MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Distributed Estimation and Detection of Cyber-Physical Attacks in Power Systems

Minot, A.; Sun, H.; Nikovski, D.N.; Zhang, J.

TR2019-035 June 25, 2019

Abstract

Dynamic state estimation, enabled by phasor measurement units (PMUs), opens new opportunities to improve detection of cyber-physical attacks in power networks. Distributed approaches to estimation and attack detection have many advantages, such as reduced processing times and increased security, and are arguably necessary for large size networks. In this work, we present a fully-distributed dynamic state estimation algorithm using PMU measurement data. The dynamic state estimation is jointly designed with an innovation-based attack detection scheme to limit communication overhead. An attractive feature of our work is that each control area utilizes a local model of reduced dimension. The design of our algorithm uses an approximation to the state covariance matrix, which allows for a trade-off between computation, communication, and accuracy. In numerical experiments, we demonstrate the effectiveness of this approach.

IEEE International Conference on Communications Workshops (ICC)

Distributed Estimation and Detection of Cyber-Physical Attacks in Power Systems

Ariana Minot
Lincoln Laboratory
Massachusetts Institute of Technology
Lexington, MA 02421, USA
ariana.minot@ll.mit.edu

Hongbo Sun, Daniel Nikovski, and Jinyun Zhang
Data Analytics Group
Mitsubishi Electric Research Laboratories
Cambridge, MA 02139, USA
{hongbosun, nikovski,jzhang}@merl.com

***Abstract*—Dynamic state estimation, enabled by phasor measurement units (PMUs), opens new opportunities to improve detection of cyber-physical attacks in power networks. Distributed approaches to estimation and attack detection have many advantages, such as reduced processing times and increased security, and are arguably necessary for large size networks. In this work, we present a fully-distributed dynamic state estimation algorithm using PMU measurement data. The dynamic state estimation is jointly designed with an innovation-based attack detection scheme to limit communication overhead. An attractive feature of our work is that each control area utilizes a local model of reduced dimension. The design of our algorithm uses an approximation to the state covariance matrix, which allows for a trade-off between computation, communication, and accuracy. In numerical experiments, we demonstrate the effectiveness of this approach.**

Keywords—cyber-physical attacks, distributed state estimation, dynamic state estimation, innovation-based attack detection, state covariance approximation

I. INTRODUCTION

With the advent of phasor measurement unit (PMU) sensors, it is becoming possible to do truly dynamic state estimation of electric power systems. One of the main motivations of promoting dynamic state estimation is for real-time monitoring and predictive capability to anticipate system problems. In particular, the dynamic state estimate can be used to more effectively monitor for line outages, faulty sensors, and cyber-physical attacks[1]. A cyber-physical attack may be undetectable from tampered measurements if there is a set of normal operating conditions consistent with the tampered measurements. Dynamic state estimation aids such circumstances, because although there may be a set of consistent normal operating conditions at any point in time, it is less likely that the set will remain consistent over many time steps [1].

For real-time security applications, distributed algorithms are critically important: 1) the reduced processing time compared to centralized schemes allows for quicker detection and remediation, which is critical for power systems where blackouts can spread quickly, and 2) by avoiding centralized processing, it is more difficult for attackers to stage a global attack on the entire system at a given time instant. In [2], a distributed algorithm is presented for detecting cyber-physical attacks in power networks using a sparse residual filter on a descriptor system model for the power system without consideration of noise presence. Given that noise is ubiquitous in real systems, we are interested in extending [2] to a noisy setting. The introduction of noise fundamentally changes certain aspects of the attack detection problem. Unlike in the noiseless case, the question of attack detectability no longer has a binary answer which can be assessed with deterministic filters. The noise introduces a level of ambiguity under which we assess how likely an attack has occurred. Rather than deterministic filters, the noise necessitates the use of inference algorithms, such as Kalman filters. There has been much recent interest on how to develop fully distributed Kalman filters. Several works have investigated the use of consensus-based [3] and diffusion-based [4] algorithms to accomplish this goal. One drawback of such methods is that they require each control area to process, communicate, and carry out computations on quantities of global dimension. For large systems, this becomes especially prohibitive. Instead, one would like to decompose the problem in such a way that each area solves a problem of reduced dimension. For an interconnected power system, the estimated state covariance used in the Kalman filter is in general full. Therefore, it is a non-trivial problem to devise local approximations to the state covariance. In [5], an approximate information filter is developed to achieve local, low-order filters. The use of iterative linear solvers and banded approximations in [5] inspires portions of our work. There are however several key differences. We focus specifically on the dynamic state estimation problem for power systems and consider the structure- preserving model for power system dynamics studied in various works [2], [6]. We find that the information filter formulation in [5] is not amenable for distributed processing under the structure-preserving model for power systems due to the resulting global coupling of the bus voltages with the generator rotor angles. Our formulation achieves a distributed solution by considering measurement noise with a non-block diagonal covariance matrix, as well as correlated process and measurement noise. These generalizations do not allow for a simple extension of the ideas in [5].

To this end, we propose a fully distributed dynamic state estimation and attack detection scheme. Both the generator state (*i.e.*, rotor angle and frequency of each generator) and the network state (*i.e.*, voltage phase angle at each bus) are considered in the estimation. Our dynamic state estimation uses iterative linear solvers and an approximation to the estimated state covariance matrix based on proximity. Such approximations are useful for modeling systems where correlations are expected to decay with distance [7], [8]. Under certain assumptions [9], [10], matrices related to power networks have a local, sparse structure. The main contributions of this work include: 1) A new decoupled state-space formulation for power system dynamic state estimation is proposed to avoid communication requirements between all

Ariana Minot conducted this research during her internship at MERL.

generators. 2) A fully distributed algorithm for attack detection based on dynamic state estimation is proposed that takes into account measurement noise. Such work is lacking under the structure-preserving dynamic model for power systems. 3) A local attack detection statistic is designed jointly with the dynamic state estimation in order to limit communication requirements. The communication requirements consist of buses' sharing their weighted measurement residuals within a user-specified neighborhood. The size of the neighborhood can be tuned to allow for a tradeoff between accuracy and communication overhead. Numerical experiments verify the effectiveness of using limited neighborhoods (e.g., "2-hop" neighbors).

