Phasor Measurement Units Optimal Placement and Performance Limits for Fault Localization

Mahdi Jamei, Raksha Ramakrishna, Teklemariam Tesfay, Reinhard Gentz, Ciaran Roberts, Anna Scaglione, Sean Peisert

***Abstract*—In this paper, the performance limits of faults localization are investigated using synchrophasor data. The focus is on a non-trivial operating regime where the number of Phasor Measurement Unit (PMU) sensors available is insufficient to have full observability of the grid state. Proposed analysis uses the Kullback Leibler (KL) divergence between different fault location hypotheses, which are associated with the observation model. This analysis shows that the most likely locations are concentrated in clusters of buses more tightly connected to the actual fault site akin to graph communities. Consequently, a PMU placement strategy is derived that achieves a near-optimal resolution for localizing faults for a given number of sensors. The problem is also analyzed from the perspective of sampling a graph signal, and how the placement of the PMUs i.e. the spatial sampling pattern and the topological characteristic of the grid affect the ability to successfully localize faults is studied. To highlight the superior performance of presented fault localization and placement algorithms, the proposed strategy is applied to a modified IEEE 34, IEEE-123 bus test cases and to data from a real distribution grid. Additionally, the detection of cyber-physical attacks is also examined where PMU data and relevant Supervisory Control and Data Acquisition (SCADA) network traffic information are compared to determine if a network breach has affected the integrity of the system information and/or operations [2].**

***Index Terms*—Cluster Detection, Cyber-Physical Security, Fault Location, Identification, Intrusion Detection, Optimal PMU Placement, Phasor Measurement Units.**

I. Introduction

One of the valuable uses of synchrophasor data from Phasor Measurement Units (PMU) is for fault localization. However, when the number of PMUs available are fewer than those required for full observability, pin-pointing the location can be problematic, as locating faults amounts to solving an under-determined system of linear equations that has infinitely many solutions.

There is rich literature on detecting and localizing events on the grid using relatively few number of measurements, but in general the existing methods fall short of providing performance guarantees and insights on the structure of the errors incurred while dealing with unobserved lines or buses in the grid. This paper aims to fill this gap through a statistical model that unveil fundamental limits in fault localization in the incomplete observability regime.

This research was supported in part by the Director, Cybersecurity, Energy Security, and Emergency Response, Cybersecurity for Energy Delivery Systems program, of the U.S. Department of Energy, under contract DE-AC02-05CH11231 and DE-OE0000780. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsors of this work. Preliminary version of this work was published in the proceedings of 2018 IEEE SmartGridComm [1].

A. Background

Event detection and localization is an active area of research in power grid. For example, Ardekanian et al., [3] exploit PMU data to detect and localize a change in the admittance matrix of a power grid. Zhou et al., [4] use PMU measurements for event detection in the distribution grid with partial information. Our previous work [5] describes a hierarchical architecture as the host for event detection rules in distribution system using PMU data when scarce measurements are available. Farajollahi et al., [6] use distribution level PMUs for general event detection and localization by generating an equivalent circuit so that the event can be represented by the voltage and current phasors.

Event detection by itself is necessary but not sufficient, since it is important for grid operators to be able to locate a faulty section in order to apply corrective measures for isolation and service restoration.

For example, Zhu et al., [7] describe an automated fault localization and diagnosis for distribution grids. Using measurements from the substation, the method first finds a set of plausible locations for the fault. During the diagnosis, the set of possibilities are ranked for the operator. Min and Santoso [8] investigate the effect of the DC component in the phasor data and how it can affect momentary fault localization process. Kashyap et al., [9] implement a fault location and isolation algorithm in a distributed fashion. Lee [10] uses PMU voltage data to search for a fault in a radial network in a timely manner. Dzafic et al., [11] take a graph marking approach to spot the location of a fault. In [12], a Gaussian Markov random field (GMRF) is defined for phasor angles along with a fault detection algorithm and subsequently a decentralized approach is used for fault localization. Authors in [13] use voltage measurements from few PMUs on the transmission network that are placed according to a "one-bus spaced deployment strategy" that ensures protection to all the lines in the grid. There are also a number of non-parametric methods for fault localization that use spatio-temporal (location and time) patterns of the measurements. For example, Jian et al., [14] extract the time-frequency features of frequency and voltage from a dictionary by *matching pursuit* [15], which is then followed by a clustering method for fault detection. Borghetti et al., [16] perform wavelet analysis on the voltage waveform generated during fault-induced transients to obtain the location of a fault in the distribution network. While the exploitation of temporal patterns helps in the localization, they do not provide an understanding on how the performance is affected by the grid parameters and the sensor deployment.

Although there exist several methods in the literature for

optimal placement of sensors [17], their goals are to ensure observability in the placement. Tackling the localization problem becomes challenging when one considers a regime where the sensors are insufficient to have full observability, which is acknowledged in [18] for transmission networks. However, their fault detection scheme and a placement scheme for PMUs aims for the identification of the exact fault location and ensures enough PMUs in the system to do the same. In the same vein, work in [19] deals with PMU placement for fault location observability. Only in [20], algorithms for placement under incomplete observability is even considered. Nonetheless, the goal of the aforementioned paper is not fault localization.

With inadequate number of PMUs, which is usually the case in distribution feeders, the optimal placement of sensors in such a scenario would not guarantee a good reconstruction of voltage or current signal but rather be aimed at detecting the fault with the maximum possible resolution on the graph topology. By resolution in localizing an event on the network, we refer to the ability to localize the correct ℓ-hop neighborhood, with ℓ as small as possible. In this work we use pre-fault and post-fault sample measurements along with the knowledge of admittance/impedance matrix for fault localization. The regime of interest is that where the number of sensors available is less than the observability limit, as in [21]. We refer to this operating condition as an *under-sampled grid* regime.

B. Contributions

The paper contributions are threefold: 1) it develops a maximum-likelihood method for fault localization; 2) it provides a direct connection between the errors made in fault localization with the *community structure* in the underlying system admittance matrix that acts as the Laplacian of the graph defined by the buses and lines in the grid, and 3) it derives an optimum PMU placement algorithm so that highest possible resolution in fault localization is achieved with fewer PMUs.

The paper also interprets *cluster-level fault localization* using concepts from *graph signal processing* [22], [23] and ties it to the theory of under-sampling [24] graph signals that have a sparse excitation [25]. Specifically, it is shown that the sampling pattern that provides the highest localization resolution is similar to finding the optimum sampling pattern to recover low-pass graph signals. To the best of our knowledge, both the connection between the graph clustering and fault localization resolution as well as the derivation of the corresponding optimum sampling pattern (i.e., PMU placement) are novel.

In addition to the analysis and placement method, the comparison of proposed method with state of the art algorithm in [21] shows that the proposed statistical approach is superior in localization performance. Apart from validating the proposed approaches on two IEEE test cases and real-life distribution grid, the numerical section also showcases the application of our method in the forensic analysis of cyber-physical attacks to a distribution Fault Location, Isolation, and Service Restoration (FLISR) system [26].

Preliminary results about the algorithm and error analysis were presented in the conference paper [1]. Additional contributions include:

- Employing a statistical model to find the fault location using a multiple hypotheses testing approach as opposed to the least-square method in the previous work. The statistical approach helps to consider the error in approximating the fault current and show the relationship between the clusters in the power grid graph used for low-resolution fault localization with the locations of the sensors.
- Performing extensive analysis of the sampling problem,
- Formulating an optimal PMU sensor placement strategy that was mentioned as future work in the conference version of [1],
- Presenting results by running fault localization and optimal placement algorithms on data from a real distribution grid,
- Integrating the proposed fault localization method in the cyber-physical intrusion detection architecture developed by the authors in [27], and show how such a methodology can be used to detect intrusions on fault detectors.

Paper organization: Section II describes the fault localization problem along with the formulation of the corresponding hypothesis testing problem to distinguish between faults at different locations. Section III elucidates upon the analysis of fault localization through KL divergence. In Section IV, the problem of optimal placement for highest possible resolution for fault localization is presented. Section V contains the numerical results of fault localization and placement for synthetic and real distribution grids, comparison with state of the art and application of the proposed methods for intrusion detection. Finally, conclusions are presented in Section VI.

Notation: In the paper, $\mathbb{I}_N$ is the $N \times N$ identity matrix. $\mathbf{A}^T, \mathbf{A}^H$ refer to transpose and conjugate transpose of $\mathbf{A}$. $||\mathbf{a}||$ is the 2-norm of vector $\mathbf{a}$. $||\mathbf{A}||_F$ is the Frobenius norm of matrix $\mathbf{A}$, $|\mathbf{A}|, tr(\mathbf{A})$ are determinant and trace of $\mathbf{A}$ respectively. Vector $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is distributed multivariate normal with expected value $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.

II. The Fault Localization Problem

Fault localization is a two-step process that first detects a fault and then seek to determine the location of said fault. This paper focuses on the latter, when a fault has already been detected for instance using the method presented in [5], that also provides an estimate of the beginning and end of the transient event associated with the fault. To keep the formulation general for balanced and unbalanced grids with transposed/untransposed lines, the three phase formulation of the Ohm's law (instead of the sequence domain) is used so that single and two phase lines can be considered. For a network of size N, let V_i and I_i denote the nodal voltage and injection current vectors at node i where depending on the number of phases connected to node i, V_i and I_i can be vectors of size $1, 2$ or 3. Thus, concatenated vectors $\mathbf{V}, \mathbf{I}$ of size M are defined as follows:

$$\mathbf{V} = [V_1^T, V_2^T, ..., V_N^T]^T, \quad \mathbf{I} = [I_1^T, I_2^T, ..., I_N^T]^T, \quad \mathbf{V}, \mathbf{I} \in \mathbb{C}^{M \times 1}$$

From Ohm's law:

$$\mathbf{I} = \mathbf{Y}\mathbf{V} \tag{1}$$

where $\mathbf{Y}$ is the admittance matrix. The sources in the grid are modeled with their Norton equivalent and their internal admittances are included in the $\mathbf{Y}$ matrix as well. Even though the formulation is general, all the numerical analyses are done using distribution grid test cases, and the algorithm has not been tested against transmission cases in this work.

