# Distributed Joint Cyber Attack Detection and State Recovery in Smart Grids

Ali Tajer*, Soummyar Kar*, H. Vincent Poor*, and Shuguang Cui†
* Department of Electrical Engineering, Princeton University, Princeton, NJ 08544
† Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843

***Abstract*—Dynamic state estimation in power systems plays a central role in controlling the system operations. State estimation, however, is vulnerable to malicious cyber attacks that contaminate the controller's observation through injecting false data into the system. It is, therefore, of paramount significance not only to detect the attacks, but also to recover the system state from the contaminated observation when an attack is deemed to be present. This paper offers a *distributed* framework for an optimal such *joint* attack detection and system recovery. This framework has two main features: 1) it provides the network operator with the freedom of striking any desired balance between attack detection and state recovery accuracies, and 2) it is distributed in the sense that different controlling agents distributed across the network carry out the attack detection and system recovery tasks through iterative *local* processing and message passing.**

## I. Introduction

### A. Background

State estimation in electricity grids was initially developed as a data processing algorithm for converting redundant meter readings, and possibly other available information, into an estimate of the state of the grid [1]. The state typically refers to bus voltage magnitudes and phase angles, and state estimation is a key function to build a real-time network model in the energy management system (EMS) [2], [3]. The main task of the traditional state estimators is conducted *centrally* in a control center with the following three functions [4]:

- Observability analysis: To determine whether a unique estimate can be found for the system state. This is generally performed prior to state estimation.
- State estimation: To determine the optimal estimate for the complex voltages at each bus based on real-time analog measurements.
- Bad data processing: To detect measurement errors, and identify and eliminate them if possible.

Conventional bad data detection techniques often determine gross measurement errors based on the measurement residuals [4]. While relatively effective against random noises, these conventional detectors lack the ability to detect *highly structured* bad data that conforms to the network topology and physical laws. This raises serious security concerns about intentional coordinated attacks that can tamper with the measurements without being detected. In [5] false data injection attacks against power grid state estimation were introduced. By leveraging the knowledge of the power network topology, it was shown that well-designed injected false data could bypass bad data detectors in today's control systems and become unobservable by the controllers. In [6] two possible approaches were proposed that aim at quantifying the effort that the attackers require for implementing this class of attacks. In [7] and [8] efforts have been made to develop computationally efficient algorithms to detect small but highly damaging structured attacks against state estimation.

This research was supported in part by the Air Force Office of Scientific Research under MURI Grant FA9550-09-1-0643 and in part by DTRA under Grant HDTRA-1-07-1-0037.

Besides the challenges related to unobservable attacks, electricity industry deregulation has led to the creation of many regional transmission organizations (RTOs) within a large interconnected power system [2]. Hence, a substantially larger number of smart elements are appearing in the grid such that the grid requires significantly more computational, processing, and communication resources. As such, both technical and institutional drivers suggest the need for more distributed information processing and control in power system operations [9]. For this purpose, distributed state estimation methods have been considered for decades with the initial goal of reducing the computational burden at the central control via distributing the centralized task across the system. A simple but effective approach is decomposition-and-merge, which is essentially a two-level hierarchical method [10], [11]. At the lower level, each local area independently runs its own estimator based on its local measurements, while at the higher level, the central coordinator receives the state estimation results, with boundary measurements if necessary, from the individual areas and then combines them to obtain a system-wide solution. The *star-like* hierarchical methods are also discussed in recent literature such as [12] and [13]. More recently, a hybrid distributed state estimation using the synchronized phasor measurements as well as the conventional measurements has been proposed [14], [15], [16].

While there exists a rich literature on state estimation as well as on attack detection, the interplay between these two closely-related problems has not been addressed yet; the analysis of this interplay is the purpose of this paper. We also address the considerations and limitations of distributed processing by introducing a distributed procedure that consists of iterative local processing combined with message passing and aims to achieve network-wide optimal attack detection and state estimation.

### B. Notation

We denote the $k$-dimensional Euclidean space by $\mathbb{R}^k$. The operator $\|\cdot\|$ denotes the standard Euclidean 2-norm when applied to vectors and denotes the induced 2-norm when applied to matrices, which in the latter case is equivalent to the matrix spectral radius for symmetric matrices. $\mathsf{P}(\cdot)$ and $\mathsf{E}(\cdot)$ denote probability and expected value, respectively.

We recall some notation from graph theory [17] to be used in the description of the distributed algorithms in Section V. The communication among the RTOs (also referred to as subsystems or agents or nodes) will be modeled by an *undirected* graph $G = (V, E)$, in which $V = [1, \cdots, N]$ denotes the set of nodes with cardinality $|V| = N$, and $E$ is the set of edges with cardinality $|E| = M$. The unordered pair $(n, l)$ belongs to $E$ if there exists an edge between nodes $n \in V$ and $l \in V$. We consider only simple graphs, i.e., graphs devoid of self-loops and multiple edges. A graph is connected if there

exists a path[1], between each pair of nodes. $\forall n \in \{1,\dots,N\}$ the neighborhood of node $n$ is defined as

$$\Omega_n \triangleq \{l \in V \,|\, (n,l) \in E\} \ . \tag{1}$$

Throughout the paper, we assume that all the random objects are defined on a common measurable space. Also, all inequalities involving random variables are to be interpreted almost surely (a.s.).

## II. Malicious Data Injection Model

Assuming linearized system dynamics, a general false data injection attack is modeled as [5]:

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{Z} + \boldsymbol{B} \ , \tag{2}$$

where $\boldsymbol{Y}$ is the observation vector, $\boldsymbol{X}$ represents the power system state vector, and the matrix $\boldsymbol{H}$ denotes the Jacobian matrix that models the linearized dynamics at some operating point. Also $\boldsymbol{Z}$ and $\boldsymbol{B}$ account for the observation noise and the false data injected by the attacker, respectively. The attack could be launched by one single autonomous attacker, or by a group of coordinated attackers that inject the false data based on their collective information about the network. From the attacker's point of view, the design of the *optimal*[2] $\boldsymbol{B}$ depends on

1) the extent of the topology and system instantaneous state (abstracted in the Jacobian $\boldsymbol{H}$) known to the attackers; and
2) the placement of secure sensors, which are the ones with sufficient protection that cannot be manipulated [18].

