J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 8797--8813

Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Research on Grid Wide Area Protection Scheme Based on Data Fusion and Distributed Computing

Yu Sui[1], Xun Lu[2], Xiaoyu Deng[1], Wei Xu[1,✉]

[1] *Power Grid Planning Research Center of Guangdong Power Grid Co., Ltd., Guangzhou, Guangdong, 510220, China*
[2] *Guangdong Power Grid Co., Ltd., Guangzhou, Guangdong, 510600, China*

ABSTRACT
Wide-area protection systems are capable of eliminating or mitigating the consequences of disturbances by obtaining multi-point information about the grid system through measurement and communication techniques, and power system control and protection systems. In this study, distribution data in the grid system is collected and preprocessed, the distribution state of the grid system is estimated using data fusion methods, and an optimization method for distribution state estimation based on distributed computing methods is proposed. Then the grid wide-area protection system is designed by combining the grid system fault diagnosis method. Simulation and example analysis results show that the grid wide-area protection system based on data fusion and distributed computing has good performance in processing grid data and detecting and localizing grid faults, and the maximum localization error of faulted line points is maintained within 0.770%. In addition, this grid wide-area protection system is able to accurately detect a certain circuit fault in a regional grid system where faults are frequent, avoiding large-scale power outages and ensuring the stable operation of the grid system. This study has important scientific significance and application value for grid multivariate data fusion modeling and real-time fault detection, and provides an effective wide-area protection scheme for the grid system.

Keywords: data fusion processing, distributed computing, grid wide-area protection

1. Introduction

The concept of wide-area protection was proposed in the process of studying the conventional drawbacks of regional power grid protection, which led to the in-depth study of backup protection. Due to technical and theoretical immaturity and other reasons, the concept of wide-area protection system has not yet formed a unified standard [1-4]. However, according to the background and demand for wide-area protection system, there are two main interpretations of its fundamental meaning: one

✉ Corresponding author.
E-mail address: SuiYu202412@126.com (W. Xu).
Received 25 February 2024; Revised 05 April 2024; Accepted 15 December 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-483
© 2025 The Author(s). Published by Combinatorial Press. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

is an automatic control function system, and the other is a safe and stable control system based on online dynamic safety analysis and wide-area measurement system [5-7]. The former is a system that utilizes multi-point information to quickly and accurately eliminate faults and analyze the impacts suffered by the system after the elimination of faults, as well as take corresponding control measures to realize the protection and automatic control functions. Although there are some differences in the expression, but the understanding of the system function is the same [8-11].

The design and realization of the grid wide area protection system is inseparable from the computer technology, network technology and communication technology and other modern information technology. Advanced computer technology can provide large-capacity storage equipment for the grid wide-area protection system, which provides the possibility of storing and efficiently processing large amounts of data and information [12-15]. The establishment of the database for a variety of measurement data, online calculation control strategies and other data storage, calculation and processing to provide a huge space, so that the data unity, peace of mind and reliability are better maintained. In addition, the continuous improvement of computer operation speed can also provide corresponding technical support for the improvement of the calculation efficiency of the algorithm of wide-area protection system [16-19].

Literature [20] proposed IWAPC, a system capable of providing protection and control for wide-area power grids using new synchronous high-speed communication technology, which includes key technologies such as wide-area intelligent reclosing and wide-area load shedding, which not only integrates the three lines of defense for power system protection and control, but also improves the security of the power grid. Literature [21] describes important contributions to improve the security and resilience of WAMPAC networks. Describes a security framework and defense-in-depth architecture for WAMPAC applications, as well as an exploration of attack-resistant algorithms for WAMPAC. It is emphasized that its findings will open new paths for research in this area. Literature [22] provides the latest components of the WAMS architecture and innovations in WAPAC research development using PMUs. Based on the literature review a wide area protection scheme based on PMUs is proposed, the concept is based on describing the dynamic operation of a transmission line with differential equations and when a fault occurs the set of equations becomes nonlinear and helps to ensure the sustainability of the power flow. Literature [23] provides a major categorization of the scope of WAMPAC cybersecurity and guides a survey of recent research in the field of WAMPAC security. The aim is in laying the foundation for the researchers concerned to study the relevant applications of WAMPAC in smart grids. Literature [24] proposes an LWA-based EBDP strategy in which the microgrid is divided into multiple protection zones at the core of the bus. A microgrid model with multiple dgs is constructed through the PSCAD platform. Simulation results show that the proposed LWADP exhibits good performance and has the potential to be widely applied to large-scale microgrids. Literature [25] emphasized the critical bus data fault identification method based on Wide Area Measurement System (WAMS) data and three metrics based on grid currents. PSSE simulation and OPAL-RT real-time simulation were also applied to test the bus system, and the results pointed out that the algorithm is not only capable of detecting faulty lines in the network, but also identifying faults from power swing and load encroachment situations. Literature [26] introduced a wide area protection scheme in a 400kv network, in order to detect and initiate remedial measures, and applied SCADA with phase measurement units for wide area protection, thus ensuring the robustness of the facility network. Literature [27] points to the “intelligent” transformation of the power system, and the adaptation of this transformation and the implementation of innovative solutions is not possible without the continuous development of the grid, which requires special attention to the protection devices, so as to ensure the successful coordination of the power system.

In this paper, we first collect multi-source data in the distribution grid system, and pre-process the grid data by using missing value processing and normalization processing methods. Combined with the measurement information and bus load data in the grid system, the relationship between the grid data volume measurement and the state volume is solved according to the weighted least squares algorithm to realize the fusion of data and the estimation of the state of the distribution grid, and the data fusion module of multi-intelligent terminal is constructed based on the data platform. Then, according to the characteristics of the grid system, the distributed computing method is used to optimize and improve the grid state estimation method, which provides an accurate data base for the grid wide-area protection system. Subsequently, the grid fault diagnosis and localization method is proposed, and the grid wide-area protection system is constructed by combining the data fusion module and the grid fault diagnosis module. This study analyzes the real-time processing performance and effect of the grid data fusion module and the fault detection and localization module through simulation experiments, and then explores the empirical effect of the grid wide-area protection system by using example analysis.

2. Method

2.1. Methods for estimating the state of the grid system

2.1.1. State estimation methods based on data fusion:

1) Data pre-processing. Various types of terminal data of distribution network have different characteristics and advantages, while the distribution data measured by different terminals overlap, resulting in a waste of resources, so multi-intelligent terminal data fusion will certainly become one of the future trends in the development of distribution network. Based on the above, this paper proposes a multi-intelligent terminal data fusion scheme, i.e., the data middle station system provides data support for the grid wide area protection system through data pooling and completing data fusion based on data fusion of distribution network state estimation.

