Scientific Research Publishing

Energy and Power Engineering, 2017, 9, 19-29
http://www.scirp.org/journal/epe
ISSN Online: 1947-3818
ISSN Print: 1949-243X

Fault Identification of Power Grid Based on Wide-Area Differential Current and K-Means Clustering

Hao Wu[1,2], Qunzhan Li[1]

[1]Electrical Engineering Institute, Southwest Jiaotong University, Chengdu, China
[2]Automation and Electronic Information Engineering, Sichuan University of Science & Engineering, Zigong, China
Email: 11305076@qq.com

How to cite this paper: Wu, H. and Li, Q.Z. (2017) Fault Identification of Power Grid Based on Wide-Area Differential Current and K-Means Clustering. *Energy and Power Engineering*, **9**, 19-29.
https://doi.org/10.4236/epe.2017.94B003

Received: March 9, 2017
Accepted: March 30, 2017
Published: April 6, 2017

Abstract

A new method of fault domain identification is proposed based on K-means clustering analysis theories using the wide-area information of power grid. In the method, the node Intelligent Electronic Device (IED) associated domain is defined, and the relationship of positive sequence current fault component for the association domain boundaries is sought, then the conception of positive sequence fault component differential current for node IED association domains is introduced. The information of the positive sequence fault component differential current gathered by node IEDs is selected as the object of K-means clustering. The node IEDs of fault associated domains can be classified into one category, and the node IEDs of non-fault associated domains are classified into another category. With the fault area minimum principle, the group of node IEDs about fault associated domains can be obtained. The overlap of fault associated domains for different nodes is the fault area. A large number of simulations show that the algorithm proposed can identify fault domains with high accuracy and no influence by the operating mode of the system and topological changes.

Keywords

Positive Sequence Fault Component Differential Current, K-Means Clustering, Fault Association Domain, The Node IED, Fault Domain Identification

1. Introduction

With the increasingly complex structure and the continuously extended scale of power grid, the traditional backup protection based on local information can not satisfy requirements of complex and various operation modes of power grid. The rapid development of computer technologies and the wide-area measurement technologies make global information being introduced into protection possible. In recent years, extensive researches on wide-area backup protection have carried on at home and abroad, mainly concentrating in tripping strategies and fault areas identification of wide-area protection, etc. [1] [2] [3] [4].

The wide-area relay protection system given in reference [5] is based on the current differential principle, and problems such as protection domain division rules for wide-area protection are discussed in the reference. A wide-area current differential protection principle based on multi Agent is proposed in the reference [6], where an expert system is used to realize the area division of current differential protection, and the wide-area differential protection is achieved through coordination between protection Agents.

In order to further study the application of artificial intelligence algorithm in wide-area backup protection and improve the accuracy of fault identification with wide-area information under different working conditions, a new method for identifying failure areas of power grid based on k-means clustering according to wide-area positive sequence fault component differential current information is proposed on the basis of previous studies.

2. K-Means Clustering

The K-means clustering algorithm is to cluster based on the objective function of a prototype. In the algorithm, the sum of distances from data to corresponding clustering centers is the optimized objective function and adjusting rules for iterative operations are obtained by finding the extremum solution of the function. The mean value of data samples of each cluster subset is selected as the clustering center of the corresponding cluster. The main idea of the algorithm is to divide data into different classes through iteration processes, and makes the clustering criterion function used to evaluate the clustering performance to achieve its optimum, so that each cluster generated can be compact inside and independent to others. The number k of clusters and a database contains n objects are needed to be input first of all, and then n objects are divided into k clusters, which can make the minimum square error criterion [7] [8]. For a given data set $Y = (Y_1, Y_2, \ldots, Y_n,)$, processes of K-means clustering algorithm are as follows [8] [9]:

1) Select K initial clustering centers: $C_1, C_2, \ldots, C_k$.