For setting up the attack detection/estimation problem from the network operator's viewpoint, we assume that in the general situation, the network operator has no knowledge about the attacker's strategy, which is a function of the network topology information available to the attacker. Also, for definiteness, we assume that the system topology and dynamics $\boldsymbol{H}$ are fully and instantaneously known to the network operator.

## III. Information Constraints of the Attacker

The effectiveness of an attack depends heavily on how informed the attacker is about the network dynamics. We indeed assume that the attacker always selects the *best* attack strategy based on their information. Due to security constraints, some critical sensors of the network cannot be compromised and the attacker has (or the coordinated attackers collectively have) only partial information of $\boldsymbol{H}$. Moreover, even the elements of $\boldsymbol{H}$ corresponding to the unsecured sensors might not be perfectly known to the attacker and only noisy estimates of are available to the attacker. Such lack of perfect knowledge is due to the unpredictable fluctuations in the system load and the generated power. The attacker, however, can potentially collect some side statistical information about the unsecured elements of $\boldsymbol{H}$ by observing and accumulating the system fluctuations pattern over time.

In order to account for the above two sources of uncertainty about $\boldsymbol{H}$ (estimation inaccuracy and secure nodes) the instantaneous Jacobian matrix can be decomposed as

$$\boldsymbol{H} = \bar{\boldsymbol{H}} + \Delta\boldsymbol{H}_1 + \Delta\boldsymbol{H}_2 \ , \tag{3}$$

where $\bar{\boldsymbol{H}}$ denotes the part of $\boldsymbol{H}$ fully known to the attacker, and $\Delta\boldsymbol{H}_1$ and $\Delta\boldsymbol{H}_2$ represent the unknown instantaneous fluctuations. $\Delta\boldsymbol{H}_1$ captures the part of $\boldsymbol{H}$ that is not known to the attacker instantaneously, but for which the statistics of its fluctuations over time are known. Finally, $\Delta\boldsymbol{H}_2$ models the part of the fluctuations that, due to the protections in the system, are neither instantly nor statistically known to the attacker. Therefore, we assume that while the network operator has access to the entire $\boldsymbol{H}$, the attacker has access to $\bar{\boldsymbol{H}}$ and $\Delta\boldsymbol{H}_1$ only.

The attacker aims to design an attack strategy $\phi$ based on its knowledge of $\boldsymbol{H}$. In general, an attack strategy $\phi$ may be viewed as a measurable function, mapping $\bar{\boldsymbol{H}}$ and $\Delta\boldsymbol{H}_1$ to the attack vector as $\boldsymbol{B} = \phi(\bar{\boldsymbol{H}}, \Delta\boldsymbol{H}_1)$. Further, such an attack function is constrained by the fact some sensors are secured and cannot be compromised, i.e., $\phi(\bar{\boldsymbol{H}}, \Delta\boldsymbol{H}_1)$ may not influence all the sensor measurements. By taking into account that the attack strategy $\phi$ is not known to the network operator and also by invoking the randomness of $\Delta\boldsymbol{H}_1$, the attack vector is clearly a random vector from the viewpoint of the network operator. We denote the probability density function of the attack vector by $\pi_\phi(\boldsymbol{B})$, which is a function of the attack strategy $\phi$.

Under the circumstances that the attacker has full knowledge of $\boldsymbol{H}$, the attacker can become undetectable if it aligns its injections within the range space of $\boldsymbol{H}$ [5]. With partial knowledge of $\boldsymbol{H}$, however, the attacker is not guaranteed to align its injections within the range space of $\boldsymbol{H}$ and, in fact, it leaves a non-zero projection in the kernel of $\boldsymbol{H}$ *almost surely*. Motivated by this premise, the best strategy of the attacker is formalized in the following theorem. The underlying justification for the best strategy relates to the fact that the attacker tries to maximize the likelihood that its energy is more focused in the range space of $\boldsymbol{H}$ rather than its null space.

*Theorem 1:* Among all attack vectors with a given energy $E_B$, i.e., $\|\boldsymbol{B}\|_2^2 = E_B$, the one with the smallest projection in the null space of $\boldsymbol{H}$ belongs to the range space of $\bar{\boldsymbol{H}} + \Delta\boldsymbol{H}_1$, i.e.,

$$\arg \min_{\boldsymbol{B}:\ \|\boldsymbol{B}\|_2^2 = E_B} \mathsf{E}_{\Delta\boldsymbol{H}_2}[\|\boldsymbol{B}\|^2] \ \in \ \text{range}\,(\bar{\boldsymbol{H}} + \Delta\boldsymbol{H}_1) \ . \tag{4}$$

Given the direction of the attack vector (from the theorem above), the attack vector $\boldsymbol{B}$ can be decomposed as

$$\boldsymbol{B} = (\bar{\boldsymbol{H}} + \Delta\boldsymbol{H}_1)\tilde{\boldsymbol{B}} \ , \tag{5}$$

based on which (2) becomes

$$\boldsymbol{Y} = \boldsymbol{H}(\boldsymbol{X} + \tilde{\boldsymbol{B}}) + \boldsymbol{Z} - \Delta\boldsymbol{H}_2\tilde{\boldsymbol{B}} \ . \tag{6}$$

Therefore, the residual of $\boldsymbol{B}$ in the kernel of $\boldsymbol{H}$, which is determined by $\Delta\boldsymbol{H}_2\tilde{\boldsymbol{B}}$, can be reliably detected and estimated based on the combined detection and estimation framework developed in Section IV. When certain conditions are satisfied this reliable estimate of $\tilde{\boldsymbol{B}}$ can lead to a reliable estimate of the attack vector $\boldsymbol{B}$.

## IV. Centralized Joint Attack Detection and Injection Estimation

In this section we consider the hypothetical scenario in which the power system has a central controller with a global and instantaneous access to all the network topology abstracted in $\boldsymbol{H}$. We solve the attack detection and system recovery problem in this setting, which in turn provides an upper bound on the performance to be expected from a system with a set of non-fully coordinated controllers distributed across the network. In the next section we offer a distributed procedure for implementing the optimal detectors and estimators that are presented in this section.

