A METHOD TO LOCATE BAD DATA IN LARGE POWER SYSTEM USING DISTRIBUTED APPROACH

SREELAL N[1], Dr. S.P KARTHIKEYAN[2], POLLY THOMAS[3], EMIL NINAN SKARIAH[4]
[1,3,4]Mahatma Gandhi University, India
[2]VIT University, India
[1]sreelal021@gmail.com

Abstract**:** *It is essential to determine the presence of bad data in power system networks and to identify the reasonable meters which cause bad data error. In this paper, a distributed approach is adopted which enables distribution of complex system network into different area and apply meter identification procedures separately in each area. The reasonable meters identified using individual error determination technique after grouping into distributed areas. It improves sensitivity as well as reliable detection of bad data along with locating the affected meters. The system works with MATLAB platform to verify meter identification.*

Key words: *State estimation, Distributed detection, Meter Identification*

Nomenclature

H	Function Matrix
R	Co-variance Matrix
Zm	Measurement Matrix
θ	Voltage angle
J(x)	Residue
Z	Power flow Meter
Nm	Number of Measurements
Ns	Number of State variables
E	Total error
e	Individual Meter error
t	Threshold value

1. Introduction

The presence of bad data in metering devices can arise severe power system security issues like blackouts. While considering the role of metering infrastructures involving measuring devices and sensors, the indication of system to fall into blackout may or may not be done by the meters. In such cases the presence of bad data in measurement plays an important role to make the system into blackout condition. It is considered that the bad data in meters may arise either due to the internal meter error or due to the wrong connection of meters. The internal meter error introduces deviation in actual meter readings exceeding the given tolerance level whereas the reverse/wrong connection of meters can produce negative readings. In modern power system, the entire load distribution is done by means of Demand Side Management (DSM). The part of DSM which performs the above task is DSM controller which has its main input parameters as metered data from the measuring devices. Based on the metered data, the DSM controller will optimize the load distribution as well as line flows. Hence the concern is that, the controller will operate such that it keeps the line flow below line limits if the line tends to exceed the limit. Also the controller may stay inactive even under load change when the line is still within line flow limits. But the occurrence of errors or bad data in metered readings cannot be avoided. As a consequence, the DSM controller will perform improper operation along with ineffective distribution of line flow.

There are several methods adopted to detect the presence of data error on the basis of data security. In [1], a lightweight message authentication scheme introduced so that malicious users were prevented to influence the secrecy or privacy of the information exchanged. A detector based on GLRT [2] investigates error data in the view of an adversary. But in nonlinear models [4], it is difficult to form data error since it requires online data to make attacks in the data system. In [8],a machine learning technique along with signal processing is introduced to enhance the data security. Another scheme involves identification of bad data [11] using state estimation with voltage magnitude and angle. Most methods adopted to detect the bad data error are limited to determine its presence in the system and not to locate the exact origin of the error. It is important that the measured data provided to DSM controller should be error corrected and verified. So such data errors or bad data is to be detected if it exists around the meters. Hence state estimation should perform to determine the estimated measurements and allowed to compare with measured reading. In this paper, a bad data detection scheme which is capable of determining whether there exists or not, the presence of bad data in metering devices. Also it checks for individual error detection and identifies the meter or combination of meters undergoes bad data around them. Here uses, least square estimation for determining the estimated values and chi-square detector for detection of bad data.

