# Monitoring and analysis of SCADA and WAMS data for EPS digitalization

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

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

**Abstract.** The properties of electric power systems (EPSs) are currently in the process of digital transformation, which should be taken into account when controlling them. Despite the numerous advantages of the digital transition, there are still problems with quality of the data used to control the EPS, and they are to a greater extent associated with the cybersecurity threats to the EPS information and communication infrastructure. The paper demonstrates the effect of changes in cybersecurity properties of the information and communication infrastructure on the quality of data streams coming from SCADA and WAMS, and reveals their complex interaction. The need has arisen to assess the quality of data during cyberattacks on systems for collecting, transmitting and processing the information. An algorithm is proposed to assess the quality of measurements based on the fuzzy logic.

**Keywords.** Quality of data, SCADA, WAMS, cyber attacks, fuzzy logic.

## 1 Introduction

The transition to the models of cyber-physical electric power systems (EPSs) is due to the digital transformation of the electric power industry, the interaction processes in which rest on new information and communication technologies, and digital models [1]. The cyber-physical system acquires new properties and distinctive features that must be taken into account to control it. In this context, such systems not only keep on facing the stability problems in terms of cybersecurity, but these problems become even more pronounced in terms of reliability requirements because of the increased vulnerability to cyberattacks on the information and communication subsystems [2,3]. The relevance of providing the EPS control with timely and reliable information is emphasized by the need to develop new methods and models for data representation based on artificial intelligence technologies.

Currently, the electric power system control is based on both SCADA measurements and synchronized vector measurements coming from WAMS measuring devices. The quality of SCADA and WAMS measurements is crucial not only for the development of automated control systems but also for the smooth operation of EPSs. The quality of information data flows is understood as the degree of completeness and reliability of information providing the required accuracy in the EPS control.

The paper proposes a method for processing measurement information based on the theory of fuzzy sets, given such cyber security requirements of the SCADA system and WAMS as timeliness, integrity, availability, and cyber resilience and confidentiality [4]. The development of the approach involved an analysis of the cyber-physical system properties and the identification of possible cyberattacks that reduce the quality of information data flows.

The paper shows that the incompleteness and inaccuracy of information increase due to cyberattacks on the SCADA system and WAMS, which can lead to the development and implementation of incorrect control actions and the adverse consequences for the EPS operation [4].

In the proposed method, an algorithm is developed to analyze the operating parameters based on fuzzy rules. This algorithm can also be used as a preliminary stage for data processing as a barrier to "bad" data in the state estimation of an EPS.

In the proposed method, an algorithm is developed to analyze the operating parameters based on fuzzy rules. This algorithm can also be used as a preliminary stage for data processing as a barrier to "bad" data in the state estimation of an EPS.

The use of artificial intelligence technologies in the analysis and processing of information flows will improve the efficiency of the EPS control and reliability.

## 2 Digital transformation of the properties of cyber-physical electric power systems

In Russia, as in other countries, the development of EPS is aimed at creating a cyber-physical system based on a single digital environment (CIM model), and introducing cybersecurity technologies and intelligent control methods in order to improve the reliability and transparency of EPS operation.

* Corresponding author: gurina@isem.irk.ru

The CIM (Common Information Model) based on the ODM (Open Model for Exchanging Power System Simulation Data) data format allows building models of any complexity, which can then be converted to any known data format or any new data format using additional plug-ins. ODM is an open model for data exchange in the modeling of power systems. ODM is an international open standard for data exchange in modeling and calculation of EPS, which supports dynamic calculations [5]. Based on CIM models, the IT infrastructure integrating the intelligent information, computing, and telecommunication environments should provide two-way communication between the information-communication and process subsystems of the cyber-physical EPS.

Transparency of the EPS operation requires the implementation of new systems for the collection, transmission and processing of information flows; the development of technologies and methods for modeling the studied processes and obtaining reliable real-time data on operating conditions for the control of EPSs.

The transition to the intelligent control of EPSs and the growing needs to monitor and analyze data, call for digital data processing technologies based on artificial intelligence methods:
- artificial neural networks and genetic algorithms;
- logical programming;
- ontological engineering;
- fuzzy logic, etc.

