A Real Time Event Detection, Classification and Localization using Synchrophasor Data

S Pandey, *Student Member, IEEE,* A. K. Srivastava, *Senior Member, IEEE,* and B. G. Amidan, *Senior Member, IEEE*

***Abstract*—With an increasing number of extreme events, grid components and complexity, more alarms are being observed in the power grid control centers. Operators in the control center need to monitor and analyze these alarms to take suitable control actions, if needed, to ensure the system's reliability, stability, security , and resiliency. Although existing alarm and event processing tools help in monitoring and decision making, synchrophasor data along with the topology and component location information can be used in detecting, classifying and locating the event, which is the focus of this work. Phasor Measurement Unit's (PMU's) data quality issue is also addressed before using data for event analysis. The developed algorithms include statistic, clustering, and Maximum Likelihood Criterion (MLE) based anomaly detection, Density-based spatial clustering of applications with noise (DBSCAN) for event detection and physics-based rule/ decision tree for event classification. Further, topology information, statistical techniques, and graph search algorithms are used for event localization. Developed algorithms have been validated with satisfactory results for IEEE 14 bus and 39 Bus as well as with real PMU data from the western US interconnection (WECC).**

***Index Terms*—PMU, Anomaly Detection, Event Detection, DBSCAN, MLE, Graph Theory.**

I. Introduction

THE power system can become more resilient and less susceptible to large outages if the operators have a wide area, real-time view of the system [1], [2]. The operators need to take control actions based on several alarms that appear in the Energy Management System (EMS). The motivation behind this study is to develop a tool using PMU measurements, which can help enhance the decision making capability of operators in the control room supplementing EMS [3]. In this work, goal is to detect, classify, and locate the physical power system events occurring frequently in the system. Example of these events include line faults, load changes, capacitor bank switching, and generator outages. Proposed algorithm has multiple steps: 1) detect anomalies, 2) classify anomalies into an active power event, reactive power event or fault events and 3) locate event. It is proposed to first detect and then classify events into an active power event, reactive power event or fault events. A generator drop is an example of active power event, whereas switching of a capacitor bank is a reactive power event. Short circuits are classified as faults. In an active power event, the active power flow between the buses changes. A reactive power event affects the reactive power flowing in the line and the voltage of the buses. In case of faults, the current dramatically increases and the voltage can drop down invariably. There will be large frequency variation and the rate of change of frequency increases to a very high value near the fault location [4]. The first step of the event detection algorithm is to detect these changes. Changes in the operating conditions are an indicator of a power system event.

There have been several studies about event detection using moving average value [5], linear principal component analysis based approach to analyze dimensional reduction of synchrophasor data [6], geographical visualization of PMU data [7], wavelet-based event detection [8], dynamic programming swinging door trending (SPSDT) [9], and [10] attempts to classify events using energy similarities and a temporal localization. Past work in literature has been reported to classify limited number of events as faults [11], cascading events [12], and cyber events [13]. Cyber event detection and classification has been studied in [14], [15] , however, in this work, focus is only to detect, classify and locate the physical power system events.

Most of the existing work related to event detection and classification methods do not classify events in a comprehensive and automated manner, and hence do not provide the exact location of events.

PMUs are known to have bad data present in the streaming data due to several reasons such as time synchronization error, communication problem sampling or simply due to error in field measurements [16], [17]. There can be data loss or offset in the measurements. Efforts have been made in the past to curb the effect of data anomalies in the PMU measurements [18], [19]. We have developed an ensemble technique that uses statistical and clustering algorithms as base detectors and unsupervised machine learning [20], [21]. Prony analysis [22] has been used to differentiate between bad data and event data to further enhance precision of the developed algorithm.

This paper is motivated to provide real-time event detection, classification, and localization with root causes for decision making by operators along with possible data anomalies in the PMU data. Processed data has been used for event detection. It is more than likely that multiple PMUs capture the same event in the system. For example, if a capacitor bank is switched on at a bus, it changes the steady-state voltage magnitude

S. Pandey, and A. K. Srivastava are with the School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA, 99163 USA. B. Amidan is with the Pacific Northwest National Lab. (e-mail: anurag.k.srivastava@wsu.edu)

The authors would also like to thank Pacific Northwest National Lab (PNNL) and Bonneville Power Administration (BPA) for providing the industrial PMU data and engineering support, which was necessary for validation of this work. The authors would like to acknowledge the National Science Foundation grant 1840192 for partially supporting this work. Thanks to Dr. Param Banarjee for simulated events and generating PMU data.

of that bus and can also change the voltages of neighboring buses. A PMU placed at the neighboring bus will also detect the capacitor bank switching. Based on this, we compute scores for each event seen by each PMUs. This helps in the formation of a sub-network in a large network and identifying the location of the event. The key contributions are itemized below:

- Synchrophasor anomaly detection (SyncAD) tool based on statistical and 'Maximum Likelihood Estimation (MLE)' technique aided with prony analysis to increase precision of anomaly detection.
- Event detection based on the data-driven 'cluster change' technique and classification into active power, reactive power and fault event types based on physics-based rule/ decision tree. In our understanding, this work is one of the first to integrate physics and data analytics for anomaly detection, event classification and localization.
- Real time event localization using graph theoretic approach and statistical score computation including Shannon entropy, standard deviation, range, mean difference, crest factor.

Developed methodology for detecting anomalies in the PMU data set are presented in section II and the proposed event detection, classification and localization technique in section III. Test cases are discussed in section IV and the simulation and results are provided in section V. Finally, conclusions are provided in section VI.

II. Synchrophasor Data Anomaly Detection

PMUs may have anomalies e.g. outliers or missing data. Outliers may look like certain measurements being off by say, 10 $\%$ of the actual measurement whereas, the missing data might just appear as '0'. Synchrophasor Anomaly Detection tool called SyncAD has been developed to address these as shown in figure 1. The tool uses Linear regression [23], Chebyshev [24], and DBSCAN [25] as statistical and clustering-based base detectors. These base detectors are independent and make decisions on a data point being an anomaly or good. Since they follow a different scale their outlier scores are normalized, followed by an ensemble-based learning algorithm. Once the model is learned the inference algorithm helps in comparing new data with the learned data and decide on anomaly data. Since the event data points can also be detected as anomalous points, we use prony analysis to detect the transient window and segregate event points form anomaly point to obtain bad data. Once the transient window is detected all the data points which were said to be bad, are unflagged and considered as a normal data point, if they lie in this transient window. Finally, depending upon the application the data can either be flagged as bad data or can be replaced by imputing as explained in [26].

A. Base Detectors

The base detectors used are Linear regression, Chebyshev, and DBSCAN method. The DBSCAN method is discussed later in this paper. The thresholds for DBSCAN, when used as a base detector, is different as the idea is not to detect event points as anomalies.

