Distributed Quickest Detection of Cyber-Attacks in Smart Grid

Mehmet Necip Kurt, Yasin Yılmaz, *Member, IEEE,* and Xiaodong Wang, *Fellow, IEEE*

***Abstract*— In this paper, online detection of false data injection (FDI) attacks and denial of service (DoS) attacks in the smart grid is studied. The system is modelled as a discrete-time linear dynamic system and state estimation is performed using the Kalman filter. The generalized CUSUM algorithm is employed for quickest detection of the cyber-attacks. Detectors are proposed in both centralized and distributed settings. The proposed detectors are robust to time-varying states, attacks, and set of attacked meters. Online estimates of the unknown attack variables are provided, that can be crucial for a quick system recovery. In the distributed setting, due to bandwidth constraints, local centers can only transmit quantized messages to the global center, and a novel event-based sampling scheme called level-crossing sampling with hysteresis (LCSH) is proposed that is shown to exhibit significant advantages compared with the conventional uniform-in-time sampling (US) scheme. Moreover, a distributed dynamic state estimator is proposed based on information filters. Numerical examples illustrate the fast and accurate response of the proposed detectors in detecting both structured and random attacks and their advantages over the existing methods.**

***Index Terms*— Smart grid, Kalman filter, false data injection attack, denial of service attack, quickest detection, generalized CUSUM, distributed algorithm, level-crossing sampling.**

I. Introduction

With the recent advancements in monitoring, sensing, signal processing, control, and communication, advanced technologies are being integrated into the next-generation power systems, i.e., the smart grid. Due to such features, the smart grid depends on a critical cyber infrastructure which makes it vulnerable to hostile cyber threats [1]. This raises safety and security concerns about the smart grid since any outages or failures in this system may lead to wide-area power blackouts and significant financial losses. Among many types of cyber-attacks, we pay special attention to false data injection (FDI) and denial of service (DoS) attacks in this study. The aim of FDI attacks is to compromise meter measurements with additive malicious data and the aim of DoS attacks is to block system functionality or intervene normal system operation to some extent.

A. Literature Survey on Cyber-Attacks and Counter-Measures in Smart Grid

In practice, FDI and DoS attacks can be performed by manipulating/jamming the network communication channels, hacking (physically or through cyber infrastructure) the smart grid components (smart meter, control center, etc.) or accessing and manipulating the database of a control center [1]–[4]. The DoS attack can also be performed by repeatedly sending huge amounts of packets to the network communication channels to prevent the useful system message packets from being received by the legitimate receivers (flooding) [5]. For an illustration of how an attacker can perform cyber-attacks in a smart grid, we present Fig. 1.

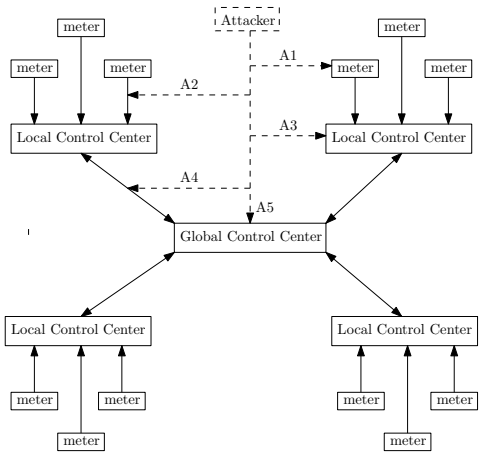

Fig. 1: An illustration of the vulnerabilities of a smart grid. The solid lines illustrate the network communication channels and the dashed lines illustrate the possible cyber-attacks. The attacker can (i) hack a smart grid component to block or manipulate its operation (A1, A3, A5), (ii) manipulate, jam or block the communication channels (A2, A4).

Cyber attacks in the smart grid mainly target against state estimation. Since the power system is regulated based on estimated states, any deviation from the actual state estimates leads to wrong decisions in energy management system, manipulated electricity market prices [6], and other unpredictable detrimental consequences. Traditionally, state estimation in power systems is based on the least squares (LS) methods. Bad data detection methods based on the LS estimators are successful in identifying bad data due to random noise and faults but unable to identify structured (called "stealth" in the literature) FDI attacks [7]. Moreover, it is inconvenient to use LS estimators for real-time monitoring of power system that is highly dynamic due to changes in load, power generation, and system topology over time [8].

Classical methods for bad data detection checks either the l_2-norm of measurement residual or the largest normalized

This work was supported in part by the U.S. National Science Foundation (NSF) under grant ECCS-1405327, and in part by the U.S. Office of Naval Research (ONR) under grant N000141410667. Y. Yilmaz's work was supported in part by NSF under grant CNS-1737598 and in part by the SCEEE-17-03 grant.

M. N. Kurt and X. Wang are with the Department of Electrical Engineering, Columbia University, New York, NY 10027, USA (e-mail: m.n.kurt@columbia.edu; wangx@ee.columbia.edu).

Y. Yılmaz is with the Department of Electrical Engineering, University of South Florida, Tampa, FL 33620, USA (e-mail: yasiny@usf.edu).

residue and if these values are above a certain threshold, an attack or a fault is detected. It has been first shown in [7] that it is possible to inject false data without changing the measurement residual if the attacker knows the system topology. Moreover, in [9], it is shown that even if the attacker has incomplete knowledge of the grid topology, it can still carry out stealth FDI attacks if the system parameters vary in a small dynamic range. This vulnerability of the LS estimator-based detectors has opened up new research directions. For instance, in [10], the attacker is constrained such that it can compromise only a limited number of meters and then an algorithm for meter selection is proposed. In [11], the problem of constructing a stealth attack by minimizing the number of attacked meters is investigated. In some studies, it is assumed that the attacker has incomplete knowledge about the system and it is shown that the attacker can estimate the topology by collecting online and offline data [12] or by exploiting electricity market data [13]. Furthermore, in [14], it is shown that if an attacker knows the topology of only a local region in a power system, it can successfully perform stealth FDI attacks to the local meters.

In response to FDI attacks, some techniques for defending the grid have been proposed. In [10], it is shown that if a certain number of meters are protected, then FDI attacks can always be detected and in [15], an algorithm to specify such meters is proposed. Meters can be protected by using Phasor Measurement Units (PMUs). PMUs are advanced devices that make use of the Global Positioning System (GPS) and provide highly precise phasor measurements synchronized over the whole grid [16]. They are expensive devices, hence the number of PMUs and their locations should be optimized. In this direction, PMU placement algorithms are proposed in [17] and [18]. It is shown in [19], however, if the GPS receiver is spoofed, the advantages of PMUs will disappear. In such a case, coordination between different parts of the power grid is lost and the whole power system might even collapse.

For a timely and reliable response, detecting cyber attacks as quickly as possible and with an acceptable level of false alarm rate has a critical importance in real-time operation of the smart grid. Hence, the framework of sequential change detection, also known as quickest detection, is very suitable for this setup. In this framework, data is sequentially observed and after each observation time, a decision is taken: it either stops and a change is declared or continues to observe more data. Moreover, there exists a tradeoff between the detection speed and the detection accuracy. As the desired detection accuracy increases, the detection delay also increases, or equivalently the detection speed decreases. There are several studies exploiting this framework for detection of FDI attacks [20], [21]. The detector in [20] outperforms the so-called adaptive CUSUM test [21] and has a quick and accurate response if the attacker has an incomplete information about the system topology. However, it cannot detect the stealth FDI attacks.

The LS methods for state estimation depend only on the present measurements. Adopting a state-space model enables a dynamic state estimator that combines present and past measurements so that the system state can be inferred in a more accurate and robust way. If the noise has a Gaussian distribution, the Kalman filter is the optimal linear estimator that minimizes the mean squared error [22]. Moreover, the Kalman filter provides predicted measurements which can be exploited to improve attack detection performance. Some simple bad data detectors based on the Kalman filter have been proposed in the literature. In particular, the Euclidean detector [23], cosine similarity metric based detector [24], and chi-square detector [25] are the existing techniques that check the difference between actual measurements and the predicted measurements. However, such detectors are essentially outlier detection methods making sample-by-sample decisions, i.e., they declare a sample measurement as either normal or anomalous. On the other hand, in sequential change detection, negative/positive evidence for a change (e.g., an attack or failure) in the system are accumulated over time and a change is declared only if the evidence supporting change is reliably high. Hence, attack/anomaly detectors based on the sequential change detection theory are more reliable compared to outlier detection techniques. Furthermore, detection-only schemes such as the detectors presented in [23]–[25] do not provide any estimates for the magnitude of the injected malicious data or the set of compromised meters, which may be critical to know for an effective attack mitigation and system recovery, e.g., to recover the attack-free states or to isolate the compromised meters during the recovery process. On the other hand, in detection schemes including an estimation mechanism, estimation errors may worsen the detection performance.

Due to limited communication resources, e.g., energy and bandwidth, collection of measurements in a single node may not be practically feasible. Moreover, in a large power system, processing huge amount of data in a single node is infeasible and susceptible to single node failure. Therefore, a resource-effective distributed implementation is required in practice. In such a system, the computation is distributed over the whole network and the communication overhead is reduced as much as possible. The information filter, which is an algebraic equivalence to the Kalman filter, is convenient for a distributed setting due to its simple update rules [26]. Furthermore, event-based sampling techniques are convenient for sequential change detection, see e.g., [20], [27], [28]. In our case, the decision statistics are expected to vary in a small range before the attack, hence non-informative transmissions during this period can be eliminated with event-based sampling techniques.

B. Contributions

In this paper, the smart grid is modeled as a discrete-time linear dynamic system and the Kalman filter is employed for state estimation. State and measurement forecasts/predictions provided by the Kalman filter are exploited to improve the attack detection performance. We list our main contributions as follows:

- Novel low-complexity real-time detection schemes are proposed for both FDI and DoS attacks in smart grid in both centralized and distributed settings. The proposed schemes are robust to unknown and time-varying attack magnitudes, set of attacked meters, and the system state.

- Online estimates of the attack variables are provided that can be crucial for a quick attack mitigation and system recovery. In particular, simple closed-form maximum likelihood estimate (MLE) expressions are derived for the attack magnitudes and the set of attacked meters.
- The stealth FDI attacks described in [7] can be detected with the proposed FDI detection mechanisms.
- A novel fully distributed dynamic state estimator is proposed.

Further, in the distributed setting, to use communication resources more effectively and to improve the distributed attack detection performance, a novel event-triggered sampling scheme called level-crossing sampling with hysteresis (LCSH) is proposed for sampling and transmission of local statistics, that is shown to exhibit significant advantages over the conventional uniform-in-time sampling (US) scheme.

C. Organization

The remainder of the paper is organized as follows. In Section II, we present the system model, the attack models under consideration, and the problem formulations. In Section III, we present the cyber-attack detectors in the centralized setting, where all measurements in the system are collected and processed by a single node. In Section IV, we present the system model in the distributed setting, the distributed state estimation technique, and the corresponding cyber-attack detectors. We illustrate the performance of the proposed detectors via extensive simulations in Section V. Finally, Section VI concludes the paper. Throughout the paper, we use boldface letters for vectors and matrices. Moreover, $\boldsymbol{o}^T$ denotes the transpose of a vector or matrix $\boldsymbol{o}$.

II. System Model and Problem Formulations

Suppose that there are K meters in a power system consisting of $N+1$ buses, where usually $K \geq N$ [29]. System state $\mathbf{x}_t = [x_{1,t}, \ldots, x_{N,t}]^T$ represents phase angles of N buses at time t where one of the buses is chosen as the reference bus. Measurement taken at meter $k \in \{1, \ldots, K\}$ at time t is denoted with $y_{k,t}$ and the set of measurements is denoted with $\mathbf{y}_t = [y_{1,t}, \ldots, y_{K,t}]^T$. In a centralized setup, a single controller node observes all the measurements in $\mathbf{y}_t$. On the other hand, in a distributed setup, measurements in $\mathbf{y}_t$ are distributed over the network, i.e., each node in the system observes $\mathbf{y}_t$ in part.

In the actual power system, relationship between the measurements and the state variables is based on a nonlinear function [1]. We consider the commonly used approximate direct current (DC) model, see e.g., [7], [29], [30], and a discrete-time linear dynamic system with the state-space equations

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{v}_t, \tag{1}$$

$$\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \mathbf{w}_t, \tag{2}$$

where $\mathbf{A} \in \mathbb{R}^{N \times N}$ is the state transition matrix, $\mathbf{H} \in \mathbb{R}^{K \times N}$ is the measurement matrix, $\mathbf{v}_t = [v_{1,t}, \ldots, v_{N,t}]^T$ is the process noise, and $\mathbf{w}_t = [w_{1,t}, \ldots, w_{K,t}]^T$ is the measurement noise. We assume that $\mathbf{v}_t$ and $\mathbf{w}_t$ are independent additive white Gaussian random processes where $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \sigma_v^2 \mathbf{I}_N)$, $\mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \sigma_w^2 \mathbf{I}_K)$, and $\mathbf{I}_K$ is a $K \times K$ identity matrix. Next, we explain the considered cyber-attack models and the corresponding problem formulations. Henceforth, we use the superscripts f, d, and 0 to denote quantities related to FDI attacks, DoS attacks, and no-attack, respectively.

A. FDI Attack

We consider that the attacker is initially inactive and at an unknown time τ, it starts to manipulate the measurements by injecting additive false data so that the measurement vector takes the following form:

$$\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \mathbf{a}_t + \mathbf{w}_t, \quad t \geq \tau \tag{3}$$

where $\mathbf{a}_t = [\mathrm{a}_{1,t}, \ldots, \mathrm{a}_{K,t}]^T$ is the false data created by the attacker at time $t \geq \tau$. Let $\mathbf{h}_k^T \in \mathbb{R}^N$ be the kth row of the measurement matrix, i.e., $\mathbf{H}^T = [\mathbf{h}_1, \ldots, \mathbf{h}_K]$. Based on (3), in case of an FDI attack, $y_{k,t}$ takes the following form:

$$y_{k,t} = \begin{cases} \mathbf{h}_k^T \mathbf{x}_t + \mathrm{a}_{k,t} + w_{k,t}, & \text{if } k \in \mathcal{S}_t^f, \\ \mathbf{h}_k^T \mathbf{x}_t + w_{k,t}, & \text{if } k \notin \mathcal{S}_t^f, \end{cases}, \quad t \geq \tau, \tag{4}$$

where $\mathcal{S}_t^f \subset \{1, \ldots, K\}$ is the unknown set of compromised meters at time t. Then, the null and alternative hypotheses can be written as

$$\mathcal{H}_0 : \; y_{k,t} \sim \mathcal{N}(\mathbf{h}_k^T \mathbf{x}_t, \sigma_w^2), \quad \forall k \in \{1, 2, \ldots, K\}, \; \forall t \tag{5}$$

$$\mathcal{H}_1^f : \; y_{k,t} \sim \begin{cases} \mathcal{N}(\mathbf{h}_k^T \mathbf{x}_t, \sigma_w^2), & \forall k \in \{1, 2, \ldots, K\}, \quad t < \tau \\ \begin{cases} \mathcal{N}(\mathbf{h}_k^T \mathbf{x}_t, \sigma_w^2), & \forall k \notin \mathcal{S}_t^f \\ \mathcal{N}(\mathbf{h}_k^T \mathbf{x}_t + \mathrm{a}_{k,t}, \sigma_w^2), & \forall k \in \mathcal{S}_t^f. \end{cases}, & t \geq \tau, \end{cases} \tag{6}$$

where we define the change event of interest as

$$|\mathrm{a}_{k,t}| \geq \gamma, \; \forall k \in \mathcal{S}_t^f, \; t \geq \tau, \tag{7}$$

i.e., γ is a predefined lower bound for the absolute value of $\mathrm{a}_{k,t}$ that draws security attentions[1]. Since it is hard to distinguish noise with low-magnitude false data, the value of γ should be selected such that the number of false positives due to noise is reduced to an acceptable level. Moreover, FDI attacks with small magnitudes are expected to affect the system minimally.

