Large Power Network Signature Analysis with PMU Signal-Dynamic Clustering Approach

R.K. Pandey, *Senior Member IEEE,* Satish Kumar and Chandan Kumar

Abstract— **The paper presents signature analysis of large power network with Phasor Measurement Unit (PMU) signal using dynamic clustering approach. This concept may become more useful for system operators having large numbers of PMUs deployment in utility power network for observing the variations due to dynamical changes. It is well known as more numbers of PMUs are introduced in the network, there is an exponential increase in data which may require a special technique to understand the behaviour of entire system within time for generating quick control actions for system regulation. In present work, the concept of Clustering Algorithms (CA) to mine the synchrophasor data from PMUs has been applied with practical utility data (POWERGRID). Due to large data generated by PMUs, it becomes imperative to mine data so that important information from dataset can be utilized for monitoring of grid health followed by modified control in order to realize effective system operation. The procedure for extracting information from this data has been discussed through clustering algorithms.**

Keywords—: Phasor Measurement Uunits (PMUs), Cluster, Wide Area Monitoring System (WAMS), SCADA

I. INTRODUCTION

The challenges and complexity in the power system operation is increasing manifold day by day as a result of various internal and external factors. Internal factors include increased system size, brisk pace of capacity addition and congestion in transmission corridors whereas the external factors may be continuously increasing competition in electricity market, weather effects, and large scale integration of renewable energy resources concentrated in certain areas. The increasing gap between demand and supply may have additional effect [1].

The grid parameters are conventionally monitored by Supervisory Control and Data Acquisition (SCADA) system. But in present monitoring scenario SCADA is not efficient enough to address the above problems as this system suffers from many drawbacks. The SCADA systems provide data at 2-5 samples per second and are affected by time skew. Monitoring capability of SCADA is dependent on signal processing from RTUs with observability of the steady state nature only. Also, the online phasor angle measurement is not possible in real time. Due to these limitations operators in the power grid had to resort to the state estimation in order to determine the state of the grid and hence a delay with imprecise result. In the meantime, some important events may go unnoticed and in the worst case, this may lead to the collapse of entire power network. Above problems can be addressed by PMUs and related signal processing. The PMU produces synchronized measurements of phasor (i.e., its amplitude and phase), frequency, ROCOF (Rate of Change of Frequency) from voltage and/or current signals based on a common time source that typically is the one provided by the Global Positioning System UTC-GPS.

The phasor data transmitted by the PMUs are time tagged with the UTC time known as synchrophasor. Synchrophasors are sent by each PMU every second to a Phasor Data Concentrator (PDC) which arranges the data according to their time stamp and finally collected by a super data concentrator. Since all the PMUs are synchronised with the UTC and data is being updated at a very high rate, this data can be used for monitoring grid health. The measurements recorded by PMUs help the operator to identify the problems of power network. With this multi-dimensional high resolution data, even minute changes in the grid can be visualized by the grid operators like never before which may help to generate control actions in real time. With the introduction of Synchrophasor technology, it is now possible to monitor the phase angles difference at the control centre. Also, it enables visualization of voltage and current phasors, frequency, rate of change of frequency and angular separation at every few millisecond interval (in India at 40 milliseconds) in the Load Despatch Centre [1-5]. Thus the dynamical behavior of power system can be observed in almost near real-time at the control centre which hitherto was possible only in offline mode in the form of substation Disturbance Records (DR) or through offline dynamic simulations performed on network model. Hence a paradigm shift from state estimation to state determination can has been easily achieved. So, the problems aroused due to SCADA limitations and complexity of the power networks are successfully addressed by the PMUs.

But the data collected by the phasor measurement units (PMUs) is huge. The rate of generation of this data is about 40-250 observations per second, resulting in large volumes of phasor data. As a greater number of PMUs are brought online, the growth in the size of the data occurs exponentially. In larger wide-area monitoring systems containing a large number of PMUs, the amount of data generated can be so large that it may be difficult to analyze this data. For example, the Tennessee Valley Authority (TVA) presently handles 120 online PMUs with 3.6 billion measurements archived per day with a storage size of 36GB [3].While these data will help an utility to understand its regional/national grid operations at a granular level and enable more targeted operational decisions, ultimately leading to more efficient grid operations, managing the data flow and its analysis is very tedious and time consuming in classical way.