Denote the pre-fault voltage and current as $\mathbf{V}_0$ and $\mathbf{I}_0$. If it is assumed that a fault happens at bus j, the nodal injection current after the fault can be decomposed as $\mathbf{I}_F + \mathbf{I}_E$, where $\mathbf{I}_E$ is a sparse vector containing the injected fault current at bus j in its non-zero entries. The localization problem amounts to finding the most likely set of indexes for these non-zero entries, given the observed measurements.

From (1), $\mathbf{I}_0 = \mathbf{Y}\mathbf{V}_0$ and when the measurements have stabilized during fault, $\mathbf{I}_F + \mathbf{I}_E = \mathbf{Y}\mathbf{V}_F$ where $\mathbf{V}_F$ is the post-fault voltage[1]. By subtracting the pre-fault current from the post-fault current provides the equation:

$$(\mathbf{I}_F - \mathbf{I}_0) + \mathbf{I}_E = \boldsymbol{\delta}\mathbf{I} + \mathbf{I}_E = \mathbf{Y}(\mathbf{V}_F - \mathbf{V}_0) = \mathbf{Y}\boldsymbol{\delta}\mathbf{V} \tag{2}$$

A. The Statistical Measurements Model

The current and voltages in (2) can be parsed as available and unavailable PMUs measurements. Let K denote the total number of phases, for which the nodal voltage and injection current are measured with PMUs. The matrix $\mathbf{\Pi}_a \in \{0,1\}^{K\times M}$ is a selection matrix that picks the available measurements in the voltage/current vector and $\mathbf{\Pi}_u \in \{0,1\}^{(M-K)\times M}$ selects the unavailable measurements. Pre-multiplying both sides of (2) by $\mathbf{\Pi} = (\mathbf{\Pi}_a^T \mid \mathbf{\Pi}_u^T)^T$, and replacing $\mathbf{Y}$ with $\mathbf{Y}\mathbf{\Pi}^{-1}\mathbf{\Pi}$ (noting that $\mathbf{\Pi}^{-1} = \mathbf{\Pi}^T$):

$$\begin{aligned} \begin{pmatrix}\mathbf{\Pi}_a\\ \mathbf{\Pi}_u\end{pmatrix}(\boldsymbol{\delta}\mathbf{I} + \mathbf{I}_E) &= \begin{pmatrix}\mathbf{\Pi}_a\\ \mathbf{\Pi}_u\end{pmatrix}\mathbf{Y}\begin{pmatrix}\mathbf{\Pi}_a^T & \mathbf{\Pi}_u^T\end{pmatrix}\begin{pmatrix}\mathbf{\Pi}_a\\ \mathbf{\Pi}_u\end{pmatrix}\boldsymbol{\delta}\mathbf{V} \\ \begin{pmatrix}\boldsymbol{\delta}\mathbf{I}_a\\ \boldsymbol{\delta}\mathbf{I}_u\end{pmatrix} + \begin{pmatrix}\mathbf{I}_E^a\\ \mathbf{I}_E^u\end{pmatrix} &= \begin{pmatrix}\mathbf{Y}_{aa} & \mathbf{Y}_{au}\\ \mathbf{Y}_{au}^T & \mathbf{Y}_{uu}\end{pmatrix}\begin{pmatrix}\boldsymbol{\delta}\mathbf{V}_a\\ \boldsymbol{\delta}\mathbf{V}_u\end{pmatrix}. \end{aligned} \tag{3}$$

From (3), the following set of equations is obtained:

$$\boldsymbol{\delta}\mathbf{I}_a + \mathbf{I}_E^a = \mathbf{Y}_{aa}\boldsymbol{\delta}\mathbf{V}_a + \mathbf{Y}_{au}\boldsymbol{\delta}\mathbf{V}_u \tag{4}$$

$$\boldsymbol{\delta}\mathbf{I}_u + \mathbf{I}_E^u = \mathbf{Y}_{au}^T\boldsymbol{\delta}\mathbf{V}_a + \mathbf{Y}_{uu}\boldsymbol{\delta}\mathbf{V}_u \tag{5}$$

Substituting $\boldsymbol{\delta}\mathbf{V}_u$ in (4) using (5) and reordering some terms, observation $\mathbf{z}$ is defined as:

$$\begin{aligned} \mathbf{z} &\triangleq \overbrace{\left(\mathbb{I} \mid -(\mathbf{Y}_{aa} - \mathbf{Y}_{au}\mathbf{Y}_{uu}^{-1}\mathbf{Y}_{au}^T)\right)}^{\mathbf{H}}\overbrace{\begin{pmatrix}\boldsymbol{\delta}\mathbf{I}_a\\ \boldsymbol{\delta}\mathbf{V}_a\end{pmatrix}}^{\mathbf{s}} \\ &= (-\mathbb{I} \mid \overbrace{\mathbf{Y}_{au}\mathbf{Y}_{uu}^{-1}}^{\mathbf{C}})\overbrace{\begin{pmatrix}\mathbf{I}_E^a\\ \mathbf{I}_E^u\end{pmatrix}}^{\mathbf{x}} + \overbrace{\mathbf{Y}_{au}\mathbf{Y}_{uu}^{-1}}^{\mathbf{C}}\overbrace{\boldsymbol{\delta}\mathbf{I}_u}^{\boldsymbol{\epsilon}} \\ \mathbf{z} &= [-\mathbb{I} \mid \mathbf{C}]\,\mathbf{x} + \mathbf{C}\boldsymbol{\epsilon} \end{aligned} \tag{6}$$

The term $\mathbf{C}\boldsymbol{\epsilon}$ can be viewed as a noise term, while the term $[-\mathbb{I} \mid \mathbf{C}]\,\mathbf{x}$, in which $\mathbf{x}$ is a sparse vector (i.e. the vector $\mathbf{I}_E$ reordered) whose non-zero entries point to the sites of the fault.

Remark. *Generally speaking, the noise term $\mathbf{C}\boldsymbol{\epsilon}$ is relatively small, if the constant impedance loads and capacitors/reactors are also included in the $\mathbf{Y}$ matrix, instead of being modeled in the current injection vector. For constant power loads, their equivalent admittance at the nominal voltage can be included in the bus admittance matrix, and the deviation of the actual consumed power from the nominal can be included in the nodal injection vector, so that a similar model can be adopted when the assumption of constant power loads is more appropriate. This modeling implies that the vectors $\mathbf{I}_F$ and $\mathbf{I}_0$ are small and, accordingly, their difference $\boldsymbol{\delta}\mathbf{I}$ is small as well.*

Assuming that the noise term $\boldsymbol{\epsilon}$ is drawn from complex normal distribution [2], i.e. $\boldsymbol{\epsilon} \sim \mathcal{CN}(\mathbf{0}, \sigma_\epsilon^2\mathbb{I})$, whitening the noise term is appropriate. Let the economy-size singular value decomposition [28] of $\mathbf{C} \in \mathbb{C}^{K\times M-K}$ be:

$$\mathbf{C} = \mathbf{U}\mathbf{\Sigma}\mathbf{W}^H, \quad \mathbf{\Sigma} \in \mathbb{C}^{\tilde{K}\times\tilde{K}} \tag{7}$$

where $\mathbf{U}$ is a $(K \times \tilde{K})$, $\tilde{K} \leq K$ containing the columns spanning the column space of $\mathbf{C}$, $\mathbf{\Sigma}$ is a diagonal matrix with non-zero singular values of $\mathbf{C}$ on its diagonal entries and $\mathbf{W}$ is of size $(M-K)\times\tilde{K}$ containing the columns spanning the row space of $\mathbf{C}$. Pre-multiplying $\mathbf{z}$ by $\mathbf{\Sigma}^{-1}\mathbf{U}^H$:

$$\mathbf{d} = \mathbf{\Sigma}^{-1}\mathbf{U}^H\mathbf{z} = \mathbf{F}\mathbf{x} + \boldsymbol{\varepsilon} \tag{8}$$

where $\mathbf{F} = \left[-\mathbf{\Sigma}^{-1}\mathbf{U}^H | \mathbf{W}^H\right]$ and the noise $\boldsymbol{\varepsilon} = \mathbf{W}^H\boldsymbol{\epsilon}$ and, since $\mathbf{W}$ is semi-unitary, $\boldsymbol{\varepsilon} \sim \mathcal{CN}(\mathbf{0}, \sigma_\epsilon^2\mathbb{I})$. Model in (8) is a well-known linear model with a sparse input and additive Gaussian noise. Fault localization is then equivalent to recovering the support of $\mathbf{x}$.

The fault location hypotheses testing problem is formulated based on (8). Let $\mathcal{H}_\ell$ denote the hypothesis that a fault has occurred at location $\ell \in \mathcal{F}_t$, where $\mathcal{F}_t$ is a set of candidate locations for a detected fault t. To understand the structure of $\mathcal{F}_t$ better, consider the three-phase diagram in Fig. 1 of a sample 5-bus system including 3 three-phase nodes and two single phase nodes: suppose that detected fault t indicates a

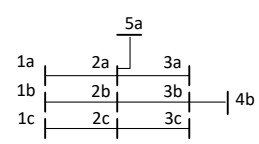

Fig. 1. Sample Three-Phase Line Diagram.

three phase fault. The candidate fault locations are:

$$\mathcal{F}_t = \left\{\{1a, 1b, 1c\}, \{2a, 2b, 2c\}, \{3a, 3b, 3c\}\right\}$$

and buses 4 and 5 are excluded since they are single phase.

