2017 IEEE 33rd International Conference on Data Engineering

Robust Power Line Outage Detection with Unreliable Phasor Measurements

Jose Cordova-Garcia and Xin Wang
Department of Electrical and Computer Engineering
Stony Brook University, Stony Brook, New York 11794
Email: jose.cordova@stonybrook.edu, x.wang@stonybrook.edu

***Abstract*—Phasor Measurement Units (PMUs) provide high precision data at high sampling rates to support Smart Grid applications. Power Line Outage detection mechanisms can enhance the grid reliability by assisting power operators in taking proper control actions. Despite the potential provided by PMUs, there are very limited efforts on exploiting data available to more effectively detect outages. Conventional outage detection schemes are mostly designed based on simplified power models, and the limited work on detection with data either assume all the measurement samples are available or ignore the missing entries. Their performance suffers in the complex grid conditions in the presence of missing data. In this paper, we design a detection mechanism considering unreliable data, in the form of missing data samples. Detection is performed through the grouping of nodes according to their data availability and their learned detection capabilities. To enable the robust detection of power line outages, we propose learning outage characteristics for each individual node instead of specific single line outage scenarios. Our results show that the outages detected are highly consistent with the evaluated failures under different scenarios, with high accuracy and low false positive rates. Moreover, the detection application is resilient to unreliable data, and can properly differentiate data problems from physical power line failures.**

I. Introduction

The installation of modern sensors with high reporting rates is arguably the most important update power grids can implement within the Smart Grid framework. Phasor Measurement Units (PMUs) are increasingly deployed to drastically enhance the grid monitoring capabilities for Smart Grid applications. PMUs measure the grid at the rate of 30-60 samples per second, much higher than legacy devices which provide input data for applications at the rate of seconds to minutes [1]. Thus, PMUs introduce massive high precision data, which calls for the revisions and updates of traditional grid management applications, as well as the introduction of new ones.

The newly available data can be used to update the power models needed for model-based applications. For viable applications, a model is often simplified based on assumptions, albeit this may compromise the accuracy. While the information from newly available data can help to refine power applications, it may also increase their complexity. Moreover, the reporting rate of PMU data is much faster than the rate complex model-based applications are executed. Thus PMU data may have to be processed in an offline fashion to prepare them for applications. On the other hand, data-based (model-less) applications attempt to enable grid monitoring by extracting information directly from the data, taking advantage of the high reporting rate of PMU without relying on the accuracy of a power model. They can be designed for online monitoring based on the most updated PMU data without preprocessing through complex models.

In this paper, we focus on the design of a data-based method for the effective detection of power-line outage in the presence of missing PMU measurement data. A power-line outage refers to the temporal or permanent disconnection of both ends of a line. This type of power event can drastically affect the power system. The incurred topology change, due to even a few line failures, may lead the power grid to reach an unplanned operational state that develops into a cascade failure [2], [3]. Discovering power-line outages, not captured by an outdated model, helps to make system-wide decisions and prevent the propagation of failures. Despite its importance, timely outage detection has received much less attention in PMU research [1]. We take advantage of the information extracted from the high-rate PMU data for outage detection without relying on model-based assumptions.

While the introduction of PMUs has attracted a lot of attentions from the power community to the design of different applications, either model-based or data-based, much less attention has been paid to problems introduced when grid applications increase their dependability on data networks. In a Smart Grid, the control center collects data and runs energy management applications. Data packets not received on time or lost represent missing entries in the data input to grid applications. A temporal malfunction of a PMU or high noise in the communication channel used to transmit the measurements will cause time-correlated missing entries, while space-correlated missing entries can be observed when data from different channels of the same PMU device are compromised by its malfunctioning or transmissions from PMUs at different locations are affected by intentional physical or cyber attacks to the network. Also, the impact of a power failure over a certain region of the grid can cause temporal or permanent malfunction of PMU devices. Various sources of unreliability in PMU data can affect the performance of time-sensitive applications such as outage detection, outage location identification and control, which in turn compromises the power grid reliability.

Despite the significant impact of unreliability of PMU data, the limited existing data-based studies generally ignore this problem. In this paper, we propose the design of a robust online power-line outage detection mechanism which is resilient to missing data. Power line outages may occur all over the grid and there may exist multiple outages. Our goal is to find a set of outage locations based on the available PMU data. Rather

2375-026X/17 $31.00 © 2017 IEEE
DOI 10.1109/ICDE.2017.173

than learning the characteristics of specific outage scenarios, whose number could be extremely large with different outage combination formats, we propose to learn the detection capabilities of every power node. To protect against the impact due to data missing from specific PMUs, we propose a subspace-based method to properly create detection groups to assist the detection of failures around different nodes. Our scheme will not need the reconstruction of missing samples, which will help avoid the time and computation costs as well as avoid the potential compromise of online characteristic due to the inaccuracy in the missing data recovery.

The rest of the paper is organized as follows. Section II reviews the related work on outage detection, particularly data-based approaches. Section III describes the data models used by our proposed detection mechanism. In Section IV, we present the proposed methodology for learning appropriate node-based subspaces and forming the detection groups for outage identification. Section V describes the test cases and metrics of interest, and presents results for different scenarios. Particularly, different missing data scenarios are discussed and evaluated. Finally, Section VI concludes the work.

II. Related Work

The problem of failure detection in power grids has been studied for a long period of time, however, it is only in the recent years attentions are attracted to data-driven mechanisms. In a model-based design, PMU measurements only constitute the input for the model used. Constrained by specific models assumed, these schemes can not take full advantage of the data available with the fast PMU rate to well capture the grid characteristics in failure detection. On the other hand, model-free applications attempt to process the data directly and extract the information that may not be captured in a model. Hence, detection mechanisms may have different levels of PMU data integration. In this section we review some detection mechanisms with high level of PMU data integration, and discuss their limitations when unreliable (i.e. with measurement missing) synchrophasors are encountered.

For a comprehensive review of PMU related research, including fault/event detection, the reader can refer to [1]. Existing studies of power line events can be broadly classified according to whether they address fault or outage detection. The majority of studies discussed in [1] correspond to the model-based methods to detect faulted sections of a power line. In such studies, current synchrophasors are assumed to be available at least at one end of every power line and used to estimate and/or extract parameters of power models that describe the faults.

Despite the large amount of effort on fault detection, as reported in [1], much less work addresses the issue of outage discovery. An outage is considered to happen if there exists a complete disconnection/fault of the line, which can be caused by reasons such as natural disasters, intentional attacks, infrastructure aging, and accidents. The occurrence of power line outages is one of the most important vulnerabilities of power grids. Undiscovered outages of a relatively small number of power lines have been reported as the source of large system-wide blackouts, i.e. cascading failures [2], [3]. Our proposed design addresses this type of power line events, and we will use the terms event, failure, and outage interchangeably. More specifically, we aim to detect and localize an outage in the presence of missing PMU data due to the unreliability of PMUs or communication network. The following discussions review some related data-based outage detection algorithms, while they generally overlook or ignore the PMU reliability problem.

In [4], the authors use a multinomial regression classification method to learn a classifier that can determine if the PMU samples taken can be classified as one of the outage scenarios learned during the training stage. The results presented show that the trained classifier can accurately identify most of the failures. However, the possible data missing has been ignored and the performance for this scenario cannot be guaranteed. The work in [5] uses an SVM to learn and classify single line outages. Similarly, it is not specified how PMU data with missing entries will be considered by the learned classifier. A Decision Tree (DT) event classifier with a transformed feature space is proposed in [6]. While promising results are shown, authors explicitly decide to leave out testing and training measurement instances that contain missing entries. This constrains the proposed classification technique to not being used for outage detection.