1) Linear regression-based detector: A regression line model is obtained for the data selected in the window (typically 1 second) by minimizing the sum of squared residuals (i.e. the vertical distance between data points of the window and the regression line). The regression line is represented as below:

$$\text{Regression Line} = \beta x + \alpha \tag{1}$$

where, β is the slope, α is the y-axis intercept of the regression line, and x is the closest point on the regression line from the actual data point. Based on this regression line, data points lying outside the low or high thresholds are considered as possible bad data. The high and low thresholds are set based on (2) and (3) below:

$$\text{High Threshold} = \beta x + \alpha + k * \text{dev} \tag{2}$$

$$\text{Low Threshold} = \beta x + \alpha - k * \text{dev} \tag{3}$$

where, parameter 'k' is the number of standard deviations, a preset number that decides the high and low thresholds and 'dev' is the root mean squared value of y-distance from the regression line as given by (4).

$$\text{dev} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\Big[x(i) - \text{regression}(i)\Big]^2} \tag{4}$$

where N is the number of points within the window, and $x(i)$ and $regression(i)$ are the actual and regression line data points respectively.

2) Chebyshev-based detector: This detector is often used when the distribution of the data set is unknown. It is a two-step process. A strict threshold is applied in the first step. Data points crossing this threshold are omitted for the next step. In the second step, the threshold is again computed but 'k' is larger than the first step. Now the omitted data points are checked to be within this relaxed threshold. If they still fail to lie within this wider threshold, it is considered to be an anomaly point. Chebyshev's inequality is shown in (5):

$$P(|X - \mu| \leq k\sigma) \geq (1 - \frac{1}{k^2}) \tag{5}$$

where 'X' represents the input PMU data, 'μ' is the mean of data within a window, 'σ' is the standard deviation of the data within the window, and 'k' represents the number of standard deviations from the mean.

B. Ensemble Method

The architecture of the bad data detection algorithm is shown in Fig-1, called as Synchrophasor Anomaly Detection (SyncAD). The base detectors make anomaly detection assessment independently on the PMU data and followed by the normalization of the scores using expectation maximization (EM) algorithm [27]. Once the outlier scores from the base detectors are normalized, it is then used to determine the MLE-ensemble model. The unsupervised learned MLE model is then

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2020.2986019, IEEE Transactions on Power Systems

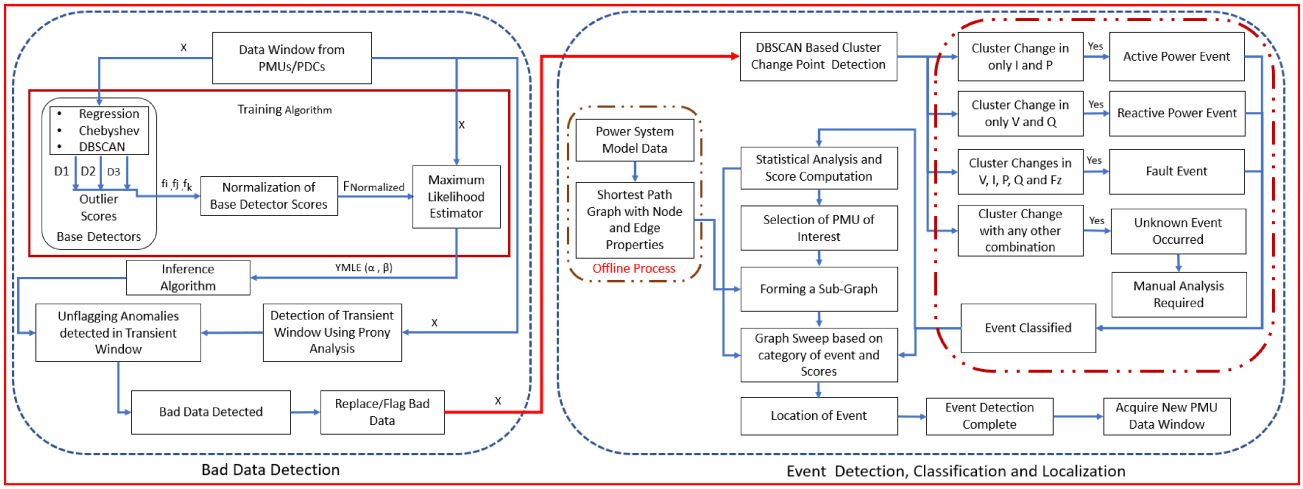

Fig. 1: Developed Data Flow Architecture for Anomaly Detection and Event Detection

fed to an inference algorithm which also takes as input the new set of normalized scores and makes an assessment on points to be an anomaly.

C. Prony-based transient window estimation

The Prony analysis [22] is used to determine the transient window, which is usually a result of an event in the system. Here the Prony analysis is used to determine the steady-state window. A steady-state window of voltage magnitude might have small oscillations due to PMU measurement uncertainty. These small oscillations are modeled as noise by the window selection filter. The window selection filter is designed by arranging the sampled values of the voltage magnitude measurements in the Hankel matrix Y as shown in (6). The method is rigorously tuned on simulated data as well as industry data (PMU) to obtain the transient window. A 2.5-second window (this size window selection provides a trade-off between speed and accuracy) of voltage measurement data is used, the total number of samples for a 2.5-second window having a PMU reporting rate of 120 frames per second is J=300.

$$Y = \begin{bmatrix} y(0) & y(1) & ... & y(\frac{J}{2}-1) \\ y(1) & y(2) & ... & y(\frac{J}{2}) \\ . & . & . & . \\ . & . & . & . \\ y(\frac{J}{2}-1) & y(\frac{J}{2}) & ... & y(J-1) \end{bmatrix} \tag{6}$$

The rank of the Hankel matrix where the elements of the matrix such as y(0), is the first element of the measurement window of 2.5 seconds, is estimated by the eigen-decomposition of the sample correlation matrix as follows:

$$QSQ' = YY' = R_{\mathrm{YY}} \tag{7}$$

where Q is an orthogonal matrix whose columns are the eigenvectors of R_{YY}; and S is the diagonal matrix, which has the singular values of the sample correlation matrix in descending order of magnitude, and can be expressed as:

$$S = diag(d_1 > d_2 > ... > d_{\mathrm{k}} > ... > d_{\frac{J}{2}}) \tag{8}$$

The logarithm of the singular values $d_2, d_3, \cdots, d_{\frac{J}{2}}$ are divided by the logarithm of the first singular value, which is denoted by:

$$\sigma_k = \frac{ln(d_k)}{ln(d_1)} \tag{9}$$

The p singular values from d_1 to d_p corresponds to the complex sinusoidal presented in the signal. The remaining $\frac{J}{2} - p$ singular values from d_{p+1} to $d_{\frac{J}{2}}$ corresponds to noise. The values of σ_k for $k = 2, 3, \cdots, p$ depend on the amplitude frequency and damping of each component.

