{YY})$ *is*

$$D(P_{Y_A^m} \| P_{Y^m}) = \frac{1}{2}\left(\log \frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{Y_AY_A}|} - m + tr\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{Y_AY_A}\right) + tr\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\mu}_A\boldsymbol{\mu}_A^{\mathsf{T}}\right)\right), \quad (16)$$

where the mean vector $\boldsymbol{\mu}_A$ *and the matrix* $\boldsymbol{\Sigma}_{Y_AY_A}$ *are in* (8)*, the matrix* $\boldsymbol{\Sigma}_{YY}$ *is in* (4)*.*

Proof. The proof of Proposition 3.1.2 is presented in Appendix B. □

Corollary 3.2. *The KL divergence between* m*-dimension multivariate Gaussian distribution* $Y_{A^m} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{Y_AY_A})$ *and* $Y^m \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{YY})$ *is given by*

$$D(P_{Y_A^m} \| P_{Y^m}) = \frac{1}{2}\left(\log \frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{Y_AY_A}|} - m + \mathrm{tr}(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{Y_AY_A})\right), \quad (17)$$

where the matrix $\boldsymbol{\Sigma}_{YY}$ *and* $\boldsymbol{\Sigma}_{Y_AY_A}$ *are in* (4) *and* (8)*, respectively.*

The following lemma shows that the optimal mean vector $\boldsymbol{\mu}_A$ of the Gaussian attack construction in (5) is a null vector.

Lemma 3.3. *The optimal Gaussian attack construction is with a null mean vector, that is,*

$$A^m \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{AA}), \quad (18)$$

where $\boldsymbol{\Sigma}_{AA} \in \mathcal{S}_+^m$*.*

Proof. From Proposition 3.1.1, the mutual information in (13) does not depend on the mean vector $\boldsymbol{\mu}_A$. From Proposition 3.1.2, the following holds

$$\begin{aligned} D(P_{Y_A^m} \| P_{Y^m}) =& \frac{1}{2}\left(\log \frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{Y_AY_A}|} - m + \mathrm{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{Y_AY_A}\right) + \mathrm{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\mu}_A\boldsymbol{\mu}_A^{\mathsf{T}}\right)\right) & (19)\\ =& \frac{1}{2}\left(\log \frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{Y_AY_A}|} - m + \mathrm{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{Y_AY_A}\right) + \boldsymbol{\mu}_A^{\mathsf{T}}\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\mu}_A\right) & (20)\\ \geq& \frac{1}{2}\left(\log \frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{Y_AY_A}|} - m + \mathrm{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{Y_AY_A}\right)\right) & (21) \end{aligned}$$

where the equality in (20) follows from the fact that

$$\mathrm{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\mu}_A\boldsymbol{\mu}_A^{\mathsf{T}}\right) = \mathrm{tr}\left(\boldsymbol{\mu}_A^{\mathsf{T}}\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\mu}_A\right) = \boldsymbol{\mu}_A^{\mathsf{T}}\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\mu}_A, \quad (22)$$

and the equality in (21) follows from $\boldsymbol{\Sigma}_{YY}^{-1} \in \mathcal{S}_+^m$. Note that the equality in (21) holds only when $\boldsymbol{\mu}_A = \mathbf{0}$. Therefore, for all $\boldsymbol{\Sigma}_{AA} \in \mathcal{S}_+^m$, the optimal mean vector for Gaussian attack construction is $\boldsymbol{\mu}_A = \mathbf{0}$. This completes the proof. □

Within this framework, stealth attacks are constructed as random vectors $A^m \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{AA})$ whose probability distribution functions are the solution to the following optimization problem:

$$\min_{P_{A^m}} I(X^n; Y_A^m) + \lambda D(P_{Y_A^m} \| P_{Y^m}), \tag{23}$$

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 (23) is a multivariate Gaussian distribution for the attack vector. It is shown in [17] that the optimal Gaussian attack is given by

$$\bar{P}_{A^m} \sim \mathcal{N}(\mathbf{0}, \bar{\mathbf{\Sigma}}), \tag{24}$$

with

$$\bar{\mathbf{\Sigma}} = \lambda^{-1/2} \mathbf{H} \mathbf{\Sigma}_{XX} \mathbf{H}^{\mathsf{T}}. \tag{25}$$

Note that the optimal Gaussian stealth attacks in (24) yields a stealth attack vector that is not sparse, indeed all the entries of the attack realizations are nonzero with probability one, that is, $\mathbb{P}\left[|\text{supp}(A^m)| = m\right] = 1$, where we define the support of the attack vector A^m as

$$\text{supp}(A^m) \triangleq \left\{ i : \mathbb{P}\left[A_i = 0\right] = 0 \right\}. \tag{26}$$

3.2 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, that is, 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 entries.

In the 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

$$\mathcal{P}_k \triangleq \left\{ P_{A^m} \sim \mathcal{N}(\mathbf{0}, \bar{\mathbf{\Sigma}}) : |\text{supp}(A^m)| = k \right\}. \tag{27}$$

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

$$\min_{P_{A^m} \in \mathcal{P}_k} I(X^n; Y_A^m) + \lambda D(P_{Y_A^m} \| P_{Y^m}). \tag{28}$$

The optimization domain including the sparsity constraint in (27) implies an additional difficulty in the construction of stealth attacks with respect to the construction proposed in [17]. 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.

3.3 Gaussian Sparse Stealth Attack Construction

From Lemma 3.3, the probability distribution function of a random vector is determined by one parameters, that is, the covariance matrix. Hence, from Corollary 3.1 and Corollary 3.2, writing the objective function of the optimization problems in (23) and (28) in terms of the 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,

$$J(\mathbf{\Sigma}_{AA}) \stackrel{\Delta}{=} (1-\lambda) \log |\mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_{AA}| - \log |\sigma^2 \mathbf{I}_m + \mathbf{\Sigma}_{AA}| + \lambda \mathrm{tr}(\mathbf{\Sigma}_{YY}^{-1} \mathbf{\Sigma}_{AA}), \tag{29}$$

where $\lambda \geq 1$ is introduced in (23); and the matrix $\mathbf{\Sigma}_{YY}$ is defined by (4). Hence, the optimization problem in (23) is equivalent to the following optimization problem:

$$\min_{\mathbf{\Sigma}_{AA} \in S_+^m} J(\mathbf{\Sigma}_{AA}). \tag{30}$$

In order to write the optimization domain of the problem in (28) in terms of the mean vector and covariance matrix of the attack random vector, it suffices to observe that the sparsity constraint in (27) 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:

$$\mathcal{S}_k \stackrel{\Delta}{=} \left\{ \mathbf{S} \in S_+^m : \|\mathrm{diag}(\mathbf{S})\|_0 = k \right\}, \tag{31}$$

where $\mathrm{diag}(\mathbf{S})$ denotes the vector formed by the diagonal entries of $\mathbf{S}$. Solving (30) within the optimization domain specified by (31) re-casts the equivalent k-sparse stealth attack construction problem in (28) as:

$$\min_{\mathbf{\Sigma}_{AA} \in \mathcal{S}_k} J(\mathbf{\Sigma}_{AA}). \tag{32}$$

4 Independent Sparse Stealth Attacks

4.1 Independent Structure

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

$$P_{A^m} = \prod_{i=1}^{m} P_{A_i}, \tag{33}$$

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

The assumption of independence relaxes the correlation requirements between the entries of the attack vector. As a result, the set of covariance matrices given by (31), with $k \leqslant m$, that arises from considering Gaussian attacks is the set

$$\tilde{\mathcal{S}}_k \triangleq \bigcup_{\mathcal{K}} \left\{ \mathbf{S} \in S_+^m \colon \mathbf{S} = \sum_{i \in \mathcal{K}} v_i \mathbf{e}_i \mathbf{e}_i^{\mathsf{T}} \text{ with } v_i \in \mathbb{R}_+ \right\}, \tag{34}$$

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 (30) boils down to the following problem:

$$\min_{\mathbf{\Sigma}_{AA} \in \tilde{\mathcal{S}}_k} J(\mathbf{\Sigma}_{AA}), \tag{35}$$

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 $\mathrm{supp}(A^m)$ in (26) and determines the corresponding entry in the diagonal of the matrix $\mathbf{\Sigma}_{AA}$ in (5).

4.2 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 [19]. 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 $\mathrm{supp}(A^m)$ in (26).