[1] A path between nodes $n$ and $l$ of length $m$ is a sequence $(i_0 = n, i_1, \cdots, i_m = l)$ of vertices, such that, $\forall k \in \{0,\dots,m-1\}$ : $(i_k, i_{k+1}) \in E$.

[2] The term optimal from the attacker's viewpoint refers to the best possible attack strategy that maximizes a relevant metric, say the probability of attack detection error.

### A. Sufficient Statistics

The objective of the defender (system operator) is to simultaneously detect an injection attack and recover the estimate of the system state from the attack-contaminated observations. The latter could be carried out through estimating the attack vector $\boldsymbol{B}$ and eliminating it from the observation vector $\boldsymbol{Y}$. The problem of joint detection and estimation is challenging, and we present general strategies in the subsequent subsections with various claims of optimality (from a defender's viewpoint).

In this subsection, we present a theoretical understanding of the delectability of $\boldsymbol{B}$. Also, we will see that not all components of the observation vector $\boldsymbol{Y}$ contain meaningful information about the attack vector $\boldsymbol{B}$. As a result, we will seek to identify the portion of $\boldsymbol{Y}$ that contains all information about $\boldsymbol{B}$. In statistical language, such a portion (or functional) of $\boldsymbol{Y}$ containing the attack information is referred to as a sufficient statistic for the detection-estimation problem, and, is, in fact, the only part of $\boldsymbol{Y}$ meaningful to a decision making process.

The difficulty in detection and estimation tasks mainly stems from the fact that the system state $\boldsymbol{X}$ is unknown to the network operator (and in fact is sought to be estimated). This, in turn, could be exploited by a smart attacker, who injects the data into the subspace spanned by the column vectors of $\boldsymbol{H}$, making the attack vector $\boldsymbol{B}$ indistinguishable from the state $\boldsymbol{X}$. In other words, by defining $\boldsymbol{y}^\dagger$, $\boldsymbol{b}^\dagger$, and $\boldsymbol{z}^\dagger$ as the projections of $\boldsymbol{Y}$, $\boldsymbol{B}$ and $\boldsymbol{Z}$, respectively, on the subspace spanned by $\boldsymbol{H}$, we find that $\boldsymbol{y}^\dagger$ given by

$$\boldsymbol{y}^\dagger = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{b}^\dagger + \boldsymbol{z}^\dagger \ , \tag{7}$$

does not leave a detectable trace of the attack as $\boldsymbol{y}^\dagger$ carries information only about the superposition $\boldsymbol{H}\boldsymbol{X} + \boldsymbol{b}^\dagger$. More specifically, $\boldsymbol{X}$ being unknown to the network operator makes recovering $\boldsymbol{b}^\dagger$ and $\boldsymbol{X}$ from the superposition $\boldsymbol{H}\boldsymbol{X} + \boldsymbol{b}^\dagger$ impossible. Hence, all the data about $\boldsymbol{B}$ will be embedded in its projection on the null space (kernel) of the matrix $\boldsymbol{H}$. The projection of the observation vector $\boldsymbol{Y}$ on the null space of $\boldsymbol{H}$ is given by

$$\boldsymbol{y} = \boldsymbol{b} + \boldsymbol{z} \ , \tag{8}$$

where $\boldsymbol{b}$ and $\boldsymbol{z}$ are the projections of the vectors $\boldsymbol{B}$ and $\boldsymbol{Z}$, respectively, on the null space of $\boldsymbol{H}$.

*Remark 1:* Given that the network operator has full knowledge of the network instantiation, by using the knowledge of $\boldsymbol{H}$ and $\Delta\boldsymbol{H}_2$ it can recover $\tilde{\boldsymbol{B}}$ defined in (5) from the estimate on $\boldsymbol{b}$. Feasibility of this procedure is mainly governed by whether the linear transformation of $\boldsymbol{B}$ to $\tilde{\boldsymbol{B}}$ and then to $\boldsymbol{b}$ is reversible. The feasibility problem of interest can be abstracted as analyzing and ensuring the non-singularity conditions of the pseudo-inverses of the matrices involved in the transforming $\boldsymbol{B}$ to $\boldsymbol{b}$. One approach for ensuring such non-singularity is the method developed in [18], which secures some sensors in the network. The data of the secured sensors is neither revealed to nor compromised by the attacker. Implementing such secure sensors provides certain structures for $\Delta\boldsymbol{H}_2$ that guarantee the inverse transformation of interest is non-singular, which in turn ensures that $\boldsymbol{B}$ can be recovered from $\boldsymbol{b}$ uniquely.

*Remark 2:* Given the attack-contaminated observation $\boldsymbol{Y}$, after estimating $\boldsymbol{B}$, the network operator (controller) can obtain the attack-free observations $\tilde{\boldsymbol{Y}} \triangleq \boldsymbol{Y} - \boldsymbol{B}$ by subtracting the attack vector $\boldsymbol{B}$ from the observation $\boldsymbol{Y}$. Given the reliable attack-free observations $\tilde{\boldsymbol{Y}}$, the controller can employ a wide range of estimators (e.g., the minimum mean-square error estimator) to recover the state vector $\boldsymbol{X}$ from $\boldsymbol{Y} - \boldsymbol{B}$.

