Published in final edited form as:

IEEE Trans Smart Grid. 2018 ; 9: . doi:10.1109/TSG.2017.2680742.

Space-Time Approach for Disturbance Detection and Classification

H. Gharavi [Life Fellow, IEEE] and **B. Hu [Senior Member, IEEE]**

Advanced Network Technology division (ANTD), National Institute of Standards and Technology,
Gaithersburg, MD 20899, USA

Abstract

A major challenge facing future grid systems is to identify the source of abnormal behaviors caused by faults or voltage instability. In this paper Phasor Measurement Units (PMUs) have been considered for detecting disturbances and degradation in the grid. Considering that the source of voltage instability mainly impacts neighboring areas, we present a simple and yet efficient algorithm that can identify affected areas. The algorithm is based on K-Mean optimization that classifies PMUs into different classes of power quality. We introduce the concept of a space-time solution to identify multiple faults. Since this is a multi-objective problem, we extend the K-mean algorithm to achieve space-time optimization in two successive stages (i.e., time and space). As an alternative approach to the space-time solution, we also present a hierarchical network that is based on a distributed mapping technique. This approach may only require time-based clustering as it limits the PMU population search to a small local region. We used our real-time Virtual PMU (VPMU) embedded into our software in the loop testbed to verify the performance of the proposed schemes. For the IEEE 39-bus transmission system, it is shown that the proposed space-time synchrophasor data classification scheme is capable of detecting and isolating areas in the grid that suffer from multiple disturbances, such as faults.

Index Term

Smart Grid; fault detection; synchrophasors; K-mean clustering; WAMS; SCADA; Virtual PMU (VPMU); PDC; Emulab; emulation

I. Introduction

A major challenge for future wide area measurement systems is how to respond securely and reliably to various disturbances in the next generation of power grid networks. Power systems are subjected to a wide range of disturbances, which may cause voltage and frequency instability. In fact, frequency and voltage are the most important parameters, where frequency is a system-wide characteristic whilst voltage is a local feature [1]. For instance, the disturbance leading to a voltage collapse may be initiated by a variety of causes, such as a severe change of system conditions, or when the reactive power demand exceeds the sustainable capacity of the available reactive power resource. Frequency deviations mainly result from an imbalance between energy supply and demand. In addition, voltage stability can be the result of a gradual variation of load, which could also affect the frequency.

Traditionally, disturbance detection relies on customer calls to the utility and through the SCADA (Supervisory Control and Data Acquisition) system [2]. The main disadvantage of SCADA is that the measurements may not be able to provide a timely assessment of the dynamic state of the power system [3, 4]. This is mainly due to low scanning rates, as well as a lack of time synchronization. In other words, the SCADA method may not be able to handle immediate event detection at the onset of an unstable event [4]. However, synchrophasor measurements offer real-time visibility to the dynamics of the power system that complement traditional SCADA measurements [5, 6].

Indeed, with the ever increasing deployment of Phasor Measurement Units (PMUs) in the power system, the dynamic has shifted towards situational awareness through a wide area monitoring system using synchronous measurement data generated by a network of PMUs [4, 7]. While the main objective of PMU deployment is to improve monitoring and dynamic control of the power grid, a new opportunity has recently emerged in how to utilize PMU measurement data for transmission line fault detection and location [8, 9]. For example, there has been considerable effort in using travelling waves (also known as telegrapher's equations) to locate faults in a single transmission line [10–13]. Bear in mind that some faults may cause voltage instability and others lead to voltage collapse. Furthermore, in many cases the source of voltage instability may not be the result of a single fault, but rather an imbalance between generation and consumption.

The presence of renewables can also contribute to voltage instability. Indeed, there can be multiple sources of such instabilities caused by changing load demands, excessive reactive power, or faults such as line tripping, short circuits, etc. Although methods can be applied to distinguish between faults and voltage fluctuations [14–16], the main challenge is to locate the proximity of the problem, regardless of its nature. Since voltage instability is a local phenomenon, our approach is based on identifying neighboring PMUs that are most affected by voltage instability. Furthermore, if demand and supply are not balanced due to a sudden overload condition at a particular location, it could lead to a frequency instability, unless it is quickly detected and mitigated.

In order to detect regions that have been impacted by faults or sudden overloads, we introduce a time-based voltage or frequency observation vector that uses K-mean optimization [17–19] within the space of large populations of PMUs. However, in a wide area monitoring system it is possible that multiple disturbances may occur simultaneously. As we will show later, the proposed k-mean optimization, despite its ability to accurately detect a single fault, is unable to identify multiple faults that may occur simultaneously in diverse regions throughout the grid network. Therefore, space-dependency is another factor that needs to be incorporated into the optimization process. Solving problems in both space and time requires a suitable multi-objective optimization solution, such as NSGA-II [20, 21]. In this paper, and for the sake of simplicity, we extend the K-mean solution into **two** successive multi-objective clustering processes. Furthermore, as an alternative approach to overcoming the space-dependency, we also consider a hierarchical synchrophasor network [22] to limit the space search associated with each local region. To achieve this, we propose a local mapping strategy to detect local disturbances, particularly on buses interconnecting neighboring regional networks.

In the following, we first present our proposed K-mean clustering technique by defining a time-based observation vector as the input. Furthermore, to offer a general solution to the problem of multiple disturbances (e.g., faults) in a large network, we introduce a recursive two-stage K-mean clustering process. In section III, we also present an alternative solution to overcome space-dependency with a local mapping scheme using a hierarchical network structure [22, 23]. In section IV, the proposed scheme is then evaluated for the IEEE 39-bus system. This is followed by final remarks and conclusion.