In [7] the authors propose a disturbance triggering approach based on the threshold of variability of singular values, empirically set through visual observation of certain historical disturbances. The performance of singular value decomposition is greatly undermined by the occurrence of missing data, which in turn compromises the functionality of this solution. The authors later in [8] consider recovering missing samples for correlated data erasures using the recovery error as a disturbance indicator, while it does not provide a way to discover the outage location. Similarly, the work in [9] presents a PCA-based detection mechanism to identify the buses with “dominant variance” as the line outages, where the variance threshold is set manually based on existing data which is not scalable. Furthermore, traditional PCA-based methods depend on singular value decomposition, which carries over the vulnerability to missing data entries. The work in [10] propose an event indicator based on the selection of “pilot” PMUs with a dimensionality reduction criteria, and with a small number of PMUs selected, the scheme may fail to function when data from one of such pilots are missing. These schemes are generally designed to detect events without providing their locations. In addition, these techniques assume the full availability of data, and/or unreliable data should be ignored [7]. They may fail to function or suffer from significant performance degradation upon data missing, while waiting for missing data would make them not applicable for online grid applications.

Despite the significant impact of unreliability of PMU data, the limited existing data-based studies generally ignore this problem. The problem of unreliable synchrophasor measurements has only started to be analyzed and exposed in [11], and considered for (offline) State Estimation [1].

Different from the literature works, we design an outage detection mechanism considering the occurrence of data unreliability in the form of missing data samples. Detection is performed through the grouping of nodes according to their data availability and their learned detection capabilities. Instead of leaning based on a specific single line outage scenario, our proposed mechanism learns the outage characteristics of each individual node, which helps to enable robust outage detection against unreliable PMU data.

III. PMU Data and Monitoring

In this section, we present the PMU data model used in our design of the robust power-line outage identification mechanism. We are interested in completing the detection task using data gathered by PMU sensors without relying on power models. We briefly present a general layout of a PMU network to identify the source of data, and then introduce the data characteristics and structures that will be processed in the detection mechanism. Finally, we describe how missing data occurs in PMU networks, and how it is modeled in the detection application.

A. Phasor Measurement Units and Data

The physical infrastructure of a power grid along with its communication network, used for monitoring and control, constitute a high-level view of a Smart Grid. The most commonly deployed monitoring networks consist of Phasor Measuremnet Units (PMUs). PMUs provide high data sampling rates compared to their legacy counterpart Remote Terminal Units (RTU) sensors used in the Supervisory Control and Data Acquistion (SCADA) system. Measurements gathered by PMUs are commonly referred as synchrophasors due to their GPS-synchronized sampling. The costs of PMU devices and communications have limited an extensive deployment of PMUs. We will not consider the optimal PMU deployment, but rather focus on the analysis of PMU data for more efficient failure detection in the presence of data missing.

The transmission level of the power grid can be modeled by a graph $\mathcal{P}(\mathcal{N}, \mathcal{E})$, where nodes in $\mathcal{N}$ are power generators, consumer loads, and substations, among other power grid elements. We refer to any power grid element being monitored as a power node. The set of edges $\mathcal{E}$ represents all the physical power lines existing in the grid. An outage of line $e_{i,j} \in \mathcal{E}$ represents the removal of its corresponding edge from the graph, $\mathcal{P}(\mathcal{N}, \mathcal{E}\backslash\{e_{i,j}\})$. The data gathered from the PMUs at the control center are represented by a matrix $\mathbf{X}$, whose rows represent different sensors and columns different time instants. The vector $\mathbf{X_i} \in \mathbb{R}^T$ contains the PMU measurements of power node i in the time window T, and the entry $x_{i,t}$ is its measurement at the instant t. The input of an online Smart Grid (SG) application at time t is a data sample with measurements from all observable power nodes, represented by $\mathbf{X_{:,t}} \in \mathbb{R}^N$, for a grid with $N = |\mathcal{N}|$ nodes.

As phasors help describe the state of the grid, we consider phasor monitoring and $\mathbf{X}$ corresponds to either voltage magnitude or phase measurements. We will extract information from historical data of both normal operations and power line outages to assist the detection of future outages. Normal operation data refer to the expected steady states of voltage phasor measurements when no failures occur in a time period. In case historical data are not available for a period, simulations can be carried out to obtain the estimates of the expected synchrophasors. Normally, such simulations use a day-ahead forecast of load distribution, and the forecasted

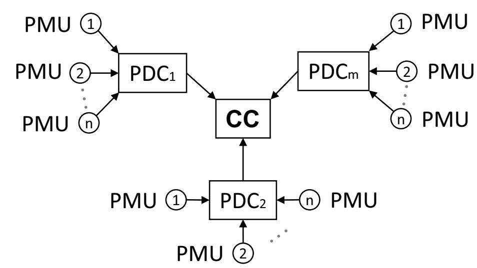


Fig. 1: PMU Data Network example

measurements are used for several other SG applications as well, e.g. generation dispatch. Similarly, an outage detection application can use these synthetic data to learn the phasor values during normal operations and data deviations due to outages. Thus, either historical or forecasted power-line outage data logs can be used to identify such deviations.

B. PMU Monitoring and Data Reliability

As illustrated in Figure 1, generally, PMU networks are highly hierarchical. The group of PMUs monitoring a region of the power grid share a common data collection point, the Phasor Data Concentrator (PDC). PDCs are commonly directly connected to the Control Center (CC) where data are stored and processed. With this monitoring infrastructure, data are readily accessible to the grid operator at the control center, and can be used as input for a variety of applications. As an example application, data can be used in the Energy Management System (EMS) for failure and outage detection. In the case of PDC malfunctioning, the whole data cluster would be compromised. With more dependability on data, Smart Grids are becoming more vulnerable to planned cyber attacks to their measurement units and communication networks. A planned simultaneous packet drop attack to several geographically correlated PMUs could cause missing entries spatially and temporally correlated. For example, a PDC constitutes an attractive target for cyber attacks.

Therefore, on the data packet level, PMU measurements can suffer from problems such as delay, loss, false data injections. Cyber attacks to the communication network can also compromise the availability of PMU data. There exist some works on estimation/imputation of missing PMU measurements in order to prepare the data for an application. For example, there is an extensive body of work on updating State Estimators (SE) where the power grid state variables, i.e. voltage phasors, can be directly obtained from PMU measurements. SE is commonly regarded as an application that does not need to be executed at the fast rate of data collection. Thus, it can tolerate the time consumption in processing backlogged measurements to reconstruct the missing data. This might not be the case for delay-sensitive applications, such as outage detection.

Timely detection of power outage is of critical importance, and the delay in control and restriction of the failure effect may lead to cascading failure. Although the correct reconstruction of missing data may help for detection, the inaccuracy and delay in the reconstruction process and/or the waiting for data with no missing samples will compromise the process of outage detection. Despite the possibly large negative effect, missing data in failure detection applications have been particularly overlooked. It is common to find mechanisms that either assume PMU data with no erasures are available at the control center, or ignore the missing samples.

Problems with the communication channel used to transmit PMU data may cause temporary unavailability of some measurements, which leads to temporally correlated missing entries along the rows of $\mathbf{X}$. On the other hand, spatially correlated missing entries along the columns of $\mathbf{X}$ may be observed when a power line outage impacts PMUs in a neighborhood or there is a planned infrastructure attack to the grid.

As a missing data pattern of particular interest, the set of data measurements from PMUs near an outage location are all not available. In this case, the control center neither receives samples from the devices at the failure location nor from its immediate neighborhood, for the duration of the line outage. Online outage detection is made extremely challenging in this case. The outage detection should be able to differentiate between data problems and physical failures, and its performance should not be compromised by data problems.