II. Proposed Formulation

A. Preliminaries on the Structure-Preserving Model

Consider the linearized structure-preserving model from [6],[11]:

$$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M_{iner}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \dot{\boldsymbol{X}}(t) = \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{L_{gg}} & \boldsymbol{D_{damp}} & \boldsymbol{L_{gl}} \\ \boldsymbol{L_{lg}} & \boldsymbol{0} & \boldsymbol{L_{ll}} \end{bmatrix} \boldsymbol{X}(t) + \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{P^G}(t) \\ \boldsymbol{P^D}(t) \end{bmatrix}. \tag{1}$$

The state $\boldsymbol{X}(t) = [\boldsymbol{\delta}(t) \quad \boldsymbol{\omega}(t) \quad \boldsymbol{\theta}(t)]^T$ consists of the rotor angle $\boldsymbol{\delta}$ and frequency $\boldsymbol{\omega}$ at every generator and the voltage phase angle $\boldsymbol{\theta}$ at every bus. The matrix $\boldsymbol{M_{iner}}$ is a diagonal matrix with whose ith entry is M_i, the inertia of the ith generator. Similarly, the matrix $\boldsymbol{D_{damp}}$ is a diagonal matrix whose ith entry is D_i, the damping coefficient of the ith generator. The matrix $\boldsymbol{L} = \begin{bmatrix} \boldsymbol{L_{gg}} & \boldsymbol{L_{gl}} \\ \boldsymbol{L_{lg}} & \boldsymbol{L_{ll}} \end{bmatrix}$ is a Laplacian matrix with a sparsity structure related to the underlying network. We note that $\boldsymbol{L_{gg}}$ is a diagonal matrix, $\boldsymbol{L_{ll}}$ is related to the bus admittance matrix, and $\boldsymbol{L_{gl}} = \boldsymbol{L_{lg}^T}$. The control input $\boldsymbol{U}(t) = [\boldsymbol{0} \quad \boldsymbol{P^G}(t) \quad \boldsymbol{P^D}(t)]^T$ is given by the mechanical power output at the generators, $\boldsymbol{P^G}$, and the electrical power demand at each bus, $\boldsymbol{P^D}$, including those connected to a generator.

To extend upon the work in [1], [2], we consider process and measurement noise. Additionally, the measurements are taken to be a discrete-time process rather than a continuous-time process, which is more realistic for digitally sampled systems. The process noise $\boldsymbol{v}(t) \sim \mathcal{N}(\boldsymbol{0}, \overline{\boldsymbol{Q}})$) and measurement noise $\boldsymbol{n}[k] \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R})$) are assumed to be Gaussian noise processes. With the introduction of noise, we cannot apply a deterministic filter for attack detection as in [2]. For the noisy setting, we appeal to Kalman filtering techniques. However, since the dynamic system matrix in (1) is a singular matrix, the dynamic model in (1) is not directly applicable for Kalman filtering. One solution is to eliminate $\boldsymbol{\theta}$ from the state dynamics through the equation

$$\boldsymbol{\theta}(t) = \boldsymbol{L_{ll}^{-1}}(\boldsymbol{P^D}(t) - \boldsymbol{L_{lg}}\boldsymbol{\delta}(t)). \tag{2}$$

Although $\boldsymbol{L_{ll}}$ is sparse, its inverse is not, which leads to a coupling of the dynamics for $\boldsymbol{\omega}(t)$ amongst *all* generators that is not present in the original formulation (1). Such a global coupling makes developing a distributed solution with reasonable communication requirements infeasible.

B. Dynamic Model

To solve the problem described above, we propose to treat the voltage phase angle as a control input rather than eliminating it from the dynamics. Consider a network with a total of n buses. Let the subset of buses with generators be denoted G and the number of generators $|G| = n_g$. Then, the state given by the generator variables is

$$\boldsymbol{x}(t) = \begin{bmatrix} \boldsymbol{x}_1(t) & \boldsymbol{x}_2(t) & \cdots & \boldsymbol{x}_{n_g}(t) \end{bmatrix}^T, \tag{3}$$

$$\boldsymbol{x}_i(t) = [\delta_i(t) \quad \omega_i(t)]^T, \quad \forall i \in \{1, \ldots, n_g\}. \tag{4}$$

The dynamics for the generator rely only on local and neighboring quantities,

$$\dot{\delta}_i(t) = \omega_i(t) + v_{i,\delta}(t), \tag{5}$$

$$\dot{\omega}_i(t) = \tfrac{-1}{M_i Z_i}\delta_i(t) - \tfrac{D_i}{M_i}\omega_i(t) + \tfrac{1}{M_i}\left(P_{g,i} - \tfrac{1}{Z_i}\theta_i\right) + v_{i,\omega}(t), \tag{6}$$

where $v_{i,\delta}(t)$ and $v_{i,\omega}(t)$ are the process noise on δ_i and ω_i, respectively, and Z_i is the transient reactance. Therefore, these equations can be collected in matrix-form:

$$\dot{\boldsymbol{x}}(t) = \overline{\boldsymbol{A}}\boldsymbol{x}(t) + \boldsymbol{u}(t) + \boldsymbol{v}(t), \tag{7}$$

where $\overline{\boldsymbol{A}}$ is block diagonal. The control input is

$$\boldsymbol{u}(t) = \begin{bmatrix} u_{1,\delta}(t) & u_{1,\omega}(t) & \cdots & u_{n_g,\delta}(t) & u_{n_g,\omega}(t) \end{bmatrix}, \tag{8}$$

$$u_{i,\delta}(t) = 0, u_{i,\omega}(t) = \tfrac{1}{M_i}\left(P_{g,i} - \theta_i/Z_i\right). \tag{9}$$

The process noise is assumed to be $\boldsymbol{v}(t) \sim \mathcal{N}(\boldsymbol{0}, \overline{\boldsymbol{Q}})$ and uncorrelated in time. After digital sampling with sampling period $\overline{T}$, the continuous-time dynamic system is converted to a discrete-time system [12] as,

$$\boldsymbol{x}[k+1] = \boldsymbol{A}\boldsymbol{x}[k] + \boldsymbol{B}\boldsymbol{u}[k] + \boldsymbol{v}[k], \tag{10}$$

where

$$\boldsymbol{A} = e^{\overline{T}\overline{\boldsymbol{A}}}, \boldsymbol{B} = \int_{\boldsymbol{0}}^{\overline{T}} e^{\tau\overline{\boldsymbol{A}}}\, d\tau, \tag{11}$$

and $\boldsymbol{v}[k] \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q})$. The sampled process noise covariance matrix $\boldsymbol{Q}$ is related to the unsampled covariance matrix $\overline{\boldsymbol{Q}}$ via

$$\boldsymbol{Q} = \int_{\boldsymbol{0}}^{\overline{T}} e^{\boldsymbol{A}\tau}\, \overline{\boldsymbol{Q}} e^{\boldsymbol{A}^T\tau} d\tau. \tag{12}$$