### B. Joint Detection and Estimation

Upon obtaining $\boldsymbol{y}$ as the projection of the observation vector onto the null space of $\boldsymbol{H}$, the objective is to detect the presence of an attack and simultaneously estimate the injected false data, when an attack is deemed to be present. This combined detection and estimation problem can be posed as the following composite binary hypothesis testing problem with $\mathsf{H}_0$ reflecting the no attack hypothesis and $\mathsf{H}_1$ representing the attack hypothesis.

$$\begin{aligned} \mathsf{H}_0 : \quad & \boldsymbol{y} \sim f_0(\boldsymbol{y}) \\ \mathsf{H}_1 : \quad & \boldsymbol{y} \sim f_1(\boldsymbol{y} \mid \boldsymbol{b}), \quad \text{and} \quad \boldsymbol{b} \sim \pi(\boldsymbol{b}) \ , \end{aligned} \tag{9}$$

where $f_0(\boldsymbol{y})$, $f_1(\boldsymbol{y} \mid \boldsymbol{b})$, and $\pi(\boldsymbol{b})$ are known probability density functions (pdfs). In particular, we assume that under $\mathsf{H}_0$ (ie., no attack, or $\boldsymbol{b} = \boldsymbol{0}$), the distribution of $\boldsymbol{y}$, which inherits all its randomness from the random noise, is fully known. It is noteworthy that the system model in (2) and the pertinent analysis provided in the next sections hold for any noise model, including the Gaussian model. Also, under $\mathsf{H}_1$ (i.e., an attack exists, or $\boldsymbol{b} \neq \boldsymbol{0}$), the distribution of $\boldsymbol{y}$ depends on the unknown attack vector $\boldsymbol{b}$, for which we know a prior distribution $\pi(\boldsymbol{b})$ based on the attack patterns accumulated over time. We seek to devise a mechanism that distinguishes between $\mathsf{H}_0$ and $\mathsf{H}_1$ reliably, and furthermore, whenever it decides in favor of $\mathsf{H}_1$ provides an accurate estimate for $\boldsymbol{b}$.

There are two general approaches in the literature for treating combined estimation and detection problems. In one direction, the combined problem is decomposed into two subproblems and each treated optimally. For instance one can use the Neyman-Pearson optimum test for detection and the optimum Bayesian estimator for parameter estimation to solve the two subproblems. Treating each subproblem independently does not necessarily yield the optimum overall performance. The second methodology employs the generalized likelihood ratio test (GLRT) which performs detection and estimation at the same time. Neither approach is capable of emphasizing each subproblem according to the needs of the corresponding application.

Here we formulate the combined problem in a more natural way for the problem in hand. In particular, when an attack occurs in the power network, the ultimate objective of the network operator is beyond a reliable detection of the attack. In fact, detecting the attack will be used as an intermediate step towards obtaining a reliable estimate about the injected false data, which in turn facilitates eliminating the disruptive effects of the false data. Serving this purpose necessitates that the operator obtain the *best* estimate about the false data injected. Therefore, assuring good estimation performance lies at the core of our combined estimation and detection problem. To account for the significance of estimation quality, we define an estimation performance measure and seek to optimize it while ensuring, in parallel, satisfactory detection performance. More precisely, we minimize the estimation-related costs subject to appropriate constraints on the tolerable levels of detection errors (missed-detection and false-alarm). This approach will lead to a novel combined test that provides the operator with the freedom to strike any desired balance between the quality of estimation and detection.

### C. Definitions

We denote the true hypothesis and the decision of the detector by $\mathsf{T} \in \{\mathsf{H}_0, \mathsf{H}_1\}$ and $\mathsf{D} \in \{\mathsf{H}_0, \mathsf{H}_1\}$, respectively. Therefore, the probabilities of missed-detection and false alarm are given by

$$\begin{aligned} \mathsf{P}_{\text{mis}} &= \mathsf{P}(\mathsf{D} = \mathsf{H}_0 \mid \mathsf{T} = \mathsf{H}_1) \ , \\ \text{and} \quad \mathsf{P}_{\text{fa}} &= \mathsf{P}(\mathsf{D} = \mathsf{H}_1 \mid \mathsf{T} = \mathsf{H}_0) \ , \end{aligned}$$

respectively. Once we decide that the observation $\boldsymbol{y}$ is drawn from hypothesis $\mathsf{H}_1$, we are also interested in providing an estimate $\hat{\boldsymbol{b}}(\boldsymbol{y})$ for $\boldsymbol{b}$. We capture the quality of the estimate by defining a cost function $C(\boldsymbol{b}, \hat{\boldsymbol{b}}(\boldsymbol{y}))$ corresponding to estimate $\hat{\boldsymbol{b}}(\boldsymbol{y})$. Two popular cost functions corresponding to the minimum mean-square error (MMSE) and maximum a-posteriori probability (MAP) estimation criteria are

$$\text{MMSE}: \quad C(\boldsymbol{b}, \boldsymbol{u}) = \|\boldsymbol{b} - \boldsymbol{u}\|^2 \ ,$$

$$\text{and} \quad \text{MAP}: \quad C(\boldsymbol{b}, \boldsymbol{u}) = \begin{cases} 0 & \|\boldsymbol{b} - \boldsymbol{u}\| \leq \delta \ll 1 \\ 1 & \text{otherwise} \end{cases} \ .$$

For a given cost function $C(\boldsymbol{u}, \boldsymbol{b})$ we also define the average *posterior* cost function, which assesses the estimation error cost after observing $\boldsymbol{y}$. The posterior cost function is

$$\begin{aligned} C_{\mathrm{p}}(\boldsymbol{u} \mid \boldsymbol{y}) &\triangleq \mathsf{E}_{\boldsymbol{b}}[C(\boldsymbol{b}, \boldsymbol{u}) \mid \boldsymbol{y}] \\ &= \frac{\int C(\boldsymbol{b}, \boldsymbol{u}) f_1(\boldsymbol{y} \mid \boldsymbol{b}) \pi(\boldsymbol{b}) \, d\boldsymbol{b}}{\int f_1(\boldsymbol{y} \mid \boldsymbol{b}) \pi(\boldsymbol{b}) \, d\boldsymbol{b}} \\ &= \frac{\int C(\boldsymbol{b}, \boldsymbol{u}) f_1(\boldsymbol{y} \mid \boldsymbol{b}) \pi(\boldsymbol{b}) \, d\boldsymbol{b}}{f_1(\boldsymbol{y})} \ , \end{aligned} \tag{10}$$

where the expectation is with respect to $\boldsymbol{b}$ for given $\boldsymbol{y}$. Therefore, the minimum posterior cost is

$$C_{\mathrm{p}}^*(\boldsymbol{y}) \triangleq \inf_{\boldsymbol{u}} C_{\mathrm{p}}(\boldsymbol{u} \mid \boldsymbol{y}) \ , \tag{11}$$

and the minimizer of the posterior cost, which is the well-known Bayesian estimator, is [19, pp. 142]

$$\hat{\boldsymbol{b}}^*(\boldsymbol{y}) \triangleq \arg\inf_{\boldsymbol{u}} C_{\mathrm{p}}(\boldsymbol{u} \mid \boldsymbol{y}) \ . \tag{12}$$

Note that $C_{\mathrm{p}}^*(\boldsymbol{y})$ plays an important role in designing the estimator and detector as it captures the quality of estimation.