IV. Robust Outage Detection resilient to Missing data samples

In order to perform timely identification of power line outages, we will leverage the node-level information of the power system to identify the nodes affected by the outages(s). Moreover, the performance of the outage identification mechanism should not be affected by the occurrence of any pattern or number of missing data points. The extracted node-level information will provide robustness against missing data.

We introduce our scheme in three steps. We first present how to learn from the data appropriate subspace information to identify outages. Then, we describe how to enhance the subspace method in order to provide robustness against missing data. Finally, we describe the proposed mechanism that identifies line outages by properly processing data that may contain missing measurements.

A. Node-based Subspace Learning

The mechanism we present in this section is based on learning node information that can help identify outages. Given that a power line outage occurs at the line $e_{i,j}$ and time t_0, any column $t > t_0$ of the data matrix $\mathbf{X}$ (represented as $\mathbf{X}_{:,\mathbf{t}}$) contains a "signature" of the outage that may be exploited for identification. Based on the learning principle, if there are enough training samples, important signatures can be identified to determine if a power-line is at fault according to the similarity of new samples to past signatures. For an outage at line $e_{i,j}$, nodes i and j are directly affected, and intuitively they are good candidates to identify the failure as along as their PMU data (i.e. voltage phasors) are available at the time of classification. In other words, certain node measurements in $\mathbf{X}_{:,\mathbf{t}}$ contain enough information to characterize an outage signature. This motivates us to design a node-based failure identification methodology. We consider a linear approximation of the power model:

$$\mathbf{X}_{:,\mathbf{t}} = \mathbf{Y}^{+}\mathbf{P}_{:,\mathbf{t}}, \tag{1}$$

where $\mathbf{P}_{:,\mathbf{t}} \in \mathbb{R}^N$ corresponds to the amount of power supplied and consumed by different power nodes of the grid and determines the grid behavior. $\mathbf{Y}$ is the admittance matrix which contains the topology information (i.e. line status) of the grid along with electrical parameters, and the "$+$" represents its pseudo-inverse. Thus, $\mathbf{Y}$ is a weighted laplacian of the power grid topology graph. From Equ. (1), the synchrophasor data collected in $\mathbf{X}$ contain enough information that characterizes the behavior and topology of the grid, i.e. the connection status of all power lines. Denoting $\mathbf{X^0}$ as the matrix containing samples of the normal operation case, and $\mathbf{X}^{\setminus e_{i,j}}$ the one that contains samples with the outage line $e_{i,j}$, we would like to investigate if it is possible to exact the features from these matrices to identify if a power-line experiences an outage. The singular value decomposition of a collection of measurements $\mathbf{X}$ in a time duration is

$$\mathbf{X} = \mathbf{U\Sigma V^T}. \tag{2}$$

As shown in [12], the vectors of $\mathbf{U}$ corresponding to the lowest singular values in $\mathbf{\Sigma}$ can represent the subspace that describes the line status thus the grid topology. We let $\mathcal{S}^0$ to represent the subspace of normal operations. We can similarly find a subspace $\mathcal{S}^{\setminus e_{i,j}}$ for each outage matrix available $\mathbf{X}^{\setminus e_{i,j}}$.

These learned subspaces provide outage-based information, thus can be used to define node-based subspaces for outage identification. We define a node-based subspace for each node $i \in \mathcal{N}$ as:

$$\begin{aligned} \mathcal{S}_i^{\cup} &= \bigcup_{k \in \mathcal{N}_i} \mathcal{S}^{\setminus e_{i,k}}, \\ \mathcal{S}_i^{\cap} &= \bigcap_{k \in \mathcal{N}_i} \mathcal{S}^{\setminus e_{i,k}} \end{aligned} \tag{3}$$

$\mathcal{S}_i^{\cup}$, and $\mathcal{S}_i^{\cap}$ correspond to the "union" and "intersection" subspaces of power node i, each is carried out over the subspaces corresponding to the failure of each line of i.

At a time t, we exploit the relative proximity of a sample $\mathbf{X}_{:,t}$ to the subspaces $\mathcal{S}^0$, $\mathcal{S}_i^{\cup}$, and $\mathcal{S}_i^{\cap}$ to identify the outage lines, as will be described below.

B. Robust Subspace Method for Outage Identification

To determine the proximity from a sample to a subspace, we can use the projection of the sample to the subspace. This requires that the sample $\mathbf{X}_{:,t}$ be complete. To account for possible missing of measurement $x_{i,t}$ in the sample, we define a group of nodes to be the detecting nodes. To detect the proximity from $x_{i,t}$ to a subspace, only the measurements from these nodes need to be present in the sample set. Thus, a "detection group" for a node i consists of nodes that can be used to approximate the distance of a sample to a subspace, in case that measurements from node i are missing. Out of the nodes that can provide $x_{i,t}$ measurements for $\mathbf{X}_{:,t}$, a detection group should satisfy two conditions: 1) Their measurements should capture approximately the same data variation as the one experienced by i when it is under the outage but without missing data, and 2) These measurements should be highly available in different missing data scenarios, as defined by the

PMU network. For example, given the possibility of missing data from sensors at an outage location, the nodes in the detection group need to be relatively far away from the affected location so that their data have higher availability than sensor nodes nearby.

It is simple to see that we can capture the contributions of different nodes to the variability induced by failures of i by means of PCA. The detection group can be formed by finding the most orthogonal nodes according to their data projections onto the PCA space. This naive choice of detection group members ideally should provide a basis for the subspace of interest. However, the number of orthogonal nodes found is usually small, and can be different for each node-based subspace. We consider an enhanced method. We will search through the past training data, and augment the detection group with nodes that have a high probability of detecting the events occurring at the node i. With this procedure, we would like that each detection group contains a sufficient number of nodes from separated sensing regions (i.e. data clusters defined by PDCs in Figure 1), as the chance of simultaneous data missing from all regions will be very small.

The past phasor data to be used, as described in Section III, present variability in both normal operations and during a power-line outage. During the normal operation of the grid, such data behaviors can be modeled by fitting expected phasors to normal operation ellipse. Then, an outage represents a deviation from the normal operation ellipse, such a deviation can be permanent or temporal. Our goal is to properly identify such deviations.

Each node i defines a normal operation ellipse:

$$\Omega_i = \{x_{i,t} \in \mathbb{R}^2 \mid (x_{i,t} - c_i)^T A_i (x_{i,t} - c_i) \leq 1\}, \quad (4)$$

where all PMU voltage phasor data are inside the ellipse, i.e., $x_{i,t} \in \Omega_i, \forall t > t_0$, during the normal operations of node i. Measurements received during a failure at line $e_{i,j}$ may fall outside, that is $x_{i,t} \notin \Omega_i$. Intuitively, any point x_i and x_j should fall sufficiently outside their ellipses, Ω_i and Ω_j respectively. Hence, the membership of a point to its corresponding ellipse can be used as a failure detection criteria.

For a case (e.g., the failure of line $\{e_{i,j}\}$) of the set of past outage data, the recorded capability of a node k to detect failure $F = \{e_{i,j}\}$ is defined as:

$$p_k(F = \{e_{i,j}\} \mid \mathbf{X}_k^{\setminus e_{i,j}}) = \frac{\sum_{t=1}^{T} \mathbb{I}(x_{i,t} \notin \Omega_k)}{\sum_{t=1}^{T} \mathbb{I}(x_{i,t}^0 \in \Omega_k)}, \quad (5)$$

where $\mathbf{X}_k^{\setminus e_{i,j}} = [x_{k,1} \dots x_{k,T}]$ contains the set of samples from historical data when there exist failure at the line $e_{i,j}$, and $x_{k,t}^0$ corresponds to a sample of normal operation at time t.