C. Measurement Model

We consider measurements of the generator rotor frequencies $\boldsymbol{\omega}$ and the voltage phase angles $\boldsymbol{\theta}$. Similar to the dynamics, using $\boldsymbol{\theta}$ directly in the measurement equations yields the following measurement model:

$$\widehat{\omega_i}[k] = \omega_i[k] + n_{i,\omega}[k], \tag{13}$$

$$\widehat{\theta_i}[k] = \left[-\boldsymbol{L_{ll}^{-1}}\boldsymbol{L_{lg}}\right]_i \boldsymbol{\delta}[k] + [\boldsymbol{L_{ll}^{-1}}]_i \boldsymbol{P}_D[k] + n_{i,\theta}[k], \tag{14}$$

where $\boldsymbol{P}_D[k]$ is a known input. This formulation is not desirable for distributed processing since $\boldsymbol{L_{ll}^{-1}}$ couples the measurement of the phase angle $\widehat{\boldsymbol{\theta}}$ at any given bus to the rotor angle $\boldsymbol{\delta}$ at *all* generators and the electrical power demand $\boldsymbol{P}_D$ at *all* buses. We handle this by instead considering measurement of

$$\theta_i^*[k] \triangleq [\boldsymbol{L_{ll}}]_i \widehat{\boldsymbol{\theta}}[k] - P_i^D[k] \tag{15}$$

$$= \delta_i[k]/Z_i + n_{i,\theta^*}[k], \tag{16}$$

Let $\boldsymbol{y}[k] = [\boldsymbol{y}_1[k] \quad \ldots \quad \boldsymbol{y}_n[k]]^T$ where

$$\boldsymbol{y}_i[k] = \begin{cases} \widehat{\omega_i}[k]\hat{\theta}_i^*[k] & i \in G \\ \hat{\theta}_i^*[k] & i \notin G \end{cases} \tag{17}$$

is the measurement set local to bus i. Then $\boldsymbol{y}[k] = \boldsymbol{H}\boldsymbol{x}[k] + \boldsymbol{n}[k]$, yields a decoupled measurement model matrix $\boldsymbol{H}$ since in (13) and (16), the measurements $\boldsymbol{y}_i$ involve only local variables ω_i and δ_i. We stress that this new formulation maintains the original sparse, localized coupling inherent to power systems rather than a global coupling. The quantity θ_i^* in (15) is a linear combination of voltage phase angles at *neighboring* buses and thus can be calculated in a distributed fashion with limited communication. Furthermore, only the local electric power demand P_i^D is needed at each bus rather than the global $\boldsymbol{P^D}$.

However, measurement of $\boldsymbol{\theta}^*$ introduces the following complication. If $\boldsymbol{R}_{\theta}$ is the covariance matrix for the phase angle measurements, then $\boldsymbol{R}_{\theta^*} = \boldsymbol{L}_{ll}\boldsymbol{R}_{\theta}\boldsymbol{L}_{ll}^T$. In particular, if $\boldsymbol{R}_{\theta}$ is diagonal, this is no longer the case for $\boldsymbol{R}_{\theta^*}$. In Section III, we will show how $\boldsymbol{R}_{\theta^*}$ introduces coupling and communication requirements that depend on neighboring rather than global information. In summary, we have a new formulation that transfers all of the coupling to the measurement covariance matrices and process-measurement covariance.

D. Attack Model

We introduce the vectors $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$, which represent additive state and measurement attack vectors, respectively.

$$\boldsymbol{x}[k+1] = \boldsymbol{A}\boldsymbol{x}[k] + \boldsymbol{B}\boldsymbol{u}[k] + \boldsymbol{v}[k] + \boldsymbol{f}_1[k] \quad (18)$$

$$\boldsymbol{y}[k] = \boldsymbol{H}\boldsymbol{x}[k] + \boldsymbol{n}[k] + \boldsymbol{f}_2[k]. \quad (19)$$

In terms of hypothesis testing, our aim is to distinguish between

$$H_0(No\ Attack): \forall k, \boldsymbol{f}_1[k] = \boldsymbol{0}, \boldsymbol{f}_2[k] = \boldsymbol{0}.$$

$$H_1(Attack): There\ exist\ \{k^*\}\ for\ which\ \boldsymbol{f}_1[k^*] \neq \boldsymbol{0} \text{ and/or } \boldsymbol{f}_2[k^*] \neq \boldsymbol{0}. \quad (20)$$

Bad data from a faulty sensor can be viewed as one particular attack in this framework. This approach allows for a more general problem than bad data detection by including the possibility of directly attacking the state [2].

III. Distributed Algorithm For Estimation and Online Attack Detection

In this section, we present our distributed algorithm for joint dynamic state estimation and attack detection. The distributed and dynamic nature of our algorithm facilitates detecting attacks in an online fashion as new measurements become available, making this particularly attractive for monitoring the health of critical cyber-physical systems, such as power grids. The first task is to carry out dynamic state estimation, because our criterion for detecting attacks is a statistic based on the output of the dynamic state estimator.

A. Preliminaries on Kalman Filtering with Correlated Process and Measurement Noise

Due to the presence of $\boldsymbol{\theta}$ as a control input in (9) and in the measurements in (15), the process noise $\boldsymbol{v}[k]$ is correlated with the measurement noise $\boldsymbol{n}[k+1]$. Let

$$\mathbb{E}\left[\boldsymbol{v}[k]\boldsymbol{n}[j]\right] \stackrel{\text{def}}{=} \boldsymbol{M}\delta_{j,k+1}, \quad (21)$$

where δ is the Kronecker delta. The Kalman gain and state covariance estimation update formulas with process and measurement noise correlated according to $\boldsymbol{M}$ [13] are given below. The Kalman filter proceeds in two steps:

1) Dynamic Update:

$$\widehat{\boldsymbol{x}}_k^- = \boldsymbol{A}\widehat{\boldsymbol{x}}_{k-1}^+ + \boldsymbol{B}\boldsymbol{u}[k-1], \quad (22)$$

$$\boldsymbol{P}_k^- = \boldsymbol{A}\boldsymbol{P}_{k-1}^+\boldsymbol{A}^T + \boldsymbol{Q}. \quad (23)$$

2) Measurement Update:

$$\widehat{\boldsymbol{x}}_k^+ = \widehat{\boldsymbol{x}}_k^- + \boldsymbol{K}_k(\boldsymbol{y}[k] - \boldsymbol{H}\widehat{\boldsymbol{x}}_k^-), \quad (24)$$

$$\boldsymbol{S}_k = \boldsymbol{H}\boldsymbol{P}_k^-\boldsymbol{H}^T + \boldsymbol{H}\boldsymbol{M} + \boldsymbol{M}^T\boldsymbol{H}^T + \boldsymbol{R}, \quad (25)$$

$$\boldsymbol{K}_k = (\boldsymbol{P}_k^-\boldsymbol{H}^T + \boldsymbol{M})\boldsymbol{S}_k^{-1}, \quad (26)$$

$$\boldsymbol{P}_k^+ = \boldsymbol{P}_k^- - \boldsymbol{K}_k(\boldsymbol{H}\boldsymbol{P}_k^- + \boldsymbol{M}^T). \quad (27)$$

In the Dynamic Update step, the estimated state, $\widehat{\boldsymbol{x}}_{k-1}^+$, and estimated covariance, $\boldsymbol{P}_{k-1}^+$, are updated according to the system dynamics. In the Measurement Update step, the predicted state estimate, $\widehat{\boldsymbol{x}}_k^-$, and the predicted covariance estimate, $\boldsymbol{P}_k^-$, are updated with the measurements to produce the current estimate, $\widehat{\boldsymbol{x}}_k^+$ and $\boldsymbol{P}_k^+$.