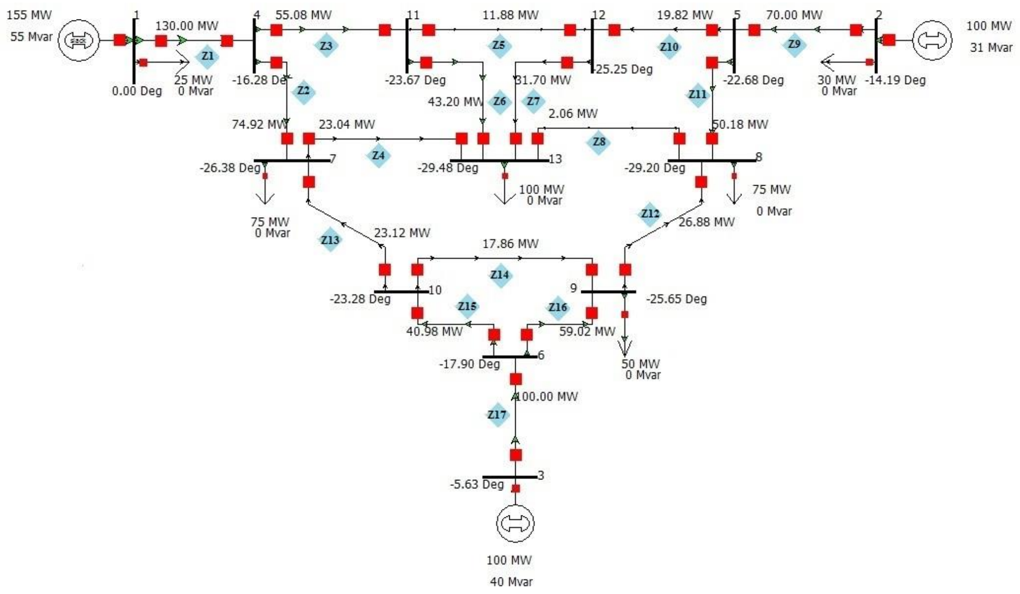


Fig.1 Base case system with normal(all line flows are within limits) line flow condition

2. System Description

Consider a 13-bus system modeled as shown in Fig. 1 which consists of three generation units and six loads such that, the system operates under normal condition. There are 17 transmission lines which were provided with certain line flow limits (in MW).Seventeen meters Z_1, Z_2, Z_3… Z_{17} indicates the instantaneous line flow and sends the measured reading to DSM controller. Based on the line flow data input from meters, the DSM controller will execute an effective load distribution in the system, without making line out flow. The system design case parameters are shown in Table 1 and 2. Consider that all the meters in Fig. 1 indicates accurate measurement, hence the case is taken as design or base case. As a test case, all the lines are taken into consideration.

Table 1
System Base Case Parameters

Parameter	Value
Generator 1 (bus 1)	155 MW, 55 Mvar
Generator 2 (bus 2)	100 MW, 31 Mvar
Generator 3 (bus 3)	100 MW, 40 Mvar
Load 1 (bus 1)	25 MW, 0 Mvar
Load 2 (bus 2)	30 MW, 0 Mvar
Load 3 (bus 7)	75 MW, 0 Mvar
Load 4 (bus 8)	75 MW, 0 Mvar
Load 5 (bus 9)	50 MW, 0 Mvar
Load 6 (bus 13)	100 MW, 0 Mvar
Line reactance	0.2 p.u (All lines)

Table 2
Line Flow Values

Line	Base case flow (MW)	Line flow limit (MW)
1-4	130.00	150.00
2-5	70.00	80.00
3-6	100.00	120.00
4-7	74.92	80.00
4-11	55.08	60.00
5-8	50.18	60.00
5-12	19.82	25.00
6-9	59.02	60.00
6-10	40.98	50.00
7-10	23.12	25.00
7-13	23.04	25.00
8-9	26.88	30.00
8-13	2.06	5.00
9-10	17.86	25.00
11-12	11.88	20.00
11-13	43.20	50.00
12-13	31.70	40.00

3. State Estimation

The state estimation form estimates which are required to obtain the respective quantity to be estimated. For complex systems with Renewable energy resources, optimization methods[12] are available to estimate the state variables. In the above test case, meter readings are the quantity to be estimated using state estimation. There are some parameters should be provided while perform state estimation, viz. the Function Matrix, the Co-variance Matrix and the Measurement Matrix.

The Function Matrix, H remains same for a given system network. It is formed form a set of expressions that obtains the quantity to be estimated. If there are 12 state variables and 17 measurements, then H is 12x17 Matrix. In case of DC power flow method, the elements in H matrix are the inverse of line reactance. The Co-variance Matrix, R represents the probability of variation of measured data from its actual value. The elements of R Matrix are obtained from the manufacturer itself but, when consider for a test case we can assume values as 1% or 2%. The Measurement Matrix, Zm is directly obtained from the measuring device which may be accurate or inaccurate. The measured data is the prime element for evaluating the nature of estimate, whether it is good or bad estimate.