Missing data, that is, one or more attribute values in the sample set are lost, which is mainly manifested in the sample data in this paper as the data of a certain moment or several consecutive moments are null values. Usually the missing data must be filled in manually. Part of the missing data can be deduced from the trend of the original sample set, and the commonly used filling methods are clustering, mean, hot card, special value, etc. In this paper, considering the small amount of missing data in the sample, the relatively simple method of mean filling [28] is used for processing, with the following formula:

$$x_i = \frac{x_{i-1} + x_{i+1}}{2} \tag{1}$$

where, x_i is missing data.

Data normalization processing [29], due to the influence of the factors of the scale, units and load data are not the same, in the calculation process may appear different degrees of influence on the role. In order to eliminate the different effects between the influencing factors and load data, need to be specification, so that each factor indicator and load data are in a certain common range of numerical characteristics, easy to calculate, set the normalization formula for:

$$x_i^{'} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{2}$$

where x_i is the original data, $x_{\max}, x_{\min}$ is the maximum and minimum values of the sample data, respectively, and x_i is the normalized value. After the above normalization process, each data is transformed to the value range of [0, 1] and is a dimensionless quantity with the same scale. Finally, for the output training, testing and prediction data, the inverse normalization process is performed again and restored to the actual values by reduction calculation.

2) Grid state estimation based on data fusion. Grid state estimation refers to the use of corresponding algorithms to estimate the state of the distribution network system with high accuracy, completeness and reliability based on the measurement information of the distribution system and supplemented with pseudo-measurement data such as bus load prediction and non-remote sensing telemetry data. The accuracy of each type of terminal data in the distribution network is different, and a reasonable allocation of the weight of each measurement can increase the accuracy of state estimation, realize effective data fusion, and provide a basis for the state assessment and identification of weak links in the distribution network. In state estimation, the relationship between quantitative measurements and state quantities is:

$$z = h(x) + n \tag{3}$$

where z is defined as the measurement vector, x is defined as the state vector, and n is defined as the measurement noise.

The weighted least squares algorithm [30] is one of the most widely used methods for solving the state estimation problem at present, and according to the relational equation between the state variables and the measurement vectors, its essence is to solve the following mathematical problem:

$$\min J(x) = (z - h(x))^T W (z - h(x)) \tag{4}$$

$$\frac{\partial J(x)}{\partial x} = -2H^T(x)W[z - h(x)] = 0 \tag{5}$$

$$f(x) = H^T(x)W[z - h(x)] = 0 \tag{6}$$

where h is the measurement equation and W is the measurement weight. In distribution network state estimation, different physical quantities can be selected as the state variables to be sought by the system, including node voltage, branch current and branch power. Each quantity measurement will have a corresponding nonlinear measurement function, and the measurement accuracy can be expressed as:

$$\varepsilon' = \delta_d + \delta_m \tag{7}$$

where ε is the error of the measurement data as a whole, δ_d is the error due to time synchronization, and δ_m is the measurement error of the measurement device. Where:

$$\delta_d = Rt_d \tag{8}$$

where R is the rate of change and t_d is the measurement delay. The measurement error δ_m is given by the manufacturer. Determining the weights based on the measurement time difference calculation error is a more appropriate solution. The time of difference between the calculation moment and the signal reception moment is t_s, and the error ε is:

$$\varepsilon_i = k_i \cdot t_s \cdot \varepsilon_i' \tag{9}$$

Assuming the existence of three systems with different characteristics (dispatch automation system, power usage information collection system, and distribution automation system), the estimation weights are set:

$$\begin{cases} a_i^R = \dfrac{(\varepsilon_i^S)^2(\varepsilon_i^A)^2}{(\varepsilon_i^S)^2(\varepsilon_i^A)^2+(\varepsilon_i^R)^2(\varepsilon_i^A)^2} \\ a_i^A = \dfrac{(\varepsilon_i^R)^2(\varepsilon_i^A)^2}{(\varepsilon_i^S)^2(\varepsilon_i^A)^2+(\varepsilon_i^R)^2(\varepsilon_i^S)^2+(\varepsilon_i^R)^2(\varepsilon_i^A)^2} \\ a_i^S = \dfrac{(\varepsilon_i^R)^2(\varepsilon_i^A)^2}{(\varepsilon_i^S)^2(\varepsilon_i^A)^2+(\varepsilon_i^R)^2(\varepsilon_i^S)^2+(\varepsilon_i^R)^2(\varepsilon_i^A)^2} \end{cases} \tag{10}$$

where a_i^R, a_i^A, a_i^S is the estimated weight of dispatching automation system, power usage information collection system, and distribution automation system, respectively, and $\varepsilon_i^R, \varepsilon_i^A, \varepsilon_i^S$ is the integrated error of each measurement system, respectively.

2.1.2. Multi-intelligent terminal data fusion program. Data center is a complete concept and system, which can better support the rapid innovation and iterative development of front-end applications by pooling data from various business systems to form a big data platform infrastructure, which mainly includes various aspects such as data collection, storage management, computation and presentation, and data circulation and transaction, and consists of three levels: technology, platform and service. The flow of the multi-intelligent terminal data fusion application program based on the data center system is shown in Figure 1. The multi-intelligent terminal data fusion application scheme based on the data center system utilizes the data center system to acquire data collected by the intelligent terminals of each system, form a big data platform after processing and fusing the data, and then provide the data required for evaluation according to the data demand and application of the real-time state evaluation system.