Since this data contains vital information which can be beneficial for the normal operation of grid, it becomes imperative to process this data timely. The processing of such a

R. K. Pandey, and Satish Kumar are with Electrical Engineering Dept., Indian Institute of Technology (Banaras Hindu University), Varanasi, India (e-mail: rpsneh@yahoo.co.in, Satish.kumar.eee11@iitbhu.ac.in).
Chandan Kumar is with Power Grid Corporation of India Limited (PGCIL). (E-mail: chandan8240000@gmail.com).

large volume of data is a big challenge yet. Such data cannot be mined using simple techniques because it needs to be processed immediately and cannot be stored owing to its size. So, algorithms that process this data reliably are required. The algorithms should be good enough with the rate of arrival of the data else the ultimate goal of installing PMUs is lost. Data mining techniques combined with statistical analysis can be used to extract important information from this data. In this paper, two methods have been discussed which can be used for analysis of synchrophasor data. The first half of the paper deals with k- means clustering algorithm and the second half deals with moving -clusters algorithm. The algorithms have been tested on MATLAB and experimental results have been discussed.

Section II provides the background for data mining and in section III Moving Clusters Algorithm is developed. The analytical studies have been discussed in section IV while section V concludes the work.

II. DATA MINING

Data mining is the process of extracting useful information from large data sets by utilizing different techniques such as statistical techniques, artificial intelligence etc. A fundamental idea is to discover patterns in the data which generally involves the identification of unusual data records that might be interesting or contain data errors or finding how the variables are associated to each other or clustering into groups of "similar" objects or use of functions to model available data with the least error. It also involves a compact representation of the data for proper visualization and report generation.

Data mining has been applied to several power system applications. The application of clustering techniques to decompose the power networks based on load flow analysis was earlier described by Saleh et al [6]. "Data mining" term in power system domain has been used by Madan et al [7] which brainstorms data mining applications in power systems. Data mining algorithms were used for describing power network states as normal or problematic. Also, authors mentioned the decision trees to classify a power system as stable or unstable, discovering change of data values from previously stored ones to detect unusual patterns, load forecasting and diagnostic expert systems for contingency analysis.

Following this, several researchers have used data mining techniques in different applications. Asheibi et al. [8] have made use of clustering techniques to identify classes of harmonic data from medium and low voltage distribution systems. The clusters thus obtained were again merged into super clusters by the further application of data mining techniques. The main aim has been to find the relation between the patterns of harmonic currents and voltages at different sites (substation, residential, commercial and industrial) for the interconnected super clusters. Groups of generators are identified which gain ability to increase revenue without increasing dispatch and hence have the potential for market mechanism. Mori [9] presents an overview of data mining papers in power systems and supports the arguments for application.

While data mining techniques have been applied in different power system applications, these methods have gained a renewed interest in the context of Smart Grids due to the fact that integration of data and information systems is one of the key advantages of the Smart Grid [10]. Over the last 10-15 years, there is an increasing trend of introducing new sensors which is anticipated to continue in coming years as well [11]. There are currently several hundred PMUs already deployed in various regions such as North American grid and there are plans for expansion also [12]. In the Western Interconnected system, there are plans to increase the no. of PMUs significantly over the years. In the Eastern Interconnect, 60 PMUs have been installed and 8 Phasor Data Concentrators (PDC) have been deployed to aggregate this data. These PDCs stream their data to a super PDC at the Tennessee Valley Authority (TVA). To cover the grid adequately, it is estimated that at least one third of the bulk power systems locations should be monitored by PMUs, thus requiring about thousands of PMUs and other resources for processing billions of data samples per day. Hence, the most important work is to handle the increasing volumes of data acquired by the network with new sensors and measurement devices. As volumes of such data are becoming gigantic, terms such as "data explosion" or "big data" are being used more frequently to describe in the expanding power system. Such data may be utilized for improved performance of the power system. However, this will require proper mining and right interpretation of the data. Analytical methods based on advanced concepts are statistical tools, pattern recognition and intelligent controls which will increasingly become need of the hour for system improvement.

