Power system transmission line fault diagnosis based on combined data analytics

Honghao Wu, Junyong Liu, Youbo Liu, Gao Qiu
School of Electrical Information and Engineering
Sichuan University
Chengdu, China
Honghao Wu@stu.scu.edu.cn

Gareth A. Taylor
College of Engineering, Design and Physical Sciences
Brunel University London
London, UK
Gareth.Taylor@brunel.ac.uk

***Abstract*—As a consequence of the recent development of situational awareness technologies for smart grids, the gathering and analysis of data from multiple sources offers a significant opportunity for enhanced fault diagnosis. In order to achieve improved accuracy for both fault detection and classification, a novel combined data analytics technique is presented and demonstrated in this paper. The proposed technique is based on a segmented approach to Bayesian modelling that provides probabilistic graphical representations of both electrical power and data communication networks. In this manner the reliability of both the data communication and electrical power networks are considered in order to improve overall power system transmission line fault diagnosis.**

Keywords—Fault diagnosis; Fault classification; Combined data analytics; Bayesian modelling; Reliability.

I. INTRODUCTION

The traditional primary power grid-centric electrical power system has gradually evolved into a CPS (Cyber Physical System [1]) which regards both power grid and information network as equally important and even closely integrated. Due to the application of advanced Information Communications Technology (ICT) at large scale, it has greatly improved the controllability and observability of the electrical power system. The powerful function of the information system provides technical support to the operation of the power grid, but on the other hand, the failure of information system also leads to even more serious consequences. In 2001, a black out happened to El Paso Electric (EE) in US[3], because the signal degradation of communication system and abnormal delay of communication data caused the malfunction of protection system, which led to the shut-down of the key electric line. The impact of abnormality of communication and information system (including natural fault and malicious attack) on the safe and steady operation of the power grid has started to be seen, and will become even more far-reaching in the future. The power system fault diagnosis can no longer be a simple mode recognition mechanism based on fault feature. The faults of Cyber Physical System include both the uncertain factors [2], (such as refused-operation and mal-operation of protection circuit-breaker), which the physical system originally has, and new issues (such as, information transmission delay, data error, communication failure etc.). The interaction between physical system and information system creates higher demand for the instantaneity of fault diagnosis, and the robustness in the presence of noise and uncertainties.

In the past, a lot of methods [4-6] have already taken into consideration the influence of uncertain factors on fault diagnosis. These methods can reduce the interference caused by error message as far as possible under the circumstance that if the number of correct messages is larger than that of the error messages. Bayesian network [7-8] is exactly an acknowledged pattern recognition method which shows good fault tolerance. In addition, it can describe the logical relation between different elements very well during the modeling. Introducing the redundant information is another important aspect in fault tolerance improvement. Information integration [9-10] is an effective way to filter the error messages by importing more data for observation. However, overlarge redundant information quantity also denotes the increase of interference information, which poses strict requirements on the method itself. Still, without consideration of information system, the improvement of fault tolerance [11] of the fault diagnosis will be very limited. Because single interference of communication system can often result in problems of multiple fault messages, so it is not appropriate for the understanding of traditional “noise”. Therefore, in the view of CPS, the availability of communication path whose power is supported by the physical system and the fault data reliability play important roles in the whole process of fault diagnosis.

This paper proposes a transmission line diagnostic method based on combined data analytics. Chapter II constructs a quick line fault diagnostic model based on the Bayesian network. In chapter III, a method for analyzing the reliability of data and a strategy of data correction are proposed along with an easy way of fault classification. Chapter IV describes the framework of combined data analytics. With case study, Chapter V analyzes the influence of data reliability on fault diagnosis by data mining (SVM) and the effectiveness of the

This work is part of the research project entitled ‘Big Data Based Power System Operation Behavior Recognition in Enabling Characterization and Visualization.’ that is founded by the national natural science foundation of China under grant no. 51437003.

method in this article is approved. The conclusion made in chapter VI summarizes the article and expounds the possibilities for future research.