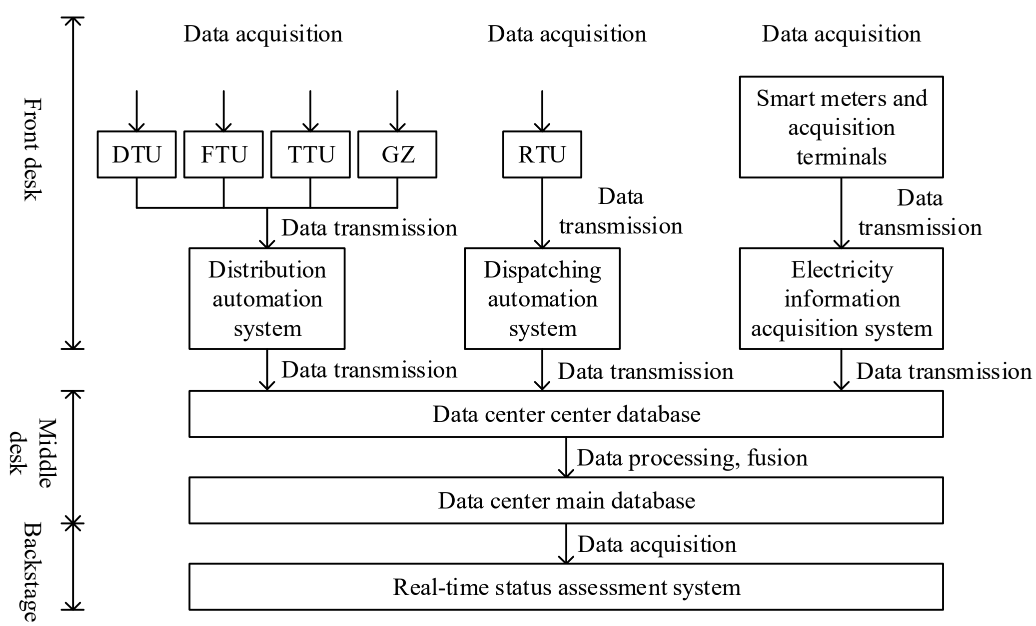

Fig. 1. Network data fusion process

2.1.3. State estimation optimization based on distributed computing. For ease of presentation, the grid data state equation is first given in the form of equation (11):

$$\begin{cases} P_{z.ij}^t = \sum_{h \in H_j} P_{z.jh}^t + (I_{z.ij}^t)^2 r_{z.ij} + P_{L.z.j}^t - P_{z.j}^t \\ Q_{z.ij}^t = \sum_{h \in H_j} Q_{z.jh}^t + (I_{p.ij}^t)^2 x_{z.ij} + Q_{L.z.j}^t - Q_{z.j}^t \\ (v_{z.j}^t)^2 = (v_{z.i}^t)^2 - 2(P_{z.ij}^t r_{z.ij} + Q_{z.ij}^t x_{z.ij}) + (I_{z.ij}^t)^2 (r_{z.ij}^2 + x_{z.ij}^2) \\ (v_{z.i}^t)^2 (I_{z.ij}^t)^2 = (P_{z.ij}^t)^2 + (Q_{z.ij}^t)^2 \end{cases}, z \in P \tag{11}$$

where P is the set of all regions in the distribution network, subscripts $z.i$ and $z.ij$ are nodes i and lines ij in region z, and H_j is the set of nodes in the region downstream of node j and connected to node j. P_{zij}^t, Q_{zij}^t is the active as well as reactive power transmitted in each line, I_{zij}^t is the line current, and $P_{Lzj}^t Q_{Lzj}^t$ is the active as well as reactive power of each node load, respectively. $P_{z,j}^t, Q_{z,j}^t$ are the active power as well as the reactive power of each distributed power source, respectively, and $v_{z,i}^t$ is the voltage magnitude at each node. r_{zij}, x_{zij} are the resistance as well as the reactance on each line, respectively.

First, according to the grid system characteristics, the line power sensitivity matrix within the region itself is calculated one by one, starting from the end line of the region z and following the order of calculation from the end to the beginning. Therefore, the bias derivation of the active injected power for the nodes inside the region can be obtained:

$$\begin{cases} \frac{\partial P_{n-1,n}^t}{\partial P_n^t} - \frac{\left(\left(2\left(P_{n-1,n}^t \frac{\partial P_{n-1,n}^t}{\partial P_n^t} + Q_{n-1,n}^t \frac{\partial Q_{n-1,n}^t}{\partial P_n^t} \right) (v_{n-1}^t)^2 - \right) 2((P_{n-1,n}^t)^2 + (Q_{n-1,n}^t)^2) v_{n-1}^t \frac{\partial v_{n-1}^t}{\partial P_n^t} \right)}{(v_{n-1}^t)^4} r_{n-1,n} + 1 = 0 \\ \frac{\partial P_{n-1,n}^t}{\partial P_n^t} - \frac{\left(\left(2\left(P_{n-1,n}^t \frac{\partial P_{n-1,n}^t}{\partial P_n^t} + Q_{n-1,n}^t \frac{\partial Q_{n-1,n}^t}{\partial P_n^t} \right) (v_{n-1}^t)^2 - \right) 2((P_{n-1,n}^t)^2 + (Q_{n-1,n}^t)^2) v_{n-1}^t \frac{\partial v_{n-1}^t}{\partial P_n^t} \right)}{(v_{n-1}^t)^4} x_{n-1,n} + 1 = 0 \end{cases} \tag{12}$$

The $2((P_{n-1,n}^t)^2 + (Q_{n-1,n}^t)^2) v_{n-1}^t (\partial v_{n-1}^t / \partial P_n^t)$ in Eq. is 2 to 3 orders of magnitude smaller in comparison with the rest, so it can be ignored in simplifying the calculations and, in turn, can be characterized as:

$$\begin{cases} (1 - \frac{2P_{n-1,n}^t r_{n-1,n}}{(v_{n-1}^t)^2}) \frac{\partial P_{n-1,n}^t}{\partial P_n^t} - \frac{2Q_{n-1,n}^t r_{n-1,n}}{(v_{n-1}^t)^2} \frac{\partial Q_{n-1,n}^t}{\partial P_n^t} = -1 \\ -\frac{2P_{n-1,n}^t x_{n-1,n}}{(v_{n-1}^t)^2} \frac{\partial P_{n-1,n}^t}{\partial P_n^t} + (1 - \frac{2Q_{n-1,n}^t x_{n-1,n}}{(v_{n-1}^t)^2}) \frac{\partial Q_{n-1,n}^t}{\partial P_n^t} = 0 \end{cases} \tag{13}$$

The line sensitivity matrix of the end line of the region to the active injected power of the node at the end of the region, which will be calculated in the order from the end to the beginning of the region can be obtained from the sensitivity matrix of all lines within the region to the active injected power of all the nodes, and its specific calculation expression is shown in Eq:

$$\frac{\partial P_{ij}^t}{\partial P_k^t}=\frac{d_{ij}^P\sum_{z\in H_j}\frac{\partial P_{jz}^t}{\partial P_k^t}+b_{ij}^P\sum_{z\in H_j}\frac{\partial Q_{jz}^t}{\partial P_k^t}}{b_{ij}^P c_{ij}^P-a_{ij}^P d_{ij}^P} \tag{14}$$

$$\frac{\partial Q_{ij}^t}{\partial P_k^t}=\frac{c_{ij}^P\sum_{z\in H_j}\frac{\partial P_{jz}^t}{\partial P_k^t}+a_{ij}^P\sum_{z\in H_j}\frac{\partial Q_{jz}^t}{\partial P_k^t}}{b_{ij}^P c_{ij}^P-a_{ij}^P d_{ij}^P} \tag{15}$$