Then, for each available outage training case F, the detection capability of any $k \in \mathcal{N}$ can be calculated using (5). We can have a node-based vector of detection capabilities, $\mathbf{p_i} \in \mathbb{R}^N$, to represent how accurate each $k \in \mathcal{N}$ can detect an outage of any line of another node i:

$$\mathbf{p_i} = [p_{i,1}, \dots, p_{i,k}, \dots p_{i,N}], \quad (6)$$

$$p_{i,k} = p(\cup_{F \in \mathcal{F}_i} \{F\}) = \sum_{l=1}^{|\mathcal{F}_i|} \left((-1)^{l-1} \sum_{\mathcal{F}_l \subset \{\mathcal{F}_i\} | l = |\mathcal{F}_l|} \prod_{F' \in \mathcal{F}_l} p_k(F' | \mathbf{X}_k^{F'}) \right), \quad (7)$$

where $\mathcal{F}_i$ is a super set that contains all cases F in the training data set that involve node i. $\mathcal{F}_l$ is a super set that contains a combination of l cases of $\mathcal{F}_i$, where each F' is an outage case involving node i. That is, every F' contains failed lines in $\mathcal{E}_i$ and $\mathcal{E}_i \subseteq \mathcal{E}$ contains all the power lines of node i.

Each $p_{i,k}$ evaluates the detection capability of node k to identify any failure case involving node i based on the available training data. Using $\mathbf{p_i}$, the control center can determine which nodes are useful to identify the outages of any power lines attached to the node i, and these nodes are candidates to form the detection group for the node i. Intuitively node i and its immediate neighbors should have the highest detection accuracy in $\mathbf{p_i}$. Hence, nodes in the same region (or cluster) will have similar detection groups. Instead of forming one detection group for each node, these detection groups may be consolidated to assist the outage detection in a grid region.

The measurements of nodes that belong to the detection group for a node i can be used for identifying outages occurring at lines connected to the node i. Some of these candidate detection nodes could also suffer from data missing. To make a detection group robust to data missing, the group should have alternative members.

In Figure 1, each PMU cluster $\mathcal{C}$ collects data of the geographical region it covers. Measurements from PMUs in the same cluster can exhibit temporal and spatial correlations. Then, in a detection group $\mathcal{D}$, the set of members from cluster $\mathcal{C}$ can be considered as one possible source of missing data, and we only need to provide an alternative detection group for each data cluster defined in the PMU network.

Hence, to build a detection group for the cluster $\mathcal{C}$, some nodes in the detection group should be members of the cluster which should provide high detection capabilities for every $k \in \mathcal{C}$. In case of data missing from nodes in a cluster, to ensure reliable detection, we consider including in the detection group the members not belonging to the cluster. Both categories of members can be found with the corresponding vectors formed in (6):

$$\begin{aligned} \mathcal{D}_{\mathcal{C}}(\mathcal{C}) &= \bigcap_{k \in \mathcal{C}} \{i \in \mathcal{C} \mid p_{k,i} \approx 1\} \\ \mathcal{D}_{\mathcal{C}}(\overline{\mathcal{C}}) &= \bigcap_{k \in \mathcal{C}} \{i \notin \mathcal{C} \mid p_{k,i} \approx 1\} \end{aligned} \quad (8)$$

where the detection group provided by cluster $\mathcal{C}$ when data are available is denoted as $\mathcal{D}_{\mathcal{C}}(\mathcal{C})$, and $\mathcal{D}_{\mathcal{C}}(\overline{\mathcal{C}})$ when there exist missing data. That is, when there is data missing from a cluster, we use nodes from other clusters for detection.

Figure 2 illustrates the case of selecting the appropriate detecting subspace for a sample at time t which contains the missing data. In the figure, the information of cluster C is missing, highlighted in grey color, and an outage has occurred near the cluster C, highlighted in red. Hence, outage information

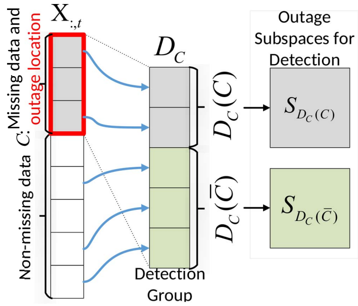


Fig. 2: Detection group formation and selection.

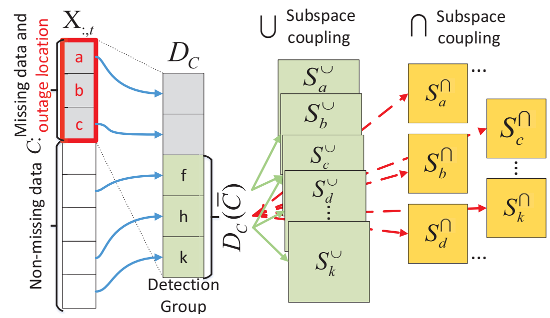


Fig. 3: Scaling proximities to detect failure at node a.

coming directly from the outage location is not available at the control center. As described before, the definition of a detection group anticipates the possible case of unreliable measurements and $\mathcal{D}_C$ can use data from nodes inside and outside the missing data cluster. As highlighted in green in the figure, the subspace selected for detection is the one associated to the available data in the detection group, i.e. $\mathcal{D}_\mathcal{C}(\bar{\mathcal{C}})$. On the other hand, while an outage detection subspace based on $\mathcal{S}_{\mathcal{D}_\mathcal{C}(\mathcal{C})}$ could have been formed in the past, such a subspace is not selected due to the unavailability of data from $\mathcal{D}_\mathcal{C}(\mathcal{C})$ at time t.

C. Detecting Power Line Outages

The outage detection is made hard in the presence of simultaneous data missing in both temporal and spatial directions. To enable the outage detection with data missing, we propose to form the detection groups with the support of node-based subspaces learnt from data. According to the data cluster a node i is the member of, every node i has three associated subspaces $\{\mathcal{S}^0, \mathcal{S}_i^\cup, \mathcal{S}_i^\cap\}$, and detection groups $\{\mathcal{D}_\mathcal{C}(\mathcal{C}), \mathcal{D}_\mathcal{C}(\bar{\mathcal{C}})\}$.

The location of an outage is identified by the proximity of the test data sample to a learned subspace. As defined in (3), the subspaces are determined based on measurements from all nodes in the grid, i.e. N dimensionality. For a subspace $\mathcal{S}$, we can divide the rows of $\mathbf{U}$ in (2) for $\mathcal{S}$ into two different matrices: $\mathcal{S}(\mathcal{D})$ containing the rows that correspond to members of the detection group, and $\mathcal{S}(\mathcal{N}\backslash\mathcal{D})$ with the rest of the nodes.

Based on [12], a regressor $\mathbf{\Phi}(\mathcal{S}) = -(\mathcal{S}(\mathcal{D})^T)^+\mathcal{S}(\mathcal{N}\backslash\mathcal{D})^T$ can assist the approximation of the distance of samples in $\mathcal{D}$ to a subspace $\mathcal{S}$. Hence, the proximity of a sample to a subspace $\mathcal{S}$ can be estimated as

$$\text{prox}_\mathcal{S}(\mathbf{X}_{:,t}) = \|\, \mathbf{X}_{\mathcal{D},t} - \mathbf{\Phi}(\mathcal{S})\mathbf{\Phi}(\mathcal{N}\backslash\mathcal{S})\mathbf{X}_{\mathcal{D},t} \,\|_2^2 \,. \quad (9)$$

When $\mathcal{S}$ contains the information of a failure, the proximity can be used to identify such failure. To determine the proximity above, the only requirement for the sample $\mathbf{X}_{:,t}$ is that there are no missing data in the measurements taken by nodes in $\mathcal{D}$. With our proposed cluster-based detection grouping, depending on if there are data missing in the measurements of nodes in cluster $\mathcal{C}$ at time t, we propose two alternative groups of members for $\mathcal{D}$ below:

$$\mathcal{D}_\mathcal{C} = \begin{cases} \mathcal{D}_\mathcal{C}(\bar{\mathcal{C}}), & \text{if any } x_{k\in\mathcal{C},t} \text{ is missing} \\ \mathcal{D}_\mathcal{C}(\mathcal{C}), & \text{otherwise.} \end{cases} \quad (10)$$