Our aim is to detect the attack as quickly as possible with the desired level of false alarm rate. In the sequential change detection literature, there are two main approaches: Bayesian and non-Bayesian [31]. In the Bayesian approach, the change point τ is considered as a random variable with a known prior (geometric) distribution [32]. However, in practice (also in our case), it is difficult to know a prior distribution for τ. Hence, we proceed with a non-Bayesian approach where τ is considered as a deterministic unknown quantity. In particular, we use Lorden's definition for the worst-case detection delay [33]

$$J(T^f) = \sup_{\tau} \operatorname*{ess\,sup}_{\mathcal{F}_\tau} \mathbb{E}_\tau \big[(T^f - \tau)^+ \, | \mathcal{F}_\tau \big], \tag{8}$$

where T^f is the stopping time of a detection scheme, $\mathcal{F}_\tau$ is the filtration, i.e., all measurements obtained until time τ, and $(\cdot)^+ = \max(\cdot, 0)$. Moreover, $\mathbb{E}_k$ is the expectation under $\mathbb{P}_k$ that is defined as the probability measure when $\tau = k$. In

[1] γ is only a detector parameter. It is not a restriction for an attacker's strategy.

(8), the essential supremum, which is a concept in measure theory, is in practice equivalent to the supremum of a set. Note that $J(T^f)$ is called the worst-case detection delay since it is equal to the average detection delay calculated under the least favorable attack time and the least favorable history of measurements until the attack time. The stopping time is chosen to minimize the detection delay subject to false alarm constraints. The optimization problem is then expressed as

$$\inf_{T^f} \; J(T^f) \;\; \text{subject to} \;\; \mathbb{E}_\infty[T^f] \geq \alpha, \tag{9}$$

where $\mathbb{E}_\infty[T^f]$ is the mean time between false alarms in case of no attack, i.e., $\tau = \infty$, and α is a predetermined lower bound for $\mathbb{E}_\infty[T^f]$.

Let the probability density functions (pdfs) of the measurements corresponding to the measurement models given in (2) and (3) be denoted with $p^0(\mathbf{y}_t \,|\, \mathbf{x}_t)$ and $p^f(\mathbf{y}_t \,|\, \mathbf{x}_t, \mathbf{a}_t)$, respectively. If these two pdfs can be completely specified, the optimal solution of the quickest detection problem in (9) can be found using the CUSUM test [34]:

$$T^f = \inf \left\{ m \in \mathbb{N} : \max_{1 \leq j \leq m} \sum_{t=j}^{m} \log \frac{p^f(\mathbf{y}_t \,|\, \mathbf{x}_t, \mathbf{a}_t)}{p^0(\mathbf{y}_t \,|\, \mathbf{x}_t)} \geq h^f \right\}, \tag{10}$$

where h^f is the threshold of the test, which controls the tradeoff between minimizing the average detection delay and the false alarm rate.

Since (i) the system state $\mathbf{x}_t$ evolves over time and is not observed and (ii) the attack vector $\mathbf{a}_t$ and the set of meters under attack $\mathcal{S}_t^f$ are time-varying and unknown, it is not possible to directly apply the CUSUM test given in (10). However, we can follow the generalized likelihood ratio method [35, Sec. 5.3] and replace the unknown quantities with their estimates [20]. In particular, since we have a discrete-time linear dynamic system, the Kalman filter can be used to obtain the optimal state estimates $\hat{\mathbf{x}}_t$ [22]. Moreover, the MLEs of the attack vector, i.e., $\hat{\mathbf{a}}_t$, and the set of attacked meters, i.e., $\hat{\mathcal{S}}_t^f$, can be derived[2]. Then, the generalized CUSUM test can be used to obtain a solution to (9).

B. DoS Attack

We assume that in case of DoS attack, communication between some unknown set of meters and the control center is lost so that the control center has no knowledge about the measurements taken at the attacked meters. Let $\mathcal{S}_t^d \subset \{1, \ldots, K\}$ be the set of attacked meters at time t. At an unknown time τ, the attack starts and a meter measurements in case of a DoS attack take the following form:

$$y_{k,t} = \begin{cases} n_{k,t}, & \text{if } k \in \mathcal{S}_t^d, \\ \mathbf{h}_k^T \mathbf{x}_t + w_{k,t}, & \text{if } k \notin \mathcal{S}_t^d, \end{cases} \qquad t \geq \tau, \tag{11}$$

where $n_{k,t} \sim \mathcal{N}(0, \sigma_n^2)$ is the i.i.d. noise observed in case of a DoS attack.

The null hypothesis is as given in (5) and the alternative hypothesis can be written as

$$\mathcal{H}_1^d : \; y_{k,t} \sim \begin{cases} \mathcal{N}(\mathbf{h}_k^T \mathbf{x}_t, \sigma_w^2), & \forall k \in \{1, 2, \ldots, K\}, \quad t < \tau \\ \begin{cases} \mathcal{N}(\mathbf{h}_k^T \mathbf{x}_t, \sigma_w^2), & \forall k \notin \mathcal{S}_t^d \\ \mathcal{N}(0, \sigma_n^2), & \forall k \in \mathcal{S}_t^d. \end{cases}, & t \geq \tau. \end{cases} \tag{12}$$

Again, we use Lorden's definition for the worst-case detection delay [33]

$$J(T^d) = \sup_{\tau} \operatorname*{ess\,sup}_{\mathcal{F}_\tau} \mathbb{E}_\tau \big[(T^d - \tau)^+ \,|\mathcal{F}_\tau \big], \tag{13}$$

where T^d is the stopping time. The optimization problem is then stated as

$$\inf_{T^d} \; J(T^d) \;\; \text{subject to} \;\; \mathbb{E}_\infty[T^d] \geq \alpha. \tag{14}$$

Let the pdfs of the measurements corresponding to the measurement models given in (2) and (11) be denoted with $p^0(\mathbf{y}_t \,|\, \mathbf{x}_t)$ and $p^d(\mathbf{y}_t \,|\, \mathbf{x}_t)$, respectively. As before, as a solution to (14), the generalized CUSUM algorithm can be used using the MLE of the set of attacked meters and estimating the state variables using the Kalman filter.

Remark 1: In the literature, DoS attack is modeled as the lack of availability of meter measurements and in case of DoS attack, either a zero signal or a random signal is observed [3], [4], [36]. The former is more appropriate if the attacker hacks some subset of meters (or some control centers) and prevents the data transmission from these meters to the control centers. On the other hand, the latter is more appropriate if the attacker jams the network communication channels. We address both kinds of DoS attacks by modeling the signal received from the attacked meters as a zero-mean Gaussian noise with a generic variance σ_n^2.

For the DoS attacks performed by preventing the data transmission from smart meters, σ_n^2 can be set to a very small value (close to zero) so that the received signal is almost the zero signal. On the other hand, for the DoS attacks performed by jamming the network communication channels, the attacker can increase the noise variance to a very high level so that the actual message is lost (very low signal to noise ratio (SNR)). In this case, we consider that the attacker jams the communication channel by constantly emitting the Gaussian noise since (i) it is a common model for jamming in the literature [37], (ii) for an additive noise channel with a Gaussian input, among all distributions with a given mean and variance, the Gaussian noise maximizes the mean squared error (MSE) of estimating the input given the channel output [38], [39]. Hence, in order to maximize damages on the state estimation mechanism, the attacker can transmit additive white Gaussian noise for jamming. σ_n^2 can then be set to the minimum possible noise variance such that the actual signal can be neglected compared to the noise signal (SNR close to zero).

III. Centralized Attack Detectors

In the centralized setting, a central controller has all the system-wide information. In particular, a single node collects and processes all the measurements in the system. This can be achieved for a power system consisting of small number

[2]Both $\mathbf{a}_t$ and $\mathcal{S}_t^f$ are time-varying, non-random unknown parameters. Their values at each time depend only on the attacker's strategy.

of meters and located in a geographically small region. In the following, we present the generalized CUSUM test structures, the Kalman filter equations, the MLEs of the unknown attack parameters, and the proposed centralized detection algorithms for FDI and DoS attacks, respectively.

A. Centralized Detector for FDI Attacks

Since the measurement models in the null (cf. (5)) and the alternative (cf. (6)) hypotheses are different, state estimates corresponding to different hypotheses need to be calculated based on their respective measurement models. For this purpose, two parallel state estimators need to be simultaneously employed. The Kalman filter consists of two steps at each iteration: the prediction step and the measurement update step. At the prediction step, the state estimates at time t are based on all measurements up to $t-1$, and at the measurement update step, the state estimates at time t are based on all measurements up to t. Let the state estimates at time t for the null and the alternative hypotheses be denoted with $\hat{\mathbf{x}}^0_{t|t'}$ and $\hat{\mathbf{x}}^f_{t|t'}$, respectively ($t'=t-1$ for prediction and $t'=t$ for measurement update). The stopping time based on the generalized CUSUM test is then given in (15) (shown at the top of the next page) where g^f_m is the decision statistic at time m, and β_t is the generalized log-likelihood ratio (GLLR) calculated at time t.

Note that in (15), the state estimates of the prediction step, i.e., $\hat{\mathbf{x}}^0_{t|t-1}$ and $\hat{\mathbf{x}}^f_{t|t-1}$ are used. One of our purposes here is to block the effect of the attack vector at time t on the state estimates at time t. In this way, we aim to improve the detection of the time-varying attacks and also to obtain closed-form expressions for the MLEs of the unknown attack variables. Moreover, the prediction step of the Kalman filter, in fact, provides state and measurement forecasts/predictions, and the deviation of the actual measurements from the predicted ones is an indication of an unusual event, e.g., a fault or an attack. Hence, it is also exploited to improve the attack detection performance.

The Kalman filter equations at time t are given as follows:

Prediction:

$$\begin{aligned}\hat{\mathbf{x}}^0_{t|t-1} &= \mathbf{A}\hat{\mathbf{x}}^0_{t-1|t-1},\\ \hat{\mathbf{x}}^f_{t|t-1} &= \mathbf{A}\hat{\mathbf{x}}^f_{t-1|t-1},\\ \mathbf{P}_{t|t-1} &= \mathbf{A}\mathbf{P}_{t-1|t-1}\mathbf{A}^T+\sigma_v^2\,\mathbf{I}_N,\end{aligned} \tag{16}$$

Measurement update:

$$\begin{aligned}\mathbf{G}_t &= \mathbf{P}_{t|t-1}\mathbf{H}^T(\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^T+\sigma_w^2\,\mathbf{I}_K)^{-1},\\ \hat{\mathbf{x}}^0_{t|t} &= \hat{\mathbf{x}}^0_{t|t-1}+\mathbf{G}_t(\mathbf{y}_t-\mathbf{H}\hat{\mathbf{x}}^0_{t|t-1}),\\ \hat{\mathbf{x}}^f_{t|t} &= \hat{\mathbf{x}}^f_{t|t-1}+\mathbf{G}_t(\mathbf{y}_t-\mathbf{H}\hat{\mathbf{x}}^f_{t|t-1}-\hat{\mathbf{a}}_t),\\ \mathbf{P}_{t|t} &= \mathbf{P}_{t|t-1}-\mathbf{G}_t\mathbf{H}\mathbf{P}_{t|t-1},\end{aligned} \tag{17}$$

where $\mathbf{P}_{t|t-1}$ and $\mathbf{P}_{t|t}$ denote the estimates of the state covariance matrix based on the measurements up to $t-1$ and t, respectively. Moreover, $\mathbf{G}_t$ is the Kalman gain matrix and $\hat{\mathbf{a}}_t$ is the MLE of $\mathbf{a}_t$ (cf. (19)). Note that $\hat{\mathbf{a}}_t$ is used in the measurement update step of the Kalman filter for the alternative hypothesis. The following proposition presents the MLEs of the unknown attack parameters and the GLLR at time t.

Proposition 1: Let $\mathbf{e}_t=[e_{1,t},\dots,e_{K,t}]^T\triangleq\mathbf{y}_t-\mathbf{H}\hat{\mathbf{x}}^f_{t|t-1}$. Then, $e_{k,t}=y_{k,t}-\mathbf{h}_k^T\hat{\mathbf{x}}^f_{t|t-1}$. The most likely set of attacked meters at time t is given by

$$\hat{\mathcal{S}}^f_t=\{k:|e_{k,t}|>\frac{\gamma}{2},\ k=1,\dots,K\}, \tag{18}$$

the MLE of the attack vector, i.e., $\hat{\mathbf{a}}_t=[\hat{\mathrm{a}}_{1,t},\dots,\hat{\mathrm{a}}_{K,t}]^T$, is given by

$$\hat{\mathrm{a}}_{k,t}=\begin{cases}e_{k,t}, & \text{if}\ \ |e_{k,t}|\geq\gamma\\ \gamma, & \text{if}\ \ \frac{\gamma}{2}<e_{k,t}<\gamma\\ -\gamma, & \text{if}\ \ -\gamma<e_{k,t}<-\frac{\gamma}{2}\\ 0, & \text{else,}\end{cases} \tag{19}$$

and the GLLR at time t is given by

$$\begin{aligned}\beta_t=\frac{1}{2\sigma_w^2}\sum_{k=1}^{K}\Big(&(y_{k,t}-\mathbf{h}_k^T\hat{\mathbf{x}}^0_{t|t-1})^2\\ &-(y_{k,t}-\mathbf{h}_k^T\hat{\mathbf{x}}^f_{t|t-1}-\hat{\mathrm{a}}_{k,t})^2\Big).\end{aligned} \tag{20}$$

Proof: See Appendix A.