In the formula:

$$\begin{aligned} a_{ij}^P &= 1-2P_{ij}^t r_{ij}^{'}/(v_i^t)^2 \\ b_{ij}^P &= 2Q_{ij}^t r_{ij}^{'}/(v_i^t)^2 \\ c_{ij}^P &= 2P_{ij}^t x_{ij}^{'}/(v_i^t)^2 \\ d_{ij}^P &= 1-2Q_{ij}^t x_{ij}^{'}/(v_i^t)^2 \end{aligned} \tag{16}$$

The mathematical expression for the derivation of the resulting line to all nodes reactive power sensitivity matrix is shown in Eq:

$$\frac{\partial Q_{ij}^t}{\partial Q_k^t}=\frac{d_{ij}^0\sum_{z\in \mathrm{H}_j}\frac{\partial Q_{jz}^t}{\partial Q_k^t}+b_{ij}^0\sum_{z\in \mathrm{H}_j}\frac{\partial P_{jz}^t}{\partial Q_k^t}}{b_{ij}^0 c_{ij}^0-a_{ij}^0 d_{ij}^0} \tag{17}$$

$$\frac{\partial P_{ij}^t}{\partial Q_k^t}=\frac{c_{ij}^Q\sum_{z\in H_j}\frac{\partial Q_{jz}^t}{\partial Q_k^t}+a_{ij}^Q\sum_{z\in H_j}\frac{\partial P_{jz}^t}{\partial Q_k^t}}{b_{ij}^Q c_{ij}^Q-a_{ij}^Q d_{ij}^Q} \tag{18}$$

In the formula:

$$\begin{aligned} a_i^{\mathrm{Q}} &= 1-2Q_i^t x_{ij}/(v_i^t)^2 \\ b_i^{\mathrm{Q}} &= 2P_i^t x_i/(v_i^t)^2 \\ c_i^{\mathrm{Q}} &= 2Q_{ij}^t r_i/(v_i^t)^2 \\ d_i^{\mathrm{Q}} &= 1-2P_{ij}^t r_{ij}/(v_i^t)^2 \end{aligned} \tag{19}$$

In calculating the nodal voltage sensitivity, the voltage squared term is viewed as a whole and denoted by v in order to simplify the solution process. Thus, the mid-node voltage equation can be described again as:

$$\begin{cases} V_{z,j}^t = V_{z,i}^t - 2(P_{z,ij}^t r_{z,ij} + Q_{z,ij}^t x_{z,ij}) + (I_{z,ij}^t)^2 (r_{z,ij}^2 + x_{z,ij}^2) \\ V_{z,i}^t (I_{z,ij}^t)^2 = (P_{z,ij}^t)^2 + (Q_{z,ij}^t)^2 \end{cases}, z\in \mathrm{P} \tag{20}$$

The sensitivity of the node voltage at the beginning of the region to the node injected power at a point in its own region, which is calculated in the order from the beginning to the end of the region can be obtained by the sensitivity matrix of all the node voltages within the region to the injected power of all the nodes, which is calculated mathematically as:

$$
\begin{cases}
\dfrac{\partial V_j^t}{\partial P_k^t} = a_{ij}^v \dfrac{\partial V_i^t}{\partial P_k^t} + b_{ij}^v \dfrac{\partial P_{ij}^t}{\partial P_k^t} + c_{ij}^v \dfrac{\partial Q_{ij}^t}{\partial P_k^t} \\
\dfrac{\partial V_j^t}{\partial Q_k^t} = a_{ij}^v \dfrac{\partial V_i^t}{\partial Q_k^t} + b_{ij}^v \dfrac{\partial P_{ij}^t}{\partial Q_k^t} + c_{ij}^v \dfrac{\partial Q_{ij}^t}{\partial Q_k^t}
\end{cases} \tag{21}
$$

In the formula:

$$
\begin{aligned}
a_{ij}^v &= 1 - ((P_{ij}^t)^2 + (Q_{ij}^t)^2)(r_{ij}^2 + x_{ij}^2) / (V_i^t)^2 \\
b_{ij}^v &= 2P_{ij}^t (r_{ij}^2 + x_{ij}^2) / V_i^t - 2r_{ij}^v \\
c_{ij}^v &= \&2Q_{ij}^t (r_{ij}^2 + x_{ij}^2) / V_i^t - 2x_{ij}
\end{aligned} \tag{22}
$$

2.2. Grid fault diagnosis and localization methods

In the distribution network, a threshold value can be set for the fault starting element in the grid, and this protection is activated when the fault current in the grid is greater than this value, and the method of calculating this starting element can be can use the fault component algorithm:

$$
\Delta I_\varphi \geq I_{\mathrm{SET}} \tag{23}
$$

where I_{SET} is the threshold value, which can also be referred to as the setting value, $I_{\mathrm{SET}} = (0.1 - 0.3) I_{\mathrm{N}}$ and I_{N} are the rated currents:

$$
\Delta I_\varphi = | i(k) - i(k-N) | - i | (k-N) - i(k-2N) | \tag{24}
$$

where ΔI_φ is the starting amount of phase current, φ indicates the three phases of A, B and C, $i(k)$ is the sampling value of current at k, and N is the number of sampling points of an industrial frequency cycle. Moreover, as a backup protection, a time limit can be set, when after a certain time limit, the original main protection does not act, i.e., the protection based on the local electrical information acquisition does not act or the circuit breaker refuses to act, then this backup protection will act.

Whether a two-phase or three-phase short circuit occurs in the downstream or upstream line of the grid at a boundary node, the amplitude of the current flowing through the grid increases, but the direction of the short-circuit power is different. Therefore, the following two conditions need to be satisfied in fault area localization.

1) Determine whether the current in the boundary area crosses the boundary About the threshold value of overcurrent protection can be adjusted according to the maximum load current of the line, as Eq:

$$
I_\varphi > I_{MAX} \tag{25}
$$

where I_φ is the effective value of phase current. I_{MAX} is the maximum load current value.

2) Judge whether the power direction is consistent with that flowing through the upstream and downstream switches

If the power direction is positive and is consistent with the downstream power direction, this area can be considered as the fault area. When the power direction is negative and is opposite to the upstream power direction, the region can also be considered as a fault region.

2.3. Grid wide protection system design

The grid wide-area protection system based on data fusion and distributed computing is shown in Fig. 2, which divides the region in the figure into an islanded region and a non-islanded region, both of which can operate independently after a fault in a non-domestic region. First, information is exchanged between $IFTU_1$ and $IFTU_2$ to determine the fault region and determine whether it is within the circle. Second, depending on the location of the fault point, the grid at each location sends a trip signal to the corresponding IB_2. If the fault area is inside the circle, the circuit breaker at the bus that the IIDG is connected to trips (IB_2). If the fault area is outside the circle, the protection at that section of the line should trip (B_1) and the in-circle area continues to supply power to the in-circle area. Whether the fault is in the park or out-of-circle, it is due to be equipped with the appropriate automatic reclosing device, if it is a transient fault, the recloser recloses and then closes. If it is a permanent type fault, the reclosing gate is reclosed and then tripped.