B. Distributed Dynamic State Estimation

In addition to eliminating the need for communication with a centralized control center, we would like each control area to solve a problem of reduced dimension with respect to the original global problem. In this aim, we introduce the notion of local states and local measurements. The network buses are partitioned into a set of N control areas. For example, Fig. 1 is a 14 bus system with 4 control areas. The state local to control area I is the generator voltage angle and frequency at the buses in control area I, $\boldsymbol{x}_I = [\boldsymbol{\delta}_I \quad \boldsymbol{\omega}_I]^T$. There is no overlap between neighboring areas' states. The local measurements for control area I, $\boldsymbol{y}_I$, are the frequencies of all generators contained in the control area and the $\boldsymbol{\theta}_I^*$ at buses contained in the control area. We note that $\boldsymbol{\theta}_I^*$ depends on measurements of phase angles at neighboring buses, so border buses must exchange their measurements with their neighbors in other control areas.

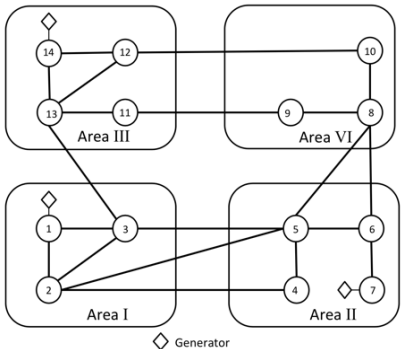


Fig. 1. Experimental setup of IEEE 14-bus network with 4 control areas.

One remarkable feature of the Kalman filter is that the estimated covariance matrices do not depend on the measurements. Therefore, they can be computed in advance offline. We stress again that the dynamic system matrix $\boldsymbol{A}$ and measurement model matrix $\boldsymbol{H}$ are both decoupled (*i.e.* no mixing is introduced between states in different areas). Assuming buses are labelled consecutively across control areas, the state-space model can be written as,

$$\boldsymbol{x}_I[k+1] = \boldsymbol{A}_I\boldsymbol{x}_I[k] + \boldsymbol{B}_I\boldsymbol{u}_I[k] + \boldsymbol{v}_I[k],$$

$$\boldsymbol{y}_I[k] = \boldsymbol{H}_I\boldsymbol{x}_I[k] + \boldsymbol{n}_I[k],$$

$$I = \{1, \dots, N\}.$$

Since the power network is an interconnected system, we expect a need to exchange information between control areas. Indeed, this need is reflected in the fact that the Kalman gain $\boldsymbol{K}_k$ is not a block-diagonal matrix. The innovation, or measurement residual, for control area I is defined as:

$$\boldsymbol{\gamma}_I[k] \stackrel{\text{def}}{=} \boldsymbol{y}_I[k] - \boldsymbol{H}_I[k]\widehat{\boldsymbol{x}}_I[k]^-. \quad (28)$$

The measurement update to the local estimate is then

$$\widehat{\boldsymbol{x}}_I[k]^+ = \widehat{\boldsymbol{x}}_I[k]^- + \boldsymbol{K}_I[k]\gamma_I[k], \quad (29)$$

where $\boldsymbol{K}_I[k]$ are the rows of $\boldsymbol{K}[k]$ corresponding to control area I. Since $\boldsymbol{K}[k]$ is not a block-diagonal matrix, entries of $\boldsymbol{\gamma}_I[k]$ will need to be communicated for a control area to update its local estimate $\widehat{\boldsymbol{x}}_I[k]^+$. Using our formulation, this is the remaining key challenge for developing a distributed algorithm. The formula for $\boldsymbol{K}[k]$ in (26) contains a matrix inverse $\boldsymbol{S}_k^{-1}$, which is difficult to calculate in a distributed way. Instead,

iterative linear solvers can be used as follows. Consider the linear system

$$\boldsymbol{S}_k \boldsymbol{a}[k] = \boldsymbol{\gamma}[k]. \quad (30)$$

Then, the measurement update in (29) is given as

$$\hat{\boldsymbol{x}}_I[k]^+ = \hat{\boldsymbol{x}}_I[k]^- + (\boldsymbol{P}_k^- \boldsymbol{H}^T + \boldsymbol{M})\boldsymbol{a}_I[k] \quad (31)$$

The key to dealing with the inverse in a distributed way is to iteratively solve (30) for $\boldsymbol{a}[k]$ without explicitly calculating the inverse and use the result in (31). We will show that the damped Jacobi method allows for a fully distributed solution to (30). Since the method is iterative, an inner-loop of iterations is introduced for each outer-loop k of the Kalman filter. We drop the outer-loop index k here for simplicity. The matrix $\boldsymbol{S}$ can be decomposed into the difference of a diagonal matrix $\boldsymbol{D}$ and a matrix containing the remaining off-diagonal entries,

$$\boldsymbol{S} = \boldsymbol{D} - \boldsymbol{E}. \quad (32)$$

Iteratively solving for $\boldsymbol{a}$ using the damped Jacobi method [14] amounts to finding the fixed point of

$$\boldsymbol{a}^{t+1} = \boldsymbol{a}^t + \alpha \boldsymbol{D}^{-1}(\boldsymbol{\gamma} - \boldsymbol{S}\boldsymbol{a}^t), \quad (33)$$

where α is the damping parameter. Since $\boldsymbol{D}$ is diagonal, its inverse is diagonal, and each block can be computed locally. The sparsity of $\boldsymbol{S}$ determines the entries from $\boldsymbol{a}^t$ that need to be communicated with neighboring areas. The following proposition guarantees the method's convergence when applied in our formulation for dynamic state estimation.