Hence, there is a critical requirement for development of highly efficient algorithms that could uncover important patterns and information from power system data and thus derive the advantage of Smart Grid. In order to address this upcoming problem of data explosion, data mining comes up as a natural choice due various inherent features. In particular, these applications concern the use of model data (sensitivities) to identify different pattern, the data enhanced estimation of network models, event identification and dealing with the challenges of real-world data and data quality.

A specific type of data of interest in power systems is transient stability data collected over a period of time in a wide area measurement system (WAMS) subject to certain disturbances. For example, consider a large power system with a large no. of buses and generators. If there is a large disturbance in the system, frequencies in the system will start oscillating and depending on the size and type of the system, a propagation of these disturbances will occur from one area of the system to the other. Given a data sheet with voltage and frequency measurements of these buses over the duration of disturbance, the pertinent questions are: what useful information can be obtained about the system just by analyzing this data? How one can interpret system conditions from data and hidden information at a glance? Since transient stability data is dynamically changing and, there is an added challenge of showing not just the value of a variable or state at a single point of time but rather to capture the variation. Normally such variations can be shown by strip-charts and multi-color plots, but it will lead to the loss of information embedded in the data about the geographical location of source. From the view point of power grid, information of the location is important since it helps in understanding how one area of the power system behaves as compared to another. This may be benefitted significantly by data mining techniques and is one of the key challenges that need to be addressed. Clustering techniques have been found to be very useful in solving such problem.

Cluster analysis is a major part of data mining. The ultimate aim of clustering algorithms is to group objects based on the information that describes them. Clusters can be considered as classes, to which the clustering process automatically assigns objects. Clustering is considered as an unsupervised learning since it does not require class labels to be known ahead of time. Several clustering algorithms are available in data mining literature [13]. Fig. 1 shows few data points and one possible way to divide them into clusters on the basis of their colour. It is important to note that there can be multiple ways to cluster objects since the same set of objects can be classified differently according to different attributes. So selection of the appropriate attribute and clustering algorithm for any application may prove significant.

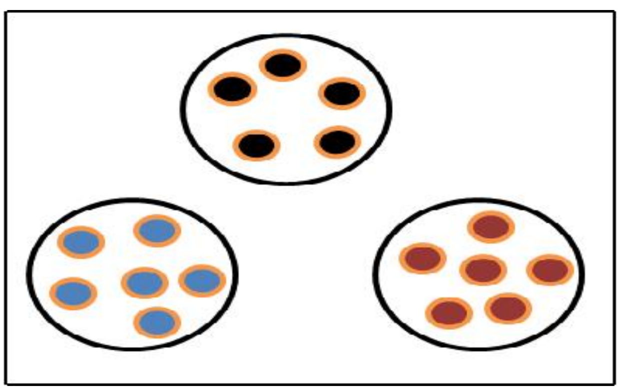

Fig.1 Cluster concept

Clustering is a powerful tool as it can identify groups of similar attributes or similar patterns. If there is any unusual pattern or object, it will usually form a single-object cluster. Such unusual behavior can be used for spotting "interesting" objects and patterns.

III. Development of Algorithm from K Means Clustering Algorithm

This sections deals with development of clustering algorithms. Primarily K-means clustering technique has been given and compared with moving cluster algorithm for a practical system data.

1. K-means Clustering Algorithm

K-means clustering is an algorithm for grouping a given data set into a certain K number of clusters. Each data point in the dataset belongs to the cluster with the nearest mean. It is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The K-means algorithm is an evolutionary algorithm that gains its name from its method of operation. The algorithm clusters observations into K groups, where K is provided as an input parameter. It then assigns each observation to clusters based upon the observation's proximity to the mean of the cluster. The cluster's mean is then recomputed and the process begins again. The steps of algorithm are given below:

Step1 – Set the value of K.