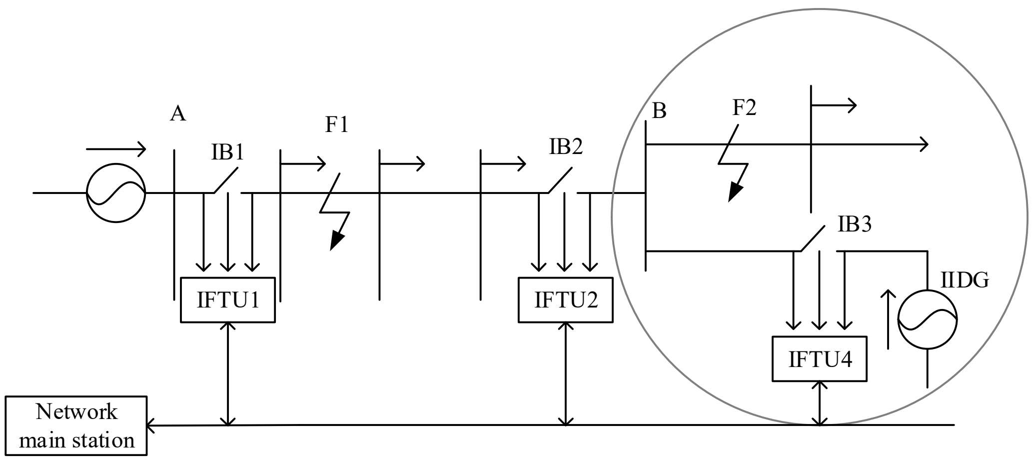

Fig. 2. Grid wide range protection scheme

3. Results and discussion

3.1. Performance analysis of real-time processing of grid data fusion

3.1.1. Simulation experiment design. The purpose of this experiment is to verify the real-time processing performance of distribution network data in the grid wide protection system based on data fusion and distributed computing proposed in this paper. This experiment restricts the cluster physical resources, and firstly determines the maximum throughput rate and average processing latency under the current environment when Flink parallelism is 6 for executing a simple data processing task, keeping all the operators of the data processing task without generating backpressure and with a good load condition. The measured data is used as the processing performance standard under the current physical resource constraints. After obtaining the processing performance standard, the maximum throughput rate and average processing delay of each Flink task and each stage of the task for real-time multi-source data processing in the grid wide-area protection system under the same physical resource limitations are measured under the same operating state. Compared with the processing performance standard, the impact of each data processing stage of real-time processing distribution system on throughput rate and processing delay is judged. At the same time, combining Checkpoint status, Flink Metrics parameters and other data to judge the operation status of each component of the real-time processing system, the performance indexes of each link of the real-time processing system are comprehensively derived. The performance index of the real-time processing system has two parts, the first part is the data processing efficiency of the distribution grid data of the grid wide-area protection system, including the data throughput per second, the average processing latency of each piece of data and the data and status of the backpressure situation and the load situation of the arithmetic operator of each link in the data processing process. The second part is the overall processing delay of data other than distribution network data and the delay of processing links.

3.1.2. Experimental results and analysis. The experiments were conducted by limiting the data input throughput rate from 10000 items per second to 50000 items per second under the current physical resource constraints and with a parallelism degree of 6. The results of the data processing throughput rate experiments are shown in Fig. 3. When the data input throughput rate is 35000 items per second and below, the average processing throughput rate and the average output throughput rate of the grid wide area protection system are close to the data input throughput rate. When the data input throughput rate reaches 35,000 items per second and above, the difference between the data processing throughput rate and the output throughput rate and the input throughput rate widens, indicating that the data processing throughput rate cannot keep up with the input throughput rate, resulting in a data pileup between the input side and the processing side. It can be seen that the processing of data in the distribution network by the grid wide area protection system can be maintained in a stable state at 35,000 items per second, without generating high backpressure and high load, and maintaining a low latency of data processing while ensuring a high processing speed. At the same time, it can be seen that under different data input throughput rates, the data processing throughput rate is almost the same as the data output throughput rate, which shows that the performance bottleneck of the system does not appear in the data output stage.

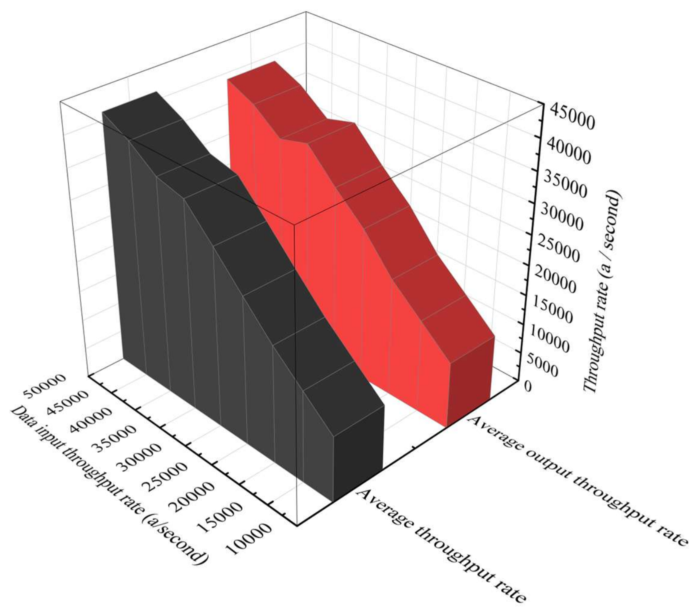

Fig. 3. Data processing rate test results

The experimental results of the load cases at various stages of the data processing task in the wide-area protection system of the grid are shown in Fig. 4. Without limiting the rate at which data is sent into the wide area protection system, the data fusion processing task shows very high backpressure (88.26%) and load conditions (92.63%) at the processing end, and high backpressure (26.48%) and load (59.67%) at the output end. When backpressure occurs in one part of the data processing task of the system grid, it indicates that data pileup occurs upstream or downstream of that part, and the higher the percentage of backpressure indicates that the data pileup is more serious, and not only the average processing delay of the data will be seriously affected, but also the data throughput rate of the system will be reduced. Similar to the results in the above data processing throughput rate analysis, when the data input throughput rate is below 35,000 items/second, the backpressure (0.29%-15.64%, 0%-0.09%) and high load (4.88%-22.11%, 2.93%-7.04%) situation between the processing end and the output end are also better, and the data fusion and distributed algorithms-based grid wide-area protection system of data processing performance is in a stable state.