Proposition 1. *The damped Jacobi method in (33) converges if* $\alpha < min_i(2S_{ii}/\sum_j |S_{ij}|)$.(The proof is in Appendix VI-A.)

Our distributed dynamic state estimation algorithm is presented in Algorithm 1.

Algorithm 1: Distributed Algorithm for Dynamic State Estimation and Attack Detection Statistic Calculation

1: Initialization: Using (23), (25), and (27), $\{\boldsymbol{P}_k^-\}_{k=0}^K$, $\{\boldsymbol{P}_k^+\}_{k=0}^K$ and $\{\boldsymbol{S}_k\}_{k=0}^K$ are calculated offline and communicated to each control area.
2: **for** k **do** = 1 to K
3: Each control area I calculates $[\hat{\boldsymbol{x}}_k^-]_I = \boldsymbol{A}_I[\hat{\boldsymbol{x}}_{k-1}^+]_I + \boldsymbol{B}_I[\boldsymbol{u}_{k-1}]_I$.
4: **for** $t = 1{:}T_{inner}$ **do** $\boldsymbol{a}_I^{t+1} = \boldsymbol{a}_I^t + \alpha[\boldsymbol{D}_k]_I^{-1}([\boldsymbol{\gamma}_k]_I - [\boldsymbol{S}_k\boldsymbol{a}^t]_I)$.
5: Using a neighbor-limited approximation to $\boldsymbol{P}_k^-$, $\widetilde{\boldsymbol{P}}_k^-$, each area calculates $[\hat{\boldsymbol{x}}_k^+]_I = [\hat{\boldsymbol{x}}_k^-]_I + (\widetilde{\boldsymbol{P}}_k^- \boldsymbol{H}^T + \boldsymbol{M})\boldsymbol{a}_I^t$.
6: Use $\boldsymbol{a}_I$ to update the local attack detection statistic $d_I[k]$ in (37).

After completing T_{inner} inner iterations, we obtain $\boldsymbol{a}^{T_{inner}}$ which is multiplied by $(\boldsymbol{P}_k^- \boldsymbol{H}^T + \boldsymbol{M})$ in order to calculate (31). Since $\boldsymbol{P}_k^-$ is in general a full matrix, in order to avoid communication between all generators, we propose the following masking approximation. Let

$$\mathbb{N}_l \stackrel{\text{def}}{=} \begin{cases} 1, & i,j \text{ are } l-\text{hop neighbors} \\ 0, & \text{otherwise.} \end{cases} \quad (34)$$

$$\widetilde{\boldsymbol{P}}_k^- \stackrel{\text{def}}{=} \mathbb{N}_l \odot \boldsymbol{P}_k^-. \quad (35)$$

where, $\odot$ denotes the entry-wise matrix multiplication. Thus $[\widetilde{\boldsymbol{P}}_k^-]_{ij}$ is nonzero if and only if the buses corresponding to $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are at most l-hops away (e.g., direct neighbors are l-hop neighbors) . By tuning l, there is a tradeoff between accuracy of estimation and communication requirements.

C. Attack Detection

Assumed that the attack vectors $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ from the attack model in Section II-D are unknown, a sliding window attack statistic can be defined based on the Kalman innovation (*i.e.*, measurement residual) [7]. The *global* attack statistic at time k is defined as follows:

$$d[k] = \sum_{j=k-W+1}^{k} \boldsymbol{\gamma}[j]^T \boldsymbol{S}_k^{-1} \boldsymbol{\gamma}[j], \quad (36)$$

where the sliding window is of length W and m is the number of measurements. The Kalman innovation is a zero-mean Gaussian random variable [13], and the statistic $d[k] \sim \chi^2(Wm)$) is a chi-squared random variable with Wm degrees of freedom [15]. This is due to the following proposition.

Proposition 2. *The global innovation* $\boldsymbol{\gamma}[k]$ *has covariance matrix* $\boldsymbol{S}_k$. (The proof is in Appendix VI-B.)

In Algorithm 1, the quantity $\boldsymbol{a}_I[j] = [\boldsymbol{S}_k^{-1}\boldsymbol{\gamma}[j]]_I$ is calculated locally in each control area I during the dynamic state estimation. Therefore, no additional communication is required to calculate the *local* attack statistic

$$d_I[k] = \sum_{j=k-W+1}^{k} \boldsymbol{\gamma}_I[j]^T \boldsymbol{a}_I[j]. \quad (37)$$

If one had access to the global detection statistic, a classic chi-squared detection test could be used. For real-time attack detection in large networks, it is not feasible to collect $\sum_k d_I[k]$ over all areas. Instead, we propose that each area base its attack detection on its local attack statistic information. If $\boldsymbol{S}_k^{-1}$ were block diagonal, then $d_I[k]$ would be distributed as a chi-squared random variable with Wm_I degrees of freedom, where m_I is the number of measurements in area I. However, in general $\boldsymbol{S}_k^{-1}$ is a full matrix, and $d_I[k]$ does not have an easily characterized distribution. We have the following proposition for the analytical mean and variance of $d_I[k]$ under hypothesis H_0 (No Attack) using a sliding window W=1. For simplicity of notation, the timestep index k is omitted.

Proposition 3. *Let* $\boldsymbol{Y}$ *be the inverse of* $\boldsymbol{S}$. *Using* W = 1, *the mean and variance of the local attack statistic without an attack can be quantified as follows*

$$E[d_I] = \sum_{i \in I} \sum_{j=1}^{n} Y_{il} S_{il}, \quad (38a)$$

$$Var(d_I) = E[d_I^2] - E[d_I]^2, \quad (38b)$$

where,

$$E[d_I^2] = \sum_{i,j \in I} \sum_{k,l=1}^{n} Y_{il} Y_{jk} (S_{il} S_{jk} + S_{ij} S_{lk} + S_{ki} S_{lj}). \quad (38c)$$

The proof follows from Proposition 2. Given the analytical value for the variance, a threshold τ_I is set such that if $|d_I[k]| > \tau_I Var(d_I)$ an attack is declared. For example, the threshold can be a multiple of $Var(d_I[k])$. A nice feature of our algorithm is that different false alarm probabilities can be set per area based on the areas' noise characteristics, and extra information about the location is available since we monitor the local partial sums of the global attack variable.

D. Communication Analysis

We analyze the communication requirements of our distributed algorithm in terms of the sparsity patterns of relevant matrices and the l-hop neighbor approximation in (35). Note that Steps 3 and 6 of Algorithm 1 do not require any communication since $\boldsymbol{A}$ is block diagonal and $[\boldsymbol{u}]_I$ only depends on local information.

To iteratively solve for $\boldsymbol{a}_I$ in Step 4, neighbors need to communicate their entries of the vector $\boldsymbol{a}$ according to the sparsity pattern of $\boldsymbol{S}_k$.