More specifically, given the sparsity constraint in (31), for some $k \leqslant m$, the construction can be divided into k epochs. At each epoch a new element is added to $\mathrm{supp}(A^m)$. 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,

$$\mathcal{A}_i = \{j \in \{1, 2, \ldots, m\} : \mathbf{e}_j^\mathsf{T} \mathbf{\Sigma}_i \mathbf{e}_j > 0\}. \tag{36}$$

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,

$$\mathbf{\Sigma}_i = \mathbf{\Sigma}_{i-1} + v\mathbf{e}_j\mathbf{e}_j^\mathsf{T}, \tag{37}$$

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, that is, the entries of the vector $\text{diag}(\mathbf{\Sigma}_{i-1})$ that are zero, as

$$\mathcal{A}_{i-1}^\mathsf{c} \triangleq \{1, 2, \ldots, m\} \setminus \mathcal{A}_{i-1}. \tag{38}$$

Our proposition to choose both $j \in \mathcal{A}_{i-1}^\mathsf{c}$ and $v > 0$ at epoch i as described in (37) is based on the following optimization problem

$$\min_{(j,v) \in \mathcal{A}_{i-1}^\mathsf{c} \times \mathbb{R}_+} J(\mathbf{\Sigma}_{i-1} + v\mathbf{e}_j\mathbf{e}_j^\mathsf{T}). \tag{39}$$

The following lemma sheds light on the solution to the optimization problem in (39).

Lemma 4.1. *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 (29) satisfies that*

$$J(\mathbf{\Sigma}_2) = J(\mathbf{\Sigma}_1) + f(\mathbf{\Sigma}_1, \mathbf{\Delta}), \tag{40}$$

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

$$\begin{aligned} f(\mathbf{\Sigma}_1, \mathbf{\Delta}) = &(1 - \lambda) \log \left| \mathbf{I}_m + \left(\mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_1\right)^{-1} \mathbf{\Delta} \right| - \log \left| \mathbf{I}_m + \left(\sigma^2 \mathbf{I}_m + \mathbf{\Sigma}_1\right)^{-1} \mathbf{\Delta} \right| \\ &+ \lambda \text{tr}\left(\mathbf{\Sigma}_{YY}^{-1} \mathbf{\Delta}\right), \end{aligned} \tag{41}$$

where the matrix $\mathbf{\Sigma}_{YY}$ is defined by (4); and $\lambda \geq 1$ is introduced in (23).

Proof. The proof of Lemma 4.1 is presented in Appendix C. □

The relevance of Lemma 4.1 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 (39). Indeed, the selection problem results in

$$\min_{(j,v) \in \mathcal{A}_{i-1}^\mathsf{c} \times \mathbb{R}_+} f(\mathbf{\Sigma}_{i-1}, v\mathbf{e}_j\mathbf{e}_j^\mathsf{T}), \tag{42}$$

where the function f is defined in (41). Theorem 4.1 provides the solution to the optimization problem in (42).

Theorem 4.1. *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* (39)*. Then, the following holds*

$$j^\star = \underset{j \in \mathcal{A}_{i-1}^{\mathsf{c}}}{\operatorname{argmin}} J(\mathbf{\Sigma}_{i-1} + v_j \mathbf{e}_j \mathbf{e}_j^{\mathsf{T}}) \quad \textit{and} \tag{43}$$

$$v^\star = v_{j^\star}, \tag{44}$$

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

$$v_{j^*} = \left(\frac{\beta_j - \alpha_j + \beta_j \alpha_j \sigma^2}{2 \beta_j \alpha_j} \right) \left(\sqrt{1 - \frac{4 \beta_j \alpha_j \left(\beta_j \sigma^2 - \alpha_j \sigma^2 - \frac{\alpha_j \sigma^2 + 1}{\lambda} \right)}{\left(\beta_j - \alpha_j + \beta_j \alpha_j \sigma^2 \right)^2}} - 1 \right),$$

with

$$\alpha_j \stackrel{\Delta}{=} \operatorname{tr} \left(\left(\mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_{i-1} \right)^{-1} \mathbf{e}_{j^*} \mathbf{e}_{j^*}^{\mathsf{T}} \right), \tag{45}$$

$$\beta_j \stackrel{\Delta}{=} \operatorname{tr} \left(\mathbf{\Sigma}_{YY}^{-1} \mathbf{e}_{j^*} \mathbf{e}_{j^*}^{\mathsf{T}} \right), \tag{46}$$

and the real $\sigma > 0$ in (45) *is introduced in* (2)*.*

Proof. The proof of Theorem 4.1 is presented in Appendix D. □

The proposed greedy construction is described in Algorithm 1.

Algorithm 1 k-sparse independent attack construction

Require: $\mathbf{H}$ in (1);
σ^2 in (2);
$\mathbf{\Sigma}_{XX}$ in (3);
λ in (28); and
k in (31).
Ensure: $\mathbf{\Sigma}_{AA}$ 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 (45)
  end for
  Compute $j^\star$ in (43)
  Compute $v^\star$ in (44)
  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
$\mathbf{\Sigma}_{AA} = \sum_{i \in \mathcal{A}_k} v_i \mathbf{e}_i \mathbf{e}_i^{\mathsf{T}}$

5 Correlated Sparse Stealth Attacks

5.1 Correlation Structure

In this section, the assumption of independence in (33) is dropped. This case boils down to the attack construction given in (32), that is, 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 4.1. More precisely, the difference introduced by selecting the index i is given by $\Delta_i \in \mathcal{D}_i$ with

$$\mathcal{D}_i = \bigcup_{\mathbf{s}\in\mathbb{R}^m} \left\{ \mathbf{D} \in \mathbb{R}^{m\times m} : \mathbf{D} = \mathbf{s}^\mathsf{T} \otimes \mathbf{e}_i + \mathbf{s} \otimes \mathbf{e}_i^\mathsf{T}, \right\}. \tag{47}$$

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:

$$\begin{aligned} \min_{j,\mathbf{\Delta}} \ & J\left(\mathbf{\Sigma}_{i-1} + \mathbf{\Delta}\right) \\ \text{s.t.} \ & j \in \mathcal{A}_{i-1}^\mathsf{c}, \\ & \mathbf{\Delta} \in \mathcal{D}_j, \\ & \mathbf{\Sigma}_{i-1} + \mathbf{\Delta} \in S_+^m. \end{aligned} \tag{48}$$

In the following we show that when the choice of the next index selected for attacks is fixed, the optimization in (48) is convex in the matrix difference.

Theorem 5.1. *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*

$$\begin{aligned} \min_{\mathbf{\Delta}} \ & J\left(\mathbf{\Sigma}_{i-1} + \mathbf{\Delta}\right) \\ \text{s.t.} \ & \mathbf{\Delta} \in \mathcal{D}_j, \\ & \mathbf{\Sigma}_{i-1} + \mathbf{\Delta} \in S_+^m, \end{aligned} \tag{49}$$

is a convex optimization problem.

Proof. The proof of Theorem 5.1 is presented in Appendix E. □

5.2 Greedy Correlated Attack Construction

The proposed greedy construction for correlated attack case is described in Algorithm 2. Note that the matrix obtained in the optimization problem in

Theorem 5.1 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

Require: $\mathbf{H}$ in (1);
σ^2 in (2);
$\mathbf{\Sigma}_{XX}$ in (3);
λ in (28); and
k in (31).
Ensure: $\mathbf{\Sigma}_{AA}$ in (5).
1: Set $\mathcal{A}_0 = \{\emptyset\}$
2: Set $\mathbf{\Sigma}_0 = \mathbf{0}$
3: **for** $j = 1$ to k **do**
4: **for** $\ell \in \mathcal{A}_{j-1}^{\mathsf{c}}$ **do**
5: Compute $\mathbf{\Delta}_\ell = \underset{\mathbf{\Delta} \in \mathcal{D}_\ell}{\operatorname{argmin}} J(\mathbf{\Sigma}_{j-1} + \mathbf{\Delta})$
6: **end for**
7: Compute $j^\star = \underset{\ell \in \mathcal{A}_{j-1}^{\mathsf{c}}}{\operatorname{argmin}} J(\mathbf{\Sigma}_{j-1} + \mathbf{\Delta}_\ell)$
8: Set $\mathcal{A}_j = \mathcal{A}_{j-1} \cup \{j^\star\}$
9: Set $\mathbf{\Sigma}_j = \mathbf{\Sigma}_{j-1} + \mathbf{\Delta}_{j^\star}$
10: **end for**
11: Compute $\mathbf{\Sigma}_{AA} = \underset{\mathbf{S} \in \mathcal{S}_+^m}{\operatorname{argmin}} \|\mathbf{\Sigma}_k - \mathbf{S}\|_{\mathsf{F}}$

6 Numerical Results

In this section, we first 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 [29]. 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 [30] 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 [13, 17, 31]. 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

$$\text{SNR} \triangleq 10 \log_{10} \left(\frac{\text{tr}(\mathbf{H}\boldsymbol{\Sigma}_{\text{XX}}\mathbf{H}^{\mathsf{T}})}{m\sigma^2} \right). \tag{50}$$

Please note that the attack construction can be generalized to a linearized AC model at a certain nominal operation point. The simulation of the linearized power flow model is provided to verify the generality of the proposed attacks. Let $\mathbf{x}_0$ be the state variables of the nominal operation point when the system is operating under optimal power flow. We use MATPOWER [28] to obtain the optimal power flow where the nominal operation point lies on. The corresponding Jacobian matrix is

$$\mathbf{H}_0 = \frac{\partial}{\partial \mathbf{x}} \mathbf{h}(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_0}, \tag{51}$$

where $\mathbf{h}(\mathbf{x}) \in \mathbb{R}^m$ denotes the vector of random variables induced by the non-linear relation between the state variables and the measurements and $\mathbf{H}_0$ is the corresponding Jacobian matrix in linearized AC model when system is operating under optimal power flow.