### D. Problem Formulation

We next propose a performance measure that incorporates both estimation and detection qualities. Optimizing this performance measure defines the detection strategy and the estimator. In order to model the detection strategy we deploy a randomized test $\{\delta_0(\boldsymbol{y}), \delta_1(\boldsymbol{y})\}$, where $\delta_i(\boldsymbol{y})$, for $i = 1, 2$, denotes the randomization probability for deciding in favor of $\mathsf{H}_i$. Clearly $\delta_i(\boldsymbol{y}) \geq 0$ and $\delta_0(\boldsymbol{y}) + \delta_1(\boldsymbol{y}) = 1$. For given randomization probabilities $\{\delta_0(\boldsymbol{y}), \delta_1(\boldsymbol{y})\}$ and an estimator $\hat{\boldsymbol{b}}$, we define the following performance measure.

$$\begin{aligned} J(\delta_0, \delta_1, \hat{\boldsymbol{b}}) &\triangleq \mathsf{E}_{1,\boldsymbol{b}}[C(\boldsymbol{b}, \hat{\boldsymbol{b}}(\boldsymbol{y})) \mid \mathsf{D} = \mathsf{H}_1] \\ &= \frac{\mathsf{E}_{1,\boldsymbol{b}}[C(\boldsymbol{b}, \hat{\boldsymbol{b}}(\boldsymbol{y})) \, \mathbb{1}_{\{\mathsf{D}=\mathsf{H}_1\}}]}{\mathsf{P}(\mathsf{D} = \mathsf{H}_1 \mid \mathsf{T} = \mathsf{H}_1)} \ , \end{aligned} \tag{13}$$

where the expectation is over $\boldsymbol{b}$ and $\boldsymbol{y}$ under $\mathsf{H}_1$. Note that the estimate $\hat{\boldsymbol{b}}(\boldsymbol{y})$ should be provided only when we decide in favor of $\mathsf{H}_1$. Moreover, the average of $C(\boldsymbol{b}, \hat{\boldsymbol{b}}(\boldsymbol{y}))$ is reasonable only under the alternative hypothesis $\mathsf{H}_1$ as there is no true $\boldsymbol{b}$ under nominal $\mathsf{H}_0$. Therefore, we compute the average estimation cost over the event $\{\mathsf{D} = \mathsf{H}_1\}$, which is the only case for which an estimate is available. In order to incorporate mechanisms for controlling the detection performance, we constrain the estimation cost optimization problem with the upper bound constraints on the missed-detection and false-alarm probabilities as:

$$\mathsf{P}_{\mathrm{mis}} \leq \beta \quad \text{and} \quad \mathsf{P}_{\mathrm{fa}} \leq \alpha \qquad \text{for} \quad \alpha, \beta \in (0, 1) \ .$$

Therefore, we can pose the joint estimation and detection problem as the following optimization problem:

$$\mathcal{P} \triangleq \begin{cases} \min_{\delta_0, \delta_1, \hat{\boldsymbol{b}}} & J(\delta_0, \delta_1, \hat{\boldsymbol{b}}) \\ \text{s.t.} & \mathsf{P}_{\mathrm{mis}} \leq \beta \\ & \mathsf{P}_{\mathrm{fa}} \leq \alpha \end{cases} \ . \tag{14}$$

### E. Solution

In this section we briefly provide the structures of the estimators and the detectors. The proofs are omitted for the sake of brevity and follow the same line of arguments as the results in [20].

*Lemma 1:* Solving (14) can be decoupled into minimizing over $\hat{\boldsymbol{b}}$ and over $\{\delta_0, \delta_1\}$ separately. In other words, we can rewrite $\mathcal{P}$ as

$$\mathcal{P} = \begin{cases} \min_{\delta_0, \delta_1, \hat{\boldsymbol{b}}} & \tilde{J}(\delta_0, \delta_1) \\ \text{s.t.} & \mathsf{P}_{\mathrm{mis}} \leq \beta \\ & \mathsf{P}_{\mathrm{fa}} \leq \alpha \end{cases} \ , \tag{15}$$

where

$$\tilde{J}(\delta_0, \delta_1) = \min_{\hat{\boldsymbol{b}}} J(\delta_0, \delta_1, \hat{\boldsymbol{b}}) \ . \tag{16}$$

Solving (14)-(15) is not necessarily feasible for any arbitrary choices of $\alpha$ and $\beta$. The following remark summarizes the coupled valid choices of $\alpha$ and $\beta$ that make $\mathcal{P}$ feasible.

*Remark 3:* By setting the tolerable level of false alarm probability at $\alpha$, the probability of missed-detection is known to be minimized by the Neyman-Pearson test. Let us define $\beta^*(\alpha)$ as the minimum missed-detection probability yielded by the Neyman-Pearson test. Clearly the two constraints

$$\mathsf{P}_{\mathrm{fa}} \leq \alpha \quad \text{and} \quad \mathsf{P}_{\mathrm{mis}} \leq \beta \ ,$$

are feasible simultaneously only if $\beta \geq \beta^*(\alpha)$ .

*Remark 4:* The observation above indicates that the proposed framework trades some detection quality, by tolerating errors beyond that achievable by the Neyman-Pearson test, in favor of enhancing the estimation quality. Allowing for such a tradeoff between estimation and detection qualities offers the freedom of putting appropriate/desired emphasis on the detection or the estimation part, depending on the application.