Then when determining the proximity of $\mathbf{X}_{:,t}$ to a subspace, we only need to use the portion of data corresponding to nodes in the detection group selected and $\mathbf{X}_{\mathcal{D},t}$ becomes $\mathbf{X}_{\mathcal{D}_\mathcal{C},t}$ in (9). The chance for both groups to miss data is very small, thus the proximity of a test sample $\mathbf{X}_{:,t}$ to a subspace can be calculated even with missing data. The proximity $\text{prox}_{\mathcal{S}^0}(\mathbf{X}_{:,t})$ indicates how likely it is that the sample $\mathbf{X}_{:,t}$ corresponds to normal operation of the grid, without outages. If at least one line of a node i experiences an outage, sample $\mathbf{X}_{:,t}$ should be relatively closer to $\mathcal{S}_i^\cup$ than to $\mathcal{S}^0$, i.e. $\text{prox}_{\mathcal{S}_i^\cup}(X_{:,t})$ is lower. If there exists a severe outage around node i with multiple lines of i disconnected, the proximity to the node-based intersection subspace $\text{prox}_{\mathcal{S}_i^\cap}(\mathbf{X}_{:,t})$ is lower.

In Figure 3, we illustrate the proximity-based detection scheme. An outage is located in the cluster $\mathcal{C}$ and data from all member nodes (a, b, c) are missing at the time of detection. Thus, the non-missing portion of the sample, corresponding to data from nodes outside of $\mathcal{C}$, is used to calculate the proximity to the detecting subspaces, highlighted in green. Due to the coupling among power lines, the proximities to subspaces of nodes not directly affected by a failure may have values close to those of nodes affected (as shown with solid green arrows), which may mislead the subspace method to make a wrong detection. This is a consequence of the similarity of detecting subspaces of nodes geographically (or electrically) close to each other as determined by the topology of the grid. On the other hand, an intersection subspace $\mathcal{S}_i^\cap$ captures the impact of node i and all its possible outages on the grid. Proximities to these subspaces, shown with dashed red arrows, are less ambiguous than the ones corresponding to $\mathcal{S}_i^\cup$. However, subspaces $\mathcal{S}_i^\cap$ may misguide the detection when the failure to detect does not correspond to the total outage subspace. To address this issue, we propose to increase the difference between proximities to subspaces of different nodes by scaling each proximity $\mathcal{S}_i^\cup$ with the ratio between the proximities of the sample to the total outage subspace and the normal operational subspace:

$$\hat{\text{prox}}_{\mathcal{S}_i^\cup}(\mathbf{X}_{:,t}) = \text{prox}_{\mathcal{S}_i^\cup}(\mathbf{X}_{:,t}) \frac{\text{prox}_{\mathcal{S}_i^\cap}(\mathbf{X}_{:,t})}{\text{prox}_{\mathcal{S}_0}(\mathbf{X}_{:,t})} \quad (11)$$

The above calculation will be performed for all $i \in \mathcal{N}$ and can be carried out in parallel, and the results will form a proximity vector of the test sample: $\hat{\mathbf{prox}}(\mathbf{X}_{:,\mathbf{t}})$. Intuitively, for an outage at line $e_{i,j}$, the nodes being mostly impacted are i, j and their immediate 1-hop neighbors. The impact decreases when the nodes are located farther away from the outage location. This causes a monotonic decreasing trend in the proximities of neighbor nodes, and we call it a proximity rule.

The proximities found in $\hat{\mathbf{prox}}(\mathbf{X}_{:,t})$ describe how close the test sample is to all outage subspaces. Let us define $\mathcal{N}_t$, a list of size N where power nodes appear in an order according to the sorted version of $\hat{\mathbf{prox}}(\mathbf{X}_{:,t})$. According to the proximity rule, contiguous nodes in $\mathcal{N}_t$ that show a decreasing proximity trend should form a connected sub-component of the original grid. This group of nodes that follow the proximity rule allow us to identify the power line outages. Finally, the set of candidate lines identified as outaged, $\hat{F}$, is formed by adding the power lines $e_{i,j}, \forall i, j \in \mathcal{N}_t$.

V. Performance Evaluation

In this section, we test the proposed data-based outage identification mechanism. We are interested in evaluating the effectiveness of the proposed scheme on correctly identifying power line outages when PMU measurements are unreliable. Outage identification is performed using data samples containing phasor measurements of $N = |\mathcal{N}|$ power nodes in the grid. We consider that there exist a proper deployment of PMUs in the grid in order to provide complete observability. The problem of placement and deployment of PMUs to achieve full observability is out of the scope of this paper and we refer the readers to [13]. For the case of reliable PMU devices and communication links, data samples are complete; otherwise some data points are missing. Missing data will be tested by removing certain measurements, i.e. data points, according to the specified missing data pattern. Performance is studied using the 14, 30, 57, and 118 IEEE bus systems. These systems have 20, 41, 80, and 186 power lines available for outage evaluation, respectively.

The metrics of interest are: identification accuracy and false alarm rate. Identification accuracy (IA) is defined as the ratio of the number of correctly identified line outages to the total number of outage test samples. An outage sample is considered to be correctly identified if $\hat{F} \subseteq F$. False alarm (FA) rate is the average mismatch between F and $\hat{F}$ relative to the latter. Calculation of both metrics is performed as follows:

$$\begin{aligned} IA &= \frac{\mid \hat{F} \cap F \mid}{\mid F \mid}, \\ FA &= 1 - \frac{\mid \hat{F} \cap F \mid}{\mid \hat{F} \mid} \end{aligned} \tag{12}$$

Furthermore, we compare the proposed methodology against other learning-based and model-free outage identification mechanisms found in the literature [4], [14]. These works use slightly different variations of Multinomial Logistic Regression and will be referred as MLR in the tests presented below.

A. Data Sets

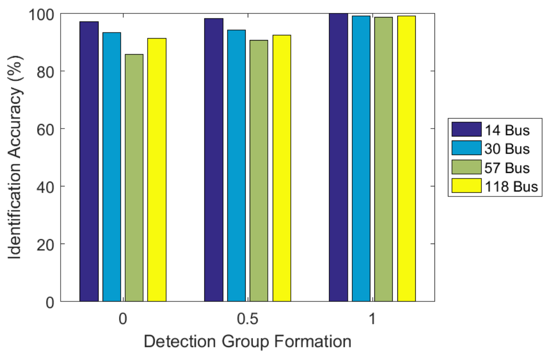
For training purposes, the proposed detection scheme uses historical data of the normal operation of the grid and operation during power line outages. Training data are synthetically generated using the topology and electrical information provided by the IEEE Bus systems [15]. An IEEE test case consists of information of parameters such as line impedance, rating and node (power bus) connectivity, among others. Node data contain information of the type of power bus (supply/generator, demand/load, slack), power demand, power output, and voltage phasors among others. The amount of power demand and/or output power of a node as specified in the test cases defines a single state of the power grid during normal operations. Such single state of the power grid corresponds to a single data sample. Different load variations are generated according to an Ornstein-Uhlenbeck process [16] to account for the dynamic and stochastic behavior of power demand in the grid over a period of time. The time period considered is 24 hours. Thus, the values of power demand specified in the test cases are considered to be the expected demand during one day of normal operations. Power generation (output) is adjusted accordingly. The different power demands, and adjusted power outputs, define different variations of the state of the power grid during normal operations. Using the generated load and generation levels, MATPOWER [17] is used to solve power flows and the resulting voltage phasors are considered to be PMU measurements. The AC model is used, instead of the DC approximation, when calculating synchrophasors and Gaussian noise is added to the voltage phasors [16] so that the obtained data can represent real PMU measurements.

