Detection of False Data Injection Attacks in Power Systems Using a Secured-Sensors and Graph-Based Method

Gal Morgenstern[1], Lital Dabush[1], Jip Kim[2], James Anderson[3], Gil Zussman[3], and Tirza Routtenberg[1]

[1] Ben Gurion University of the Negev, Beer-Sheva 84105, Israel
galmo@post.bgu.ac.il
[2] KENTECH, South Korea
[3] Columbia University, New York, NY

Abstract. False data injection (FDI) attacks pose a significant threat to the reliability of power system state estimation (PSSE). Recently, graph signal processing (GSP)-based detectors have been shown to enable the detection of well-designed cyber attacks named unobservable FDI attacks. However, current detectors, including GSP-based detectors, do not consider the impact of secured sensors on the detection process; thus, they may have limited power, especially in the low signal-to-noise ratio (SNR) regime. In this paper, we propose a novel FDI attack detection method that incorporates both knowledge of the locations of secured sensors and the GSP properties of power system states (voltages). We develop the secured-sensors-and-graph-Laplacian-based generalized likelihood ratio test (SSGL-GLRT) that integrates the secured data and the graph smoothness properties of the state variables. Furthermore, we introduce a generalization of the method that allows the use of different high-pass GSP filters together with prior knowledge of the locations of the secured sensors. Then, we develop the SSGL-GLRT for a distributed PSSE based on the alternating direction method of multipliers (ADMM). Numerical simulations demonstrate that the proposed method significantly improves the probability of detecting FDI attacks compared to existing GSP-based detectors, achieving an increase of up to 30% in the detection probability for the same false alarm rate by integrating secured sensor location information.

Keywords: Graph signal processing (GSP) · false data injection (FDI) attack detection · secured sensors · power system state estimation (PSSE) · cyber-physical systems · distributed detection.

1 Introduction

Smart grids integrate traditional power system components with advanced information and communication technology (ICT), providing critical cyber-physical infrastructure [43]. However, this also makes them vulnerable to cyber attacks [40–42], particularly false data injection (FDI) attacks, where an attacker corrupts measurements and injects fake information into the system. FDI attacks may inflict severe damage that ranges from economic consequences to the destruction of grid devices [14, 23, 24, 47, 48] by influencing the critical power system state estimation (PSSE) process, which provides grid monitoring signals for power system operations [26, 27]. PSSE is typically equipped with residual-based bad data detection (BDD) capabilities and, therefore can identify faulty data and random faults [27]. However, a well-designed, unobservable FDI attack can bypass the conventional residual-based BDD [21, 25]. Therefore, developing advanced tools to detect unobservable FDI attacks is crucial to maintaining high power supply quality and stable system operation.

In the past decade, various methods have been proposed for the detection of unobservable FDI attacks. Some methods utilize a set of protected measurements or synchronized phasor measurement units [2, 6, 19, 20]. Specifically, these works aim to find the best locations for the protected sensors. Machine learning-based methods have been proposed, but they require a large, stationary, and reliable database of data, which

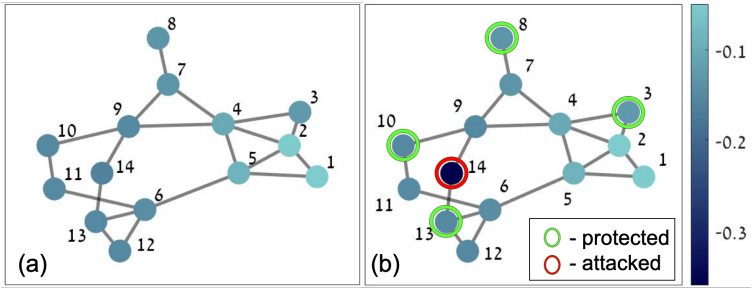


Fig. 1: Graph representation of IEEE 14-bus system. The node color represents the value of the states (voltage phases). In (a), the grid is not under attack, whereas in (b), node 14 is attacked (red circle), and nodes $\{3, 8, 10, 13\}$ are protected (green circles). It can be seen that the unattacked grid state is much smoother than the attacked grid state, i.e., the states of connected buses tend to be similar.

is often not available [10, 12, 18, 44]. Sparse methods were proposed in [28, 34]. However, these methods impose assumptions on the stationary and structural characteristics of the system loads, such as the lack of correlation with the system topology and the sparse nature of the attack in the time domain, which may not be true in real-world situations. Additionally, previous studies, such as [17, 31, 46], investigated the use of BDD and cyber attacks to compromise the distributed PSSE. Furthermore, graph signal processing (GSP) methods have been demonstrated to be useful for the detection of failures, topology changes, and FDI attacks [4, 5, 9, 11, 29, 33, 37, 38]. Despite this, incorporating information on secured sensor locations into FDI detection designs has not yet been explored either in centralized or in distributed frameworks. Additionally, the use of GSP properties for FDI detection remains at a preliminary stage and has not been fully investigated.

In this study, we present a novel approach for the detection of unobservable FDI attacks in power systems in the presence of secured sensors that are assumed to be immune to adversarial cyber attacks. These sensors with secured measurements can be obtained by additional validation processes by methods such as encryption, continuous monitoring, and separation from the Internet [20]. Our approach leverages the fact that the system states are known to be smooth graph signals [8, 9, 33], as illustrated in Fig. 1. Moreover, our approach is distinguished from existing GSP-based detectors by its ability to incorporate prior knowledge on the locations of the secured sensors. We formulate the hypothesis testing for this setting and derive the secured-sensors-and-graph-Laplacian-based generalized likelihood ratio test (SSGL-GLRT) that incorporates both the information on the locations of the secured-sensors and the graph smoothness properties of the system states. Furthermore, we introduce a generalization of the SSGL-GLRT by replacing the graph smoothness measure with any high-pass graph filter. The considered model can also accommodate distributed power system operation. In this approach, the network is divided into interconnected areas that are controlled separately, but share partial information. To this end, we derive the distributed SSGL-GLRT, that utilizes the alternating detection method of multipliers (ADMM) optimization algorithm in [3]. The numerical results indicate that the proposed SSGL-GLRT with secured sensors achieves a higher probability of detection and a lower false alarm rate, compared to existing methods, in the presence of secured sensors. This is due to the fact that the SSGL-GLRT exploits the graph smoothness property of the states as well as the knowledge of unattacked measurements.

In the following, vectors and matrices are denoted by boldface lowercase and uppercase letters, respectively. The mth element of the vector $\mathbf{a}$ and the $(m, q)^{\text{th}}$ element of the matrix $\mathbf{A}$ are denoted by a_m and $A_{m,q}$, respectively. Similarly, $\mathbf{a}_\Lambda$ is a subvector of $\mathbf{a}$ with the elements indexed by Λ. The matrix $\mathbf{I}$ and the vector $\mathbf{0}$ denote the identity matrix and the zero vector, respectively, with appropriate dimensions, and $||\cdot||$ denotes the Euclidean l_2-norm of vectors.

2 Model

The power system is represented by an undirected weighted graph, $\mathcal{G}(\mathcal{V},\xi)$, where $\mathcal{V}$ is the set of N nodes (bus and/or generators), and ξ is the set of edges (transmission lines) between the nodes. In this graph representation of the power system, it can be shown that the nodal admittance matrix is a graph Laplacian matrix. The (k,l)th element of $\mathbf{B}$ is given by [27]

$$B_{k,l}=\begin{cases}-\sum_{n\in\mathcal{N}_k} b_{k,n}, & k=l\\ b_{k,l}, & (k,l)\in\xi\\ 0, & \text{otherwise}\end{cases},\ \forall k,l=1,\ldots,N, \tag{1}$$

where $\mathcal{N}_k$ is the set of buses connected to bus k and $b_{k,n}<0$ is the susceptance of line $(k,n)\in\xi$.

The power system is governed by the nonlinear power flow equations, which are often approximated by the linearized DC model [27]. We consider the attacked and noisy DC model:

$$\mathbf{z}=\mathbf{H}\boldsymbol{\theta}+\mathbf{a}+\mathbf{e}, \tag{2}$$

where the active power measurements, $\mathbf{z}\in\mathbb{R}^M$, are corrupted by an additive FDI attack, $\mathbf{a}\in\mathbb{R}^M$, and by measurement noise, $\mathbf{e}\in\mathbb{R}^M$, which is assumed to be a zero-mean Gaussian vector with covariance matrix $\mathbf{R}$. The matrix $\mathbf{H}\in\mathbb{R}^{M\times N}$ is a known full-rank matrix, which is determined by the network topology and by the admittance matrix [27]. It should be noted that the matrix $\mathbf{B}$ from (1) is a submatrix of $\mathbf{H}$ from (2) that is associated with the power injection meters. Finally, the system states, i.e., the voltage phases, are denoted by $\boldsymbol{\theta}\in\mathbb{R}^N$.