II. Space-Time Synchrophasor Data Classification

In this section we propose an approach that aims at partitioning a grid monitoring area into differing levels of power quality ranging from safe to unsafe. To achieve this, we offer a clustering technique to classify the synchrophasor data into a certain number of clusters, according to their degree of voltage fluctuation. The method is based on the K-means algorithm, which has been extensively used for vector quantization of speech and video signals [17–19]. The basic idea is to obtain an optimum set of K -cluster centers where every observation vector can be assigned to the nearest center. This is an iterative process with the object of minimizing the sum of the squared distance of all such assignments. However, in a wide area monitoring system it is possible that multiple disturbances, such as faults, may occur in different regions of the grid. This would require defining a position-based observation vector. Furthermore, to analyze the effect of a single disturbance we also consider a time-based observation criterion that looks at the consecutive time-sample of each synchrophasor date. To achieve both space and time optimization we consider a two-stage solution (shown in Fig. 1), which is described below.

In the first-stage we consider the time synchronized samples of positive sequence voltage to form the observation vector. For an N number of PMUs in the network, let's define a voltage-based vector for the i th node ($i=1, 2, \dots, N$) as:

$$V_i = [V_{t-n+1,i}, V_{t-n+2,i}, \dots, V_{t,i}]. \quad (1)$$

The vector has a dimension of n , where $t-n+1, t-n+2, \dots, t$ corresponding to the n consecutive time stamps in which the synchrophasor data has been measured (it should be noted that a frequency-based observation, $f_i = (f_{t-n+1,i}, f_{t-n+2,i}, \dots, f_{t,i})$ can also be used, as will be discussed in Section IV).

Let K be the number of clusters representing the power quality levels (e.g., $K=2$ in the case of fault detection) with n -dimensional cluster centers as

$$C = \{c_1, c_2, \dots, c_K\} \quad (2)$$

The clustering algorithm is run iteratively to minimize the error function defined as:

$$E(V) = \sum_{i=1}^K \sum_{V \in S_i} \|c_i - V\|^2 \quad (3)$$

where S_i is the cluster with the centroid being C_i .

The iteration process is done according to the following steps:

1. Select a set of K arbitrary n -dimensional center vectors within the observation vector domain.
2. Obtain the Euclidean distance between observation vectors and each cluster center.
3. Form a set of K clusters $S_i (i=1, 2, \dots, K)$ with the centroid being C_i and map each observation vector to the nearest cluster center, where

$$S_i = \{V_k: \|c_i - V_k\|^2 \leq \|c_j - V_k\|^2, 1 \leq j \leq K, j \neq i, 1 \leq k \leq N\} \quad (4)$$

4. Update the centroid of each cluster to represent a new set of cluster centers,

$$C_i = 1/m_i \sum_{V_k \in S_i} V_k \quad (5)$$

where m_i is the number of observation vectors assigned to the cluster S_i .

5. Repeat steps 2–4 until the assignments in each cluster remain unchanged.
6. Select the cluster with the longest distance to the positive sequence vector of the non-faulted balanced system (i.e., $V_{balance} = [V_{t-n+1,i} = 1, V_{t-n+2,i} = 1, \dots, V_{t,i} = 1]$) as the relatively Most Unstable Cluster (MUC).

Following the first stage in which the MUC has been identified, the second stage begins with the objective of localizing the synchrophasor data that has fallen into the MUC. In this case, K-mean clustering aims at further splitting the MUC in which the nodes within each sub-cluster can be as close to each other as possible, and yet as far from nodes in other sub-clusters as possible. In this case, for the M number of nodes in MUC, the space-based observation vector for the j th node ($j = 1, 2, \dots, M$) is defined as:

$$S_j = (x_j, y_j). \quad (6)$$

The procedure of space-based partitioning is similar to time-based partitioning, except that the voltage-based observation vector $V_i = [V_{t-n+1,i}, V_{t-n+2,i}, \dots, V_{t,i}]$ is replaced by the space-based observational vector $S_j = (x_j, y_j)$. In space-based partitioning, K-mean clustering is initially centered on selecting $K=2$, regardless of the number of faults. Under these

conditions, two spaced-based sub-clusters will be generated. The first step is to measure the location distance between the two sub-clusters' centers as:

$$D_{x,y} = [(C_x^{S1} - C_x^{S2})^2 + (C_y^{S1} - C_y^{S2})^2]^{\frac{1}{2}} \quad (7)$$

where C_x^{S1}, C_x^{S2} and C_y^{S1}, C_y^{S2} are the horizontal and vertical components of sub-cluster centers: C^{S1} and C^{S2} .

As shown in Fig. 1, the location distance is then compared with a pre-defined threshold, $d_{threshold}$ to assess whether the two sub-clusters belong to the same neighborhood. For example, if $D_{x,y} < d_{threshold}$ the two sub-clusters will merge into one sub-cluster. The resulting sub-clusters undergo further time-space splitting until the unstable buses are located, as depicted in Fig. 1. Although $K=2$ is used in the K-means algorithm, recursive time-space classification is capable of locating more than 2 faults. The number of final sub-clusters emerging from this process corresponds to the number of faults.

It should be noted that in the case where the number of nodes in a faulty sub-cluster exceeds two, further time-based refinement will be carried out to identify the worst affected areas. Bear in mind that since the algorithm in each stage has a finite number of partitions, it will converge to a local optimum [18]. There is no guarantee that global optimum can be reached. Nonetheless, it is widely believed that the effects of the initial centroids in the k-means process, which are normally randomly selected, could have a significant impact upon the stability and optimization of the results [24–28]. Therefore, finding a better initialization method would result not only in faster convergence, but also mitigate the tendency to fall into a local minimum [24–28]. Although there are many variations in intuitively selecting the proper quantity of cluster centroids, their performance largely depends on the user's application and this is deemed important enough to warrant a separate study. Based on the nature of the data set for our application, we have intuitively developed a method of determining the cluster center for $K=2$, which is based on merging and splitting clusters.