The singular values of the matrix Y consist of signal singular values and noise singular values. For a perfect noiseless signal, the noise singular values are zero. Hence, even a DC signal will have one singular value with a finite real part and zero imaginary part. A signal with more variation will have many significant singular values. We have analyzed offline by randomly generating the matrix from continuous voltage measurements, whether the matrix is singular. We could never find a case when the determinant of the matrix was zero and the matrix to be singular for continuous PMU measurements. However, as an extra step, the algorithm could be updated to check for the singularity and if found to be singular the matrix could be formed by shifting the measurements by say 10 measurement points. This will not affect the performance of the algorithm as we are only interested in the dominant modes and we want to determine if the given window is a transient or a quasi-steady state window. Here we are only interested in the dominant modes (one or two modes) which are less than the dimension of the matrix.

A transient window has more frequency modes due to transients and oscillations. The threshold for σ_k is set as -4.27 by tuning it using the known events obtained from real-time digital simulator (RTDS) simulation and industry data for transmission PMUs but it works well for D-PMUs [28]. Data points lying in the transient window if flagged as bad data by the ensemble method are unflagged as normal data resulting in higher precision.

III. Event Detection, Classification and Localization

The operating point changes in voltages, currents, active power flow, reactive power flow, and frequency are detected. Once the events are classified the top PMUs are chosen based on statistical parameters and normalized scores. Using the top PMUs a sub-Graph is formed and a graph sweep algorithm is used to detect the location of the event. The event detection algorithm architecture is shown in Fig 1.

PMU data is first checked for data anomalies or missing data using the anomaly detection algorithm and the detected anomalies are then imputed using the average of the following and preceding data points. After the clean data is obtained, active and reactive power flows are computed using the PMU data given by equations 10 and 11, before feeding all the operating condition parameters to DBSCAN algorithm.

$$P_{ij} = 3 * |V^i_{LL}| * |I_{ij}| * cos(V_{angle} - I_{angle}) \quad (10)$$

$$Q_{ij} = 3 * |V^i_{LL}| * |I_{ij}| * sin(V_{angle} - I_{angle}) \quad (11)$$

Where,

P_{ij} and Q_{ij} are three phase active and reactive power injection (p.u) in line i & j

I_{ij} is the line i & j current (pu)

V^i_{LL} is the line to line voltage in (pu)

V_{angle} and I_{angle} are voltage and current angles.

A. Event Detection using DBSCAN

Voltage (V), current (I), active power flow (P), reactive power flow (Q) and frequency (Fz), from all the PMUs, are fed to the DBSCAN algorithm.

The DBSCAN algorithm [29] uses two parameters ϵ and the minimum number of points (MinPts). Data points lying within the ϵ radius of a cluster becomes the part of the Cluster-1. Once an event occurs, the operating points change and the next point in the time series data is out of reach of the first cluster as shown in Fig 2. As the next steady state is reached, a new cluster with a new operating point is formed. Our interest is to detect the boundary point instance of the first cluster, i.e. cluster change point.

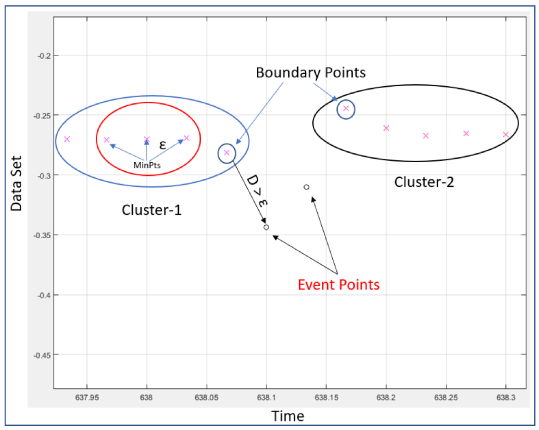

Fig. 2: DBSCAN

B. Event Classification

These instances are then fed to a physics based rule/ decision tree for event classification as explained below:

1) Active Power Event: If there are cluster change in current measurement and calculated active power, flow whereas no changes in cluster of voltage measurement and calculated reactive power.
2) Reactive Power Event: If there are cluster change in voltage measurement and calculated reactive power measurement whereas no changes in cluster of current measurement, calculated active power and frequency.
3) Fault Event: If cluster changes in voltage, current, active power, reactive power, and in frequency is observed, then it is a fault event.

This active, reactive and fault event classification is done as shown in Fig 1.

C. Event Localization

The location of PMU's placement plays an important role in better observability [30] and localization of events. When a power system event occurs it causes ripples across the system just like when a stone is thrown in calm water. The disturbance at the point of impact of stone is more, but, it dies out soon and at the other end of the lake, no such disturbance can be visibly detected. Similarly, the PMUs can detect the events if they are near the event location and in some cases, multiple PMUs can detect the events. Fault events have a larger impact and their signature can be detected in multiple PMUs, however an event such as a load change, transformer tab change are local events and their signature is not strong enough to be picked up by PMUs far in the electrical distance in the network. Therefore, we compute statistical parameters and get a combined score to decide the PMUs that have the strongest signal for a particular event. This method is useful in extracting information based on PMU data.

1) Scores of PMUs: Once the event point instance is detected we compute five statistical parameters that define the event signatures.

- Shannon Entropy: An event changes the quasi steady-state nature of the power system. The PMUs capture these events and they contain information regarding the magnitude and nature of the event. In information theory, Shannon entropy is defined as a metric that can capture the information [31]. A window of 30 samples is formed around the event instances taking 15 points before the event and 15 points after the event instance. Shannon Entropy is computed as per equation 12.

$$SE = -\sum_{i=1}^{N} X_i^2 \log(X_i^2) \quad (12)$$

where 'X' is the data stream and 'i' is the point in the window of length 'N'.

- Standard Deviation: One of the best measures that define the data distribution and in our case the magnitude of the disturbance caused by the event in PMU data. Standard Deviation is computed as per equation 13.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2020.2986019, IEEE Transactions on Power Systems

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2} \tag{13}$$

where 'X' is the data stream, 'i' is the point in the window of length 'N' and μ is the mean of data window 'N'.

- Range : This is an important measure that contains the variation in the PMU data during the event. It is computed using equation 14.

$$Range = |\max(X_i) - \min(X_i)| \tag{14}$$

where the data X ranges from 'i=1' to 'i=N' and 'N' is the window length.

- Mean difference: A mean difference is computed using data points which are before and after the event instance. Once the transient settles and steady-state is reached the mean difference of the two steady-state has information on the magnitude of the operating point shift. It is computed using equation 15

$$MD = |\mu_{Cluster1} - \mu_{Cluster2}| \tag{15}$$

where $Cluster\ 1$ is the data window before the event and $Cluster\ 2$ is the data window after the event with 15 samples each.