6.1 Performance in terms of information theoretic cost

Let $\boldsymbol{\Sigma}_i^k$ 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

$$\eta \triangleq \frac{J(\boldsymbol{\Sigma}_i^k) - J(\boldsymbol{\Sigma}_i^m)}{J(\boldsymbol{\Sigma}_i^m)}, \tag{52}$$

where $J(\cdot)$ is the cost defined in (29). Note that $J(\boldsymbol{\Sigma}_i^m)$ 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 in DC model obtained with Algorithm 1 in different IEEE test systems as a function of the proportion of compromised sensors, that is, k/m, for correlation parameter $\rho = 0.9$ and $\lambda = 8$. Similarly, Fig. 2 depicts the performance of the correlated sparse stealth attack construction in DC model 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.

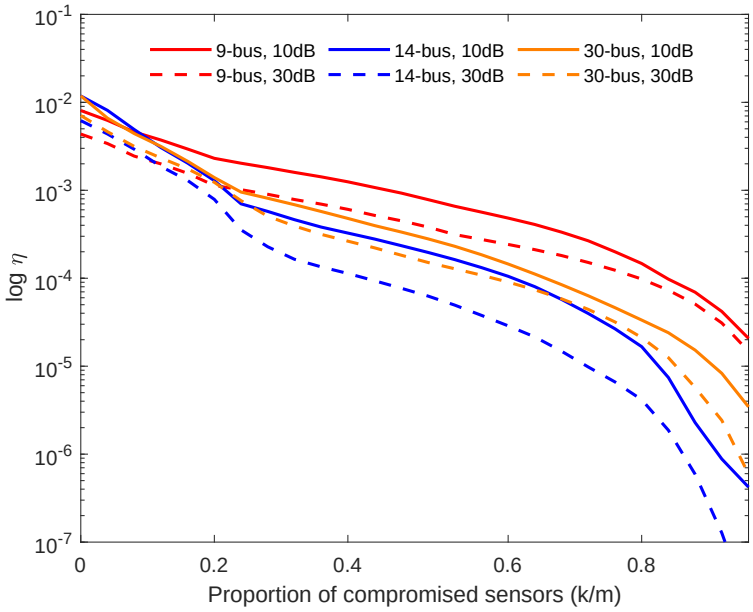


Figure 1: Performance of independent attack constructions in DC model on different IEEE test systems with $\rho = 0.9$ and $\lambda = 8$.

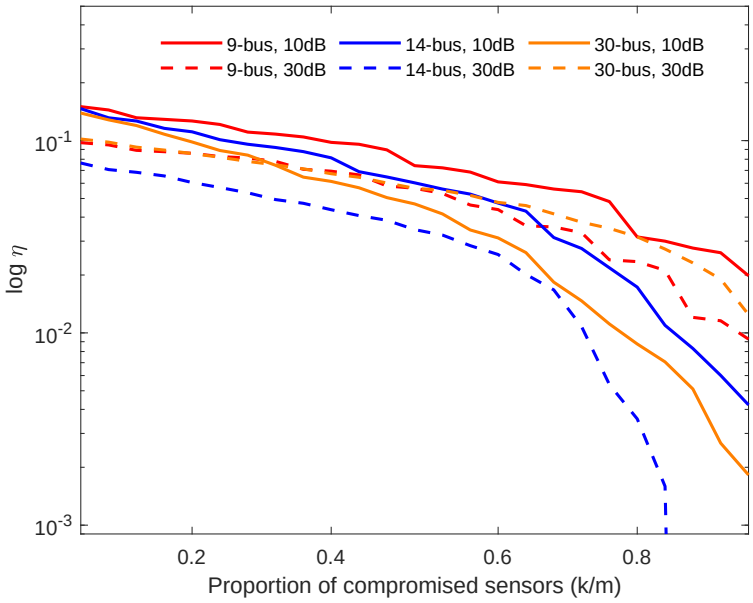


Figure 2: Performance of correlated attack constructions in DC model on different IEEE test systems with $\rho = 0.9$ and $\lambda = 8$.

Fig. 3 and Fig. 4 depict the performance of the independent sparse stealth attack construction and correlated sparse stealth attack construction in linearized AC model from Algorithm 1 and Algorithm 2, respectively, in different IEEE test systems as a function of the proportion of compromised measurements, that is, k/m, for parameters $\rho = 0.9$ and $\lambda = 8$.

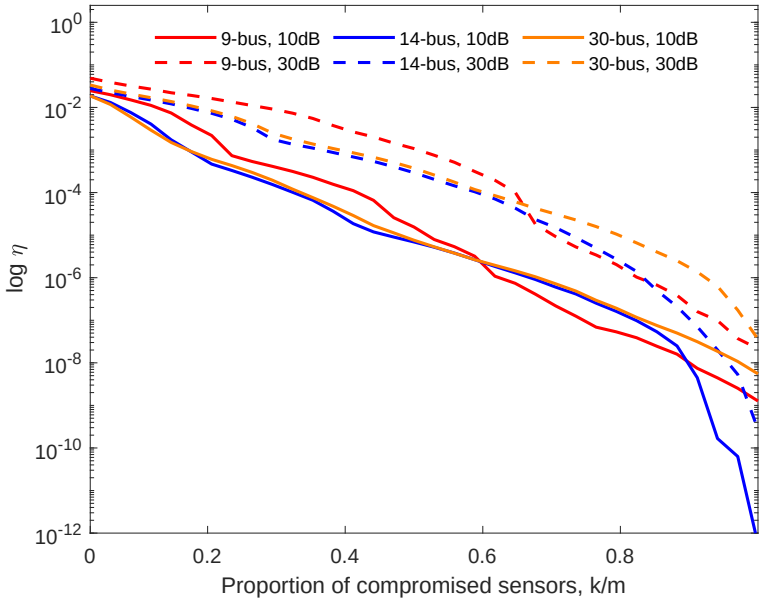


Figure 3: Performance of independent attack constructions in linearized AC model on different IEEE test systems with $\rho = 0.9$ and $\lambda = 8$.

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

Fig. 5 and Fig. 6 depict the multiobjective performance of the Algorithm 1 attack construction in DC model 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. 7 and Fig. 8 depict the same setting in DC model for the Algorithm 2 attack construction. As expected, larger values of the parameter λ yield smaller values of KL divergence, that is, 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.

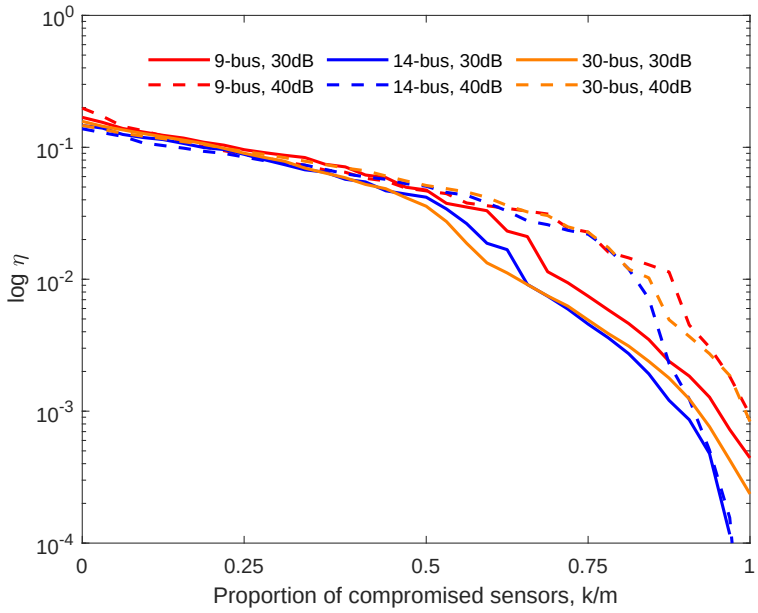


Figure 4: Performance of correlated attack constructions in linearized AC model on different IEEE test systems with $\rho = 0.9$ and $\lambda = 8$.

