



Data-driven failure diagnosis in transmission protection system with multiple events and data anomalies



Amir GHOLAMI¹, Anurag K. SRIVASTAVA¹, Shikhar PANDEY¹

Abstract To guarantee the reliable power supply, the expected operation of all the components in the power system is critical. Distance protection system is primarily responsible of isolating the faulty section from the healthy part of the grid. Failure in protection devices can result in multiple conflicting alarms at the power grid operation center and complex events analysis to manually find the root cause of the observed system state. If not handled in time, it may lead to the propagation of the faults/failures to the adjacent transmission lines and components. With availability of the synchronized measurements from phasor measurement units (PMUs), real-time system monitoring and automated failure diagnosis is feasible. With multiple adverse events and possible data anomalies, the complexity of the problem will be escalated. In this paper, a PMU based algorithm is presented and discussed to detect the root cause of the failure in transmission protection system based on the observed state, e.g. multiple line tripping, breaker failures. The failure diagnosis algorithm is further enhanced to come up with the fully functional version of the failure diagnosis tool, which is tailored for the cases in which the PMU anomalies are present. In the developed

algorithm the validity of the PMU data is critical; however, such causes as communication errors or cyber-attacks might lead to the PMU data anomalies. This issue is well-addressed in this paper and some major types of anomaly detection methods suitable for PMU data are discussed. Results show that the ensemble approach has some distinct advantages in data anomaly detection compared to the previously used standalone algorithms. Additionally, the enhanced failure diagnosis method is developed to clean the inaccurate data in case of the anomaly in measured voltage magnitudes. Finally, both original and enhanced versions of the tool are tested on 96-bus test system using the real-time OPAL-RT simulator. The results show the accuracy of the enhanced tool and its advantages over the primary version of the tool.

Keywords Failure diagnosis, Transmission protection system, Protection mis-operation, Phasor measurement unit (PMU) data anomaly and cleaning, Ensemble method

1 Introduction

In order to have a reliable delivery of electricity, it is necessary to continuously monitor the operation of the protection system in the generation, transmission, and distribution [1]. In the transmission system, distance protection relays and associated breakers are responsible for detecting the fault as well as disconnecting the faulty part from the rest of the grid [2]. This is generally done by setting hard thresholds based on the local measurements, which in turn, will trigger the relays and breakers to operate in their corresponding zones of function [3, 4]. However, any malfunction or failure in protection devices can escalate the effects of the fault to endanger the

CrossCheck date: 19 March 2019

Received: 16 November 2018 / Accepted: 19 March 2019 / Published online: 10 June 2019

© The Author(s) 2019

✉ Anurag K. SRIVASTAVA
anurag.k.srivastava@wsu.edu

Amir GHOLAMI
Amir.gholami@wsu.edu

Shikhar PANDEY
Shikhar.pandey@wsu.edu

¹ Washington State University, Pullman, USA



operation of the grid [5]. By the general definition, distance protection mis-operation can be defined as any type of operation which is not the normal expected operation of the protection device at the time of contingency [6], including failure of any protection devices to timely operate, any operation out of the designed protection zone, and unintentional operation when there is no fault [7]. Some of the established fault analysis tools, mentioned in [8–10], are primarily using either line impedances or frequency analysis of the transients to detect the location of the fault.

The possibility of the malfunctioning in protection system is addressed in [11] and a vulnerability study of protection system is conducted in [12]. However, different conflicting data and alarms in control center, which are primarily triggered by the fault and aggravated by the protection mis-operation, motivate the need for an algorithmic outage identification tool to automatically pinpoint the location of the fault and the malfunctioned devices [13]. Authors of [13] have proposed a systematic 5-digit algorithm, in which a 5-digit number will be assigned to each phasor measurement unit (PMU) data and different possible scenarios will be extracted and investigated. Furthermore, post-fault system status will be combined by the acquired PMU data and the credibility values will be assigned to each hypothesis. Finally, based on the generated credibility values, the scenario with the maximum credibility will be determined as the actual event. In this work, the authors have developed the next version of failure diagnosis tool primarily reported in [13]. However, [13] has the underlying assumption of fully accurate PMU data, which might not be practical in many situations [14, 15]. The preliminary tool has been further developed to a fully functional level to work in real time with inaccuracy in measurements. Measurement anomalies can be triggered by such several reasons as the communication failures and/or cyber-attacks, and the validity of the PMU data can be endangered accordingly [16]. Such several methods are developed in [17, 18] to detect the anomaly in PMU time series data. Reference [19] has considered the bad PMU data as the outliers and has developed an outlier detection technique to detect the presence of bad data. In terms of the accuracy and the needed manual effort, there is no major difference among different types of standalone outlier detection methods, as they all need great amount of manual work to adjust the parameters to achieve an acceptable accuracy. Moreover, due to the response time and quality restrictions, not all the outlier detection methods can be applied to the real-life PMU data [20]. State estimation based PMU bad data detection has also been widely researched. The major drawback of the state estimation methods is that they require the topology information as well as strategic placement of multiple PMUs in the system. The results presented in [21], has an average

recall of 0.95 and false detection of 0.07 whereas the average recall of our algorithm is about 0.97 and the false detection is about 0.01. The aforementioned challenges inspire the need for a more efficient and faster PMU data anomaly detection method which is developed in [20] and named as “ensemble based algorithm”. The algorithm requires less amount of manual work for parameter tuning and at the same time it keeps the accuracy at a reasonable level. After the bad data is detected, the next step is to either remove or recover the inaccurate measurements. Finally, by combining the anomaly detection and failure diagnosis, the more suitable failure diagnosis tool can be developed.

In this paper, we will first elaborate upon the ensemble method for PMU anomaly detection. Further on, Prony-based transient window estimation will be detailed and a developed methodology for invalid data recovering will be discussed. In the next step, a centralized 5-digit algorithm, proposed in [13] as a systematical tool for transmission line outage, location identification, will be reviewed. In the rest of the paper, the discussed failure diagnosis tool is extended and the presence of PMU invalid measurements is also taken into account.

Contributions of this research work beyond our previous work are:

- 1) Further enhancing the 5-digit algorithm to address the PMU bad data.
- 2) Further enhancing the Prony scheme to compensate for the inaccurate measurements and developing a procedure to clean and mitigate the detected bad data.
- 3) Exploiting the enhanced bad data processing algorithm and the failure diagnosis tool to create a complete framework for transmission line failure diagnosis tool with the bad data and PMU anomalies.
- 4) Developing real-time test cases and validating the performance of the original fault diagnosis tool in real time with the assumption of fully accurate input measurements.
- 5) Conducting a comparison between the performance of the original failure diagnosis tool and the enhanced version of the tool on 96-bus transmission system using the real-time OPAL-RT simulator and validating the performance of the extended tool.