Let the noise variance σ_ϵ^2 be assumed to be known. It is also assume that: 1) conditioned on the hypothesis $\mathcal{H}_\ell$, the sparse vector $\mathbf{x}$ has its non-zero entries normally distributed. This can be captured by $\mathbf{x}|\mathcal{H}_\ell \sim \mathcal{CN}(\boldsymbol{\mu}_\ell, \mathbf{\Phi}_\ell)$, where $\boldsymbol{\mu}_\ell$ is a sparse vector containing the expected value for the fault current under

[1]The assumption is that the post-fault data is recorded before any corrective protection measure is taken so that the admittance matrix stays the same.

[2]The pseudo-covariance matrix is assumed to be zero for all complex normally distributed vectors so it is omitted from the definition for simplicity.

hypothesis $\mathcal{H}_\ell$ at its non-zero entries and $\mathbf{\Phi}_l = \mathbb{E}[\mathbf{x}\mathbf{x}^H|\mathcal{H}_\ell]$ is a diagonal covariance matrix, which is only non-zero at entries corresponding to the candidate location ℓ. Thus:

$$\mathbf{F}\mathbf{x}|\mathcal{H}_\ell \ \sim \ \mathcal{CN}(\mathbf{F}\boldsymbol{\mu}_\ell, \mathbf{F}\mathbf{\Phi}_\ell\mathbf{F}^H)$$

Assuming that $\mathbf{F}\mathbf{x}|\mathcal{H}_\ell$ and $\boldsymbol{\varepsilon}|\mathcal{H}_\ell$ are independent, we have:

$$\mathbf{d}|\mathcal{H}_\ell \ \sim \ \mathcal{CN}(\overbrace{\mathbf{F}\boldsymbol{\mu}_\ell}^{\mathbf{m}_\ell}, \overbrace{\sigma_\epsilon^2\mathbb{I} + \mathbf{F}\mathbf{\Phi}_\ell\mathbf{F}^H}^{\mathbf{\Psi}_\ell}) \tag{9}$$

Denoting by $f(\mathbf{d}|\mathcal{H}_\ell)$ the probability density function of $\mathbf{d}|\mathcal{H}_\ell$ and by $\lambda_\ell(\mathbf{d}) = \ln f(\mathbf{d}|\mathcal{H}_\ell)$ the log-likelihood function, the maximum likelihood (ML) detector of the fault location is:

$$\begin{aligned}\ell^* &= \underset{\ell\in\mathcal{F}_t}{\text{argmax}}\ \lambda_\ell(\mathbf{d}) \\ &= \underset{\ell\in\mathcal{F}_t}{\text{argmin}}\ (\mathbf{d}-\mathbf{m}_\ell)^H\mathbf{\Psi}_\ell^{-1}(\mathbf{d}-\mathbf{m}_\ell) + \ln(\pi^K|\mathbf{\Psi}_\ell|).\end{aligned} \tag{10}$$

In this case, the values of these non-zero entries in the mean $\mathbf{m}_\ell$ and covariance matrix $\mathbf{\Psi}_\ell$ are not known but can be approximated. Suppose that there is a PMU installed at each source node in the grid (e.g., substation, generators). Let the vectors $I_{s,0}$ and $I_{s,F}$ denote the sum of the current injected by each source for each faulty phase into the grid during pre and post-fault condition, respectively. Note that the formulation is applicable to active distribution networks where distributed generators are placed in the grid as long there is a PMU connected to each source as mentioned above. Fig. 2 shows one-line diagram of a sample test case after a fault, which is supplied through sources 1 and 2. There

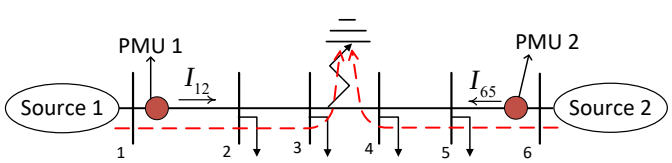

Fig. 2. Composition of Vectors $I_{s,0}$ and $I_{s,F}$.

is a PMU installed next to each one measuring the three-phase line pre-fault currents $I_{12,0} = (i^a_{12,0}, i^b_{12,0}, i^c_{12,0})^T$ and $I_{65,0} = (i^a_{65,0}, i^b_{65,0}, i^c_{65,0})^T$ and post-fault currents $I_{12,F}$ and $I_{65,F}$. Assume that the fault detectors indicate that a fault is on phase b and c. The vectors, $I_{s,0}$ and $I_{s,F}$ are formed as follows for this fault:

$$\begin{aligned}I_{s,0} &= [i^b_{12,0}, i^c_{12,0}]^T + [i^b_{65,0}, i^c_{65,0}]^T \\ I_{s,F} &= [i^b_{12,F}, i^c_{12,F}]^T + [i^b_{65,F}, i^c_{65,F}]^T\end{aligned}$$

Forming the vectors $I_{s,0}$ and $I_{s,F}$ for other types of fault follows a similar approach. Note that if P denotes the number of faulty phases, the vectors $I_{s,0}$ and $I_{s,F}$ are of size $P \times 1$. These vectors can be used to estimate the values of the mean and the covariance of the non-zero entries of $\mathbf{I}_E$. First, define matrix $\mathbf{A}_l \in \{0,1\}^{M\times P}$ that is zero everywhere except for a block equal to the identity matrix $\mathbb{I}_P$ corresponding to the fault location indexed by l. For a given fault at node ℓ^*, the value of the non-zero entries in $\boldsymbol{\mu}_\ell$ equal approximately to the sum of the currents flowing from the sources into the grid minus the current flowing into the grid before the fault occurred on the corresponding faulty phases. Denoting by $\boldsymbol{\delta} I_s = (I_{s,F} - I_{s,0})$, no matter where the fault is, the following approximation can be used:

$$\boldsymbol{\mu}_\ell = \mathbb{E}[\mathbf{x}|\mathcal{H}_l] = \mathbf{A}_l\boldsymbol{\delta} I_s \tag{11}$$

Also, the variance can be set to be a certain percentage of the value $|[\boldsymbol{\delta} I_s]_j|^2$, i.e. $[\mathbf{\Gamma}_s]_{jj} = \rho|[\boldsymbol{\delta} I_s]_j|^2$, where ρ accounts for the percentage approximation error made in assuming that the fault current is completely accounted for by the source currents change. Hence:

$$\mathbf{\Phi}_\ell = \mathbf{A}_l\mathbf{\Gamma}_s\mathbf{A}_l^T \ \rightarrow \ \mathbf{\Psi}_\ell = \sigma_\epsilon^2\mathbb{I} + \mathbf{F}\mathbf{A}_l\mathbf{\Gamma}_s\mathbf{A}_l^T\mathbf{F}^H \tag{12}$$

Even though Fig. 2 demonstrates a line to ground fault with no resistance, the formulation does not restrict the model to the type of fault. In fact, all that matters is the current withdrawn from the faulty phases that later on form the vector $\mathbf{x}$. The same thing applies to symmetric and asymmetric faults. As it was discussed, if only one or two phases are at fault, only the elements corresponding to those phases would be non-zero in the vector $\boldsymbol{\delta} I_s$ and the the methodology to locate the fault would not change.

Before going through the discussion of the fault identifiability, it should be noted that we have introduced faults at nodes, to have finite and countable number of hypotheses. In reality, a fault occurs on a line. If the method was able to locate the fault at full resolution, then the two nodes with the highest $\lambda_l(\mathbf{d})$ would determine the faulty line. However, as discussed in the next section, the fault localization method here can locate a fault with low-resolution up to a certain neighborhood of the actual location, and therefore putting a fault on a line or on its closest node would not change the end result.

III. Fault Location Identifiability

If for a given fault at location ℓ^*, the value of the metric $\lambda_\ell(\mathbf{d})$ in (10) is close to $\lambda_k(\mathbf{d})$ ($k \neq \ell^*, \ell = \ell^*$), the location of the fault can be mis-identified due to different sources of error such as approximation errors, noisy measurements, etc.

Definition. *For each possible ℓ^* a cluster is a set of nodes for which $\lambda_\ell(\mathbf{d}) \approx \lambda_{\ell^*}(\mathbf{d})$ under $\mathbf{d}|\mathcal{H}_{\ell^*}$.*

It is desirable to investigate if the properties of the grid and the connection of the nodes affect the closeness of the log-likelihood value in (10), using the Kullback-Leibler (KL) divergence [29] of the two probability density functions $f(\mathbf{d}|\mathcal{H}_\ell)$ and $f(\mathbf{d}|\mathcal{H}_k)$. The KL divergence is:

$$\begin{aligned}D_{KL}(f(\mathbf{d}|\mathcal{H}_\ell)\|f(\mathbf{d}|\mathcal{H}_k)) &= \ln\frac{|\mathbf{\Psi}_k|}{|\mathbf{\Psi}_\ell|} + tr\big(\mathbf{\Psi}_k^{-1}\mathbf{\Psi}_\ell\big) - K \\ &+ \big(\mathbf{F}(\mathbf{A}_k - \mathbf{A}_\ell)\boldsymbol{\delta}\mathrm{I}_s\big)^H\mathbf{\Psi}_k^{-1}\big(\mathbf{F}(\mathbf{A}_k - \mathbf{A}_\ell)\boldsymbol{\delta}\mathrm{I}_s\big)\end{aligned} \tag{13}$$

In this application the weighted mean of the observations is very large compared to the noise σ_ϵ^2. This leads to the following proposition:

Proposition 1. *Let $\tilde{\mathbf{F}}_k = \mathbf{F}\mathbf{A}_k\mathbf{\Gamma}_s^{1/2}$, $\boldsymbol{\xi} = \mathbf{\Gamma}_s^{-1/2}\boldsymbol{\delta} I_s$ and assume: $\forall k \ \ \frac{\|\tilde{\mathbf{F}}_k\|_F^2}{\sigma_\epsilon^2} \gg 1$. Then, the following holds:*

$$D_{KL}(f(\mathbf{d}|\mathcal{H}_\ell)\|f(\mathbf{d}|\mathcal{H}_k)) \approx \frac{\boldsymbol{\xi}^H\tilde{\mathbf{F}}_\ell^H\tilde{\mathbf{\Pi}}_k^\perp\tilde{\mathbf{F}}_\ell\boldsymbol{\xi}}{\sigma_\epsilon^2} \tag{14}$$

where $\tilde{\mathbf{\Pi}}_k^{\perp} = (\mathbb{I} - \tilde{\mathbf{F}}_k(\tilde{\mathbf{F}}_k^H \tilde{\mathbf{F}}_k)^{-1}\tilde{\mathbf{F}}_k^H)$ *is the projector onto the subspace orthogonal to the columns of* $\tilde{\mathbf{F}}_k$.