*Theorem 2:* Let $\gamma > 0$ be the solution of the equation

$$\mathsf{P}\left(C_{\mathrm{p}}^*(\boldsymbol{y}) \leq \gamma \mid \mathsf{T} = \mathsf{H}_1\right) = 1 - \beta \ . \tag{17}$$

Then the decision rule for detection is

$$\text{if} \quad \mathsf{P}\left(C_{\mathrm{p}}^*(\boldsymbol{y}) \leq \gamma \mid \mathsf{T} = \mathsf{H}_0\right) \leq \alpha \ , \text{ the decision rule is} \quad C_{\mathrm{p}}^*(\boldsymbol{y}) \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\lessgtr}} \gamma \ , \tag{18}$$

$$\text{and if} \quad \mathsf{P}\left(C_{\mathrm{p}}^*(\boldsymbol{y}) \leq \gamma \mid \mathsf{T} = \mathsf{H}_0\right) > \alpha \ , \text{ the decision rule is} \quad \frac{f_1(\boldsymbol{y})}{f_0(\boldsymbol{y})} \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\lessgtr}} \theta \ , \tag{19}$$

where $\zeta$ and $\theta$ are selected such that the two constraints $\mathsf{P}_{\mathrm{fa}} \leq \alpha$ and $\mathsf{P}_{\mathrm{mis}} \leq \beta$ are satisfied with equality. Also, the estimator is

$$\hat{\boldsymbol{b}}(\boldsymbol{y}) = \hat{\boldsymbol{b}}^*(\boldsymbol{y}) \ . \tag{20}$$

*Remark 5:* When the observation noise term $\boldsymbol{Z}$ has the standard complex Gaussian distribution, given the model in (8), the distributions $f_0(\boldsymbol{y})$ and $f_1(\boldsymbol{y})$ are complex Gaussian with means $\boldsymbol{0}$ and $\boldsymbol{b}$, respectively, and covariance matrix $\boldsymbol{I}$. Also an appropriate choice for the cost function is the MMSE cost.

## V. Distributed Detection-Estimation of Attacks

In this section we focus on a distributed solution for the joint attack detection and system state recovery problem, whose centralized counterpart was developed in Section IV. For operational purposes, the current electricity grid may be viewed as a large collection of interconnected RTOs. The sheer large size of the grid and the environmental unpredictability often make centralized information processing tasks infeasible. This problem will manifest itself more acutely in the future especially with the incorporation of numerous diverse sources of power generation and consumption distributed throughout the grid. As an alternative to centralized information processing, we propose a fully distributed approach in this section, in which the geographically distributed controlling agents in the network collaborate by means of iterative *local* processing and message passing to achieve system-wide attack detection and state recovery.

Referring to the (centralized) decomposition in (7), as shown in Section IV-A, the useful information about the attack vector $\boldsymbol{B}$ is embedded only in $\boldsymbol{y}$, the component of the observation vector $\boldsymbol{Y}$ in the null space of $\boldsymbol{H}$. For computational purposes, the orthogonal complement $\boldsymbol{y}^\dagger$ is simply the projection of $\boldsymbol{Y}$ onto the range space of $\boldsymbol{H}$. Writing

$$\boldsymbol{y}^\dagger = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{b}^\dagger + \boldsymbol{z}^\dagger \tag{21}$$

we note that $\boldsymbol{y}^\dagger = \boldsymbol{H}\boldsymbol{\theta}_{\text{lc}}$, where $\boldsymbol{\theta}_{\text{lc}}$ corresponds to the least squares estimate of $\left(\boldsymbol{H}\boldsymbol{X} + \boldsymbol{b}^\dagger\right)$ based on $\boldsymbol{Y}$. It then follows that,

$$\boldsymbol{y} = \boldsymbol{Y} - \boldsymbol{H}\boldsymbol{\theta}_{\text{lc}} = \left(I - \boldsymbol{H}(\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\right)\boldsymbol{Y} \ , \tag{22}$$

assuming $\boldsymbol{H}$ is full rank, i.e., the system is observable. The optimal centralized detection-estimation test (Theorem 2) is based on the statistic $\boldsymbol{y}$, which in turn depends on the global observation vector $\boldsymbol{Y}$.

Here we consider a framework for performing distributed detection-estimation of $\boldsymbol{B}$. To this end, let $N$ denote the number of substations (nodes), which may be quite large depending on the region of interest. Accordingly, the observation vector $\boldsymbol{Y}$ is a collection of local observation vectors, $\boldsymbol{Y} = [\boldsymbol{Y}_1^T, \cdots, \boldsymbol{Y}_N^T]^T$, where $\boldsymbol{Y}_n$ denotes the observation of the $n$-th node and is generated as follows:

$$\boldsymbol{Y}_n = \boldsymbol{H}_n\boldsymbol{X} + \boldsymbol{Z}_n + \boldsymbol{B}_n \ , \tag{23}$$

where $\boldsymbol{H}_n$ corresponds to the local Jacobian, and with $\boldsymbol{Z}_n$ and $\boldsymbol{B}_n$ being the portions of the noise and attack vector respectively influencing the measurements at node $n$. Due to the large size of the grid, it is often desirable, to compute the statistic of interest in a distributed fashion, whereby nodes iteratively exchange information to reproduce the value $\boldsymbol{y}$ locally. Such distributed computation relieves a centralized fusion center of a possible communication bottleneck and requires each node to know its local $\boldsymbol{H}_n$ and $\boldsymbol{Y}_n$ only. In the absence of malicious data attacks, the problem of distributed state estimation in smart grids was addressed in [21] (see also [22]), which introduced a distributed least-squares algorithm $\mathcal{M}-\mathcal{CSE}$ for computing $\boldsymbol{\theta}_{\text{lc}}$ at each node. However, in the presence of injection attacks, as shown in Section IV, the optimal detection-estimation task is no longer a direct computation of $\boldsymbol{\theta}_{\text{lc}}$ from $\boldsymbol{y}$, but involves the computation of the attack sufficient statistic $\boldsymbol{y}$. This sufficient statistic $\boldsymbol{y}$ is used for the joint detection-estimation of the attack (Theorem 2). The attack estimate thus obtained is used to eliminate the effect of $\boldsymbol{B}$ from the observation vector $\boldsymbol{Y}$ and subsequently the system state $\boldsymbol{X}$ is recovered by performing a least squares estimation using the residual observation $\boldsymbol{y}^\dagger$. In this paper, we introduce a non-trivial variant of $\mathcal{M}-\mathcal{CSE}$, the $\mathcal{DA}$ algorithm, for the distributed computation of the attack sufficient statistic $\boldsymbol{y}$ at each node, which is the fundamental building block in joint attack detection-estimation and subsequent state-recovery. This, in turn, allows each node in the network to compute the optimal detector-estimator of the attack and subsequently recover the system state.