2) Calculate the distance d from every data to each clustering center, and divide the data to the corresponding cluster I_j with the minimum distance d.

$$d = \left\| y_i - C_j \right\| = \min_{1<r<k} \left\| y_i - C_r \right\| \quad (i \in \{1, 2, \ldots, n\}, 0 < j < k).$$

3) Calculate the new clustering center vector C_j of each cluster,

$$C_j = [C_{j1}, C_{j2}, \ldots, C_{jp}]^T, C_{jq} = \frac{1}{N_j} \sum_{y \in I_j} y (j = 1, 2, \ldots, k)$$

In which, q is the attribute number of data, N_j is the number of data that the j-th cluster I_j included.

4) Repeat processes 2 and 3, until each cluster is no longer changes.

3. Fault Domain Identification Based on K-Means

3.1. The Analysis of Clustering Objects

Node IEDs of power grid are installed at substation nodes, corresponding to substations. Each node IED has the same status, whose works are mainly to collect electric information sent from related line IEDs, and upload them to wide-area decision center after preliminary processing. Line IEDs mainly acquire positive sequence current fault component information at installation places, and upload the information to the corresponding grid node IEDs. Fault domains of power grid can be identified by the fault recognition algorithm to process the data uploaded by node IEDs. The associated domain of node IED is defined in this paper. As shown in **Figure 1**, the domain surrounded by dotted line 2 is the associated domain of the node IED_{B2}, which consists of line L_1, L_2 and bus B_2 with two boundary IEDs, IED_1 and IED_4. Similarly, associated domains of other nodes are domains surrounded by dotted lines 1, 3, 4, 5.

The positive sequence fault component differential current of the node IED associated domain is defined as the sum of phasors of all positive sequence fault current components measured at boundary line IEDs. For example, at the node IED_{B2}, the positive sequence fault component differential current $\Delta\dot{I}_{B2}$ of the node associated domain is $\Delta\dot{I}_{B2} = \Delta\dot{I}_1 + \Delta\dot{I}_4$. Any fault occurs in the associated domain, the value of $\Delta\dot{I}_{B2}$ which is the total positive sequence fault current component will be very large and associated domain ② is the fault associated domain for the moment. When normal operation or external fault of the associated domain, the positive sequence fault component differential current $\Delta\dot{I}_{B2}$ whose ideal value is zero is actually an unbalanced current with small value. Thus, when a short-circuit fault occurs at K_1 point in the **Figure 1**, domains ①, ④ and ⑤ are non-fault domains, domains ② and ③ are fault domains and corresponding fault nodes are B_2 and B_3. Therefore, it can be assured that the fault domain is the overlapped part of two node IEDs associated domains (as the shaded part shown in **Figure 1**), that is the line L_2.

When a fault occurs at bus B_3 in the **Figure 1**, for domain ① we have $\Delta\dot{I}_{B1} = \Delta\dot{I}_1 + \Delta\dot{I}_2 \approx 0$, for domain ② we have $\Delta\dot{I}_{B2} = \Delta\dot{I}_1 + \Delta\dot{I}_4 \approx 0$, for domain ③ we have $\Delta\dot{I}_{B3} = \Delta\dot{I}_3 + \Delta\dot{I}_6 = \Delta\dot{I}_{B3k}$, for domain ④ we have $\Delta\dot{I}_{B4} = \Delta\dot{I}_5 + \Delta\dot{I}_8 \approx 0$, for domain ⑤ we have $\Delta\dot{I}_{B5} = \Delta\dot{I}_7 + \Delta\dot{I}_8 \approx 0$. It can be

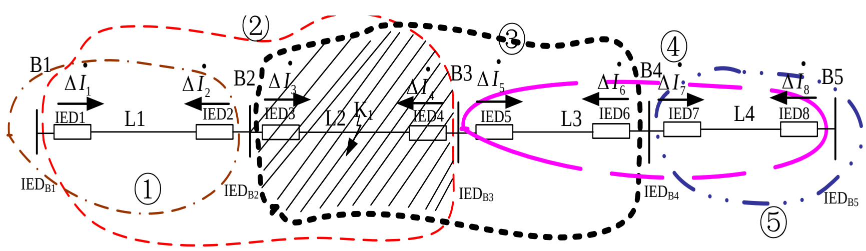


Figure 1. IED associated domain analysis.

assured that the domain ③ is the fault associated domain. Hence, when a single independent fault associated domain appears, the bus in the associated domain is thought to be failed.