In our approach the initial centroid values are determined by first calculating the mean vector value of the entire data set. For example; for a grid system with an N number of PMUs we form an observation matrix based on N observation vectors as:

$$V^{PDC} = \begin{bmatrix} V_{t-n+1,1} & V_{t-n+2,1} & \dots & V_{t,1} \\ V_{t-n+1,2} & V_{t-n+2,2} & \dots & V_{t,2} \\ \dots & \dots & \dots & \dots \\ V_{t-n+1,N} & V_{t-n+2,N} & \dots & V_{t,N} \end{bmatrix} \quad (8)$$

we then obtain a row vector, M , containing the mean value of each column as follows:

$$M = [m_{t-n+1}, m_{t-n+2}, \dots, m_t] \quad (9)$$

where $m_{t-n+j} = \frac{(V_{t-n+j,1} + V_{t-n+j,2} + \dots + V_{t-n+j,N})}{N}$, $j=1, 2, \dots, n$, is the mean value of column j . We then calculate the corresponding standard deviation of each column shown below

$$\begin{aligned} \delta_1 &= \sqrt{\frac{(V_{t-n+1,1} - m_{t-n+1})^2 + (V_{t-n+1,2} - m_{t-n+1})^2 + \dots + (V_{t-n+1,N} - m_{t-n+1})^2}{N}} \\ \delta_2 &= \sqrt{\frac{(V_{t-n+2,1} - m_{t-n+2})^2 + (V_{t-n+2,2} - m_{t-n+2})^2 + \dots + (V_{t-n+2,N} - m_{t-n+2})^2}{N}} \\ &\dots\dots\dots \\ \delta_n &= \sqrt{\frac{(V_{t,1} - m_t)^2 + (V_{t,2} - m_t)^2 + \dots + (V_{t,N} - m_t)^2}{N}} \end{aligned} \quad (10)$$

Based on the mean vector M , as well as $\delta_1, \delta_2, \dots$ and δ_n , two ($K=2$) initial n -dimensional center vectors can be generated as:

$$\begin{aligned} C_1 &= [m_{t-n+1} - \frac{\delta_1}{2}, m_{t-n+2} - \frac{\delta_2}{2}, \dots, m_t - \frac{\delta_n}{2}] \quad (11) \\ C_2 &= [m_{t-n+1} + \frac{\delta_1}{2}, m_{t-n+2} + \frac{\delta_2}{2}, \dots, m_t + \frac{\delta_n}{2}]. \end{aligned}$$

The above two initial centroids, which are based on the mean vector and the corresponding standard deviations, can prevent the centroids getting too close to each other, but not too far away from the center of the whole entity (the mean vector). The results presented in Section IV verify that such selections can reduce the K-means computational complexity with respect to the number of iterations.

As shown in [24, 28], the computation complexity of the K-means algorithm is $O(N^{nK+1} \log N)$, where N is number of vectors to be clustered, n is the dimension of the vectors, and K is the number of clusters. Lloyd's K-means algorithm, one of the heuristic algorithms, has a complexity of $O(NnKI)$ [24], where I is the number of iterations.

III. Space Localization via Hierarchical Network

To overcome space dependency, we also consider a hierarchical synchrophasor network as an alternative approach to second stage clustering. Bear in mind that for wide area

measurement systems a Phasor Data Concentrator (PDC) first receives GPS-synchronized phasor measurement data from its local PMUs [23]. The aggregated synchrophasor data collected at each local PDC is then forwarded through a wide area network to the so-called Regional PDC (RPDC) or Master PDC (MPDC), which is normally located at the grid regional control center [22].

Fig. 2 shows an example of a hierarchical synchrophasor network that consists of two groups of PMUs communicating with their respective local PDCs. Here, our main objective is to process the PMU data at lower local PDC levels in order to identify buses/lines that suffer from power disturbances. It should be noted that a power disturbance can also affect the power profile of the neighboring regions. In order to detect links with poor power quality, it is important to recognize whether the source of voltage instability is confined within a single PDC or involves any of its neighboring PDCs.

If it is within the same local area, the respective PDC would initiate an appropriate action to overcome the problem. If the unstable region is shared between two neighboring areas, such actions need to be taken at the next level PDC and, in the case of a hierarchical PDC structure [22], this would require creating a local map at a lower PDC level. The map is constructed by assigning a so-called neighboring ID list to PMU nodes reporting to their local PDC. In the example shown in Fig. 2, there are two local PDCs: PDC-A and PDC-B both reporting to the next level PDC, i.e., Regional PDC (RPDC). According to this figure, the PDC-A and PDC-B each have 6 and 5 PMUs respectively. The local map of PDC-A is furnished with the neighbor IDs of its associated PMUs. Furthermore, if the neighbor PMU reports to another PDC, its ID is followed by its respective PDC, as shown in Fig. 2.

At each instance, as soon as a local PDC receives synchrophasor data from its associate PMUs, it performs a power quality assessment to identify the PMU nodes with abnormal behavior. Specifically, time-based partitioning will be performed and repeated when the MUC contains more than two observation vectors, in order to locate the source node or link of the disturbance. If such nodes are not edge nodes according to the local map (e.g., every node in PDC-A, except A6), the PDC would be able to classify the area between the two most abnormal nodes as an unstable region. In this case the PDC will then forward their corresponding PMU IDs, together with their synchrophasor data, to the RPDC. The PDC will also use the optional field to flag the status of the combined data frame as unstable [22]. If any abnormal PMU node is an edge node, the PDC will then forward the combined data to the RPDC by flagging its status as regionally unstable. Under these conditions and, if the combined data frame from the neighboring PDC has also been flagged as unstable, a further PMU data analysis will be carried out at the RPDC by invoking time-based partitioning on the involved nodes. Bear in mind that the main objective of such an analysis is to identify any two neighboring PMUs with the worst abnormal behavior. It should be noted that any form of disturbance, such as a fault occurrence or voltage fluctuation, can have a different impact on the synchrophasor data of the nearby PMU nodes. Although the above mapping strategy helps to localize the source of a disturbance, the main challenge is how to analyze, process, and classify the PMU data in order to detect voltage fluctuations.