Fig. 1 shows a basic 13-bus system with measurements of all the transmission line flows were taken into consideration. In this case, the state estimation is done with voltage angles are taken as estimation variables (state variables). The bus at which generator 1 connected is taken as the slack bus. i.e. $\theta_1 = 0.00^0$ and the voltage angles in other buses viz. θ_2 , θ_3 , θ_4 ,… θ_{13} are the estimation variables. Now, the degree of freedom, m-n = 17-12 = 5. The level of significance of each of the measurement is considered to be 1% (i.e. 0.01). The general expression for line flow can be written as[13],

$$f(i,j) = \frac{1}{X(i,j)}(\theta i - \theta j) \tag{1}$$

Since the meters represents line flow, then the expression for measurements Z_1 to Z_{17} can be determined using the above general formula. From the set of equations, the function matrix H can be formed by substituting the coefficients of estimates in expression for measurement. Since each expression involve function of any two estimates, the elements of each row have at most two non-zero elements. The reactance of each transmission line is considered to be 0.2pu. Hence the non-zero elements in H will be always 5 or -5. The Co-variance Matrix, R represents the probability of variation of measured data from its actual value. The R matrix is a diagonal matrix formed by variance of each meters in the system. The variance is represented as 'σ_i' and the co-variance matrix R becomes,

$$R = \begin{matrix} \sigma_1^2 & 0 & 0 & ... & 0 \\ 0 & \sigma_2^2 & 0 & ... & 0 \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ 0 & 0 & 0 & ... & \sigma_{Nm}^2 \end{matrix} \tag{2}$$

The Zm Matrix is directly obtained from the meter itself. It is column matrix of all the meter readings in a specific order. The readings may accurate or inaccurate, despite that they are provided to the state estimator. The detection process will start only if the estimates from the estimator should be accurate. The measurement matrix can be given as,

$$Zm = \begin{matrix} Zm(1) \\ Zm(2) \\ \vdots \\ Zm(17) \end{matrix} \tag{3}$$

The general expression for residue,

$$J(x) = \sum_{i=1}^{Nm} \frac{(Zmi - fi(x))^2}{\sigma i^2} \tag{4}$$

Where, $Zm = \begin{matrix} Zm(1) \\ Zm(2) \\ \vdots \\ Zm(17) \end{matrix}$, $fi(x) = \begin{matrix} f1(1) \\ f2(2) \\ \vdots \\ fNm(17) \end{matrix} = Hx$

In matrix form, the expression for residue can be given as,

$$J(x) = [Zm - f(x)]^T [R]^{-1} [Zm - f(x)] \tag{5}$$

$$= Zm^T [R]^{-1} Zm \text{ - } f(x)^T [R]^{-1} Zm \text{ - } Zm^T [R]^{-1} f(x) \text{ - } f(x)^T [R]^{-1} f(x) \tag{6}$$

Substituting the value $f(x) = Hx$ in (6)

$$J(x) = Zm^T [R]^{-1} Zm - x^T H^T [R]^{-1} Zm \text{ - } Zmeas^T Hx \text{ - } x^T H^T [R]^{-1} Hx \tag{7}$$

Best estimate is obtained by minimizing the residue J(x). Hence partial derivative is applied and equate to zero.

$$\nabla J(x) = 0 - H^T [R]^{-1} Zm \text{-} H[R]^{-1^T} Zm^T \text{-} 2H^T [R]^{-1} Hx = 0$$

$$= -2H^T [R]^{-1} Zm + 2H^T R^{-1} Hx = 0$$

Hence the value of estimate can be defined as,

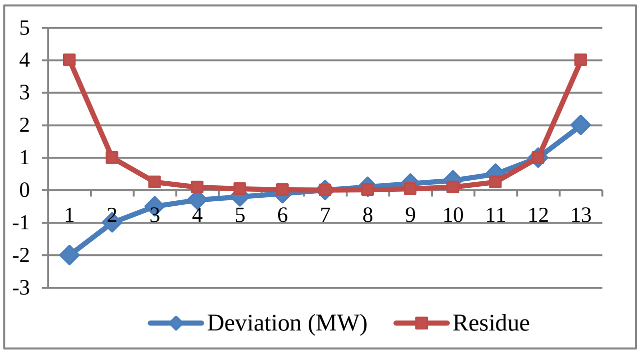
$$x^{est} = [H^T [R]^{-1} H]^{-1} [H^T [R]^{-1} Zm] \tag{8}$$

Now the estimates are determined and hence the estimated measurement vector [Z] is calculated. Note that [Zm] is the measured value which directly taken from the measuring devices and the estimates formed are based on Zm. Thus comparison of estimated value with design parameters results the total error due to inaccuracy in estimator input. This value ultimately indicates the nature of estimate. It is also possible to identify the reason for bad estimate if any occurred in estimation.