Clustering status characteristic values selected in this paper are the RMS ΔI_1 in the first cycle and the RMS ΔI_2 in second cycle of the positive sequence fault component differential current at the node *IED* associated domain after the fault, that is, the status information vector for the *i*-th node *IED* is $IED_{Bi} = [\Delta I_{i1} \quad \Delta I_{i2}]$. If there are n substations (n nodes) in power grid, the wide-area information matrix A $(n \times 2)$ could be

$$A = \begin{bmatrix} IED_{B1} \\ IED_{B2} \\ \vdots \\ \vdots \\ IED_{Bn} \end{bmatrix} = \begin{bmatrix} \Delta I_{11} & \Delta I_{12} \\ \Delta I_{21} & \Delta I_{22} \\ \vdots & \vdots \\ \vdots & \vdots \\ \Delta I_{n1} & \Delta I_{n2} \end{bmatrix}$$

Row vectors of the matrix A correspond to node *IED* status information, that is clustering objects of K-means.

3.2. The Fault Domain Identification of Power Grid Based on K-Means

The wide-area information matrix A of power grid is the input of K-means clustering for the clustering analysis of the associated domain of each grid node. Still the circuit in **Figure 1**, for example, the wide-area information matrix A is

$$A = \begin{bmatrix} IED_{B1} \\ IED_{B2} \\ IED_{B3} \\ IED_{B4} \\ IED_{B5} \end{bmatrix} = \begin{bmatrix} \Delta I_{11} & \Delta I_{12} \\ \Delta I_{21} & \Delta I_{22} \\ \Delta I_{31} & \Delta I_{32} \\ \Delta I_{41} & \Delta I_{42} \\ \Delta I_{51} & \Delta I_{52} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \Delta I_{21} & \Delta I_{22} \\ \Delta I_{31} & \Delta I_{32} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

where, nodes corresponded to fault domains are IED_{B2} and IED_{B3}, nodes corresponded to non-fault domains are IED_{B1}, IED_{B4} and IED_{B5}. Characteristic information of associated node IEDs in fault domains are all the whole fault current at fault points in domains with similar vector information. And all characteristic information of associated node IEDs in non-fault domains are merely unbalanced currents with small values and their vector information are similar. But vectors information are different vigorously between node IEDs of fault domain s and non-fault domain. Based on a large number of simulations, wide-area information samples acquired by node IEDs are divided into two groups: the *IED* class of fault domain associated nodes and the *IED* class of non-fault domain associated nodes.

In a large multi-station power system, the principle of minimum fault area is satisfied, based on which, the cluster with the least node *IED* number in clustering results is chosen as the associated node *IED* class of fault domains in this paper. In the class, the overlapped domain of associated fault domains of each node *IED* is thought as the fault domain. If there is no overlapped domain, bus failure at associated node is thought to happen in corresponding fault domain. The process of the fault identification based on K-means algorithm is shown in **Figure 2**.

4. Example Analysis

As shown in **Figure 3**, simulations with the fault identification method based on K-means are carried on IEEE-3 machine 9-node system. Several typical fault situations are analyzed and tested on this paper.