IV. Performance Evaluation

In the following experiments we consider the IEEE 39-bus grid model (see Fig. 3) using 30 PMUs: 29 installed on bus 1 to 29 and 1 on bus 39. While PMUs can be placed optimally [29–34], our main objective has been mainly to assess the performance of the space-time scheme under various test conditions. We used our real-time EMULAB based synchrophasor network testbed [35, 36] which utilizes a licensed software package at the application layer and includes the IEEE 39-bus grid system [37, 38]. The testbed has the flexibility to install a Virtual PMU (VPMU) at any desirable location in the IEEE 39-bus grid system. The VPMU has been carefully calibrated according to the C37.118 [39, 23] for both measurements and real-time communications. The frame rate is 60 frames per second.

We first examine the performance of the space-time clustering technique for double and triple faults in a non-hierarchical (flat) synchrophasor network. Bear in mind that the main objective of space-time partitioning is first to detect disturbance occurrences and then identify their approximate whereabouts in the grid. To achieve this, K-mean partitioning is performed in two successive stages: time and space. In both cases $K=2$ is used for clustering, which is based on merging and splitting the unsafe cluster. For time-based partitioning, by assuming that the time of impact is short, we consider five consecutive measurements to be sufficient to detect a severe overload occurrence, such as fault. Under these conditions the observation vector for node “ i ” corresponds to: $V_i = (V_{-4,i} \ V_{-3,i} \ V_{-2,i} \ V_{-1,i} \ V_{0,i})$. We assume that all the PMUs communicate with a single PDC. For multiple disturbance detection two sudden overload changes (e.g., faults) are imposed simultaneously on buses 7 and 16 as shown in Fig. 3. The PDC upon receiving all the PMU data, first performs time based (first stage) partitioning to find the worse cluster (MUC). The positive sequence voltage vectors for all buses and the corresponding cluster center for two clusters are displayed in Table I. The initial centroid vectors have been selected according to Eq. (11) in order to carry out the two-cluster time-based classification. As indicated in Table I, cluster 1, covers buses 6, 7, 8, 15, 16, 17, 21, 22, 23 and 24. It has a cluster center: $C_1 = [1, 0.2218, 0.2191, 0.2182, 0.8734]$, which is far away from the non-faulted balanced vector $V_{balance}$ (i.e., distance is 1.3574), hence is classified as MUC (an unstable area). Cluster 2 with cluster center $C_2 = [1, 0.5538, 0.5488, 0.5484, 0.9336]$, is deemed as a stable area.

In the second stage, the PDC carries out space-based partitioning on the MUC (cluster 1) to further locate and isolate multiple faults. From Table II we can observe that two sub-clusters, namely sub-cluster-1 (buses 6, 7, 8) and sub-cluster-2 (buses 15, 16, 17, 21, 22, 23, 24), are generated. Since the distance between the two sub-cluster centers (i.e., $D_{x,y} = 3.1423$) is larger than the pre-defined threshold $d_{threshold} = 2.1$, both sub-clusters are classified as non-neighbors. Consequently, both have to be treated as two separate faulty clusters and will undergo further space-based splitting. In the second round, two sub-clusters that emerge from sub-cluster-1 will be merged back to one cluster indicating a single fault since the distance between the centroids of the two sub-clusters (i.e., 0.8766) is far less than $d_{threshold}$. Similar to sub-cluster-1, sub-cluster-2 will be merged back to one cluster, indicating a single fault. We should point out that the pre-defined threshold, $d_{threshold}$ is experimentally selected in such a way as to ensure that distances between faulty areas are far from each

other. More specifically, the threshold determines the size of the geographical area even though the sub-cluster may end up with more than one fault.

In the case where the number of nodes in a faulty sub-cluster contains more than two nodes, further time-based refinement can be carried out to identify the worst affected areas. As shown in Fig. 4, bus 7 is categorized as the faulty bus in sub-cluster-1 having a distance of 1.4871 to the non-faulted balanced vector $V_{balance}$. At the same time buses 16 and 24 are classified as the faulty area having a distance of 1.5019 to the non-faulted balanced vector $V_{balance}$. These results indicate that the space-time solution can effectively isolate faulty areas.

Fig. 5 shows the performance of the proposed space-time clustering scheme in a scenario where the digital data waveform generated at the application layer is corrupted by white Gaussian noise. Specifically, the signal to noise ratio (SNR) is set to 15 dB. As can be seen from Fig. 5, after time-based partitioning, bus 5 has fallen in the MUC due to the effect of noise. This is different than the result in Table I where there is no noise and bus 5 is excluded from MUC. However, after space-based partitioning and further time-based refinement, bus 5 is excluded from MUC and bus 7 is correctly categorized as the faulty bus. This result shows that low SNR may affect some edging buses (such as bus 5) which is far from the cluster center in the first round of recursion in space-time classification. With the help of further space-time refinement, the effect of noise will be eliminated and with an increase in SNR, the impact of noise is reduced and is negligible when the SNR is higher than 25 dB.

Table III depicts the performance of frequency based K-means clustering for a non-hierarchical (flat) IEEE 39-bus system when a sudden overload change happens on bus 16. Specifically, a 3-phase RLC load (a EMTP component with voltage being 25 kV (RMS, L-L), active power being 100 MW and reactive power being 50 MVAR) is suddenly connected to Bus 16 at time point A (3.0s) and disconnected at time point B (3.05 s). In this case, the frequent-base observation vector $f_i = (f_{4,i} f_{3,i} f_{2,i} f_{1,i} f_{0,i})$ is used for node: “i”. As can be seen from Table III, cluster 1, covering buses 15, 16, 17, 18, 21, 24 and 27 has a cluster center: $C_1 = [59.3154 \ 58.7892 \ 44.9277 \ 59.1118 \ 57.3436]$, which is far from the frequency balanced vector $f_{balance} = [60 \ 60 \ 60 \ 60 \ 60]$. After further frequency based K-means clustering on MUC, buses 16 and 17 ($C_1 = [59.3114 \ 58.7844 \ 41.0215 \ 58.6391 \ 56.8626]$) are classified as faulty areas having a distance of 19.3347 to the non-faulted balanced vector $f_{balance}$. These results indicate that frequency based K-means clustering can effectively isolate faulty areas.