2 PMU anomaly detection and cleaning

Figure 1 shows the general outline of the failure diagnosis with bad data which is developed in this paper, where CB represents circuit breaker.

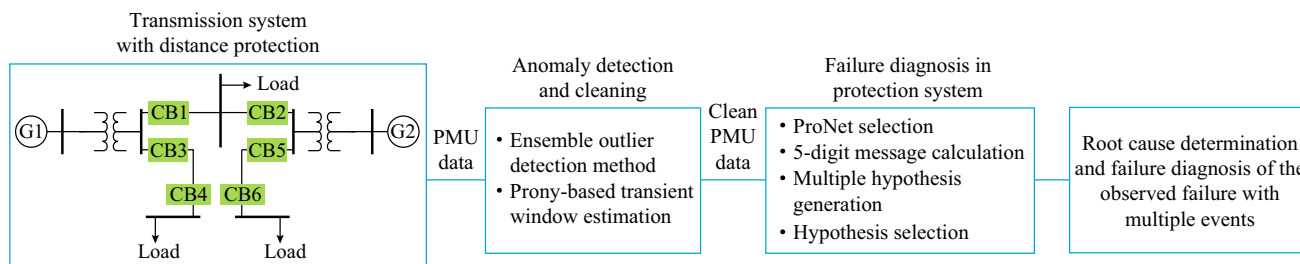


Fig. 1 Proposed architecture for failure diagnosis in distance protection system with data anomalies

The synchrophasor anomaly detection (SyncAD) tool developed by our research group for PMUs [10], has been designed to analyze the anomalies in PMU data at different reporting rates. Users can browse and input PMU data in an Excel or comma-separated value (CSV) format. The tool will recognize the PMU ID and users can choose a PMU for anomaly detection analysis. Users can choose to flag the detected anomaly instances or can choose an option to clean the anomalies. Anomaly flagging will simply create an extra column in the output file with a flag associated to the anomaly instances. Clean data anomalies option will impute the anomalies by taking the average of data points which lies before and after the detected anomaly point. This tool provides the option to view the graph of original data, anomaly plot and clean data plot. Finally, an output file can be saved to be used for different tools. The SyncAD algorithm first uses three base detectors for the anomaly detection individually. The scores of these base detectors are then used by the maximum likelihood estimator (MLE) ensemble model to get a final score and bad data assessment. Further, the Prony analysis is used to determine if the bad data is a misinterpretation of transients in the system.

2.1 Base detectors

The base detectors used are based on linear regression, Chebyshev, and density-based spatial clustering of applications with noise (DBSCAN) methods.

- 1) Linear regression-based detector: a window size of one second is chosen based on experiments which provides a balance between computational expenditure and accuracy. From the selected window the regression line model is obtained by minimizing the sum of squared residuals, i.e. the vertical distance between data points of the window and the regression line [22]. The regression line f_r is represented as below:

$$f_r = \beta x + \alpha \tag{1}$$

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, any data point lying outside the low or high thresholds can be considered as possible bad data. The high and low thresholds D_h and D_l are set based on (2) and (3).

$$D_h = \beta x + \alpha + kV_{dev} \tag{2}$$

$$D_l = \beta x + \alpha - kV_{dev} \tag{3}$$

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

$$V_{dev} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x(i) - x_r(i))^2} \tag{4}$$

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

- 2) Chebyshev-based detector: this detector is a two-step process [23] and is often used when the distribution of the data set is unknown. In the first step, a strict threshold is usually set and the detected data points are omitted for the next step. In the second step, the threshold is again computed but k is larger than the first step. If the data point still fails to lie within this wider threshold, it is considered to be an anomaly. Chebyshev 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; and σ is the standard deviation of the data within the window.

- 3) DBSCAN-based clustering: DBSCAN is applied to detect outliers and missing data. The DBSCAN algorithm [24] uses two parameters ϵ and the minimum number of points. Data points lying within the ϵ radius of a cluster becomes the part of the cluster. Data points which are outside the reach of cluster points are considered as anomalies as shown in Fig. 2.

Once the existing cluster cannot be expanded, the algorithm forms new clusters.



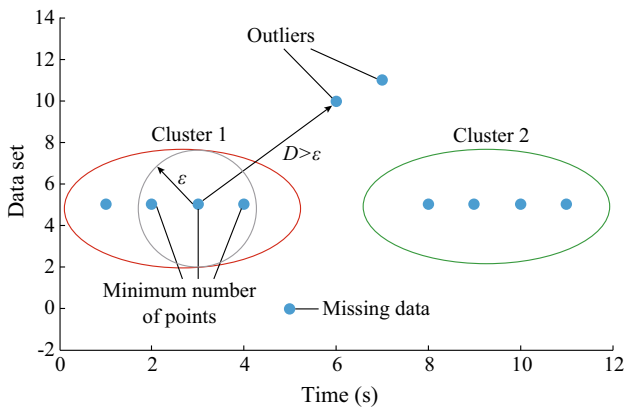


Fig. 2 DBSCAN

2.2 Ensemble method

An architectural schematic of the anomaly detection algorithm is shown in Fig. 3, where PDC represents phasor data concentrator. The base detectors find anomalies independently on the PMU data (D_1, D_2 and D_3), followed by the normalization of the scores (f_1, f_2 and f_3) using expectation maximization (EM) algorithm [25]. Once the outlier scores from the base detectors are normalized ($F_{normalized}$), they are then integrated using the MLE ensemble model. The output from unsupervised MLE ensemble model ($Y_{MLE}(\alpha, \beta)$) is then fed to an inference algorithm for anomaly assessment. It was observed that almost all the inserted bad data were detected as data anomalies by the algorithm correctly. However, a few transient data points were also flagged as data anomalies leading to low precision. To overcome this issue Prony analysis is used. The details of the ensemble based anomaly detection is discussed with equations in [20].

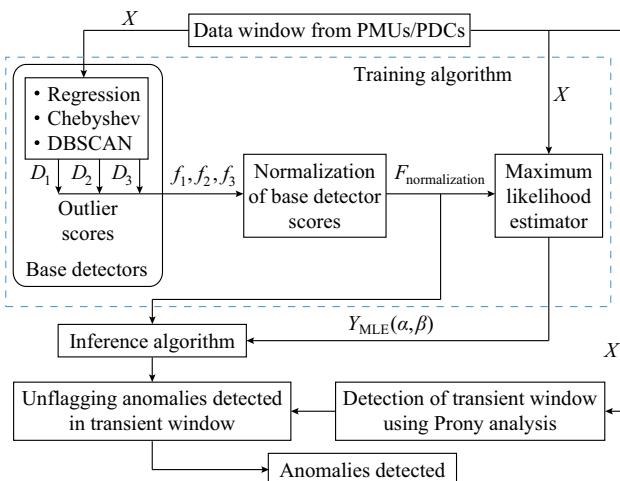


Fig. 3 Anomaly detection architecture

2.3 Prony-based transient window estimation

The Prony analysis [26] 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. 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 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 $J = 300$.