Based on (15), recursion of the decision statistic $g^f_t, t\in\mathbb{N}$ can be written as

$$g^f_t=(g^f_{t-1}+\beta_t)^+, \tag{21}$$

where $g^f_0=0$ and β_t is as given in (20). Note that if $g^f_t=0$ for any time t, then the change-point estimate in the (generalized) CUSUM algorithm is updated to time t [35, Sec. 2.2]. Since the alternative hypothesis $\mathcal{H}^f_1$ assumes the normal (no-attack) measurement model up to the change-point (cf. (6)), the Kalman filter for the alternative hypothesis needs also to be employed based on the normal measurement model up to the change-point. Hence, whenever the change-point estimate is updated, the Kalman filter estimates for the alternative hypothesis are updated by setting $\hat{\mathbf{x}}^f_{t|t}\leftarrow\hat{\mathbf{x}}^0_{t|t}$.

We summarize the proposed centralized attack detector in Algorithm 1 and present a graphical representation of the algorithm in Fig. 2. At each time t, we first employ the Kalman filters to estimate the states through the prediction step. We then calculate the MLE of the attack vector and specify the most likely set of attacked meters. Using the MLEs, we then implement the measurement update step of the Kalman filter. Then, we calculate the decision statistic. If the decision statistic crosses the predefined threshold, we declare an attack, otherwise we continue to collect measurements in the next time cycle.

Note that we deal with the false alarms due to outliers (e.g., high noise) through selecting γ and h^f sufficiently large. Since we gather attack statistics both in space and time, sufficiently high thresholds ensure small false alarm rates, equivalently high false alarm periods. On the other hand, higher thresholds will cause larger detection delays. Hence, there is a tradeoff in selecting γ and h^f.

B. Centralized Detector for DoS Attacks

Let the state estimates under the null (cf. (5)) and the alternative (cf. (12)) hypotheses be denoted with $\hat{\mathbf{x}}^0_{t|t'}$ and $\hat{\mathbf{x}}^d_{t|t'}$,

$$T^f = \inf\left\{ m \in \mathbb{N} : \underbrace{\max_{1\le j\le m} \sum_{t=j}^{m} \underbrace{\sup_{\mathcal{S}_t^f} \log \frac{\sup_{|a_{k,t}|\ge\gamma,\, k\in\mathcal{S}_t^f} p^f(\mathbf{y}_t \mid \hat{\mathbf{x}}_{t|t-1}^f, \mathbf{a}_t)}{p^0(\mathbf{y}_t|\hat{\mathbf{x}}_{t|t-1}^0)}}_{\beta_t} \ge h^f}_{g_m^f} \right\} \tag{15}$$

Algorithm 1 The centralized attack detector

1: Initialization: $t \leftarrow 0$, $g_0^f \leftarrow 0$
2: **while** $t < T^f$ **do**
3: $t \leftarrow t+1$
4: Implement the prediction step of the Kalman filter using (16).
5: Compute $\hat{\mathbf{a}}_t$, β_t, and g_t^f using (19), (20), and (21), respectively.
6: Implement the measurement update step of the Kalman filter using (17).
7: **if** $g_t^f = 0$ **then**
8: $\hat{\mathbf{x}}_{t|t}^f \leftarrow \hat{\mathbf{x}}_{t|t}^0$
9: **else if** $g_t^f \ge h^f$ **then**
10: $T^f \leftarrow t$
11: **end if**
12: **end while**
13: Declare the attack and stop the procedure.

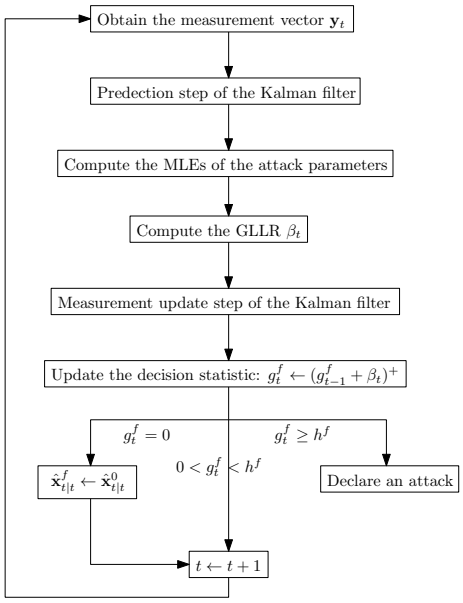

Fig. 2: A graphical description of Algorithm 1.

respectively. The generalized CUSUM test is given by

$$T^d = \inf\left\{ m \in \mathbb{N} : \underbrace{\max_{1\le j\le m} \sum_{t=j}^{m} \underbrace{\sup_{\mathcal{S}_t^d} \log \frac{p^d(\mathbf{y}_t \mid \hat{\mathbf{x}}_{t|t-1}^d)}{p^0(\mathbf{y}_t|\hat{\mathbf{x}}_{t|t-1}^0)}}_{\rho_t}}_{g_m^d} \ge h^d \right\}, \tag{22}$$

where g_m^d is the decision statistic at time m, ρ_t is the GLLR calculated at time t, and h^d is the test threshold.

The noise levels in the measurement models given in (2) and (11) are different. Hence, in addition to the state estimates, estimates of the state covariance matrices are also different for different hypotheses. Let $\mathbf{P}_{t|t'}^0$ and $\mathbf{P}_{t|t'}^d$ be such estimates for the null and alternative hypotheses, respectively. The Kalman filter equations at time t are then given as follows:

Prediction:

$$\begin{aligned}
\hat{\mathbf{x}}_{t|t-1}^0 &= \mathbf{A}\hat{\mathbf{x}}_{t-1|t-1}^0,\\
\hat{\mathbf{x}}_{t|t-1}^d &= \mathbf{A}\hat{\mathbf{x}}_{t-1|t-1}^d,\\
\mathbf{P}_{t|t-1}^0 &= \mathbf{A}\mathbf{P}_{t-1|t-1}^0\mathbf{A}^T + \sigma_v^2\,\mathbf{I}_N,\\
\mathbf{P}_{t|t-1}^d &= \mathbf{A}\mathbf{P}_{t-1|t-1}^d\mathbf{A}^T + \sigma_v^2\,\mathbf{I}_N,
\end{aligned} \tag{23}$$

Measurement update:

$$\begin{aligned}
\mathbf{G}_t^0 &= \mathbf{P}_{t|t-1}^0\mathbf{H}^T(\mathbf{H}\mathbf{P}_{t|t-1}^0\mathbf{H}^T + \sigma_w^2\,\mathbf{I}_K)^{-1},\\
\mathbf{G}_t^d &= \mathbf{P}_{t|t-1}^d\mathbf{H}^T(\mathbf{H}\mathbf{P}_{t|t-1}^d\mathbf{H}^T + \mathbf{\Lambda}_t)^{-1},
\end{aligned}$$

$$\begin{aligned}
\hat{\mathbf{x}}_{t|t}^0 &= \hat{\mathbf{x}}_{t|t-1}^0 + \mathbf{G}_t^0(\mathbf{y}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1}^0),\\
\hat{\mathbf{x}}_{t|t}^d &= \hat{\mathbf{x}}_{t|t-1}^d + \mathbf{G}_t^d(\mathbf{y}_t - \mathbf{u}_t),\\
\mathbf{P}_{t|t}^0 &= \mathbf{P}_{t|t-1}^0 - \mathbf{G}_t^0\mathbf{H}\mathbf{P}_{t|t-1}^0,\\
\mathbf{P}_{t|t}^d &= \mathbf{P}_{t|t-1}^d - \mathbf{G}_t^d\mathbf{H}\mathbf{P}_{t|t-1}^d,
\end{aligned} \tag{24}$$

where $\mathbf{\Lambda}_t = \mathrm{diag}(\lambda_{1,t}, \ldots, \lambda_{K,t})$ is a diagonal matrix with the following diagonal terms:

$$\lambda_{k,t} = \begin{cases} \sigma_n^2, & \text{if } \ k \in \hat{\mathcal{S}}_t^d \\ \sigma_w^2, & \text{if } \ k \notin \hat{\mathcal{S}}_t^d. \end{cases} \tag{25}$$

Furthermore, $\mathbf{u}_t = [u_{1,t}, \ldots, u_{K,t}]^T$ can be determined as

$$u_{k,t} = \begin{cases} 0, & \text{if } \ k \in \hat{\mathcal{S}}_t^d \\ \mathbf{h}_k^T\hat{\mathbf{x}}_{t|t-1}^d, & \text{if } \ k \notin \hat{\mathcal{S}}_t^d. \end{cases} \tag{26}$$

Note that the measurement update step of the Kalman filter for the alternative hypothesis is performed based on the MLE of the set of attacked meters. The following proposition presents the MLE of $\mathcal{S}_t^d$ and the GLLR at time t.

Proposition 2: The most likely set of attacked meters can be determined as

$$\hat{\mathcal{S}}_t^d = \Big\{ k : \frac{1}{\sigma_n^2} y_{k,t}^2 - \frac{1}{\sigma_w^2}(y_{k,t} - \mathbf{h}_k^T\hat{\mathbf{x}}_{t|t-1}^d)^2 < \log\left(\frac{\sigma_w^2}{\sigma_n^2}\right),\ k = 1, \ldots, K \Big\}, \tag{27}$$

and the GLLR can be computed as follows:

$$\rho_t = \frac{1}{2}\Bigg(K\log(\sigma_w^2) + \frac{1}{\sigma_w^2}\sum_{k=1}^{K}(y_{k,t} - \mathbf{h}_k^T\hat{\mathbf{x}}_{t|t-1}^0)^2 - \sum_{k\in\hat{\mathcal{S}}_t^d}\log(\sigma_n^2) + \frac{1}{\sigma_n^2}y_{k,t}^2 - \sum_{k\notin\hat{\mathcal{S}}_t^d}\log(\sigma_w^2) + \frac{1}{\sigma_w^2}(y_{k,t} - \mathbf{h}_k^T\hat{\mathbf{x}}_{t|t-1}^d)^2\Bigg). \tag{28}$$

Proof: See Appendix B.

Based on (22), recursion of the decision statistic $g_t^d, t \in \mathbb{N}$ can be written as

$$g_t^d = (g_{t-1}^d + \rho_t)^+, \tag{29}$$

where $g_0^d = 0$ and ρ_t is as given in (28). As before, if $g_t^d = 0$ for any $t \in \mathbb{N}$, then the change-point estimate is updated and consequently the Kalman filter estimates for the alternative hypothesis are updated as $\hat{\mathbf{x}}_{t|t}^d \leftarrow \hat{\mathbf{x}}_{t|t}^0$ and $\mathbf{P}_{t|t}^d \leftarrow \mathbf{P}_{t|t}^0$, respectively. We summarize the centralized DoS attack detector in Algorithm 1 after a few changes. Particularly, we replace T^f with T^d, g_t^f with g_t^d, and h^f with h^d. Moreover, we replace lines 4-6, and 8 of Algorithm 1 with the following lines.

4: Implement the prediction step of the Kalman filter using (23).

5: Compute $\hat{\mathcal{S}}_t^d$, $\mathbf{\Lambda}_t$, $\mathbf{u}_t$, ρ_t, and g_t^d based on (27), (25), (26), (28), and (29), respectively.

6: Implement the measurement update step of the Kalman filter using (24).

8: $\hat{\mathbf{x}}_{t|t}^d \leftarrow \hat{\mathbf{x}}_{t|t}^0$, $\mathbf{P}_{t|t}^d \leftarrow \mathbf{P}_{t|t}^0$

IV. Distributed Attack Detectors

In practice, the interconnected power grid is composed of several geographically separated subregions and each subregion contains a different set of meters. Thus, gathering and processing measurements obtained by all meters at a single place is infeasible, especially in large grids. Hence, a distributed implementation is needed. In the distributed setting, we consider a hierarchical structure where there are several local (control) centers and a global (control) center (cf. Fig. 4). In particular, each subregion is supervised by a local center that collects the measurements in its subregion, performs some computational tasks, and communicates with the neighboring local centers and with the global center through ideal (error-free) communication channels. We assume that (i) there is an ample bandwidth between any two neighboring local centers, (ii) a local center has knowledge of only its measurements and the configuration of the whole power grid.

The global center is responsible for detecting an attack and it needs to compute decision statistics based on all measurements, as discussed in Section III. We assume that there are parallel communication channels between local centers and the global center and the resources, e.g., bandwidth, for communication are scarce. Hence, we propose that each local center calculates a local statistic based on the measurements collected in its subregion, then transmits a quantized version of it to the global center. The global center then detects and declares an attack (if any) based on the received messages from the local centers. Furthermore, the global center sends feedback signals to the local centers when necessary. Note that for transmission of local statistics from the local centers to the global center and for the feedback signals received from the global center, we assume instantaneous communication.

In the following, we firstly describe the system model in the distributed setting, then explain the distributed state estimation procedure, and finally present the distributed detection algorithms for FDI attacks and DoS attacks, respectively.

A. System Model in the Distributed Setup

Suppose that there exist L subregions in the power grid and let $\mathcal{R}^\ell$ denote the set of meters inside the ℓth subregion, $\ell = 1, 2, \ldots, L$. We assume that a meter reports its measurements to only one local center. Hence, if the number of meters in the ℓth subregion is denoted with $K^\ell \triangleq |\mathcal{R}^\ell|$, then $\sum_{\ell=1}^{L} K^\ell = K$. The measurement vector $\mathbf{y}_t$ is then decoupled into L subvectors where $\mathbf{y}_t^\ell$, consisting of the set of measurements $\{y_{k,t} \,|\, k \in \mathcal{R}^\ell\}$, denotes the measurements collected in the ℓth subregion at time t.

We next need to determine the state vector of a local center. Based on the measurement model given in (2), $y_{k,t}, k \in \{1, \ldots, K\}$ can be written as

$$y_{k,t} = \mathbf{h}_k^T\mathbf{x}_t + w_{k,t},$$

where $\mathbf{h}_k^T = [h_{k,1}, \ldots, h_{k,N}]$ is the kth row of the measurement matrix $\mathbf{H}$, as defined before. Then, $y_{k,t}$ depends on, equivalently bears information about, the state variables corresponding to the nonzero entries of $\mathbf{h}_k$. Let the set of such state variables be denoted with $\mathcal{X}_{y_k} \triangleq \{x_n \,|\, n \in \{1, 2, \ldots, N\}, h_{k,n} \neq 0\}$. The state vector of the ℓth local center, which is denoted with $\mathbf{x}_t^\ell$ at time t, includes the state variables in $\mathcal{X}_{y_k}$ for all $k \in \mathcal{R}^\ell$. In fact, $\mathbf{x}_t^\ell$ may include further state variables due to dependencies between state variables over time.