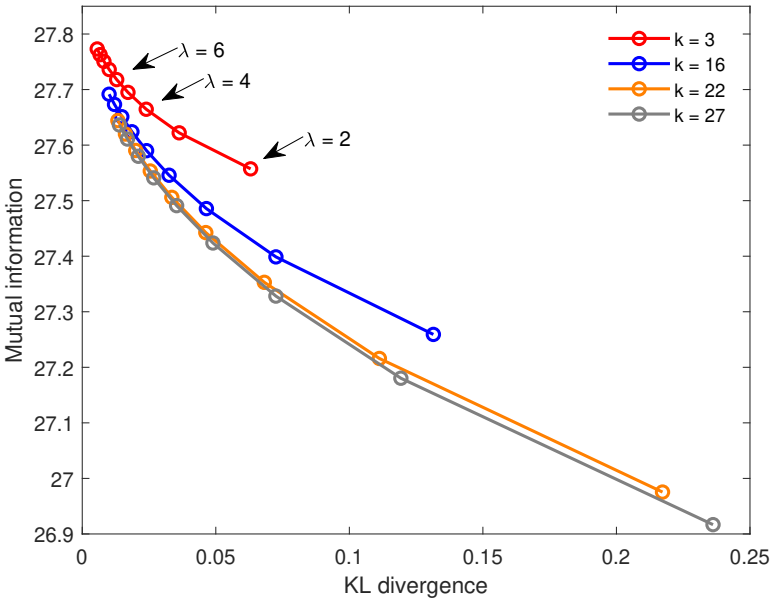


Figure 5: Performance of independent sparse attack construction in DC model in terms of mutual information and KL divergence for different values of λ on the IEEE 9-bus system with SNR = 30 dB and $\rho = 0.9$.

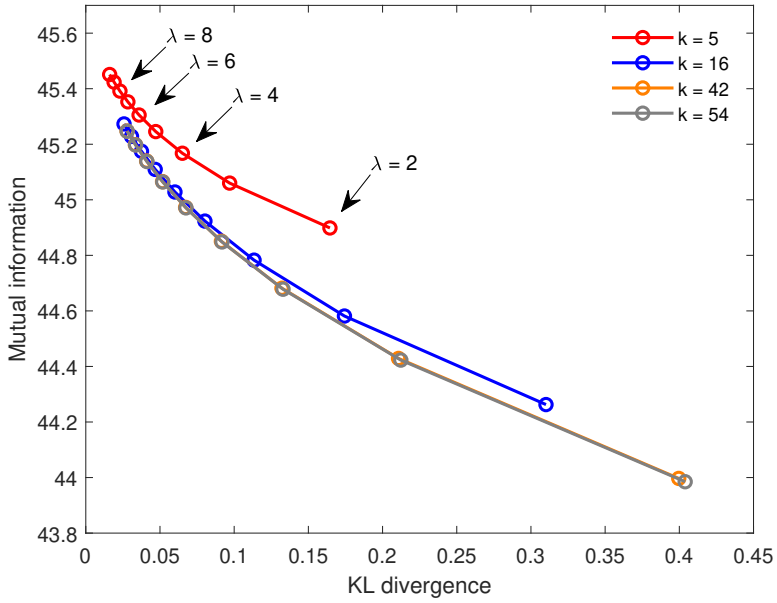


Figure 6: Performance of independent sparse attack construction in DC model in terms of mutual information and KL divergence for different values of λ on the IEEE 14-bus system with SNR = 30 dB and $\rho = 0.9$.

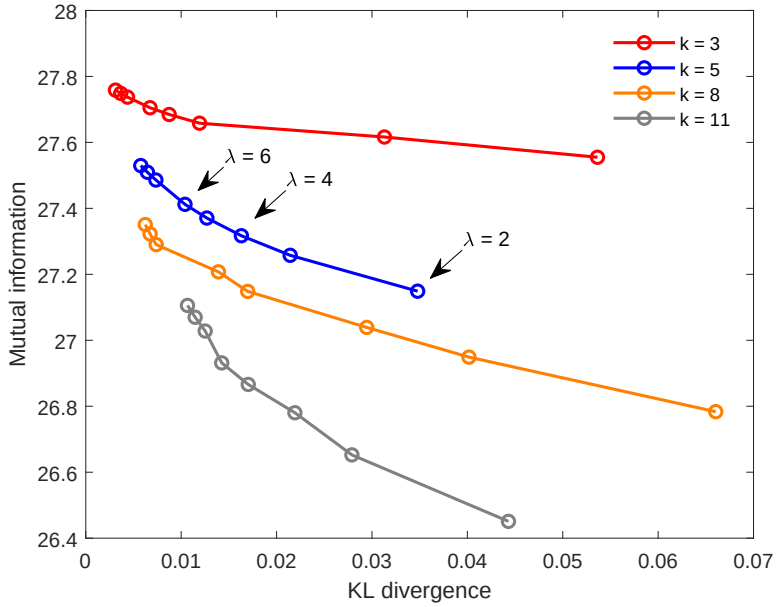


Figure 7: Performance of correlated sparse attack construction in DC model in terms of mutual information and KL divergence for different values of λ on the IEEE 9-bus system with SNR = 30 dB and $\rho = 0.9$.

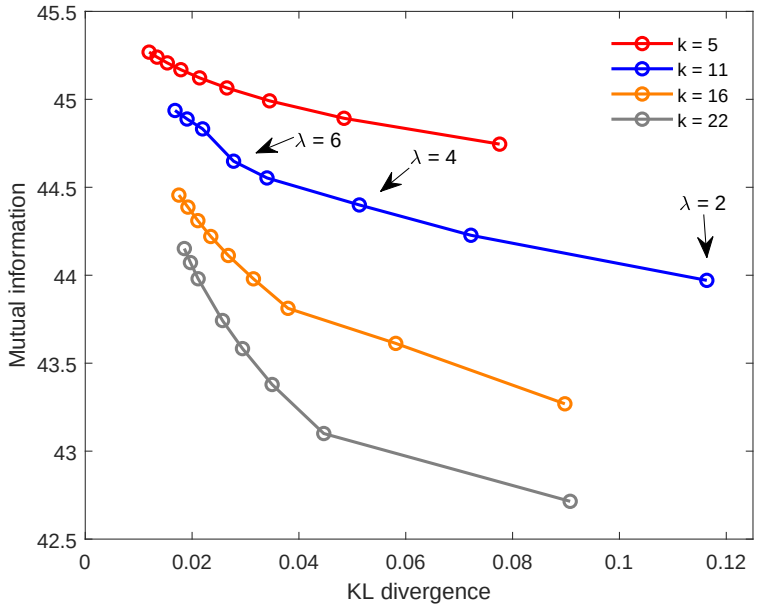


Figure 8: Performance of correlated sparse attack construction in DC model in terms of mutual information and KL divergence for different values of λ on the IEEE 14-bus system with SNR $=$ 30 dB and $\rho = 0.9$.

Fig. 9 and Fig. 10 depict the multiobjective performance of the Algorithm 1 attack construction in linearized AC model 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. 11 and Fig. 12 depict the same setting in linearized AC model for the Algorithm 2 attack construction.

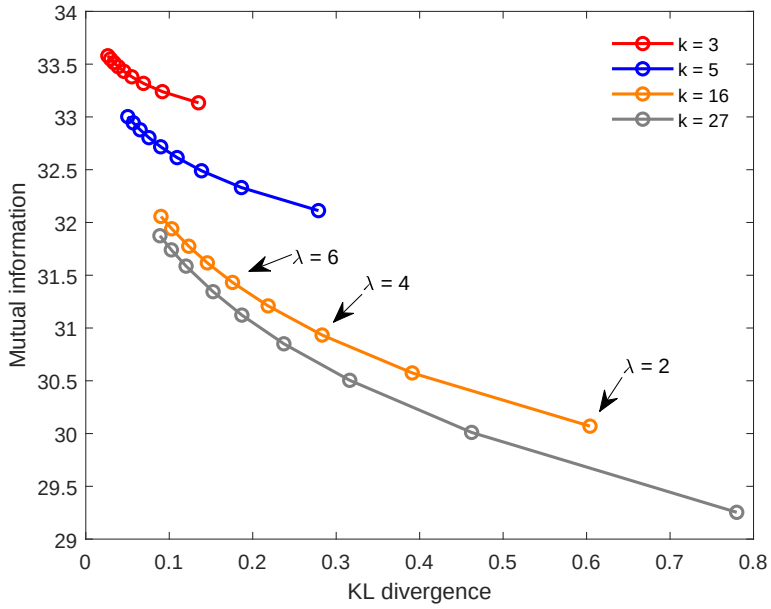


Figure 9: Performance of independent sparse attack construction in linearized AC mdoel in terms of mutual information and KL divergence for different values of λ on the IEEE 9-bus system with SNR = 30dB and $\rho = 0.9$.

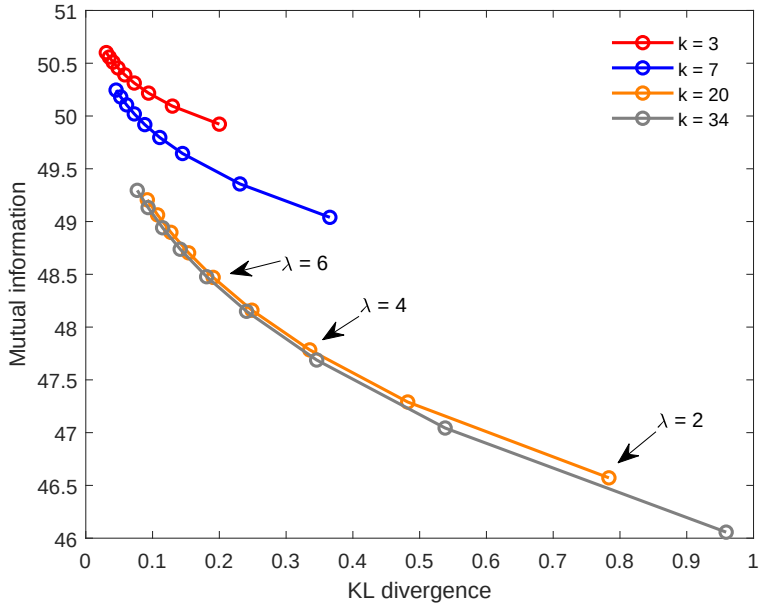


Figure 10: Performance of independent sparse attack construction in linearized AC mdoel in terms of mutual information and KL divergence for different values of λ on the IEEE 14-bus system with SNR = 30dB and $\rho = 0.9$.