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

where $y(\cdot)$ is the element of the 2.5-second measurement window.

The rank of the Hankel matrix is estimated by the eigen decomposition of the sample correlation matrix as follows:

$$QSQ' = YY' = R_Y Y \tag{7}$$

where Q is an orthogonal matrix whose columns are the eigenvectors of $R_Y Y$; 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:

$$S = \text{diag}(d_1 > d_2 > \dots > d_k > \dots > d_{\frac{J}{2}}) \tag{8}$$

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

$$\sigma_k = \frac{\lg d_k}{\lg d_1} \tag{9}$$

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

The singular values of 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 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 was 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.

Events cause ripples throughout the system. PMUs located near to the event might see a larger variation in measurements than the ones which are farther. 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. 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.

2.4 Data cleaning by SyncAD

SyncAD tool provides two methods of addressing data anomalies, flagging anomalies and replacing the anomalous data by taking the average of the data points that lie before and after the anomaly data point. On a clean PMU data set anomalies were randomly inserted as seen in Fig. 4. This set of anomalous data had single point anomalies, which was detected and cleaned by the averaging method. This produces a data set which is close to the clean data set. The

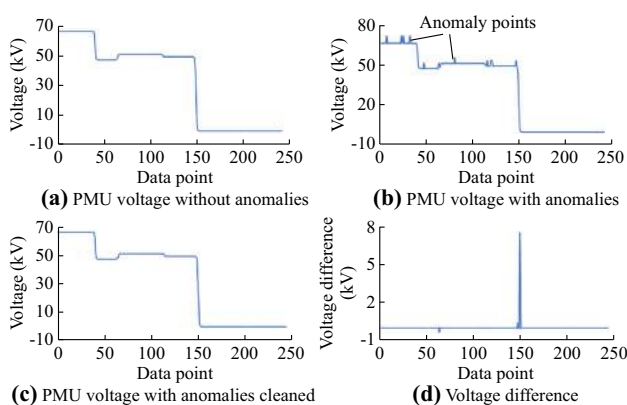


Fig. 4 Outliers imputed by SyncAD

difference is also plotted and it can be seen that except a couple of data points, the cleaned data set matches the data set without anomalies. However, in case where missing packet data anomalies are inserted, the averaging technique is not effective as can be seen in Fig. 5. The missing packet data occurs during transients and replacing these missing data by averaging results into a plot, i.e. data set which is not similar to the original data plot. The difference between this clean data and original data plot is very large. Therefore, for these cases flagging the anomaly instances will be useful.

2.5 SyncAD performance

The SyncAD tool for PMU anomaly detection was tested by adding anomalies to clean data set obtained from real PMUs in Smart Grid Demonstration and Research Investigation Lab (SGDRIL) at Washington State University (WSU). The results are presented in Table 1 based on the precision P_p and true positive P_t as discussed by (10) and (11).

$$P_p = \frac{D_D \cap D_A}{D_D} \quad (10)$$

$$P_t = \frac{D_D \cap D_A}{D_A} \quad (11)$$

where D_D and D_A are the detected bad data and actual bad data, respectively.

3 Automated failure diagnosis

This section discusses the centralized 5-digit algorithm [13], to systematically identify the location of the transmission line outage based on the PMU measurements in the presence of the protection system malfunctioning. This tool is developed to run in the control center and reduces the

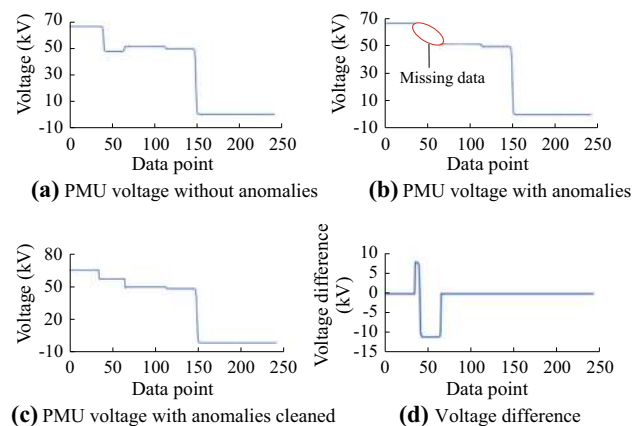


Fig. 5 Missing data imputed by SyncAD



Table 1 Performance of anomaly detectors

| Data set | Without Prony | | With Prony | |
|----------|---------------|--------|------------|--------|
| | Precision | Recall | Precision | Recall |
| 1 | 0.9859 | 0.9333 | 0.9560 | 0.9437 |
| 2 | 0.9745 | 0.9562 | 0.9928 | 0.9437 |
| 3 | 0.9803 | 0.9677 | 0.9934 | 0.9612 |
| 4 | 0.9875 | 0.9575 | 0.9939 | 0.9939 |
| 5 | 0.9820 | 0.9761 | 0.9940 | 0.9880 |

confusion of the conflicting alarms triggered by the outage. The tool needs at least one cycle of PMU data before the first distance relay tripping is reported. By running the tool, the operator can easily find the malfunctioned protection devices as well as the accurate location of the fault. The automated failure diagnosis tool consists of four different stages which are protection net (ProNet) selection and data collection, 5-digit data calculation, multiple hypothesis generation, and hypothesis selection. Each of the four stages are described as the following subsections.

3.1 ProNet selection and data collection

The very first step of the 5-digit algorithm is to determine the ProNet. The ProNet can be generally defined as all the relays of the whole grid that can see the fault in each of their three zones of protection. In this work, zone 1, zone 2, and zone 3 cover up to 0.8, 1.2, and 2 times of the length of the transmission line, respectively. Once the first tripping relay is reported, the ProNet selection process will begin and will include all the lines with an open breaker accompanied by all the neighbouring lines and buses. The main idea of selecting the ProNet is to confine the number of PMUs that are of interest in the algorithm and to expedite the fault location identification process.

