A bad data detection approach to EPS state estimation based on fuzzy sets and wavelet analysis

Irina Kolosok[1], *Liudmila* Gurina[1,*]

[1] Melentiev Energy Systems Institute of SB RAS, Irkutsk, Russia

Abstract. The paper offers an algorithm for detection of erroneous measurements (bad data) that occur at cyberattacks against systems for data acquisition, processing and transfer and cannot be detected by conventional methods of measurement validation at EPS state estimation. Combined application of wavelet analysis and theory of fuzzy sets when processing the SCADA and WAMS measurements produces higher accuracy of the estimates obtained under incomplete and uncertain data and demonstrates the efficiency of proposed approach for practical computations in simulated cyberattacks.

Keywords. WAMS, SCADA; measurements; data quality; certainty; cyberattacks; fuzzy sets; wavelet analysis.

1 Introduction

Enhancement of information and communication infrastructure during EPS digitalization is ensured by development of sensor and network technologies that are based on introduction of digital equipment, application of intelligent technologies in the systems for data measurement, interpretation and transfer that are needed for EPS operation control. They raise the efficiency and flexibility of EPS control and monitoring [1]. At the same time the problems of data quality occur during combined application of SCADA and WAMS measurements in the conditions of growing number of cyberattacks against cyber-physical EPS. Noted is the negative impact of the above problems on the accuracy of solving the EPS state estimation problems due to erroneous measurements that are not detected by conventional methods [2,3], and due to lack of sufficiently scope of measurements [4].

Therefore, development of algorithms for data processing and interpretation at low quality of SCADA and WAMS measurements as a preliminary step of EPS state estimation is of practical importance.

EPS state estimation includes such functions as analysis of EPS observability, analysis of the network configuration (topological network analysis), identification and filtration of 'bad data', additional computation of non-measured parameters [5]. Availability of excessive measurements and the number of available measurements play an important role in obtaining all the estimates of conditions variables.

For the purpose of bad data identification, including those in the algorithms for bad data detection on the base of test equations [5], all the measurements are divided into the following groups:
- valid measurements;
- erroneous measurements whose values can be replaced by computed ones;
- doubtful measurements, i.e., measurements included into critical groups that may contain bad data, but their values cannot be computed based on valid data, thus increasing the dispersion;
- unchecked or critical measurements; they are measurements that were not included into test equations and errors in them cannot be detected [6].

The use of PMU measurements along with SCADA measurements improved circumstances related to "bad data" and to validation of measurements [7]. Nevertheless, [8] demonstrates vulnerability of EPS state estimation towards unidentifiable cyberattacks.

The paper analyzes the quality of SCADA and WAMS measurements at cyberattacks against information-communication EPS infrastructure. An algorithm for identification of erroneous measurements under data uncertainty using wavelet analysis and fuzzy sets is proposed. Implementation of the algorithm is demonstrated on the simulated cyberattacks.

2 Quality of SCADA and WAMS measurements

Introduction of new information and communication technologies during EPS conversion into a Smart Grid would, along with SCADA measurements, provide control systems with high-frequency PMU measurements.

Interrelations between an information-communication subsystem and a physical subsystem of a cyber-physical EPS are in practice constrained by data quality and security. Reliable EPS operation can be compromised by incomplete and unreliable SCADA and WAMS measurements.

* Corresponding author: gurina@isem.irk.ru

Under data quality we mean the degree of their completeness and reliability [9].

During EPS digitalization it is important to take into account the problem of data quality for EMS applications used in EPS control as risks of external and internal disturbances for cyber security and due to peculiarities of the existing systems for measured data acquisition, transfer and processing are growing; the risks are originated by the following devices:
- SCADA RTU;
- WAMS PMU;
- RTU and PMU [10].

PMU measurement technologies are currently applied in EPS, which allows timely control of the system state. But for a number of reasons, including economic ones, replacement of all RTU by PMU is not currently feasible. Therefore, EPS state estimation is made either using the data of SCADA measurements, or mixed measurements of SCADA and PMU. EPS state estimation using mixed measurements of SCADA and PMU gives rise to certain difficulties. Partially the states can be measured directly using PMU, the remaining states shall be estimated using RTU, which, in turn, requires the development and modification of conventional EPS state estimation methods whose algorithms are based on the integration of SCADA and PMU measurements [4].