- Crest Factor: It is a measure that defines how severe is a peak during events. A PMU close to the event will have a higher crest factor than a PMU which is farther from the event location. Crest factor is computed using equation 16.

$$CF = \frac{|\max(X_i)|}{\frac{1}{N}\sum_{i=1}^{N}(X_i)^2} \tag{16}$$

where the data length is the same as used to compute all the other factors except MD.

The nearest PMU from event location in terms of electrical distance will see more prominent event signatures. The respective values of the above discussed statistical measures would be higher.

It is hard to compare the scores of each measure computed above across PMUs, so it is useful to normalize the score. Normalization helps in the identification of prominent PMUs for a particular event by comparing the single point scores. In our case, the normalized score is computed as per equation 17.

$$NS = (SE + CF) * (\sigma + Range + MD) \tag{17}$$

To understand better, the statistics based score comparison for a transformer tap change event that occurred at 253 seconds of the simulation are plotted in Fig 3. Voltages and reactive power flow for bus 6 and bus 13 are plotted on the primary axis and secondary axis respectively.

It can be seen that the nature of the graphs look similar but there operating points are different and so are the statistical factors. These factors are presented in Table I. The normalized score for this case points that bus 6 PMU has a higher score as compared to the PMU placed on bus 13. Therefore it can be concluded that the PMU on bus 6 is closer to the event as compared to bus 13 PMU.

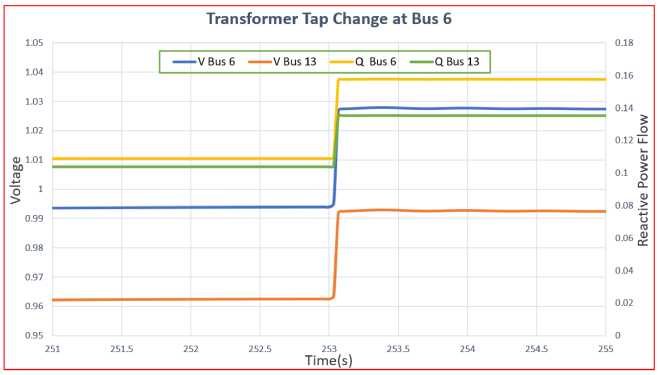

Fig. 3: Transformer Tap Change Event

TABLE I: SCORE COMPARISON FOR EVENT

Bus	SE	μ	Range	MD	CF	NS
6	1.15	0.042	0.08	0.083	2.16	0.68
13	1.15	0.031	0.061	0.062	2.12	0.51

2) Forming Sub Graph and Graph Sweep: The next step after event classification and normalized score computation is to select the top PMUs based on the scores. Certain events are detected by several PMUs as their system-level impact is higher. A fault event would be seen by multiple PMUs and a transformer tap change or a load change would be sensed locally by PMUs at those buses, at adjacent buses or PMUs placed close to the event bus in terms of electrical distance. We use Dijkstra's Algorithm [32] to find the shortest path and store it offline for a particular system. The process of the selection of a subgraph is explained in Algorithm 1. It has to be noted here that a subgraph formation can be an offline process. Based on the PMUs of interetst the subgraph can be chosen for graph sweep analysis and updated only when network changes.

IV. TEST-BED AND TEST CASES

A. Test-bed architecture

Fig. 4 shows the architecture of the test bed setup created using Real Time Digital Simulator (RTDS) [33] to test the algorithm. The setup is comprised of four layers. The first layer is the physical layer which is the first stage in the setup to model the power grid. The second layer is the sensor layer that includes all the measurement devices (PMUs). The third layer, real-time data archival layer, collects all the data from each substation and is located centrally in the control center. OpenPDC software has been used to archive the data for test cases in the lab. The final layer is the application layer where PMU data cleaning, and event detection algorithm is run. It delivers the output to an action layer which is not part of the test setup. However, during the field implementation the operator can have a better situational awareness or take control actions to operate the power system in a resilient and reliable manner.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2020.2986019, IEEE Transactions on Power Systems

Algorithm 1 Sub Graph and Graph Sweep

Input: $Shortest\ Path\ Graph: S_G;\ Event\ Type$
Output: Sub Graph, Event Location

Compute Normalized Score [eq.17] for each PMUs
Form Sub Graph
if Number of PMU > 1 **then**
  Select top 2 PMUs based on NS
  $Radius$= $Shortest\ Distance(PMU^{Bus1}, PMU^{Bus2})$
  $Sub\ Graph$= Buses from PMU^{Bus1}, PMU^{Bus2} within $Radius$
else
  Number of PMU $= 1$
  $Sub\ Graph$= $Adjacent\ Buses\ from\ PMU^{Bus1}$
end if
Graph Sweep
$Total\ Bus$= $Buses(Sub\ Graph)$
for each bus n $\in$ $Total\ Bus$ **do**
  if $Event\ Type = Reactive$ **then**
    $Location = Reactive\ Power\ Capability\ Bus$
  else if $Event\ Type = Active$ **then**
    $Location = Active\ Power\ Capability\ Bus$
  else if $Event\ Type = Fault$ **then**
    $Location = PMU^{Bus}$ $with\ highest\ NS$
  end if
  $Event\ Location = Location$
end for
return $Sub\ Graph, Event\ Location$

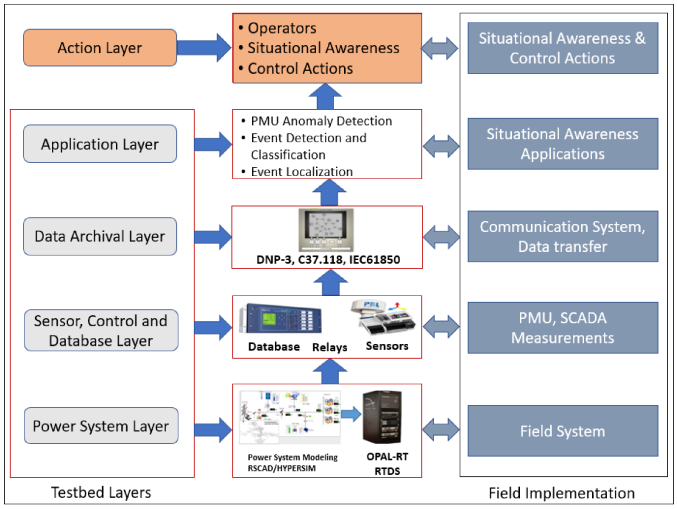

Fig. 4: Test-bed architecture and Field Implementation

B. Test Cases

IEEE 14-bus and IEEE 39-bus system were modeled in RTDS and different events were simulated. 5 PMUs were placed on different buses. The sampling rate of 30 samples/ sec was set for each PMUs. Three cases were modeled with different events and PMU placement as discussed in the test cases below:

1) Case 1: As seen in Fig 5 the IEEE 14 bus system, 5 PMUs on buses 2,6,8,9 and 10 were placed which monitored the currents flowing in lines 2-Gen, 6-11, 8-Load, 9-7 and 10-9 respectively.