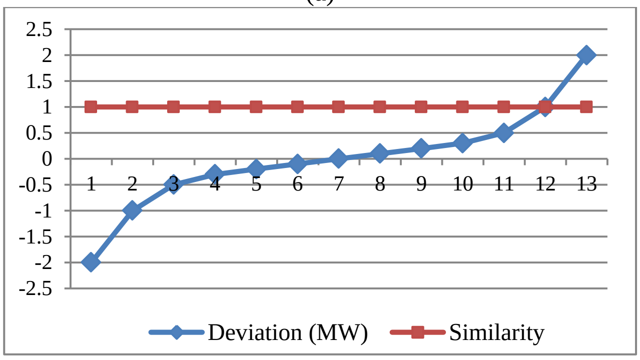
4. Bad Data Detection

The secure operation of power system networks enables error free data communication. Thus the data provided to the controllers should be accurate and error free with minimum possible tolerance of deviation. Hence it is required to detect the bad data errors in meters as well as specify the meters(s) involving data errors associated with them. Such a way, it is possible to provide error free measurements to DSM controller. The detection of bad data can be done with help of main two parameters which involves the estimated value and measured value and their difference gives the estimated error. By analyzing the estimated error with the help of predefined threshold value (from standard chi-square table), presence of bad data can be effectively detected. Now the detection requires state estimation using Nm number of measurements and Ns number of state variables.

Standard chi-square technique is used to detect the presence of bad data. It is also possible to perform detection using another method called cosine similarity matching. The following graphs indicate the comparison between chi-square and cosine similarity methods.



(a)



(b)

Fig.2 Comparison between (a) chi-square and (b) cosine similarity methods

From Fig.2, it is clear that the variation of residue against deviation for same system network is higher for chi-square method than cosine similarity matching approach. Hence it is easier to provide threshold values to filter the error value, thereby detection can be done more accurately. In case of CSM, the deviation of similarity value varies slightly causes difficulty in setting threshold values. In other words, chi-square obtains more reliable detection than cosine similarity matching technique.

The threshold value is the ultimate limit for detecting the nature of estimate. The threshold value should be selected with standard methods which are commonly acceptable. Here the threshold value is set using standard chi-square table. In the above test case, there are 13 estimates (θ_2, θ_3….. θ_{13}) represented as Ns and 17 no.s of measurement (Z_1, Z_2…., Z_{17}) denoted as Nm. Now the degree of freedom can be defined as,

Degree of freedom, k = Nm-Ns =5

Let the level of significance α be taken as 1% (i.e 0.01). Thus the chi-square value corresponding above condition will be obtained from standard chi-square table as 15.086. Hence the threshold value t = 15.086 is used to compare with total error calculated. Now the method enables detection of presence of bad data in power flow meters. But it does not sufficient to provide any information regarding the location bad data error. Hence individual meter errors are to be determined to specify those reasonable meters. The individual meter errors e_1,e_2…e_{17} are calculated and check individually and followed by combination of meters. Finally, display the status of presence of bad data as well as the meters reasonable for bad data error.

5. Distributed Detection and Meter Identification

The simultaneous detection of bad data error in a complex system network having more number of buses, have some problems in reliable operation. Such detection approaches may cause reduced sensitivity [5], software program become more complex and the program delay increases. If there is any error occurred in the program, will affect the entire system operation. To avoid such difficulties, we can divide the system network into different area/zone and it is possible to apply the detection procedure in each area separately in a distributed manner. The division of above test case can be made into 4 zones as shown in Fig.3. It is clear that, the division of zones done by grouping all the meters into different zones which were treated independently. Note that, state estimation is done once for entire measurements and the detection is performed separately. The distribution of meters which comes under different zone can be tabulated in Table 3. In this approach, the estimator takes the input values as measured data and performs state estimation to determine the estimated measurements. After detecting the presence of bad data in the whole system, the network system is divided into different areas. The region of error is identified first by calculating the total error occurred in the area due to the concerned meters. If the error exceeds threshold value, it is required to identify the reasonable meters. Hence thorough checking is adopted inside each area which undergoes bad data error. The meter identification involves determining individual meter errors in each area separately for the purpose identifying the location of error in the whole system. The flow chart shown in Fig. 5 indicates the proposed identification technique using distributed approach. The error may occur in more than one meter simultaneously in the system. Hence combinational meter identification is required in the case of multiple meter errors.