In the GSP literature, signals measured over the nodes of the graph are assumed to be smooth w.r.t. the Laplacian matrix [7, 15, 22, 35, 39, 45, 49]. In the context of power systems, it was shown in [4, 9, 32] that the system states are smooth graph signals, i.e.

$$TV_{\mathcal{G}}(\boldsymbol{\theta})\triangleq\boldsymbol{\theta}^T\mathbf{B}\boldsymbol{\theta}\leq\varepsilon_1, \tag{3}$$

where $\varepsilon_1>0$ is small relative to all other parameters in the system. By substituting (1) in (3), we obtain

$$TV_{\mathcal{G}}(\boldsymbol{\theta})=\frac{1}{2}\sum_{k=1}^{N}\sum_{n\in\mathcal{N}_k}B_{k,n}\big(\theta_k-\theta_n\big)^2. \tag{4}$$

Roughly speaking, the smoothness property in (3), also referred to as graph total variation (TV), implies that the signal values (states in power systems) associated with the end nodes of edges with high weights in the graph (buses with large susceptance values) tend to be similar. In particular, the voltage angles of connected buses are similar.

The FDI attack, $\mathbf{a}\in\mathbb{R}^M$, is considered to be an unobservable FDI attack [25], i.e. it satisfies

$$\mathbf{a}=\mathbf{H}\mathbf{c}, \tag{5}$$

where $\mathbf{c}\in\mathbb{R}^N$ is an arbitrary vector. As a result, the attack $\mathbf{a}$ is in the range of $\mathbf{H}$. It is known that the attack described in (5) surpasses classical BDD methods [21].

3 GSP-Based FDI Detection with Secured Sensors

In this section, we design the SSGL-GLRT for detecting unobservable FDI attacks in the presence of secured measurements. In particular, it is assumed that a subset of the measurements is more reliable as these

measurements are equipped with additional protection measures, e.g. encryption, continuous monitoring, and separation from the Internet [20]. This set of protected sensors may encompass generator nodes, which are typically highly secured, and/or specific locations that were chosen based on a defense policy against FDI attacks. The SSGL-GLRT is based on the generalized likelihood ratio test (GLRT). Specifically, we consider the following hypothesis test associated with the model from Section 2:

$$\begin{cases} \mathcal{H}_0 : \mathbf{a} = \mathbf{0} \\ \mathcal{H}_1 : \mathbf{a} \neq \mathbf{0}. \end{cases}$$

To this end, we derive the secured-sensors-and-graph-Laplacian-based maximum likelihood estimator (SSGL-ML) of the states in Subsection 3.1. Subsequently, we use the SSGL-ML to derive the SSGL-GLRT in Subsection 3.2, and discuss its properties in Subsection 3.3.

3.1 SSGL-MLE

As stated at the beginning of this section, a subset of measurements, $\Lambda \subset \{1, \ldots, M\}$, is assumed highly secured. From the point of view of an adversary, this assumption implies that the measurements in the subset Λ cannot be attacked:

$$\mathbf{a}_\Lambda = \mathbf{0}. \tag{6}$$

From the defender's perspective, we assume that constraint (6) is relaxed and replaced by the following assumption:

$$||\mathbf{a}_\Lambda||^2 = ||\mathbf{M}\mathbf{a}||^2 \leq \varepsilon_2, \tag{7}$$

where ε_2 is small relative to the other parameters in the system and $\mathbf{M}$ is a diagonal mask matrix with the diagonal elements

$$M_{i,i} = \begin{cases} 1 & i \in \Lambda \\ 0 & i \notin \Lambda. \end{cases}$$

Assumption (7) implies that the attack, $\mathbf{a}$, has relatively small absolute values over the sensors in the set $\Lambda \subset \mathcal{M}$. This assumption permits flexibility in the case where some sensors in the set Λ are affected by random bad data (not originated by an attack), and makes the system more robust to small misspecifications or perturbations of Λ.

The SSGL-ML is a PSSE method with prior knowledge about the locations of the secured measurements and the graph smoothness properties of the system states [4, 9, 33]. The SSGL-ML is solved by maximizing the following regularized log-likelihood function over the system state variables $\boldsymbol{\theta}$ and the FDI attack $\mathbf{a}$:

$$\begin{aligned} \mathcal{Q}^{SSGL}(\boldsymbol{\theta}, \mathbf{a}) = &- (\mathbf{z} - \mathbf{H}\boldsymbol{\theta} - \mathbf{a})^T \mathbf{R}^{-1} (\mathbf{z} - \mathbf{H}\boldsymbol{\theta} - \mathbf{a}) \\ &- \mu_1 \boldsymbol{\theta}^T \mathbf{B}\boldsymbol{\theta} - \mu_2 ||\mathbf{M}\mathbf{a}||^2, \end{aligned} \tag{8}$$

where $\mu_1 > 0$ and $\mu_2 > 0$ are regularization parameters. These parameters enable the system operator to adjust the importance of each of the regularization functions. Note that the log-likelihood function in (8) is a concave function (see Appendix), and thus, the solution to the SSGL-ML is obtained by solving the normal equations. This function is equivalent to the standard PSSE log-likelihood function with two additional regularization terms:

R.1 Graph-Laplacian regularization ($\mu_1 \boldsymbol{\theta}^T \mathbf{B}\boldsymbol{\theta}$): A graph smoothness regularization term that incorporates the smoothness of the states in (3). This allows us to make a distinction between the system states, which are considered smooth, and the non-smooth FDI attack.

R.2 Secured-sensors regularization ($\mu_2 ||\mathbf{M}\mathbf{a}||^2$): An energy regularization function that incorporates the information on the locations of the secured sensors by using (7). This allows further distinction between the signal $\mathbf{H}\boldsymbol{\theta}$, which is a non-sparse signal with energy across all sensor positions, and the low-energy attack.

We now derive the SSGL-ML for the state vector, $\boldsymbol{\theta}$, and the attack vector, $\mathbf{a}$, based on the regularized log-likelihood function in (8). Later, these estimators will be used for deriving the GLRT in Subsection 3.2. We first consider the null hypothesis, $\mathcal{H}_0$, i.e. there is no attack ($\mathbf{a} = \mathbf{0}$). By substituting $\mathbf{a} = \mathbf{0}$ in (8), we obtain that under hypothesis $\mathcal{H}_0$, the SSGL-ML of $\boldsymbol{\theta}$ is

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{|\mathcal{H}_0}^{\text{SSGL-ML}} &= \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^N} -\mathcal{Q}^{SSGL}(\boldsymbol{\theta}, \mathbf{a}=\mathbf{0}) \\ &= \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^N} (\mathbf{z}-\mathbf{H}\boldsymbol{\theta})^T\mathbf{R}^{-1}(\mathbf{z}-\mathbf{H}\boldsymbol{\theta}) + \mu_1\boldsymbol{\theta}^T\mathbf{B}\boldsymbol{\theta} \\ &= \mathbf{K}^{\boldsymbol{\theta}}\mathbf{z}, \end{aligned} \tag{9}$$

where the gain matrix is given by

$$\mathbf{K}^{\boldsymbol{\theta}} \triangleq (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mu_1\mathbf{B})^{-1}\mathbf{H}^T\mathbf{R}^{-1}. \tag{10}$$

The SSGL-ML estimator in (9)-(10) coincides with the GSP weighted least squares (GSP-WLS) estimator from [4].

Under hypothesis $\mathcal{H}_1$, when it is known that $\mathbf{a} \neq \mathbf{0}$, the SSGL-ML for both $\boldsymbol{\theta}$ and $\mathbf{a}$ is given by

$$(\hat{\boldsymbol{\theta}}_{|\mathcal{H}_1}^{\text{SSGL-ML}}, \hat{\mathbf{a}}_{|\mathcal{H}_1}^{\text{SSGL-ML}}) = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^N, \mathbf{a}\in\mathbb{R}^M} -\mathcal{Q}^{SSGL}(\boldsymbol{\theta}, \mathbf{a}). \tag{11}$$

Since $-\mathcal{Q}^{SSGL}(\boldsymbol{\theta}, \mathbf{a})$ from (8) is convex (see Appendix), the estimators of $\boldsymbol{\theta}$ and $\mathbf{a}$ can be computed by the following normal equations:

$$\mathbf{a} = \mathbf{K}^{\mathbf{a}}(\mathbf{z} - \mathbf{H}\boldsymbol{\theta}) \tag{12}$$

$$\boldsymbol{\theta} = \mathbf{K}^{\boldsymbol{\theta}}(\mathbf{z} - \mathbf{a}), \tag{13}$$

where $\mathbf{K}^{\boldsymbol{\theta}}$ is defined in (10) and

$$\mathbf{K}^{\mathbf{a}} = (\mathbf{R}^{-1} + \mu_2\mathbf{M})^{-1}\mathbf{R}^{-1}. \tag{14}$$

Substituting (12) into (13) results in

$$\hat{\boldsymbol{\theta}}_{|\mathcal{H}_1}^{\text{SSGL-ML}} = \mathbf{A}^{\boldsymbol{\theta}}\mathbf{z}, \tag{15}$$

where

$$\mathbf{A}^{\boldsymbol{\theta}} \triangleq \left(\mathbf{I} - \mathbf{K}^{\boldsymbol{\theta}}\mathbf{K}^{\mathbf{a}}\mathbf{H}\right)^{-1}\mathbf{K}^{\boldsymbol{\theta}}(\mathbf{I} - \mathbf{K}^{\mathbf{a}}). \tag{16}$$

Substituting (15) in (12) results in

$$\hat{\mathbf{a}}_{|\mathcal{H}_1}^{\text{SSGL-ML}} = \mathbf{K}^{\mathbf{a}}(\mathbf{I} - \mathbf{H}\mathbf{A}^{\boldsymbol{\theta}})\mathbf{z}. \tag{17}$$

The MLEs of $\boldsymbol{\theta}$ and $\mathbf{a}$ given in (9), (15), and (17), are used in the next subsection to derive the SSGL-GLRT.