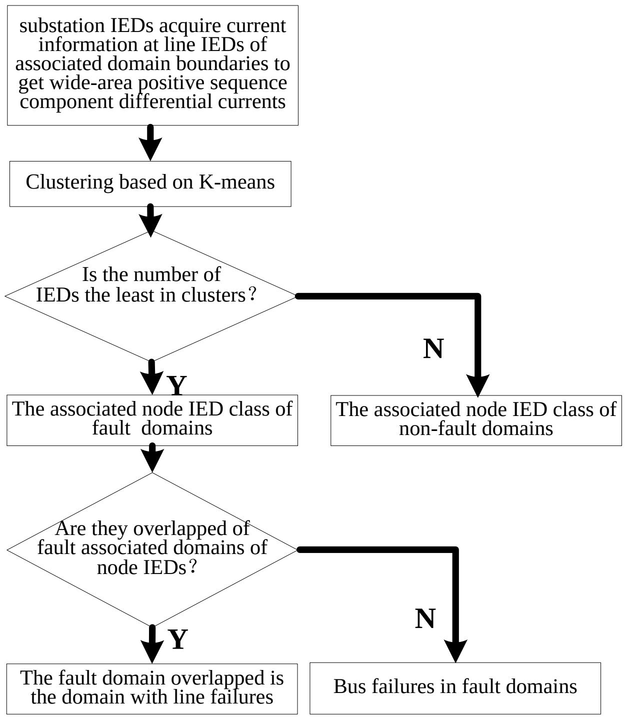


Figure 2. Fault domain identification flow based on K-means.

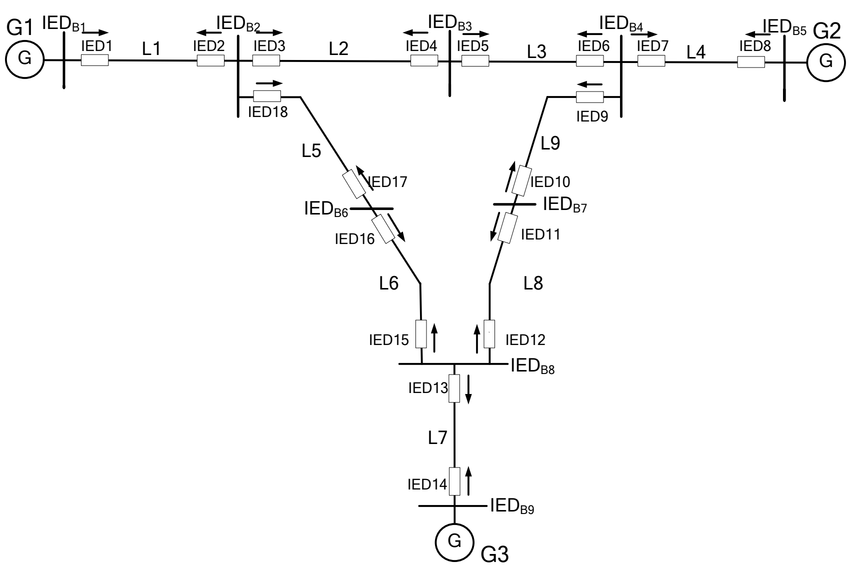


Figure 3. IEEE 3-machine 9-node system.

According to the definition above, the positive sequence fault component differential current of the node *IED* associated domain is referred as the sum of current phasors measured by boundary line IEDs in the associated domain. Calculations of positive sequence fault component differential currents for node IED associated domains of the IEEE-3 machine 9-node system are as shown in **Table 1**.

After calculating all positive sequence fault component differential currents of the node *IED* associated domains, the RMS value ΔI_{i1} in first circle and the RMS value ΔI_{i2} in second circle of differential currents after fault are selected as wide-area information vector for the *i*-th node IED_{Bi}. Hence, the node IED wide-area information matrix A (9×2) of IEEE-3 machine 9-node system is represented as

$$A=\begin{bmatrix} IED_{B1} \\ IED_{B2} \\ IED_{B3} \\ IED_{B4} \\ IED_{B5} \\ IED_{B6} \\ IED_{B7} \\ IED_{B8} \\ IED_{B9} \end{bmatrix}=\begin{bmatrix} \Delta I_{11} & \Delta I_{12} \\ \Delta I_{21} & \Delta I_{22} \\ \Delta I_{31} & \Delta I_{32} \\ \Delta I_{41} & \Delta I_{42} \\ \Delta I_{51} & \Delta I_{52} \\ \Delta I_{61} & \Delta I_{62} \\ \Delta I_{71} & \Delta I_{72} \\ \Delta I_{81} & \Delta I_{82} \\ \Delta I_{91} & \Delta I_{92} \end{bmatrix}$$

4.1. A Fault Occurs at Line L_9

Assume three-phase short circuit fault occurs at line L_9, the wave of positive sequence fault component differential currents measured at part node IEDs is shown in **Figure 4**.