Despite all the obvious advantages of the digitalization of electric power systems, it makes them more vulnerable to cyberattacks, which is associated with a large scale (including spatial fragmentation) of the process part and the multicomponent nature (devices for collecting, transmitting and processing information at all levels of control) of the information and communication infrastructure of cyber-physical EPSs and their information interaction. At the level of hardware and software support of control, the risk of the occurrence of hidden threats is growing. The integration of IT infrastructure technologies contributes to the increase in the number of cyberattacks [4].

EPS control is based on data from the SCADA and WAMS. Cyberattacks aimed at the components of these systems or two-way data flows of information-communication and process systems can disrupt not only the control functions but also cause failures in the EPS operation.

The study in [4] demonstrates the effect of poor information quality, which results in false visualization of EPS operating conditions and generation of incorrect control actions due to cyberattacks on SCADA system and WAMS. Data quality analysis can determine the type of cyberattack and identify overlooked vulnerabilities.

## 3 The quality of data flows of the SCADA system and WAMS

The large-scale EPS monitoring involves both SCADA system and WAMS, in which the measurements come from PMU devices. In these conditions, the EPS can be controlled based on
- SCADA measurements;
- WAMS measurements;
- Mixed measurements.

Technologies of synchronized phasor measurements can increase the observability of the system and provide more accurate and timely information for control.

The quality of information data flows is understood as an extent to which the information is complete and reliable, which provides the required accuracy of solutions for control of EPS operating conditions.

In [6], the information for the EPS control is classified as follows:
- deterministic;
- probabilistic;
- uncertain.

Deterministic information is based on the laws of cause and effect relationships and is conditioned by a numerically unambiguous specification of the types of equipment, its composition and rated parameters.

Probabilistic information describes the stochastic nature of a change in the operating condition, the totality of the electric network components, which corresponds to a given behavior of EPS.

Uncertain information is divided into four groups:
- ambiguous;
- unknown;
- insufficient;
- unreliable.

The ambiguity of information refers to its multivariance due to various methods used to obtain and describe it. The lack of information about the components and operating parameters due to technical and physical factors leads to uncertainty. Various extents to which the information is unknown and insufficient reflect the incompleteness of information. Information unreliability arises when the model does not correspond to the modeled process, when there are measurement errors, data inaccuracy, etc.

However, the joint use of SCADA and WAMS measurements requires that the following issues be solved:
- high computational load;
- big data;
- weak conditionality of covariance matrices.

There arise serious data quality and cybersecurity problems that have complex interactions. A decrease in data quality, for example, can be a consequence of a successful cyberattack. At the same time, an analysis of data quality can determine the type of cyberattack and identify overlooked vulnerabilities [7]. To check the cybersecurity properties, it is necessary to develop methods for analyzing the data quality of SCADA and WAMS.

In this regard, the influence of cyberattacks on data quality was analyzed taking into account violations of cybersecurity properties [8] (Fig. 1).

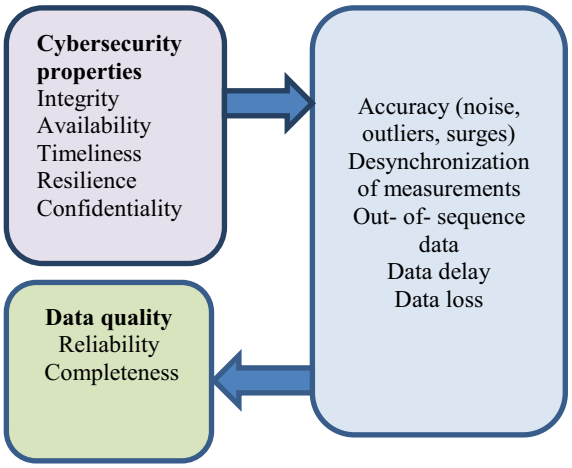


**Fig. 1.** The impact of cyberattacks on data quality.

In [5], a criterion for the quality of information and a method for its determination on the basis of the theory of fuzzy sets were introduced. The authors of [9], depending on the level of SCADA and WAMS measurements completeness and reliability, propose the measurement models for EPS state estimation. Consideration of the impact of cyberattacks on the completeness and reliability of information required the extension of a list of factors to assess the data quality. The study in [10–14] examines possible cyberattacks on information collection, transmission, and processing systems, reveals their vulnerabilities, and shows how the violations of the cyber security properties of the SCADA systems and WAMS affect the EPS control functions [4]. These studies indicate the need to take into account the following factors of the impact the cyberattack on the SCADA system and WAMS has on data quality:

- sequence;
- timeliness of data;
- data consistency.