Proof. First, the first and the second terms in the KL divergence expression are just a function of the covariance matrices. Therefore, the following approximation holds:

$$D_{KL}(f(\mathbf{d}|\mathcal{H}_\ell)\|f(\mathbf{d}|\mathcal{H}_k)) \approx \|\mathbf{F}(\mathbf{A}_k - \mathbf{A}_\ell)\boldsymbol{\delta}\mathrm{I}_s\|^2_{\mathbf{\Psi}_k^{-1}} \quad (15)$$

Second, note that $\mathbf{F}(\mathbf{A}_k - \mathbf{A}_\ell)\boldsymbol{\delta}\mathrm{I}_s = (\tilde{\mathbf{F}}_k - \tilde{\mathbf{F}}_\ell)\boldsymbol{\xi}$ and by applying the matrix inversion lemma while using the assumption $\frac{\|\tilde{\mathbf{F}}_k\|_F^2}{\sigma_\epsilon^2} \gg 1$:

$$\mathbf{\Psi}_k^{-1} = \frac{1}{\sigma_\epsilon^2}(\mathbb{I} - \tilde{\mathbf{F}}_k(\sigma_\epsilon^2\mathbb{I} + \tilde{\mathbf{F}}_k^H\tilde{\mathbf{F}}_k)^{-1}\tilde{\mathbf{F}}_k^H) \approx \frac{1}{\sigma_\epsilon^2}\tilde{\mathbf{\Pi}}_k^{\perp} \quad (16)$$

Noting that by definition $\tilde{\mathbf{\Pi}}_k^{\perp}\tilde{\mathbf{F}}_k = \mathbf{0}$, the statement follows. □

Note that $\tilde{\mathbf{F}}_k$ include a weighted subset of columns of the matrix $\mathbf{F}$ defined in (8). Let $\mathbf{f}_k$ denote a column of the matrix $\mathbf{F}$. Their correlation is the absolute value of the cosine of the angle between them:

$$r_{\ell k} = \frac{|\mathbf{f}_\ell^H \mathbf{f}_k|}{||\mathbf{f}_\ell||\ ||\mathbf{f}_k||} \quad (17)$$

From what has been shown, as $r_{\ell k}$ increases for two locations ℓ and k, the divergence of their corresponding hypotheses reduces and therefore it is more likely that they are mistaken with each other. Locations with high correlation coefficient belong to the same cluster.

A. Sparse recovery interpretation

One can consider (8) as a sparse recovery problem. From the theory of recovery of sparse signals [30], it is well known that, when

$$\|\mathbf{x}\|_0 \leq \frac{1}{2}\left(1 + \mu(\mathbf{F})^{-1}\right),\ \mu(\mathbf{F}) = \max_{k,\ell,k\neq\ell} r_{\ell k} \quad (18)$$

where $\mu(\mathbf{F})$ is the mutual coherence of matrix $\mathbf{F}$, then the ℓ_1 minimization of the support can recover $\mathbf{x}$. In general, the lower is the mutual coherence of a matrix, the better the recovery performance as it is similar to unitary matrices. To compute the mutual coherence, consider

$$\mathbf{F}^H\mathbf{F} = \begin{bmatrix} \mathbf{U}\mathbf{\Sigma}^{-2}\mathbf{U}^H & -\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{W}^H \\ -\mathbf{W}\mathbf{\Sigma}^{-1}\mathbf{U}^H & \mathbb{I} \end{bmatrix}. \quad (19)$$

The magnitude of each entry of the matrix $\mathbf{F}^H\mathbf{F}$, with suitable normalization, corresponds to $r_{\ell k}$ except for the diagonal entries. Naturally, the magnitude depends on the spectrum $\mathbf{\Sigma}$ of the matrix $\mathbf{C} = \mathbf{Y}_{au}\mathbf{Y}_{uu}^{-1}$ which must be as *flat* as possible so as to lower the mutual coherence of the matrix $\mathbf{F}$. This suggests to minimize the so called *Shatten* infinity norm of the matrix $\mathbf{Y}_{au}\mathbf{Y}_{uu}^{-1}$, which is the infinity norm $\|\boldsymbol{\sigma}\|_\infty$ of the vector $\boldsymbol{\sigma}$ containing the non-zero singular values of $\mathbf{Y}_{au}\mathbf{Y}_{uu}^{-1}$. The design strategy, detailed in the next section, is aimed precisely at providing a flat spectrum for $\mathbf{C}$.

IV. Optimal Phasor Measurement Unit Placement

Following from the discussion in the previous section, for better resolution of faults, it is desirable for a placement of sensors such that it can provide a flat spectrum for $\mathbf{C}$ or in other words minimize the *Shatten* infinity norm of the matrix $\mathbf{Y}_{au}\mathbf{Y}_{uu}^{-1}$. Let $\mathfrak{P}$ denote the set of all permutation matrices. Then, the objective of optimal placement of PMUs can be realized by solving the following problem:

$$\begin{aligned} \mathbf{\Pi}^{\text{opt}} = &\ \arg\min_{\mathbf{\Pi}\in\mathfrak{P}} \ \|\boldsymbol{\sigma}(\mathbf{Y}_{au}\mathbf{Y}_{uu}^{-1})\|_\infty \\ \text{s.t.} \quad &\ \mathbf{\Pi}\mathbf{Y}\mathbf{\Pi}^T = \begin{pmatrix} \mathbf{Y}_{aa} & \mathbf{Y}_{au} \\ \mathbf{Y}_{au}^T & \mathbf{Y}_{uu} \end{pmatrix} \end{aligned} \quad (20)$$

From (20), the placement problem is combinatorial and exponentially complex so it does not scale well. We use a *greedy search* as an alternative to reduce the time complexity to *polylog*. The pseudo-code of the used greedy search is shown in Algorithm. 1.

Algorithm 1: Greedy Search Pseudo-Code for PMU Placement.

```
Initialization
    P := Number of available PMUs;
    S := Number of source nodes;
    𝒫 := {Source Nodes} // Nodes with PMU;
    ℳ := Set of candidate placement buses;
begin
    for p = 1..P − S do
        Cost ← ∞;
        for each m ∈ ℳ do
            𝒫 := 𝒫 ∪ {m};
            given 𝒫, calculate ‖σ(Y_au Y_uu^{-1})‖_∞;
            if ‖σ(Y_au Y_uu^{-1})‖_∞ < Cost then
                m_opt ← m;
                Cost ← ‖σ(Y_au Y_uu^{-1})‖_∞;
            𝒫 := 𝒫 \ {m};
        𝒫 := 𝒫 ∪ {m_opt};
        ℳ := ℳ \ {m_opt};
```

Complexity of the fault localization algorithm: The computational complexity of the proposed fault localization method is composed of two parts: optimal placement of PMUs and likelihood computation for fault localization. Time complexity of the greedy search that is employed for optimal placement is of the order of $|\mathcal{M}|(P - S)$. It depends linearly on the number of candidate buses as well as number of PMUs to be placed. The likelihood computation for fault localization is inexpensive for small network sizes and is of the order of number of possible fault locations, $|\mathcal{F}_t|$. Also, this computation can be done in parallel to save time for large networks.

A. Connection to graph signal sampling and graph clustering

In addition to interpretation using sparse recovery, one can also understand the loss of resolution and consequent best possible placement using sampling theory from graph signal

processing [24]. Placing PMU sensors can be thought of as the problem of sampling a graph signal in a fixed spatial pattern. This is because voltage or current signals are graph signals that occur on the electrical network. Not having sufficient PMUs means not sampling from enough nodes in the graph. This under-sampling leads to poor recovery or reconstruction. However, if the graph signal is sparse in the graph Fourier domain, then measurements of the order of the number of approximately non-zero components in the Fourier domain are enough for perfect reconstruction. For rigorous theory, please see [31]. The same theory can be applied to signals from the electrical grid like the measurement $\mathbf{d}$.

In [25] it was established that voltage phasor measurements are an output of a low-pass graph filter defined using the graph-shift operator of system admittance matrix $\mathbf{Y}$ which means that voltage signals are *sparse* or band-limited in the graph Fourier domain where graph Fourier transform is defined as $\mathbf{U}_Y^T\mathbf{a}$ where $\mathbf{U}_Y$ are the eigenvectors of $\mathbf{Y}$ and $\mathbf{a}$ is a graph signal. It means that measurements from a certain number of nodes say $\tilde{M}$ are needed for perfect reconstruction [24]. Since sparsity in graph Fourier domain is only approximate, optimal placement of $\tilde{M}$ measurement units on the network aims to minimize the reconstruction error. This criterion leads to the optimization problem of maximizing the singular value of $\mathcal{D}_{\tilde{M}}\mathbf{U}_Y$ i.e. $\sigma_{min}(\boldsymbol{\mathcal{D}}_{\tilde{M}}\mathbf{U}_Y)$ where $\mathcal{D}_{\tilde{M}}$ is a diagonal matrix with values from $\{0,1\}$ indicating 1 if the node is chosen for placement and it is the same size as $\mathbf{Y}$. Intuitively, this amounts to choosing the rows of $\mathbf{U}_Y$ with the smallest possible coherence or, in simpler terms, as close as possible to being orthogonal.