3.2 5-digit data calculation

5-digit number is a string message consisting of combinations of 0 and 1 as each of the digits. The first digit is considered as the “trust digit”. Based on the PMU status flag, the trust digit will be 1 when the PMU data are accurate and will be 0 when the measurements are not valid. The second and third digits are together called as the fault digits. Based on the accurate PMU measurements, the line impedance Z can be obtained by:

$$Z = \frac{(V_1 - V_2)(V_1 + V_2)}{V_1 I_2 + V_2 I_1} \quad (12)$$

where V_1 and V_2 represent the sending and receiving end measured voltages, respectively; and I_1 and I_2 represent the

sending and receiving end measured currents, respectively.

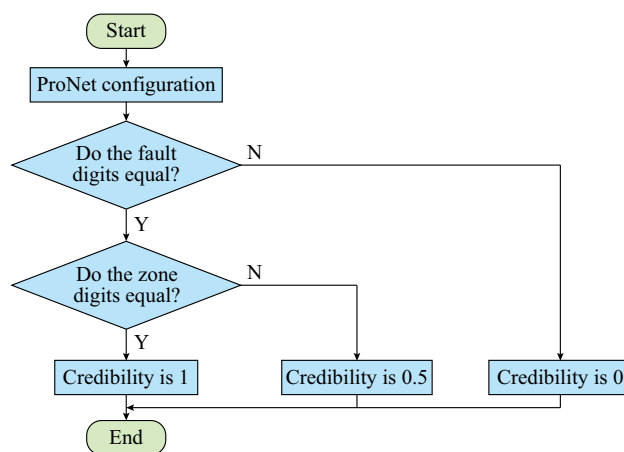
If Z is within any protection zones of the relay, then the second digit will be 1, meaning that the corresponding relay can see the fault in any of its protection zones. Otherwise, it will be 0. The third digit is the breaker digit, which is equal to 1 if the corresponding breaker is open and is equal to 0 if the breaker is closed. The fourth and fifth digits are called zone digits together and will represent the zone of protection of the corresponding relay, in which 00, 01, 10, and 11 mean zones 1, 2, 3, and none of the zones, respectively.

3.3 Multiple hypothesis generation

This step is to examine all the possible scenarios of the existing status of the ProNet. All the transmission lines within the ProNet will be tested for the fault location and the compatibility of the status of relays and breakers with those assumptions will be examined. As the result of this step, all the scenarios of possible fault locations and combinations of protection malfunctioning will be extracted. The detailed calculations of this step can be found in [8].

3.4 Hypothesis selection

This step is to determine the actual event by comparing the several different possible hypothesizes, developed in previous step. A credibility value will be assigned to each hypothesis by comparing the last four digits of the data string for all the PMUs within the ProNet, as shown in Fig. 6. The comparison is between the last four digits of the relays in each hypothesis with the last four digits of relays constructed by the PMU measurements. The scenario with the maximum credibility will be chosen as the actual event,

**Fig. 6** Failure diagnosis with accurate PMU data

as it explains the status of the relays and breakers within the ProNet more accurately.

$C_k(H)$ represents the credibility for the PMU k in the hypothesis H . Considering n number of PMUs in the ProNet, the total credibility value for the hypothesis H is as follows:

$$C_{tot}(H) = \frac{\sum_{k=1}^n C_k(H)}{n} \times 100 \quad (13)$$

The presented failure diagnosis algorithm is proven to have a higher accuracy of finding the location of the faults and protection device failures/malfunctioning. Also the time for the identification is considerably less than the other similar algorithms. However, the main specific advantage of the 5-digit algorithm compared to the other existing failure diagnosis methods lies in the automation process and ability to work with multiple failures. Manual work always accompanies with high chances of making mistakes, which can primarily affect the final decision of finding the location of the fault or the malfunctioned protective devices. To address the aforementioned issue, the 5-digit algorithm is an automated tool with the minimal participation from the operator. The only task needed by the operator is to launch the tool and provide the PMU time series data including at least one cycle of data before and after the event/fault happens. Considering different categories of existing fault location identification methods, the following features of the 5-digit algorithm makes it a better option for many cases to be run in the control center: ① based on multiple hypothesis; ② being quantitative; ③ as a result of being based on multiple hypothesis and being quantitative, the algorithm operation time is faster than the existing methods; ④ accurate performance of finding the location of the fault even with presence of protection devices malfunction; ⑤ being automated in real time.

4 Failure diagnosis with data anomalies

After the detection of outliers in the PMU measurements, the outliers can be either totally discarded or recovered, depending on the time length and the location of the invalid data. As a general rule, recovering the inaccurate data is preferred and discarding the data is prevented.

In order to clean the bad data, the measured values of the neighbouring buses and lines can be exploited to get an estimation of the error of invalid measurements. On the other hand, in case of large time periods of missing/inaccurate data, or several neighbouring faulty measurements, the elimination of the bad data would be inevitable.

As for the purpose of the practical situations in power system, at most of the time there are some types of unreliable measurements in the PMU streaming data which are generally flagged by the PMU. Therefore, the need for an automatic PMU invalid data detection and cleaning is of paramount importance in any event detection procedure. This paper develops an algorithmic extension, to compensate for the gap in the presented failure diagnosis tool. The presented tool, expects the full authentic PMU streaming data as the input. However, as discussed earlier, unavoidable presence of the bad measurements in the time series data, highly impacts the performance of the failure diagnosis tool. Figure 7 shows the block diagram of the extended failure diagnosis tool with addressing the presence of invalid data.

Once the credibility value for each PMU is calculated, the rest of the failure diagnosis algorithm is the same as discussed in Section 3.4.

5 Real-time validation

The performance of the presented 5-digit algorithm is tested on the 96-bus transmission system using the hardware in the loop (HIL) OPAL-RT real-time simulator. Protection malfunctioning case study for the 96-bus test system is developed and the possible hypotheses have been investigated. Then the developed case study is simulated in