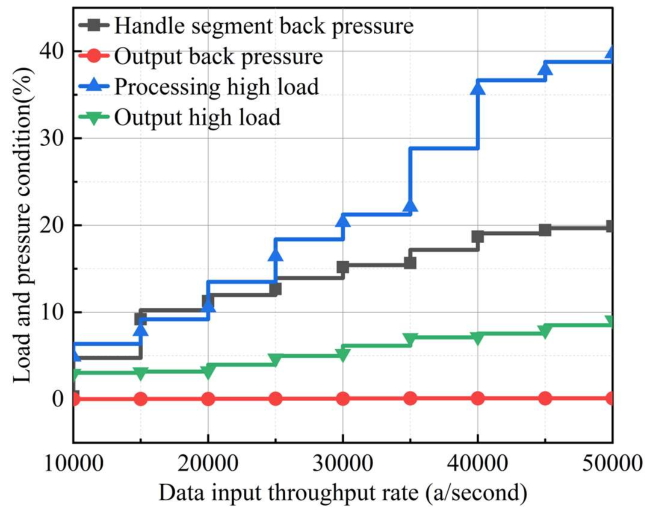

Fig. 4. Data processing load

3.2. Simulation Analysis of Grid Fault Detection and Localization

3.2.1. Simulation grid system establishment. In this paper, the IEEERTs-24 system and the actual operation data of a provincial grid were used to simulate and test and analyze the system respectively.IEEERTS-24 is a reliability standard test system with a total installed capacity of 3,552 MW and a maximum annual load of 2,965 MW, containing 30 nodes with 18 load nodes, a total of 36 lines, 7 transformers, and 30 generator sets. Since the components involved in sampling include four categories of lines, nodes, generators and transformers, a transformer is added between the generators and loads in the system above 220Kv and the high-voltage bus to connect it to the high-voltage grid, and it is set that the failure of each category of components shall not be greater than two, and the total number of failures shall not be greater than three. The reliability component modeling of a provincial power grid adopts the mode in which the high-voltage (above 220Kv) grid is the main one and the low-voltage grid is attributed to the corresponding parameters of the high-voltage. In order to adapt to the needs of operational reliability assessment, the modeling of loads, generators, lines, and buses are not merged, especially the buses, which are modeled as unmerged bus segments with small branches, which makes the reliability assessment of the grid wide protection system more accurate, and the assessment results are more suitable for online applications. Nevertheless, this modeling nearly doubles the number of nodes and branches. After modeling, this grid includes a total of 211 HV nodes, 195 lines, 64 generators, 182 transformers, and 32 section constraints (stability constraints).

3.2.2. Fault Scenario Simulation Results. In order to verify the diagnostic analysis and localization effect of the wide-area protection system on the location of fault points and fault types in the simulated grid system, the IEEERTs-24 grid and a provincial grid simulation system are set up to have faults on the L22-25 lines and L18-19 lines, respectively. The fault locations are at 5%, 20%, 50% and 90% from the circuit bus, respectively, and the fault types are A-phase ground fault AG, A-phase and B-phase short-circuit fault AB, A-phase and B-phase short-circuit ground fault ABG, and three-phase short-circuit fault ABC, and the setups of the AG and ABG faults have a transition resistance of 300Ω. The real-time analysis results of injected currents in the grid wide-area protection system are shown in Table 1, and the fault detection results in a provincial grid simulation system are shown in Table 2, when various faults occur in the lines in the IEEERTs-24 grid simulation system, respectively. In the table, $dI22$, $dI23$, $dI24$, $dI25$, $dI18$, and $dI19$ represent the line node fault injection current amplitude detected in the wide-area protection system, and their values are taken as the steady state values after 1 week wave of the fault. The distance error (%) of the grid wide-area protection system for the diagnosis and localization of the fault point in the simulated grid = (distance of the localized fault -

distance of the real fault)/line length × 100%. In the case of various faults occurring in lines L22-25 and L18-19, respectively, the grid fault detection and localization module in the grid wide-area protection system based on data fusion and distributed computing in this paper is able to accurately identify nodes on both sides of the faulty line and detect the faulty line. The localization distance errors for various fault points of lines occurring in the two simulated grid systems are 0.458% and 0.266% on average, which will not cause misjudgment of normal lines in the region and have a high margin.

Table 1. The results of the failure analysis in IEEERTs-24 grid simulation system

Fault type	Fault position	R_f / Ω	*dI*22	*dI*23	*dI*24	*dI*25	Fault line	Distance error (%)
AG	5%	0.01	0.0049	0.1896	0.0975	0.1456	L22-25	0.325
		300	0.0303	0.0496	0.0921	0.1382	L22-25	0.151
	20%	0.01	0.0937	0.1253	0.0786	0.1664	L22-25	0.751
		300	0.1151	0.1964	0.052	0.1179	L22-25	0.77
	50%	0.01	0.0381	0.1768	0.1837	0.0268	L22-25	0.474
		300	0.157	0.0068	0.125	0.0382	L22-25	0.068
	90%	0.01	0.0475	0.1017	0.0227	0.1267	L22-25	0.489
		300	0.1369	0.009	0.1515	0.0823	L22-25	0.479
ABG	5%	0.01	0.0732	0.1832	0.1346	0.0203	L22-25	0.47
		300	0.0373	0.1966	0.0389	0.0452	L22-25	0.07
	20%	0.01	0.1309	0.1225	0.134	0.0514	L22-25	0.11
		300	0.1262	0.1984	0.1213	0.1076	L22-25	0.556
	50%	0.01	0.0926	0.1514	0.0538	0.1033	L22-25	0.754
		300	0.0631	0.0662	0.1442	0.1649	L22-25	0.642
	90%	0.01	0.0824	0.0999	0.1071	0.0749	L22-25	0.357
		300	0.0854	0.1638	0.0903	0.0162	L22-25	0.69
AB	5%	——	0.1348	0.1202	0.1398	0.1834	L22-25	0.743
	20%	——	0.1949	0.1214	0.1722	0.1468	L22-25	0.654
	50%	——	0.1725	0.0251	0.0348	0.1938	L22-25	0.711
	90%	——	0.0134	0.1055	0.041	0.0553	L22-25	0.118
ABC	5%	——	0.1873	0.0482	0.0607	0.151	L22-25	0.651
	20%	——	0.0629	0.0365	0.039	0.0176	L22-25	0.408
	50%	——	0.0404	0.0602	0.0141	0.0534	L22-25	0.254
	90%	——	0.089	0.1066	0.1967	0.0069	L22-25	0.285