Proposition 4. *Using a* l*-hop neighbor mask in (35), the sparsity pattern of* $\boldsymbol{S}_k$ *has non-zero entries only at pairs of measurements corresponding to buses that are at most* l*-hops away.* (The proof is in Appendix VI-C.)

After calculating $\boldsymbol{a}_I$, there are additional communication requirements for calculating $(\widetilde{\boldsymbol{P}}_k^- \boldsymbol{H}^T + \boldsymbol{M})\,\boldsymbol{a}_I$ in Step 5 of Algorithm 1. Before discretization the control input to δ_i is zero,

and the control input to ω_i depends only on θ_i. After discretization, the matrix $\boldsymbol{B}$ in (11) is block-diagonal introducing a coupling between u_{δ_i} and θ_i. Therefore, we specify $\boldsymbol{M}$ to have non-zero entries only at:

$$\boldsymbol{M}\left(u_{\delta_i}, \hat{\theta}_j^*\right) \text{ if } j = i \text{ or } j \in \mathcal{N}_i \quad (39a)$$

$$\boldsymbol{M}\left(u_{\omega_i}, \hat{\theta}_j^*\right) \text{ if } j = i \text{ or } j \in \mathcal{N}_i \quad (39b)$$

$\mathcal{N}_i$ is the set of neighbors of bus i.

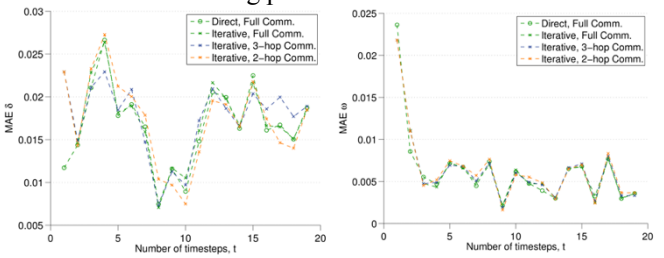
The sparsity of $\boldsymbol{P}_k^- \boldsymbol{H}^T$ is such that the columns corresponding to measurements of $\hat{\boldsymbol{\theta}}^*$ at a load bus are zero. In order to calculate the entry of vector $[\boldsymbol{P}_k^- \boldsymbol{H}^T]\boldsymbol{a}$ corresponding to measurement $\hat{\omega}_i$, the entries of $\boldsymbol{a}$ corresponding to the measurements of $\boldsymbol{\omega}$ and $\boldsymbol{\theta}^*$ at all other generators are needed. If an l-hop neighbor mask is applied, then only the entries of $\boldsymbol{a}$ corresponding to measurements at generators at most l-hops away are needed. In summary, to approximately calculate local entries of the vector $[\boldsymbol{P}_k^- \boldsymbol{H}^T + \boldsymbol{M}]\boldsymbol{a}$, areas must communicate local entries of $\boldsymbol{a}$ with at most their l-hop neighbors.

IV. Numerical Results

We have tested the proposed distributed estimation and cyber- attack detection method on the IEEE 14-bus network shown in Fig. 1.

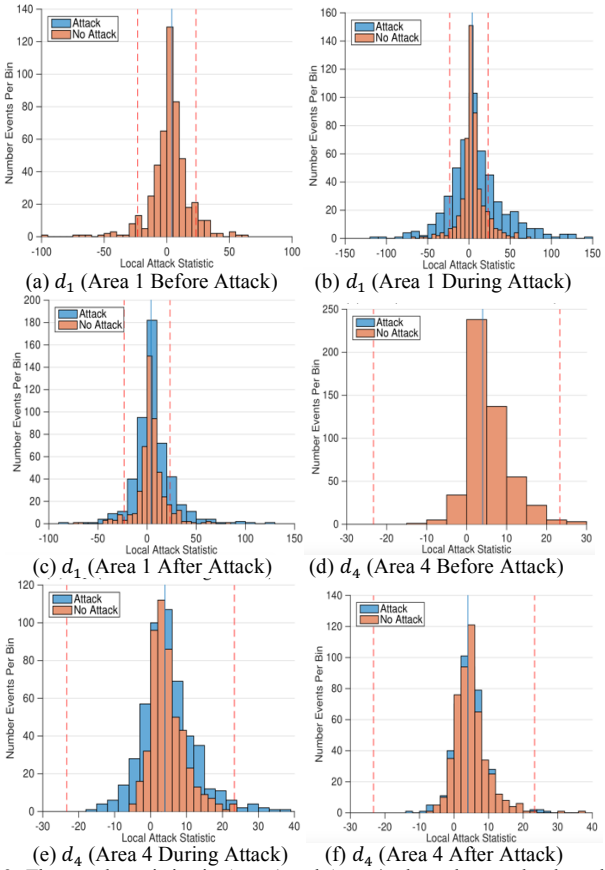
Fig. 2 shows the performances of the dynamic state estimation using different communication structures. The case where all generators communicate with each other is labeled ("Full Comm."), and the case where generators at most l-hops away communicate with each other is labeled ("l -hop Comm."). In the 14-bus network in Fig. 1, the generators are at buses 1, 7, and 14. Since buses 1 and 14 are 3-hop neighbors, they will communicate under the ("3- hop Comm.") scenario but not under the ("2-hop Comm.") scenario. Estimation performance is not significantly degraded using the l-hop approximations. In addition, Fig. 2 verifies the accuracy of the iterative (distributed) inversion with respect to the direct (centralized) inversion for the ("Full Comm.") scenario. In our simulations, the measurement noise for PMUs is $\sigma = 10^{-4}$ [16], the process noise $\bar{\boldsymbol{Q}}$ is diagonal with a variance of 10^{-3}, and the entries of $\boldsymbol{M}$ are 10^{-5} for the correlation between the process and measurement noise. The inner-loop tolerance is set to $\varepsilon = 10^{-6}$ for the matrix-splitting iterations. To evaluate the integrals in (11)-(12), we use a second-order approximation. Since the sampling rate of the PMUs is high enough to track the system dynamics, we assume the second-order approximation is accurate.