3.2 SSGL-GLRT

The SSGL-GLRT is the difference between the regularized log-likelihood function from (8) under $\mathcal{H}_1$ and under $\mathcal{H}_0$ [16]:

$$T^{\text{SSGL-GLRT}}(\mathbf{z}) = \mathcal{Q}^{SSGL}(\hat{\boldsymbol{\theta}}_{|\mathcal{H}_1}^{\text{SSGL-ML}}, \hat{\mathbf{a}}_{|\mathcal{H}_1}^{\text{SSGL-ML}}) - \mathcal{Q}^{SSGL}(\hat{\boldsymbol{\theta}}_{|\mathcal{H}_0}^{\text{SSGL-ML}}, \mathbf{0}). \tag{18}$$

By using (15) and (17), we obtain

$$\begin{aligned}\mathcal{Q}^{SSGL}(\hat{\boldsymbol{\theta}}_{|\mathcal{H}_1}^{\text{SSGL-ML}}, \hat{\mathbf{a}}_{|\mathcal{H}_1}^{\text{SSGL-ML}}) = &- (\mathbf{z} - \mathbf{H}\mathbf{A}^{\boldsymbol{\theta}}\mathbf{z} - \mathbf{K}^{\mathbf{a}}(\mathbf{I} - \mathbf{H}\mathbf{A}^{\boldsymbol{\theta}})\mathbf{z})^T\mathbf{R}^{-1} \\ &\times (\mathbf{z} - \mathbf{H}\mathbf{A}^{\boldsymbol{\theta}}\mathbf{z} - \mathbf{K}^{\mathbf{a}}(\mathbf{I} - \mathbf{H}\mathbf{A}^{\boldsymbol{\theta}})\mathbf{z}) \\ &- \mu_1(\mathbf{A}^{\boldsymbol{\theta}}\mathbf{z})^T\mathbf{B}\mathbf{A}^{\boldsymbol{\theta}}\mathbf{z} - \mu_2||\mathbf{M}\mathbf{K}^{\mathbf{a}}(\mathbf{I} - \mathbf{H}\mathbf{A}^{\boldsymbol{\theta}})\mathbf{z}||^2. \end{aligned} \tag{19}$$

Similarly, using (9), we obtain

$$\begin{aligned}\mathcal{Q}^{SSGL}(\hat{\boldsymbol{\theta}}^{\text{SSGL-ML}}_{|\mathcal{H}_0}, \mathbf{0}) = &- (\mathbf{z} - \mathbf{H}\mathbf{K}^{\boldsymbol{\theta}}\mathbf{z})^T\mathbf{R}^{-1}(\mathbf{z} - \mathbf{H}\mathbf{K}^{\boldsymbol{\theta}}\mathbf{z}) \\ &- \mu_1(\mathbf{K}^{\boldsymbol{\theta}}\mathbf{z})^T\mathbf{B}\mathbf{K}^{\boldsymbol{\theta}}\mathbf{z}.\end{aligned} \tag{20}$$

Substituting (19) and (20) in (18), results in

$$T^{\text{SSGL-GLRT}}(\mathbf{z}) = \mathbf{z}^T\mathbf{G}\mathbf{z}, \tag{21}$$

where

$$\begin{aligned}\mathbf{G} \triangleq &(\mathbf{I} - \mathbf{H}\mathbf{K}^{\boldsymbol{\theta}})^T\mathbf{R}^{-1}(\mathbf{I} - \mathbf{H}\mathbf{K}^{\boldsymbol{\theta}}) - (\mathbf{I} - \mathbf{H}\mathbf{A}^{\boldsymbol{\theta}})^T(\mathbf{I} - \mathbf{K}^{\mathbf{a}})^T\mathbf{R}^{-1} \\ &\times (\mathbf{I} - \mathbf{K}^{\mathbf{a}})(\mathbf{I} - \mathbf{H}\mathbf{A}^{\boldsymbol{\theta}}) + \mu_1\big((\mathbf{K}^{\boldsymbol{\theta}})^T\mathbf{B}\mathbf{K}^{\boldsymbol{\theta}} - (\mathbf{A}^{\boldsymbol{\theta}})^T\mathbf{B}\mathbf{A}^{\boldsymbol{\theta}}\big) \\ &- \mu_2(\mathbf{I} - \mathbf{H}\mathbf{A}^{\boldsymbol{\theta}})^T(\mathbf{K}^{\mathbf{a}})^T\mathbf{M}\mathbf{K}^{\mathbf{a}}(\mathbf{I} - \mathbf{H}\mathbf{A}^{\boldsymbol{\theta}}).\end{aligned} \tag{22}$$

The SSGL-GLRT in (21) is a weighted energy detector, where the weight matrix $\mathbf{G}$ in (22) is composed of five components: The first and second components evaluate the estimation accuracy of the SSGL-ML under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively, w.r.t. the input measurements. The third and fourth components evaluate the smoothness of the estimated state vector under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Finally, the fifth component evaluates the compliance of the estimated attack with the assumption in (7).

The computational complexity of the detector proposed in (21)-(22) can be separated into two parts: the online and offline operations. Online, it is required to compute (21) given the $M \times M$ matrix $\mathbf{G}$ and the $M \times 1$ vector $\mathbf{z}$. In this case, the number of multiplications is in order of $O(M^2)$ when $\mathbf{G}$ is dense and unstructured. Offline, it is required to calculate the matrix $\mathbf{G}$ defined in (22). In this case, the most demanding procedure is the inverse of $\mathbf{R}$, which is an $M \times M$ matrix. Thus, the computational complexity is in order of $O(M^3)$ when $\mathbf{R}$ is dense and unstructured.

3.3 Special Cases

In the following, we present a few special cases of the SSGL-GLRT.

C.1 No regularization ($\mu_1 = \mu_2 = 0$): By substituting $\mu_1 = 0$ and $\mu_2 = 0$ in (10) and (14), we obtain

$$\mathbf{K}^{\boldsymbol{\theta}} = \mathbf{K} \triangleq (\mathbf{H}^T\mathbf{R}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}$$

and $\mathbf{K}^{\mathbf{a}} = \mathbf{I}$, respectively. Substituting these results and $\mu_1 = \mu_2 = 0$ in (22), results in

$$\mathbf{G} = (\mathbf{I} - \mathbf{H}\mathbf{K})^T\mathbf{R}^{-1}(\mathbf{I} - \mathbf{H}\mathbf{K}). \tag{23}$$

By substituting (23) in (21), one obtains the $J(\boldsymbol{\theta})$-test [27]:

$$T^{\text{BDD}}(\mathbf{z}) = \mathbf{z}^T(\mathbf{I} - \mathbf{H}\mathbf{K})^T\mathbf{R}^{-1}(\mathbf{I} - \mathbf{H}\mathbf{K})\mathbf{z}. \tag{24}$$

It is known that the BDD detector in (24) cannot detect unobservable FDI attacks as defined in (5) (see e.g. [21, 25]).

C.2 Only Laplacian-based regularization ($\mu_1 > 0$, $\mu_2 = 0$): When $\mu_2 = 0$, similarly to in **C.1**, we obtain that $\mathbf{K}^{\mathbf{a}} = \mathbf{I}$. By substituting this result and $\mu_2 = 0$ into (16), we get $\mathbf{A}^{\boldsymbol{\theta}} = \mathbf{0}$. Thus, in this case, (22) is reduced to

$$\mathbf{G} = (\mathbf{I} - \mathbf{H}\mathbf{K}^{\boldsymbol{\theta}})^T\mathbf{R}^{-1}(\mathbf{I} - \mathbf{H}\mathbf{K}^{\boldsymbol{\theta}}) + \mu_1(\mathbf{K}^{\boldsymbol{\theta}})^T\mathbf{B}\mathbf{K}^{\boldsymbol{\theta}}. \tag{25}$$

Finally, substitution of (25) in (21) results in

$$\begin{aligned}T^{\text{GL-GLRT}}(\mathbf{z}) = &\mathbf{z}^T(\mathbf{I} - \mathbf{H}\mathbf{K}^{\boldsymbol{\theta}})^T\mathbf{R}^{-1}(\mathbf{I} - \mathbf{H}\mathbf{K}^{\boldsymbol{\theta}})^T\mathbf{z} \\ &+ \mu_1\mathbf{z}^T(\mathbf{K}^{\boldsymbol{\theta}})^T\mathbf{B}\mathbf{K}^{\boldsymbol{\theta}}\mathbf{z},\end{aligned} \tag{26}$$

which is the graph-Laplacian-regularized GLRT (GL-GLRT) from [5]: that only considers the prior on the smoothness of the states.