Successful attacks against SCADA and WAMS also have an impact on the quality of measurements and occurrence of mistakes in measurements, data loss and loss of synchronization being their consequences. Ref. [9] shows dependence of the cyber security properties loss on the data quality.

For identification of consequences of successful cyberattacks the focus is made on accuracy, adequacy, timeliness, synchronization, consistency and sequence of measurements as they characterize their quality.

Reliability requires accuracy and synchronization of measurements in time within permissible mistakes without violating the sequence of data occurrence. Accuracy of estimation requires account of such a factor as consistency of measurements. Completeness is characterized by availability of data and requires that data of measurements be without losses and were timely, i.e., delivered within permissible time limits.

Consequences of successful cyberattacks for factors characterizing the data completeness and reliability are analyzed in Ref. [9] based on the algorithm developed by the authors for data quality evaluation at EPS state estimation.

Most frequent cyberattacks against cyber-physical electric power systems whose consequences are misleading for EPS state estimation are False Data Injection (FDI) and Denial of Service (DoS) attacks [11]. FDI attacks are targeted at changing the measurement data and can bypass the routes for identification of bad data in EMS. Successful DoS-attacks may cause considerable loss of measurements thus making the system unobservable, and application of conventional EPS state estimation methods becomes impossible.

For facilitating the solution of the EPS state estimation problems in the conditions of cyberattacks that deteriorate the data quality, for identification of erroneous measurements the data should be processed as a preliminary stage of EPS state estimation on the base of wavelet analysis and fuzzy sets.

3 An algorithm for identification of erroneous measurements

This algorithm should be developed for assessing the measurement accuracy required for validating the reliability of data used for EPS state estimation.

FDI attacks against random processes of changes in the conditions parameters are more difficult to detect than FDI attacks against static models as attacks can be mixed both with errors of the measurement route and with noises of communication channels. The model of measurements in this case can be described as

$$\bar{y}(t) = y(t) + \xi_y(t) + a(t), \quad (1)$$

where $y(t)$ is a flow of true values of measured parameters; $\xi_y(t)$ is a vector of measurement noise with normal distribution $\xi_y(t) \to (0, \sigma_y^2)$ with zero mathematical expectation and dispersion σ_y^2 characterizing the accuracy of measurements; $a(t)$ is a cyberattack [12].

Attacks $a(t)$ can be launched by injection of false data into the flow of measurements and/or noise.

The proposed Bad Data Identification (BDId) algorithm includes two stages:
1. Wavelet analysis of information flows on the base of the validation scheme proposed in [12];
2. Identification of erroneous measurements at the i-th time moment based on a fuzzy system of the logical conclusion.

The advantage of wavelet conversion of measurement flows is reduction of the impact of cyberattacks on the data reliability by noise filtration and elimination (smoothing) of errors in measurements.

Furthermore, use of wavelet analysis enhances the accuracy of measurement flow characteristics that are required for developing the Fuzzy Inference System (FIS) at the second stage of erroneous measurements identification.

For developing the FIS, the following characteristics of measurement flows shall be determined:
- mathematical expectation m_y;
- standard deviation σ_y;
- minimum $\min_y$ and maximum $\max_y$ values.

"Measurement" of the form «$\tilde{y} = y_{tr} + \xi_y$» and "Consistency" characterized by observance of the laws of electric circuits for the considered measurements are assumed to be the input linguistic variables (LV). "Accuracy of measurements" characterizing availability or absence of false data injected by cyberattacks is an output variable. Basic term-manifolds of linguistic variables are defined, and membership functions (MF)

are described (Table 1-3). A FIS to determine the level of measurement accuracy has been developed (Fig. 1).

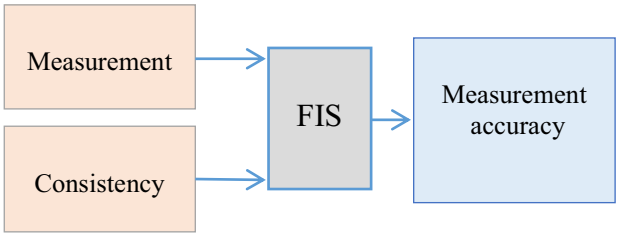

Fig. 1. A FIS to determine the level of measurement accuracy.

Table 1. Basic term-manifold of LV *"Measurement"*.