Fig. 3 demonstrates the local attack detection statistic behavior. A 3-hop communication approximation with iterative inversion is used. In each of the six figures, the histogram of the attack statistic, d_l without an attack present (red) and with an attack present (blue) is overlaid. Each event in the histogram corresponds to different values for the correlated process and measurement noise. We examine the attack statistic at three different timesteps (before, during, and after the attack) and in two different control areas. The attack is a corruption of the signal reading the power demand at bus 2 in Control Area 1 in Fig. 1. The power demand at bus 2 is taken to be ten times its actual value in the state estimation algorithm. Fig. 3 (a)-(c) show the attack statistics in Control Area 1,where the attack takes place. The analytic values for the mean and variance from Proposition 3 are calculated using the covariance matrix from our simulations. The blue vertical line is the analytic mean, and the dashed red lines are at plus and minus 1 analytic standard deviation. As expected, before the attack the histogram for d_1 matches exactly with or without an attack. At the time step where the attack occurs, the variance of the statistic d_1 is increased with respect to the case when no attack is present. Fig.3 (d)-(f) show the histogram for the attack statistic d_4 in neighboring Control Area 4. We see that during the attack, the variance of d_4 is only slightly increased. This points to another potential advantage of using the local attack statistic d_l in (37) rather than the global attack statistic d in (36). Since the variance of the local attack statistic where the attack is taking place is increased with respect to the local statistics in other areas, this suggests that using the local attack statistic helps not only in detecting the presence of an attack but also in *identifying* where the attack is taking place.



(a) MAE for δ (b) MAE for ω

Fig. 2: The mean absolute error (MAE) for δ and ω averaged over all generators of the 14-bus system as shown in Fig. 1.



(a) d_1 (Area 1 Before Attack) (b) d_1 (Area 1 During Attack)

(c) d_1 (Area 1 After Attack) (d) d_4 (Area 4 Before Attack)

(e) d_4 (Area 4 During Attack) (f) d_4 (Area 4 After Attack)

Fig. 3: The attack statistics in Area 1 and Area 4 where the attack takes place in Area 1 as corrupting the value of the power demand at bus 2.

V. Conclusion

In summary, we present a new formulation of the dynamic state estimation problem for power networks under the structure-preserving model that preserves the sparse coupling of the dynamics. We propose a new distributed Kalman filtering algorithm based on iterative linear solvers and a neighborhood-approximation to the estimated state covariance matrix. Our dynamic state estimation allows for calculation of a local attack detection statistic without any additional communication. Numerical results demonstrate the effectiveness of the estimation scheme and the utility of the local attack detection statistic.

VI. Appendix

A. Proof of Proposition 1

From Proposition 2, since $\boldsymbol{S}$ is the covariance matrix for the innovations, it is symmetric and positive semi-definite. Furthermore, by the standard assumptions for Kalman filtering, $\boldsymbol{S}$ is an invertible matrix, thus $\boldsymbol{S}$ is positive definite, i.e. $\boldsymbol{S} \succ \boldsymbol{0}$. For $\boldsymbol{S} \succ \boldsymbol{0}$, the damped Jacobi method converges if and only $\boldsymbol{0} \prec \boldsymbol{S} \prec (2/\alpha)\boldsymbol{D}$ [14]. A sufficient condition is to choose α so that $(2/\alpha)\boldsymbol{D} - \boldsymbol{S}$ is diagonally dominant which requires that

$$\left[\frac{2}{\alpha}\boldsymbol{D} - \boldsymbol{S}\right]_{ii} = (\frac{2}{\alpha} - 1)S_{ii} > \sum_{j\neq i}\left|\left[\frac{2}{\alpha}\boldsymbol{D} - \boldsymbol{S}\right]_{ij}\right| = \sum_{j\neq i}|S_{ij}| \tag{40}$$

$$\Leftrightarrow \alpha < \frac{2S_{ii}}{\sum_j |S_{ij}|} \ \forall i. \tag{41}$$

The result utilizes the fact that $S_{ii} > 0$ since it is the value of a variance.

B. Proof of Proposition 2

Here we show that the global innovations covariance matrix is given by $\boldsymbol{S}_k$.

$$E[\gamma[k]\gamma[k]^T] = E[(\boldsymbol{y}[k] - \boldsymbol{H}\widehat{\boldsymbol{x}}_{\boldsymbol{k}}^{-})(\boldsymbol{y}[k] - \boldsymbol{H}\widehat{\boldsymbol{x}}_{\boldsymbol{k}}^{-})^T] \tag{42a}$$

$$= E\{[\boldsymbol{H}(\boldsymbol{x}_k - \widehat{\boldsymbol{x}}_{\boldsymbol{k}}^{-}) + \boldsymbol{n}_k][\boldsymbol{H}(\boldsymbol{x}_k - \widehat{\boldsymbol{x}}_{\boldsymbol{k}}^{-}) + \boldsymbol{n}_k]^T\} \tag{42b}$$

$$= \boldsymbol{H}\boldsymbol{P}_k^{-}\boldsymbol{H}^T + \boldsymbol{R} + E[\boldsymbol{H}(\boldsymbol{x}_k - \widehat{\boldsymbol{x}}_{\boldsymbol{k}}^{-})\boldsymbol{n}_k^T] + E[\boldsymbol{n}_k(\boldsymbol{x}_k - \widehat{\boldsymbol{x}}_{\boldsymbol{k}}^{-})^T\boldsymbol{H}^T] \tag{42c}$$

From (22) we have that

$$\boldsymbol{x}_k - \widehat{\boldsymbol{x}}_{\boldsymbol{k}}^{-} = (\boldsymbol{A}_{k-1}\boldsymbol{x}_{k-1} + \boldsymbol{B}_{k-1}\boldsymbol{u}_{k-1} + \boldsymbol{v}_{k-1}) - (\boldsymbol{A}_{k-1}\widehat{\boldsymbol{x}}_{k-1}^{+} + \boldsymbol{B}_{k-1}\boldsymbol{u}_{k-1}) \tag{43a}$$

$$= \boldsymbol{A}_{k-1}(\boldsymbol{x}_{k-1} - \widehat{\boldsymbol{x}}_{k-1}^{+}) + \boldsymbol{v}_{k-1} \tag{43b}$$

$$E[(\boldsymbol{x}_k - \widehat{\boldsymbol{x}}_{\boldsymbol{k}}^{-})\boldsymbol{n}_k^T] = E[\boldsymbol{A}_{k-1}(\boldsymbol{x}_{k-1} - \widehat{\boldsymbol{x}}_{k-1}^{+})\boldsymbol{n}_k^T] + E[\boldsymbol{v}_{k-1}\boldsymbol{n}_k^T] = \boldsymbol{M} \tag{44}$$

where the last line follows from the fact measurement noise at time k is uncorrelated with the measurements and state at the previous time step. Plugging into (42c), we obtain the desired result

$$E[\gamma[k]\gamma[k]^T] = \boldsymbol{H}\boldsymbol{P}_k^{-}\boldsymbol{H}^T + \boldsymbol{R} + \boldsymbol{H}\boldsymbol{M} + \boldsymbol{M}^T\boldsymbol{H} = \boldsymbol{S}_k \tag{45}$$