C.3 Only secured-sensors-based regularization ($\mu_1 = 0$, $\mu_2 > 0$): By substituting $\mu_1 = 0$ in (10) and (16) we obtain $\mathbf{K}^{\boldsymbol{\theta}} = \mathbf{K}$ and

$$\mathbf{A}_2^{\boldsymbol{\theta}} \triangleq \big(\mathbf{I} - \mathbf{K}\mathbf{K}^{\mathbf{a}}\mathbf{H}\big)^{-1}\mathbf{K}(\mathbf{I} - \mathbf{K}^{\mathbf{a}}).$$

By substituting these results in (22), we obtain the weighting matrix for this case:

$$\begin{aligned}\mathbf{G} = &-(\mathbf{I} - \mathbf{H}\mathbf{A}_2^{\boldsymbol{\theta}})^T(\mathbf{I} - \mathbf{K}^{\mathbf{a}})^T\mathbf{R}^{-1}(\mathbf{I} - \mathbf{K}^{\mathbf{a}})(\mathbf{I} - \mathbf{H}\mathbf{A}_2^{\boldsymbol{\theta}})\\ &-\mu_2(\mathbf{I} - \mathbf{H}\mathbf{A}_2^{\boldsymbol{\theta}})^T(\mathbf{K}^{\mathbf{a}})^T\mathbf{M}\mathbf{K}^{\mathbf{a}}(\mathbf{I} - \mathbf{H}\mathbf{A}_2^{\boldsymbol{\theta}})\\ &+(\mathbf{I} - \mathbf{H}\mathbf{K})^T\mathbf{R}^{-1}(\mathbf{I} - \mathbf{H}\mathbf{K}).\end{aligned}$$

The resulting detector only takes into account the prior information of the secured measurements. However, this detector is not practical because if Λ does not include all measurements, i.e. some measurements are not secured, then $(\mathbf{I} - \mathbf{K}\mathbf{K}^{\mathbf{a}}\mathbf{H})$ is not invertible. Moreover, by substituting (13) in (12) and then substituting $\mathbf{K}^{\boldsymbol{\theta}} = \mathbf{K}$ we see that (17) can also be written as

$$\hat{\mathbf{a}}^{\text{SS-ML}}_{|\mathcal{H}_1} = (I - \mathbf{K}^{\mathbf{a}}\mathbf{H}\mathbf{K})^{-1}\mathbf{K}^{\mathbf{a}}(\mathbf{I} - \mathbf{H}\mathbf{K})\mathbf{z}.$$

This indicates that for unobservable attacks, $\mathbf{a} = \mathbf{H}\mathbf{c}$, we obtain that $\hat{\mathbf{a}}^{\text{SS-ML}}_{|\mathcal{H}_1}$ is the same for input $\mathbf{z}$ and its corrupted version $\mathbf{z} + \mathbf{H}\mathbf{c}$, because

$$(\mathbf{I} - \mathbf{H}\mathbf{K})\mathbf{H}\mathbf{c} = \mathbf{H}\mathbf{c} - \mathbf{H}\mathbf{c} = \mathbf{0}.$$

Hence, this detector is not effective against unobservable FDI attacks.

	Regularization term	
Detector	**Secured sensors**	**Graph Laplacian**
SSGL-GLRT	v	v
GL-GLRT	x	v
PP-GLRT	v	x
BDD	x	x

Table 1: Classification of the different GLRTs based on the regularization functions used.

3.4 General Graph High Pass Filter (GHPF)

The SSGL-GLRT exploits the smoothness property of the states in (3). Other approaches in [9, 32] are built upon the idea that the states can be thought of as graph signals with low energy in the high-frequency range of the graph spectrum, as defined in the GSP literature [39]. Similarly, we can generalize the proposed SSGL-GLRT as follows. Since the states can be considered low-pass graph signals [32], the smoothness term, $\boldsymbol{\theta}^T\mathbf{B}\boldsymbol{\theta}$, can be replaced by any term of the form

$$\boldsymbol{\theta}^T\mathbf{U}_B f^{\frac{1}{2}}(\boldsymbol{\Phi}_B)\mathbf{U}_B^T\boldsymbol{\theta}, \tag{27}$$

where $\mathbf{U}_B$ and $\boldsymbol{\Phi}_B$ are the eigenvector and eigenvalue matrices of $\mathbf{B}$, i.e. $\mathbf{B} = \mathbf{U}_B\boldsymbol{\Phi}_B\mathbf{U}_B^T$. The graph filter $f(\cdot)$ is assumed to be a nonnegative analytic function, defined by its graph frequency response [30], $f(\boldsymbol{\Phi}) = \text{diag}(f(\phi_1), \ldots, f(\phi_N))$. Roughly speaking, $f(\boldsymbol{\Phi})$ is a GHPF if the frequency response $f(\phi_n)$ increases

as the eigenvalue ϕ_n increases. Thus, using the GHPF in (27) results in a penalty on signal content in the high graph frequencies that can be used to detect outliers/anomalies w.r.t. the graph [36], or, in our case, FDI attacks. The practical implementation results in the same SSGL-GLRT, where $\mathbf{B}$ is replaced by $\left(\mathbf{U}_B f^{\frac{1}{2}}(\boldsymbol{\Phi}_B)\mathbf{U}_B^T\right)$ everywhere.

For example, using the graph frequency response

$$f(\phi_n) = \sqrt{\phi_n}, \quad n = 1, \ldots, N$$

in (27), results in the smoothness criterion $\boldsymbol{\theta}^T\mathbf{B}\boldsymbol{\theta}$ used in the CP-GLRT. An alternative GHPF is the following ideal-GHPF:

$$f^{\text{GHPF}}(\phi_n) = \begin{cases} 0 & \phi_n \leq \phi_{cut} \\ 1 & \phi_n > \phi_{cut} \end{cases}, \quad n = 1, \ldots, N, \tag{28}$$

where ϕ_{cut} is the cutoff frequency. This GHPF is used for FDI detection in [9, 33], but without using protected measurements.

4 Distributed Detection

In the previous section, we derived the SSGL-GLRT for the centralized approach in which a single control center operates the system. However, a centralized approach may incur impractical computational and communication load, increased vulnerability, and disclosure of the internal system structure. Therefore, in this section, we discuss the modification of the SSGL-GLRT, and a special case, the GL-GLRT, for distributed frameworks. Our derivation is based on the distributed PSSE approach described in [17], in which the PSSE is performed with measurements corrupted by bad data. This section is organized as follows. In Subsection 4.1, we review the distributed PSSE from [17]. Then, in Subsection 4.2, we derive the proposed distributed SSGL-GLRT and GL-GLRT detectors.

4.1 Distributed PSSE

We consider an interconnected power system comprising L control areas. The measurement model for the lth area, based on the DC power flow model given in (2), can be expressed as

$$\mathbf{z}_l = \mathbf{H}_l\boldsymbol{\theta}_l + \mathbf{a}_l + \mathbf{e}_l, \quad l = 1, \ldots, L, \tag{29}$$

where $\boldsymbol{\theta}_l \in \mathbb{R}^{N_l \times 1}$ represents the subset of interconnected power system states (i.e. a subvector of $\boldsymbol{\theta}$) associated with the measurements in $\mathbf{z}_l$, $\mathbf{H}_l \in \mathbb{R}^{M_l \times N_l}$ is the appropriate submatrix topology matrix (a submatrix of $\mathbf{H}$), $\mathbf{a}_l \in \mathbb{R}^{M_l \times 1}$ is the attack on the sensors in the lth area (a submatrix of $\mathbf{H}$), and $\mathbf{e}_l \in \mathbb{R}^{M_l \times 1}$ represents the system noise in this area, modeled as a zero-mean Gaussian noise with covariance matrix $\mathbf{R}_q \in \mathbb{R}^{M_l \times M_l}$ (a submatrix of $\mathbf{R}$). The distributed PSSE can be written as the following optimization problem [17]:

$$\begin{aligned} \{\hat{\boldsymbol{\theta}}_l\}_{l=1}^L = \arg \min_{\substack{\boldsymbol{\theta}_l \\ l=1,\ldots,L}} & \sum_{l=1}^{L} \mathcal{Q}_l \\ s.t.\ & \boldsymbol{\theta}_l[l'] = \boldsymbol{\theta}_{l'}[l],\ \forall l' \in \mathcal{A}_l,\ \forall l, \end{aligned} \tag{30}$$

where the cost function of the different areas, $\mathcal{Q}_l$, is jointly minimized subject to the constraint that the state vectors of each area partially overlap. Specifically, we assume that the state vector of area l includes all buses in that area and their first-order neighbors, and the set $\mathcal{A}_l$ includes all areas that share state variables with area l. The notation $\boldsymbol{\theta}_l[l']$ represents the subvector of $\boldsymbol{\theta}_l$ that includes all state variables shared with area l'.