Step 2 - Define K centroids, one for each cluster. Centroids from the given data set should be selected in a way that they are as much as possible far away from each other.

Step 3 - Take each point belonging to the given data set and associate it to the nearest centroid forming clusters. When all the points have been grouped, the first iteration is completed.

Step 4 - Now K new centroids are recalculated as centers of the clusters resulting from the previous step.

Step 5 - Once again the same data set points are grouped so that each one is now associated with its nearest new centroid.

Step 6 - This process continues till there is no change in the location of the centroids. In other words, centroids do not move any more.

<u>Constraints</u>

The above algorithm is good when a fixed data is provided. The value of K will be predefined and K means clustering algorithm can be implemented for real time data of power grid, it is noted that no fixed set of data is available as PMUs are continuously generating new data in real time. So, the above algorithm cannot be directly applied and some modifications will have to be done so that data is processed in real time and control action could be initiated accordingly, if required. So, another algorithm with a sliding window which addresses above problem has been developed.

2. Moving Clusters Algorithm (MCA)

In real time PMU generate a continuous stream of data (in India it is produced every 40 milliseconds). Here data is not fixed, so a sliding window is used. An upper and lower threshold of the quantity to be observed is set for data window variation. If the value recorded at any point is beyond the tolerable limits, the algorithm should initiate a control action.

Let the quantity on which the algorithm is being applied be X and the lower and upper limits (considering a pre-defined tolerance) be X_L and X_U respectively. A sliding data window is used which shifts in the forward direction by one data point (for example, if the algorithm is executed on 0-20 data points for a particular window width, it would now be executed on 1 - 21 data points) when the algorithm has finished execution on the previous data window. Every data is recorded after 40 milliseconds and hence the time frame will be decided by the width of the window taken.

The following steps form the process of algorithm development. The flow chart has been given below in Fig.2.