II. FAST SCREENING OF POTENTIAL LINE FAULTS

A. *Segmented Bayesian Network model for fast screening of potential line faults*

A simple structure of transmission line model is proposed as shown in Fig.1. For the switch state data, the breakers related to the transmission line could be classified into two categories which are the component breaker layer and failure protection layer respectively. The breaker CP1 and CP2 directly connected to the transmission line belongs to the component breaker layer, the other breakers CR1, CR2…CRN on the adjacent lines of the same bus belong to failure protection layer. “MP” and “BP” are the output information of main protection and backup protection respectively. Fault probability could be obtained from the weighted sum of the probability of breaker mode and protection mode. The primary and prior probabilities of breaker and protection nodes are obtained from the expert knowledge of history database. Where ω1 and ω2 are the weight coefficient, which can be adjusted according to the reliability ranking of breaker and protection in the power grid. This paper assume that the reliability of the protection is higher than breaker, ω1 is 0.45 and ω2 is 0.55.

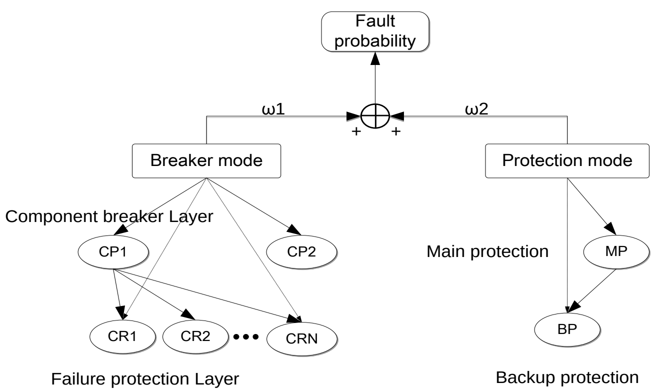


Fig.1 The Bayesian model of transmission line

For the root node of Bayesian network, 0 means normal, 1 represents the state is fault. For the protection and breaker node, 0 represents no action/no tripped, 1 represents action/tripped. The probability of root node breaker mode/protection mode equals to 1 could be calculated on the basis of evidence set E. After input the fault information, the fault probability could be obtained by (1).

$$P(C=1|E=e)=\frac{P(C=1,E=e)}{P(C=0,E=e)+P(C=1,E=e)} \quad (1)$$

Where, C is the root node, E is the practical state of each nodes. The evidence value e is also 0 or 1.

The joint probability is the multiplication of conditional probabilities of each nodes as shown in (2). According to (1) and (2), the probability of root node C can be calculated.

$$\mathrm{P}(X_1,X_2\cdots,X_i)=\prod_{i=1}^{n}P(X_i|Parent(X_i)) \quad (2)$$

Where the node variables of Bayesian network are $X_1, X_2, \cdots, X_i$. And $Parent(X_i)$ is the parent node of X_i.

III. FAULT DETECTION AND CLASSIFICATION

A. *Data communication network for diagnosis*

1) *Reliability analysis of fault data*

As an important segment in power system operation and control, the electric power communication and information reliability is of great significance especially in online fault diagnosis. In order to nip in the bud and actively cope with the possible long delay, information loss and other communication system problems, this paper has considered the monitoring of the communication system and established the corresponding index simultaneously to evaluate the fault data on the basis of the traditional fault analysis data.

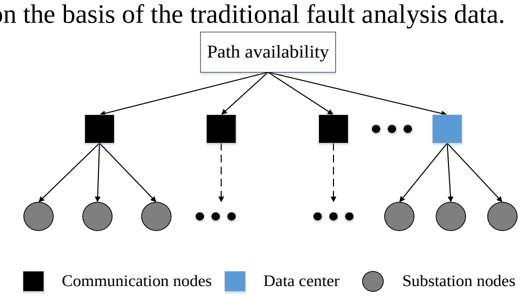


Fig.2 The analytical framework of path availability

In a centralized power system diagnosis system, the communication network can be regarded as a set of nodes and links, and each path is also a set of nodes and links. The reliability of the line fault data can be equivalent to that of the communication path between the data center and the substation information nodes. Suppose that the monitoring data of the substation communication status is ranked highest in the business channel. The monitoring data when uploaded to the data center will not occupy the service channel for fault diagnosis information. In the process of data communication, the effectiveness of the communication path can be illustrated as shown in the Fig.2. The effectiveness of the communication path is related to information nodes as well as the substation nodes supporting power supply for the information nodes. It could be represented by the multiplication of a series conditional probabilities.

$$P_A=P(\mathrm{N1}|\mathrm{S}_1)\cdot P(\mathrm{N2}|\mathrm{S}_2)\cdots P(C|\mathrm{S}_{c1},\mathrm{S}_{c2}...S_{cn}) \quad (3)$$