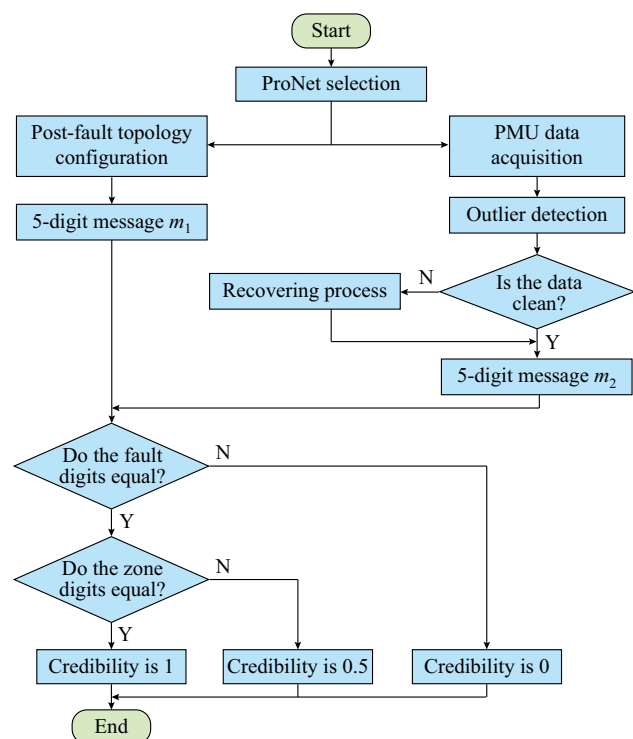


Fig. 7 Anomaly detection architecture

OPAL-RT real-time simulator and the acquired PMU data are provided for the fault location identifier. Finally, the results are compared with the actual scenario and the credibility values are assigned. Further on, the case study is extended and the PMU data anomaly is taken into account. As a result, the flag digit for some of the PMUs gets equal to 0, meaning the presence of bad data. Then the ensemble-based algorithm is applied to detect and clean the outliers of the measurements. As the last step, the fault location identifier tool is provided with the recovered PMU data and using the generated credibility values, the actual scenario is flagged.

In this work, the protection scheme is set to be non-pilot and with the distance relays. The ProNets are programmed to operate within their three zones of operation. As for the settings of the distance relays, the observed admittance in the three zones of operation are set to be 80%, 120% and 200% of the length of the transmission line, respectively, for the first, second, and third zones of operation.

For the HIL simulation of the system, the current transformer (CT) sampling rate is assumed to be 2000 Hz and the PMU reporting rate is 60 Hz, and all the further analysis on the system for all the case studies have been done based on the same rate of data acquisition and report.

5.1 Protection malfunctioning case 1

A three-phase symmetrical line to ground fault is simulated at $t = 0.6$ s in the transmission line 1 as shown in Fig. 8. The relays 1 and 2 at the transmission line 1, detect the fault in their first zone of protection and are expected to send the trip signal to their corresponding breakers. Relay 1 operates at $t = 1$ s and sends the trip command to the breaker 1, and as a result of the appropriate operation of the breaker 1, the fault isolates from this end. However, relay 2 is assumed to malfunction and refuses to send the trip command. As a consequence, the breaker 2 will not be able to timely operate, and the fault current will be propagated

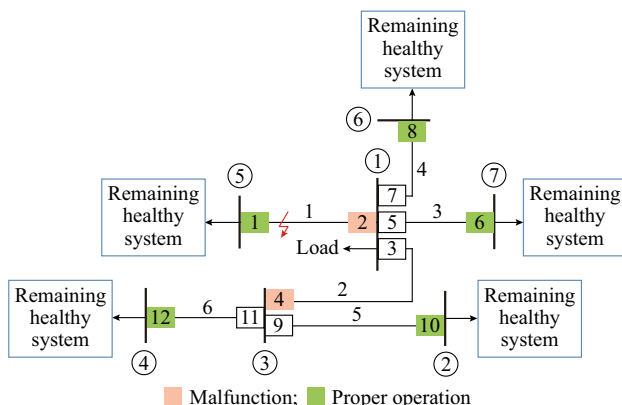


Fig. 8 ProNet configuration

into the neighbouring lines. In the second zone of distance relay protection, relays 4, 6, and 8 are in charge of the detection of the fault. All three relays detect the fault and send the trip command to their corresponding breakers. Breakers 6 and 8 are expected to properly operate and disconnect the fault. However, the breaker 4 is assumed to malfunction and thus the fault current spreads into the transmission lines 5 and 6. To get the fault current disconnected from the rest of the grid, the relays 12 and 10 will operate in their third zones of protection and will properly send the trip command to their corresponding breakers. After receiving the trip command in the third zone of protection, all the corresponding breakers are assumed to work appropriately and as a result the fault will get isolated at $t = 2.4$ s.

Figure 9a shows the fault current. At $t = 0.6$ s, the fault happens, and the breaker 1 is assumed to operate at $t = 1$ s and disconnects the fault by its end. At $t = 2.4$ s, After operation of all the affected breakers, the line gets completely isolated and the fault current drops to zero.

Due to the malfunction of the breaker 2, the fault current propagates to the transmission lines 2, 3, and 4. After a delay of 1.2 s, which is the designed second zone delay in this work, the breakers 6 and 8 trip at 0.6 s + 1.2 s = 1.8 s. Figure 9b and c shows the phase A currents at lines 3 and 4, respectively.

After the isolation of fault at lines 3 and 4 at $t = 1.8$ s, the mis-operation of breaker 4 causes a huge overload current at line 2, as shown in Fig. 9d.

At the final step of the isolation process, breakers 10 and 12 get actuated and trip at $t = 2.4$ s, and the faulty section segregates from the remaining healthy part of the system. Currents at lines 5 and 6 are shown in Fig. 9e and f, respectively.

After $t = 2.4$ s, the ProNet is completely isolated from the healthy part of the network, thus, the current for all the transmission lines in the ProNet will be equal to 0.

Figure 10 summarizes all the expected operations and malfunctions of the breakers with the corresponding timeline of the activities.

The existing status of the system, after operating all the relays and breakers, can be explained by several different possibilities.

The presented failure diagnosis tool, aims to investigate all the possible hypotheses which can account for the current status of the system. It exploits the post-fault topology of the system shown in the Fig. 11 to calculate the 5-digit data for each of the relays in all the suspect lines, and at the next step, by comparing the actual PMU measurements with the calculated values, the most possible hypothesis will be adopted as the actual happening.

The tool tests the possibility of fault at all the transmission lines 1, 2, 3, 4, 5, and 6 in Fig. 11, and examines