Table 2. Analysis of failure analysis in provincial grid simulation system

Fault type	Fault position	R_f/Ω	$dI18$	$dI19$	Fault line	Distance error (%)
AG	5%	0.01	0.1146	0.1321	L18-19	0.245
		300	0.0798	0.1203	L18-19	0.304
	20%	0.01	0.1975	0.0699	L18-19	0.388
		300	0.0932	0.0388	L18-19	0.237
	50%	0.01	0.1693	0.1324	L18-19	0.334
		300	0.1127	0.0612	L18-19	0.333
	90%	0.01	0.0507	0.0237	L18-19	0.248
		300	0.0175	0.1029	L18-19	0.343
ABG	5%	0.01	0.0692	0.0513	L18-19	0.263
		300	0.1817	0.1776	L18-19	0.241
	20%	0.01	0.1953	0.1246	L18-19	0.23
		300	0.0607	0.0942	L18-19	0.233
	50%	0.01	0.0188	0.1197	L18-19	0.376
		300	0.1378	0.1008	L18-19	0.201
	90%	0.01	0.0628	0.1442	L18-19	0.282
		300	0.1314	0.1694	L18-19	0.228
AB	5%	——	0.1957	0.0218	L18-19	0.272
	20%	——	0.071	0.1331	L18-19	0.254
	50%	——	0.1494	0.0686	L18-19	0.129
	90%	——	0.1305	0.1221	L18-19	0.289
ABC	5%	——	0.1855	0.1989	L18-19	0.398
	20%	——	0.129	0.0912	L18-19	0.309
	50%	——	0.1706	0.1612	L18-19	0.16
	90%	——	0.1611	0.0499	L18-19	0.097

3.3. Analysis of application examples of power grid wide area protection system

3.3.1. Operation of a regional power grid. In this paper, the empirical case study selected a city regional power grid as the object, through the recent years of the regional power grid operation process occurred during the fault record statistics, due to faults and triggered by the occurrence of regional blackout accidents in the statistical results shown in Table 3. A comparison of the data from 2015 to 2021 shows that the average number of regional power outages caused by grid faults is about 13 times per year. It is due to the limitations of the traditional grid relay protection system that it cannot intelligently remove fault-induced regional blackouts. The differential protection in the traditional relay protection system of the regional power grid can realize the protection of the entire length of the line, but there is no upper and lower level cooperation. Quick-break protection can be up and down with, but the number of system distribution level for 3 or even 4, quick-break with high sensitivity, it is easy to overstep trip, and the protection range is only 25% ~ 40% of the line. The power supply in the traditional relay protection system cannot maintain the isolated network operation of the regional grid, and in the case of double circuit tripping of the regional grid feeder, the generating units in the system will be unlisted, and cannot remove part of the load to maintain the security load of the grid. In addition, most of the distribution stations do not have fault recording system, even if there is a good or bad degree, the integrity of the recording can not be guaranteed, which is not conducive to stability and accident analysis. It can be seen that the regional power grid belongs to a typical case, and the analysis results obtained by applying the grid wide-area protection system designed in this paper to the regional power grid are representative.

Table 3. Statistical analysis of the area failure of the grid

Month/year	2015	2016	2017	2018	2019	2020	2021
January	2	3	1	0	2	0	1
February	0	2	0	1	2	3	1
March	1	1	2	2	0	0	1
April	2	0	0	0	1	3	1
May	2	0	0	1	3	2	0
June	0	1	0	1	1	1	0
July	0	0	1	2	1	3	2
August	0	2	0	0	1	0	1
September	3	0	0	3	1	2	0
October	0	0	2	1	0	2	2
November	1	0	0	1	2	3	0
December	3	0	0	0	3	2	1
Total	14	9	6	12	17	21	10
	89						

3.3.2. Results of the application of the grid wide protection system. In this study, the designed grid wide-area protection system was applied to the regional power grid, and it was found that the wide-area protection system monitored the overall grid system from the time period of January 2022-January 2023, and the specific monitoring results are shown in Table 4. The grid wide-area protection system found a total of 10 line faults and was able to accurately monitor the point where the line fault occurred, assisting the staff to repair the faults in time and preventing large-scale power outages from occurring.

Table 4. Grid failure detection results

Time	Fault line	Voltage rating (kV)	Fault type	Failure phase	The departure of the line	Fault line	Failure phase	Trip avoidance
2022.4.5	Hengjing line Ⅱ	220	Ground	A	0.51	√	√	√
2022.5.2	Wuma line	220	Interphase	AB	0.65	√	√	√
2022.5.19	Cangyi line	220	Ground	BC	0.68	√	√	√
2022.6.8	Sungao line	220	Interphase	ABC	0.41	√	√	√
2022.8.9	Cangyi line	220	Interphase	A	0.72	√	√	√
2022.9.6	Changqing line Ⅱ	220	Ground	A	0.87	√	√	√
2022.10.8	Changqing line Ⅰ	220	Ground	BC	0.35	√	√	√
2022.10.19	Fangci line	220	Interphase	AB	0.22	√	√	√
2022.11.14	Beiqing line Ⅰ	220	Ground	ABC	0.65	√	√	√
2022.12.3	Antong line	220	Interphase	C	0.81	√	√	√

4. Conclusion

In this paper, a data fusion strategy is proposed to process large-scale data in the distribution network system, and the performance test results of the distributed algorithm data fusion processing module are used to find that the processing of data in the distribution network by the power grid wide area protection system can maintain a stable state when there are 35,000 data per second, and the back pressure (0.29%-15.64%, 0%-0.09%) and high load (4.88%-22.11%, 2.93%-7.04%) of the processing end and the output end are also good. There is no high back pressure and high load situation, and the high processing speed is maintained while maintaining the low latency of data processing. The verification of the grid fault detection method in simulation experiments shows that it can accurately identify the nodes on both sides of the faulty line in the grid system and detect the faulty line, and the error in the distance of locating the fault point is 0.458% and 0.266% on average, which will not cause miscalculation of the normal lines in the region. In the example application of a city regional power grid system, the grid wide area protection system proposed in this paper is able to monitor the existence of circuit faults in the power grid system and warn the staff in time to stabilize the good operation of the power grid system.

Compared with the traditional relay protection system, the proposed wide-area protection system has strong advantages in preventing the expansion of faults and accidents in the power grid system and taking timely measures, which theoretically and practically proves that the system is very beneficial to the maintenance of the safe and stable operation of the power grid.

Funding

This work was supported by Special Topic 9 of Guangdong's "15th Five Year Plan" New Power System Planning: Research on Guangdong Power Grid's "15th Five Year Plan" Secondary System Planning (Project Number: 031000QQ00240021).