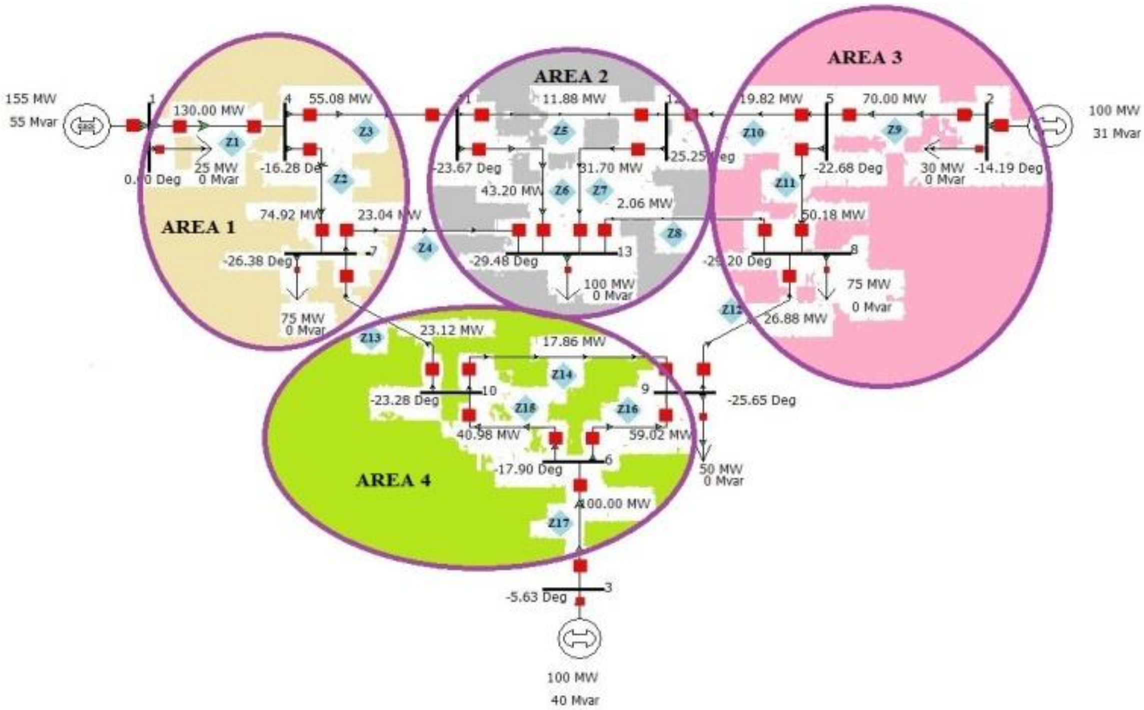


Fig.3 Bus system with distributed detection approach

For example, if bad data error occurred at meter 8 in line 8-13, it will displays that bad data error occurred at Zone 2 and check meter 8. Similarly it is applicable to all the meters in all zones. If there is no error in any zone, it will indicate separately. Thus the region or area associated with reasonable meter can be located easily for further correction proceedings. The average area error is determined inorder to identify the area which possesses dominant error in the system. While meters were identified, they are displayed in the descending order of error deviation. This implies that the firstly indicated meter in a respective area undergoes most deviation error among others.

The Error-deviation graph shown in Fig. 4 shows that when there is deviation in any meter the error become non zero and keep increasing as the deviation in both directions. When the error value meets the threshold line provided, it reaches the maximum allowable limit of deviation and beyond that value the meter will be grouped into error measurement device. Similarly, all the meters inside each area is evaluated to identify those meters undergoes error and the corresponding location of error.

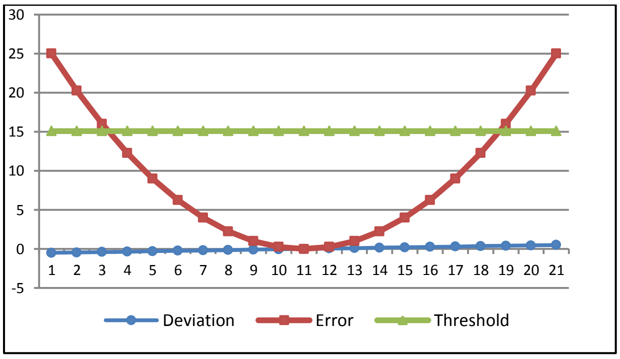


Fig. 4 Error-deviation graph for meter identification

Table 3
Distribution of Meters into Different Zones

AREA 1		AREA 2		AREA 3		AREA 4	
Meters	Line	Meters	Line	Meters	Line	Meters	Line
Z_1	1-4	Z_5	11-12	Z_9	2-5	Z_{13}	7-10
Z_2	4-7	Z_6	11-13	Z_{10}	5-12	Z_{14}	9-10
Z_3	4-11	Z_7	12-13	Z_{11}	5-8	Z_{15}	6-10
Z_4	7-13	Z_8	8-13	Z_{12}	8-9	Z_{16}	6-9
						Z_{17}	3-6