The solution to (30) by the ADMM algorithm [3] consists of the following iterative steps [17]:

$$\boldsymbol{\theta}_l^{(t+1)} = \arg\min_{\boldsymbol{\theta}} \ \mathcal{Q}_l(\boldsymbol{\theta}_l) + \frac{\zeta}{2}\sum_{i=1}^{N_l} \mathbb{1}_{\{\mathcal{A}_l^i \neq \emptyset\}} |\mathcal{A}_l^i| (\boldsymbol{\theta}_l(i) - \mathbf{p}_l^{(t)}(i))^2, \tag{31a}$$

$$\mathbf{s}_l^{(t+1)}(i) = \frac{1}{|\mathcal{A}_l^i|}\sum_{l\in\mathcal{A}_l^i} \boldsymbol{\theta}_l^{(t+1)}[i], \ \forall i \text{ with } \mathcal{A}_l^i \neq \emptyset, \tag{31b}$$

$$\mathbf{p}_l^{(t+1)}(i) = \mathbf{p}_l^{(t)}(i) + \mathbf{s}_l^{(t+1)}(i) - \frac{\boldsymbol{\theta}_l^{(t)}(i) - \mathbf{s}_l^{(t)}(i)}{2}, \ \forall i \text{ with } \mathcal{A}_l^i \neq \emptyset. \tag{31c}$$

Here, the auxiliary vectors $\mathbf{s}_l$ and $\mathbf{p}_l$ are used, and $\mathbb{1}_{(\cdot)}$ denotes the indicator function, which equals 1 if its condition is met and 0 otherwise. The set $\mathcal{A}_l^i$ represents the areas that share variable $\boldsymbol{\theta}_l(i)$ with area l. Additionally, the parameter ζ represents the user-defined step size. We use here the least squares cost function, $Q^{LS}(\theta) = (\mathbf{z} - \mathbf{H}\boldsymbol{\theta})^T\mathbf{R}^{-1}(\mathbf{z} - \mathbf{H}\boldsymbol{\theta})$, which can be modified for each area l to $Q_l^{LS}(\boldsymbol{\theta}) = (\mathbf{z}_l - \mathbf{H}_l\boldsymbol{\theta}_l)^T\mathbf{R}_l^{-1}(\mathbf{z}_l - \mathbf{H}_l\boldsymbol{\theta}_l)$. In this case, as shown in [17], the problem is solved using (31) while replacing (31a) with:

$$\boldsymbol{\theta}_l^{(t+1)} = (\mathbf{H}_l\mathbf{R}_l^{-1}\mathbf{H}_l + \zeta\mathbf{D}_l)^{-1}(\mathbf{R}_l^{-1}\mathbf{H}_l^T\mathbf{z}_l + \zeta\mathbf{D}_l\mathbf{p}_l^{(t)}), \tag{32}$$

where $\mathbf{D}_l$ is the diagonal matrix with the (i,i) entry $|\mathcal{A}_l^i|$. As for initialization, the state variables $\boldsymbol{\theta}_l$ are set to arbitrary values $\theta_l^{(0)}$, variables $\mathbf{s}_l^{(0)}$ are initialized as in (31b), and $\mathbf{p}_l^{(0)}(i)$ is initialized as $(\mathbf{x}_l^{(0)}(i) - \mathbf{s}_l^{(0)}(i))/2$. The ADMM iterative step converges when the objective function and constraints functions are convex, closed, and proper, and the augmented Lagrangian has a saddle point [3].

4.2 Distributed SSGL-GLRT and GL-GLRT

The cost function for the SSGL-ML in (8) is obtained by solving the standard PSSE, which is defined as an unconstrained LS problem along with two regularization terms. One term, $\mu_1\boldsymbol{\theta}^T\mathbf{B}\boldsymbol{\theta}$, imposes prior knowledge on the smoothness property of the state variables, as defined in (3). The other term, $\mu_2\|\mathbf{Ma}\|^2$, imposes prior knowledge on the secured sensors, as defined in (7). We modify the regularization terms to recast the optimization problem as the minimization of a regional cost function. Specifically, we introduce the local smoothness measure defined in [39], which is given by the inner summation of the global smoothness measure in (4):

$$S_i(\theta) = \sum_{j\in\mathcal{N}_i} B_{i,j}(\theta_i - \theta_j)^2, \tag{33}$$

where $\mathcal{N}_i$ is the first-order neighborhood of bus i. We measure the smoothness over each region by summing the local smoothness of all buses in that region, resulting in $\sum_{i=1}^{N_l} S_i(\boldsymbol{\theta})$. It can be verified that this sum satisfies

$$\sum_{i\in R} S_i(\boldsymbol{\theta}) = \boldsymbol{\theta}_l^T\mathbf{B}_l\boldsymbol{\theta}_l,$$

where $\mathbf{B}_l$ is the submatrix of $\mathbf{B}$ associated with the state variables in the lth region. Moreover, since the prior knowledge on the location of the secured sensors is local to each sensor, we modify the prior assumption in (34) for each area l to

$$\|\mathbf{M}_l\mathbf{a}_l\|^2 \leq \varepsilon_l, \tag{34}$$

where $\mathbf{M}_l$ is the $M_l \times M_l$ submatrix of the diagonal matrix $\mathbf{M}$ associated with the power measurements in the lth area. Using (29) and (33)-(34), we can modify the log-likelihood function in (8) to measure the cost function of the lth area as

$$\mathcal{Q}_l^{SSGL}(\boldsymbol{\theta}_l, \mathbf{a}_l) = -(\mathbf{z}_l - \mathbf{H}_l\boldsymbol{\theta}_l - \mathbf{a}_l)^T\mathbf{R}_l^{-1}(\mathbf{z}_l - \mathbf{H}_q\boldsymbol{\theta}_l - \mathbf{a}_l) - \mu_{1,l}\boldsymbol{\theta}_l^T\mathbf{B}_l\boldsymbol{\theta}_l - \mu_{2,l}\|\mathbf{M}_l\mathbf{a}_l\|^2. \tag{35}$$

Algorithm 1: Distributed SS-GLRT in area l

Input: Fix detection threshold γ_l and step size ζ Set initial guess: $\boldsymbol{\theta}_{|\mathcal{H}_0,l}^{(0)}$, $\mathbf{s}_l^{(0)}$, and $\mathbf{p}_l^{(0)}$

for $t = 0, 1, \ldots$ **do**

  Update:

  $\boldsymbol{\theta}_{|\mathcal{H}_0,l}^{(t+1)} = (\mathbf{H}_l\mathbf{R}_l^{-1}\mathbf{H}_l + \mu_{1,l}\mathbf{B}_l + \zeta\mathbf{D}_l)^{-1}(\mathbf{R}_l^{-1}\mathbf{H}_l^T\mathbf{z}_l + \zeta\mathbf{D}_l\mathbf{p}_l^{(t)})$

  $\mathbf{s}_l^{(t+1)}(i) = \frac{1}{|\mathcal{A}_l^i|}\sum_{l\in\mathcal{A}_l^i}\boldsymbol{\theta}_{|\mathcal{H}_0,l}^{(t+1)}[i],\ \forall i \text{ with } \mathcal{A}_l^i \neq \emptyset$

  $\mathbf{p}_l^{(t+1)}(i) = \mathbf{p}_l^{(t)}(i) + \mathbf{s}_l^{(t+1)}(i) - \frac{\boldsymbol{\theta}_{|\mathcal{H}_0,l}^{(t)}(i) - \mathbf{s}_l^{(t)}(i)}{2},\ \forall i \text{ with } \mathcal{A}_l^i \neq \emptyset$

Set $\hat{\boldsymbol{\theta}}_{|\mathcal{H}_0,l}^{\text{SSGL-ML}} = \boldsymbol{\theta}_{|\mathcal{H}_0,l}^{(t+1)}$

Set initial guess: $\boldsymbol{\theta}_{|\mathcal{H}_1,l}^{(0)}$, $\mathbf{s}_l^{(0)}$, $\mathbf{p}_l^{(0)}$, and $\mathbf{a}_{|\mathcal{H}_1,l}^{(0)}$

for $t = 0, 1, \ldots$ **do**

  Update:

  $\boldsymbol{\theta}_{|\mathcal{H}_1,l}^{(t+1)} = (\mathbf{H}_l\mathbf{R}_l^{-1}\mathbf{H}_l + \mu_{1,l}\mathbf{B}_l + \zeta\mathbf{D}_l)^{-1}(\mathbf{R}_l^{-1}\mathbf{H}_l^T(\mathbf{z}_l - \mathbf{a}_l^{(t)}) + \zeta\mathbf{D}_l\mathbf{p}_l^{(t)})$

  $\mathbf{s}_l^{(t+1)}(i) = \frac{1}{|\mathcal{A}_l^i|}\sum_{l\in\mathcal{A}_l^i}\boldsymbol{\theta}_{|\mathcal{H}_0,l}^{(t+1)}[i],\ \forall i \text{ with } \mathcal{A}_l^i \neq \emptyset$