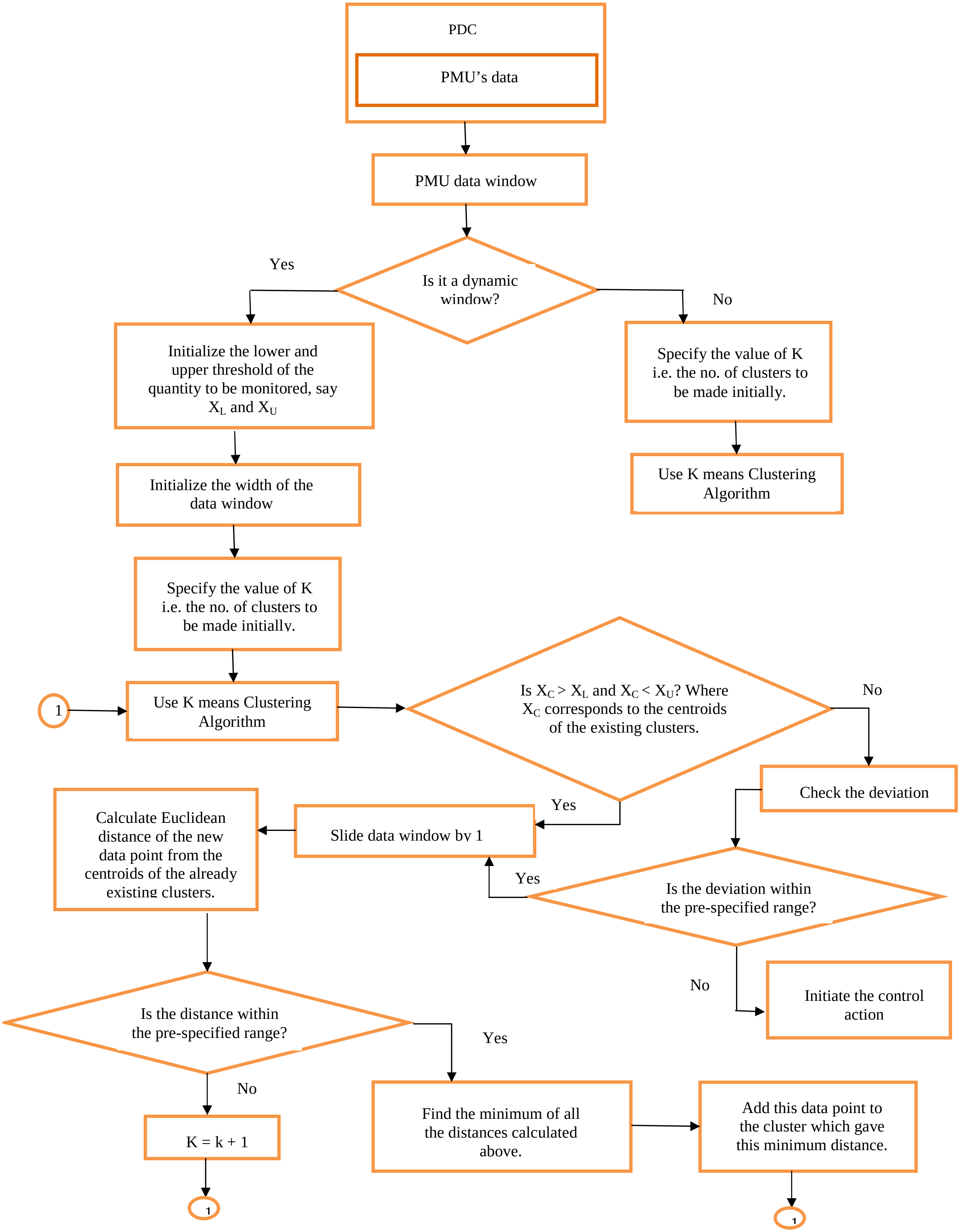
PDC
PMU's data
PMU data window
Is it a dynamic window?
Yes
No
Initialize the lower and upper threshold of the quantity to be monitored, say X_L and X_U
Specify the value of K i.e. the no. of clusters to be made initially.
Use K means Clustering Algorithm
Initialize the width of the data window
Specify the value of K i.e. the no. of clusters to be made initially.
1
Use K means Clustering Algorithm
Is $X_C > X_L$ and $X_C < X_U$? Where X_C corresponds to the centroids of the existing clusters.
No
Check the deviation
Yes
Slide data window by 1
Calculate Euclidean distance of the new data point from the centroids of the already existing clusters.
Yes
Is the deviation within the pre-specified range?
No
Initiate the control action
Is the distance within the pre-specified range?
Yes
No
K = k + 1
Find the minimum of all the distances calculated above.
Add this data point to the cluster which gave this minimum distance.
1
1

Fig. 2 Flow chart of dynamic clustering algorithm

The algorithm can be written as follow:

1. Initialize the lower and upper threshold of the quantity being monitored.

2. Initialize the width of the data window.

3. Specify the value of K.

4. Apply K means Clustering Algorithm.

5. If (X_C>X_L and X_C<X_U): (X_C is the centroid of the cluster)

   Flag=0 (Status OK)

6. Else (Boundary is violated and control initiated based upon signal):

   Flag=Flag+1

   i) check the deviation, d

   ii) if (d < allowable deviation)

   Go to step 7

   Else

   Execute Control action.

7. Slide data window by one.

8. Calculate the Euclidean distance of the new data from all the existing centroids.

9. If this distance is within pre-specified range:

   i) Find the minimum value of distance and note the cluster which gives this minimum value.

   ii) This new data will be added to the above cluster.

   iii) Go to step 4.

10. Else k = k + 1

    i) Go to step 4.

IV. ANALYTICAL RESULTS

The analysis with Moving Clusters Algorithm (MCA) has been done using data sets made available from Power System Operation Corporation of India (POSOCO), a subsidiary of Power Grid Corporation of India Limited (PGCIL) for different real time situations. The value of deviation, d can be calibrated depending on data set from history. The threshold was set depending on data set. The plots have been obtained by using the dataset of Dehgam region. Fig. 1 shows the voltage for cluster 1. In cluster 1, the most dip in phase c has been observed, followed by recovery at time 8:38:44:920, whereas in Fig.4 cluster 1'shows a different result as compared to the cluster 1, which indicates a little rise in all phases at 8:38:55:080.