In Table IV the performance of the traditional K-means algorithm with randomly selected initial centroids and the proposed K-means algorithm are compared using the multivariate iris data set from the UCI repository of machine learning databases [40, 41]. In these tests, the accuracy of clustering is assessed by comparing the clusters obtained in our experiments with the clusters already available in the UCI data set [41]. The percentage accuracy and the average number of iterations are displayed in Table IV. As can be observed from this table, the proposed K-mean algorithm can achieve faster convergence and much better accuracy, despite the additional computation required to obtain a suitable set of centroid values.

In our next experiment, we will examine the performance of space-localization via hierarchical networks for the same double and triple faults. As shown in Fig. 3, the IEEE 39-bus transmission system is divided into local regions: A and B. Time-based K-means clustering has been carried out simultaneously within the local PDC-A (i.e., buses 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 18 and 39) and PDC-B (i.e., buses 14, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28 and 29), respectively. In area A its local PDC classifies all the measurements into two clusters as shown in Fig. 6. Cluster 1 includes buses 6, 7, and 8 has a cluster center $C_1 = [1.0214, 0.213, 0.213, 0.862]$, which is far from the non-faulted balanced vector $V_{balance}$ (distance is 1.61) and can be viewed as the MUC. At the same time, cluster 2 with a center $C_2 = [1.05513, 0.5465, 0.5462, 0.9392]$ is closer to the non-faulted balanced vector $V_{balance}$ (distance is 0.7852) hence is classified as a stable cluster. Since there are 3 buses in MUC, the time-based K-means clustering algorithm is repeated to obtain the final MUC (cluster 1 containing bus 7). The PDC-B performs a two-cluster K-means algorithm in area B and the results are shown in Fig. 6. The cluster center of Cluster 1 with buses: 15, 16, 17, 21, 22, 23 and 24, has a longer distance to the non-faulted balanced vector $V_{balance}$ (distance is 1.3522) and is therefore considered an MUC. Similarly, since there are 7 buses in the MUC (cluster 1), the K-means clustering algorithm is repeated to get the final MUC (cluster 1 containing buses 16 and 24), as shown in Fig. 6.

In the following experiment, a single fault occurring on the interconnecting bus 17, is investigated by space location via a hierarchical network. In area A, its local PDC classifies all measurements into two clusters, as shown in Fig. 7. Cluster 1 includes buses 3, 4, and 18 is far from the non-faulted balanced vector $V_{balance}$ (distance is 0.96) and can be viewed as an MUC. Cluster 2 is closer to the non-faulted balanced vector $V_{balance}$ (distance is 0.48) and can be deemed a stable cluster. Since there are 3 buses in MUC (cluster 1), a K-means clustering algorithm is performed again to get the final MUC (cluster 1 containing bus 18). By checking the local map, PDC-A will recognize that bus 18 is an edge node connecting to bus 17 of area B. After labeling bus 18 as abnormal, PDC-A will then forward its synchrophasor data to the RPDC for further classification. PDC-B performs a two-cluster K-means algorithm in area B and the results are shown in Fig. 7. The cluster center of cluster 1 has a longer distance to the non-faulted balanced vector $V_{balance}$ (distance is 1.24) and is therefore considered an MUC. Again, a K-means clustering algorithm is repeated to get the final MUC (cluster 1 containing buses 16 and 17), as shown in Fig. 7. Since bus 17 is connected to bus 18 of area A (an edge Bus), PDC-B will then forward the synchrophasor data of buses: 16 and 17, to RPDC for further cluster analysis. The RPDC will then reassess the synchrophasor data for buses 16, 17, and 18 by carrying out a two-cluster classification. As displayed in Fig. 7, bus 17 is categorized in the MUC and can then be recognized as the faulty bus. This result demonstrates that the hierarchical network is capable of locating a single fault in the edge area.

For the single fault on bus 17, space-time partitioning will first generate an MUC including buses 3, 4, 15, 16, 17, 18, 21, 22, 23, 24 and 27 by performing time based partitioning. After space-based partitioning, two sub-clusters namely, sub-cluster-1 (buses 3, 4, 17, 18, 27 and sub-cluster-2 (buses 15, 16, 21, 22, 23, 24), are generated. However, since the distance (1.4576) between their centroids is less than $d_{threshold}$ both sub-clusters will be merged back

into one, indicating a single fault. Time-based partitioning will be performed again to locate faulty bus 17.

As can be observed from the results, both space-time and time-based hierarchical approaches can identify multiple faults. However, while the hierarchical approach offers less computational complexity, it is unable to detect multiple faults in each local region, including interconnecting buses. Nonetheless, in most practical applications the synchrophasor network may have a two-level hierarchical structure, as shown in Fig. 6. In order to incorporate multiple fault detection capabilities, a space-time solution may also need to be considered instead of time-based alone. This could consequently add to the computational complexity of the hierarchical approach.

Finally, we should point out that the results presented here are based on severe disturbances, such as faults. Nonetheless, the proposed space-time approach can also be used for power quality classification by selecting $K > 2$. Consequently, this can result in a higher number of clusters where each can represent a distinct power-quality level.

V. Conclusion

Power quality disturbances and ways to control it has been one of the most challenging issues in smart grid. Our objective in this paper is to develop a K-means algorithm based strategy that is capable of identifying regions that suffer from faults or undesirable voltage fluctuations. We have proposed a new concept called space-time classification. As a proof of concept we use a simple solution to a multi-objective problem using a two-stage K-mean optimization. As an alternative approach to localize a disturbance, we also proposed a mapping strategy capable of detecting unstable regions in hierarchical networks to overcome the space-dependency that can also reduce computation. The results indicate that the proposed strategy is very effective in detecting multiple sources of voltage instability.