The RMS values ΔI_{i1} in first circle and the RMS values ΔI_{i2} in second circle of positive sequence fault component differential currents in the associated domain of each node IED are as shown in **Table 2**.

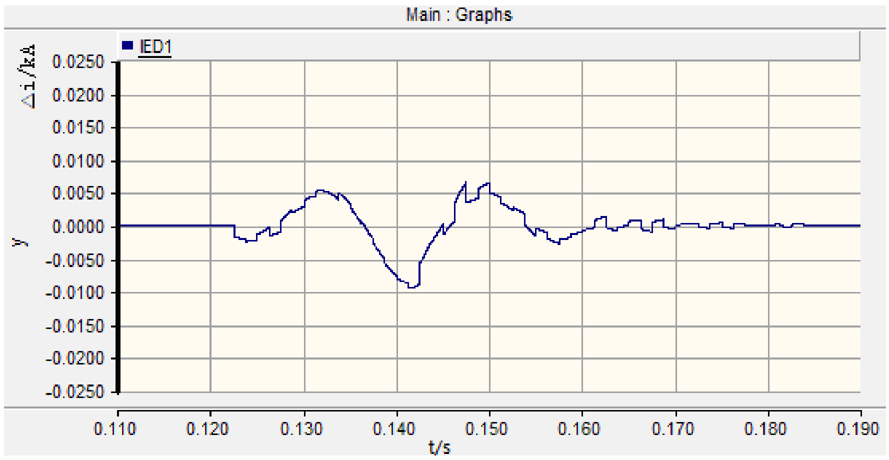
Therefore, the wide-area information matrix A (9×2) of the IEEE-3 machine 9-node system is represented as

Table 1. Calculations of positive sequence fault component differential currents.

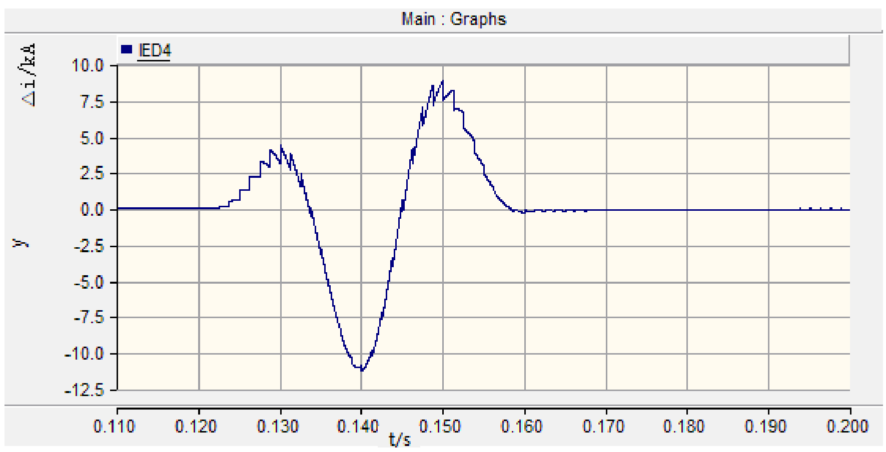
Node IED	Calculations of differential currents in associated domains	Node IED	Calculations of differential currents in associated domains
IED_{B1}	$\Delta\dot{I}_{B1}=\Delta\dot{I}_{1}+\Delta\dot{I}_{2}$	IED_{B6}	$\Delta\dot{I}_{B6}=\Delta\dot{I}_{18}+\Delta\dot{I}_{15}$
IED_{B2}	$\Delta\dot{I}_{B2}=\Delta\dot{I}_{1}+\Delta\dot{I}_{4}+\Delta\dot{I}_{17}$	IED_{B7}	$\Delta\dot{I}_{B7}=\Delta\dot{I}_{12}+\Delta\dot{I}_{9}$
IED_{B3}	$\Delta\dot{I}_{B3}=\Delta\dot{I}_{3}+\Delta\dot{I}_{6}$	IED_{B8}	$\Delta\dot{I}_{B8}=\Delta\dot{I}_{16}+\Delta\dot{I}_{11}+\Delta\dot{I}_{14}$
IED_{B4}	$\Delta\dot{I}_{B4}=\Delta\dot{I}_{5}+\Delta\dot{I}_{8}+\Delta\dot{I}_{10}$	IED_{B9}	$\Delta\dot{I}_{B9}=\Delta\dot{I}_{13}+\Delta\dot{I}_{14}$
IED_{B5}	$\Delta\dot{I}_{B5}=\Delta\dot{I}_{8}+\Delta\dot{I}_{7}$		