  $\mathbf{p}_l^{(t+1)}(i) = \mathbf{p}_l^{(t)}(i) + \mathbf{s}_l^{(t+1)}(i) - \frac{\boldsymbol{\theta}_{|\mathcal{H}_0,l}^{(t)}(i) - \mathbf{s}_l^{(t)}(i)}{2},\ \forall i \text{ with } \mathcal{A}_l^i \neq \emptyset$

  $\mathbf{a}_{|\mathcal{H}_1,l}^{(t+1)} = (\mathbf{R}_l^{-1} + \mu_{2,l}\mathbf{M}_l)^{-1}\mathbf{R}_l^{-1}(\mathbf{z}_l - \mathbf{H}_l\boldsymbol{\theta}_l^{(t+1)})$

Set $\hat{\boldsymbol{\theta}}_{|\mathcal{H}_1,l}^{\text{SSGL-ML}} = \boldsymbol{\theta}_{|\mathcal{H}_1,l}^{(t+1)}$ and $\hat{\mathbf{a}}_{|\mathcal{H}_1,l}^{\text{SSGL-ML}} = \mathbf{a}_{|\mathcal{H}_1,l}^{(t+1)}$

if $\mathcal{Q}_l^{SSGL}(\hat{\boldsymbol{\theta}}_{|\mathcal{H}_1,l}^{SSGL\text{-}ML}, \hat{\mathbf{a}}_{|\mathcal{H}_1,l}^{SSGL\text{-}ML}) - \mathcal{Q}_l^{SSGL}(\hat{\boldsymbol{\theta}}_{|\mathcal{H}_0,l}^{SSGL\text{-}ML}, \mathbf{0}) > \gamma_l$ **then**

  return "The area is under an FDI attack"

else

  return "The area is under normal operation"

As presented in Section 3.2, the SSGL-GLRT is a detector derived from (18). For the distributed case, the SSGL-GLRT can be adapted by defining L detectors, denoted as $T_l^{\text{SSGL-GLRT}}$, where each detection test is performed in the corresponding control center. These detectors are defined as follows:

$$T_l^{\text{SSGL-GLRT}} = \mathcal{Q}_l^{SSGL}(\hat{\boldsymbol{\theta}}_{|\mathcal{H}_1,l}^{\text{SSGL-ML}}, \hat{\mathbf{a}}_{|\mathcal{H}_1,l}^{\text{SSGL-ML}}) - \mathcal{Q}_l^{SSGL}(\hat{\boldsymbol{\theta}}_{|\mathcal{H}_0,l}^{\text{SSGL-ML}}, \mathbf{0}),\ l = 1, \ldots, L, \tag{36}$$

where $\hat{\boldsymbol{\theta}}_{|\mathcal{H}_1,l}^{\text{SSGL-ML}}$ and $\hat{\mathbf{a}}_{|\mathcal{H}_1,l}^{\text{SSGL-ML}}$ are the ML estimates for the state variables and the attack in the lth area under the $\mathcal{H}_1$ hypothesis, and $\hat{\boldsymbol{\theta}}_{|\mathcal{H}_0,l}^{\text{SSGL-ML}}$ are the ML estimates for the state variable in the lth area under the $\mathcal{H}_0$ hypothesis.

For hypothesis $\mathcal{H}_0$, we seek to estimate $\hat{\boldsymbol{\theta}}_{|\mathcal{H}_0,l}^{\text{SSGL-ML}}$, which is obtained by replacing $\mathcal{Q}_l(\boldsymbol{\theta}_l)$ in (31) with (35) when $\mathbf{a}_l$ is replaced with $\mathbf{0}$. Therefore, we can estimate $\hat{\boldsymbol{\theta}}_{|\mathcal{H}_0,l}^{\text{SSGL-ML}}$ by applying the results from (9)-(10) to (31), which results in replacing (31a) with

$$\boldsymbol{\theta}_{|\mathcal{H}_0,l}^{(t+1)} = (\mathbf{H}_l\mathbf{R}_l^{-1}\mathbf{H}_l + \mu_{1,l}\mathbf{B}_l + \zeta\mathbf{D}_l)^{-1}(\mathbf{R}_l^{-1}\mathbf{H}_l^T\mathbf{z}_l + \zeta\mathbf{D}_l\mathbf{p}_l^{(t)}). \tag{37}$$

Note that the inclusion of the term $\zeta\mathbf{D}_l$ is motivated by the same reasons as in (32). For hypothesis $\mathcal{H}_1$, we want to estimate $(\hat{\boldsymbol{\theta}}_{|\mathcal{H}_1,l}^{\text{SSGL-ML}}, \hat{\mathbf{a}}_{|\mathcal{H}_1,l}^{\text{SSGL-ML}})$, which is obtained by replacing $\mathcal{Q}_l(\boldsymbol{\theta}_l)$ in (31) with (35), a function of both $\boldsymbol{\theta}_l$ and $\mathbf{a}_l$. In this case, we can estimate $(\hat{\boldsymbol{\theta}}_{|\mathcal{H}_1,l}^{\text{SSGL-ML}}, \hat{\mathbf{a}}_{|\mathcal{H}_1,l}^{\text{SSGL-ML}})$ by applying the results from (9)-(14) to (31), which results in replacing (31a) with

$$\boldsymbol{\theta}_{|\mathcal{H}_1,l}^{(t+1)} = (\mathbf{H}_l\mathbf{R}_l^{-1}\mathbf{H}_l + \mu_{1,l}\mathbf{B}_l + \zeta\mathbf{D}_l)^{-1}(\mathbf{R}_l^{-1}\mathbf{H}_l^T(\mathbf{z}_l - \mathbf{a}_l^{(t)}) + \zeta\mathbf{D}_l\mathbf{p}_l^{(t)}) \tag{38}$$

Algorithm 2: Distributed GL-GLRT in area l

Input: Fix detection threshold γ_l and step size ζ Set initial guess: $\boldsymbol{\theta}_{|\mathcal{H}_0,l}^{(0)}$, $\mathbf{s}_l^{(0)}$, and $\mathbf{p}_l^{(0)}$

for $t = 0, 1, \ldots$ **do**

  Update:

  $\boldsymbol{\theta}_{|\mathcal{H}_0,l}^{(t+1)} = (\mathbf{H}_l\mathbf{R}_l^{-1}\mathbf{H}_l + \mu_{1,l}\mathbf{B}_l + \zeta\mathbf{D}_l)^{-1}(\mathbf{R}_l^{-1}\mathbf{H}_l^T\mathbf{z}_l + \zeta\mathbf{D}_l\mathbf{p}_l^{(t)})$

  $\mathbf{s}_l^{(t+1)}(i) = \frac{1}{|\mathcal{A}_l^i|}\sum_{l\in\mathcal{A}_l^i}\boldsymbol{\theta}_{|\mathcal{H}_0,l}^{(t+1)}[i],\ \forall i$ with $\mathcal{A}_l^i \neq \emptyset$

  $\mathbf{p}_l^{(t+1)}(i) = \mathbf{p}_l^{(t)}(i) + \mathbf{s}_l^{(t+1)}(i) - \frac{\boldsymbol{\theta}_{|\mathcal{H}_0,l}^{(t)}(i) - \mathbf{s}_l^{(t)}(i)}{2},\ \forall i$ with $\mathcal{A}_l^i \neq \emptyset$

Set $\hat{\boldsymbol{\theta}}_{|\mathcal{H}_0,l}^{\text{SSGL-ML}} = \boldsymbol{\theta}_{|\mathcal{H}_0,l}^{(t+1)}$ **if** $-\mathcal{Q}_l^{SSGL}(\hat{\boldsymbol{\theta}}_{|\mathcal{H}_0,l}^{SSGL\text{-}ML}, \mathbf{0}) > \gamma_l$ **then**

  return "The area is under an FDI attack"

else

  return "The area is under normal operation"

and adding

$$\mathbf{a}_{|\mathcal{H}_1,l}^{(t+1)} = (\mathbf{R}_l^{-1} + \mu_{2,l}\mathbf{M}_l)^{-1}\mathbf{R}_l^{-1}(\mathbf{z}_l - \mathbf{H}_l\boldsymbol{\theta}_l^{(t+1)}). \tag{39}$$

Note that steps (31b)-(31c) are not modified, ensuring that the agreement between shared states is unrelated to the local functions $\mathcal{Q}_l$. Moreover, the inclusion of the term $\zeta\mathbf{D}_l$ is motivated by the same reasons as in (32) and (37). The distributed SS-GLRT is summarized in Algorithm 1.