VI. References

1. Frequency Instability Problems in North American Interconnections, DOE/NETL-2011/1473, Final Report, May 1, 2011, National Energy Technology Laboratory, www.netl.doe.gov
2. TECHNICAL INFORMATION BULLETIN 04-1, Supervisory Control and Data Acquisition (SCADA) Systems, NCS TIB 04-1 (2004).
3. Chenine, M; Karam, E; Nordstrom, L. Modeling and Simulation of Wide Area Monitoring and Control Systems in IP-based Networks. IEEE Power & Energy Society General Meeting; 26–30 July 2009; 2009. PES '09
4. Naduvathuparambil, B; Valenti, MC; Feliachi, A. Communication delays in wide area measurement systems. Proc. Southeastern Symp. System Theory; 2002. 118–122.
5. Giri J, Parashar M, Trehern J, Madani V. The Situation Room: Control Center Analytics for Enhanced Situational Awareness. IEEE Power and Energy Magazine. 10(5):24–39.Sep-Oct;2012
6. Phadke AG. The Wide World of Wide-area Measurement. IEEE Power and Energy Magazine. 6(5): 52–65.Sep-Oct;2008
7. Fang X, Misra S, Xue G, Yang D. Smart grid - the new and improved power grid: A survey. IEEE Commun Surveys Tutorials. 2012
8. Jiang, Joe-Air; Yang, Jun-Zhe; Lin, Ying-Hong; Liu, Chih-Wen; Ma, Jih-Chen. An adaptive PMU based fault detection/location technique for transmission lines. I. Theory and algorithms. IEEE Transactions on Power Delivery. 15(2):486–493.Apr; 2000

9. Chen, Ching-Shan; Liu, Chih-Wen; Jiang, Joe-Air. A new adaptive PMU based protection scheme for transposed/untransposed parallel transmission lines. *IEEE Transactions on Power Delivery*. 17(2):395–404.Apr; 2002
10. Magnago FH, Abur A. Fault location using wavelets. *IEEE Transactions on Power Delivery*. 13(4): 1475–1480.Oct; 1998
11. Chen W, Malik OP, Yin X, Chen D, Zhang Z. Study of Wavelet-Based Ultra High Speed Directional Transmission Line Protection. *IEEE Transactions on Power Delivery*. 18(4):1134–1139.Oct; 2003
12. Jafarian P, Sanaye-Pasand M. Traveling-Wave-Based Protection Technique Using Wavelet/PCA Analysis. *IEEE Transactions on Power Delivery*. 25(2):588–599.Apr; 2010
13. Glik, K, , et al. Electrical Review. Detection, classification and fault location in HV lines using travelling waves. R. 88 NR 6/2012
14. Jonsson M, Daadler JE. An Adaptive Scheme to Prevent Undesirable Distance Protection Operation during Voltage Instability. *IEEE Transaction on Power Delivery*. 18(4):1174–1180.Oct; 2003
15. Shen, G; Ajjarapu, V. A Novel Algorithm Incorporating System Status to Prevent Undesirable Protection Operation during Voltage Instability. 39th North American Power Symposium, 2007, NAPS'07; Sept. 30 2007–Oct. 2 2007; 373–378.
16. Haque, MH. Use Of Local Information To Determine The Distance To Voltage Collapse. *International Power Engineering Conference; 2007; IPEC; 2007*. 1–6.
17. MacQueen, J. Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics; Berkeley, California: University of California Press; 1967*. 281–297.
18. Lloyd, SP. Least Squares Quantization in PCM. *Bell Telephone Laboratories Paper; Murray Hill, NJ: 1957*.
19. Linde Y, Buzo A, Gray RM. An algorithm for vector quantizer design. *IEEE Trans Commun*. COM-28:84–95.1980;
20. Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm. *IEEE Transactions on Evolutionary Computation*. 6(2):182–197.2002;
21. Bilil, H, Aniba, G, Gharavi, H. Dynamic Appliances Scheduling in Collaborative MicroGrids System. accepted for publication, *IEEE Transactions on Power Systems*
22. Gharavi H, Hu B. Scalable Synchrophasors Communication Network Design and Implementation for Real-Time Distributed Generation Grid. *IEEE Transactions on Smart Grid*. 6(5):2539–2550.Sep; 2015
23. IEEE Std. C37.118.2011-2, IEEE Standard for Synchrophasor Data Transfer for Power Systems, revision of the IEEE Std. C37.118.2005. Dec, 2011
24. Inaba, M; Katoh, N; Imai, H. Applications of weighted Voronoi diagrams and randomization to variance-based k-clustering. *Proceedings of 10th ACM Symposium on Computational Geometry; 1994*. 332–339.
25. Arthur, D; Vassilvitskii, S. How Slow is the k-Means Method?. *SCG'06; June 5–7, 2006; Sedona, Arizona, USA*.
26. Har-Peled S, Sadri B. How fast is the k-means method? *Algorithmica*. 41(3):185–202.2005;
27. Arthur, D; Vassilvitskii, Sergei. k-means lower bound implementation. 2006. <http://www.stanford.edu/~darthur/kMeansLbTest.zip>
28. Velmurugan T, Santhanam T. Computational Complexity between K-Means and K Medoids Clustering Algorithms for Normal and Uniform Distributions of Data Points. *J Computer Sci*. 6(3): 363–368.2010;
29. Chakrabarti S, Kyriakides E. Optimal Placement of Phasor Measurement Units for Power System Observability. *IEEE Transactions on Power Systems*. 23(3):1433–1440.Jul; 2008
30. Kekatos V, Giannakis GB, Wollenberg B. Optimal Placement of Phasor Measurement Units via Convex Relaxation. *IEEE Transactions on Power Systems*. 27(3):1521–1530.Aug; 2012