Similar intuition continues while looking at $\mathbf{C} = \mathbf{Y}_{au}\mathbf{Y}_{uu}^{-1}$. With $\mathbf{U}_Y$, it is choosing a representative row per 'cluster' in a graph. Similarly, in $\mathbf{C}$, choosing the rows of $\mathbf{Y}_{au}\mathbf{Y}_{uu}^{-1}$ to be uncorrelated is attained by having non overlapping support among of the rows in $\mathbf{Y}_{au}\mathbf{Y}_{uu}^{-1}$ and having $\mathbf{Y}_{uu}$ as close as possible to a block diagonal matrix, where the diagonal blocks as matched to the non zero portions of rows $\mathbf{Y}_{au}$.

Therefore, it is possible to predict how to obtain good placements looking at the structure of the graph, and particularly of its natural clustering in sub-graphs with higher connectivity within themselves. In network science, these sub-graphs are often referred to as *communities*. This is because the spectral properties we are seeking can be tied to selecting nodes in such a way that the sparsity patterns of the rows of $\mathbf{Y}_{au}$ and of $\mathbf{Y}_{uu}$ separate in clusters. Since there are few observable nodes, the rank of the matrix $\mathbf{Y}_{au}$ is limited by its number of rows. The algorithm performs best if the rows of $\mathbf{Y}_{au}\mathbf{Y}_{uu}^{-1}$ are as uncorrelated from each other as possible. Given the similarity of line parameter values, low correlation is mostly attained by having non overlapping support among of the rows in $\mathbf{Y}_{au}\mathbf{Y}_{uu}^{-1}$. As the next example illustrates, this can be attained by having $\mathbf{Y}_{uu}$ as close as possible to a block diagonal matrix, where the diagonal blocks are matched to the non zero portions of rows $\mathbf{Y}_{au}$. This is possible if the neighborhoods of the buses where the sensors are located, have the smallest intersection possible. For example, for the system in Fig. 3, if the PMUs are placed at bus 2 and 5 as shown, the admittance

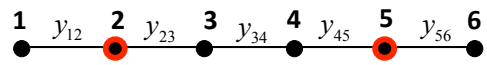

Fig. 3. One-Line Diagram of a Sample Radial Network

matrix is partitioned in the following desirable way:

$$\mathbf{Y} = \left(\begin{array}{cc|cc|cc} y_2 & 0 & -y_{12} & -y_{23} & 0 & 0 \\ 0 & y_5 & 0 & 0 & -y_{45} & -y_{56} \\ \hline -y_{12} & 0 & y_1 & 0 & 0 & 0 \\ -y_{23} & 0 & 0 & y_3 & -y_{34} & 0 \\ \hline 0 & -y_{45} & 0 & -y_{34} & y_4 & 0 \\ 0 & -y_{56} & 0 & 0 & 0 & y_6 \end{array}\right)$$

This design is clearly distancing the PMUs in the graph and dividing the network in neighborhoods, each associated with one of the sensors immediate and two hop neighbors where measurements are unavailable. The remaining ambiguity is confined to a connected set of buses that are topologically close to the PMU sensors. Hence, the optimum resolution for a certain grid is tied to the same intrinsic topological properties that are studied in graph clustering.

V. Numerical results and applications

In this section, technical discussion is firstly corroborated and then its application in the forensic analysis of cyber attacks to FLISR systems is showcased. In the numerical simulations the IEEE-34 bus and 123 test feeders (c.f. [32]) are used as synthetic grids with publicly available data, Additionally, we extend the analysis to a real distribution grid.

A. Sensor Placement and Identified Clusters

Before analyzing the results of the fault identification algorithm based on (10), the *greedy-based* placement for the PMUsis identified based on (20). The clustering is highlighted by showing the values of $r_{\ell k}$ in (17) that exceed a certain threshold τ through images where all $r_{\ell k} < \tau$ are set to zero.

1) IEEE-34 Bus Test Feeder: The one-line diagram of the test-case is shown in Fig. 4, where a 100 kW generator is added at bus 848. This grid is unbalanced and radial and has untransposed lines and one phase laterals. Also, without loss of generality, only the nodes at the same voltage level have been studied. During the placement, only three-phase buses

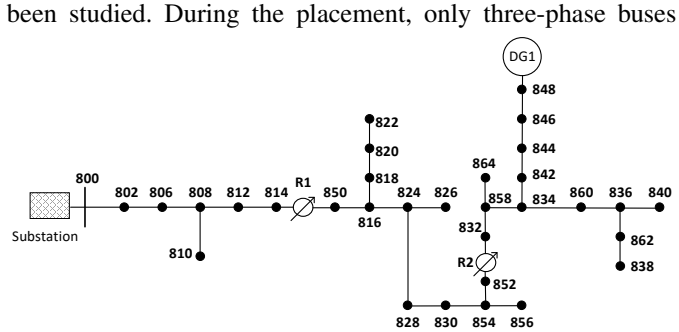

Fig. 4. Reduced IEEE-34 Test Case with Added Generator.

are kept as the candidate location for PMUs to better utilize the resources. Assuming that there are 4 PMUs available i.e.,

$P = 4$, Table. I shows the greedy-based location of the sensors and the random placement used as a comparison. Since this

TABLE I
PMU LOCATIONS FOR IEEE-34 BUS

Test Case	#PMUs	Greedy-based sites	Random sites
IEEE-34	4	800-830-848-840	800-814-816-848

is a small case, one is able to perform exhaustive search to understand how well the greedy search has performed. For this purpose, the calculated relative optimality gap shows a value of 0.75% for the greedy-based search versus 7.61% for the random case that puts the greedy-based strategy to fall below 1% optimality gap.

It is clear from the locations in Fig. 4 according to Table. I that the sensors are placed to cover the grid. Fig. 5(a) shows the correlation (17) of the columns of $\mathbf{F}$ after thresholding corresponding to phase-A for the greedy-based placement in Table I and Fig. 5(b) is the same quantity for the $\mathbf{F}$ that corresponds to the random choice. Since high correlation leads to low KL divergence, the clusters of light gray values effectively represent fault locations that are hard to discriminate. The larger the cluster, the lower the resolution. Hence, Fig. 5(b) clearly illustrates the impact on the resolution of a bad sensor placement, when compared to the clustering that emerges in Fig. 5(a). The correlation coefficients heat-map

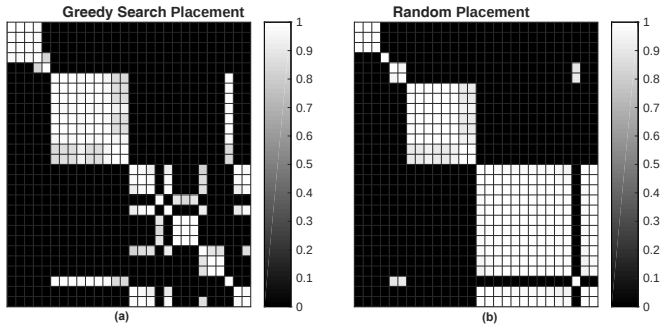

Fig. 5. Correlation of Columns of $\mathbf{F}$ for IEEE-34 Bus Case related to Phase-A for a) Greedy-based Location b) Random Location with $\tau = 0.814$

for the columns corresponding to phase-B and phase-C in the matrix $\mathbf{F}$ have a similar pattern as in Fig. 5. To corroborate the conjecture that the placement is affected by the intrinsic topology of the graph and its clusters, in Fig. 6 we highlight the set of buses that exhibit high correlation based on the greedy-based placement, by building a graph adjacency matrix such that its ℓk entry $[\mathbf{A}]_{\ell k} = u(r_{\ell k} - \tau)$[3] using the correlation coefficients of columns of $\mathbf{F}$ corresponding to phase-A and overlaying it on the IEEE-34 test case topology. As predicted, the highly correlated nodes are those that are located in a neighborhood of each other.

2) IEEE-123 Bus Test Feeder: The analysis of the larger IEEE-123 test case, with a greedy-based assignment of 10 PMUs, returns the results in Table II. Once again, the placement is done after rolling up the single and two-phase laterals

[3] $u(x)$ is the step function $u(x) = 1$ if $x > 0$ and $u(x) = 0$ else.

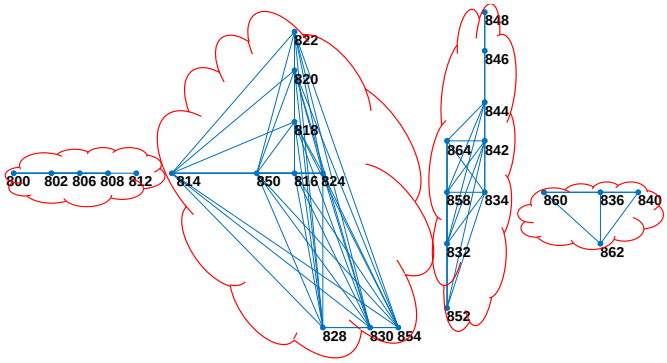

Fig. 6. Adjacency Matrix Graph for Correlation Coefficients of Columns of $\mathbf{F}$ with $\tau = 0.814$ Corresponding to Phase-A.

TABLE II
GREEDY-BASED SENSOR LOCATIONS FOR IEEE-123 BUS

Test Case	#PMUs	Greedy-based Location
IEEE-123	10	149-81-61-56-105-250-86-151-72-57

and only nodes at the same voltage level are analyzed. The one-line diagram and node numbering of this case can be found in [32]. The PMUs in this case also are spread over different areas of the grid topology to form communities of neighboring nodes of comparable sizes. The thresholded heat-map given in Fig. 7 based on the placement of Table II, highlights the balance across clusters.