Moreover, from (20) and (26) we observe that the GL-GLRT, which is a special case of the SSGL-GLRT, can be expressed as $T^{\text{GL-GLRT}} = \mathcal{Q}^{SSGL}(\hat{\boldsymbol{\theta}}_{|\mathcal{H}_0}^{\text{SSGL-ML}}, \mathbf{0})$. Similar to the SSGL-GLRT, the GL-GLRT can be adjusted for the distributed scenario by applying L detectors, represented as $T_l^{\text{GL-GLRT}}$, where each test is performed in the appropriate control center. These detectors are defined as

$$T_l^{\text{GL-GLRT}} = \mathcal{Q}_l^{SSGL}(\hat{\boldsymbol{\theta}}_{|\mathcal{H}_0,l}^{\text{SSGL-ML}}, \mathbf{0}),\ l = 1, \ldots, L,$$

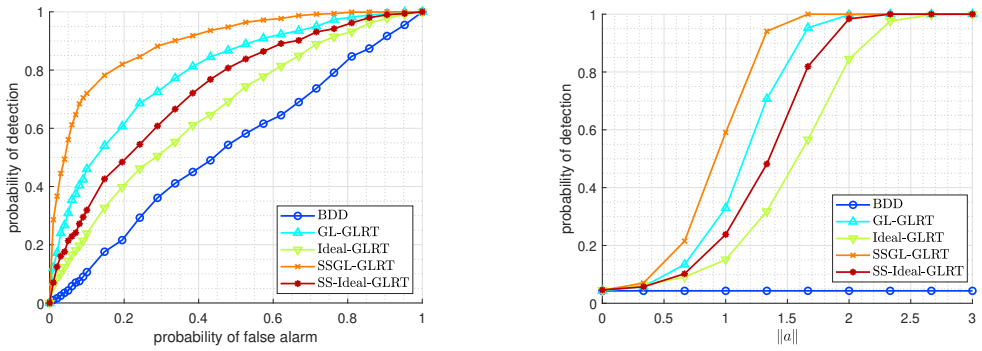
where $\hat{\boldsymbol{\theta}}_{|\mathcal{H}_0,l}^{\text{SSGL-ML}}$ estimation is described in (37). The distributed GL-GLRT is summarized in Algorithm 2.

5 Simulations: IEEE 57-Bus Test Case

The performance of the SSGL-GLRT from (21) is evaluated and compared with the following detectors:

1. The $J(\boldsymbol{\theta})$ test in (24) [1], which is the conventional BDD method.
2. GSP-based methods: the GL-GLRT in (26). and the Ideal-GLRT introduced in [9, 33].
3. The SSGL-GLRT obtained by using $\mathbf{B} = f^{\frac{1}{2}}(\boldsymbol{\Phi}_B)$ in (21), where $f(\boldsymbol{\Phi}_B) = 1 + 99 \times f^{\text{GHPF}}(\boldsymbol{\Phi}_B)$, which is the perturbed ideal GHPF defined in (28).

These methods were selected to demonstrate the advantage of incorporating both the physical and the GSP information. For the SSGL-GLRT, SS-Ideal-GLRT, GL-GLRT, and Ideal-GLRT, we chose the regularization parameter $\mu_1 = 900$; in addition, for the SSGL-GLRT and SS-Ideal-GLRT we also set $\mu_2 = 10$. We conducted $1,000$ Monte-Carlo simulations based on the IEEE 57-bus test case network using the DC-PF model in (2) to evaluate performance. For each trial, we randomly drew the load demand from a Gaussian distribution with the mean set to the load values provided in the test case. We computed the system states using the Matpower command $runpf(\cdot)$ [50]. We set the noise covariance matrix to $\mathbf{R} = \sigma^2\mathbf{I}$ with $\sigma^2 = 0.01$. We generated an unobservable FDI attack using (5) with $c_{33} \neq 0$, and then normalized it to satisfy $||\mathbf{a}|| = 1$. In addition, we defined the set of secured sensors $\mathcal{S}$ by constraining 80 power measurements (36% of the total measurements) such that it was ensured that the state variables in the generator buses and their first-order



(a) (b)

Fig. 2: The probability of detection is measured versus: (a) the probability of false alarm (ROC), and (b) the strength of the attack $\|\mathbf{a}\|$

.

neighbors are not affected by the attack. This set includes the power injection measurements in these buses and the power flow measurement in the lines entering these buses.

The performance of the different detectors is exhibited in Fig. 2. In Fig. 2.(a), the receiver operating characteristic (ROC) curves demonstrate the balance between the probability of detection and the probability of false alarms. The results show that the proposed SSGL-GLRT outperforms all other detectors in terms of the probability of detection for any level of false alarm probability. In particular, the inclusion of prior information about protected measurements gives the SSGL-GLRT an advantage over the GL-GLRT. Similarly, the SS-Ideal-GLRT, which benefits from incorporating the additional information on the locations of the secured sensors, outperforms the Ideal-GLRT. The results also show that detectors based on the smoothness of the states, i.e., the SSGL-GLRT and GL-GLRT, perform better than those based on the graph-bandlimited assumption, i.e., the SS-Ideal-GLRT and Ideal-GHPF. This is because the smoothness assumption provides a better description of the states' behavior than the graph-bandlimited assumption. Finally, it can be seen that the that the conventional BDD method - the $J(\boldsymbol{\theta})$ test - has the same power as random chance ("coin flipping"). Thus, it cannot detect the unobservable FDI attack, as expected.

In Fig. 2.(b) the detection probability is shown versus the attack strength, which is measured by $\|\mathbf{a}\|$. As expected, it can be seen that the detection probability of all the detectors except the BDD detector increases with an increase in $\|\mathbf{a}\|$. In a similar manner to Fig. 2. (a), it can be observed that incorporating the additional information on the locations of the secured sensors improves the probability of detection, where the SSGL-GLRT and SS-Ideal-GLRT outperforms the GL-GLRT and the Ideal-GLRT, respectively. Moreover, it can be observed that the SSGL-GLRT shows the best performance. Finally, as expected, the BDD detector fails to detect the unobservable FDI attack for any selection of $\|\mathbf{a}\|$ presented.

6 Conclusions

We introduce SSGL-GLRT, which is a new detection method against FDI attacks based on the well-known GLRT. The SSGL-GLRT is derived while incorporating knowledge of secured sensors' locations and graph smoothness properties of power system state variables. We provide a generalization of the method that allows the use of different high-pass GSP filters instead of using the graph smoothness measure. Moreover, we also consider the case where the power system is operated in a distributed manner and provide the distributed SSGL-GLRT detector. Numerical simulation show that incorporating the knowledge of the locations of the secured sensors alongside the graph smoothness properties in the design of the detector significantly improves

the detection capabilities against FDI attacks. Future work may focus on expanding the proposed detector to the alternating current (AC) power flow model, which is often used in power systems.

Appendix: Concavity of $\mathcal{Q}(\boldsymbol{\theta}, \mathbf{a})$

In order to show that the function $\mathcal{Q}(\boldsymbol{\theta}, \mathbf{a})$ from (8) is a concave function w.r.t $\boldsymbol{\theta}$ and $\mathbf{a}$, we need to show that the Hessian matrix of the second-order partial derivatives of $-\mathcal{Q}(\boldsymbol{\theta}, \mathbf{a})$ is a positive semidefinite matrix. It can be seen that the Hessian matrix of $-\mathcal{Q}(\boldsymbol{\theta}, \mathbf{a})$ w.r.t. the vector $[\boldsymbol{\theta}^T, \mathbf{a}^T]^T$ is

$$\begin{pmatrix} \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}+\mathbf{B} & \mathbf{H}^T\mathbf{R}^{-1} \\ \mathbf{R}^{-1}\mathbf{H} & \mathbf{R}^{-1}+\mathbf{M} \end{pmatrix} = \begin{pmatrix} \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} & \mathbf{H}^T\mathbf{R}^{-1} \\ \mathbf{R}^{-1}\mathbf{H} & \mathbf{R}^{-1} \end{pmatrix} + \begin{pmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{M} \end{pmatrix}.$$

The Hessian is a sum of two matrices. In the following, we show that each one of these matrices is positive semidefinite, which implies that the Hessian is a positive semidefinite matrix. First, it can be seen that the matrix $\begin{pmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{M} \end{pmatrix}$ is a positive semidefinite matrix because it is a block diagonal matrix of two positive semidefinite matrices (see the definitions of $\mathbf{B}$ and $\mathbf{M}$ in (1) and (8), respectively). Second, the matrix $\begin{pmatrix} \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} & \mathbf{H}^T\mathbf{R}^{-1} \\ \mathbf{R}^{-1}\mathbf{H} & \mathbf{R}^{-1} \end{pmatrix}$ is a positive semidefinite matrix since it can be verified that its Schur complement,

$$\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} - \mathbf{H}^T\mathbf{R}^{-1}\mathbf{R}\mathbf{R}^{-1}\mathbf{H} = \mathbf{0},$$

is a positive semidefinite matrix [13].