In particular, due to the state transition matrix $\mathbf{A}$, evolution of a state variable over time depends on a set of state variables. Let $\boldsymbol{a}_n^T = [a_{n,1}, \ldots, a_{n,N}]$ be the nth row $\mathbf{A}$, i.e., $\mathbf{A}^T = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]$. Based on the state update equation given in (1), $x_{n,t}, n \in \{1, \ldots, N\}$ can be written as follows:

$$x_{n,t} = \boldsymbol{a}_n^T\mathbf{x}_{t-1} + v_{n,t}.$$

Thus, $x_{n,t}$ depends directly (over one-time period) on the state variables in the set $\{x_i \,|\, i \in \{1, 2, \ldots, N\}, a_{n,i} \neq 0\}$. Since the state variables belonging to this set may depend on several other state variables through (1), $x_{n,t}$ may indirectly (over multiple time periods) depend on a larger set of state variables. Including all direct and indirect dependencies, we finally obtain a set of state variables $\mathcal{X}_n$ that do not depend on any state variable outside the set $\mathcal{X}_n$. Hence, the evolution of x_n in time depends only on $\mathcal{X}_n$.

The state vector of the ℓth local center then consists of the state variables in the set:

$$\mathcal{X}_{\mathbf{y}^\ell} \triangleq \bigcup_{k\,\in\,\mathcal{R}^\ell}\Bigg\{\bigcup_{x_n\,\in\,\mathcal{X}_{y_k}}\{\{x_n\}\cup\mathcal{X}_n\}\Bigg\}.$$

Let $N^\ell \triangleq |\mathcal{X}_{\mathbf{y}^\ell}|$, then $\mathbf{x}_t^\ell \in \mathbb{R}^{N^\ell}$. Note that $\sum_{\ell=1}^{L} N^\ell \geq N$.

As a simple and illustrative example, consider a system with the state vector $\mathbf{x}_t = [x_{1,t}, x_{2,t}, x_{3,t}, x_{4,t}]^T$. Suppose that the system matrix and the measurement matrix are given as

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & -0.3 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0.5 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & -1 & 2 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

respectively. Moreover, let the system be composed of two subregions and the measurements obtained at the first and the second local centers at time t be $\mathbf{y}_t^1 = [y_{1,t}, y_{2,t}]^T$ and $\mathbf{y}_t^2 = [y_{3,t}, y_{4,t}, y_{5,t}]^T$, respectively. Then, $\mathcal{X}_{y_1} = \{x_1\}$, $\mathcal{X}_{y_2} = \{x_1, x_2\}$, $\mathcal{X}_{y_3} = \{x_2, x_3, x_4\}$, $\mathcal{X}_{y_4} = \{x_3, x_4\}$, and $\mathcal{X}_{y_5} = \{x_4\}$. Furthermore, $\mathcal{X}_1 = \{x_1, x_2, x_3\}$, $\mathcal{X}_2 = \{x_2\}$, $\mathcal{X}_3 = \{x_2, x_3\}$, and $\mathcal{X}_4 = \{x_4\}$. Therefore, $\mathcal{X}_{\mathbf{y}^1} = \{x_1, x_2, x_3\}$ and $\mathcal{X}_{\mathbf{y}^2} = \{x_2, x_3, x_4\}$. Then, the local state vectors are obtained as $\mathbf{x}_t^1 = [x_{1,t}, x_{2,t}, x_{3,t}]^T$ and $\mathbf{x}_t^2 = [x_{2,t}, x_{3,t}, x_{4,t}]^T$.

For the ℓth local center, the state update equation and the normal measurement model are then given by

$$\mathbf{x}_t^\ell = \mathbf{A}^\ell \mathbf{x}_{t-1}^\ell + \mathbf{v}_t^\ell, \tag{30}$$

$$\mathbf{y}_t^\ell = \mathbf{H}^\ell \mathbf{x}_t^\ell + \mathbf{w}_t^\ell, \tag{31}$$

where $\mathbf{A}^\ell \in \mathbb{R}^{N^\ell \times N^\ell}$ is the local system matrix, $\mathbf{v}_t^\ell$, a subvector of $\mathbf{v}_t$, is the local process noise vector corresponding to $\mathbf{x}_t^\ell$, $\mathbf{H}^\ell \in \mathbb{R}^{K^\ell \times N^\ell}$ is the local measurement matrix, and $\mathbf{w}_t^\ell \sim \mathcal{N}(\mathbf{0}, \sigma_w^2 \mathbf{I}_{K^\ell})$ is the local measurement noise vector. Note that $\mathbf{A}^\ell$ and $\mathbf{H}^\ell$ can be simply obtained from $\mathbf{A}$ and $\mathbf{H}$, respectively. Recall that for any two subregions ℓ and j, $\mathbf{y}_t^\ell$ and $\mathbf{y}_t^j$ do not overlap but $\mathbf{x}_t^\ell$ and $\mathbf{x}_t^j$ may overlap.

B. Distributed State Estimation Assuming Null Hypothesis

The information filter, or the inverse covariance filter, is an algebraic equivalence to the Kalman filter and its update rules are simpler and more convenient for a distributed setup [26], [40]. Hence, we use information filters for state estimation in the distributed setting. In particular, we employ two information filters (one for the null hypothesis and one for the alternative hypothesis) at each local center. With the exchange of necessary information between local centers, state estimation is performed in a fully distributed manner. In this section, we explain the proposed distributed state estimation procedure for one of the local centers, say the ℓth one, in case of no attack (null hypothesis). Note that the procedure is the same for all local centers.

Among all measurements taken system-wide, the state estimator of the ℓth local center needs to exploit the measurements that bear information about at least one of the state variables in $\mathbf{x}_t^\ell$. Clearly, $\mathbf{y}_t^\ell$ are among such measurements. Note that due to the tie-lines between neighboring subregions, some state variables can be shared between neighboring local centers. Moreover, due to the dependencies between state variables over time, some (neighboring or non-neighboring) local centers may have common state variables with the ℓth local center. Hence, it is possible that some measurements collected at the other local centers bear information about a nonempty subset of $\mathbf{x}_t^\ell$. Another challenge is that such measurements may be partially related to $\mathbf{x}_t^\ell$.

Let the set of local centers that share at least one state variable with the ℓth local center be denoted with $\mathcal{C}^\ell$, i.e.,

$$\mathcal{C}^\ell = \{i \,|\, i \in \{1, 2, \ldots, L\}, \mathcal{X}_{\mathbf{y}^i} \cap \mathcal{X}_{\mathbf{y}^\ell} \neq \emptyset\},$$

where $\emptyset$ is an empty set. Suppose that $j \in \mathcal{C}^\ell$. Then, let $\mathcal{X}_{\mathbf{y}^{\ell,j}}$ be the set of shared state variables between the ℓth and the jth local centers, i.e.,

$$\mathcal{X}_{\mathbf{y}^{\ell,j}} \triangleq \mathcal{X}_{\mathbf{y}^\ell} \cap \mathcal{X}_{\mathbf{y}^j},$$

and let the set of state variables included in the jth local center but not included in the ℓth local center be

$$\mathcal{X}_{\mathbf{y}^{\bar{\ell},j}} \triangleq \mathcal{X}_{\mathbf{y}^j} \backslash \mathcal{X}_{\mathbf{y}^\ell}.$$

Furthermore, let the subset of measurements of the jth local center at time t that bear information about $\mathcal{X}_{\mathbf{y}^{\ell,j}}$ be denoted with $\mathbf{y}_t^{\ell,j}$, which is given, with an abuse of notation, by

$$\mathbf{y}_t^{\ell,j} \triangleq \{y_{k,t} \,|\, k \in \mathcal{R}^j, \mathcal{X}_{y_k} \cap \mathcal{X}_{\mathbf{y}^\ell} \neq \emptyset\}.$$

If $\mathcal{X}_{\mathbf{y}^{\bar{\ell},j}}$ is a nonempty set, then $\mathbf{y}_t^{\ell,j}$ may also depend on some state variables that are not included in $\mathcal{X}_{\mathbf{y}^\ell}$. Since our aim is to use $\mathbf{y}_t^{\ell,j}$ in the estimation of $\mathbf{x}_t^\ell$, the non-included state variables cannot be considered as unknown from the perspective of the state estimator of the ℓth local center. Hence, the corresponding measurements should be adjusted by subtracting the part about the non-included states. To that end, the non-included states can be replaced with their estimates. Note that the non-included states need to be replaced with different state estimates under different hypotheses.

Let $\{{\mathbf{h}_k^j}^T \,|\, k \in \mathcal{R}^j\}$ denote the rows of the local measurement matrix $\mathbf{H}^j$ and suppose there exists a measurement $y_{k,t} \in \mathbf{y}_t^{\ell,j}$. Then, based on (31), $y_{k,t}$ can be written as follows:

$$y_{k,t} = {\mathbf{h}_k^j}^T \mathbf{x}_t^j + w_{k,t}. \tag{32}$$

Let $\mathbf{x}_t^{\bar{\ell},j}$ be the state vector consisting of the state variables in $\mathcal{X}_{\mathbf{y}^{\bar{\ell},j}}$. We then decompose the term ${\mathbf{h}_k^j}^T \mathbf{x}_t^j$ in (32) into two parts as follows:

$${\mathbf{h}_k^j}^T \mathbf{x}_t^j = {\mathbf{h}_k^{\ell,j}}^T \mathbf{x}_t^\ell + {\mathbf{h}_k^{\bar{\ell},j}}^T \mathbf{x}_t^{\bar{\ell},j}, \tag{33}$$

where the vectors $\mathbf{h}_k^{\ell,j}$ and $\mathbf{h}_k^{\bar{\ell},j}$ are determined to satisfy the equality in (33) for all t. The processed measurement, by the jth local center for the estimation purposes of the ℓth local center under the null hypothesis, is denoted with $\tilde{y}_{k,t}^{0,\ell,j}$ and given as follows:

$$\begin{aligned} \tilde{y}_{k,t}^{0,\ell,j} &= y_{k,t} - {\mathbf{h}_k^{\bar{\ell},j}}^T \hat{\mathbf{x}}_{t|t-1}^{0,\bar{\ell},j} \\ &\simeq {\mathbf{h}_k^{\ell,j}}^T \mathbf{x}_t^\ell + w_{k,t}, \end{aligned} \tag{34}$$

where $\hat{\mathbf{x}}_{t|t-1}^{0,\bar{\ell},j}$ is the estimate of $\mathbf{x}_t^{\bar{\ell},j}$, calculated under the null hypothesis and at the prediction step of the information filter of the jth local center at time t.

Let $\tilde{\mathbf{y}}_t^{0,\ell,j}$, consisting of the set of processed measurements $\{\tilde{y}_{k,t}^{0,\ell,j} \,|\, k \in \mathcal{R}^j, y_{k,t} \in \mathbf{y}_t^{\ell,j}\}$, denote the processed measurement vector under the null hypothesis, which takes the following form:

$$\tilde{\mathbf{y}}_t^{0,\ell,j} \simeq \mathbf{H}^{\ell,j} \mathbf{x}_t^\ell + \mathbf{w}_t^{\ell,j}, \tag{35}$$

where $\mathbf{H}^{\ell,j}$ and $\mathbf{w}_t^{\ell,j}$ are the corresponding measurement matrix and measurement noise vector, respectively. Note that $\{{\mathbf{h}_k^{\ell,j}}^T \,|\, k \in \mathcal{R}^j, y_{k,t} \in \mathbf{y}_t^{\ell,j}\}$ correspond to the rows of $\mathbf{H}^{\ell,j}$.

Considering the same example given in the previous section, we have $\mathbf{y}_t^{1,2} = [-x_{2,t} - x_{3,t} + 2\,x_{4,t},\, x_{3,t} - x_{4,t}]^T + \mathbf{w}_t^{1,2}$, where $\mathbf{w}_t^{1,2} = [w_{3,t}, w_{4,t}]^T$. Then,

$$\begin{aligned}\tilde{\mathbf{y}}_t^{0,1,2} &= \mathbf{y}_t^{1,2} - [2\,\hat{x}_{4,t|t-1}^{0,j},\, -\hat{x}_{4,t|t-1}^{0,j}]^T \\ &\simeq [-x_{2,t} - x_{3,t},\ x_{3,t}]^T + \mathbf{w}_t^{1,2} \\ &= \begin{bmatrix}0 & -1 & -1\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x_{1,t}\\ x_{2,t}\\ x_{3,t}\end{bmatrix} + \mathbf{w}_t^{1,2} \\ &= \mathbf{H}^{1,2}\mathbf{x}_t^1 + \mathbf{w}_t^{1,2},\end{aligned}$$

where $\hat{x}_{4,t|t-1}^{0,j}$ is estimate of $x_{4,t}$, calculated at the jth local center under the null hypothesis. Next, we present the information filter equations and explain the distributed state estimation procedure (under the null hypothesis $\mathcal{H}_0$) for the ℓth local center.

Let the state estimate of the ℓth local center under the null hypothesis at time t be denoted with $\hat{\mathbf{x}}_{t|t'}^{0,\ell}$. Moreover, let $\mathbf{Z}_{t|t'}^{\ell}$ be the information matrix of the ℓth local center and $\mathbf{z}_{t|t'}^{0,\ell} = \mathbf{Z}_{t|t'}^{\ell}\hat{\mathbf{x}}_{t|t'}^{0,\ell}$ be the information vector of the ℓth local center under the null hypothesis. The information filter equations at time t at the ℓth local center under the null hypothesis are then given as follows:

Prediction:

$$\begin{aligned}\mathbf{Z}_{t|t-1}^{\ell} &= (\mathbf{I}_{N^\ell} - \mathbf{F}_{t-1}^{\ell})\mathbf{E}_{t-1}^{\ell}, \\ \mathbf{z}_{t|t-1}^{0,\ell} &= (\mathbf{I}_{N^\ell} - \mathbf{F}_{t-1}^{\ell})\mathbf{A}^{\ell-T}\mathbf{z}_{t-1|t-1}^{0,\ell},\end{aligned} \tag{36}$$

Measurement update:

$$\begin{aligned}\mathbf{Z}_{t|t}^{\ell} &= \mathbf{Z}_{t|t-1}^{\ell} + \frac{1}{\sigma_w^2}\Bigg(\mathbf{H}^{\ell T}\mathbf{H}^{\ell} + \sum_{j\,\in\,\mathcal{C}^\ell}\underbrace{\mathbf{H}^{\ell,j\,T}\mathbf{H}^{\ell,j}}_{\boldsymbol{\Upsilon}^{\ell,j}}\Bigg), \\ \mathbf{z}_{t|t}^{0,\ell} &= \mathbf{z}_{t|t-1}^{0,\ell} + \frac{1}{\sigma_w^2}\Bigg(\mathbf{H}^{\ell T}\mathbf{y}_t^{\ell} + \sum_{j\,\in\,\mathcal{C}^\ell}\underbrace{\mathbf{H}^{\ell,j\,T}\tilde{\mathbf{y}}_t^{0,\ell,j}}_{\boldsymbol{\upsilon}_t^{0,\ell,j}}\Bigg), \\ \mathbf{E}_t^{\ell} &= \mathbf{A}^{\ell-T}\mathbf{Z}_{t|t}^{\ell}\mathbf{A}^{\ell-1}, \\ \mathbf{F}_t^{\ell} &= \mathbf{E}_t^{\ell}\big(\mathbf{E}_t^{\ell} + (1/\sigma_v^2)\mathbf{I}_{N^\ell}\big)^{-1},\end{aligned} \tag{37}$$

where $\mathbf{E}_t^{\ell}, \mathbf{F}_t^{\ell} \in \mathbb{R}^{N^\ell \times N^\ell}$ are auxiliary matrices at time t, $\boldsymbol{\Upsilon}^{\ell,j} \triangleq \mathbf{H}^{\ell,j\,T}\mathbf{H}^{\ell,j}$, and $\boldsymbol{\upsilon}_t^{0,\ell,j} \triangleq \mathbf{H}^{\ell,j\,T}\tilde{\mathbf{y}}_t^{0,\ell,j}$. Note that the structure of the information filter requires the matrices $\{\mathbf{A}^{\ell}\}_{\ell=1}^{L}$ to be invertible.