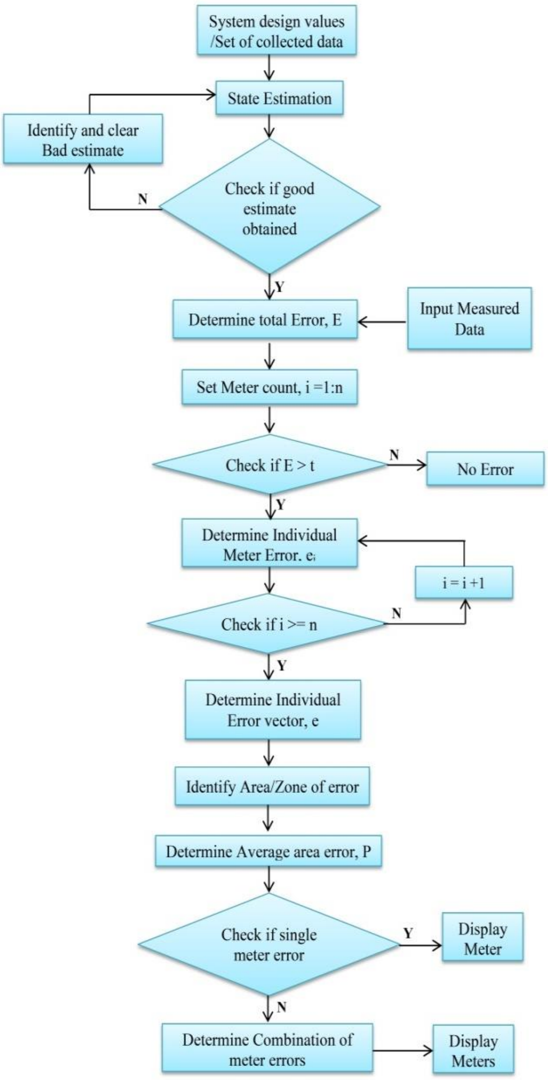


Fig. 5 Flow chart of proposed meter identification scheme

6. Results

A MATLAB program was developed and the base case values are obtained from 13-bus system modeled in POWER WORLD simulator software package. Program verified the method to determine the estimates as well as the meters that causes bad data error if any. The bad data is detected in distributed network system enables identification of location of system from which the error occurs and the concerned meter(s) causes error. Fig. 6 shows the case for "No error" condition, means that the measured data from all power flow meter connected in various locations of the system network gives accurate reading or the readings are within tolerance limit. Fig. 8a shows the graph of deviation in measurements and it is clear that all the measured reading overlaps the estimated value which indicates the accurate measurement. Here it is not required to identify the location or meter(s) since there is no error happened in the system.

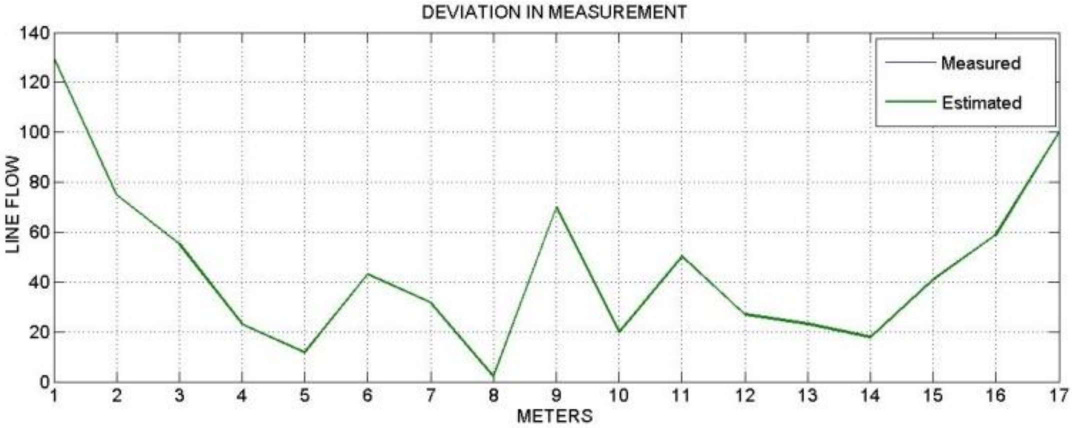
```
        BAD DATA DETECTION IN METERS
       ******************************
"Good estimate obtained"
ENTER THE METER READINGS
 [130;74.92;55.08;23.04;11.88;43.20;31.70;2.06;70;
          19.82;50.18;26.88;23.12;17.86;40.98;59.02;100]
The Meter readings are..
    Meter 1 = 130.000000 MW      Meter 2 = 74.920000 MW
    Meter 3 = 55.080000 MW       Meter 4 = 23.040000 MW
    Meter 5 = 11.880000 MW       Meter 6 = 43.200000 MW
    Meter 7 = 31.700000 MW       Meter 8 = 2.060000 MW
    Meter 9 = 70.000000 MW       Meter 10 = 19.820000 MW
    Meter 11 = 50.180000 MW      Meter 12 = 26.880000 MW
    Meter 13 = 23.120000 MW      Meter 14 = 17.860000 MW
    Meter 15 = 40.980000 MW      Meter 16 = 59.020000 MW
    Meter 17 = 100.000000 MW

    "No Error Found in ZONE 1"

    "No Error Found in ZONE 2"

    "No Error Found in ZONE 3"

    "No Error Found in ZONE 4"
```