Where N represents information node of communication system, S is substation node, C is the data center and cn is the number of substations supporting the data center. P is the effectiveness probability of information node with the support of substation node.

$$L_{zone}=\frac{L_c+\sum_{i=1}^{n-1}\frac{1}{N_i}\cdot L_i+\sum_{m,k\in S}(\frac{1}{N_m}+\frac{1}{N_k})\cdot L<m,k>}{1+\sum_{i=1}^{n-1}\frac{1}{N_i}+\sum_{m,k\in L}\frac{1}{N_m}+\frac{1}{N_k}} \quad (4)$$

Where the L_{ZONE} is load factor of the communication subsystem, L_c is the load factor of data center, N_i is the shortest path to the data center. n is the number of nodes, L_i is the load factor of node i, S is the set of communication link, L_{mk} is the load factor of the link between m and k.

$$R_{node}(i)=\frac{N_i}{N_{i\max}}\cdot L_{zone}L_i \tag{5}$$

$$R_{line}(i)=\frac{1}{n}[R_{node}(1)+R_{node}(2)...+R_{node}(n)] \tag{6}$$

$L_{node}(i)$ is the load factor of node i. $R_{node}(i)$ is the reliability that the information transmitted from node i to the data center. $R_{line}(i)$ is the reliability of line, R_{node} (n) is the information node of the line.

2) Correction strategy of missing information

When the data reliability R_{line} is less than 0.95, the data delay or loss will affect the performance of fault diagnosis. So the data need to be modified accordingly. Assuming the state matrix of fault information is:

$$layer\_c=\{x_1,x_2,\ldots,x_n\} \tag{7}$$

$$layer\_f=\{X_1,X_2\ldots X_n\} \tag{8}$$

$$Pm=\{pm_1,pm_2,\ldots pm_n\} \tag{9}$$

$$Pb=\{pb_1,pb_2,\ldots,pb_n\} \tag{10}$$

Layer_c is the state matrix of component breaker layer, layer_f is the failure protection layer, Pm is main protection, and Pb is backup protection. The element of the state matrix is 0 or 1.

For the fault data from a same substation, the modification of data follows the principles below:

- Relevance principle of protection and breaker: if any elements of Pm or Pb is not zero, then one of the breaker layers should have state of 1.
- The consistency principle: all the elements of layer_c or all the column vector of layer_f should have a same state. If more than 50% elements is 1, than the left should be modified to 1.
- If all the protection data is missing, but the breaker data meet the consistency principle, then modification of protection could reference to the fault information from the other side of line. If Pm or Pb from the other side of line is 1 and all the fault information is in line with the relevance principle and consistency principle, then the state of Pm could be revised to 1.

B. *Fault classification based on symmetrical components of reactive power*

The symmetrical components of reactive power [12] are calculated by using the sequential components of voltage and current. As shown in Fig.3, a fault inception is declared, the fault classification could be completed by checking the quantity criterion of $|Q_0|/|Q_2|$ and ΔQ_{12}.

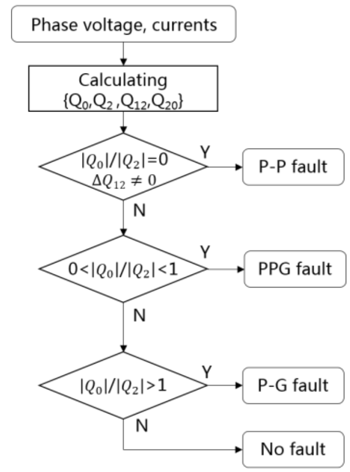


Fig.3 The flow of fault classification

C. *Behavioral analysis of protection and breakers*

After fault classification, input the diagnostic results and the state of protection and circuit breaker to the logic analysis module in Fig.4.

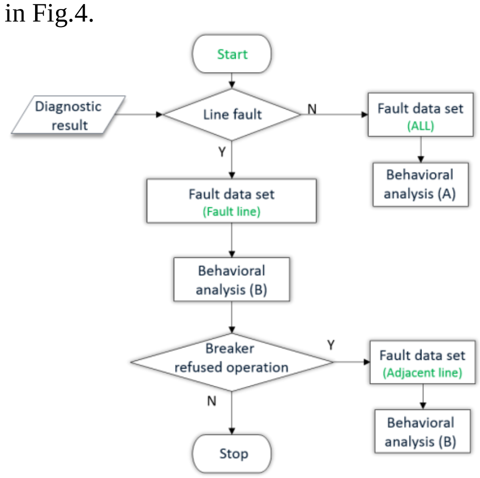


Fig.4 The flow of behavioral analysis

The logic analysis module can classify the state of protection and breaker quickly with the behavior analysis of the sub module in Fig.5.