Since the ℓth local center knows the grid topology (and hence the matrices $\{\mathbf{H}^{\ell,j}\}_j$), it can easily compute the matrices $\{\boldsymbol{\Upsilon}^{\ell,j} \,|\, j \in \mathcal{C}^\ell\}$ defined above. However, $\boldsymbol{\upsilon}_t^{0,\ell,j}$ is calculated based on the measurements collected at the jth local center. Thus, each local center $j \in \mathcal{C}^\ell$ needs to compute and report $\boldsymbol{\upsilon}_t^{0,\ell,j}$ to the ℓth local center. Because only neighboring local centers are allowed to communicate with each other, multiple hops might be needed to send the required information entities. After receiving $\{\boldsymbol{\upsilon}_t^{0,\ell,j} \,|\, j \in \mathcal{C}^\ell\}$, the ℓth local center performs its measurement update step.

Remark 2: The proposed distributed state estimation procedure requires that all necessary communications between local centers are done before the next measurement interval. This can be achieved in practice since measurements in real power grid are currently taken with 15-minutes time intervals [41]. Hence, the possible communication delays due to multiple hops are not expected to affect the proposed procedure.

C. Distributed Detector for FDI Attacks

The attack vector $\mathbf{a}_t$ is decomposed into L sub-vectors where $\mathbf{a}_t^{\ell}$ consists of $\{\mathrm{a}_{k,t} \,|\, k \in \mathcal{R}^\ell\}$. The measurement vector of the ℓth local center in case of FDI attack is then given by

$$\mathbf{y}_t^{\ell} = \mathbf{H}^{\ell}\mathbf{x}_t^{\ell} + \mathbf{a}_t^{\ell} + \mathbf{w}_t^{\ell}.$$

Furthermore, let the processed measurements at the jth local center, $j \in \mathcal{C}^\ell$, under the alternative hypothesis ($\mathcal{H}_1^f$) be denoted with $\tilde{\mathbf{y}}_t^{f,\ell,j}$ and given by

$$\tilde{\mathbf{y}}_t^{f,\ell,j} = \mathbf{H}^{\ell,j}\mathbf{x}_t^{\ell} + \mathbf{a}_t^{\ell,j} + \mathbf{w}_t^{\ell,j}, \tag{38}$$

where $\tilde{\mathbf{y}}_t^{f,\ell,j}$ consists of $\{\tilde{y}_{k,t}^{f,\ell,j} \,|\, k \in \mathcal{R}^j, y_{k,t} \in \mathbf{y}_t^{\ell,j}\}$, that are obtained similar to (34) as follows:

$$\begin{aligned}\tilde{y}_{k,t}^{f,\ell,j} &= y_{k,t} - \mathbf{h}_k^{\bar{\ell},j\,T}\hat{\mathbf{x}}_{t|t-1}^{f,\bar{\ell},j} \\ &\simeq \mathbf{h}_k^{\ell,j\,T}\mathbf{x}_t^{\ell} + \mathrm{a}_{k,t} + w_{k,t},\end{aligned} \tag{39}$$

where $\hat{\mathbf{x}}_{t|t-1}^{f,\bar{\ell},j}$ is the estimate of $\mathbf{x}_t^{\bar{\ell},j}$, calculated at the jth local center under the alternative hypothesis. Moreover, $\mathbf{a}_t^{\ell,j}$ is the attack vector corresponding to $\mathbf{y}_t^{\ell,j}$.

Let the state estimate of the ℓth local center under the alternative hypothesis at time t be denoted with $\hat{\mathbf{x}}_{t|t'}^{f,j}$ and let $\mathbf{z}_{t|t'}^{f,\ell} = \mathbf{Z}_{t|t'}^{\ell}\hat{\mathbf{x}}_{t|t'}^{f,\ell}$ be the corresponding information vector at time t. Together with (36) and (37), the following equations form the information filter equations of ℓth local center at time t:

Prediction:

$$\mathbf{z}_{t|t-1}^{f,\ell} = (\mathbf{I}_{N^\ell} - \mathbf{F}_{t-1}^{\ell})\mathbf{A}^{\ell-T}\mathbf{z}_{t-1|t-1}^{f,\ell}, \tag{40}$$

Measurement update:

$$\begin{aligned}\mathbf{z}_{t|t}^{f,\ell} = \mathbf{z}_{t|t-1}^{f,\ell} + \frac{1}{\sigma_w^2}\Bigg(&\mathbf{H}^{\ell T}(\mathbf{y}_t^{\ell} - \hat{\mathbf{a}}_t^{\ell}) \\ &+ \sum_{j\,\in\,\mathcal{C}^\ell}\underbrace{\mathbf{H}^{\ell,j\,T}(\tilde{\mathbf{y}}_t^{f,\ell,j} - \hat{\mathbf{a}}_t^{\ell,j})}_{\boldsymbol{\upsilon}_t^{f,\ell,j}}\Bigg),\end{aligned} \tag{41}$$

where $\hat{\mathbf{a}}_t^{\ell}$ and $\hat{\mathbf{a}}_t^{\ell,j}$ are the MLEs of $\mathbf{a}_t^{\ell}$ and $\mathbf{a}_t^{\ell,j}$, respectively, and $\boldsymbol{\upsilon}_t^{f,\ell,j} \triangleq \mathbf{H}^{\ell,j\,T}(\tilde{\mathbf{y}}_t^{f,\ell,j} - \hat{\mathbf{a}}_t^{\ell,j})$.

At the ℓth local center, after the prediction step at time t, the state estimates are calculated as $\hat{\mathbf{x}}_{t|t-1}^{0,\ell} = \mathbf{Z}_{t|t-1}^{\ell^{-1}}\mathbf{z}_{t|t-1}^{0,\ell}$ and $\hat{\mathbf{x}}_{t|t-1}^{f,\ell} = \mathbf{Z}_{t|t-1}^{\ell^{-1}}\mathbf{z}_{t|t-1}^{f,\ell}$. Then, $\hat{\mathbf{a}}_t^{\ell}$ is calculated based on (19) where $e_{k,t} = y_{k,t} - \mathbf{h}_k^{\ell\,T}\hat{\mathbf{x}}_{t|t-1}^{f,\ell}$. Then, the ℓth local center calculates and transmits the information entities $\{\boldsymbol{\upsilon}_t^{0,j,l}\}_j$ and $\{\boldsymbol{\upsilon}_t^{f,j,l}\}_j$ to the corresponding local centers through its neighboring local centers. Moreover, it performs its measurement update step after receiving $\{\boldsymbol{\upsilon}_t^{0,\ell,j} \,|\, j \in \mathcal{C}^\ell\}$ and $\{\boldsymbol{\upsilon}_t^{f,\ell,j} \,|\, j \in \mathcal{C}^\ell\}$.

Sampling and Transmission of Local Statistics

Based on (20), the local statistic at the ℓth local center is calculated as follows:

$$\begin{aligned}\beta_t^{\ell} \triangleq \frac{1}{2\sigma_w^2}\sum_{k\,\in\,\mathcal{R}^\ell}\Bigg(&\big(y_{k,t} - \mathbf{h}_k^{\ell\,T}\hat{\mathbf{x}}_{t|t-1}^{0,\ell}\big)^2 \\ &-\big(y_{k,t} - \mathbf{h}_k^{\ell\,T}\hat{\mathbf{x}}_{t|t-1}^{f,\ell} - \hat{\mathrm{a}}_{k,t}^{\ell}\big)^2\Bigg).\end{aligned} \tag{42}$$

The local center then sends a summary of $\{\beta_t^\ell\}_t$ to the global center. Recall that the communication channels between the local centers and the global center are bandlimited and therefore only quantized versions of the local statistics can be transmitted to the global center. We propose two sampling schemes for the local centers: US and LCSH [42]. If the conventional sampling is used, the range of the possible values of the local statistic and the quantization levels are determined. Then, the local statistic is sampled at each predetermined sampling time, quantized according to the quantization levels, and the corresponding finite bit sequence is transmitted to the global center.

In the LCSH scheme, the amplitude axis is uniformly partitioned with a spacing level Δ and the corresponding amplitude levels are determined a priori. A local center transmits a message to the global center only when the local statistic crosses a new amplitude level. If a lower/upper level is crossed, a sign bit 0/1 is transmitted. If more than one level are crossed simultaneously, then 1/0 is transmitted for each additional double/single crossings. For instance, let the most recently crossed level is Δ and the new value of the local statistic is 5.7Δ at a local center. Then, the bit sequence 110 is transmitted to the fusion center, where the first bit denotes the sign of the first crossing and the subsequent bits represent additional 3 crossings.

Let the maximum and minimum possible values of the local statistic be $\beta_{\max}^\ell$ and $\beta_{\min}^\ell$, respectively. If the conventional sampling is used, the interval between $\beta_{\min}^\ell$ and $\beta_{\max}^\ell$ is divided into 2^ν quantization intervals and ν bits are transmitted for β_t^ℓ indicating its quantization interval. The length of a quantization interval is $\eta^\ell \triangleq (\beta_{\max}^\ell - \beta_{\min}^\ell)/2^\nu$. The transmitted bit sequence for β_t^ℓ is the binary representation of

$$\zeta_t^\ell \triangleq \left\lfloor \frac{\beta_t^\ell - \beta_{\min}^\ell}{\eta^\ell} \right\rfloor \tag{43}$$

in ν bits. The global center, upon receiving the bit sequence from the ℓth local center, converts the bit sequence into its decimal form and obtain ζ_t^ℓ. Then, it determines the quantization level $\beta_{q,t}^\ell$ for the ℓth local center as follows:

$$\beta_{q,t}^\ell \triangleq \begin{cases} 0, & \text{if } \beta_{\min}^\ell + \zeta_t^\ell \eta^\ell \leq 0 < \beta_{\min}^\ell + (\zeta_t^\ell + 1)\eta^\ell \\ \beta_{\min}^\ell + (\zeta_t^\ell + 0.5)\eta^\ell, & \text{else.} \end{cases} \tag{44}$$

Note that the local statistics belonging to the quantization interval containing zero are mapped to zero as the quantization level. This is due to the fact that the local statistics before the attack are expected to take values around zero (cf. (42)). After receiving bit sequences from all local centers, the global center updates the decision statistic at time t based on (21) as follows:

$$g_t^f = \left(g_{t-1}^f + \sum_{\ell=1}^{L} \beta_{q,t}^\ell \right)^+. \tag{45}$$

If the LCSH is used as the sampling scheme, the interval between $\beta_{\min}^\ell$ and $\beta_{\max}^\ell$ is uniformly partitioned into subintervals with spacing Δ. Let the most recently crossed amplitude level by the local statistic in terms of Δ be ψ^ℓ and let the most recent sampling time be ϱ_i^ℓ. The next sampling instant

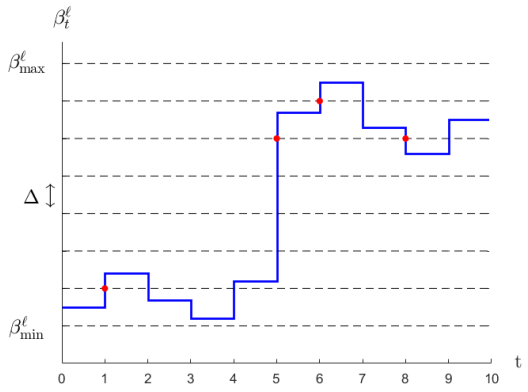

Fig. 3: LCSH scheme at the ℓth local center. The pairs of sampling times and the corresponding crossed levels are indicated with red points.

is determined as

$$\varrho_{i+1}^\ell = \min\left\{ t \in \mathbb{N} \,|\, t > \varrho_i^\ell,\, |\beta_t^\ell - \psi^\ell \Delta| \geq \Delta \right\}.$$

At $t = \varrho_{i+1}^\ell$, the corresponding sign bit is given by

$$\pi_t^\ell \triangleq \operatorname{sgn}\left(\beta_t^\ell - \psi^\ell \Delta\right), \tag{46}$$

where $\operatorname{sgn}(\cdot)$ is the sign function, the number of level crossings is given by

$$\phi_t^\ell \triangleq \left\lfloor \frac{|\beta_t^\ell - \psi^\ell \Delta|}{\Delta} \right\rfloor \geq 1, \tag{47}$$

and the number of transmitted bits equal to the following:

$$\varpi_t^\ell \triangleq \left\lceil \frac{\phi_t^\ell - 1}{2} \right\rceil + 1. \tag{48}$$

Furthermore, the most recently crossed amplitude level is updated as $\psi^\ell \leftarrow \psi^\ell + \pi_t^\ell \phi_t^\ell$ at both the ℓth local center and the global center. LCSH scheme at the ℓth local center is illustrated in Fig. 3. The global center, upon receiving the bit sequences from the local centers, updates the global decision statistic at time t based on (21) as follows:

$$g_t^f = \left(g_{t-1}^f + \Delta \sum_{\ell=1}^{L} \psi^\ell \right)^+. \tag{49}$$

In both sampling schemes, if $g_t^f = 0$ at any time t, the change-point estimate is updated and hence the Kalman filter for the alternative hypothesis need to be updated. In order to notify the local state estimators, the global center sends a feedback signal to the all local centers such that upon receiving this signal, each local center updates its information vector for the alternative hypothesis as being equal to the information vector for the null hypothesis. The proposed algorithms for a local center and for the global center are summarized in Algorithm 2 and Algorithm 3, respectively.