Term name	MF representation
About $\tilde{y}$	Gaussian MF
Approximately $\tilde{y}$	Gaussian MF
Much higher than $\tilde{y}$	Z-shaped MF
Much lower than $\tilde{y}$	S-shaped MF

Gaussian MF of measurements is determined as

$$\mu_y = \begin{cases} 0, y < -3\sigma_y; \\ e^{-k\delta_y^2}; \\ 0, y > 3\sigma_y; \end{cases} \quad (2)$$

where k characterizes the slope of a membership function, $\delta_y = \frac{(y - m_y)}{2\sigma_y}$ is dispersion of measurement values.

Z-shaped and S-shaped membership functions are used for utmost terms [13].

Table 2. Basic term-manifold of LV *"Consistency"*.

Level	Description
0.75-1	Measurements are consistent
0.25-0.75	Measurements are consistent with a permissible error conditioned by technological peculiarities
0-0.25	Measurements are inconsistent

Table 3. Basic term-manifold of LV *"Measurement Accuracy"*.

Level	Description
0.75-1	High (reliable measurements)
0.25-0.75	Middle (doubtful measurements)
0-0.25	Low (erroneous measurements)

A bad data identification scheme obtained is given in Fig. 2.

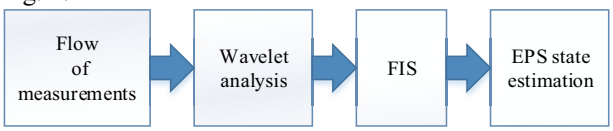

Fig. 2. Identification of erroneous measurements at EPS state estimation.

4 Case study

For validating the efficiency of using the developed BDId algorithm, we analyzed PMU measurements of a real diagram of the electric network (Fig. 3) where PMUs are located. The scope of sampling for every measurement was $n = 30000$ with digitization interval of $\Delta t = 20$ ms. Wavelet analysis has shown that measurements have not gross errors.

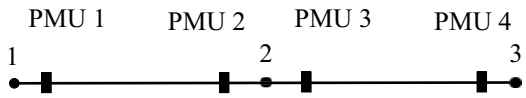

Fig. 3. A section of the electric network.

Figs. 4-7 present the initial graphs of changes in the active power flows P_{2-3} and P_{3-2}, in reactive power flows Q_{2-3} and Q_{3-2} in lines 2-3.

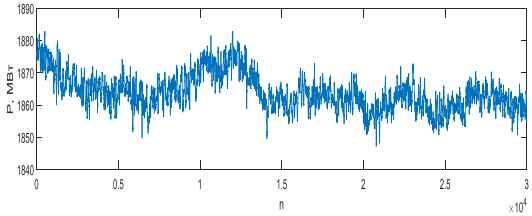

Fig. 4. Changes of the active power flow P_{2-3}.

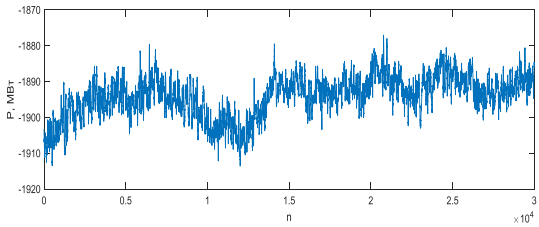

Fig. 5. Changes of the active power flow P_{3-2}.

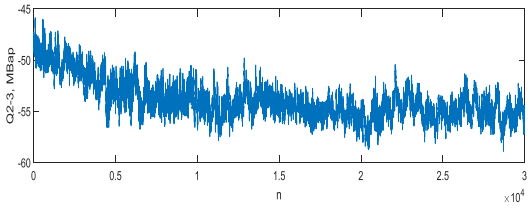

Fig. 6. Changes of the reactive power flow Q_{2-3}.

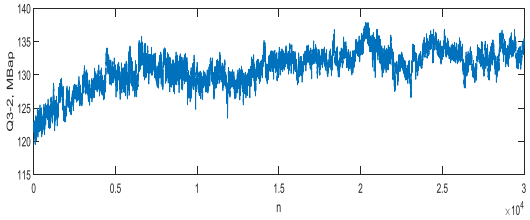

Fig. 7. Changes of the reactive power flow Q_{3-2}.

Characteristics needed for constructing a fuzzy inference system for BDId were computed for these measurement flows (Table 4).