In Fig.5, cluster 1" however maintains almost constant voltages in all phases which can be seen as the system has recovered with operator's decision (of course manual). It can be concluded that in situations of the control strategy if implemented which is linked with the change in phase voltages may not result the system behaviour as observed in Figs.3-4, i.e., cluster 1 and 1'. This indicates that the real time data processing if done with the present method, the entire system may remain in good health and this concept may be more beneficial for designing controllers for self healing grid behaviour. Therefore, it is recommended that PMU signal should be used for controller activation and if possible PMU may form an integrated part of controller as ancillary services module for correcting voltages/frequencies at different locations for dynamically changing network. The authors have proposed the similar recovery procedure for the system restoration and effective regulation [14].

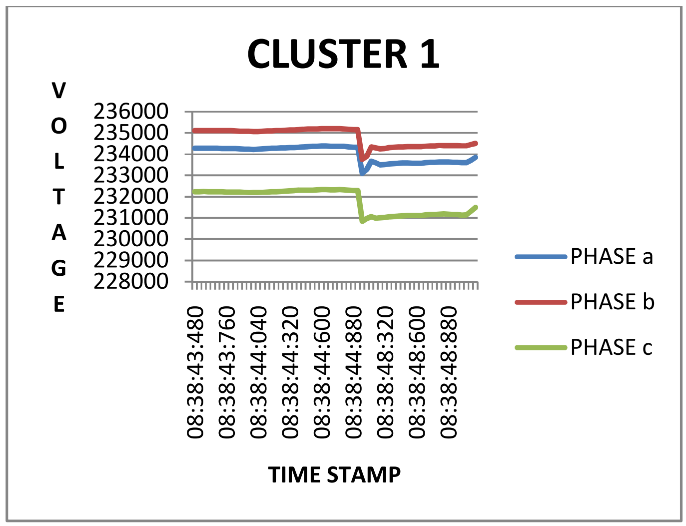

Fig. 3 Voltage for cluster 1

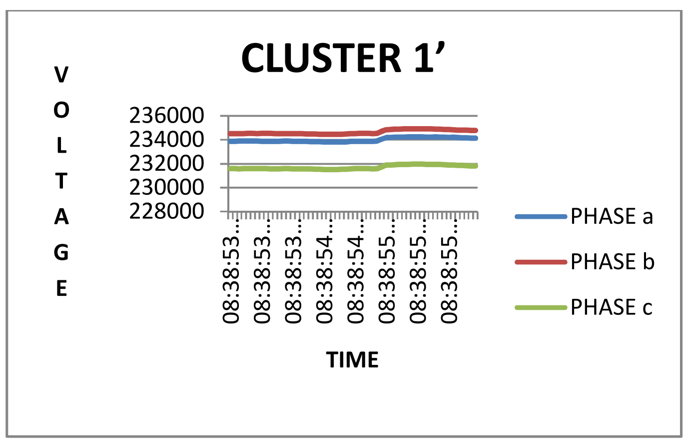

Fig. 4 Voltage for cluster 1'

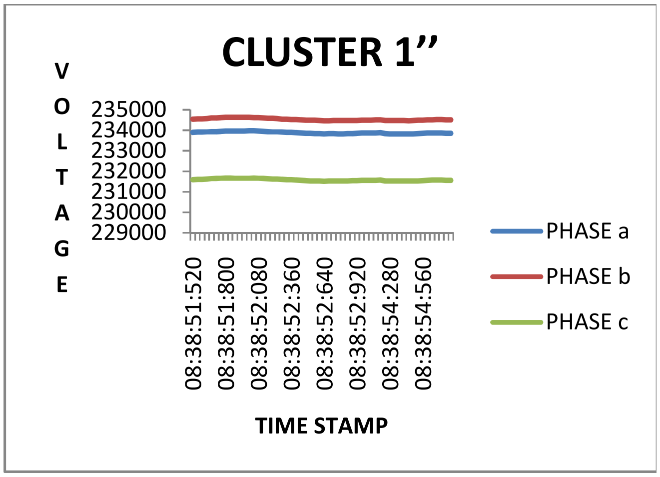

Fig.5 Voltage for cluster 1"