As described in section IV-A, our proposed detection mechanism requires outage data of every line in the grid. Each set of outage data is generated by removing the corresponding line from the bus system topology information and rerun the power flow solver, thus obtaining outage PMU measurements. Cases that do not converge or result in disconnecting the grid after line removal, i.e. islanding, are not considered. The number of lines, out of the set of all lines in the system $\mathcal{E}$, that provide valid cases will be denoted by $E \leq |\mathcal{E}|$. Similarly as it is performed for other energy management applications, the generated data along with real PMU measurements collected during a period of time can be used to train the proposed outage identification system, and become part of the day-ahead planning and forecasting tasks performed at the control center. Generated data are divided into training and testing sets following the procedure described in [14].

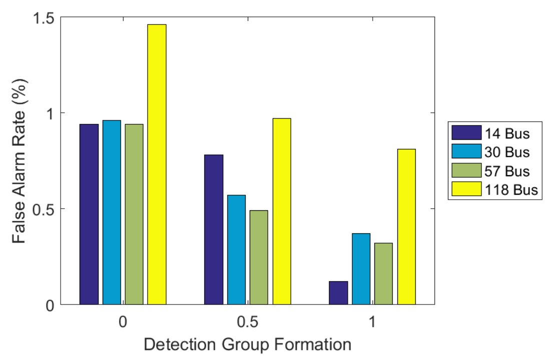
B. Complete Data Case

The proposed scheme identifies the location of the line outage using the proximity of the test sample to node-based outage subspaces. To demonstrate the effectiveness of the proposed subspace method, we evaluate the identification accuracy for several test samples when detection groups are fully formed and each test sample contains complete data.

Results in Figure 5 are the average accuracy of all single line outage cases, where 100 randomly selected test samples were evaluated for each outage case. Then, each bus system topology is evaluated using 100*E test samples, where E is the number of lines whose disconnection provides a valid test case as defined in Section V-A. Results are presented in comparison with the previously described Multinomial Logistic Regression methods, marked as MLR. It is worth noting that, different from MLR methods that learn identification rules based on

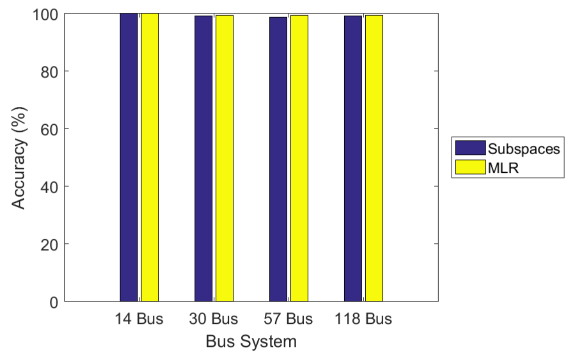


(a) Effect on Identification Accuracy

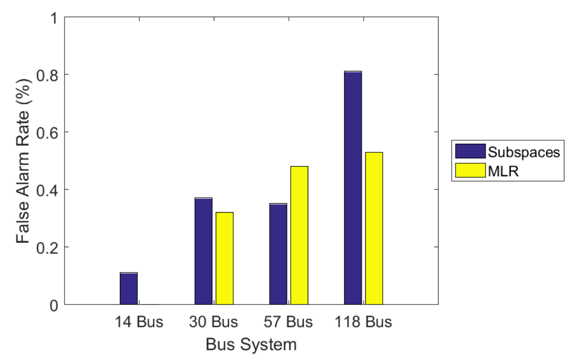


(b) Effect on False Alarm rate

Fig. 4: Effect of Detection Groups formation



(a) Identification Accuracy for Single Line Outages.



(b) False Alarm Rate for Single Line Outages.

Fig. 5: Complete Data case

specific single line outages, our proposed subspace method learns to identify outages based on the behavior of every node in the system when any of its lines is faulted. Therefore, while there is a very small identification accuracy difference between MLR and the proposed detection mechanism, the performance of both methodologies are comparable.

In Section IV-B we discussed that a naive way to form a detection group is to include nodes that have orthogonal loadings in the outage subspace. Furthermore, we described how to add robustness to a detection group by including node members determined with equation (8). As described in IV-A, ideally all nodes with high detection capabilities in $\mathcal{D}_C$ should be included in the detection group. Furthermore, as described in IV-C, the node subspaces through their detection groups contain additional information of the detection capabilities of the power nodes. Thus, in Figure 4, we evaluate the effect of the proposed choice of detection groups. Results for a case are obtained after the evaluation of 100 randomly selected data samples from the case testing set. The average identification accuracy of all single line outage cases is shown in the figure.

The values shown in the x-axis represent the portion of nodes in the detection group that were found using equation (8). Thus, with 0 the detection group used contains only orthogonal members, and with 1 outage identifcation is performed with the proposed robust detection group.

When no detection capabilities learning is performed, the detection groups only include nodes that showed low correlations in each intersection subspace. It can be seen in the figure, that when our methodology uses a naive detection group, i.e. 0 on the x-axis, identification performance is compromised with high false alarm rate. On the other hand, as we increase the portion of nodes found by (8) into a detection group, the performance is improved. Once we use a detection group formed as proposed, i.e. 1 on the x-axis, identification accuracy is relatively high for any power system evaluated, and the false alarm rate is also significantly reduced.

This demonstrates two things: 1) learning subspaces for identifying outages is effective, 2) by adding nodes with historic high detection capabilities to the detection groups, we improve the performance of the subspace method. Moreover, a detection group formed as proposed contains more nodes than the naive choice of orthogonal nodes. In the missing data case, some measurements of member nodes will not be available, then these larger detection groups can provide alternatives from reliable nodes with similar detection capabilities.

C. Missing Data Case

The previous study demonstrated that the proposed subspace method along with properly formed detection groups can effectively identify outage locations. As explained in Section I, data issues are often overlooked when designing applications based on PMU data. The design of our proposed detection

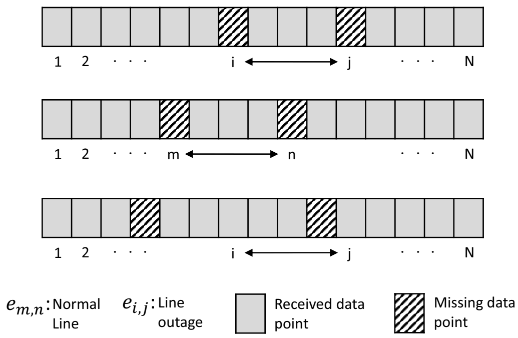
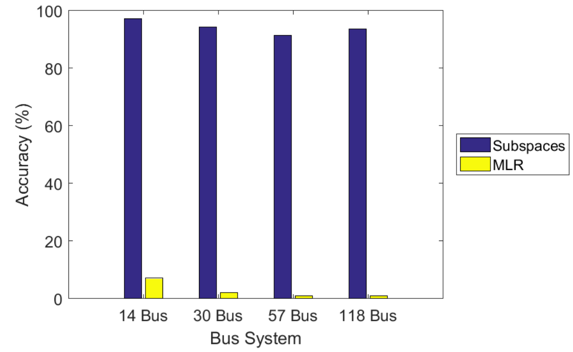
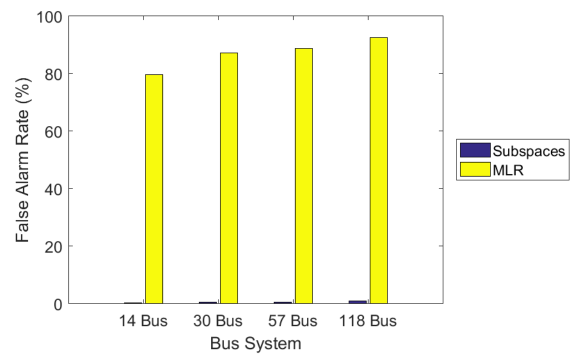


Fig. 6: PMU Data Missing patterns



(a) Identification Accuracy for Missing Outage Data



(b) False Alarm rate for Missing Outage Data

Fig. 7: Missing Outage data case

scheme was motivated by the possible occurrence of missing data. Now we evaluate the effect of missing data on the performance of the proposed solution. First, we evaluate the important case of missing data originated precisely at the outage location. Then, we study the impact of other missing data patterns that can occur in the PMU network. Namely, we evaluate the impact of another random pattern of relatively small number of missing measurements. Then we evaluate the systems under a generalization of possible missing data patterns.