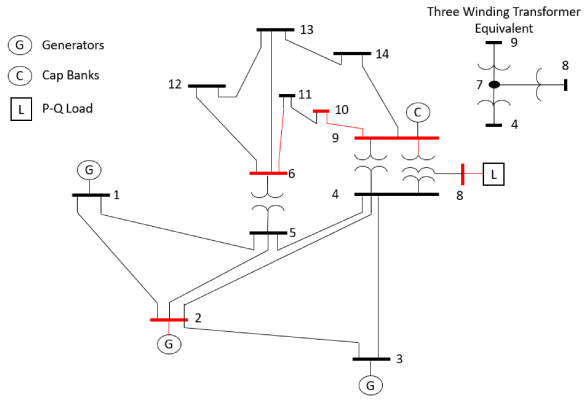

Fig. 5: Generating Test Cases using RTDS-Case1

A cap bank was placed on bus 9 which was operated at different time instances whereas a load was placed on bus 8 and the 3-phase fault was modeled on bus 10.

2) Case 2: The location of PMUs and events were changed for this case. In IEEE 14 bus system, PMUs were placed on buses 2, 6, 7, 9 and 13, which monitored the currents flowing in lines 2-Gen, 6-11, 7-4, 9-7 and 13-14 respectively. Capacitor banks were placed on bus 9 and 3, the fault was modeled on bus 13, the tap of the transformer between 6 and 5 was operated and the load was placed on bus 8.

3) Case 3: As seen in Fig. 6 the IEEE 39 bus system, 5 PMUs on buses 4,12,17,22 and 39 were placed which monitored the currents flowing in lines 4-3, 12-11, 17-16, 22-21 and 39-9 respectively.

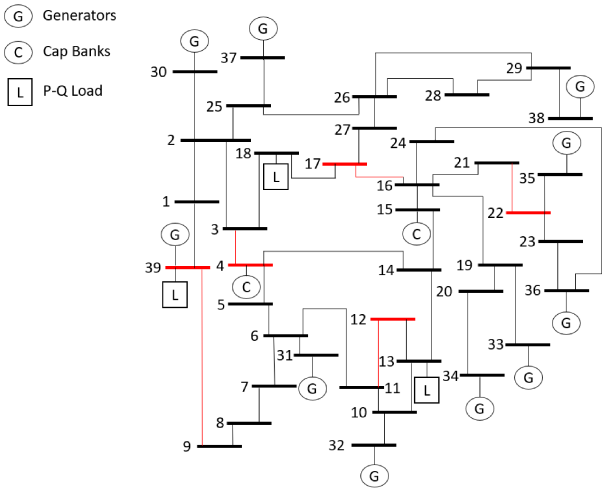

Fig. 6: Generating Test Cases using RTDS-Case3

Capacitor banks were placed on bus 4 and bus 15. Adjustable loads were placed on bus 13, 18 and 39. Faults were modeled on bus 8 and bus 19. The generators on Bus 38 and bus 32 were dropped during the simulation at different time instances.

4) Real Data from Industry Partners: A 10-month PMU data from different PMUs provided by Bonneville Power Administration (BPA) was analyzed. An event list of frequency

lists was recorded by BPA and the data were analyzed using the developed techniques at Pacific Northwest National Labs (PNNL). The Field Plot of a PMU data with calculated active and reactive power flow is shown in Fig 7. The topology and location of the PMU are not presented due to confidentiality.

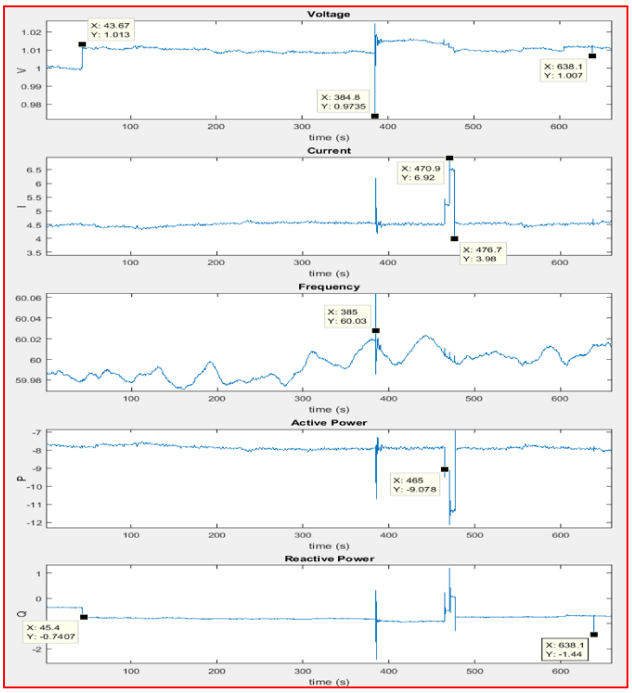

Fig. 7: Field Data Plot showing step changes in parameters

V. Simulation and Results

A. Anomaly Detection

Anomalies were manually inserted in the two test cases and the anomaly detection algorithm was applied. The results are presented in terms of precision and recall. Recall or true positive, given by equation 18, is the proportion of inserted bad data that were correctly identified by the algorithm. Precision is the proportion of inserted bad data that were correctly identified over the total number of times data were identified as bad data by the algorithm computed using equation 19.

$$Recall = \frac{Detected\ Bad\ Data \cap Actual\ Bad\ Data}{Actual\ Bad\ Data} \quad (18)$$

$$Precision = \frac{Detected\ Bad\ Data \cap Actual\ Bad\ Data}{Detected\ Bad\ Data} \quad (19)$$

First 5 values are for 5 PMUs of case-1 and next 5 are for 5 PMUs of case 2. In table II, it can be seen that the ability to detect the inserted bad data given by recall is above 98 % in most cases and to precisely detect it, i.e. without false detection is around 95 %.

In table III, a comparison of SyncAD vs other methods is presented. A case 3 data set for PMU on bus 4, with 10% bad data was analyzed for bad data by different methods. The recall for MLE Ensemble and SyncAD is similar and better than all the other methods, however, SyncAD has better precision compared to other methods as SyncAD uses Prony analysis to differentiate between event points and bad data points.