C. Proof of Proposition 4

The matrix $\boldsymbol{S}_k$ consists of the sum of four terms. For convenience, we drop the time step index k. Consider the sparsity pattern of $\boldsymbol{A} = \boldsymbol{H}\widetilde{\boldsymbol{P}}^{-}\boldsymbol{H}^T$.

$$A_{ij} = \sum_k H_{ik}\sum_l \tilde{P}_{kl}^{-}H_{jl} \tag{46}$$

For concreteness, take row i of $\boldsymbol{H}$ to correspond to measurement θ_i^* and take row j to correspond to measurement $\widehat{\omega}_\mathrm{j}$.

$$A_{ij} = \sum_k H_{\theta_i^*,k}\sum_l \tilde{P}_{kl}H_{\omega_\mathrm{j},l} \tag{47}$$

$$= \sum_k H_{\theta_i^*,k}\tilde{P}_{\mathrm{k},\omega_\mathrm{j}}^{-} = H_{\theta_i^*,\delta_\mathrm{i}}\tilde{P}_{\delta_\mathrm{i},\omega_\mathrm{j}} \tag{48}$$

is nonzero if and only if $\tilde{P}_{\delta_\mathrm{i},\omega_\mathrm{j}}$ is nonzero. From the definition of l-hop mask in (35), this is equivalent to having bus i and bus j be at most l-hop neighbors. A similar argument follows for the other measurement types. In conclusion, the calculation of $\boldsymbol{A}$ requires l-hop neighbor communication. Without the masking approximation, the estimate of the state covariance matrix $\boldsymbol{P}^{-}$ is full, and each generator would need to communicate with every other generator.

The measurement noise covariance matrix is

$$\boldsymbol{R} = \begin{bmatrix}\boldsymbol{R}_\omega & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{L}_{ll}\boldsymbol{R}_\theta\boldsymbol{L}_{ll}^T\end{bmatrix}, \tag{47}$$

where $\boldsymbol{R}_\omega$ and $\boldsymbol{R}_\theta$ are assumed to be diagonal. The sparsity pattern of $\boldsymbol{L}_{ll}$ is the same as the adjacency matrix of the underlying network. The matrix $[\boldsymbol{L}_{ll}]_{ij}$ is nonzero if and only if bus i and bus j are neighbors, and $[\boldsymbol{L}_{ll}\boldsymbol{R}_\theta\boldsymbol{L}_{ll}^T]_{ij}$ is nonzero if and only bus i and bus j are at most 2-hop neighbors. Therefore, $\boldsymbol{R}$ requires at most 2-hop neighbor communication. The last component to analyze is the matrix $\boldsymbol{HM}$. The entry $[\boldsymbol{HM}]_{ij}$ is nonzero if and only if the states that measurement i depends upon have overlap with the control inputs correlated to measurement j. Measurement i can either be $\widehat{\omega}_\mathrm{i}$ or θ_i^* which depends on ω_i and δ_i, respectively. From (39), a measurement of ω_j is not correlated to the control inputs, so $[\boldsymbol{HM}]_{ij}$ is zero for columns j corresponding to measurements of ω. Measurements of θ_i^* are correlated with the control input at bus i and neighboring buses. Therefore, $\boldsymbol{HM}$ only requires neighbor to neighbor communication. The same argument holds for $\boldsymbol{M}^T\boldsymbol{H}^T$. In summary, at most l-hop neighbor communication is needed due to calculation of $\boldsymbol{H}\boldsymbol{P}^{-}\boldsymbol{H}^T$.

References

[1] F. Pasqualetti, F. Drfler, and F. Bullo, "Cyber-Physical Attacks in Power Networks: Models, Fundamental Limitations and Monitor Design," in 50th IEEE Conf. on Decision and Control and European Control Conference (CDC-ECC), Dec 2011, pp. 2195–2201.

[2] F. Dorfler, F. Pasqualetti, and F. Bullo, "Distributed Detection of Cyber-Physical Attacks in Power Networks: A Waveform Relaxation Approach," in 49th Annual Conference on Communication, Control, and Computing (Allerton), Sept 2011, pp. 1486–1491.

[3] S. Kar and J. Moura, "Gossip and Distributed Kalman Filtering: Weak Consensus Under Weak Detectability," IEEE Trans. on Signal Processing, vol. 59, no. 4, pp. 1766–1784, April 2011.

[4] F. Cattivelli and A. Sayed, "Diffusion Strategies for Distributed Kalman Filtering and Smoothing," IEEE Trans. on Automatic Control, vol. 55, no. 9, pp. 2069–2084, Sept 2010.

[5] U. Khan and J. Moura, "Distributing the Kalman Filter for Large-Scale Systems," IEEE Trans. on Signal Processing, vol. 56, no. 10, pp. 4919–4935, Oct 2008.

[6] P. W. Sauer and M. A. Pai, Power System Dynamics and Stability. Prentice Hall, 1998.

[7] T. Chin, W. Karl, and A. Willsky, "Sequential Filtering for Multi Frame Visual Reconstruction," Signal Processing on Multidimensional Signal Processing, vol. 28, pp. 311–333, Aug 1992.

[8] P. J. Bickel and E. Levina, "Regularized Estimation of Large Covariance Matrices," Ann. Statist., vol. 36, no. 1, pp. 199–227, 2 2008.

[9] A. Bergen and D. Hill, "A Structure Preserving Model for Power System Stability Analysis," IEEE Trans. on Power Apparatus and Systems, no. 1, pp. 25–35, Jan 1981.

[10] M. Ilic, L. Xie, U. Khan, and J. Moura, "Modeling Future Cyber- Physical Energy Systems," in 2008 IEEE Power and Energy Society General Meeting, July 2008, pp. 1–9.

[11] F. Pasqualetti, A. Bicchi, and F. Bullo, "A Graph-Theoretical Characterization of Power Network Vulnerabilities," in 2011 American Control Conference (ACC), June 2011, pp. 3918–3923.

[12] F. L. Lewis, L. Xie, and D. Popa, Optimal and Robust Estimation. CRC Press, 2007.

[13] D. Simon, Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches. Wiley-Interscience, 2006.

[14] D. M. Young, Iterative Solution of Large Linear Systems . Academic Press New York, 1971.

[15] A. S. Willsky, "Paper: A Survey of Design Methods for Failure Detection in Dynamic Systems," Automatica, vol. 12, no. 6, pp. 601– 611, Nov. 1976.

[16] L. Zhao and A. Abur, "Multiarea State Estimation Using Synchronized Phasor Measurements," IEEE Trans. on Power Systems, vol. 20, no. 2, pp. 611–617, 2005.