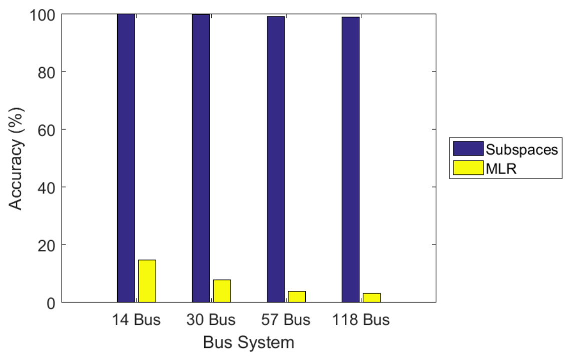
1) Missing Outage Data: Consider the missing data pattern shown at the top row of Figure 6. Given an outage of line $e_{i,j}$, the phasor data points of the corresponding nodes i and j are considered to be missing for this pattern. This missing data pattern is of particular importance as it can be originated as a consequence of PMU malfunctioning, or communication link unavailability, due to the outage itself. We account for other missing data patterns in the studies in the later cases.

The results are presented in Figure 7. Similar to the performance study with complete data, each system tests all possible line outages and 100 realizations of each outage case are evaluated. It can be seen in the figure, the performance of the subspace method is only slightly impacted by the presence of missing data from the outage locations. In fact, the performance is comparable with that when complete data are available for fault identification, using either our detection mechanism or peer solutions. On the other hand, the performance of MLR-based solutions is greatly degraded as such methods work specifically based on the data signatures of single line outages when complete data are available. Thus, it is expected that the identification of outage $e_{i,j}$ would require the data points from nodes i and j. The subspace method, instead, can exploit the use of available data samples from other nodes for the identification of an outage location. In this study, we have restricted that the data are only missed from the outage locations. Our results show that the performance of the peer methodologies is largely impacted even when there is only a relatively small number of missing data points.

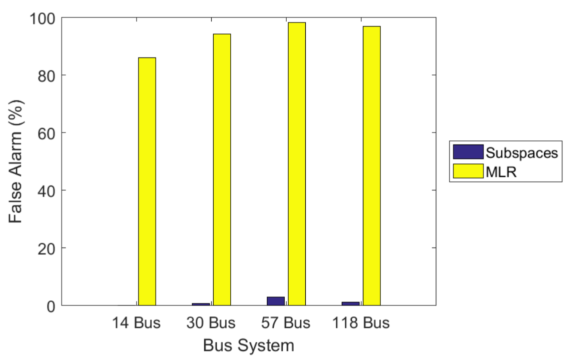
2) Missing Random Data: The missing data pattern used in Section V-C1 affect the data samples of the nodes associated with the outage being evaluated. Given that such data anomaly locations coincide with the outage locations, it could be argued that a data-based mechanism could claim to identify an outage only when encountering missing data points. In other words, the algorithm could classify missing data as a power outage. The following scenario will be used to evaluate the capability of an outage detection application to differentiate data anomalies, in the form of missing data, from power anomalies due to an outage.

Test samples are drawn exclusively from the normal operations case. Then, some data points are dropped as illustrated on the pattern of the middle row of Figure 6. That is, data points of existing lines are missing but no outage has occurred. Hence, $|F| = 0$ and $FA = 1$ in Equation (12) when $|\hat{F}| \neq 0$. Similarly, $IA = 1$ when $|\hat{F}| = 0$. In Figure 8, we can see that the false alarm of the subspace method is negligible and the performance is still comparable with that of the no missing data scenario, which demonstrate its robustness on differentiating missing data from actual outages. On the other hand, the extremely high FA rate associated with MLR methods further confirms that conventional methods based on data learning can confuse small missing data patterns with outages as explained before.

In Figure 9, we evaluate the missing data pattern shown at the bottom row of Figure 6. In this scenario, test samples are drawn from line outage cases and we restrict the random missing data are not from the outage locations. This is the case where missing data and outages are not correlated. In the figure, while peer solutions have a small improvement on their performance compared with the previous case, its

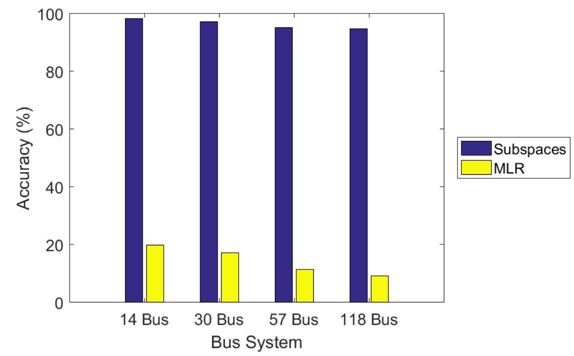


(a) Identification Accuracy for Normal Operations samples

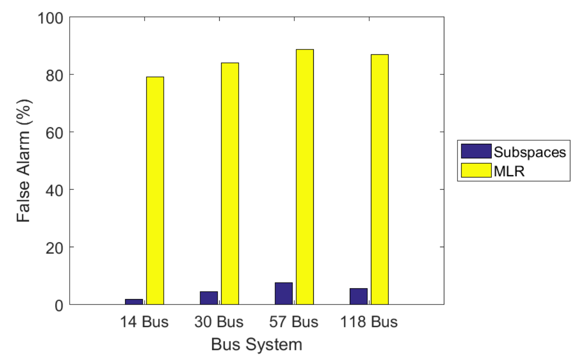


(b) False Alarm for Normal Operations samples

Fig. 8: Random Missing Data Normal Operations case



(a) Identification Accuracy for Outage samples



(b) False Alarm for Outage samples

Fig. 9: Random Missing Data Outage case

performance is far from that of the proposed detection scheme. These results indicate that the subspace method can effectively handle different missing data patterns of interest.

3) PMU Network Reliability: So far, we have studied missing data from outage and random locations, and their impacts on identifying the correct situation: outage (and its location) or normal operations. Now we generalize the patterns of missing data that can occur in the PMU network. We are interested in studying the impact of any number of missing data points in a test sample with no restriction on the location of the missing data points. In this study, we simultaneously consider all data patterns shown in Figure 6 with an arbitrary number of missing points.