TABLE II: Performance of Anomaly Detector

S. No.	Precision	Recall
1	0.9560	0.9842
2	0.9478	0.9842
3	0.9436	0.9836
4	0.9468	0.9864
5	0.9688	0.9842
6	0.9280	0.9948
7	0.9764	0.9640
8	0.9834	0.9860
9	0.9356	0.9836
10	0.9815	0.9932

TABLE III: Comparision of SyncAD vs other anomaly detection methods

S. No.	Precision	Recall
Linear Regression [23]	0.8565	0.9021
DBSCAN [34]	0.8821	0.8821
Chebyshev [24]	0.8754	0.9154
K Means [35]	0.8554	0.9298
MLE Ensemble [21]	0.8913	0.9351
SyncAD	0.9648	0.9351

B. Detection and Characterization of Events

The simulation results for case 1 and case 2 can be seen from table IV and VI. It is clear from Table IV that all the events were detected successfully, except for an active load change on bus 8 at 188 seconds in case 1. The reason for no detection is that the load change was of a very low magnitude i.e. less than 0.05 pu. In case 2 the same load change at the instant of 188 sec was undetected and a cap bank operation on bus 3 at the 240 seconds was undetected. The cap bank operation was not detected by any of the PMU's because in the same instance there was a 3 phase fault on bus 13. Therefore the fault signature dominated the cap bank operation signature. In case 3, the number of buses increased whereas PMUs used were only 5. The PMUs were placed at buses which were closer to the modeled events to detect the signature of events. It detected all the events except the P load increase at 325 seconds. This event's signature was shadowed by the three-phase fault event.

C. Score Computation, Sub-graph and Graph Sweep

Normalized scores for PMUs that detected a particular event are computed and the top 2 PMUs are chosen and sub-graph is formed. If only 1 PMU detects the event the sub-graph consists of the PMU bus and adjacent buses, as shown in Fig 8. The sub-graphs can also be computed and stored offline to make the process fast. The formation of sub graphs can be changed according to the power system model.

Once the sub graph is formed and the type of event is known, a Graph Sweep algorithm is run which queries each bus for the capability that they possess. In case 1, at 109 s, PMUs at Bus 9 and 10 detected the events but the score of Bus

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2020.2986019, IEEE Transactions on Power Systems

TABLE IV: SIMULATION RESULT- CASE 1

S No.	Time (s)	Actual Event	Actual Location	Detection Bus		Classified Event	Normalized Score	
				PMU 1	PMU 2		PMU 1	PMU 2
1	109	Cap bank closed	Bus 9	9	10	Reactive Power Event	0.22	0.14
2	119	Three phase fault	Bus 10	10	9	Fault Event	37.25	23.54
3	132	Cap bank opened	Bus 9	9	10	Reactive Power Event	0.23	0.13
4	148	P load decreased	Bus 8	8	-	Active Power Event	2.11	-
5	158	Cap bank closed	Bus 9	9	10	Reactive Power Event	0.20	0.12
6	168	Three phase fault	Bus 10	10	9	Fault Event	48.06	26.86
8	179	Cap bank opened	Bus 9	9	10	Reactive Power Event	0.24	0.16
8	188	P load increased	Bus 8	-	-	No Detection	-	-
9	198	Q load increased	Bus 8	8	9	Reactive Power Event	0.90	0.36
10	209	P load decreased	Bus 8	8	9	Active Power Event	2.07	0.72
11	219	Q load decreased	Bus 8	8	9	Reactive Power Event	1.15	0.62
12	229	Gen drop	Bus 2	2	-	Active Power Event	2.73	-

TABLE V: SIMULATION RESULT- CASE 2

S No.	Time (s)	Actual Event	Actual Location	Detection Bus		Classified Event	Normalized Score	
				PMU 1	PMU 2		PMU 1	PMU 2
1	109	Cap bank closed	Bus 9	9	7	Reactive Power Event	2.15	2.11
2	119	Three phase fault	Bus 13	13	9	Fault Event	26.56	11.09
3	132	Cap bank opened	Bus 9	9	7	Reactive Power Event	1.89	1.84
4	148	P load decreased	Bus 8	7	-	Active Power Event	2.61	-
5	158	Cap bank closed	Bus 9	9	7	Reactive Power Event	2.26	2.25
6	168	Three Phase fault	Bus 13	13	9	Fault Event	40.32	12.92
7	179	Cap bank opened	Bus 9	9	7	Reactive Power Event	1.58	1.57
8	188	P load increased	Bus 8	-	-	No Detection	-	-
9	198	Q load increased	Bus 8	7	9	Reactive Power Event	0.74	0.67
10	209	P load decreased	Bus 8	7	9	Active Power Event	2.71	0.68
11	219	Q load decreased	Bus 8	7	9	Reactive Power Event	0.85	0.79
12	229	Tap down	Bus 6	6	-	Reactive Power Event	0.62	-
13	240	Three phase fault	Bus 13	13	9	Fault Event	20.72	15.37
14	240	Cap bank closed	Bus 3	-	-	No Detection	-	-
15	253	Tap up	Bus 6	6	13	Reactive Power Event	0.68	0.51
16	263	Gen drop	Bus 2	2	-	Active Power Event	2.06	-

TABLE VI: SIMULATION RESULT- CASE 3

S No.	Time (s)	Actual Event	Actual Location	Detection Bus		Classified Event	Normalized Score	
				PMU 1	PMU 2		PMU 1	PMU 2
1	50	P load decrease	Bus 18	17	4	Active Power Event	0.63	0.52
2	80	Cap bank closed	Bus 4	4	17	Reactive Power Event	2.34	2.07
3	138	Q load decrease	Bus 13	12	-	Reactive Power Event	0.47	-
4	170	P load decreased	Bus 39	39	-	Active Power Event	1.53	-
5	200	Three phase fault	Bus 8	39	4	Fault Event	39.32	24.86
6	235	Gen drop	Bus 32	12	-	Active Power Event	3.08	-
7	260	Cap bank Closed	Bus 15	17	22	Reactive Power Event	1.68	1.43
8	270	P load increased	Bus 18	17	4	Active Power Event	0.86	0.73
9	300	Q load increased	Bus 13	12	-	Reactive Power Event	0.74	-
10	325	Three phase fault	Bus 19	17	22	Fault Event	34.59	29.65
11	325	P load increase	Bus 18	-	-	No Detection	-	-
12	330	Cap bank opened	Bus 4	4	17	Reactive Power Event	2.24	1.93

TABLE VII: EVENT DETECTION AND ANOMALY DETECTION ALGORITHM TIMELINES

Case	PMUs	Data Length(s)	Anomaly Detect.(s)	Event Detect.(s)	Event Localiz.(s)	Total Time(s)	Total Time per Window(s)
Case-1	5	269	2.10	20.01	3.01	25.13	0.09
Case-2	5	300	2.42	22.65	3.2	28.27	0.09
Case-3	5	350	2.68	23.89	5.4	31.97	0.09
Industry	5	272	4.01	16.97	-	20.98	0.07

9 was higher and only Bus 9 has the capability of causing a reactive power event. Therefore, it can be easily concluded that the cap bank at bus 9 was operated. In case 3, the generator drop event at 235 seconds was detected and classified as an active power event, but it could not be located. Bus 32 was not adjacent to bus 12 and it did not lie in the current sub-graph formation scheme, that was formed around bus 12. The solution for this case would be to either increase the number of PMUs so that more bus is covered by a PMU adjacent to it or the sub-graphs be expanded to more than only one adjacent bus. In this scenario, the generator drop event could have been located if 3 adjacent buses were chosen to form a sub-graph.