Table 2. The RMS values of positive sequence fault component differential currents at each node IED.

Node *IED*	$\Delta I_{i1}(kA)$	$\Delta I_{i2}(kA)$
IED_{B1}	0.003262	0.00434
IED_{B2}	0.04379	0.049084
IED_{B3}	0.039394	0.052657
IED_{B4}	4.650301	5.788424
IED_{B5}	0.019525	0.026133
IED_{B6}	0.043009	0.056496
IED_{B7}	4.70755	5.794702
IED_{B8}	0.043522	0.05657
IED_{B9}	0.015611	0.020266



(a)



(b)

Figure 4. The wave of positive sequence fault component differential currents at some node IEDs. (a) The wave of positive sequence fault component differential currents at node IED_{B1}. (b) The wave of positive sequence fault component differential currents at node IED_{B4}.

$$A = \begin{bmatrix} IED_{B1} \\ IED_{B2} \\ IED_{B3} \\ IED_{B4} \\ IED_{B5} \\ IED_{B6} \\ IED_{B7} \\ IED_{B8} \\ IED_{B9} \end{bmatrix} = \begin{bmatrix} 0.003262 & 0.00434 \\ 0.04379 & 0.049084 \\ 0.039394 & 0.052657 \\ 4.650301 & 5.788424 \\ 0.019525 & 0.026133 \\ 0.043009 & 0.056496 \\ 4.70755 & 5.794702 \\ 0.043522 & 0.05657 \\ 0.015611 & 0.020266 \end{bmatrix}$$

Row vectors of the matrix are objects analyzed according to K-means clustering algorithm. The dimension of sample characteristic values is m =2, the number of data samples is n = 9, and the initial cluster number is h = 2. Select randomly the 1-th and 6-th rows as initial clustering centers, the class centroid coordinate matrix C of two classes is

$$C = \begin{bmatrix} 1.7637 & 1.7638 \\ -0.5039 & -0.5039 \end{bmatrix}$$

The distance sum vector in classes is SUMD = [0.00039 0.0008]

The distance matrix D of each data to their class center is

$$D = \begin{bmatrix} 10.222489 & 0.000095 \\ 0.0000196 & 10.354075 \\ 10.237311 & 0.000056 \\ 10.233981 & 0.000066 \\ 0.000196 & 0.216211 \\ 10.347862 & 0.000096 \\ 10.403895 & 0.000342 \\ 10.328666 & 0.000046 \\ 10.221224 & 0.000099 \end{bmatrix}$$

The outline of K-means clustering is as shown in **Figure 5**.

Clustering results are as shown in **Table 3**.