Table 4. Characteristics of processes of active and reactive power change in Line 2-3.

	P_{2-3}	P_{3-2}	Q_{2-3}	Q_{3-2}
m_y	-1864	1894	-53 .94	130 .9
σ_y	5 .635	5 .772	1 .896	2 .826
$\min_y$	-1883	1877	-58 .9	119 .5
$\max_y$	-1847	1914	-45 .86	137 .9

Membership functions in the system of a fuzzy logical conclusion were constructed for linguistic variables "Measurement P_{2-3}", "Measurement P_{3-2}", "Measurement Q_{2-3}", "Measurement Q_{3-2}" using the characteristics obtained.

Implementation of BDId algorithm included calculations during simulation of FDI cyberattacks that cannot be detected by conventional bad data identification methods, i.e., by a method of test equations, when validation is done based on residuals of test equations, and by using a classical state estimation method, when measurements reliability is validated by weighted residuals of estimation [5].

Calculations were made in the simulation experiment that consisted in simulating the random mistakes of measurements in the standard steady-state conditions obtained by calculations using a program for computing the steady-state conditions or state estimation.

Those measurements included false data injection attacks in the form of errors b_{CA}. Model (1) in this case has the form:

$$\bar{y}_{CA1} = y(t) + \xi_y(t) + b_{CA}(t), \quad (3)$$

4.1 Simulating the cyberattacks that are not identified by test equations

For validating the measurements using the test equations method, the test equations are constructed and the following condition is verified:

$$|w_k| \le d_k \quad (4)$$

where w_k is residual of test equation, d_k is some threshold value. If condition (4) holds, then all the measurements in this test equation are assumed to be valid.

Two kinds of gross errors were simulated: -100 MW in the measurement P_{2-3} and +100 MW in the measurement P_{3-2}. Table 5 presents the results of bad data detection and state estimation using the test equation method, identification of erroneous measurements using the bad data detection method and BDId algorithm.

Table 5. Results of state estimation by test equation method, identification of erroneous measurements using bad data detection method and BDId algorithm[a].

Parameter	Etalon	Measurements		Test equation method		BDId
		without gross error	with gross error	BDD	SE	
P_{1-2}	-992.7	-993	-	V	-994	V
Q_{1-2}	-187.5	-183	-	V	-189	V
P_{2-1}	1010	1013	-	V	1011	V
Q_{2-1}	-468	-446	-	V	-464	V
P_{2-3}	**-1880**	**-1879**	**-1979**	V	**1982**	**E**
Q_{2-3}	29	34	-	V	48	V
P_{3-2}	**1896**	**1903**	**2003**	V	**2000**	**E**
Q_{3-2}	-92	-90	-	V	-85	V
The value of a objective function is 9.98						

Calculations have shown that measurements P_{2-3}, P_{3-2} corrupted by a cyberattack were defined by test equations method as reliable and were used by the state

[a] BDD – Bad Data Detection, SE – State Estimation, V- Valid, E – Erroneous.

estimation algorithm for computation of the estimated conditions. The estimates obtained considerably deviated from standard conditions though the value of objective function meets the χ-square criterion [5]. Analysis of measurements P_{2-3} and P_{3-2} using the BDId algorithm allowed identification of measurements as erroneous ones with the accuracy level of 0.12 (low level).

4.2 Simulation of cyberattacks that are not identified by the state estimation residuals

Here they give the results of calculations while simulating the 'false data injection' attack following the technique described in [2]. This technique was developed for the case when the problem of state estimation is solved using the classical method through the state vector x, and bad data are detected *a posteriori* based on the weighted residuals of estimation that are computed using the following formula:

$$\hat{r}_W = R_y^{-1/2} \left| \bar{y} - y(\hat{x}) \right|,$$

and for reliable measurements should not exceed the threshold of 3-3.5.

Based on the classical statement of the state estimation problem considering the relation between estimates of measurements $\hat{y}$ and estimates of the state vector $\hat{x}$ ($\hat{y} = H\hat{x}$, where H is a Jacobian matrix), cyberattacks were simulated according to [2]:

1. A non-zero vector c is specified that distorts the state vector components.
2. A vector of attacks $a = Hc$ of length m is formed, where m is the number of measurements.