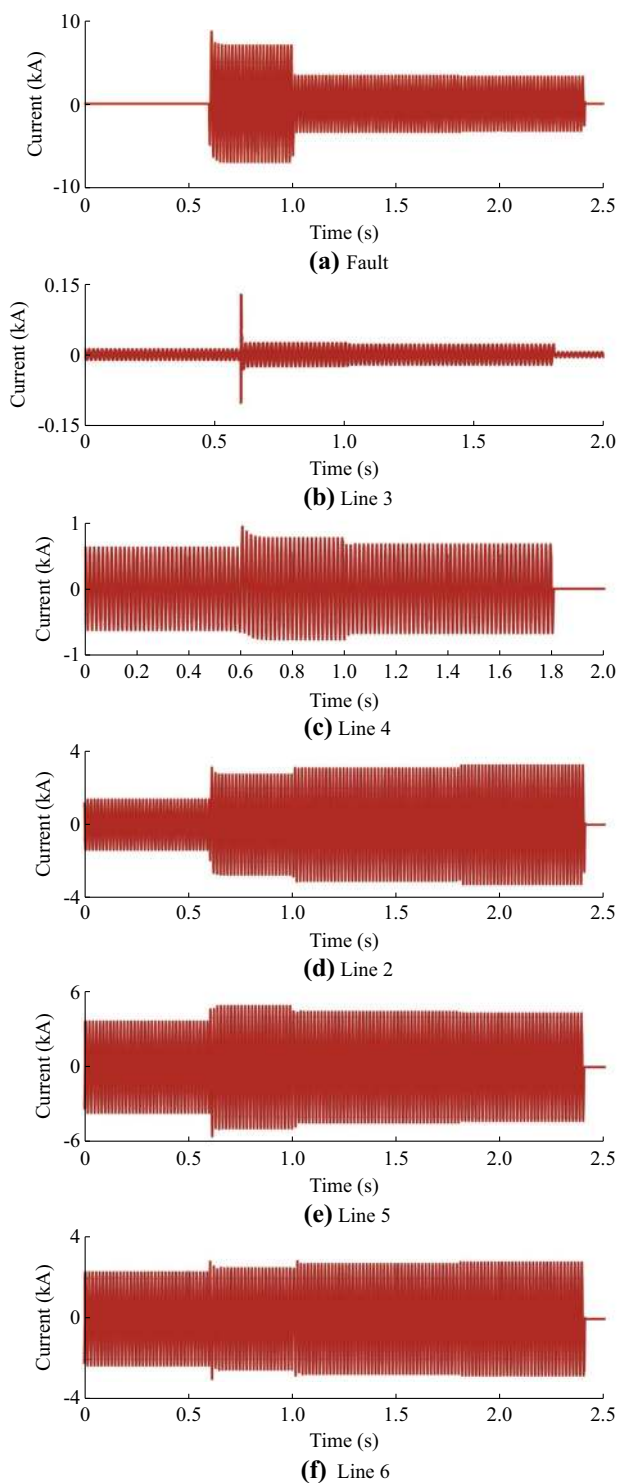


Fig. 9 Phase A current of true event

whether or not each hypothesis complies with the current status of the system.

As for the developed case study in this work, there are 6 different transmission line candidates for the location of fault and considering three different possibilities of fault

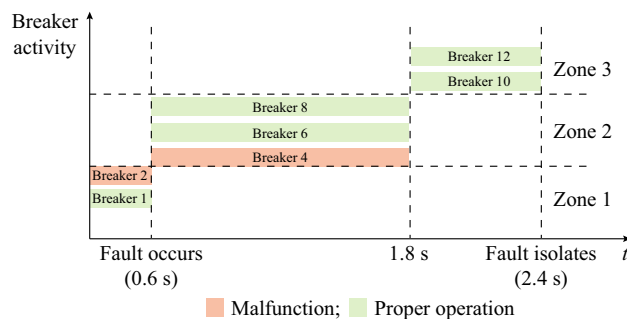


Fig. 10 Breaker activities time-line

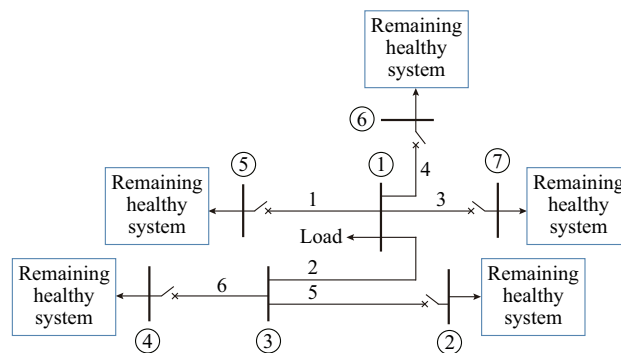


Fig. 11 Post-fault ProNet topology

happening at 0–20%, 20%–80%, 80%–100% portion of the line, there are 18 different hypotheses, which are all investigated. There are 12 total number of relays/PMUs within the final ProNet, that are all valued a 5-digit number based on each of the hypotheses as well as a 5-digit number based on the PMU measurements. At this step, bad data is not considered in the measurements, therefore the PMU flag is 1 for all the 12 PMUs.

Tables 2, 3, 4, and 5 respectively show the measurement based 5-digit data (total number of hypotheses is 18) as

Table 2 Measurement-based 5-digit message

| PMU number | Trust digits | Fault digits | Zone digits |
|------------|--------------|--------------|-------------|
| 1 | 1 | 11 | 00 |
| 2 | 1 | 10 | 00 |
| 3 | 1 | 00 | 11 |
| 4 | 1 | 10 | 01 |
| 5 | 1 | 00 | 11 |
| 6 | 1 | 11 | 01 |
| 7 | 1 | 00 | 11 |
| 8 | 1 | 11 | 01 |
| 9 | 1 | 00 | 11 |
| 10 | 1 | 11 | 10 |
| 11 | 1 | 00 | 11 |
| 12 | 1 | 11 | 10 |



Table 3 Hypothesis 1, fault at 20%–80% line 1

| PMU number | 5-digit message | Credibility |
|------------|-----------------|-------------|
| 1 | 11100 | 1 |
| 2 | 11000 | 1 |
| 3 | 10011 | 1 |
| 4 | 11001 | 1 |
| 5 | 10011 | 1 |
| 6 | 11101 | 1 |
| 7 | 10011 | 1 |
| 8 | 11101 | 1 |
| 9 | 10011 | 1 |
| 10 | 11110 | 1 |
| 11 | 10011 | 1 |
| 12 | 11110 | 1 |
| Total | | 1 |

Table 4 Hypothesis 2, fault at 20%–80% line 2

| PMU number | 5-digit message | Credibility |
|------------|-----------------|-------------|
| 1 | 11101 | 0.500 |
| 2 | 10011 | 0 |
| 3 | 11000 | 0 |
| 4 | 11000 | 0.500 |
| 5 | 10011 | 1.000 |
| 6 | 11101 | 1.000 |
| 7 | 10011 | 1.000 |
| 8 | 11101 | 1.000 |
| 9 | 10011 | 1.000 |
| 10 | 11101 | 0.500 |
| 11 | 10011 | 1.000 |
| 12 | 11101 | 0.500 |
| Total | | 0.666 |

Table 5 Hypothesis 3, fault at 20%–80% line 6

| PMU number | 5-digit message | Credibility |
|------------|-----------------|-------------|
| 1 | 11110 | 0.500 |
| 2 | 10111 | 0 |
| 3 | 11001 | 0 |
| 4 | 10111 | 0 |
| 5 | 10011 | 1.000 |
| 6 | 11110 | 0.500 |
| 7 | 10011 | 1.000 |
| 8 | 11110 | 0.500 |
| 9 | 10011 | 1.000 |
| 10 | 11101 | 0.500 |
| 11 | 11000 | 0 |
| 12 | 11100 | 0.500 |
| Total | | 0.416 |

well as the calculated values for the three hypothesized faults at lines 1, 2, and 6 in addition to their corresponding credibility values, which are calculated using (13). Due to the limitation of space, we will just create the tables and investigate the post-fault quantities for transmission lines 1, 2, 3, and 4, as these lines include all the three different zones of operation and exploit the symmetry of the ProNet.