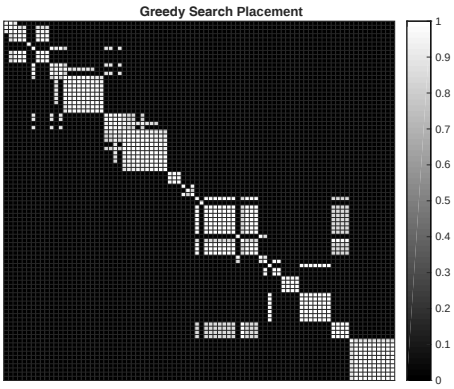

Fig. 7. Correlation of Columns of $\mathbf{F}$ for IEEE-123 Bus Case Related to Phase-A for Greedy-based Location with $\tau = 0.814$

3) Real Utility Distribution Grid: To show the performance of the sensor placement on a real grid, we use the data from a partner utility grid, which is a medium voltage (12.47 KV) network. Due to the non-disclosure agreements, the details of the grid data cannot be published. The studied grid has 1066 nodes and we choose a very small number of $P = 20$ PMUs to place. Fig. 8 shows the one-line diagram of the grid along with the location of the placed PMUs based on the greedy search that are marked with red circles. Fig. 9 demonstrates some of the formed communities of neighboring nodes for part of the studied grid[4]. As it can be seen, a very similar structure

[4] Due to large size of the grid, the full heat-map could not be shown.

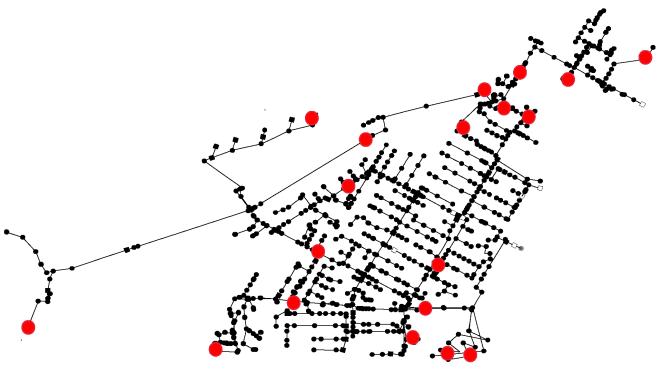

Fig. 8. Greedy Search PMU Placement with $P = 20$ PMUs.

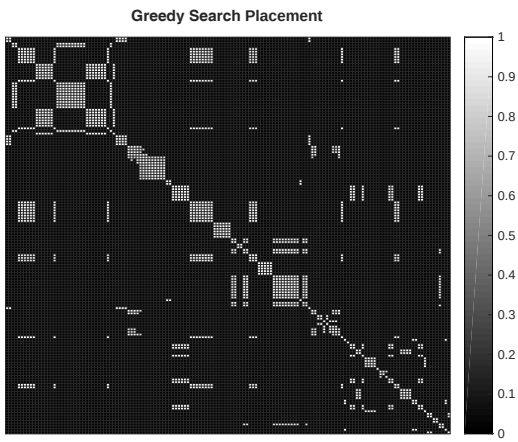

Fig. 9. Correlation of Columns of $\mathbf{F}$ for Utility Grid Related to Phase-A for Greedy-based Location with $\tau = 0.814$

of correlation matrix appears in the real-life distribution grid that confirms observations in the synthetic grids.

B. Fault Localization Algorithm

The synthetic PMU data for this experiment were generated using dynamic mode of the OpenDSS software [33] to better model the behavior of the components during fault. The voltage and current data corresponding to pre and post fault conditions are recorded. Note that the tap changers usually have a delay for 15-30 seconds in order to respond to a change so the voltage and current data should be recorded before the tap value changes, so that the admittance matrix stays the same before and after the fault.

1) IEEE-34 Bus Test Feeder: Fig. 10 shows the value of the log-likelihood function $\lambda_\ell(\mathbf{d})$ for different hypotheses when a single-phase to ground fault occurs at phase-A of bus 820. As analysis indicated, using the metric in (10) shown in Fig. 6, it becomes readily apparent that one can narrow down the location of the fault up to a certain resolution, since a cluster of nodes return very close log-likelihood value. These are also the nodes that correspond to high values of $r_{\ell k}$ with respect to the actual location of the fault. To further test and verify the fault localization method, we introduce different types of fault and show the results of the fault localization in Table. III under the greedy-based placement.

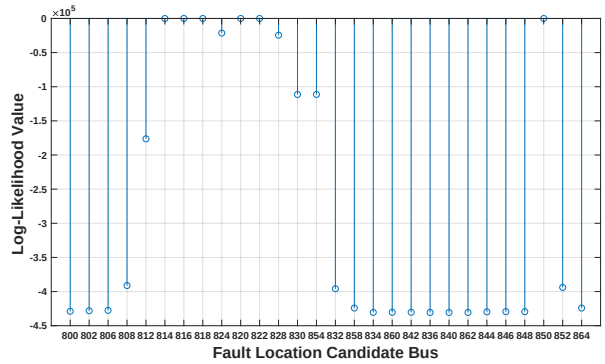

Fig. 10. Log-Likelihood Value for Fault at Bus 820-Phase-A

To demonstrate the performance of the method when a fault occurs on a line, we introduce a line-to-ground fault on the line connecting bus 850 to 816 at 30% distance from node 850. Fig. 11 shows the log-likelihood value of the candidate locations. As it was mentioned in the discussion

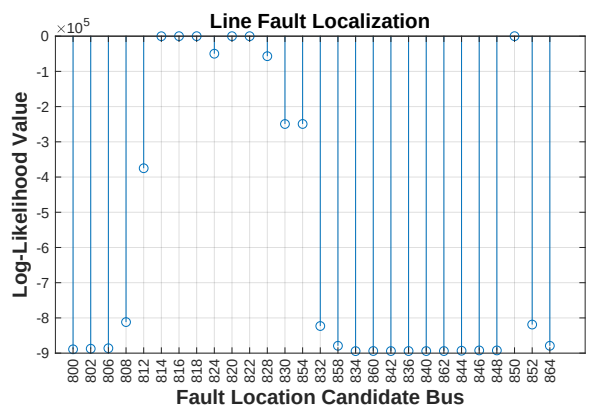

Fig. 11. Log-Likelihood Value for Phase A-to-Ground Fault on Line 850-816 with resistance=10 Ohms.

before, since our fault localization is at low-resolution level, a line fault would still pick the nodes that are within the community, where the line is located. In this sense, identifying the community, in which fault has occurred is still possible. One edge case of the algorithm is when a fault happens on a line that connects two separate communities to each other. First of all, it should be noted that separation of grid into communities has been done with a threshold (τ), otherwise there is no clear border between the communities. In order to test the method for such a case, a three-phase to ground has been introduced on a line connecting buses 812 and 814. From Fig. 12, it can be seen that the plausible locations for the fault are from the two communities that are connected via line 812-814. However, nodes that are closer to the actual location of the fault, have higher possibility based on the observed values. The fact that not all the nodes within the community have the same value of log-likelihood helps to understand that the fault should be close to the edge of the two communities of neighboring nodes.

2) IEEE-123 Bus Test Feeder: We extend the analysis of the fault localization algorithm to IEEE-123 test case for

TABLE III
IDENTIFIED FAULT LOCATIONS FOR IEEE-34 BUS CASE

Fault Type	Fault Resistance (ohms)	Exact Fault Location	Locations with Highest $\lambda_\ell(\mathbf{d})$
ABC	0	816	814,816,850
A-G	0	822	814,816,818,820,822,850
BC-G	0	852	832,852,858
AC-G	0	836	836,840,862
AB-G	0	808	800,802,806,808
AB-G	10	834	834,842
AC	5	850	850,814,816
ABC-G	5	806	800,802,806

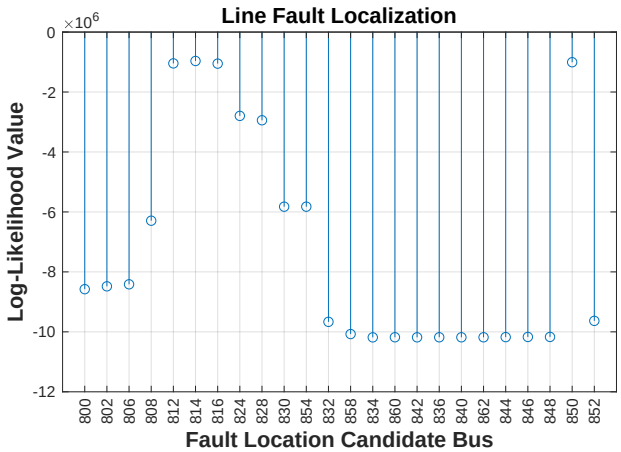

Fig. 12. Log-Likelihood Value for Three-Phase-to-Ground Fault on Line 812-814 with resistance=0 Ohms.

different types of fault after the placement of the sensors on the locations in Table II. Fig. 13(a) shows the log-likelihood value of different hypotheses for a three-phase fault at bus 160. Part of the correlation coefficients heat-map that illustrates the correlation of the columns that have high correlation with bus 160 is snipped out of Fig. 7 and is illustrated in Fig. 13(b). As expected, the neighboring nodes that also have high correlation with bus 160 are those which have the closest log-likelihood value in our metric and can be mis-detected as the fault location.