Fig.6 Case for No error condition (Command Window)

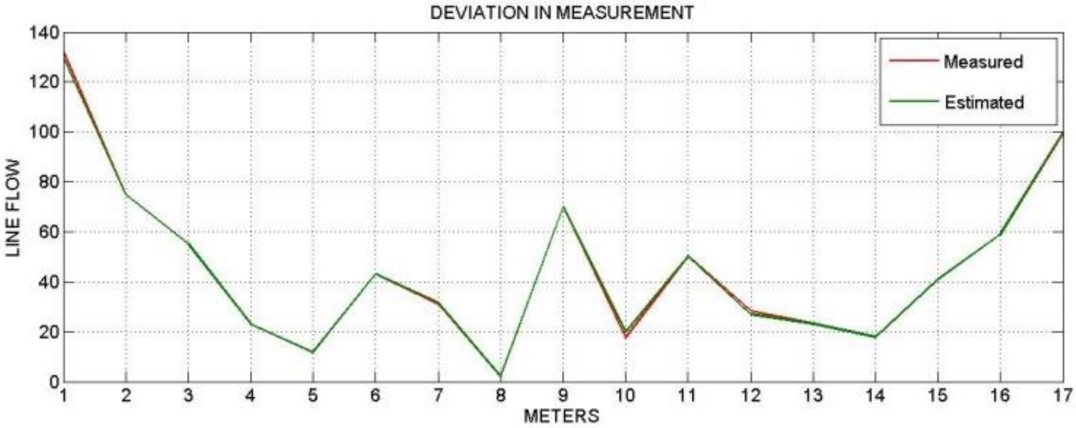
Now it is considered that bad data error occurred in some power flow meters in the system network as shown in Fig.7. The meters 1, 7,10,12,15 and 17are provided with sufficient deviation in readings and given during the run time as measured data.

```
        BAD DATA DETECTION IN METERS
       ******************************
"Good estimate obtained"
ENTER THE METER READINGS
 [132.5;74.92;55.08;23.04;11.88;43.20;31.10;2.06;70;
        17.42;50.18;28.32;23.12;17.86;41.28;59.02;99.25]
The Meter readings are..
    Meter 1 = 132.500000 MW      Meter 2 = 74.920000 MW
    Meter 3 = 55.080000 MW       Meter 4 = 23.040000 MW
    Meter 5 = 11.880000 MW       Meter 6 = 43.200000 MW
    Meter 7 = 31.100000 MW       Meter 8 = 2.060000 MW
    Meter 9 = 70.000000 MW       Meter 10 = 17.420000 MW
    Meter 11 = 50.180000 MW      Meter 12 = 28.320000 MW
    Meter 13 = 23.120000 MW      Meter 14 = 17.860000 MW
    Meter 15 = 41.280000 MW      Meter 16 = 59.020000 MW
    Meter 17 = 99.250000 MW

    "Bad data detected in ZONE 1 (0.5 percent)"

        Check Meter 1

    "Bad data detected in ZONE 2 (0.5 percent)"

        Check Meter 7

    "Bad data detected in ZONE 3 (4.4 percent)"

        Check Meters 10 and 12

    "Bad data detected in ZONE 4 (0.3 percent)"

        Check Meters 17 and 15
```

Fig.7 Case for error in meters 1, 7,10,12,15 and 17 (Command Window)



(a)



(b)

Fig,8 Deviation graph for (a) No error condition (b) Error in meters 1, 7,10,12,15 and 17

Here the system determines the presence of bad data error and identifies the meters which causes the error. Since the meter 1 involves Area 1, meter 7 involves Area 2, meters 10 and 12 involves Area 3 and meters 15 and 17 involves Area 4 respectively, the system identifies the areas which undergo bad data error and the respective metes.