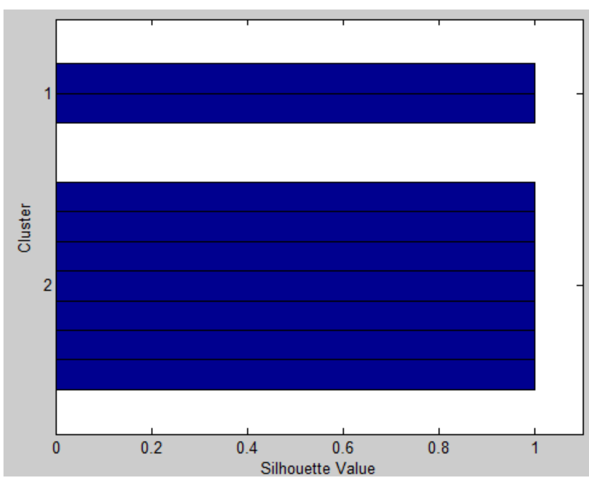


Figure 5. The outline of K-means clustering.

Table 3. K-means classification for L_9 fault.

Class 1	Class 2
"IED_{B7}"	"IED_{B6}"
"IED_{B4}"	"IED_{B3}"
	"IED_{B2}"
	"IED_{B9}"
	"IED_{B1}"
	"IED_{B5}"
	"IED_{B8}"

According clustering results, the wide-area information of 9 node IEDs are divided into two classes, in which the one with least nodes are identified as the node *IED* class of fault associated domains according to the algorithm proposed. As in **Table 3**, class 1 is the node *IED* class of fault associated domains. The overlapped domain of node IEDs associated domains is where faults occur. If there is no overlapped domain, a bus fault occurs in the domain. In the class 1, the associated domains of IED_7 and IED_4 are overlapped at line L_9, then line L_9 is the domain where the fault happens.

4.2. A Fault Occurs at Bus B2

Assume AC two-phase to ground fault occurs at bus B_2, the node *IED* wide-area information matrix A (9×2) obtained accordingly is

$$A=\begin{bmatrix} IED_{B1} \\ IED_{B2} \\ IED_{B3} \\ IED_{B4} \\ IED_{B5} \\ IED_{B6} \\ IED_{B7} \\ IED_{B8} \\ IED_{B9} \end{bmatrix}=\begin{bmatrix} 0.008099 & 0.008231 \\ 4.349681 & 5.516054 \\ 0.041813 & 0.044519 \\ 0.037209 & 0.029788 \\ 0.00867 & 0.009265 \\ 0.035244 & 0.0371 \\ 0.012672 & 0.013912 \\ 0.015652 & 0.016637 \\ 0.00448 & 0.004777 \end{bmatrix}$$

Clustering results are as shown in **Table 4**. The class 1 with least associated IED number is the node class of the fault associated domain with only one node IED, that is one independent fault associated domain. According to the algorithm, no overlapped area exists, the fault occurs at the bus in the fault associated domain, that is at bus B_2.

4.3. Clustering Analysis under Other Fault Conditions

To test accuracy of the identification algorithm based on K-means, clustering analysis are carried on when faults occur under other fault conditions, results seen in **Table 5**. Experiments shown that the algorithm proposed in this paper can identify fault domains when power grid operates under different modes and with different topology structures.

Table 4. K-means classification for bus B_2 fault.

Class 1	Class 2
"IED_{B2}"	"IED_{B6}"
	"IED_{B7}"
	"IED_{B3}"
	"IED_{B4}"
	"IED_{B9}"
	"IED_{B1}"
	"IED_{B5}"
	"IED_{B8}"

Table 5. Simulation analysis of the fault domain identification based on K-means for different faults.