### A. Algorithm $\mathcal{DA}$ and Convergence

The goal of this section is to design and analyze the distributed scheme $\mathcal{DA}$, in which each node converges almost surely to the centralized sufficient statistic $\boldsymbol{y}$. In particular, we show that the $\mathcal{DA}$ algorithm leads to such convergence at each node, under the assumption of global observability and connectivity of the inter-node communication network. Before proceeding to the details, we note that the $\mathcal{DA}$ scheme introduced here can be modified suitably to 1) yield more general observation functionals (statistics) at each node, including the $\chi^2$ statistic often used for bad data detection and identification; 2) deal with nonlinear observation models; and 3) operate in unpredictable environments with random inter-sensor communication failures or transmission noises. Some of these generalizations were studied in [23] in the context of general parameter estimation. Recall the observation model at the $n$-th node:

$$\boldsymbol{Y}_n = \boldsymbol{H}_n\boldsymbol{X} + \boldsymbol{Z}_n + \boldsymbol{B}_n \ , \tag{24}$$

where $\boldsymbol{Y}_n$ is the local observation vector at the $n$-th node. We make the following assumption on global observability.

**(E.1)**-Observability: The matrix

$$\boldsymbol{G} = \sum_{n=1}^{N} \boldsymbol{H}_n^T\boldsymbol{H}_n \tag{25}$$

is of full-rank.

Starting from an initial deterministic estimate of $\boldsymbol{y}$ (the initial state may be random, but here we assume it is deterministic for notational simplicity), denoted by $\mathbf{x}_n(0)$, each node generates by a distributed iterative algorithm a sequence of estimates, $\{\mathbf{x}_n(i)\}_{i\geq 0}$. The estimate $\mathbf{x}_n(i+1)$ of $\boldsymbol{y}$ at the $n$-th node at time $i+1$ is a function of: 1) its previous estimate; 2) the communicated estimates at time $i$ from its neighboring nodes; and 3) the local observation $\boldsymbol{Y}_n$.

**Algorithm** $\mathcal{DA}$: Based on the current state $\mathbf{x}_n(i)$, the exchanged data $\{\mathbf{x}_l(i)\}_{l\in\Omega_n}$, and the observation $\boldsymbol{Y}_n$, we update the estimate at the $n$-th node by the following distributed iterative algorithm:

$$\begin{aligned}\mathbf{x}_n(i+1) = \mathbf{x}_n(i) - \left[\gamma(i)\sum_{l\in\Omega_n}(\mathbf{x}_n(i)-\mathbf{x}_l(i))\right] \\ - \gamma(i)\mathcal{P}_n^T\left[\boldsymbol{Y}_n - \boldsymbol{H}_n\hat{\mathbf{x}}_n(i) - \mathcal{P}_n\mathbf{x}_n(i)\right] \ ,\end{aligned} \tag{26}$$

where $\mathcal{P}_n$ is an $m_n \times m$ selection matrix that selects the components of $\mathbf{x}_n(i)$ corresponding to the location of $\boldsymbol{Y}_n$ in the vector $\boldsymbol{Y}$, with $\boldsymbol{Y}_n \in \mathbb{R}^{m_n}$ and $\boldsymbol{Y} \in \mathbb{R}^m$. The auxiliary state sequence $\{\hat{\mathbf{x}}_n(i)\}$ at each node $n$ is also generated according to a distributed scheme,

$$\begin{aligned}\hat{\mathbf{x}}_n(i+1) = \hat{\mathbf{x}}_n(i) \\ - \left[\beta(i)\sum_{l\in\Omega_n}(\hat{\mathbf{x}}_n(i)-\hat{\mathbf{x}}_l(i)) - \alpha(i)\boldsymbol{H}_n^T\left(\boldsymbol{Y}_n - \boldsymbol{H}_n\hat{\mathbf{x}}_n(i)\right)\right] \ .\end{aligned} \tag{27}$$

In (26)-(27), $\{\gamma(i)\}, \{\alpha(i)\}, \{\beta(i)\}$ are appropriately chosen time-varying weight sequences. Algorithm $\mathcal{DA}$ is distributed since at node $n$ it involves only the data from the nodes in its communication neighborhood $\Omega_n$. In order to implement the $\mathcal{DA}$, each node stores and updates two states, $\mathbf{x}_n(i)$, the estimate of $\boldsymbol{y}$, and $\hat{\mathbf{x}}_n(i)$, an auxiliary state used for the update of $\mathbf{x}_n(i)$.

Now we formally refer to the recursive estimation algorithm in (26)-(27) as $\mathcal{DA}$. We note that the estimate sequence $\{\mathbf{x}_n(i)\}$ is random, due to the stochasticity of the noise. The following assumption on the connectivity of the inter-node communication network is assumed:

**(E.2)**-Connectivity: The inter-node communication network determined by the communication neighborhoods $\Omega_n$ is connected.[3]

**(E.3)**-Time varying weights: The sequences $\{\alpha(i)\}$ and $\{\beta(i)\}$ are of the form

$$\alpha(i) = \frac{a}{(i+1)^{\tau_1}} \quad \text{and} \quad \beta(i) = \frac{b}{(i+1)^{\tau_2}} \ , \tag{28}$$

where $a, b > 0$ are constants and the exponents $\tau_1$ and $\tau_2$ satisfy

$$0 < \tau_1 \leq 1 \ \text{ and } \ 0 < \tau_2 < \tau_1 \ . \tag{29}$$

The sequence $\gamma(i)$ is of the form

$$\gamma(i) = \frac{c}{(i+1)^{\tau_3}} \ , \tag{30}$$

where $c > 0$ is a constant and the exponent $\tau_3$ satisfies

$$0 < \tau_3 \leq 1 \ . \tag{31}$$

*Remark 6:* A key thing to note is that, although the weights are decaying over time, i.e., $\gamma(i), \alpha(i), \beta(i) \to 0$ as $i \to \infty$, they are persistent, i.e.,

$$\sum_{i \geq 0} \alpha(i) = \infty, \quad \sum_{i \geq 0} \beta(i) = \infty, \ \text{ and } \ \sum_{i \geq 0} \gamma(i) = \infty \ . \tag{32}$$

Whereas the decaying nature of the weight sequences guarantee convergence, the persistence condition is necessary to drive the estimators to $\boldsymbol{y}$ from arbitrary initial conditions.