31. Aminifar F, Lucas C, Khodaei A, Fotuhi-Firuzabad M. Optimal Placement of Phasor Measurement Units Using Immunity Genetic Algorithm. *IEEE Transactions on Power Delivery*. 24(3):1014–1020;Jul; 2009
32. Baldwin TL, Mili L, Boisen MB, Adapa R. Power system observability with minimal phasor measurement placement. *IEEE Transactions on Power Systems*. 8(2):707–715;May; 1993
33. Yuill, W; Edwards, A; Chowdhury, S; Chowdhury, SP. Optimal PMU placement: A comprehensive literature review. 2011 IEEE Power and Energy Society General Meeting; San Diego, CA. 2011; 1–8.
34. Nazari-pouya, H; Mehraeen, S. Optimal PMU placement for fault observability in distributed power system by using simultaneous voltage and current measurements. 2013 IEEE Power & Energy Society General Meeting; Vancouver, BC. 2013; 1–6.
35. Emulab. Network Emulation Testbed Home. www.emulab.net
36. Combined Grid Communication Testbed Implementation for Real-time Measurement, Calibration, and Communication. <https://www.nist.gov/programs-projects/combined-grid-communication-testbed-implementation-real-time-measurement>
37. Athay T, Podmore R, Virmani S. A Practical Method for the Direct Analysis of Transient Stability. *IEEE Transactions on Power Apparatus and Systems*. PAS-98(2)Mar-Apr;1979 :573–584.
38. Pai, Anantha. *Energy Function Analysis for Power System Stability*. Springer; 1989.
39. IEEE Std. C37.118.2011-1, IEEE Standard for Synchrophasor Measurements for Power Systems, revision of the IEEE Std. C37.118.2005, Dec. 2011.
40. Merz, C; Murphy, P. UCI Repository of Machine Learning Databases. Available: <ftp://ftp.ics.uci.edu/pub/machine-learning-databases>
41. Nazeer, KAA; Sebastian, MP. Improving the Accuracy and Efficiency of the k-means Clustering Algorithm. *Proceedings of the World Congress on Engineering 2009 Vol IWCE 2009*; July 1 – 3, 2009; London, U.K.

Biographies

HAMID GHARAVI received his Ph.D. degree from Loughborough University, United Kingdom, in 1980. He joined the Visual Communication Research Department at AT&T Bell Laboratories, Holmdel, New Jersey, in 1982. He was then transferred to Bell Communications Research (Bellcore), where he became a Distinguished Member of Research Staff. In 1993, he joined Loughborough University as a professor and chair of communication engineering. Since September 1998, he has been with the National Institute of Standards and Technology (NIST), U.S. Department of Commerce, Gaithersburg, Maryland. He was a core member of Study Group XV (Specialist Group on Coding for Visual Telephony) of the International Communications Standardization Body CCITT (ITU-T). His research interests include smart grid, wireless multimedia, mobile communications and wireless systems, mobile ad hoc networks, and visual communications. He holds eight U.S. patents and has over 130 publications related to these topics. He received the Charles Babbage Premium Award from the Institute of Electronics and Radio Engineering in 1986, and the IEEE CAS Society Darlington Best Paper Award in 1989. He served as a Distinguished Lecturer of the IEEE Communication Society. In 1992 he was elected a

Fellow of IEEE for his contributions to low-bit-rate video coding and research in subband coding for video applications. He has been a Guest Editor for a number of special issues, including Smart Grid: The Electric Energy System of the Future, *Proceedings of the IEEE*, June, 2011; a Special Issue on Sensor Networks & Applications, *Proceedings of the IEEE*, August 2003. He was a TPC Co-Chair for IEEE SmartGridComm in 2010 and 2012. He served as a member of the Editorial Board of *Proceedings of the IEEE* from January 2003 to December 2008. From January 2010 to December 2013 he served as Editor-in-Chief of *IEEE Transactions on CAS for Video Technology*. He is currently serving as Editor-in-Chief of *IEEE Wireless Communications*.

Bin Hu (Senior Member, IEEE) became a senior member of IEEE in 2013. He received his Ph.D. degree from the School of Electronics and Computer science at University of Southampton, Southampton, U.K., in 2006. Since September 2006, he has been with the National Institute of Standards and Technology, U.S. Department of Commerce, Gaithersburg, MD, where he is currently a research scientist in advanced network technologies division. His research interests include smart grid, video/image transmission, mobile communications and mobile ad hoc networks.

Fig. 1.

Flowchart of the proposed space-time synchrophasor data classification.

Fig. 2.
Hierarchical network with two local PDCs.

Fig. 3.
IEEE 39-bus transmission system.

Fig. 4.
Time based K-means clustering for the sub-clusters in Table II.

Fig. 5.
Time based K-means Clustering for Non-hierarchical (flat) IEEE 39-bus system where SNR=15dB

Fig. 6.
Space location via hierarchical network for IEEE 39-bus system with double faults.

Fig. 7.
Hierarchical K-means clustering algorithm for hierarchical IEEE 39-bus system.