V. CONCLUSION

This paper describes the moving clusters scheme which can be implemented on real time data of PMUs without large storage of data to avoid storage problems. The data can be processed in real time and in case of fault, this can be detected quickly. Also, it is observed that the concept of hybrid window may be more useful as hybrid window consists of both the sliding window and the static window. Every time the operator need not decide the threshold limit and the control action. If the situation in sliding window matches with the situation already present in the history of static window, then same control action can be initiated without the loss of time and propagation of perturbations can be rejected almost within no time. This strategic concept may be refined as intelligent system healing. Further extension of the moving clusters can be done by operating the algorithms in parallel for different PMU data received by the phasor data concentrator. The deviation in any of the quantity (several operating in parallel at the same time) can be suitably detected and control action can be initiated immediately at specific location with WAMS control concept. The algorithm can be applied over the data for signals like rate of change of frequency, voltage, frequency, current, angle with an appropriate choice of clusters.

VI. ACKNOWLEDGMENT

The authors acknowledge with thanks for the field interactions and support given by managements of POSOCO as well as PGCIL and for permitting the publication of this paper. The authors are also thankful to WRLDC and NRLDC personnel for their support. The views expressed in this paper are of authors and not necessarily that of the organizations they represent.

VII. REFERENCES

[1] R. K. Pandey, P. Pentayya, A. Gartia, Rajkumar A. and Chandan Kumar "Operational Reliability Enhancement with PMUs in Indian Power Network", IEEE CATCON, Kolakata, 2013.

[2] "Synchorphasors Initiative in India" POSOCO, India, June 2012

[3] "New Technology can Improve Electric Power System Efficiency and Reliability", Today in Energy, U.S Energy Information Administration (eia), March 30, 2012.

[4] PMU Enhanced Power System Operations, Interactive Energy Roadmap to Achieve Energy Delivery Systems Cyber security, Sept 2011.

[5]Chakrabarti S, Kyriakides E, "Optimal Placement of Phasor Measurement Units for Power System Observability", IEEE Transactions on Power Systems, Vol.23, No. 3, pp1433-1440.

[6] A. O. M. Saleh and M. A. Laughton "Cluster analysis of power-system networks for array processing solutions," IEE Proceedings, vol. 132, no. 4, Jul. 1985.

[7] S. Madan, W.-K. Son, and K. E. Bollinger, "Applications of data mining for power systems," in Proceedings of IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), 1997.

[8] A. Asheibi, D. Stirling, and D. Robinson, "Identification of load power quality characteristics using data mining," in Proceedings of IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), Ottawa, pp. 157-162, May 2006.

[9] H. Mori, "State-of-the-art overview on data mining in power systems," in Proceedings of IEEE Power Systems Conference and Exposition (PSCE), 2006.

[10] M. McGranaghan, "Making connections: Asset management and the smart grid," IEEE Power and Energy Magazine, vol. 8, pp. 16-22, Nov. - Dec. 2010.

[11] P. Myrda, "Optimizing assets: Smart grid-enabled asset management," IEEE Power and Energy Magazine, vol. 8, pp. 109-112, Nov. - Dec. 2010.

[12] T. Gibson, A. Kulkarni, K. Kleese-Van Dam, and T. Critchlow, "The feasibility of moving PMU data in the future power grid," in Proceedings of Cigré Conference on Power Systems, Halifax, Sep. 6-8, 2011, Canada.

[13] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," ACM Computing Surveys (CSUR), vol. 31, no. 3, pp. 264-323, Sep. 1999.

[14] R. K. Pandey, S. M. Archana, Krati Nayyar, P. Pentayya, and Chandan Kumar "PMU Signal Prioritization for Effective Control Coordination of Load Despatch Centres" Proceedings of 2014 IEEE PES General meeting, Washigton DC.