From the output window shown in Fig. 7, the location of meters which causes error was identified. While calculating the total error of each area, the average area error is determined and is specified along with it. Here it is found that, Area 3 possesses highest error occurring region among others since the average area error is highest for Area 3. This indicates that the meters which come under Area 3 are the cause for producing most error in the network. Also in area 3, Meter 10 introduces more deviating error than Meter 12 so that Meter 10 is indicated first and followed by Meter 12. Similarly in Area 4, Meter 17 causes more error than Meter 15, thus Meter 17 is indicated first. The deviation in measurement can be represented with variation in measured data from estimated data of reasonable meters as shown in Fig. 8b.

The approach can be used to analyze the rate of error occurrence of different areas in the system. Also, those regions with frequent error occurrence can be identified and further proceedings such as improving the sensitivity of meter identification as well as error correction can be done as extension.

7. Conclusion

This paper proposes a method to detect the presence of bad data and to locate the meter(s) which causes error in power flow meters connected in a 13-bus system network. A proposed distributed detection approach is used along with chi-square method for setting standard values of threshold. The method enables distribution of complex system network into different area and applies detection procedures separately instead of performing once in the entire system and identifies the location of error in the system in terms of region and concerned meter(s). It gives an idea about which region possess dominant error in the system and the meter in which most error occurs. The result shows that the system works with efficient detection of bad data error and identifies the exact location of reasonable meters.

References

1. M. M. Fouda, Z. M. Fadlullah, N. Kato, R. Lu, and X. Shen, "*A lightweight message authentication scheme*

for smart grid communications," IEEE Transactions on Smart Grid, vol. 2, no. 4, pp. 675–685, 2011.

2. O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "*Malicious data attacks on the smart grid*," IEEE Transactions on Smart Grid, vol. 2, no. 4, pp. 645–658, 2011.
3. M. Esmalifalak, G. Shi, Z. Han, and L. Song, "*Bad data injection attack and defense in electricity market using game theory study*," IEEE Transactions on Smart Grid, vol. 4, no. 1, pp. 160–169, 2012.
4. M. A. Rahman,H Mohsenian-Rad,"*False Data Injection Attacks Against Nonlinear State Estimation in Smart Power Grids*" IEEE Transactions on Smart Grid,2013
5. Y Gu,T Liu,D Wang,X Guan,Z Xu," *Bad data detection method for smart grids based on distributed state estimation*"IEEE International Conference on Communications 2013
6. J.Salmeron, K. Wood, and R. Baldick, "*Analysis of electric grid security under terrorist threat*," IEEE Transactions on Power Systems, vol. 19, no. 2, pp. 905–912, 2004.
7. D.B.Rawat and C.Bajracharya, "*Cyber security for smart grid systems: Status, challenges and perspectives*," in Proceedings of IEEE South- EastCon 2015, Fort Lauderdale, FL, USA, Apr. 9–12, 2015 , "," , vol. , pp. –, .
8. M Esmalifalak, L Liu, N Nguyen, R Zheng,Z Han, "*Detecting Stealthy False Data Injection Using Machine Learning in Smart Grid* ",IEEE systems Journal 2014
9. M.A.Rahman,H Mohsenian-Rad, "*False Data Injection Attacks with Incomplete Information Against Smart Power Grids*" IEEE 2012
10. S Li, Y Yılmaz, X Wang," *Quickest Detection of False Data Injection Attack in Wide-Area Smart Grids*" IEEE Transactions on Smart Grid, vol. 6, no. 6, pp. 2725–2735, 2015.
11. A. Sahu." *Bad Data Detection and Identification in Power Systems*",Retrieved (2016) from http://www. researchgate.net/publication/29028438.
12. L. Kumar, M. Prasanna,T. Ananthapadmanabha, "*A novel approach for distribution system state estimation with renewable energy sources based distributed generations*", Journal of Electrical Engineering, vol. 14 (3rd ed.), pp. 295-301, 2014.
13. Allen J wood and Bruce F Wollenberg. (1996, February6*). Power Generation, Operation, and Control* (2nd edition) [Online]. Available:http://www books. google.com
14. Saadat Hadi. (1999*). Power System Analysis* [Online].Available:http://www books. google.com
15. Power World Corporation (2015), *Power World Simulator online Help.* Available:http://www.power world.com/WebHelp.
16. Knight Andrew (1961), *Basics of MATLAB and beyond.* Available:http://www.crcpress.com