D. Distributed Detector for DoS Attacks

For the ℓth local center, measurement model in case of DoS attack is given by

$$y_{k,t} = \begin{cases} n_{k,t}, & \text{if } k \in \mathcal{S}_t^{d,\ell} \\ \mathbf{h}_k^{\ell\,T} \mathbf{x}_t^\ell + w_{k,t}, & \text{if } k \notin \mathcal{S}_t^{d,\ell}, \end{cases} \tag{50}$$

Algorithm 2 The distributed attack detector: procedure at the ℓth local center

Initialization: $t \leftarrow 0$, $\psi^\ell \leftarrow 0$
while $t < T^f$ **do**
  $t \leftarrow t+1$
  Implement the prediction step of the local information filter using (36) and (40).
  Compute $\hat{\mathbf{a}}_t^\ell$ based on (19) and β_t^ℓ as in (42).
  if the US scheme is used, **then**
    Compute ζ_t^ℓ as in (43) and transmit its binary equivalent in ν bits to the global center.
  else if the LCSH scheme is used, **then**
    if $|\beta_t^\ell - \psi^\ell \Delta| \geq \Delta$ **then**
      Compute π_t^ℓ and ϕ_t^ℓ as in (46) and (47), respectively.
      Transmit π_t^ℓ and ϕ_t^ℓ to the global center using ϖ_t^ℓ bits.
      $\psi^\ell \leftarrow \psi^\ell + \pi_t^\ell \phi_t^\ell$.
    end if
  end if
  Calculate and send $\{\boldsymbol{\upsilon}_t^{0,j,\ell}, \boldsymbol{\upsilon}_t^{f,j,\ell}\}_j$ to the corresponding local centers.
  Receive $\{\boldsymbol{\upsilon}_t^{0,\ell,j}, \boldsymbol{\upsilon}_t^{f,\ell,j} \mid j \in \mathcal{C}^\ell\}$.
  Implement the measurement update step of the local information filter using (37) and (41).
  if a feedback signal is received from the global center, **then**
    $\mathbf{z}_{t|t}^{f,\ell} \leftarrow \mathbf{z}_{t|t}^{0,\ell}$
  end if
end while

where $k \in \mathcal{R}^\ell$ and $\mathcal{S}_t^{d,\ell}$ denotes the set of attacked meters inside the ℓth subregion.

Let the vector of processed measurements at the jth local center, $j \in \mathcal{C}^\ell$, for the state estimator of the ℓth local center under the alternative hypothesis ($\mathcal{H}_1^d$) be denoted with $\tilde{\mathbf{y}}_t^{d,\ell,j}$, which consists of $\{\tilde{y}_{k,t}^{d,\ell,j} \mid k \in \mathcal{R}^j, y_{k,t} \in \mathbf{y}_t^{\ell,j}\}$ where

$$\tilde{y}_{k,t}^{d,\ell,j} = \begin{cases} n_{k,t}, & \text{if } k \in \hat{\mathcal{S}}_t^{d,j} \\ y_{k,t} - \mathbf{h}_k^{\bar{\ell},j^T} \hat{\mathbf{x}}_{t|t-1}^{d,\bar{\ell},j}, & \text{if } k \notin \hat{\mathcal{S}}_t^{d,j}, \end{cases} \tag{51}$$

where $\hat{\mathcal{S}}_t^{d,j}$ is the MLE of $\mathcal{S}_t^{d,j}$ and given in (55). Moreover, $\hat{\mathbf{x}}_{t|t-1}^{d,\bar{\ell},j}$ is the estimate of $\mathbf{x}_t^{\bar{\ell},j}$, calculated at the jth local center under the alternative hypothesis.

Let the information vector under the alternative hypothesis at time t for the ℓth local center be denoted with $\mathbf{z}_{t|t'}^{d,\ell}$. Moreover, let the corresponding information matrix be denoted with $\mathbf{Z}_{t|t'}^{d,\ell}$. Together with (36) and (37), the following equations form the local information filter equations:

Prediction:

$$\begin{aligned} \mathbf{Z}_{t|t-1}^{d,\ell} &= (\mathbf{I}_{N^\ell} - \mathbf{F}_{t-1}^{d,\ell}) \mathbf{E}_{t-1}^{d,\ell}, \\ \mathbf{z}_{t|t-1}^{d,\ell} &= (\mathbf{I}_{N^\ell} - \mathbf{F}_{t-1}^{d,\ell}) \mathbf{A}^{\ell^{-T}} \mathbf{z}_{t-1|t-1}^{d,\ell}, \end{aligned} \tag{52}$$

Algorithm 3 The distributed attack detector: procedure at the global center

Initialization: $t \leftarrow 0$, $g_0^f \leftarrow 0$, $\psi^\ell \leftarrow 0$
while $t < T^f$ **do**
  $t \leftarrow t+1$
  if the US scheme is used, **then**
    Compute $\{\zeta_t^\ell,\ \ell = 1, \ldots, L\}$ based on the received bit sequences and $\{\beta_{q,t}^\ell,\ \ell = 1, \ldots, L\}$ using (44).
    $g_t^f \leftarrow \left(g_{t-1}^f + \sum_{\ell=1}^L \beta_{q,t}^\ell\right)^+$
  else if the LCSH scheme is used, **then**
    if a new bit sequence is received during $(t-1, t]$ from the ℓth local center, $\ell = 1, \ldots, L$ **then**
      $\psi^\ell \leftarrow \psi^\ell + \pi_t^\ell \phi_t^\ell, \quad \ell = 1, \ldots, L.$
      $g_t^f \leftarrow \left(g_{t-1}^f + \Delta \sum_{\ell=1}^L \psi^\ell\right)^+$
    end if
  end if
  if $g_t^f = 0$ **then**
    Send a feedback signal to all the local centers.
  else if $g_t^f \geq h^f$ **then**
    $T^f \leftarrow t$
    Declare the attack and send a stop signal indicating the stopping time T^f to all the local centers.
  end if
end while

Measurement update:

$$\begin{aligned} \mathbf{Z}_{t|t}^{d,\ell} &= \mathbf{Z}_{t|t-1}^{d,\ell} + \mathbf{H}^{\ell^T} \boldsymbol{\Lambda}_t^{\ell^{-1}} \mathbf{H}^\ell + \sum_{j \in \mathcal{C}^\ell} \mathbf{H}^{\ell,j^T} \boldsymbol{\Lambda}_t^{\ell,j^{-1}} \mathbf{H}^{\ell,j}, \\ \mathbf{z}_{t|t}^{d,\ell} &= \mathbf{z}_{t|t-1}^{d,\ell} + \mathbf{H}^{\ell^T} \boldsymbol{\Lambda}_t^{\ell^{-1}} (\mathbf{y}_t^\ell + \mathbf{b}_t^\ell) \\ &\quad + \sum_{j \in \mathcal{C}^\ell} \underbrace{\mathbf{H}^{\ell,j^T} \boldsymbol{\Lambda}_t^{\ell,j^{-1}} (\tilde{\mathbf{y}}_t^{d,\ell,j} + \mathbf{b}_t^j)}_{\boldsymbol{\upsilon}_t^{d,\ell,j}}, \\ \mathbf{E}_t^{d,\ell} &= \mathbf{A}^{\ell^{-T}} \mathbf{Z}_{t|t}^{d,\ell} \mathbf{A}^{\ell^{-1}}, \\ \mathbf{F}_t^{d,\ell} &= \mathbf{E}_t^{d,\ell} (\mathbf{E}_t^{d,\ell} + (1/\sigma_v^2) \mathbf{I}_{N^\ell})^{-1}, \end{aligned} \tag{53}$$

where $\boldsymbol{\upsilon}_t^{d,\ell,j} \triangleq \mathbf{H}^{\ell,j^T} \boldsymbol{\Lambda}_t^{\ell,j^{-1}} (\tilde{\mathbf{y}}_t^{d,\ell,j} + \mathbf{b}_t^j)$. Note that $\boldsymbol{\Lambda}_t^\ell \in \mathbb{R}^{K^\ell \times K^\ell}$ is a diagonal matrix with the diagonal elements $\{\lambda_{k,t}, k \in \mathcal{R}^\ell\}$ obtained as in (25) after replacing $\hat{\mathcal{S}}_t^d$ with $\hat{\mathcal{S}}_t^{d,\ell}$. Furthermore, $\mathbf{b}_t^\ell$ is a vector consisting of $\{b_{k,t}, k \in \mathcal{R}^\ell\}$ and computed as follows:

$$b_{k,t} = \begin{cases} \mathbf{h}_k^{\ell^T} \hat{\mathbf{x}}_{t|t-1}^{d,\ell}, & \text{if } \ k \in \hat{\mathcal{S}}_t^{d,\ell} \\ 0, & \text{if } \ k \notin \hat{\mathcal{S}}_t^{d,\ell}. \end{cases} \tag{54}$$

At the ℓth local center, after the prediction step, state estimates are computed as $\hat{\mathbf{x}}_{t|t-1}^{0,\ell} = \mathbf{Z}_{t|t-1}^{\ell^{-1}} \mathbf{z}_{t|t-1}^{0,\ell}$ and $\hat{\mathbf{x}}_{t|t-1}^{d,\ell} = \mathbf{Z}_{t|t-1}^{d,\ell^{-1}} \mathbf{z}_{t|t-1}^{d,\ell}$, and the MLE of the attacked subset of meters for the ℓth local center is determined based on (27) as follows:

$$\hat{\mathcal{S}}_t^{d,\ell} = \Big\{ k : \frac{1}{\sigma_n^2} y_{k,t}^2 - \frac{1}{\sigma_w^2} \big(y_{k,t} - \mathbf{h}_k^{\ell\,T} \hat{\mathbf{x}}_{t|t-1}^{d,\ell}\big)^2 < \log\Big(\frac{\sigma_w^2}{\sigma_n^2}\Big),\ k \in \mathcal{R}^\ell \Big\}, \tag{55}$$

Every local center $j \in \mathcal{C}^\ell$ needs to compute and send the information entities $\boldsymbol{\upsilon}_t^{d,\ell,j}$ to the ℓth local center. Moreover, the information filter at the ℓth local center needs to calculate $\{\mathbf{H}^{\ell,j\,T} \boldsymbol{\Lambda}_t^{\ell,j^{-1}} \mathbf{H}^{\ell,j}, j \in \mathcal{C}^\ell\}$ in the measurement update step. The local center already knows $\{\mathbf{H}^{\ell,j}, j \in \mathcal{C}^\ell\}$ but the information regarding $\boldsymbol{\Lambda}_t^{\ell,j}$ must be sent from the jth local center. Note that $\boldsymbol{\Lambda}_t^{\ell,j}$ is a diagonal matrix with diagonal elements $\{\lambda_{k,t} \,|\, k \in \mathcal{R}^j, y_{k,t} \in \mathbf{y}_t^{\ell,j}\}$. Hence,

$$\mathbf{H}^{\ell,j\,T} \boldsymbol{\Lambda}_t^{\ell,j^{-1}} \mathbf{H}^{\ell,j} = \sum_{k:\ k \in \mathcal{R}^j,\, y_{k,t} \in \mathbf{y}_t^{\ell,j}} \big(\mathbf{h}_k^{\ell,j} \mathbf{h}_k^{\ell,j\,T}\big)/\lambda_{k,t},$$

where $\lambda_{k,t}$ is equal to σ_n^2 if $k \in \hat{\mathcal{S}}_t^{d,j}$ and σ_w^2, otherwise. Since the number of possible values of $\lambda_{k,t}$ is only two, this information can be sent with one bit for each $\lambda_{k,t}$. Hence, the jth local center first computes $\hat{\mathcal{S}}_t^{d,j}$, then for the measurements corresponding to $\mathbf{h}_k^{\ell,j}$ vectors, transmits 1 if $k \in \hat{\mathcal{S}}_t^{d,j}$ and 0, otherwise. Upon receiving the corresponding bit sequence, the ℓth local center forms $\boldsymbol{\Lambda}_t^{\ell,j}$ matrix. Note that $\{\boldsymbol{\Lambda}_t^{\ell}\}_{\ell=1}^{L}$ and $\{\boldsymbol{\Lambda}_t^{\ell,j}\}_{\ell,j}$ are invertible since they are diagonal matrices with nonzero diagonal elements.

Based on (28), the local statistic at the ℓth local center at time t is determined as follows:

$$\begin{aligned}\rho_t^\ell = \frac{1}{2}\Bigg(& K^\ell \log(\sigma_w^2) + \frac{1}{\sigma_w^2} \sum_{k \in \mathcal{R}^\ell} \big(y_{k,t} - \mathbf{h}_k^{\ell\,T} \hat{\mathbf{x}}_{t|t-1}^{0,\ell}\big)^2 \\ & - \sum_{k \in \hat{\mathcal{S}}_t^{d,\ell}} \log(\sigma_n^2) + \frac{1}{\sigma_n^2} y_{k,t}^2 \\ & - \sum_{k \in \mathcal{R}^\ell \setminus \hat{\mathcal{S}}_t^{d,\ell}} \log(\sigma_w^2) + \frac{1}{\sigma_w^2} \big(y_{k,t} - \mathbf{h}_k^{\ell\,T} \hat{\mathbf{x}}_{t|t-1}^{d,\ell}\big)^2 \Bigg). \end{aligned} \tag{56}$$

Then, each local center performs either conventional sampling or level-crossing sampling on ρ_t^ℓ, as described in Section IV-C (for β_t^ℓ). The global center, upon receiving bit sequences from the local centers, updates the decision statistic. If the decision statistic crosses the test threshold, it declares an attack. If the decision statistic gets the value of zero after an update, it immediately sends feedback signals to all the local centers. A local center, upon receiving this signal, equates its local information matrix and vector for the alternative hypothesis to the local information matrix and vector for the null hypothesis, respectively. The proposed procedure for a local center is summarized in Algorithm 2 after a few changes. In particular, T^f is changed with T^d, β_t^ℓ is changed with ρ_t^ℓ, and lines 4-5, 15-17, and 19 are changed with the following lines.

Furthermore, the procedure in the global center is summarized in Algorithm 3 after replacing T^f with T^d, g_t^f with g_t^d, and h^f with h^d.

4: Implement the prediction step of the local information filter using (36) and (52).

5: Compute $\hat{\mathcal{S}}_t^{d,\ell}$, $\boldsymbol{\Lambda}_t^\ell$, $\mathbf{b}_t^\ell$, and ρ_t^ℓ based on (55), (25), (54), and (56), respectively.

15: Calculate and send $\{\boldsymbol{\upsilon}_t^{0,j,\ell}\}_j$, $\{\boldsymbol{\upsilon}_t^{d,j,\ell}\}_j$, and the bit sequence corresponding to $\{\lambda_{k,t} \,|\, k \in \mathcal{R}^\ell, y_{k,t} \in \mathbf{y}_t^{j,\ell}, \mathbf{h}_k^{j,\ell} \neq \mathbf{0}\}_j$ to the corresponding local centers.