To account for a realistic generalization of the results, we consider the reliability reported of common PMU devices and corresponding communication links [18]. Such reliability will define the probability of occurrence of a specific missing data pattern. For example, consider the extreme missing data pattern where all data points are missing in the sample. This corresponds to the case where all PMU (or communication links) are not working. The sample at this low reliability point of the network will cause any solution to not work properly. However, such scenario is highly unlikely to happen. Hence, to properly account for each missing data scenario, we use the effective false alarm rate for a given reliability level r scenario as:

$$FA(r) = \sum_{l=1}^{2^L} FA_l p_l(r) \tag{13}$$

Where FA_l represents the false alarm rate defined in (12) when data points are missing according to the pattern specified in scenario l. L represents the number of PMU devices (and their corresponding PMU $\rightarrow$ PDC communication links). Hence, we are considering a weighted average of all possible missing data patterns. Each combination determines which data samples are missing, indicating that the PMU and/or its communication link is not working. The parameter r represents the total system-wide reliability level of a PMU network of the type shown in Figure 1 and can be calculated as:

$$r = (r_{PMU} r_{PMU \rightarrow PDC})^L \tag{14}$$

where we have considered that the reliability of the L PMUs, and their associated communication links, are independent. r_{PMU} and $r_{PMU \rightarrow PDC}$ correspond to the reliability levels of the PMU device and its communication link with the PDC, respectively. We consider links to the Control Center, PDC $\rightarrow$ CC, to be reliable. The weight $p_l(r)$ is the probability of occurrence of the $l-th$ missing data scenario, calculated as follows:

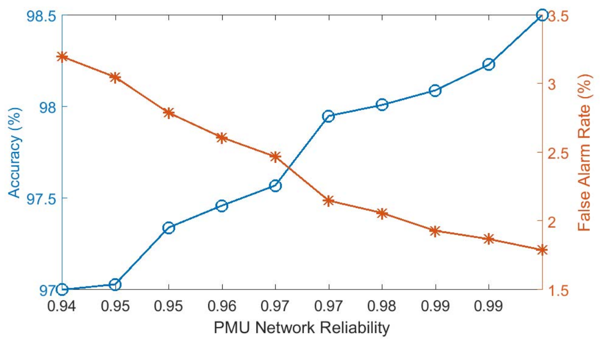


Fig. 10: Real PMU Network Reliability case

$$p_l(r) = \prod_{i=1}^{L} p_{i,l} \quad (15)$$

where $p_{i,l,r} = r_{PMU} r_{PMU \to PDC}$ if the i-th PMU device is working (no missing data) for the l-th missing data pattern, otherwise $p_{i,l,r} = 1 - r_{PMU} r_{PMU \to PDC}$. Each l-th of the 2^L data patterns considered in (13) corresponds to all possible combinations of missing data for all L PMUs. Thus, (13) provides the appropriate false alarm metric over all data patterns. The parameters information is adopted from [18].

Given the reported range of r_{PMU} and using Equation (14), we evaluate the effective false alarm rate of the proposed subspace scheme for different levels of system-wide PMU network reliability. Results are presented in Figure 10. It can be seen in the figure, the proposed methodology has a consistent performance with the scenarios of other missing data patterns previously evaluated. That is, the subspace-based detection scheme produces relatively small errors when identifying locations of outages under any possible missing data situation.

VI. Conclusion

PMU measurements provide high precision data that can be exploited to enhance traditional model-based power grid monitoring and control. We address one application of vital importance, the timely discovery of power line outages. Current detection schemes that make use of PMU data either assume that all measurements are available or ignore the missing data scenario. These considerations can largely compromise the performance of a data-based application. Particularly, without reliable data, detection mechanisms can fail to provide real-time situational awareness for grid operators to take the proper corrective control actions. In this paper, we proposed a subspace learning based detection mechanism that is robust to different possible missing data patterns. To address the missing data problem, we design detection groups that provide alternative measurements to process along with the learned subspaces. We evaluate the proposed methodology using different IEEE test systems. Our results show that, when data are complete, the subspace-based method along with properly formed detection groups can achieve comparable detection performance as peer methodologies. When tested over several common patterns of missing data, our performance is not compromised by the evaluated missing data scenarios and largely outperforms peer methodologies. Furthermore, we generalize the missing data scenario to account for different possible patterns of missing data according to the reliability of the PMU network and obtain similar robust performance.

References

[1] F. Aminifar, M. Fotuhi-Firuzabad, A. Safdarian, A. Davoudi, and M. Shahidehpour. Synchrophasor measurement technology in power systems: Panorama and state-of-the-art. *IEEE Access*, 2:1607–1628, 2014.

[2] P. Hines, K. Balasubramaniam, and E. C. Sanchez. Cascading failures in power grids. *IEEE Potentials*, 28(5):24–30, September 2009.

[3] Réka Albert, István Albert, and Gary L. Nakarado. Structural vulnerability of the north american power grid. *Phys. Rev. E*, 69:025103, Feb 2004.

[4] M. Garcia, T. Catanach, S. V. Wiel, R. Bent, and E. Lawrence. Line outage localization using phasor measurement data in transient state. *IEEE Transactions on Power Systems*, 31(4):3019–3027, July 2016.

[5] A. Y. Abdelaziz, S. F. Mekhamer, M. Ezzat, and E. F. El-Saadany. Line outage detection using support vector machine (svm) based on the phasor measurement units (pmus) technology. In *2012 IEEE Power and Energy Society General Meeting*, pages 1–8, July 2012.

[6] D. Nguyen, R. Barella, S. A. Wallace, X. Zhao, and X. Liang. Smart grid line event classification using supervised learning over pmu data streams. In *Green Computing Conference and Sustainable Computing Conference (IGSC), 2015 Sixth International*, pages 1–8, Dec 2015.

[7] M. Wang, J. H. Chow, P. Gao, X. T. Jiang, Y. Xia, S. G. Ghiocel, B. Fardanesh, G. Stefopolous, Y. Kokai, N. Saito, and M. Razanousky. A low-rank matrix approach for the analysis of large amounts of power system synchrophasor data. In *System Sciences (HICSS), 2015 48th Hawaii International Conference on*, pages 2637–2644, Jan 2015.

[8] P. Gao, M. Wang, S. G. Ghiocel, J. H. Chow, B. Fardanesh, and G. Stefopoulos. Missing data recovery by exploiting low-dimensionality in power system synchrophasor measurements. *IEEE Transactions on Power Systems*, 31(2):1006–1013, March 2016.

[9] T. Xu and T. Overbye. Real-time event detection and feature extraction using pmu measurement data. In *2015 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 265–270, Nov 2015.

[10] L. Xie, Y. Chen, and P. R. Kumar. Dimensionality reduction of synchrophasor data for early event detection: Linearized analysis. *IEEE Transactions on Power Systems*, 29(6):2784–2794, Nov 2014.

[11] S. Pal, B. Sikdar, and J. Chow. Real-time detection of packet drop attacks on synchrophasor data. In *Smart Grid Communications (SmartGridComm), 2014 IEEE International Conference on*, pages 896–901, Nov 2014.

[12] Shankar Narasimhan and Sirish L. Shah. Model identification and error covariance matrix estimation from noisy data using {PCA}. *Control Engineering Practice*, 16(1):146 – 155, 2008.

[13] Y. Zhao, J. Chen, A. Goldsmith, and H. V. Poor. Identification of outages in power systems with uncertain states and optimal sensor locations. *IEEE Journal of Selected Topics in Signal Processing*, 8(6):1140–1153, Dec 2014.

[14] T. Kim and S. Wright. Pmu placement for line outage identification via multinomial logistic regression. *IEEE Transactions on Smart Grid*, PP(99):1–1, 2016.

[15] University of Washington. Power systems test case archive. [Online] http://www.ee.washington.edu/research/pstca/, 2016.

[16] Magnus Perninge, Valerijs Knazkins, Mikael Amelin, and Lennart Sder. Modeling the electric power consumption in a multi-area system. *European Transactions on Electrical Power*, 21(1):413–423, 2011.

[17] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas. Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on Power Systems*, 26(1):12–19, Feb 2011.

[18] F. Aminifar, S. Bagheri-Shouraki, M. Fotuhi-Firuzabad, and M. Shahidehpour. Reliability modeling of pmus using fuzzy sets. *IEEE Transactions on Power Delivery*, 25(4):2384–2391, Oct 2010.