Based on the PMU measurements and the comparison between the calculated credibility values for each hypothesis, the failure diagnosis tool recognizes the first hypothesis as the actual event.

The presence of bad data in PMU measurements can lead to wrong measured 5-digit messages for some PMUs/relays within the ProNet, which in turn will lead to the inaccurate calculation of the credibility values for the different hypotheses and accordingly the final accuracy of the failure diagnosis process will be jeopardized.

In order to consider the effect of bad data in the failure diagnosis tool, anomalies are assumed to be present in the measurements of the PMU which is connected to bus 1. For the purpose of this paper, we assume that the anomaly is in the form of voltage magnitude outliers in the measurements.

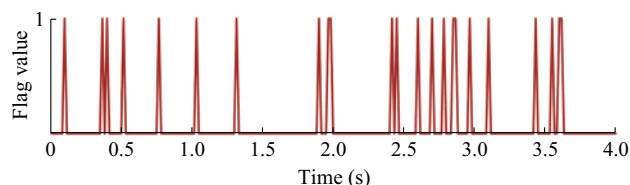
Figure 12 shows the flag values for the PMU measurements at bus 1.

Considering the anomalies of Fig. 12 in the measurements of the PMU at bus 2, the voltage waveform of this bus, is as shown in Fig. 13. In the next step, the anomalous measurements are accordingly cleaned and the corresponding credibility values for the case of anomalous and cleaned data are calculated as shown in Table 6.

Without the invalid data detection and cleaning process, the failure diagnosis tool mistakenly chooses the second hypothesis as the true event, whereas when the invalid data are recovered, the first hypothesis is correctly pinpointed as the actual happening.

5.2 Protection malfunctioning case 2

Considering a cyber-attack to relays 3 and 7, the normal operation of these relays is affected and as a result of injecting the false data to the corresponding current and voltage transformers, unexpected tripping of circuit

**Fig. 12** Flag values for PMU measurements

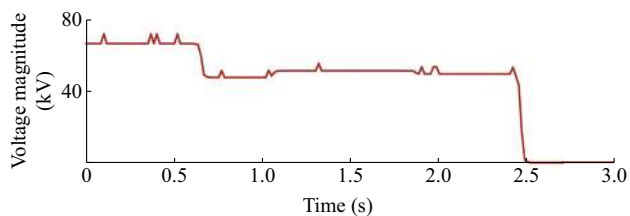


Fig. 13 Anomalous measurements

breakers 3 and 7 at $t = 2.6$ s is assumed. The ProNet for this scenario is the same as Fig. 8. Post-fault topology of the system after operating all the relays and breakers is shown in Fig. 14.

For this scenario, we continue with investigating the possibility of fault at lines 1, 2, and 6. Table 7 shows the final credibility values for each hypothesis with PMU anomalies as well as the credibility values for the recovered measurements.

In case of the anomalous data, the performance of the failure diagnosis without anomaly detection is inaccurate, and the third hypothesis is by mistake selected as the correct happening. However, after the PMU outlier detection and cleaning is applied, the credibility of the first hypothesis increases to the maximum value and is accepted by the tool as the actual event.

As for the required computation time by the developed tool, it is worth-mentioning that the complexity of the event highly impacts the required time of the fault diagnosis tool to find the solution. Such factors as number of transmission lines within the ProNet, number of malfunctioned devices, and the extent of the anomaly occurred in PMU data will determine the complexity of the event and affects the operational time of the tool accordingly. Note that time taken will still be much faster compared to the manual analysis.

6 Conclusion

In this paper, the failure diagnosis in transmission line protection system is addressed in presence of measurement data anomalies. An automated 5-digit protection system failure diagnosis tool is discussed and the dependency of the failure diagnosis to the accuracy of the PMU measurements is specified. The outlier detection methodologies tailored for PMU data are outlined and the ensemble-based tool SyncAD is presented. By integrating the 5-digit algorithm as well as the PMU outlier detection and cleaning, a complete protection failure identification algorithm is developed, which can automatically run in operation center and reduces the amount of time taken to analyze the conflicting alarms created by multiple adverse

Table 6 Credibility scores for anomalous and cleaned data

| Hypothesis | Credibility | |
|------------|----------------|--------------|
| | Anomalous data | Cleaned data |
| 1 | 0.583 | 1.000 |
| 2 | 0.750 | 0.666 |
| 3 | 0.375 | 0.416 |

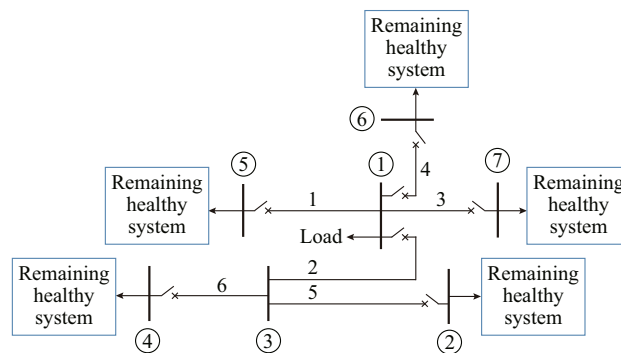


Fig. 14 Post-fault topology

Table 7 Credibility scores for more complex scenario

| Hypothesis | Credibility | |
|------------|----------------|--------------|
| | Anomalous data | Cleaned data |
| 1 | 0.500 | 0.833 |
| 2 | 0.625 | 0.583 |
| 3 | 0.875 | 0.375 |

events with data anomalies. The developed technique considers the presence of bad data in PMU measurements.

The accuracy of the developed methodology is validated using the OPAL-RT real-time simulator, and the simulation results show the superior performance of the proposed algorithms for different test cases.