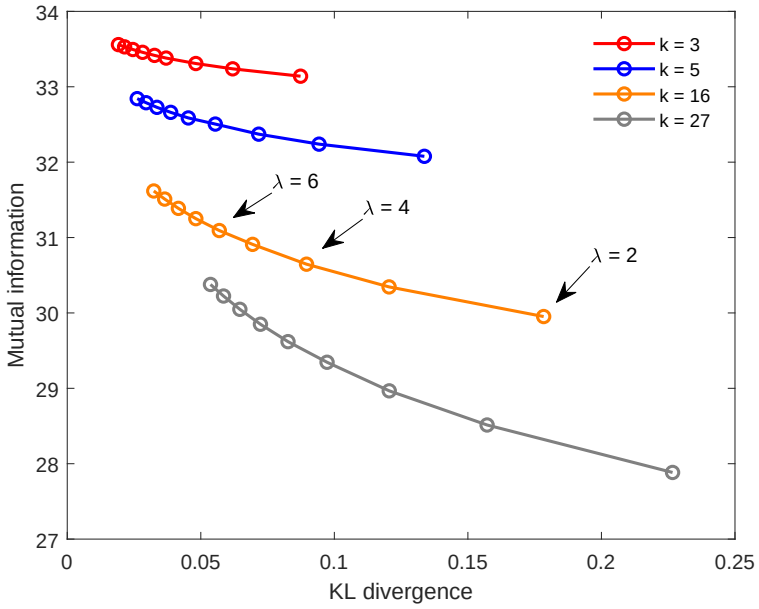


Figure 11: Performance of correlated sparse attack construction in linearized AC mdoel in terms of mutual information and KL divergence for different values of λ on the IEEE 9-bus system with SNR = 30dB and $\rho = 0.9$.

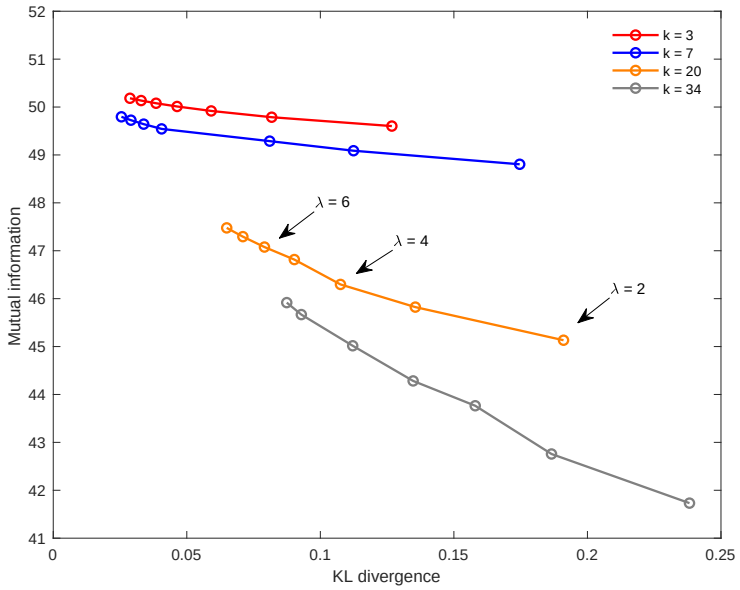


Figure 12: Performance of correlated sparse attack construction in linearized AC mdoel in terms of mutual information and KL divergence for different values of λ on the IEEE 14-bus system with SNR = 30dB and $\rho = 0.9$.

6.3 Performance in terms of mutual information and probability of attack detection

Fig. 13 and Fig. 14 depict the performance of the attack construction in DC model for different values of λ and sparse constraint k with SNR $= 30$ dB, $\rho = 0.9$ and $\tau = 2$ for the IEEE 9-bus and the IEEE 14-bus test systems, respectively. As expected, larger values of the parameter λ yield smaller values of the probability of attack detection while increasing the mutual information between the vector of state variables and the vector of observations in the systems. We note that the probability of attack detection decreases approximately linearly with respect to $\log \lambda$ for small values of λ. Simultaneously for this range of λ, mutual information increases approximately linearly with respect to $\log \lambda$. For moderate values of λ, we observe a significant decrease in the probability of detection with respect to $\log \lambda$ with a smaller rate of increase in mutual information. The comparison between independent and correlated attack constructions, shows that for the same sparsity constraint, the correlated attack construction successfully exploits the coordination between different locations to yield a smaller probability of detection and a smaller mutual information.

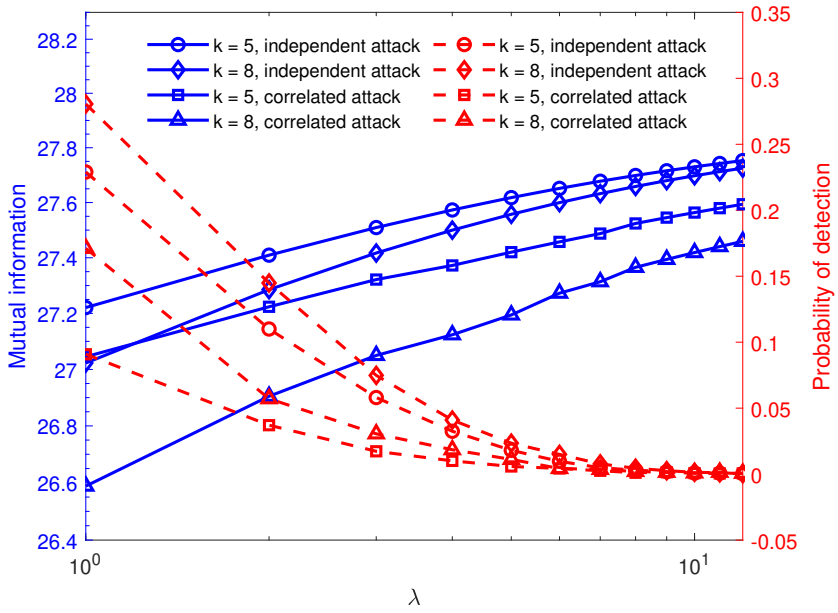


Figure 13: Performance of attack constructions in DC model on IEEE 9-bus test system with $\rho = 0.9$, SNR $= 30$dB and $\tau = 2$.

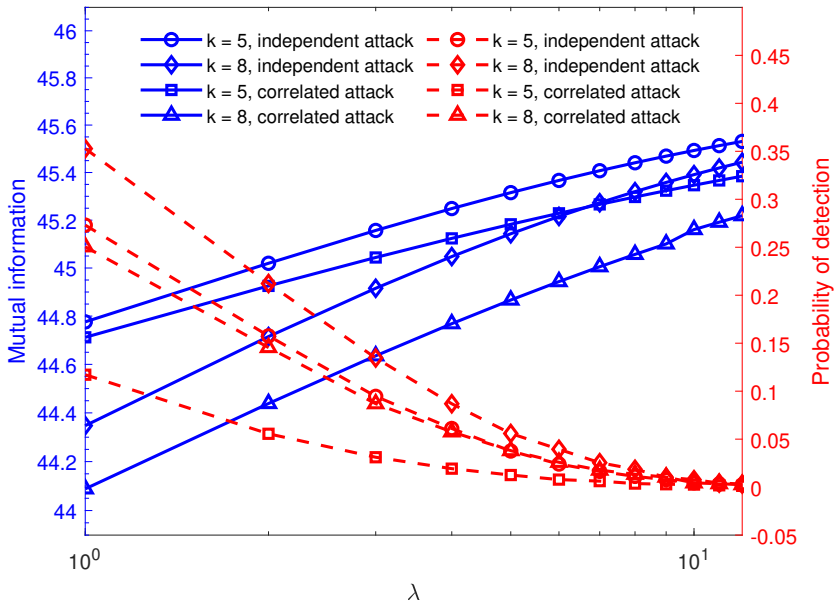


Figure 14: Performance of attack constructions in DC model on IEEE 14-bus test system with $\rho = 0.9$, SNR $= 30$dB and $\tau = 2$.

7 Conclusions

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 characterised the dependence.