TABLE I

Time based K-means Clustering for Non-hierarchical (flat) IEEE 39-bus system

	Bus	V_{-4}	V_{-3}	V_{-2}	V_{-1}	V_0
Cluster 1 (MUC)	Bus-6	1	0.278539293	0.273464124	0.272755318	0.832637523
	Bus-7	1	0.136292343	0.145771852	0.146172596	0.921900688
	Bus-8	1	0.226858147	0.220160184	0.219446326	0.83184428
	Bus-15	1	0.210647990	0.207339147	0.206968439	0.882956923
	Bus-16	1	0.128442950	0.132180462	0.131734924	0.866305842
	Bus-17	1	0.279945352	0.275384525	0.275281920	0.889133787
	Bus-21	1	0.206795588	0.201730213	0.199572400	0.874695378
	Bus-22	1	0.314184297	0.306470828	0.303925321	0.884096559
	Bus-23	1	0.293754457	0.285614693	0.284052179	0.883614914
	Bus-24	1	0.143015283	0.143165332	0.142320032	0.866907303
Cluster 2	Bus-1	1	0.761050033	0.759186393	0.759831124	0.969819666
	Bus-2	1	0.659235810	0.653302016	0.652764741	0.943289874
	Bus-3	1	0.491675871	0.484006642	0.483624602	0.920272422
	Bus-4	1	0.455104111	0.448318893	0.447709482	0.923390633
	Bus-5	1	0.466929328	0.461460580	0.460752041	0.931318341
	Bus-9	1	0.545170728	0.538685698	0.537957062	0.961887533
	Bus-10	1	0.501304763	0.497179946	0.496538690	0.932433096
	Bus-11	1	0.493134621	0.488788826	0.488134884	0.93261217
	Bus-12	1	0.486021105	0.483392527	0.482759357	0.934924596
	Bus-13	1	0.476062822	0.471847933	0.471224858	0.928613719
	Bus-14	1	0.4663539166	0.468665819	0.468062561	0.927832739
	Bus-18	1	0.461069582	0.454184471	0.454053654	0.910370082
	Bus-19	1	0.487755650	0.481564659	0.479467552	0.904822371
	Bus-20	1	0.466904629	0.457590443	0.454067916	0.890655309
	Bus-25	1	0.668680589	0.664837987	0.664567279	0.944600092
	Bus-26	1	0.558276701	0.551514738	0.551587814	0.927747456
	Bus-27	1	0.432654840	0.425119308	0.425138286	0.909813415
	Bus-28	1	0.673327267	0.667138563	0.667595378	0.944311756
	Bus-29	1	0.706947862	0.702046397	0.702730933	0.951454062
	Bus-39	1	0.819255887	0.817580970	0.819491192	0.981569447

TABLE II

Space-based K-means Clustering Result for MUC (cluster 1) in Table I

Sub-cluster-1		6			
	X	Y		X	Y
Bus-6	1.4	3.72	Bus-15	2.9	2.25
Bus-7	1.1	4.1	Bus-16	3.3	2.0
Bus-8	0.75	4.63	Bus-17	3.2	1.6
			Bus-21	4.8	2.1
			Bus-22	4.7	4.05
			Bus-23	4.13	3.4
			Bus-24	3.63	2.7

TABLE III

Frequency based K-means clustering for non-hierarchical (flat) IEEE 39-bus system

	Bus	f_{-4}	f_{-3}	f_{-2}	f_{-1}	f_0
Cluster 1 (MUC)	BUS-15	59.294611	58.764338	47.451465	59.396527	57.607447
	BUS-16	59.308641	58.781191	43.482230	59.271108	57.161009
	BUS-17	59.314198	58.787638	38.560763	58.007160	56.564165
	BUS-18	59.343970	58.822620	45.568968	58.682654	56.986947
	BUS-21	59.362255	58.845469	47.016734	61.216830	58.087679
	BUS-24	59.290575	58.760068	44.956926	58.738363	57.757216
	BUS-27	59.293351	58.763063	47.456915	58.470080	57.240887
Cluster 2	BUS-1	59.933771	59.668767	58.539237	58.919623	58.379339
	BUS-2	59.666712	59.234906	56.177549	58.806895	57.900907
	BUS-3	59.441460	58.940581	52.539675	58.873401	57.606940
	BUS-4	59.409391	58.900863	54.744631	58.971690	57.818545
	BUS-5	59.515566	59.033268	56.338081	59.030353	57.970260
	BUS-6	59.536760	59.060547	56.517241	59.022772	57.994149
	BUS-7	59.501959	59.015664	56.639068	59.139756	57.977922
	BUS-8	59.510321	59.026188	56.672031	59.210312	57.997994
	BUS-9	59.895886	59.592947	58.144447	59.646774	58.341691
	BUS-10	59.561874	59.093589	56.176927	58.910414	57.986702
	BUS-11	59.552685	59.081466	56.313145	58.915864	57.975587
	BUS-12	57.944956	57.419434	58.102526	57.138214	56.328428
	BUS-13	59.528860	59.050772	55.632519	58.923760	57.953421
	BUS-14	59.443077	58.942576	53.989556	59.033556	57.895465
	BUS-19	59.442263	58.943052	53.308136	59.211824	58.183313
	BUS-20	59.363313	58.846856	54.918276	59.048135	58.220500
	BUS-22	59.517908	59.038854	51.022485	59.783675	59.106519
	BUS-23	59.491332	59.005053	51.096034	58.631236	59.553937
	BUS-25	59.626190	59.179622	56.036914	58.870707	57.955803
	BUS-26	59.385148	58.872486	52.411835	58.473198	57.645027
	BUS-28	59.392827	58.882346	54.776271	58.725770	57.945472
	BUS-29	59.474327	58.983132	55.352718	58.828178	58.052835
	BUS-39	59.996884	59.857097	59.368822	59.168737	58.560867
	Bus	f_{-4}	f_{-3}	f_{-2}	f_{-1}	f_0
Cluster 1 (MUC)	BUS-16	59.308641	58.781191	43.482230	59.271108	57.161009
	BUS-17	59.314198	58.787638	38.560763	58.007160	56.564165
Cluster 2	BUS-15	59.294611	58.764338	47.451465	59.396527	57.607447
	BUS-18	59.343970	58.822620	45.568968	58.682654	56.986947
	BUS-21	59.362255	58.845469	47.016734	61.216830	58.087679
	BUS-24	59.290575	58.760068	44.956926	58.738363	57.757216
	BUS-27	59.293351	58.763063	47.456915	58.470080	57.240887

TABLE IV

Comparison of the K-means with Randomly Selected Initial Centroids and K-means with Standard Deviation-based Initial Centroids by Using Iris Data Set

Initial centroids	Randomly selected	Standard deviation-based
Accuracy	81.3%	94.33%
Average number of Iterations	2.56	2.33