The following result characterizes the convergence of the individual node estimates $\{\mathbf{x}_n(i)\}$ to the desired sufficient statistic $\boldsymbol{y}$.

*Theorem 3:* Consider the $\mathcal{DA}$ under **(E.1)-(E.3)**. Then, for each $n$, the estimate sequence $\{\mathbf{x}_n(i)\}$ converges a.s. to the sufficient statistic $\boldsymbol{y}$, i.e.,

$$\mathsf{P}\left(\lim_{i \to \infty} \mathbf{x}_n(i) = \boldsymbol{y}\right) = 1 \quad \forall n \ . \tag{33}$$

The convergence rate in Theorem 3 depends on the choice of the various weight sequences. The proven convergence above allows each node in a distributed way to compute the centralized sufficient statistic $\boldsymbol{y}$ needed for the construction of the optimal detector-estimator of the attack vector $\boldsymbol{B}$.

## VI. Conclusions

In this paper we have studied the problem of distributed state estimation in smart grids with malicious injection data attacks. Our approach is unified, in that, it presents a joint detection-estimation framework for simultaneous attack detection and system state recovery. Under varying scenarios of the attacker's information resources, we have characterized optimal tests and procedures for attack detection and state estimation. Furthermore, we have provided a completely distributed approach based on local processing and message passing that enables each network node to compute the global sufficient statistic for performing the optimal detection-estimation task. Convergence of our distributed scheme is guaranteed as long as the physical model is observable and the inter-node communication network is connected.

[3]We note here that the cyber communication neighborhood could be significantly different and sparser than the physical neighborhood determined by electrical connections.

## References


[1] F. C. Schweppe, J. Wildes, and A. Bose, "Power system static state estimation, Parts I, II and III," *IEEE Transactions on Power Apparatus and Systems*, vol. 89, no. 1, pp. 120–135, Jan. 1970.

[2] F. F. Wu, K. Moslehi, and A. Bose, "Power system control centers: Past, present, and future," *Proceedings of the IEEE*, vol. 93, no. 11, pp. 1890–1908, Nov. 2005.

[3] F. Monticelli, "Electric power system state estimation," *Proceedings of the IEEE*, vol. 88, no. 2, pp. 262–282, Feb. 2000.

[4] A. Abur and A. Gomez-Expósito, *Power System State Estimation: Theory and Implementation*. New York, NY: Marcel Dekker, 2004.

[5] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proceedings of the 16th ACM Conference on Computer and Communications Security*, Chicago, IL, Nov. 2009.

[6] H. Sandberg, A. Teixeira, and K. H. Johansson, "On security indices for state estimators in power networks," in *Proceedings of the First Workshop on Secure Control Systems*, Stockholm, Sweden, Apr. 2010.

[7] O. Kosut, L. Jia, R. Thomas, and L. Tong, "Limiting false data attacks on power system state estimation," in *Proceedings of the Conference on Information Sciences and Systems*, Princeton, NJ, Mar. 2010.

[8] ——, "Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures," in *Proceedings of the IEEE SmartGridComm*, Gaithersburg, MD, Oct. 2010.

[9] L. Xie, "A framework for distributed decision making in electric energy systems with intermittent resources," Ph.D. dissertation, Carnegie Mellon University, Pittsburgh PA, USA, 2009.

[10] T. V. Cutsem, J. L. Horward, and M. Ribbens-Pavella, "A two-level static state estimator for electric power systems," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-100, no. 8, pp. 3722–3732, Aug. 1981.

[11] T. V. Cutsem and M. Ribbens-Pavella, "Critical survey of hierarchical methods for state estimation of electric power systems," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-102, no. 10, pp. 247–256, Oct. 1983.

[12] A. Bose, A. Abur, K. Y. K. Poon, and R. Emami, "Implementation issues for hierarchical state estimators," *Final Project Report*, vol. PSERC Document 10-11, Aug. 2010.

[13] G. N. Korres, "A distributed multiarea state estimation," *IEEE Transactions on Power Apparatus and Systems*, vol. 41, no. 4, pp. 550–558, Mar. 2010.

[14] L. Zhao and A. Abur, "Multiarea state estimation using synchronized phasor measurements," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 611–617, May 2005.

[15] W. Jiang, V. Vittal, and G. T. Heydt, "A distributed state estimator utilizing synchronized phasor measurements," *IEEE Transactions on Power Systems*, vol. 22, no. 2, pp. 563–571, May 2007.

[16] G. Valverde, S. Chakrabarti, E. Kyriakides, and V. Terzija, "A constrained formulation for hybrid state estimation," *IEEE Transactions on Power Systems.*, accepted, Digital Object Identifier 10.1109/TPWRS.2010.2079960.

[17] F. R. K. Chung, *Spectral Graph Theory*. Providence, RI : American Mathematical Society, 1997.

[18] T. T. Kim and H. V. Poor, "Strategic protection against data injection attacks on power grids," *IEEE Transactions on Smart Grid*, vol. 3, no. 2, pp. 326–333, Jun. 2011.

[19] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York, NY: Springer, 1994.

[20] G. V. Moustakides, G. H. Jajamovich, A. Tajer, and X. Wang, "Joint detection and estimation: Optimum tests and applications," *IEEE Transactions on Information Theory*, Jan. 2011, submitted for publication, arXiv:1101.5084.

[21] L. Xie, D.-H. Choi, S. Kar, and H. V. Poor, "Distributed state estimation in wide-area power systems," to be submitted for journal publication.

[22] L. Xie, D.-H. Choi, and S. Kar, "Cooperative distributed state estimation: Local observability relaxed," in *Proceedings of the IEEE Power and Energy Society General Meeting*, Detroit, MI, Jul. 2011.

[23] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," *IEEE Transactions on Information Theory*, Aug. 2008, submitted for publication, arXiv:0809.0009.