Appendices

A Proof of Proposition 3.1.1

Proof. Let $W^{n+m} \triangleq (X^n; Y_A^m)$. It follows that $W^{n+m} \sim \mathcal{N}(\boldsymbol{\mu}_W, \boldsymbol{\Sigma})$ such that

$$\boldsymbol{\mu}_W \triangleq \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\mu}_A \end{pmatrix}, \tag{53}$$

$$\boldsymbol{\Sigma} \triangleq \begin{pmatrix} \boldsymbol{\Sigma}_{XX} & \boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T} \\ \mathbf{H}\boldsymbol{\Sigma}_{XX} & \mathbf{H}\boldsymbol{\Sigma}_{XX}\mathbf{H}^\mathsf{T} + \sigma^2\mathbf{I}_m + \boldsymbol{\Sigma}_{AA} \end{pmatrix}. \tag{54}$$

Note that

$$I(X^n; Y_A^m) \triangleq \mathbb{E}_{X^n;Y_A^m}\left[\log \frac{f_{X^n;Y_A^m}}{f_{X^n} f_{Y_A^m}}\right] \tag{55}$$

$$= \mathbb{E}_{W^{n+m}}\left[\log \frac{f_{W^{n+m}}}{f_{X^n} f_{Y_A^m}}\right], \tag{56}$$

where the functions $f_{X^n;Y_A^m}$, f_{X^n} and $f_{Y_A^m}$ in (55) are the probability density functions of $(X^n; Y_A^m)$, X^n and Y_A^m, respectively; the function $f_{W^{n+m}}$ in (56) is

the pdf of W^{n+m} and $f_{X^n;Y_A^m} = f_{W^{n+m}}$. It follows that

$$
\begin{aligned}
& I(X^n; Y_A^m) && (57)\\
=& \mathbb{E}_{W^{n+m}}\left[\log \frac{\exp(-\frac{1}{2}(W^{n+m}-\boldsymbol{\mu}_W)^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(W^{n+m}-\boldsymbol{\mu}_W))}{(2\pi)^{\frac{n+m}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\right] && (58)\\
& -\mathbb{E}_{X^n}\left[\log \frac{\exp(-\frac{1}{2}(X^n)^{\mathsf{T}}\boldsymbol{\Sigma}_{XX}^{-1}X^n)}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}_{XX}|^{\frac{1}{2}}}\right]\\
& -\mathbb{E}_{Y_A^m}\left[\log \frac{\exp(-\frac{1}{2}(Y_A^m-\boldsymbol{\mu}_A)^{\mathsf{T}}\boldsymbol{\Sigma}_{Y_AY_A}^{-1}(Y_A^m-\boldsymbol{\mu}_A))}{(2\pi)^{\frac{m}{2}}|\boldsymbol{\Sigma}_{Y_AY_A}|^{\frac{1}{2}}}\right]\\
=& -\frac{1}{2}\mathbb{E}_{W^{n+m}}\left[\left(W^{n+m}-\boldsymbol{\mu}_W\right)^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\left(W^{n+m}-\boldsymbol{\mu}_W\right)\right] - \frac{1}{2}\log(2\pi)^{n+m}|\boldsymbol{\Sigma}| && (59)\\
& +\frac{1}{2}\mathbb{E}_{X^n}\left[(X^n)^{\mathsf{T}}\boldsymbol{\Sigma}_{XX}^{-1}X^n\right] + \frac{1}{2}\log(2\pi)^n|\boldsymbol{\Sigma}_{XX}|\\
& +\frac{1}{2}\mathbb{E}_{Y_A^m}\left[(Y_A^m-\boldsymbol{\mu}_A)^{\mathsf{T}}\boldsymbol{\Sigma}_{Y_AY_A}^{-1}(Y_A^m-\boldsymbol{\mu}_A)\right] + \frac{1}{2}\log(2\pi)^m|\boldsymbol{\Sigma}_{Y_AY_A}|\\
=& -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbb{E}_{W^{n+m}}\left[\left(W^{n+m}-\boldsymbol{\mu}_W\right)\left(W^{n+m}-\boldsymbol{\mu}_W\right)^{\mathsf{T}}\right]\right)\\
& +\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}_{Y_AY_A}^{-1}\mathbb{E}_{Y_A^m}\left[\left(Y_A^m-\boldsymbol{\mu}_A\right)\left(Y_A^m-\boldsymbol{\mu}_A\right)^{\mathsf{T}}\right]\right) + \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_{XX}||\boldsymbol{\Sigma}_{Y_AY_A}|}{|\boldsymbol{\Sigma}|} && (60)\\
=& -\frac{1}{2}(n+m) + \frac{1}{2}n + \frac{1}{2}m + \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_{XX}||\boldsymbol{\Sigma}_{Y_AY_A}|}{|\boldsymbol{\Sigma}|} && (61)\\
=& \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_{XX}||\boldsymbol{\Sigma}_{Y_AY_A}|}{|\boldsymbol{\Sigma}|}, && (62)
\end{aligned}
$$

where the equality in (58) follows from taking the expression of the probability density functions of multivariate Gaussian distributions, that is, f_{X^n}, $f_{Y_A^m}$ and $f_{W^{n+m}}$ into (56); the equality in (59) follows from taking constants out of the expectation; the equality in (60) follows from the fact that

$$
\begin{aligned}
& \mathbb{E}_{W^{n+m}}\left[\left(W^{n+m}-\boldsymbol{\mu}_W\right)^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\left(W^{n+m}-\boldsymbol{\mu}_W\right)\right] && (63)\\
=& \mathbb{E}_{W^{n+m}}\left[\mathrm{tr}\left(\left(W^{n+m}-\boldsymbol{\mu}_W\right)^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\left(W^{n+m}-\boldsymbol{\mu}_W\right)\right)\right]\\
=& \mathbb{E}_{W^{n+m}}\left[\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\left(W^{n+m}-\boldsymbol{\mu}_W\right)\left(W^{n+m}-\boldsymbol{\mu}_W\right)^{\mathsf{T}}\right)\right]\\
=& \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbb{E}_{W^{n+m}}\left[\left(W^{n+m}-\boldsymbol{\mu}_W\right)\left(W^{n+m}-\boldsymbol{\mu}_W\right)^{\mathsf{T}}\right]\right),
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}_{X^n}\left[(X^n)^{\mathsf{T}}\boldsymbol{\Sigma}_{XX}^{-1}X^n\right] =& \mathrm{tr}\left(\boldsymbol{\Sigma}_{XX}^{-1}\mathbb{E}_{X^n}\left[X^n(X^n)^{\mathsf{T}}\right]\right), && (64)\\
\mathbb{E}_{Y_A^m}\left[(Y_A^m-\boldsymbol{\mu}_A)^{\mathsf{T}}\boldsymbol{\Sigma}_{Y_AY_A}^{-1}(Y_A^m-\boldsymbol{\mu}_A)\right] =& \mathrm{tr}\left(\boldsymbol{\Sigma}_{Y_AY_A}^{-1}\mathbb{E}_{Y_A^m}\left[(Y_A^m-\boldsymbol{\mu}_A)(Y_A^m-\boldsymbol{\mu}_A)^{\mathsf{T}}\right]\right). && (65)
\end{aligned}
$$

the equality in (61) follows from the fact that

$$\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbb{E}_{W^{n+m}}\left[\left(W^{n+m}-\boldsymbol{\mu}_W\right)\left(W^{n+m}-\boldsymbol{\mu}_W\right)^{\mathsf{T}}\right]\right)=\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\right)=n+m, \tag{66}$$

$$\operatorname{tr}\left(\boldsymbol{\Sigma}_{XX}^{-1}\mathbb{E}_{X^n}\left[X^n\left(X^n\right)^{\mathsf{T}}\right]\right)=\operatorname{tr}\left(\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XX}\right)=n, \tag{67}$$

$$\operatorname{tr}\left(\boldsymbol{\Sigma}_{Y_AY_A}^{-1}\mathbb{E}_{Y_A^m}\left[\left(Y_A^m-\boldsymbol{\mu}_A\right)\left(Y_A^m-\boldsymbol{\mu}_A\right)^{\mathsf{T}}\right]\right)=\operatorname{tr}\left(\boldsymbol{\Sigma}_{Y_AY_A}^{-1}\boldsymbol{\Sigma}_{Y_AY_A}\right)=m \tag{68}$$