Real fault element	Class 1	Class 2	Identification results
L_4	"$IEDB_4$" "$IEDB_5$"	"$IEDB_6$" "$IEDB_3$" "$IEDB_9$" "$IEDB_1$" "$IEDB_7$" "$IEDB_8$" "$IEDB_2$"	Line L_4
L_1	"$IEDB_1$" "$IEDB_2$"	"$IEDB_4$" "$IEDB_5$" "$IEDB_9$" "$IEDB_6$" "$IEDB_7$" "$IEDB_8$" "$IEDB_3$"	Line L_1
B_4	"$IEDB_4$"	"$IEDB_6$" "$IEDB_5$" "$IEDB_8$" "$IEDB_1$" "$IEDB_3$" "$IEDB_9$" "$IEDB_2$" "$IEDB_7$"	Bus B_4
B_8	"$IEDB_8$"	"$IEDB_9$" "$IEDB_5$" "$IEDB_6$" "$IEDB_1$" "$IEDB_3$" "$IEDB_4$" "$IEDB_2$" "$IEDB_7$"	Bus B_8
$L_3$is not in operation, B_7 faults	"$IEDB_7$"	"$IEDB_6$" "$IEDB_5$" "$IEDB_8$" "$IEDB_1$" "$IEDB_3$" "$IEDB_4$" "$IEDB_2$" "$IEDB_9$"	Bus B_7
G_2 is not in operation, L_3 faults	"$IEDB_3$" "$IEDB_4$"	"$IEDB_6$" "$IEDB_5$" "$IEDB_9$" "$IEDB_1$" "$IEDB_7$" "$IEDB_8$" "$IEDB_2$"	Line L_3

5. Conclusions

A new method for fault domain identification based on wide-area positive sequence fault component differential currents and K-means algorithm is proposed in this paper. Wide-area information of node IEDs are clustered by K-means according to the fault domain minimum principle to assure the class with least node IEDs to be the associated node class of fault associated domains. The fault identification can be realized by finding the overlapped area of fault associated domains of those node IEDs.

Simulation results show that fault domains can be identified correctly when the operational mode of power grid changes, such as one line or one source is not in operation. Fault domain identification based on wide-area status information and the intelligent algorithm are discussed in this paper, which provides a new way to diagnose faults in grid.

Acknowledgements

The research work was financially supported by the artificial intelligence key laboratory of Sichuan province (2014RYY05，2015RYY01).

References

[1] He, Z.Q., Zhang, Z., Yin, X.G. and Cheng, W. (2010) Overview of Power System Wide Area Protection. *Electric Power Automation Equipment*, **30**, 125-130. https://doi.org/10.1109/TPWRD.2009.2035202

[2] Novosel, D., Bartok, G., Henneberg, G., *et al.* (2010) IEEE PSRC Report on Performance of Relaying During Wide-Area Stressed Conditions. *IEEE Transactions on Power Delivery*, **25**, 3-16.

[3] Xu, T.Q., Yin, X.G. and You, D.H. (2009) Analysis on Functionality and Feasible Structure of Wide Area Protection System. *Power System Protection and Control*, **37**, 93-97.

[4] Yang, Z.L., Shi, D.Y. and Duan, X.Z. (2008) Wide-Area Protection System Based on Direction Comparison Principle. *Proceedings of the CSEE*, **28**, 87-93.

[5] Cong, W., Pan, Z.C., Zhao, J.G., *et al.* (2006)A Wide Area Relaying Protection Relaying System Based on Current Differential Protection Principle. *Power System Technology*, **30**, 91-95.

[6] Su, S., Duan, X.Z., Zeng, X.J., *et al.* (2005) A Multi-Agent Based Wide-Area Current Differential Protection System. *Power System Technology*, **29**, 15-19.

[7] Bai, X.F. and Jiang, G.D. (2010) Load Modeling Based on Improved K-Means Clustering Algorithm and Its Application. *Electric Power Automation Equipment*, **30**, 80-83.

[8] Liu, Y.H., Xi, Y.Y., Wu, H.Y., *et al.* (2014) Application of K-Means Clustering Algorithm in Oil-Water Interface Detection. *Control and Instruments in Chemical Industry*, **41**, 19-21.

[9] Li, C.C., Wang, B., Wang, J., *et al.* (2012) Extracting Vein of Leaf Image Based on K-Means Clustering. *Transactions of the Chinese Society of Agricultural Engineering*, **28**, 157-162.