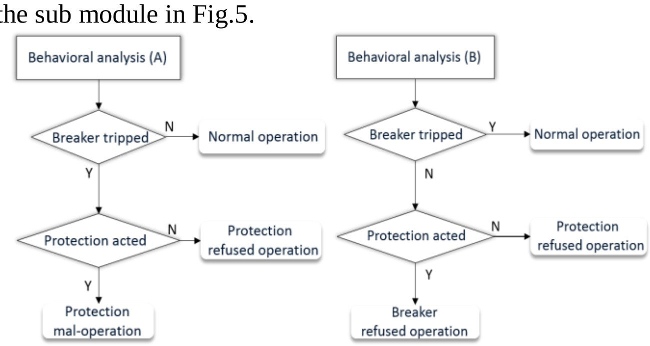


Fig.5 The behavioral analysis module

IV. MBINATED DATA ANALYTICS

A. *Combinated data analytics framework*

As shown in Fig.6, the step of combined data analytics consists of following three parts.

1）By analyzing the topology data of SCADA system, the set of transmission lines could be obtained. Then establish the Bayesian network for each transmission line;

2）According to (3), evaluate the path availability for the nodes of each lines. If the communication path is valid, then calculate the reliabilities of each lines via (4) to (6). If Rline is lower than 0.95, use the correction strategy to revise the fault data;

3）Input the state data of protection and breaker to BN model after data correction. According to (1) and (2), the fault probabilities will be obtained. The threshold value will self-adjust according to the data load factor in the control area. The fault lines will be screened by the threshold value. Finally, call the electric quantities data and calculate the fault classification criterion based on symmetrical components of reactive power. After that, input the diagnostic result to the behavioral analysis model and screen out the abnormal protections and breakers.

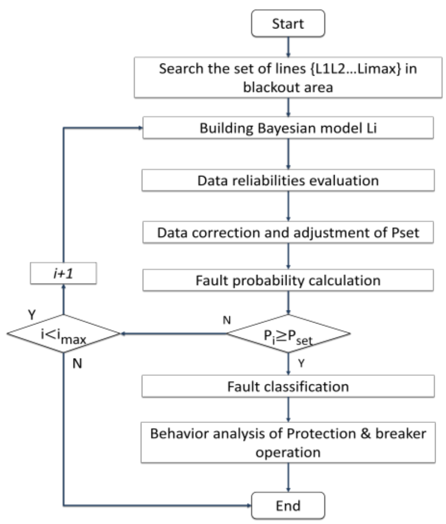


Fig.6 The framework of combined data analytics

V. CASE STUDY

A. Data reliability analysis

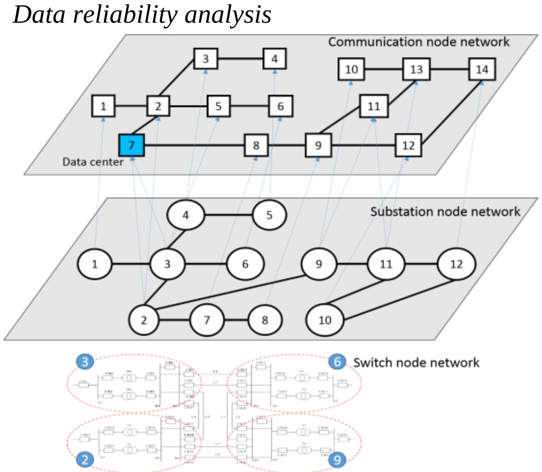


Fig.7 Three-layer network of a typical power system

As shown in Fig.7, the power grid is simulated as a cyber-physical system [13] which is comprised of switch nodes network, substation nodes network and the communication nodes network formed by the data center and information nodes. Every substation node sends status messages and receives control commands through at least one information node, while every information node gets electrical power from at least one substation node. In order to improve the robustness of fault diagnosis for missing information, this paper uses the support vector machine [14] to predict the data reliability. When the data reliability L_{ZONE} is lower than 0.9, the fault information will be revised.