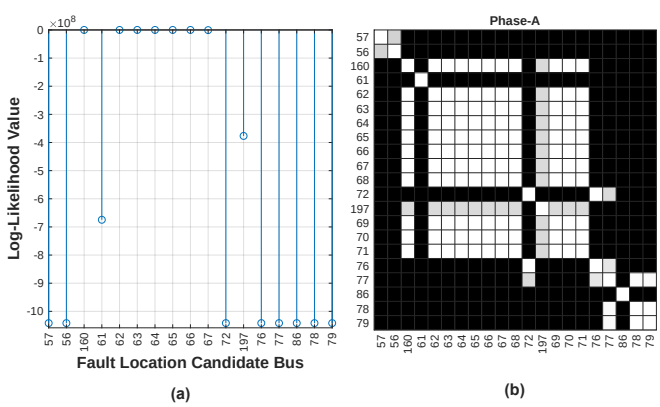

Fig. 13. a) Log-Likelihood Value for Three-Phase Fault at Bus 160 and b) Correlation of Corresponding Columns of $\mathbf{F}$-phase A.

Other types of faults and the potential locations with highest $\lambda_\ell(\mathbf{d})$ are summarized in Table IV.

Faults are introduced at different locations of the grid and the results show that the fault is identifiable up to the resolution of the clusters of neighboring nodes identified with the correlation coefficients. An interesting case is the two-phase fault at node 57 (last row in Table IV), where the fault is identifiable with a high resolution. This is consistent with our claims, since the coefficients $r_{\ell k}$ reveal that there is no other node with very high correlation with this specific node.

3) Real Utility Distribution Grid: To show the application of the fault localization method on real grids, we introduce different types of fault in the utility grid and evaluate the fault localization algorithm. As the data is not publicly available, the faults that are introduced are marked by a unique number. Table. V describes the details of the type of fault each number represents. In Fig. 14, the arrows point to the location where the fault occurred and the circles around the location are the plausible fault locations that the localization algorithm returned.

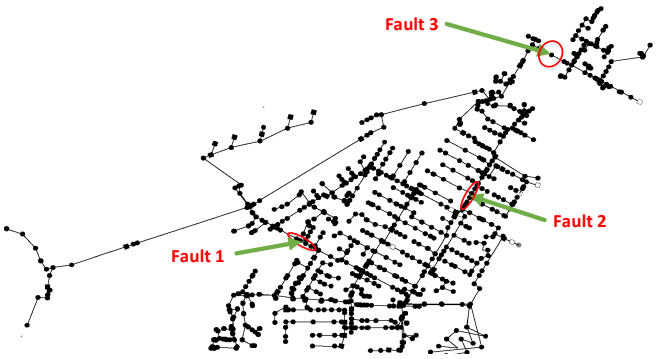

Fig. 14. Fault Localization Algorithm Results for Three Different Types of Fault in the Real Utility Grid.

4) Comparison with the State of the Art: Proposed approach in using the pre and post-fault samples and treating the fault as a current injection improves upon the work in [21] through its statistical underpinning. To show this, we introduce a two-phase fault at bus 836-A-C in IEEE-34 case and the results are shown in Fig. 15. The original method in [21] is designed as a minimization problem to find the fault location so we changed it to a maximization by adding a negative sign to make a better visual comparison with our method.

TABLE IV
IDENTIFIED FAULT LOCATIONS FOR IEEE-123 BUS CASE

Fault Type	Fault Resistance (ohms)	Exact Fault Location	Locations with Highest $\lambda_\ell(\mathbf{d})$
LLL	0	42	40,42,44,47,48
A-G	0	108	105,108,109,300,110, 111,112,113,114
BC-G	0	89	86,87,89,91,93,95
AC-G	0	50	47,48,49,50,51,151
AB-G	0	57	57
AC-G	10	42	40,42,44
AB	20	21	21

TABLE V
UTILITY GRID INTRODUCED FAULT DESCRIPTIONS

Fault Number	Fault Type	Fault Resistance (ohms)
Fault 1	LLL-G	15
Fault 2	BC-G	10
Fault 3	A-G	0

The results clearly show that the proposed method can locate the fault more accurately than the algorithm in [21].

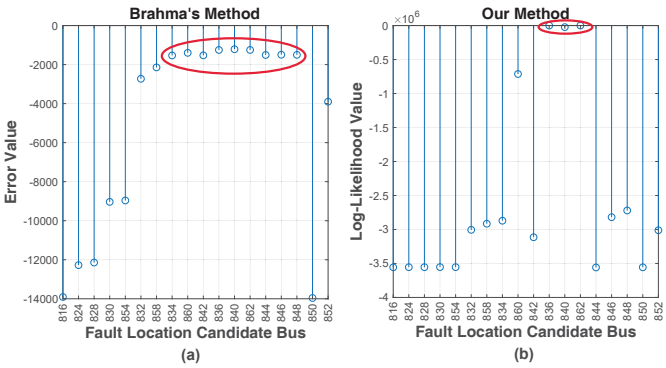

Fig. 15. Fault Localization for Fault at Bus 836-A-C Using a) Method in [21] b) Described Method.

Besides this improvement, investigating the reason behind this ambiguity is missing in [21], whereas we have investigated this matter theoretically and analytically in our work. In addition, we have presented a sensor placement strategy here to place the sensors at nodes that would return the highest fault localization resolution.

C. Application to Cyber-Physical Intrusion Detection

By leveraging the insights from this work, network intrusion detection system testbed hosted at Lawrence Berkeley National Lab (LBNL) [34] is enhanced and we incorporate additional rules to monitor for cyber attacks that interfere with the normal operations of FLISR systems [26]. Testbed (see Fig. 16) is defined as a hierarchical architecture to fast-detect the presence of attacks against networked devices controlling physical systems by correlating analytic results from SCADA traffic (which includes traces from SCADA communications)

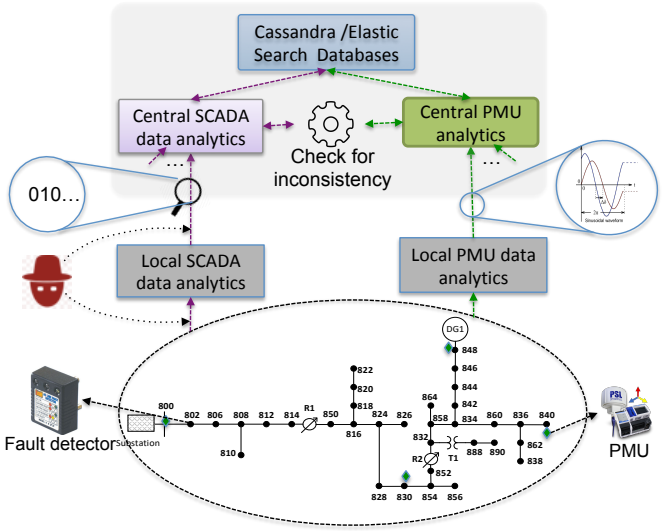

Fig. 16. LBNL Stream-Processing Architecture for Real-time Cyber-physical Security (SPARCS) used in our experiments [34]–[36]. Its components are: 1) The Bro IDS, the industry-standard, network monitoring framework; 2) a publish-subscribe messaging system called RabbitMQ used to transfer PMU and SCADA data and analysis results; 3) the Cassandra database, used for permanent archiving of all data; 4) The Elasticsearch database, which collects analysis results in real time; 5) BeagleBone Black (BBB) devices that receive synchrophasor data at a rate of 120 samples/sec, and analyze 1 sec. worth of data at a time to rapidly detect a cyber-physical event [27].

with PMUs deployed sparsely over a distribution system [34].

1) The FLISR System Operations and Vulnerabilities: FLISR systems detect the location of a permanent fault [5] in a feeder and automatically restore service to customers in the healthy section of the feeder. To do so they employ directional fault detectors, installed at every line in a distribution feeder, that communicate the occurrence of a fault and its direction to a distribution management system (DMS). The DMS analyzes these data and, once the faulted section is identified, issues commands to a predetermined set of switches to first isolate the faulted section and then restore service to the non-faulted areas [26], so as to minimize the service disruption. Consider the one-line diagram of a sample radial network shown in Fig. 17 to describe how FLISR works. The network is connected to a substation at bus 1 and a distributed generation is connected

[5]Temporary faults that clear themselves do not trigger the FLISR operation.

to bus 6. Fault detectors and normally-closed switches are connected at the end of every line. When a fault happens at line $3-4$, all the fault detectors detect the presence of this fault such that fault detectors at buses $1, 2$ and 3 indicate that the fault is in their right side, whereas fault detectors at buses $4, 5$ and 6 indicate the fault is to their left. Fault localization analysis at the DMS using this information identifies the faulted section to be line $3-4$. The DMS, therefore, sends control signals to automatically open switches to the right of bus 3 and to the left of bus 4 splitting the network into two parts and restoring service to customers on both sections.

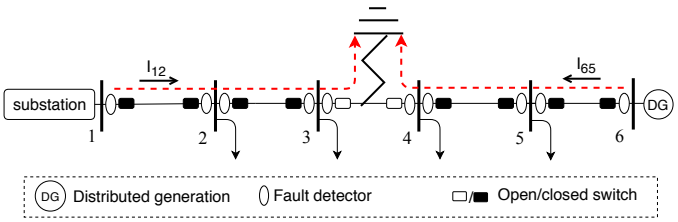

Fig. 17. Diagram of a Sample Radial Network with a Fault in Line $3-4$.

Fault detector data is transmitted to the DMS through a wide area network using industrial protocols such as DNP3 [37] or Modbus [38] which are not designed with security concerns in mind. Therefore, a dedicated attacker can tamper with fault detector data either remotely, through a compromised network device (e.g., a router), or by physically connecting herself to an exposed section of the communication network. For instance, the tampering can lead the DMS analysis to pick the wrong location for an actual fault or hide the presence of a fault.