Acknowledgements The authors gratefully acknowledge the National Science Foundation (NSF) for supporting this research project, and the help of OPAL-RT support team. We appreciate help and support from Dr. Yinghui Wu related to the anomaly detection.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



References

- [1] Aggarwal RK, Johns AT, Bo ZQ (1994) Non-unit protection technique for EHV transmission systems based on fault-generated noise, part 2: signal processing. *IEEE Proc Gener Transm Distrib* 141(2):141–147
- [2] Lin P, Lin T, Liu C (2012) Development of a transmission line fault location platform using digital relay data. In: *Proceedings of IEEE PES general meeting*, San Diego, USA, 22–26 July 2012, 5 pp
- [3] Apostolov A, Vandiver B (2008) Ensuring the correct operation of distance relays under dynamic system conditions. In: *Proceedings of 61st annual conference for protective relay engineers*, College Station, USA, 1–3 April 2008, pp 72–77
- [4] Korkali M, Lev-Ari H, Abur A (2012) Traveling wave-based fault-location technique for transmission grids via wide-area synchronized voltage measurements. *IEEE Trans Power Syst* 27(2):1003–1011
- [5] Yalcin MA, Turan M, Demir Z (1999) Effects of transmission line faults on dynamic voltage stability. In: *Proceedings of PowerTech Budapest 99*, Budapest, Hungary, 29 August–2 September 1999, 80 pp
- [6] Smetek G, Izykowski J (2016) Distance protection performance under single phase to earth fault alone and simultaneously with open conductor failure. In: *Proceedings of 2016 electric power networks*, Poland, 19–21 September 2016, 5 pp
- [7] Gray S, Haas D, McDaniel R (2018) CCVT failures and their effects on distance relays. In: *Proceedings of 71st annual conference for protective relay engineers (CPRE)*, College Station, USA, 26–29 March 2018, 13 pp
- [8] Abdi-Khorsand M, Vittal V (2017) Modeling protection systems in time-domain simulations: a new method to detect mis-operating relays for unstable power swings. *IEEE Trans Power Syst* 32(4):2790–2798
- [9] Pal D, Mallikarjuna B, Reddy RJ et al (2017) Synchrophasor assisted adaptive relaying methodology to prevent zone-3 mal-operation during load encroachment. *IEEE Sens J* 17(23):7713–7722
- [10] Mahadevan N, Dubey A, Chhokra A et al (2015) Using temporal causal models to isolate failures in power system protection devices. *IEEE Instrum Meas Mag* 18(4):28–39
- [11] Cardoso G, Rolim JG, Zurn HH (2008) Identifying the primary fault section after contingencies in bulk power systems. *IEEE Trans Power Deliv* 23(3):1335–1342
- [12] Yu X, Singh C (2003) Integrated power system vulnerability analysis considering protection failures. In: *Proceedings of IEEE PES general meeting*, Toronto, Canada, 13–17 July 2003, pp 706–711
- [13] Cui B, Srivastava AK, Banerjee P (2018) Automated failure diagnosis in transmission network protection system using synchrophasors. *IEEE Trans Power Deliv* 33(5):2207–2216
- [14] Farrokhifard M, Hatami M, Parniani M (2015) Novel approaches for online modal estimation of power systems using PMUs data contaminated with outliers. *Electr Power Syst Res* 24:74–84
- [15] Zhang L, Abur A (2012) Impact of tuning on bad data detection of PMU measurements. In: *Proceedings of IEEE PES innovative smart grid technologies*, Tianjin, China, 21–24 May 2012, 5 pp
- [16] Wang J, Shi D, Li Y et al (2018) Distributed framework for detecting PMU data manipulation attacks with deep autoencoders. *IEEE Trans Smart Grid*. <https://doi.org/10.1109/TSG.2018.2859339>
- [17] Yang Z, Chen N, Chen Y et al (2018) A novel PMU fog based early anomaly detection for an efficient wide area PMU network. In: *Proceedings of IEEE 2nd international conference on fog and edge computing (ICFEC)*, 1–3 May 2018, Washington DC, USA, 10 pp
- [18] Trachian P (2010) Machine learning and windowed sub-second event detection on PMU data via Hadoop and the openPDC. In: *Proceedings of IEEE PES general meeting*, Providence, USA, 25–29 July 2010, 5 pp
- [19] Guan H, Li Q, Yan Z et al (2015) SLOF: identify density-based local outliers in big data. In: *Proceedings of 12th web information system and application conference (WISA)*, Jinan, China, 11–13 September 2015, pp 61–66
- [20] Zhou M, Wang Y, Srivastava AK et al (2018) Ensemble based algorithm for synchrophasor data anomaly detection. *IEEE Trans Smart Grid*. <https://doi.org/10.1109/TSG.2018.2816027>
- [21] Lin Y, Abur A (2018) A highly efficient bad data identification approach for very large scale power systems. *IEEE Trans Power Syst* 33(6):5979–5989
- [22] Sedgwick P (2013) Simple linear regression. <https://doi.org/10.1136/bmj.f2340>. Accessed 12 April 2013
- [23] Amidan BG, Ferryman TA, Cooley SK (2005) Data outlier detection using the Chebyshev theorem. In: *Proceedings of 2005 IEEE aerospace conference*, Big Sky, USA, 5–12 March 2005, pp 3814–3819
- [24] Ester M, Kriegel H-P, Sander J et al (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proceedings of the second international conference on knowledge discovery and data mining*, Portland, USA, 2–4 August 1996, pp 226–231
- [25] Gao J, Tan P-N (2006) Converting output scores from outlier detection algorithms into probability estimates. In: *Proceedings of 6th international conference on data mining*, Hong Kong, China, 18–22 December 2006, pp 212–221
- [26] Hauer JF, Demeure CJ, Scharf LL (1990) Initial results in Prony analysis of power system response signals. *IEEE Trans Power Syst* 5(1):80–89

Amir GHOLAMI received his B.Sc. degree from Sharif University of Technology, Tehran, Iran, in electrical engineering, in 2017. He was a visiting researcher at University of Manitoba, Winnipeg, Canada, in 2017–2018. He is currently a Ph.D. student under the supervision of Dr. Anurag K. SRIVASTAVA at Washington State University, Pullman, USA. His areas of interest include power system transmission and distribution modeling, planning and operation monitoring.

Anurag K. SRIVASTAVA received his Ph.D. degree in electrical engineering from the Illinois Institute of Technology, USA, in 2005. He is an associate professor of electric power engineering at Washington State University and the director of the Smart Grid Demonstration and Research Investigation Lab (SGDRIL) within the Energy System Innovation Center (ESIC). He is an editor of the *IEEE Transactions on Smart Grid*, *IEEE Transactions on Power Systems*, *IET Generation, Transmission and Distribution* and *Elsevier Sustainable Computing*. He is an IEEE distinguished lecturer, and the co-author of more than 250 technical publications. His research interests include data-driven algorithm for the power system operation and control.

Shikhar PANDEY received his B.Tech. degree in electrical engineering from the National Institute of Technology Patna, India, in 2013. He is currently pursuing his Ph.D. in the field of power systems with a focus on smart grid technology from Washington State University, USA.