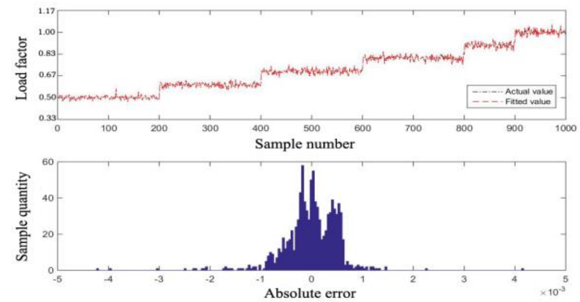


Fig.8 Comparison chart of fitting curve and error display

By the training of 1000 groups of sample data, the load factor of communication subsystem (L_{zone}) were fitted by support vector machine. Fig.9 is the comparison chart of fitting curve and error display, as seen from the chart, the average fitting precision of multi group experiments reaches to 99.98%.

B. Case simulation of typical power system

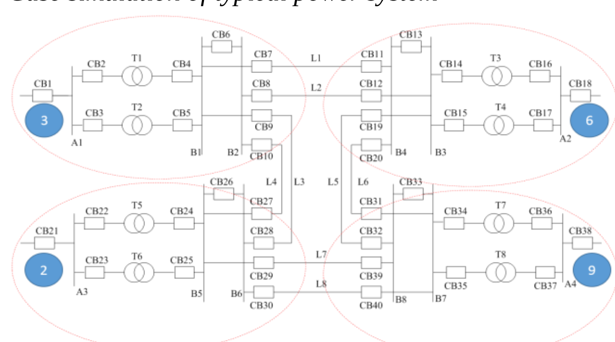


Fig. 9 Switch node network of simulation

Eight lines of the above system (Fig.9) have been simulated with 64 fault scenarios, the results of each test are derived from the average of 5000 random tests. The fault scenarios contain protection/breaker refused operation and mal-operation. Furthermore, 30 group simulations under different data loss rate are added. Fig.10 shows the misdiagnosis rate under different data loss rates. As diagnosis system miss more data, the diagnosis performance will drop quickly. When the data loss rate is 0.2, misdiagnosis rate is 0.55. If 30% of data is missing, the misdiagnosis rate will reach or more than 0.7.

Table I Threshold value under different situations

L_{ZONE}	0.33~0.83	0.83~1.00	1.00~1.17
Data loss rate	0~0.05	0.05~0.15	0.15~0.3
P_{set}	0.75	0.55	0.3

In order to ensure the validity of the fault screening, the threshold value P_{set} should be self-adjusted according to the reliability of the data. Table I shows the selection of threshold value under different load factor or data loss rate.

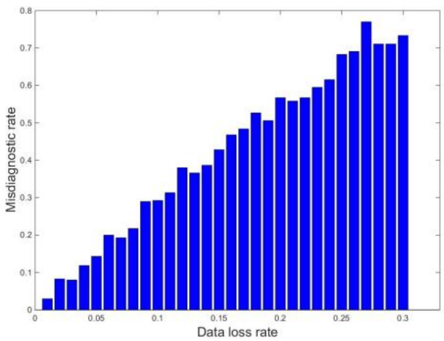
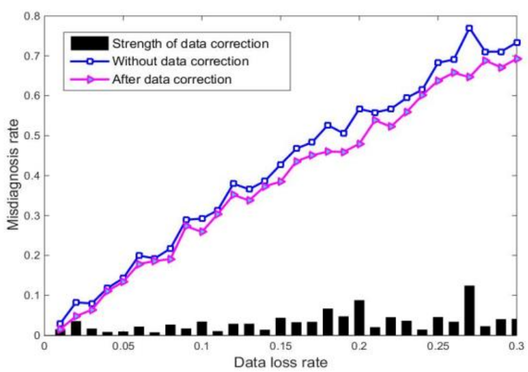
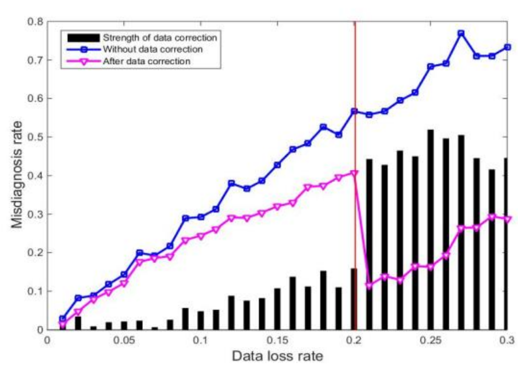


Fig.10 Misdiagnosis rate under different data loss rates

Although the consequences of data loss are difficult to recover, the assessing of data reliability could offer a better guide for related operator to choose a reasonable failure recovery strategy. In Fig.11 (a), the blue line is the misdiagnosis rate without data correction and self-adjustment of P_{set}. After data correction, the performance of diagnosis has already been improved. With the self-adjustment of P_{set}, the misdiagnosis rate drops obviously as shown in Fig.11 (b).