The computational cost for graph sweep with P vertices and complexity K measured as the number of segments of the embedding, the running time of the algorithm is $\Theta(K+PM)$, where M is the maximum number of edges cut by any vertical line. The DBSCAN algorithm has a computation complexity of $\Theta(\log(N))$, linear regression and Chebyshev have $\Theta(2(N+1))$, where N is the number of points. As shown in Table VII, the algorithm requires less time to run as compared to the data length time using an Intel i7 computer. The total run-time for the algorithm is approximately $1/10$ of data length. It must be noted that the 5 PMUs were chosen for the analysis and no parallelism was used. The average time taken for the algorithm to run on a typical window size of 30 samples data, i.e. 1 second data window for 5 PMUs is around 0.09 seconds. There are limited number of existing works addressing all three aspects of anomaly detection, event classification and localization. The work in [36], attempts to determine the location of events in real time, however the computational complexity and time taken for the algorithm to run is not discussed in the paper.

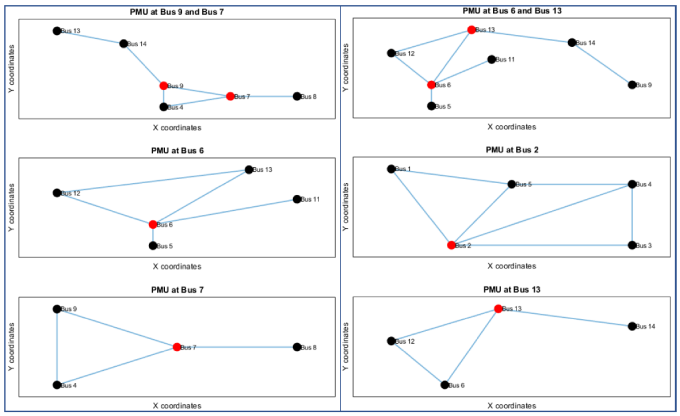

Fig. 8: Sub Graphs

D. Event Detection using Industry Data

From Fig 9, a field voltage plot, it can be seen that there are some missing data and outliers. The data that seem to be outliers are missing data. The graph is plotted on this scale to compare it with the clean data plot obtained by the anomaly detection algorithm. The clean data plot is shifted in axis and is plotted on secondary axis.

This proves the necessity of PMU data cleaning before using it for any further processing. Event Detection algorithm was applied and it was observed that the frequency events were

TABLE VIII: FIELD DATA RESULT

S No.	Time (s)	Active Event (PMU No.)	Reactive Event (PMU No.)	Fault Event (PMU No.)
1	44	-	5	-
2	111	-	1	-
3	385	-	-	1,2,3,4,5
4	465	2,5	-	-
5	471	2,5	-	-
6	477	2,5	-	-
7	558	-	3	-
8	638	-	2,5	-

detected as active events or fault events. several reactive events were detected but these events were not labeled by BPA even with engineering analysis. It can be seen from Table VIII and Fig 7 that the detection results corresponds to the step changes in the parameters.

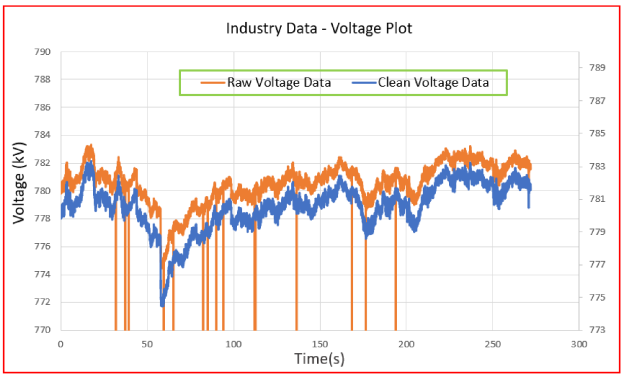

Fig. 9: Voltage Plot of Field Data

The topology for the industrial system is not presented due to confidentiality therefore, the exact location and cause of the events were not determined and presented.

VI. CONCLUSIONS

New computationally efficient algorithms have been developed for anomaly and event detection, classification and localization using PMU data and utilizing a suite of mathematical and statistical techniques include Maximum Likelihood Estimation (MLE), DBSCAN based cluster change, physics-based rule and decision tree, Shannon entropy, crest factor and graph theory. The simulation results prove the effectiveness of the algorithm to detect, classify and locate the events. Results on industrial data prove the necessity of anomaly detection along with the effectiveness of the event detection algorithm. The event detection tool can prove to be a very effective real-time PMU based enhancement for the power system operators and efficient decision making. The tool will be most effective if it can be integrated with the existing EMS and synchrophasor analysis tools so that the topology and PMU location information is available. Future work will be including machine learning-based detectors and analysis for faster results and real-time implementation tool development.

REFERENCES

[1] A. M. Shotorbani, S. Madadi, and B. Mohammadi-Ivatloo, "Wide-area measurement, monitoring and control: PMU-based distributed wide-area damping control design based on heuristic optimisation using digsilent powerfactory," in *Advanced Smart Grid Functionalities Based on PowerFactory*, pp. 211–240, Springer, 2018.

[2] M. U. Usman and M. O. Faruque, "Applications of synchrophasor technologies in power systems," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 2, pp. 211–226, 2019.

[3] S. Pandey, N. Patari, and A. K. Srivastava, "Cognitive flexibility of power grid operator and decision making in extreme events," in *2019 IEEE Power Energy Society General Meeting (PESGM)*, pp. 1–5, 2019.

[4] S. Pandey, "A real time synchrophasor data-driven approach for event detection in the power grid," Master's thesis, Washington State University, 2017.

[5] J. H. Chow, A. Chakrabortty, M. Arcak, B. Bhargava, and A. Salazar, "Synchronized phasor data based energy function analysis of dominant power transfer paths in large power systems," *IEEE Transactions on Power Systems*, vol. 22, no. 2, pp. 727–734, 2007.

[6] L. Xie, Y. Chen, and P. R. Kumar, "Dimensionality reduction of synchrophasor data for early event detection: Linearized analysis," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 2784–2794, 2014.

[7] A. Kaci, I. Kamwa, L.-A. Dessaint, and S. Guillon, "Synchrophasor data baselining and mining for online monitoring of dynamic security limits," *IEEE transactions on power systems*, vol. 29, no. 6, pp. 2681–2695, 2014.

[8] D.-I. Kim, T. Y. Chun, S.-H. Yoon, G. Lee, and Y.-J. Shin, "Wavelet-based event detection method using pmu data," *IEEE Transactions on Smart grid*, vol. 8, no. 3, pp. 1154–1162, 2015.

[9] M. Cui, J. Wang, J. Tan, A. R. Florita, and Y. Zhang, "A novel event detection method using pmu data with high precision," *IEEE Transactions on Power Systems*, vol. 34, pp. 454–466, Jan 2019.