Although such attacks can be partially prevented by enabling cryptographically authenticated communication on top of the stated communication protocols, such a measure does not prevent the whole spectrum of possible cyber attacks. Therefore, a more robust approach, that does not merely rely on cryptographic solutions, is desired that is capable of detecting the presence of an attacker. The analysis we carried out in this paper shows that, even with a very limited number of sensors scattered in the systems, the PMU localization performed using (10) enhances an operator's confidence about the fault localization information (or the lack of thereof) extracted from fault detectors data, by giving an additional means to verify the trustworthiness of the SCADA messaging, albeit at a lower spatial resolution. In the next section we illustrate the effectiveness of the forensic analysis on FLISR attacks carried out using a experiment, with simulated streaming SCADA and PMU data, to our LBNL testbed in Fig. 16.

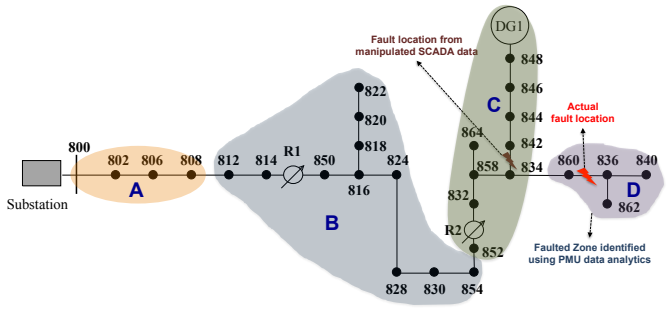

Fig. 18. Output Mismatch as a Result of an Attack on SCADA data

2) Intrusion Detection on Fault Detectors: In our simulation to demonstrate intrusion detection on fault detectors, we use the modified IEEE 34-bus system shown in Fig. 18. We introduce a two-phase AC-G fault at line $860-836$. Our PMU data analytic results correctly indicate the fault to be in Zone D. We generate simulated SCADA data for each fault detector in the network using the openDNP3 library [39] that implements the DNP3 communication protocol. In our simulation, we modify the packets from some of the fault detectors (those at buses 834 and 860) so that the SCADA data analytics indicate the fault is at line $834-842$, which is in zone C. The inconsistent results from the PMU data analytics and the SCADA data analytics raise an alarm about a possible cyber-attack on the fault detectors.

VI. Conclusions

A fault localization method is presented where a fault can be localized up to a cluster of neighboring nodes when very few sensors are available in the grid. It is proved that these clusters can be inferred from the properties of the network and where the sensors are placed. Having that identified, a PMU placement strategy to achieve fault localization with highest resolution over the grid is proposed. Fault localization method is integrated into a security detection framework, where the results of PMU-based fault localization is correlated with the SCADA-reported fault location to check the integrity of the SCADA data.

References

[1] M. Jamei, A. Scaglione, and S. Peisert, "Low-resolution fault localization using phasor measurement units with community detection," in *2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, 2018, pp. 1–6.

[2] C. McParland, S. Peisert, and A. Scaglione, "Monitoring Security of Networked Control Systems: It's the Physics," *IEEE Security & Privacy*, vol. 12, no. 6, pp. 32–39, Nov/Dec 2014.

[3] O. Ardakanian, Y. Yuan, R. Dobbe, A. von Meier, S. Low, and C. Tomlin, "Event detection and localization in distribution grids with phasor measurement units," in *2017 IEEE Power Energy Society General Meeting*, July 2017, pp. 1–5.

[4] Y. Zhou, R. Arghandeh, and C. J. Spanos, "Partial knowledge data-driven event detection for power distribution networks," *IEEE Transactions on Smart Grid*, 2017.

[5] M. Jamei, A. Scaglione, C. Roberts, E. Stewart, S. Peisert, C. McParland, and A. McEachern, "Anomaly detection using optimally-placed μPMU sensors in distribution grids," *IEEE Transactions on Power Systems*, 2017.

[6] M. Farajollahi, A. Shahsavari, E. Stewart, and H. Mohsenian-Rad, "Locating the source of events in power distribution systems using Micro-PMU data," *IEEE Transactions on Power Systems*, 2018.

[7] J. Zhu, D. L. Lubkeman, and A. A. Girgis, "Automated fault location and diagnosis on electric power distribution feeders," *IEEE Transactions on Power Delivery*, vol. 12, no. 2, pp. 801–809, 1997.

[8] K. W. Min and S. Santoso, "DC offset removal algorithm for improving location estimates of momentary faults," *IEEE Transactions on Smart Grid*, 2017.

[9] N. Kashyap, C.-W. Yang, S. Sierla, and P. G. Flikkema, "Automated fault location and isolation in distribution grids with distributed control and unreliable communication," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 4, pp. 2612–2619, 2015.

[10] J. Lee, "Automatic fault location on distribution networks using synchronized voltage phasor measurement units," in *ASME 2014 Power Conference*. American Society of Mechanical Engineers, 2014, pp. V002T14A008–V002T14A008.

[11] I. Dzafic, R. A. Jabr, S. Henselmeyer, and T. Donlagic, "Fault location in distribution networks through graph marking," *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 1345–1353, 2018.

[12] M. He and J. Zhang, "A dependency graph approach for fault detection and localization towards secure smart grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 342–351, 2011.

[13] K.-P. Lien, C.-W. Liu, C.-S. Yu, and J.-A. Jiang, "Transmission network fault location observability with minimal pmu placement," *IEEE Transactions on Power Delivery*, vol. 21, no. 3, pp. 1128–1136, 2006.

[14] H. Jiang, J. J. Zhang, W. Gao, and Z. Wu, "Fault detection, identification, and location in smart grid based on data-driven computational methods," *IEEE Transactions on Smart Grid*, vol. 5, no. 6, pp. 2947–2956, 2014.

[15] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on signal processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[16] A. Borghetti, M. Bosetti, M. Di Silvestro, C. A. Nucci, and M. Paolone, "Continuous-wavelet transform for fault location in distribution power networks: Definition of mother wavelets inferred from fault originated transients," *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 380–388, 2008.

[17] W. Yuill, A. Edwards, S. Chowdhury, and S. Chowdhury, "Optimal pmu placement: A comprehensive literature review," in *2011 IEEE Power and Energy Society General Meeting*. IEEE, 2011, pp. 1–8.

[18] Q. Jiang, X. Li, B. Wang, and H. Wang, "Pmu-based fault location using voltage measurements in large transmission networks," *IEEE transactions on power delivery*, vol. 27, no. 3, pp. 1644–1652, 2012.

[19] M. Korkali and A. Abur, "Optimal deployment of wide-area synchronized measurements for fault-location observability," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 482–489, 2012.

[20] R. F. Nuqui and A. G. Phadke, "Phasor measurement unit placement techniques for complete and incomplete observability," *IEEE Transactions on Power Delivery*, vol. 20, no. 4, pp. 2381–2388, 2005.

[21] S. M. Brahma, "Fault location in power distribution system with penetration of distributed generation," *IEEE transactions on power delivery*, vol. 26, no. 3, pp. 1545–1553, 2011.

[22] A. Sandryhaila and J. M. Moura, "Discrete Signal Processing on Graphs," *IEEE Transactions on Signal Processing*, vol. 61, no. 7, April 2013.

[23] N. Tremblay, P. Gonçalves, and P. Borgnat, "Design of graph filters and filterbanks," in *Cooperative and Graph Signal Processing*. Elsevier, 2018, pp. 299–324.

[24] P.D.Lorenzo, S. Barbarossa, and P. Banelli, *Cooperative and Graph Signal Processing*. Elsevier, June 2018, ch. 9.

[25] R. Ramakrishna and A. Scaglione, "On modeling voltage phasor measurements as graph signals," in *IEEE Data Science Workshop (DSW) 2019*, June 2019.

[26] R. Uluski, "Using distribution automation for a self-healing grid," in *Transmission and Distribution Conference and Exposition (T&D), 2012 IEEE PES*. IEEE, 2012, pp. 1–5.

[27] M. Jamei, E. Stewart, S. Peisert, A. Scaglione, C. McParland, C. Roberts, and A. McEachern, "Micro synchrophasor-based intrusion detection in automated distribution systems: Toward critical infrastructure security," *IEEE Internet Computing*, vol. 20, no. 5, pp. 18–27, 2016.

[28] D. S. Bernstein, *Matrix mathematics: Theory, facts, and formulas with application to linear systems theory*. Princeton University Press Princeton, 2005, vol. 41.

[29] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 03 1951.

[30] D. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proceedings of the National Academy of Sciences*, 2003.

[31] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Transactions on Signal Processing*, vol. 64, no. 18, pp. 4845–4860, 2016.

[32] W. H. Kersting, "Radial distribution test feeders," in *Power Engineering Society Winter Meeting, 2001. IEEE*, vol. 2. IEEE, 2001, pp. 908–912.

[33] R. C. Dugan, "Reference guide: The open distribution system simulator (opendss)," *Electric Power Research Institute, Inc*, vol. 7, 2012.

[34] R. Gentz, "Wireless sensor data transport, aggregation and security," Ph.D. dissertation, Arizona State University, 2017.

[35] S. Peisert, R. Gentz, J. Boverhof, C. McParland, S. Engle, A. Elbashandy, and D. Gunter, "LBNL Open Power Data," http://powerdata-explore.lbl.gov, 2017.

[36] "LBNL Stream-Processing Architecture for Real-time Cyber-physical Security (SPARCS)," https://github.com/lbnl-cybersecurity/sparcs.

[37] IEEE, "IEEE Std 1815-2012 (Revision of IEEE Std 1815-2010) - IEEE Standard for Electric Power Systems Communications-Distributed Network Protocol (DNP3)," 2012. [Online]. Available: http://standards.ieee.org/findstds/standard/1815-2012.html

[38] I. Modbus, "Modbus messaging on tcp," *IP Implementation Guide v1. 0a, North Grafton, Massachusetts (www. modbus. org/specs. php)*, 2004.

[39] Open DNP3 Group and others, "DNP3–distributed network protocol 3.0–google project hosting," 2012. [Online]. Available: https://www.automatak.com/opendnp3/