(a) Without adjustment of threshold value



(b) With the self-adjustment of threshold value

Fig.11 Comparison charts of diagnostic performance

VI. CONCLUSIONS AND FUTURE RESEARCH

This paper proposes a combined data analytics technique, which Bayesian network to construct diagnostic model through the switching value information. By use of support vector machine to predict the reliability of communication data and with the data correction strategy and self-adjustment of threshold value, the accuracy of fault diagnosis is improved.

The data transmission is very complex, especially in the large-scale information network, the observability is poor. It is a feasible way to evaluate the data reliability by the method of data mining. In the analysis of data reliability, this paper only considers the influence of data loss. In the future work, this paper will further extend the fault scenarios, such as the linkage effects of the information node breakdown caused by physical system failure. Through the simulation of a variety of scenarios and the means of big data analysis to further enhance the performance of reliability assessment of information system, and build a fault diagnosis model with greater robustness.

REFERENCES

[1] Qu C,Chen W, et al. "Distributed data traffic scheduling with awareness of dynamics state in cyber physical systems with application in smart grid". *IEEE Transactions on Smart Grid*, vol. 6, no.6, pp.2895-2905, 2015.

[2] G. Li, H. Wu, et al. "Bayesian network approach based on fault segregation for power system fault diagnosis", *2014 International Conference on Power System Technology*, 2014.

[3] North American Electric Reliability Councils Disturbance Analysis Working Group(DAWG) Database[DB/OL].2001.

[4] Q. LI, J. Xu, "Power System Fault Diagnosis Based on Subjective Bayesian Approach," *Automation of Electric Power Systems*, vol. 31, no. 15, pp. 46-50, 2007.

[5] R. Wang, X. Wang, "Wide Area Backup Protection Algorithm for Power Grid Based on Correlation Matrix", *Automation of Electric Power Systems*, vol. 37, no. 4, pp. 69-74, 2013.

[6] Y. Zhu, L. Huo, J. Lu, "Bayesian networks-based approach for power systems fault Diagnosis," *IEEE Transactions on Power Delivery*, vol. 21,no. 2, pp. 634-639, 2006.

[7] J. Pearl, "Propagation, and structuring in belief networks". *Artificial Intelligence*, vol. 29, no. 3, pp. 241-288, 1986.

[8] B. Cai, H. Liu, M. Xie. "A real-time fault diagnosis methodology of complex systems using object-oriented Bayesian networks", *Mechanical Systems and Signal Processing*, vol. 80, pp. 31-44, 2016.

[9] B. Cai, H. Liu. Q. Fan, et al. "Multi-source information fusion based fault diagnosis of ground-source heat pump using Bayesian network", *Applied Energy*, vol. 114, pp. 1–9, 2014.

[10] J. Xiong, Q. Zhang, et al "An Information Fusion Fault Diagnosis Method Based on Dimensionless Indicators With Static Discounting Factor and KNN", *IEEE Sensors Journal*, vol. 16, pp. 2060-2069, 2016

[11] Cieslak J, Henry D, Zolghadri A. "Fault tolerant flight control from theory to piloted flight simulator experiments", *IET Control Theory Applications*, vol. 4, no. 8, pp. 1451-1464, 2010.

[12] B. Mahamedi and J. Zhu, "Fault Classification and Faulted Phase Selection Based on the Symmetrical Components of Reactive Power for Single-Circuit Transmission Lines", *IEEE Trans. Power Del.*, vol. 28, no. 4, pp. 2326-2332, OCT. 2013.

[13] M. Parandehgheibi, E. Modiano. "Robustness of interdependent networks: The case of communication networks and the power grid", *IEEE Conference on Global Communications*, Atlanta, USA, pp. 2164-2169, 2013.

[14] Leonidas C. Resende; Luiz A. F. Manso, et al " Support Vector Machine application in composite reliability assessment", *2015 18th International Conference on Intelligent System Application to Power Systems (ISAP)*, pp.1-6, 2015.