16: Receive $\boldsymbol{\upsilon}_t^{0,\ell,j}, \boldsymbol{\upsilon}_t^{d,\ell,j}$, and the bit sequence corresponding to $\{\lambda_{k,t} \,|\, k \in \mathcal{R}^j, y_{k,t} \in \mathbf{y}_t^{\ell,j}, \mathbf{h}_k^{\ell,j} \neq \mathbf{0}\}$ from every $j \in \mathcal{C}^\ell$.

17: Implement the measurement update step of the local information filter using (37) and (53).

19: $\mathbf{z}_{t|t}^{d,\ell} \leftarrow \mathbf{z}_{t|t}^{0,\ell}$, $\mathbf{E}_t^{d,\ell} \leftarrow \mathbf{E}_t^{0,\ell}$, $\mathbf{F}_t^{d,\ell} \leftarrow \mathbf{F}_t^{0,\ell}$

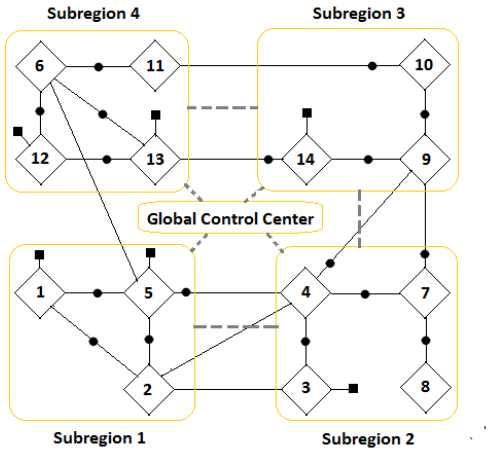

Fig. 4: IEEE-14 Bus Power System. Four subregions and the global control center are shown. Communication channels are illustrated with dashed lines. The circles on the branches represent the power-flow measurements, and the squares represent the power injection measurements. Bus 6 is chosen as the reference bus.

V. Numerical Results

In this section, performance of the proposed centralized and distributed cyber-attack detectors are evaluated in simple case studies. Throughout the section, simulations are performed on an IEEE-14 bus power system consisting of four subregions (see Fig. 4) and the measurement matrix $\mathbf{H}$ is determined accordingly. The system matrix $\mathbf{A}$ is chosen to be an identity matrix. In this system, $K = 23$ and $N = 13$. The initial state variables are obtained with the DC optimal power flow algorithm for case-14 in MATPOWER [43]. Noise variances are selected as $\sigma_v^2 = 10^{-4}$, $\sigma_w^2 = 2 \times 10^{-4}$, and $\sigma_n^2 = 4 \times 10^{-4}$. Moreover, γ is selected to be 0.18. The cyber-attacks are launched at $t = 100$.

We consider two types of false data: randomly created and carefully designed. If the attacker has incomplete knowledge about the network topology, it may randomly create the attack data. On the other hand, if it perfectly knows the topology, then it can perform structured attacks with false data lying on the column space of the measurement matrix, also known as stealth FDI attack. Further, in case of a DoS attack, the attacker can randomly choose the attacked meters. Next, we present performance of the proposed detectors in case of a random FDI attack, a structured FDI attack, and a random DoS attack,

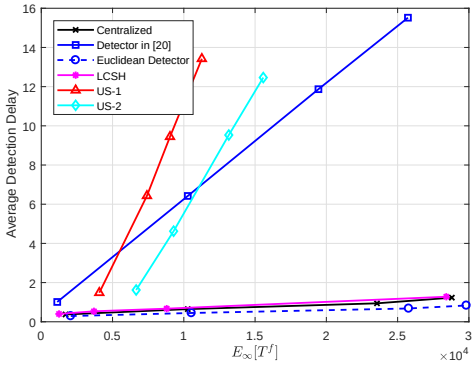

Fig. 5: Average detection delay versus average false alarm period in case of a random FDI attack for the proposed centralized and distributed FDI attack detectors and the detectors in [20] and [23].

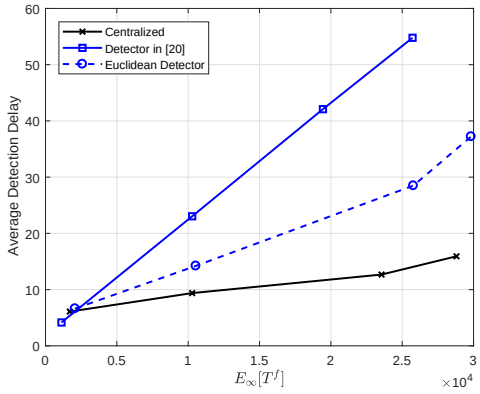

Fig. 6: Performance of the proposed centralized FDI attack detector and the detectors in [20] and [23] in case of a low-magnitude random FDI attack.

respectively and discuss the performance of the proposed centralized detectors. We then discuss the performance of the proposed distributed detectors.

A. Case 1: Random FDI Attack

Firstly, we consider a time-varying random FDI attack. Considering that the meters under the control of the attacker may be limited due to security measures, we specify 10 meters as subject to attacks out of 23 meters in the system. To make the attack randomized and time-varying, at each time the attacker first randomly determines the meters to attack (it chooses a meter with probability 0.5) and then injects realizations of a uniform random variable $\mathcal{U}[-0.2, 0.2]$ to the measurements of the compromised meters.

In Fig. 5, we present the delay vs. false alarm curves for the (centralized) detectors in [20] and [23] and our proposed centralized and distributed detectors. The detector in [20] is considered as a representative of LS-based detectors. Moreover, the detector [23] is considered as a representative of outlier detection techniques based on the Kalman filter. We observe that compared to [20], the average detection delays are significantly smaller in the proposed centralized detector for the same levels of false alarm period. This is due to (i) by using a dynamic state estimator, system/attack dynamics can be more effectively captured/detected compared to the conventional LS estimator, (ii) state forecasts/predictions provided by the Kalman filter are exploited to improve attack detection performance, and (iii) the detector in [20] is designed based on the assumption of a constant set of compromised meters over time where in the considered attack case, at each time possibly a different subset of compromised meters are selected by the attacker.

On the other hand, the Euclidean detector in [23] slightly outperforms the proposed detectors due to estimation errors incorporated into the proposed detection mechanisms. However, as explained before, detection-only schemes such as [23] do not provide any estimates about which part of the grid is attacked and in which magnitude. We note that it is, in fact, unfair to compare detection-only schemes with the schemes involving an estimation mechanism. Furthermore, for

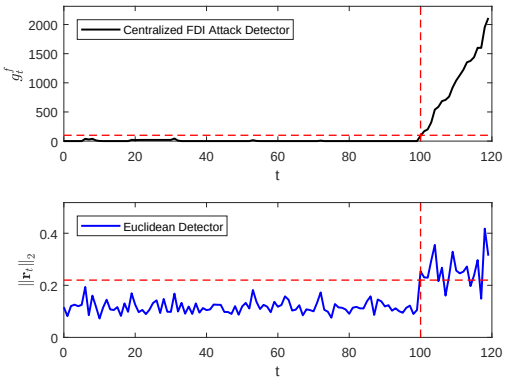

Fig. 7: Sample responses of the proposed centralized detector and the Euclidean detector in case of a random FDI attack, where $\mathbf{r}_t \triangleq \mathbf{y}_t - \mathbf{H}\hat{\mathbf{x}}^0_{t|t-1}$. The dashed lines (in red) indicate the test thresholds and the attack launch time. For both detectors, the thresholds are selected such that the average false alarm period is approximately 10^3.

lower magnitude attacks, the proposed detectors outperform the outlier detection schemes as we accumulate change (attack) statistics over time. We present in Fig. 6 the performance of the centralized proposed and benchmark detectors in case of a random FDI attack with attack magnitudes being realizations of $\mathcal{U}[-0.1, 0.1]$. Moreover, the outlier detection schemes may not be reliable due to sample-by-sample decisions, i.e., ignoring the accumulation of evidence obtained based on the history of measurements. To verify this claim, we present Fig. 7 where the sample responses of our centralized detector and the Euclidean detector are shown up to time $t = 120$ where the attack is launched at $t = 100$. Through the figure, we observe that the decision statistic of the Euclidean detector falls below its threshold during the attack period, indicating no attack and hence the corresponding measurements are declared as normal.

Further, in Fig. 8, we present the performance of the proposed centralized detector as the magnitude of the injected false data (amount of deviation from the null hypothesis $\mathcal{H}_0$) varies. We consider the random FDI attacks described

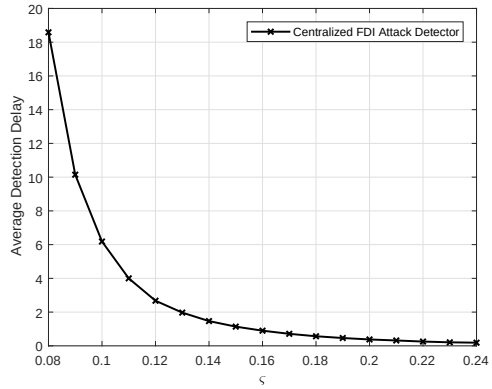

Fig. 8: Average detection delay versus magnitude of the injected false data in case of a random FDI attack for the proposed centralized detector, where $\mathbb{E}_\infty[T^f] \simeq 10^3$ and $\gamma = 0.18$.

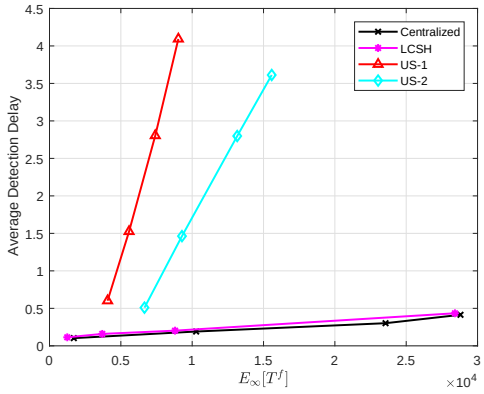

Fig. 9: Average detection delay versus average false alarm period in case of a structured FDI attack for the proposed centralized and distributed FDI attack detectors.

above but only vary the attack magnitude. Particularly, the attacker injects the realizations of $\mathcal{U}[-\varsigma, \varsigma]$ where ς takes values between 0.08 and 0.24. As expected, average detection delay gets smaller as ς increases. Note that the critical level for detection purposes is $\gamma = 0.18$. Hence, as ς gets smaller below the level of γ, the average detection delay significantly increases.

B. Case 2: Structured FDI Attack

Secondly, we evaluate the proposed centralized and distributed detectors under a time-varying structured FDI attack. In particular, the attacker chooses the false data at time t as $\mathbf{a}_t = \mathbf{H}\,\mathbf{c}_t$ where $\mathbf{c}_t = [c_{1,t}, c_{2,t}, \ldots, c_{N,t}]^T$ and $\{c_{n,t}, n = 1, 2, \ldots, N\}$ are realizations of $\mathcal{U}[-0.1, 0.1]$. Note that the attacker is assumed to have the capability of compromising any meter in the system and based on the realizations $\{c_{n,t}, n = 1, 2, \ldots, N\}$, the set of attacked meters can change over time. We present the delay vs. false alarm curves for the proposed centralized and distributed detectors in Fig. 9, which verifies that adapting a state-space model to the formulation and using the Kalman filter for state estimation enables the detection of the structured FDI attacks. Note that the detector in [20] and other LS-based detectors are only able to detect the part of the false data lying on the null space of the measurement matrix and hence they are unable to detect such structured attacks.

C. Case 3: Random DoS Attack

Finally, we consider a time-varying DoS attack where the attacked meters are randomly determined. In particular, at each time after the DoS attack is launched, any meter in the system is attacked with probability 0.1. We present the delay vs. false alarm curve in Fig. 10. Comparing the performance curves for the FDI and DoS attacks, we observe that the DoS attacks are relatively easier/quicker to detect compared to the FDI attacks. This is because in case of a DoS attack, measurements deviate more from the null hypothesis compared to an FDI attack with small to moderate attack magnitudes (cf. (6) and (12)).

D. Discussion on Performance of the Distributed Detectors

There are two sources of performance degradation in our distributed implementation: shared state variables between local centers and the quantization of local statistics. As explained in the distributed state estimation mechanism in Sec. IV, due to the shared states, a local center may need to use some information entities computed at the other local centers. However, computing such information entities contains loss of information since some state variables are replaced with their estimates. Furthermore, the global decision statistics are computed using the quantized versions of the local statistics, which leads to another loss of information. Hence, for a distributed detector, its centralized counterpart can be considered as a performance upper bound.

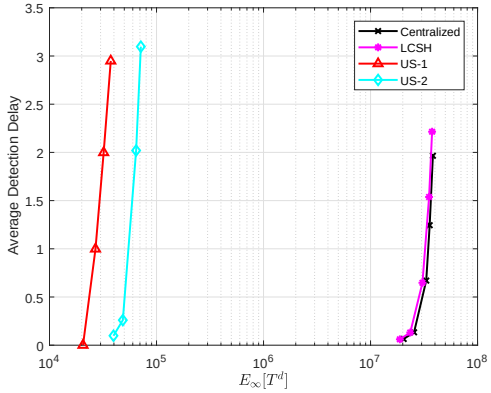

Fig. 10: Average detection delay versus average false alarm period in case of a random DoS attack for the proposed centralized and distributed DoS attack detectors.

Performance of the distributed detectors with US is expected to improve as the number of transmitted bits per time increases. However, in a distributed setting, practically the number of transmitted bits need to be limited due to resource constraints. Hence, we present results only for US with $\nu = 1$ (US-1) and $\nu = 2$ (US-2) cases. In the US-1 scheme, a local center transmits 1 bit per unit time and since there are 4 local centers in the power system (cf. Fig. 4), totally 4 bits are transmitted per unit time. Similarly, in the US-2 scheme, 8 bits are transmitted per unit time. Via an offline simulation, the Δ level in the LCSH scheme is determined such that 1 bit is

transmitted on average during the no-attack period. We observe through Figures 5, 9, and 10 that the distributed detectors with LCSH significantly outperform the distributed detectors with the conventional US-1 and US-2 schemes. This is due to the fact that the LCSH scheme is adaptive to the local statistics and hence provides a better summary of local statistics to the global center compared to the conventional sampling scheme.

VI. Conclusions

In this paper, we have studied real-time detection of cyber-attacks in the smart grid. We have used the Kalman filter for state estimation and the generalized CUSUM algorithm for a timely and reliable detection. We have proposed detection mechanisms in both centralized and distributed settings. The proposed detectors are robust to time-varying states, attack magnitudes, and the set of attacked meters, hence ideally suited for the smart grid which is a highly dynamic system. We have also provided closed-form online MLE estimates of the unknown attack variables. In the distributed setting, we have proposed a fully distributed dynamic state estimation procedure, and for resource concerns, we have proposed to use LCSH for the sampling and transmission of local decision statistics. We have illustrated via extensive simulations that the proposed detectors have quick and reliable responses to random and structured FDI attacks and DoS attacks. Moreover, we have shown that the proposed distributed detectors with LCSH perform quite closely to its centralized counterparts.