References

[1] Wang, W., Sun, K., Zeng, C., Chen, C., Qiu, W., You, S., & Liu, Y. (2021). Information and communication infrastructures in modern wide-area systems. Wide Area Power Systems Stability, Protection, and Security, 71-104.

[2] Kavya, G., Devi, M. M., & Geethanjali, M. (2018, March). Wide area backup protection scheme using optimal PMUs. In 2018 National Power Engineering Conference (NPEC) (pp. 1-5). IEEE.

[3] Rezapour, H., Jamali, S., & Siano, P. (2024). Wide-Area Protection System for Radial Smart Distribution Networks. Applied Sciences, 14(11), 4862.

[4] Eissa, M. M. (2018). Developing wide area phase plane primary protection scheme "WA4PS" for complex smart grid system. International Journal of Electrical Power & Energy Systems, 99, 203-213.

[5] Tabatabaei, N. M., Kabalci, E., & Bizon, N. (Eds.). (2019). Microgrid architectures, control and protection methods. Springer.

[6] Goli, R. K., Pappula, S. K., Sankabathula, K., Matta, H., Agrapu, D. P., & Rishi, K. (2024, July). Wide Area Power System Network Protection using Wavelet-Machine Learning Approach. In 2024 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT) (pp. 1-6). IEEE.

[7] Khalid, H., & Shobole, A. (2021). Existing developments in adaptive smart grid protection: A review. Electric Power Systems Research, 191, 106901.

[8] Shen, S., Wang, H., Jiang, K., Zhu, L., Gan, C., Markham, P. N., ... & He, B. (2018). Regional area protection scheme for modern distribution system. IEEE Transactions on Smart Grid, 10(5), 5416-5426.

[9] Mallikarjuna, B., Samir, S. D., Reddy, M. J. B., & Mohanta, D. K. (2017). An adaptive supervised wide-area backup protection scheme for transmission lines protection. Protection and Control of Modern Power Systems, 2(3), 1-16.

[10] Singh, V. K., & Govindarasu, M. (2021). A cyber-physical anomaly detection for wide-area protection using machine learning. IEEE Transactions on Smart Grid, 12(4), 3514-3526.

[11] Chaitanya, B. K., Yadav, A., & Pazoki, M. (2020). Wide area monitoring and protection of microgrid with DGs using modular artificial neural networks. Neural Computing and Applications, 32(7), 2125-2139.

[12] Dileep, G. J. R. E. (2020). A survey on smart grid technologies and applications. Renewable energy, 146, 2589-2625.

[13] Anderson, P. M., Henville, C. F., Rifaat, R., Johnson, B., & Meliopoulos, S. (2022). Power system protection. John Wiley & Sons.

[14] Miron, A., Popa, C., Bobric, C. E., & Dragomir, M. (2018, October). Considerations regarding implementing wide area monitoring, protection and control in north–east of Romanian power grid. In 2018 International Conference and Exposition on Electrical And Power Engineering (EPE) (pp. 0867-0872). IEEE.

[15] Blaabjerg, F., Yang, Y., Yang, D., & Wang, X. (2017). Distributed power-generation systems and protection. Proceedings of the IEEE, 105(7), 1311-1331.

[16] Telukunta, V., Pradhan, J., Agrawal, A., Singh, M., & Srivani, S. G. (2017). Protection challenges under bulk penetration of renewable energy resources in power systems: A review. CSEE journal of power and energy systems, 3(4), 365-379.

[17] Noreen, U., Bounceur, A., & Clavier, L. (2017, May). A study of LoRa low power and wide area network technology. In 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP) (pp. 1-6). IEEE.

[18] Ashok, V., Yadav, A., & Abdelaziz, A. Y. (2021). A comprehensive review on wide-area protection, control and monitoring systems. Wide Area power systems stability, protection, and security, 1-43.

[19] Liu, C., Tao, J., Lou, Z., & Shao, Q. (2018, March). Reliable Routing Algorithm for Wide Area Protection. In 2018 2nd International Conference on Electrical Engineering and Automation (ICEEA 2018) (pp. 196-200). Atlantis Press.

[20] Wang, Q., Bo, Z., Zhao, Y., Ma, X., Zhang, M., Zheng, H., & Wang, L. (2019). Integrated wide area protection and control for power grid security. CSEE Journal of Power and Energy Systems, 5(2), 206-214.

[21] Ashok, A., Govindarasu, M., & Wang, J. (2017). Cyber-physical attack-resilient wide-area monitoring, protection, and control for the power grid. Proceedings of the IEEE, 105(7), 1389-1407.

[22] Eissa, M. M. (2018). Challenges and novel solution for wide-area protection due to renewable sources integration into smart grid: an extensive review. IET Renewable Power Generation, 12(16), 1843-1853

[23] Vahidi, S., Ghafouri, M., Au, M., Kassouf, M., Mohammadi, A., & Debbabi, M. (2023). Security of wide-area monitoring, protection, and control (WAMPAC) systems of the smart grid: A survey on challenges and opportunities. IEEE Communications Surveys & Tutorials, 25(2), 1294-1335.

[24] Wu, S. (2017). An adaptive limited wide area differential protection for power grid with micro-sources. Protection and Control of Modern Power Systems, 2(3), 1-9.

[25] Kundu, P., & Pradhan, A. K. (2017). Power network protection using wide-area measurements considering uncertainty in data availability. IEEE Systems Journal, 12(4), 3358-3368.

[26] Nalini, A., Percis, E. S., Jijina, G. O., Raja, M., & Sendilvelan, S. (2020, December). Supervisory protection scheme for a wide area 400KV power network. In IOP Conference Series: Materials Science and Engineering (Vol. 993, No. 1, p. 012072). IOP Publishing.

[27] Al-Maitah, K., & Al-Odienat, A. (2024). Wide area protection in modern power systems. In Power System Protection in Future Smart Grids (pp. 5-29). Academic Press.

[28] Jinjun Tang,Guohui Zhang,Yinhai Wang,Hua Wang & Fang Liu. (2015). A hybrid approach to integrate fuzzy C-means based imputation method with genetic algorithm for missing traffic volume data estimation. Transportation Research Part C29-40.

[29] Jackson Barth,Yuqiu Yang,Guanghua Xiao & Xinlei Wang. (2024). MetaNorm: Incorporating meta-analytic priors into normalization of NanoString nCounter data. Bioinformatics (Oxford, England).

[30] Yongze Li,Jin Ma,Dongyang Fu,Jiajun Yuan & Dazhao Liu. (2024). Mangrove Extraction Algorithm Based on Orthogonal Matching Filter-Weighted Least Squares. Sensors(22),7224-7224.