[10] R. Yadav, A. K. Pradhan, and I. Kamwa, "Real-time multiple event detection and classification in power system using signal energy transformations," *IEEE Transactions on Industrial Informatics*, vol. 15, pp. 1521–1531, March 2019.

[11] S. W. X. Z. D. Nguyen, R. Barella and X. Liang, "Smart grid line event classification using supervised learning over pmu data streams," in *2015 Sixth International Green and Sustainable Computing Conference (IGSC)*, pp. 1–8, IEEE, 2015.

[12] D. L. a. S. M. M. Rafferty, X. Liu, "Real-time multiple event detection and classification using moving window pca," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2537–2548, 2016.

[13] T. M. S. Pan and U. Adhikari, "Classification of disturbances and cyber-attacks in power systems using heterogeneous time-synchronized data," *IEEE Transactions on Industrial Informatics*, vol. 11, pp. 650–662, June 2015.

[14] A. Giani, E. Bitar, M. Garcia, M. McQueen, P. Khargonekar, and K. Poolla, "Smart grid data integrity attacks," *IEEE Transactions on Smart Grid*, vol. 4, no. 3, pp. 1244–1253, 2013.

[15] M. Liao and A. C. , "Optimization algorithms for catching data manipulators in power system estimation loops," *IEEE Transactions on Control Systems Technology*, no. 99, pp. 1–16, 2018.

[16] S. G. Ghiocel, J. H. Chow, G. Stefopoulos, B. Fardanesh, D. Maragal, B. Blanchard, M. Razanousky, and D. B. Bertagnolli, "Phasor-measurement-based state estimation for synchrophasor data quality improvement and power transfer interface monitoring," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 881–888, 2013.

[17] P. NASPI, "PMU data quality: A framework for the attributes of pmu data quality and quality impacts to synchrophasor applications," 2017.

[18] X. Bian, X. R. Li, H. Chen, D. Gan, and J. Qiu, "Joint estimation of state and parameter with synchrophasors—part i: State tracking," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1196–1208, 2011.

[19] B. Gou and R. G. Kavasseri, "Unified PMU placement for observability and bad data detection in state estimation," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 2573–2580, 2014.

[20] P. Banerjee, S. Pandey, M. Zhou, A. Srivastava, and A. Wu, "Data Mining Based Anomaly Detection In PMU Measurements and Event Detection," in *North American Synchrophasor Initiative (NASPI), Gaithersburg, MD*, pp. 1–24, May 2017.

[21] M. Zhou, Y. Wang, A. K. Srivastava, Y. Wu, and P. Banerjee, "Ensemble based algorithm for synchrophasor data anomaly detection," *IEEE Transactions on Smart Grid*, pp. 1–1, 2018.

[22] J. F. Hauer, C. J. Demeure, and L. L. Scharf, "Initial results in prony analysis of power system response signals," *IEEE Transactions on Power Systems*, vol. 5, pp. 80–89, Feb 1990.

[23] P. Sedgwick, "Simple linear regression," *BMJ*, vol. 346, p. f2340, 2013.

[24] B. G. Amidan, T. A. Ferryman, and S. K. Cooley, "Data outlier detection using the chebyshev theorem," in *Aerospace Conference, 2005 IEEE*, pp. 3814–3819, IEEE, 2005.

[25] E. M. et al., "A density-based algorithm for discovering clusters in large spatial databases with noise.," in *Kdd*, vol. 96, pp. 226–231, 1996.

[26] A. Gholami, A. K. Srivastava, and S. Pandey, "Data-driven failure diagnosis in transmission protection system with multiple events and data anomalies," *Journal of Modern Power Systems and Clean Energy*, pp. 1–12, 2019.

[27] J. Gao and P.-N. Tan, "Converting output scores from outlier detection algorithms into probability estimates," in *Data Mining, 2006. ICDM'06. Sixth International Conference on*, pp. 212–221, IEEE, 2006.

[28] S. Pandey, A. K. Srivastava, P. Markham, M. Patel, *et al.*, "Online estimation of steady-state load models considering data anomalies," *IEEE Transactions on Industry Applications*, vol. 54, no. 1, pp. 712–721, 2017.

[29] S. Pandey, S. Chanda, A. Srivastava, and R. Hovsapian, "Resiliency-driven proactive distribution system reconfiguration with synchrophasor data," *IEEE Transactions on Power Systems*, pp. 1–1, 2020.

[30] N. M. Manousakis and G. N. Korres, "Optimal pmu placement for numerical observability considering fixed channel capacity—a semidefinite programming approach," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 3328–3329, 2015.

[31] S. Guo, M. Yan, and Y. Yang, "Processing combat information with shannon entropy and improved genetic algorithm," in *2015 8th International Conference on Biomedical Engineering and Informatics (BMEI)*, pp. 857–861, Oct 2015.

[32] Z. Fuhao and L. Jiping, "An algorithm of shortest path based on dijkstra for huge data," in *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, vol. 4, pp. 244–247, Aug 2009.

[33] "Real Time Digital Simulator (RTDS)." http://www.rtds.com.

[34] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise.," in *Kdd*, vol. 96, pp. 226–231, 1996.

[35] N. Grira, M. Crucianu, and N. Boujemaa, "Unsupervised and semi-supervised clustering: a brief survey," *A review of machine learning techniques for processing multimedia content*, vol. 1, pp. 9–16, 2004.

[36] D. Kim, A. White, and Y. Shin, "PMU-based event localization technique for wide-area power system," *IEEE Transactions on Power Systems*, vol. 33, pp. 5875–5883, Nov 2018.

Shikhar Pandey completed his MS degree (Electrical Engineering) in 2017 and is pursing his Ph.D. in Electrical Engineering at Washington State University Pullman. His Research interest includes synchrophasor technology, their measurement, data quality issues, event detection and synchrophasor applications. He graduated with undergraduate degree in Electrical Engineering from National Institute of Technology Patna in 2013 and worked as Sr. Electrical Engineer(2013-2015) at Larsen and Toubro ECC, Kullu, H.P. India.

Anurag K. Srivastava is an associate professor of electric power engineering at the Washington State University and the director of the Smart Grid Demonstration and Research Investigation Lab (SGDRIL) within the Energy System Innovation Center (ESIC). He received his Ph.D. degree in electrical engineering from the Illinois Institute of Technology in 2005. His research interests include data-driven algorithm for the power system operation and control. Dr. Srivastava is an editor of the IEEE Transactions on Smart Grid and IEEE Transactions on Power Systems. He is an IEEE distinguished lecturer, and the co-author of more than 300 technical publications.

Brett G. Amidan worked for Pacific Northwest National Laboratory for over 20 years as a statistician and data scientist. His research included big data analytics as applied to the power grid and aircraft safety. He was also involved in bio-threat sampling designs and analyses. He is currently a professor in the Mathematics department at Brigham Young University Idaho.