Appendix

A. Proof of Proposition 1

Proof. Based on the normal measurement model (cf. (2)), $\mathbf{y}_t \sim \mathcal{N}(\mathbf{H}\mathbf{x}_t, \sigma_w^2\,\mathbf{I}_K)$ and based on the measurement model in case of an FDI attack (cf. (3)), $\mathbf{y}_t \sim \mathcal{N}(\mathbf{H}\mathbf{x}_t + \mathbf{a}_t, \sigma_w^2\,\mathbf{I}_K)$. Then, based on (15) and noting that (i) for the set of meters $\{k \notin \mathcal{S}_t^f\}$, $\mathrm{a}_{k,t} = 0$ and (ii) taking supremum of a quantity is equivalent to taking infimum of the negative of the quantity, β_t can be written as in (57) (shown at the top of next page). The MLE of the attack vector, i.e., $\hat{\mathbf{a}}_t = [\hat{\mathrm{a}}_{1,t}, \ldots, \hat{\mathrm{a}}_{K,t}]^T$ is then calculated as follows:

$$\begin{aligned}\hat{\mathrm{a}}_{k,t} &= \underset{|\mathrm{a}_{k,t}|\geq\gamma,\, k\in\mathcal{S}_t^f}{\operatorname{argmin}} (e_{k,t} - \mathrm{a}_{k,t})^2 \\ &= \begin{cases} e_{k,t}, & \text{if } |e_{k,t}| \geq \gamma,\ k \in \mathcal{S}_t^f \\ \gamma, & \text{if } 0 \leq e_{k,t} < \gamma,\ k \in \mathcal{S}_t^f \\ -\gamma, & \text{if } -\gamma < e_{k,t} < 0,\ k \in \mathcal{S}_t^f \\ 0, & \text{if } k \notin \mathcal{S}_t^f. \end{cases}\end{aligned} \tag{58}$$

Then the MLE of $\mathcal{S}_t^f$ is given by

$$\hat{\mathcal{S}}_t^f = \underset{\mathcal{S}_t^f \subset \{1,\ldots,K\}}{\operatorname{argmin}} \sum_{k\in\mathcal{S}_t^f} (e_{k,t} - \hat{\mathrm{a}}_{k,t})^2 + \sum_{k\notin\mathcal{S}_t^f} e_{k,t}^2. \tag{59}$$

Based on (58) and (59), the most likely set of attacked meters can be determined as

$$\hat{\mathcal{S}}_t^f = \{k : |e_{k,t}| > \frac{\gamma}{2},\ k = 1, \ldots, K\}, \tag{60}$$

and combining (58) and (60), the MLE of the attack vector is obtained as follows:

$$\hat{\mathrm{a}}_{k,t} = \begin{cases} e_{k,t}, & \text{if } |e_{k,t}| \geq \gamma \\ \gamma, & \text{if } \frac{\gamma}{2} < e_{k,t} < \gamma \\ -\gamma, & \text{if } -\gamma < e_{k,t} < -\frac{\gamma}{2} \\ 0, & \text{else.} \end{cases}$$

Finally, we have

$$\beta_t = \frac{1}{2\sigma_w^2} \sum_{k=1}^{K} \Big((y_{k,t} - \mathbf{h}_k^T \hat{\mathbf{x}}_{t|t-1}^0)^2 - (y_{k,t} - \mathbf{h}_k^T \hat{\mathbf{x}}_{t|t-1}^f - \hat{\mathrm{a}}_{k,t})^2 \Big).$$

□

B. Proof of Proposition 2

Proof. Based on (22) and the measurement models given in (2) and (11), ρ_t can be written as in (61) (shown at the top of next page). The MLE of $\mathcal{S}_t^d$ is then given as

$$\begin{aligned}\hat{\mathcal{S}}_t^d &= \underset{\mathcal{S}_t^d \subset \{1,\ldots,K\}}{\arg\min} \Big\{ \sum_{k\in\mathcal{S}_t^d} \log(\sigma_n^2) + \frac{1}{\sigma_n^2} y_{k,t}^2 \\ &\qquad + \sum_{k\notin\mathcal{S}_t^d} \log(\sigma_w^2) + \frac{1}{\sigma_w^2}(y_{k,t} - \mathbf{h}_k^T \hat{\mathbf{x}}_{t|t-1}^d)^2 \Big\} \\ &= \Big\{ k : \frac{1}{\sigma_n^2} y_{k,t}^2 - \frac{1}{\sigma_w^2}(y_{k,t} - \mathbf{h}_k^T \hat{\mathbf{x}}_{t|t-1}^d)^2 \\ &\qquad < \log\left(\frac{\sigma_w^2}{\sigma_n^2}\right),\ k = 1, \ldots, K \Big\}.\end{aligned}$$

Then, based on $\hat{\mathcal{S}}_t^d$ and (61), ρ_t takes the following form:

$$\begin{aligned}\rho_t = \frac{1}{2}\Big(& K\log(\sigma_w^2) + \frac{1}{\sigma_w^2}\sum_{k=1}^{K}(y_{k,t} - \mathbf{h}_k^T \hat{\mathbf{x}}_{t|t-1}^0)^2 \\ & - \sum_{k\in\hat{\mathcal{S}}_t^d} \log(\sigma_n^2) + \frac{1}{\sigma_n^2} y_{k,t}^2 \\ & - \sum_{k\notin\hat{\mathcal{S}}_t^d} \log(\sigma_w^2) + \frac{1}{\sigma_w^2}(y_{k,t} - \mathbf{h}_k^T \hat{\mathbf{x}}_{t|t-1}^d)^2 \Big).\end{aligned}$$

□

References

[1] G. Liang, J. Zhao, F. Luo, S. Weller, and Z. Y. Dong, "A review of false data injection attacks against modern power systems," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2016.

[2] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, "The 2015 ukraine blackout: Implications for false data injection attacks," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3317–3318, July 2017.

[3] S. Amin, A. A. Cárdenas, and S. S. Sastry, *Safe and Secure Networked Control Systems under Denial-of-Service Attacks*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 31–45.

[4] A. Sargolzaei, K. Yen, M. Abdelghani, A. Abbaspour, and S. Sargolzaei, "Generalized attack model for networked control systems, evaluation of control methods," *Intelligent Control and Automation*, vol. 08, pp. 164–174, 2017.

[5] S. Asri and B. Pranggono, "Impact of distributed denial-of-service attack on advanced metering infrastructure," *Wireless Personal Communications*, vol. 83, no. 3, pp. 2211–2223, 2015.

[6] L. Xie, Y. Mo, and B. Sinopoli, "False data injection attacks in electricity markets," in *2010 First IEEE International Conference on Smart Grid Communications*, Oct 2010, pp. 226–231.

$$\beta_t = \sup_{\mathcal{S}_t^f} \log \frac{\sup_{|a_{k,t}|\geq\gamma,\, k\in\mathcal{S}_t^f} \exp\{\frac{-1}{2\sigma_w^2}(\mathbf{y}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1}^f - \mathbf{a}_t)^T(\mathbf{y}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1}^f - \mathbf{a}_t)\}}{\exp\{\frac{-1}{2\sigma_w^2}(\mathbf{y}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1}^0)^T(\mathbf{y}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1}^0)\}}$$

$$= \frac{1}{2\sigma_w^2}\left(\sum_{k=1}^{K}(y_{k,t} - \mathbf{h}_k^T\hat{\mathbf{x}}_{t|t-1}^0)^2 - \inf_{\mathcal{S}_t^f}\left\{\sum_{k\in\mathcal{S}_t^f}\inf_{|\mathrm{a}_{k,t}|\geq\gamma}(y_{k,t} - \mathbf{h}_k^T\hat{\mathbf{x}}_{t|t-1}^f - \mathrm{a}_{k,t})^2 + \sum_{k\notin\mathcal{S}_t^f}(y_{k,t} - \mathbf{h}_k^T\hat{\mathbf{x}}_{t|t-1}^f)^2\right\}\right) \quad (57)$$

$$\rho_t = \frac{1}{2}\left(K\log(\sigma_w^2) + \frac{1}{\sigma_w^2}\sum_{k=1}^{K}(y_{k,t} - \mathbf{h}_k^T\hat{\mathbf{x}}_{t|t-1}^0)^2 - \inf_{\mathcal{S}_t^d}\left\{\sum_{k\in\mathcal{S}_t^d}\log(\sigma_n^2) + \frac{1}{\sigma_n^2}y_{k,t}^2 + \sum_{k\notin\mathcal{S}_t^d}\log(\sigma_w^2) + \frac{1}{\sigma_w^2}(y_{k,t} - \mathbf{h}_k^T\hat{\mathbf{x}}_{t|t-1}^d)^2\right\}\right) \quad (61)$$

[7] 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*, ser. CCS '09. New York, NY, USA: ACM, 2009, pp. 21–32.

[8] S. Tan, D. De, W. Z. Song, J. Yang, and S. K. Das, "Survey of security advances in smart grid: A data driven approach," *IEEE Communications Surveys Tutorials*, vol. 19, no. 1, pp. 397–422, Firstquarter 2017.

[9] M. Esmalifalak, H. Nguyen, R. Zheng, and Z. Han, "Stealth false data injection using independent component analysis in smart grid," in *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Oct 2011, pp. 244–248.

[10] R. B. Bobba, K. M. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. J. Overbye, "Detecting false data injection attacks on dc state estimation," in *Preprints of the First Workshop on Secure Control Systems, CPSWEEK 2010*, 2010.

[11] K. C. Sou, H. Sandberg, and K. H. Johansson, "Electric power network security analysis via minimum cut relaxation," in *2011 50th IEEE Conference on Decision and Control and European Control Conference*, Dec 2011, pp. 4054–4059.

[12] M. A. Rahman and H. Mohsenian-Rad, "False data injection attacks with incomplete information against smart power grids," in *2012 IEEE Global Communications Conference (GLOBECOM)*, Dec 2012, pp. 3153–3158.

[13] V. Kekatos, G. B. Giannakis, and R. Baldick, "Grid topology identification using electricity prices," in *2014 IEEE PES General Meeting — Conference Exposition*, July 2014, pp. 1–5.

[14] X. Liu, Z. Bao, D. Lu, and Z. Li, "Modeling of local false data injection attacks with reduced network information," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1686–1696, July 2015.

[15] A. Anwar, A. N. Mahmood, and Z. Tari, "Identification of vulnerable node clusters against false data injection attack in an {AMI} based smart grid," *Information Systems*, vol. 53, pp. 201 – 212, 2015.

[16] Y. Deng and S. Shukla, "Vulnerabilities and countermeasures - a survey on the cyber security issues in the transmission subsystem of a smart grid," *Journal of Cyber Security and Mobility*, 2012.

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

[18] J. Chen and A. Abur, "Placement of pmus to enable bad data detection in state estimation," *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1608–1615, Nov 2006.

[19] S. Gong, Z. Zhang, H. Li, and A. D. Dimitrovski, "Time stamp attack in smart grid: Physical mechanism and damage analysis," *CoRR*, vol. abs/1201.2578, 2012. [Online]. Available: http://arxiv.org/abs/1201.2578

[20] S. Li, Y. Yilmaz, and X. Wang, "Quickest detection of false data injection attack in wide-area smart grids," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2725–2735, Nov 2015.

[21] Y. Huang, H. Li, K. A. Campbell, and Z. Han, "Defending false data injection attack on smart grid network using adaptive cusum test," in *2011 45th Annual Conference on Information Sciences and Systems*, March 2011, pp. 1–6.

[22] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[23] K. Manandhar, X. Cao, F. Hu, and Y. Liu, "Detection of faults and attacks including false data injection attack in smart grid using kalman filter," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 4, pp. 370–379, Dec 2014.

[24] D. B. Rawat and C. Bajracharya, "Detection of false data injection attacks in smart grid communication systems," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1652–1656, Oct 2015.

[25] B. Brumback and M. Srinath, "A chi-square test for fault-detection in kalman filters," *IEEE Transactions on Automatic Control*, vol. 32, no. 6, pp. 552–554, Jun 1987.

[26] R. Olfati-Saber, "Kalman-consensus filter : Optimality, stability, and performance," in *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, Dec 2009, pp. 7036–7042.

[27] S. Li and X. Wang, "Quickest attack detection in multi-agent reputation systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 4, pp. 653–666, Aug 2014.

[28] ——, "Cooperative change detection for voltage quality monitoring in smart grids," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 1, pp. 86–99, Jan 2016.

[29] A. Abur and A. Gomez-Exposito, *Power System State Estimation: Theory and Implementation*, 01 2004, vol. 24.

[30] S. Cui, Z. Han, S. Kar, T. T. Kim, H. V. Poor, and A. Tajer, "Coordinated data-injection attack and detection in the smart grid: A detailed look at enriching detection solutions," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 106–115, Sept 2012.

[31] H. V. Poor and O. Hadjiliadis, *Quickest Detection*. Cambridge University Press UK, 2009.

[32] A. Shiryaev, *Optimal Stopping Rules*. New York, NY: Springer, 2008.

[33] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, vol. 42, no. 6, pp. 1897–1908, 1971.

[34] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Statist.*, vol. 14, no. 4, pp. 1379–1387, 1986.

[35] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[36] Y. Li, L. Shi, P. Cheng, J. Chen, and D. E. Quevedo, "Jamming attacks on remote state estimation in cyber-physical systems: A game-theoretic approach," *IEEE Transactions on Automatic Control*, vol. 60, no. 10, pp. 2831–2836, Oct 2015.

[37] J. Gao, S. A. Vorobyov, H. Jiang, and H. V. Poor, "Worst-case jamming on mimo gaussian channels," *IEEE Transactions on Signal Processing*, vol. 63, no. 21, pp. 5821–5836, Nov 2015.

[38] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[39] S. Gezici, S. Bayram, M. N. Kurt, and M. R. Gholami, "Optimal jammer placement in wireless localization systems," *IEEE Transactions on Signal Processing*, vol. 64, no. 17, pp. 4534–4549, Sept 2016.

[40] A. G. O. Mutambara, *Decentralized Estimation and Control for Multisensor Systems*, 1st ed. Boca Raton, FL, USA: CRC Press, Inc., 1998.

[41] J. Jiang and Y. Qian, "Defense mechanisms against data injection attacks in smart grid networks," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 76–82, Oct 2017.

[42] Y. Yilmaz, G. Moustakides, X. Wang, and A. Hero, "Event-based statistical signal processing," in *Event-Based Control and Signal Processing*. CRC Press, Nov 2015, pp. 457–486.

[43] R. Zimmerman, C. Murillo-Sanchez, and R. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, Feb 2011.