This completes the proof. □

B Proof of Proposition 3.1.2

Proof. Note that

$$D(P_{Y_A^m}\|P_{Y^m}) \tag{69}$$

$$\triangleq\mathbb{E}_{Y_A^m}\left[\log\frac{f_{Y_A^m}}{f_{Y^m}}\right], \tag{70}$$

$$=\mathbb{E}_{Y_A^m}\left[\log\frac{\exp(-\frac{1}{2}(Y_A^m-\boldsymbol{\mu}_A)^{\mathsf{T}}\boldsymbol{\Sigma}_{Y_AY_A}^{-1}(Y_A^m-\boldsymbol{\mu}_A))}{(2\pi)^{\frac{m}{2}}|\boldsymbol{\Sigma}_{Y_AY_A}|^{\frac{1}{2}}}-\log\frac{\exp(-\frac{1}{2}(Y_A^m)^{\mathsf{T}}\boldsymbol{\Sigma}_{YY}^{-1}Y_A^m)}{(2\pi)^{\frac{m}{2}}|\boldsymbol{\Sigma}_{YY}|^{\frac{1}{2}}}\right] \tag{71}$$

$$=\frac{1}{2}\mathbb{E}_{Y_A^m}\left[\log\frac{\exp(-(Y_A^m-\boldsymbol{\mu}_A)^{\mathsf{T}}\boldsymbol{\Sigma}_{Y_AY_A}^{-1}(Y_A^m-\boldsymbol{\mu}_A))}{|\boldsymbol{\Sigma}_{Y_AY_A}|}-\log\frac{\exp(-(Y_A^m)^{\mathsf{T}}\boldsymbol{\Sigma}_{YY}^{-1}Y_A^m)}{|\boldsymbol{\Sigma}_{YY}|}\right] \tag{72}$$

$$=\frac{1}{2}\mathbb{E}_{Y_A^m}\left[-(Y_A^m-\boldsymbol{\mu}_A)^{\mathsf{T}}\boldsymbol{\Sigma}_{Y_AY_A}^{-1}(Y_A^m-\boldsymbol{\mu}_A)+(Y_A^m)^{\mathsf{T}}\boldsymbol{\Sigma}_{YY}^{-1}Y_A^m\right]+\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{Y_AY_A}|} \tag{73}$$

$$=-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_{Y_AY_A}^{-1}\mathbb{E}_{Y_A^m}\left[\left(Y_A^m-\boldsymbol{\mu}_A\right)\left(Y_A^m-\boldsymbol{\mu}_A\right)^{\mathsf{T}}\right]\right)+\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\mathbb{E}_{Y_A^m}\left[Y_A^m(Y_A^m)^{\mathsf{T}}\right]\right) \tag{74}$$
$$+\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{Y_AY_A}|}$$

$$=-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_{Y_AY_A}^{-1}\boldsymbol{\Sigma}_{Y_AY_A}\right)+\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}(\boldsymbol{\Sigma}_{Y_AY_A}+\boldsymbol{\mu}_A\boldsymbol{\mu}_A^{\mathsf{T}})\right)+\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{Y_AY_A}|} \tag{75}$$

$$=-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_{Y_AY_A}^{-1}\boldsymbol{\Sigma}_{Y_AY_A}\right)+\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{Y_AY_A}+\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\mu}_A\boldsymbol{\mu}_A^{\mathsf{T}}\right)+\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{Y_AY_A}|} \tag{76}$$

$$=\frac{1}{2}\left(\log\frac{|\boldsymbol{\Sigma}_{YY}|}{|\boldsymbol{\Sigma}_{Y_AY_A}|}-m+\operatorname{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{Y_AY_A}\right)+\operatorname{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\mu}_A\boldsymbol{\mu}_A^{\mathsf{T}}\right)\right) \tag{77}$$

where the functions $f_{Y_A^m}$ and f_{Y^m} in (70) are the probability density functions of Y_A^m and Y^m, respectively; the equality in (71) follows from taking the probability density functions of Y_A^m and Y^m into the definition of KL divergence; the equality in (73) follows from taking constants out of the expectation; the equality in (75) follows from (65) and the fact that

$$\mathbb{E}_{Y_A^m}\left[(Y_A^m)^{\mathsf{T}}\boldsymbol{\Sigma}_{YY}^{-1}Y_A^m\right]=\operatorname{tr}\left(\boldsymbol{\Sigma}_{YY}^{-1}\mathbb{E}_{Y_A^m}\left[(Y_A^m)^{\mathsf{T}}Y_A^m\right]\right); \tag{78}$$

the equality in (75) follows from the fact that

$$\begin{aligned}\mathbb{E}_{Y_A^m}\left[Y_A^m(Y_A^m)^{\mathsf{T}}\right]=&\mathbb{E}_{Y_A^m}\left[\left(Y_A^m-\boldsymbol{\mu}_A\right)\left(Y_A^m-\boldsymbol{\mu}_A\right)^{\mathsf{T}}\right]+\boldsymbol{\mu}_A\boldsymbol{\mu}_A^{\mathsf{T}} \\ =&\boldsymbol{\Sigma}_{Y_AY_A}+\boldsymbol{\mu}_A\boldsymbol{\mu}_A^{\mathsf{T}}.\end{aligned} \tag{79}$$

This completes the proof. □

C Proof of Lemma 4.1

Proof. 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 $\mathbf{\Delta} \in \mathbb{R}^{m\times m}$. Taking $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ into (29), then the cost difference between $J(\mathbf{\Sigma}_2)$ and $J(\mathbf{\Sigma}_1)$ is given by

$$
\begin{aligned}
&J(\mathbf{\Sigma}_2) - J(\mathbf{\Sigma}_1) && (80)\\
=&(1-\lambda)\log|\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_2| - \log|\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_2| + \lambda\mathrm{tr}(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_2)\\
&-\left((1-\lambda)\log|\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_1| - \log|\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_1| + \lambda\mathrm{tr}(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_1)\right) && (81)\\
=&(1-\lambda)\log\frac{|\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_1+\mathbf{\Delta}|}{|\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_1|} - \log\frac{\left|\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_1+\mathbf{\Delta}\right|}{|\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_1|} + \lambda\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\left(\mathbf{\Sigma}_2-\mathbf{\Sigma}_1\right)\right) && (82)\\
=&(1-\lambda)\log\left|\mathbf{I}_m+\left(\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_1\right)^{-1}\mathbf{\Delta}\right| - \log\left|\mathbf{I}_m+\left(\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_1\right)^{-1}\mathbf{\Delta}\right| + \lambda\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Delta}\right), && (83)
\end{aligned}
$$

where the equality in (81) follows from taking $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ into (29) and the equality in (82) follows from replacing $\mathbf{\Sigma}_2$ with $\mathbf{\Sigma}_1+\mathbf{\Delta}$ and the equality in (83) follows from eliminating $|\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_1|$ and $|\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_1|$. From the equality in (83), the cost difference can be written as a function of $\mathbf{\Sigma}_1$ and $\mathbf{\Delta}$, that is, $J(\mathbf{\Sigma}_2) - J(\mathbf{\Sigma}_1) = f(\mathbf{\Sigma}_1, \mathbf{\Delta})$, where $f(\mathbf{\Sigma}_1, \mathbf{\Delta}) \triangleq (1-\lambda)\log\left|\mathbf{I}_m+\left(\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_1\right)^{-1}\mathbf{\Delta}\right| - \log\left|\mathbf{I}_m+\left(\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_1\right)^{-1}\mathbf{\Delta}\right| + \lambda\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Delta}\right)$. This completes the proof. □

D Proof of Theorem 4.1

Proof. It follows from Lemma 4.1 that the optimization problem in (39) is equivalent to

$$
\min_{(j,v)\in\mathcal{A}_{i-1}^{\mathsf{c}}\times\mathbb{R}_+} J(\mathbf{\Sigma}_{i-1}) + f(\mathbf{\Sigma}_{i-1}, v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}), \tag{84}
$$

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

$$
\begin{aligned}
f(\mathbf{\Sigma}_{i-1}, v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}) =&(1-\lambda)\log\left|\mathbf{I}_m+\left(\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{i-1}\right)^{-1}v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right| - \log\left|\mathbf{I}_m+\left(\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_{i-1}\right)^{-1}v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right|\\
&+\lambda\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right). && (85)
\end{aligned}
$$

Note that $J(\mathbf{\Sigma}_{i-1})$ is a constant with respect to j and v. Hence, it holds that the optimization problem in (39) is equivalent to

$$
\begin{aligned}
&\min_{(j,v)\in\mathcal{A}_{i-1}^{\mathsf{c}}\times\mathbb{R}_+} f(\mathbf{\Sigma}_{i-1}, v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}) && (86)\\
=&\min_{(j,v)\in\mathcal{A}_{i-1}^{\mathsf{c}}\times\mathbb{R}_+} (1-\lambda)\log\left|\mathbf{I}_m+\left(\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{i-1}\right)^{-1}v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right| - \log\left|\mathbf{I}_m+\left(\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_{i-1}\right)^{-1}v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right|\\
&+\lambda\mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1}v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right). && (87)
\end{aligned}
$$

Note that $j\in\mathcal{A}_{i-1}^{\mathsf{c}}$. Therefore, it holds that

$$
\log\left|\mathbf{I}_m+\left(\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{i-1}\right)^{-1}v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right| = \log(1+\alpha_j v), \tag{88}
$$

where $\alpha_j \triangleq \mathrm{tr}\left(\left(\mathbf{\Sigma}_{YY}+\mathbf{\Sigma}_{i-1}\right)\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right)$. Note that $j \in \mathcal{A}_{i-1}^c$ and $\mathbf{\Sigma}_{i-1}$ is formed from previous i epochs. It yields that $\mathbf{\Sigma}_{i-1}$ is a diagonal matrix and

$$\log\left|\mathbf{I}_m+\left(\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_{i-1}\right)^{-1}v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right| = \log(1+\frac{v}{\sigma^2}). \tag{89}$$