Acknowledgments

This work was supported in part by the Next Generation Internet (NGI) program, the Jabotinsky Scholarship from the Israel Ministry of Technology and Science, the Israel Ministry of National Infrastructure, Energy, National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. RS-2023-00210018), NSF grants CNS-2148128, EPCN-2144634, EPCN-2231350, and by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy under the Solar Energy Technology Office Award Number DE-EE0008769. The views expressed herein do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Bibliography

[1] Abur, A., Gomez-Exposito, A.: Power System State Estimation: Theory and Implementation. Marcel Dekker (2004)

[2] Bi, S., Zhang, Y.J.: Graphical methods for defense against false-data injection attacks on power system state estimation. IEEE Trans. Smart Grid **5**(3), 1216–1227 (2014)

[3] Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine learning **3**(1), 1–122 (2011)

[4] Dabush, L., Kroizer, A., Routtenberg, T.: State estimation in partially observable power systems via graph signal processing tools. Sensors (MDPI) **23**(3), 1387 (2023)

[5] Dabush, L., Routtenberg, T.: Detection of false data injection attacks in unobservable power systems by Laplacian regularization. In: IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM). pp. 415–419 (2022)
[6] Deng, R., Xiao, G., Lu, R.: Defending against false data injection attacks on power system state estimation. IEEE Trans. Ind. Informat. **13**(1), 198–207 (2015)
[7] Dong, X., Thanou, D., Frossard, P., Vandergheynst, P.: Learning Laplacian matrix in smooth graph signal representations. IEEE Trans. Signal Processing **64**(23), 6160–6173 (Dec 2016)
[8] Drayer, E., Routtenberg, T.: Detection of false data injection attacks in power systems with graph fourier transform. In: Glob. Conf. Sig. and Info. Process. (GlobalSIP). pp. 890–894 (2018)
[9] Drayer, E., Routtenberg, T.: Detection of false data injection attacks in smart grids based on graph signal processing. IEEE Syst. J. (2019)
[10] Esmalifalak, M., Liu, L., Nguyen, N., Zheng, R., Han, Z.: Detecting stealthy false data injection using machine learning in smart grid. IEEE Syst. J. **11**(3), 1644–1652 (2017)
[11] Hasnat, M.A., Rahnamay-Naeini, M.: A graph signal processing framework for detecting and locating cyber and physical stresses in smart grids. IEEE Trans. Smart Grid **13**(5), 3688–3699 (2022)
[12] He, Y., Mendis, G.J., Wei, J.: Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism. IEEE Trans. Smart Grid **8**(5), 2505–2516 (2017)
[13] Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, New York, NY, USA, 2nd edn. (2012)
[14] Jia, L., Kim, J., Thomas, R.J., Tong, L.: Impact of data quality on real-time locational marginal price. IEEE Trans. Power Syst. **29**(2), 627–636 (2014)
[15] Kalofolias, V.: How to learn a graph from smooth signals. In: Journal of Machine Learning Research (JMLR) (2016)
[16] Kay, S.M.: Fundamentals of Statistical Signal Processing: Detection Theory, vol. 2. Prentice Hall PTR, Englewood Cliffs (N.J.) (1998)
[17] Kekatos, V., Giannakis, G.B.: Distributed robust power system state estimation. IEEE Trans. Power Syst. **28**(2), 1617–1626 (2012)
[18] Kim, J., Bhela, S., Anderson, J., Zussman, G.: Identification of intraday false data injection attack on DER dispatch signals. In: 2022 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm). pp. 40–46 (2022)
[19] Kim, J., Tong, L.: On phasor measurement unit placement against state and topology attacks. In: SmartGridComm. pp. 396–401 (2013)
[20] Kim, T.T., Poor, H.V.: Strategic protection against data injection attacks on power grids. IEEE Trans. Smart Grid **2**(2), 326–333 (2011)
[21] Kosut, O., Jia, L., Thomas, R.J., Tong, L.: Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures. In: 2010 First IEEE International Conference on Smart Grid Communications. pp. 220–225 (2010)
[22] Kroizer, A., Routtenberg, T., Eldar, Y.C.: Bayesian estimation of graph signals. IEEE Trans. Signal Processing **70**, 2207–2223 (2022). https://doi.org/10.1109/TSP.2022.3159393
[23] Liang, G., Zhao, J., Luo, F., Weller, S.R., Dong, Z.Y.: A review of false data injection attacks against modern power systems. IEEE Trans. Smart Grid **8**(4), 1630–1638 (2017)
[24] Lin, J., Yu, W., Yang, X., Xu, G., Zhao, W.: On false data injection attacks against distributed energy routing in smart grid. In: Int. Conf. Cyber-Physical. Syst. pp. 183–192. IEEE Comput. Soc (2012)
[25] Liu, Y., Ning, P., Reiter, M.K.: False data injection attacks against state estimation in electric power grids. ACM Trans. Inf. Syst. Secur. **14**(1), 13 (2011)
[26] Minot, A., Lu, Y.M., Li, N.: A distributed Gauss-Newton method for power system state estimation. IEEE Trans. Power Syst. **31**(5), 3804–3815 (2015)
[27] Monticelli, A.: State Estimation in Electric Power Systems: A Generalized Approach, pp. 39–61,91–101,161–199. Springer US, Boston, MA (1999)
[28] Morgenstern, G., Routtenberg, T.: Structural-constrained methods for the identification of unobservable false data injection attacks in power systems. IEEE Access **10**, 94169–94185 (2022)

[29] Morgenstern, G., Kim, J., Anderson, J., Zussman, G., Routtenberg, T.: Protection against graph-based false data injection attacks on power systems. (2023), https://arxiv.org/abs/2304.10801
[30] Ortega, A., Frossard, P., Kovačević, J., Moura, J.M.F., Vandergheynst, P.: Graph signal processing: Overview, challenges, and applications. Proceedings of the IEEE **106**(5), 808–828 (May 2018). https://doi.org/10.1109/JPROC.2018.2820126
[31] Primadianto, A., Lu, C.N.: A review on distribution system state estimation. IEEE Trans. Power Syst. **32**(5), 3875–3883 (2016)
[32] Ramakrishna, R., Scaglione, A.: Grid-graph signal processing (Grid-GSP): A graph signal processing framework for the power grid. IEEE Trans. Signal Process. **69**, 2725–2739 (2021)
[33] Ramakrishna, R., Scaglione, A.: Detection of false data injection attack using graph signal processing for the power grid. In: 2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP). pp. 1–5. IEEE (2019)
[34] Routtenberg, T., Eldar, Y.C.: Centralized identification of imbalances in power networks with synchrophasor data. IEEE Trans. Power Syst. **33**(2), 1981–1992 (2017)
[35] Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: nonlinear phenomena **60**(1-4), 259–268 (1992)
[36] Sandryhaila, A., Moura, J.M.F.: Discrete signal processing on graphs: Frequency analysis. IEEE Trans. Signal Processing **62**(12), 3042–3054 (June 2014)
[37] Shaked, S., Routtenberg, T.: Identification of edge disconnections in networks based on graph filter outputs. IEEE Trans. Signal Inf. Process. Netw. (2021)
[38] Shereen, E., Ramakrishna, R., Dán, G.: Detection and localization of pmu time synchronization attacks via graph signal processing. IEEE Trans. Smart Grid **13**(4), 3241–3254 (2022)
[39] Shuman, D.I., Narang, S.K., Frossard, P., Ortega, A., Vandergheynst, P.: The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. IIEEE Signal Process. Mag. **30**(3), 83–98 (May 2013)
[40] Soltan, S., Mazauric, D., Zussman, G.: Analysis of failures in power grids. IEEE Control Netw. Syst. **4**(2), 288–300 (2017). https://doi.org/10.1109/TCNS.2015.2498464
[41] Soltan, S., Yannakakis, M., Zussman, G.: Power grid state estimation following a joint cyber and physical attack. IEEE Trans. Control. Netw. Syst. **5**(1), 499–512 (2016)
[42] Soltan, S., Yannakakis, M., Zussman, G.: React to cyber attacks on power grids. IEEE Transactions on Network Science and Engineering **6**(3), 459–473 (2018)
[43] Sridhar, S., Hahn, A., Govindarasu, M.: Cyber-physical system security for the electric power grid. Proceedings of the IEEE **100**(1), 210–224 (2011)
[44] Veith, E., Fischer, L., Tröschel, M., Nieße, A.: Analyzing cyber-physical systems from the perspective of artificial intelligence. In: Proceedings of the 2019 International Conference on Artificial Intelligence, Robotics and Control. pp. 85–95 (2019)
[45] Verdoja, F., Grangetto, M.: Graph Laplacian for image anomaly detection. Machine Vision and Applications **31**(1-2), 11 (2020)
[46] Vuković, O., Dán, G.: Security of fully distributed power system state estimation: Detection and mitigation of data integrity attacks. IEEE J. Sel. Areas Commun **32**(7), 1500–1508 (2014)
[47] Xie, L., Mo, Y., Sinopoli, B.: Integrity data attacks in power market operations. IEEE Trans. Smart Grid **2**(4), 659–666 (2011)
[48] Yuan, Y., Li, Z., Ren, K.: Quantitative analysis of load redistribution attacks in power systems. IEEE Trans. Parallel Distrib. Syst. **23**(9), 1731–1738 (2012)
[49] Zhu, X., Kandola, J.S., Lafferty, J., Ghahramani, Z.: Graph kernels by spectral transforms (2006)
[50] Zimmerman, R.D., Murillo-Sanchez, C.E., Thomas, R.J.: MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education. IEEE Trans. Power Syst. **26**(1), 12–19 (Feb 2011)