A vector of erroneous measurements is determined as:

$$y_a = y + a .$$

3. State estimation is performed. State vector estimates obtained after quality assessment are equal to:

$$\hat{x}_a = \hat{x} + c$$

4. Estimation residuals are computed:

$$r_w = \left\| y_a - H\hat{x}_a \right\| = \left\| y + a - H(\hat{x} + c) \right\| =$$
$$= \left\| y - H\hat{x} + (a - Hc) \right\| = \left\| y - H\hat{x} \right\| .$$

In this case, we get the estimation residuals that are equal to residuals computed based on the state estimation results without a cyberattack.

Table 6 presents the results of state estimation and identification of erroneous measurements using the BDLd algorithm for the distorting vector $c = (0, 0, 20, 0, 0)$.

The results obtained evidence that despite the false data injected into the state vector, the method of weighted residuals analysis did not identify erroneous measurements, i.e., did not allow the cyberattack to be identified. Accuracy levels computed based on the BDId algorithm for P_{2-3} (0.126), Q_{2-3} (0.117), P_{3-2} (0.125), Q_{3-2} (0.117) allowed those measurements to be identified as erroneous.

Table 6. Results of state estimation by the classical method; identification of bad measurements using weighted residuals and BDId.

Parameter	Measurements without gross error	Attack	Measurements with gross error	Computations using classical method		BDId
				Estimates	Weighted residuals	
P_{1-2}	-993	0	-	-994	0.256	
Q_{1-2}	-183	0	-	-189	0.709	
P_{2-1}	1013	0	-	1011	0.257	
Q_{2-1}	-446	0	-	-464	1.86	
P_{2-3}	**-1879**	**-32.5**	**-1911**	**-1928**	**3.49**	**E**
Q_{2-3}	**34**	**-432.5**	**-398**	**-395**	**0.327**	**E**
P_{3-2}	**1903**	**33.3**	**1963**	**1946**	**3.05**	**E**
Q_{3-2}	**-90**	**434.8**	**345**	**339**	**0.682**	**E**
The value of a objective function is 20.17						

6 Conclusion

The paper proposes a data processing algorithm (as a preliminary stage of EPS state estimation) for identification of erroneous measurements caused by cyberattacks against SCADA and WAMS. This algorithm is based on the wavelet analysis and fuzzy sets. The findings have shown that its use can timely prevent the impact of successful cyberattacks on the EPS state estimation results and ensure reliable data control. The efficiency of this algorithm has been confirmed by experimental calculations.

Acknowledgement

The research was carried out as part of the scientific project III.17.4.2. of the basic research program SB RAS, reg. number AAAA-A17-117030310438-1.

References

1. Voropai N. I., Elektrichestvo **7**, 12-21 (2020)
2. Y. Liu, M. K. Reiter, and P. Ning, *Computer and Communications Security*, 21-32 (2009)

3. Khohlov M.V., *Methodological issues of studying the reliability of large energy systems*, 366-376 (2016)
4. L. Hu, Z. Wang, X. Liu, A. V. Vasilakos and F. E. Alsaadi, IET Control Theory & Applications **11(18)**, 3221-3232 (2017)
5. Gamm A.Z., Kolosok I.N., Identification of bad data in telemetry measurements in electric power systems, 152 (2000)
6. Glazunova A.M., Kolosok I.N., *Energy of Russia in XXI century: development, operation, and control*, 696-704 (Irkutsk, 2006)
7. A. Tarali and A. Abur, IEEE PES ISGT Europe, 1-8 (Berlin, 2012)
8. Khohlov M.V., Methodological issues of studying the reliability of large energy systems: Urgent reliability problems of energy systems, 557-566 (2015)
9. Kolosok, I.N., Gurina L.A., Informational and mathematical technologies in the science of control **1(17)**, 68-78 (2020)
10. G. A. Ortiz, D. G. Colomé and J. J. Quispe Puma, *IEEE ANDESCON*, 1-4 (Arequipa, 2016)
11. Kolosok, I.N., Gurina L.A., *Methodological issues of studying the reliability of large energy systems*, 238-247 (2019)
12. Kolosok I., Gurina L., *ICIEAM*, 1-4 (Moscow, Russia, 2018)
13. Bogatyrev L.L., Manusov V.Z., Sodnomdorzh D. Mathematical modeling of EPS conditions under uncertainty, 348 (1999)