It follows that the minimization problem in (86) can be rewritten as

$$\min_{(j,v)\in\mathcal{A}_{i-1}^c\times\mathbb{R}_+} f(\mathbf{\Sigma}_{i-1}, v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}) \tag{90}$$

$$= \min_{(j,v)\in\mathcal{A}_{i-1}^c\times\mathbb{R}_+} (1-\lambda)\log(1+\alpha_j v) - \log(1+\frac{v}{\sigma^2}) + \lambda\beta_j v, \tag{91}$$

where $\beta_j \triangleq \mathrm{tr}\left(\left(\sigma^2\mathbf{I}_m+\mathbf{\Sigma}_{i-1}\right)\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}\right)$. We break the optimization problem in (90) as follows:

$$\min_{(j,v)\in\mathcal{A}_{i-1}^c\times\mathbb{R}_+} f(\mathbf{\Sigma}_{i-1}, v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}) = \min_{j\in\mathcal{A}_{i-1}^c}\min_{v\in\mathbb{R}_+} (1-\lambda)\log(1+\alpha_j v) - \log(1+\frac{v}{\sigma^2}) + \lambda\beta_j v \tag{92}$$

Let $\lambda \geq 1$, then it holds that for all $j \in \mathcal{A}_{i-1}$, the cost function in (92) is convex in v. The only solution of the inner minimization problem in (92) is obtained by letting the first derivative equal to zero, that is,

$$\frac{\partial f(\mathbf{\Sigma}_{i-1}, v\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}})}{\partial v} = 0, \tag{93}$$

which is

$$\beta_j\alpha_j v^2 + (\beta_j - \alpha_j + \beta_j\alpha_j\sigma^2)v + \beta_j\sigma^2 - \alpha_j\sigma^2 - \frac{\alpha_j\sigma^2+1}{\lambda} = 0. \tag{94}$$

Note that equation (94) is quadratic with two solutions as follows:

$$\frac{1}{2\beta_j\alpha_j}\left(-(\beta_j-\alpha_j+\beta_j\alpha_j\sigma^2)+\sqrt{(\beta_j-\alpha_j+\beta_j\alpha_j\sigma^2)^2-4\beta_j\alpha_j(\beta_j\sigma^2-\alpha_j\sigma^2-\frac{\alpha_j\sigma^2+1}{\lambda})}\right), \tag{95}$$

and

$$\frac{1}{2\beta_j\alpha_j}\left(-(\beta_j-\alpha_j+\beta_j\alpha_j\sigma^2)-\sqrt{(\beta_j-\alpha_j+\beta_j\alpha_j\sigma^2)^2-4\beta_j\alpha_j(\beta_j\sigma^2-\alpha_j\sigma^2-\frac{\alpha_j\sigma^2+1}{\lambda})}\right). \tag{96}$$

The result follows by choosing the solution such that $v \in \mathbb{R}_+$, that is, the solution in (95). After some algebra manipulation, the solution is rewritten as

$$v_j^* = \frac{\beta_j-\alpha_j+\beta_j\alpha_j\sigma^2}{2\beta_j\alpha_j}\left(\sqrt{1-\frac{4\beta_j\alpha_j\left(\beta_j\sigma^2-\alpha_j\sigma^2-\frac{\alpha_j\sigma^2+1}{\lambda}\right)}{\left(\beta_j-\alpha_j+\beta_j\alpha_j\sigma^2\right)^2}}-1\right), \tag{97}$$

We now proceed to the outer minimization in (92). The solution j^* to the optimization problem in (92) is obtained by searching over all the possible candidate in $\mathcal{A}_{i-1}^c$, that is,

$$j^\star = \underset{j\in\mathcal{A}_{i-1}^c}{\mathrm{argmin}}\, J(\mathbf{\Sigma}_{i-1} + v_j\mathbf{e}_j\mathbf{e}_j^{\mathsf{T}}), \tag{98}$$

with v_j in (97). This completes the proof. □

E Proof of Theorem 5.1

Proof. Let $\mathbf{\Sigma}_{i-1} \in \mathcal{S}_{i-1}$ and $j \in \mathcal{A}_{i-1}^c$. From Lemma 4.1, for all $j \in \mathcal{A}_{i-1}^c$, the optimization problem in (48) is equivalent to the following optimization problem:

$$\begin{aligned} \min_{\mathbf{\Delta}} \ & J\left(\mathbf{\Sigma}_{i-1}\right) + f(\mathbf{\Sigma}_{i-1}, \mathbf{\Delta}) \\ \text{s.t.} \ & \mathbf{\Delta} \in \mathcal{D}_j, \\ & \mathbf{\Sigma}_{i-1} + \mathbf{\Delta} \in S_+^m, \end{aligned} \tag{99}$$

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

$$\begin{aligned} f(\mathbf{\Sigma}_{i-1}, \mathbf{\Delta}) =& (1-\lambda) \log \left| \mathbf{I}_m + \left(\mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_{i-1}\right)^{-1} \mathbf{\Delta} \right| - \log \left| \mathbf{I}_m + \left(\sigma^2 \mathbf{I}_m + \mathbf{\Sigma}_{i-1}\right)^{-1} \mathbf{\Delta} \right| \\ & + \lambda \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1} \mathbf{\Delta}_i\right). \end{aligned} \tag{100}$$

Therefore, the optimization problem in (99) is equivalent to

$$\begin{aligned} \min_{\mathbf{\Delta}} \ & f(\mathbf{\Sigma}_{i-1}, \mathbf{\Delta}) \\ \text{s.t.} \ & \mathbf{\Delta} \in \mathcal{D}_j, \\ & \mathbf{\Sigma}_{i-1} + \mathbf{\Delta} \in S_+^m. \end{aligned} \tag{101}$$

Note that the minimization problem in (101) is equivalent to the following optimization problem, up to a constant additive term,

$$\begin{aligned} \min_{\mathbf{\Delta}} \ & (1-\lambda) \log \left| \mathbf{\Sigma}_{YY} + \mathbf{\Sigma}_{i-1} + \mathbf{\Delta} \right| - \log \left| \sigma^2 \mathbf{I}_m + \mathbf{\Sigma}_{i-1} + \mathbf{\Delta} \right| + \lambda \mathrm{tr}\left(\mathbf{\Sigma}_{YY}^{-1} \mathbf{\Delta}_i\right) \\ \text{s.t.} \ & \mathbf{\Delta} \in \mathcal{D}_j, \\ & \mathbf{\Sigma}_{i-1} + \mathbf{\Delta} \in S_+^m. \end{aligned} \tag{102}$$

Noting that S_+^m is convex and the set $\mathcal{D}_j$ is convex for all $j \in \mathcal{A}_{i-1}^{\mathrm{c}}$. Hence, the logarithm terms are convex [32] for $\lambda \geq 1$. The trace term is a linear operation. It follows that the optimization problem in (102) is convex in $\mathbf{\Delta}$. Therefore, the optimization problem in (49) is convex in $\mathbf{\Delta}$. This completes the proof. □

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] A. Bretas, N. Bretas, J. B. London Jr, and B. Carvalho, *Cyber-physical power systems state estimation*. Elsevier, 2021.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

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

[16] 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.

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

[18] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[19] X. Ye, I. Esnaola, S. M. Perlaza, and R. F. Harrison, "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.

[20] 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.

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

[22] K. Sun and Z. Li, "Sparse data injection attacks on smart grid: An information-theoretic approach," *IEEE Sensors Journal*, vol. 22, no. 14, pp. 14 553–14 562, May 2022.

[23] 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.

[24] P. Lévy, "Propriétés asymptotiques des sommes de variables aléatoires enchaînées," *J. Math. Pures Appl.*, vol. 14, pp. 109–128, 1935.

[25] H. Cramér, "Über eine Eigenschaft der normalen Verteilungsfunktion," *Math. Z.*, vol. 41, pp. 405–414, 1936.

[26] 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.

[27] T. M. Cover and J. A. Thomas, *Elements of information theory.* John Wiley & Sons, Nov. 2012.

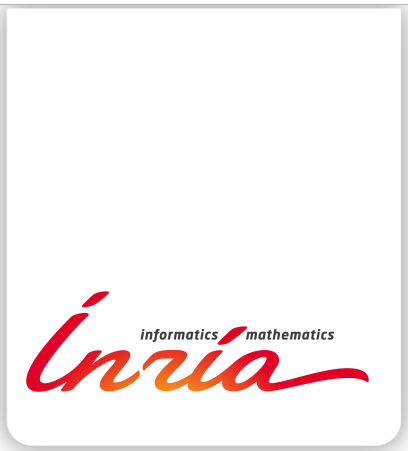
[28] H. V. Poor, *An introduction to signal detection and estimation.* Springer, 1994.

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

[30] 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.

[31] 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.

[32] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.



RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Publisher
Inria
Domaine de Voluceau -
Rocquencourt
BP 105 - 78153 Le Chesnay
Cedex
inria.fr