The timeliness of data in real time reflects the uncertainty of information. Data consistency must be considered when SCADA and PMU measurements are used jointly.

## 4 A fuzzy system for data flow processing considering the cybersecurity properties of the SCADA system and WAMS

The proposed algorithm for evaluating data quality is based on the following algorithm:

1. Determine the level of information reliability;
2. Determine the level of information completeness;
3. Assess the information quality.

To determine the levels of reliability and completeness of information, we specify the linguistic variables (accuracy, sequence, consistency, timeliness, adequacy), determine the term sets and give their semantic definition. The developed fuzzy system for assessing the quality of measurement data is presented in Fig. 2.

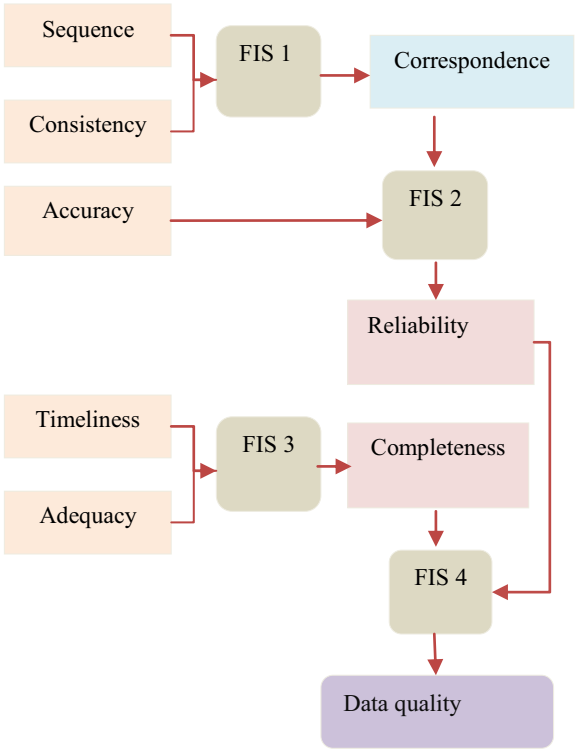


**Fig. 2.** Fuzzy system of data quality assessment.

## 5 Case study

A fuzzy system has been built to assess the quality of information, given the problems of EPS state estimation arising from cyberattacks on SCADA system and WAMS [15, 16].

The semantic description of input and output linguistic variables is presented in Tables 1-4.

**Table 1.** Levels of factors affecting the reliability of information.

| Level | Accuracy | Sequence | Consistency |
|---|---|---|---|
| Low 0-0,25 | The measurements contain errors due to cyberattacks, including those that cannot be detected | The sequence is broken | Data are not consistent |
| Medium 0,25-0,75 | The measurements contain errors due to cyberattacks, which can be detected by the bad data detection methods | Sequence is broken but there is a possibility of elimination (duplication, comparison) | Data are not consistent, but there is the possibility of duplication and restoration |

| | | | |
|---|---|---|---|
| High 0,75-1 | The measurements contain errors resulted from the errors of measuring devices, etc., which do not affect the accuracy of the EPS state estimation | Sequence is not broken | Data are consistent |

**Table 2.** Levels of factors affecting the completeness of information.

| Level | Timeliness | Adequacy |
|---|---|---|
| Low 0-0,25 | Big delay | No data |
| Medium 0,25-0,75 | Delay with the possibility of considering in measurement models | Data loss is not significant for solving the problem |
| High 0,75-1 | Measurements arrive without delay | Sufficient number of measurements are received |

**Table 3.** Completeness and reliability of data.

| Level | Completeness | Reliability |
|---|---|---|
| Low 0-0,25 | The system is not observable | Doubtful |
| Medium 0,25-0,75 | Ability to calculate missing measurement values | Erroneous |
| High 0,75-1 | Excessive measurements | Reliable |

**Table 4.** Data quality.

| Level | Quality |
|---|---|
| Low 0-0,25 | Network is unobservable and/or measurements are unreliable |
| Medium 0,25-0,75 | The use of bad data detection methods, validation of measurements, filtering errors, restoration of measurement flows, consideration of information aging will allow EPS state estimation with required accuracy |
| High 0,75-1 | Full reliable flow of information |

Figures 3-5 demonstrate the obtained three-dimensional surfaces of completeness, reliability and quality of information.

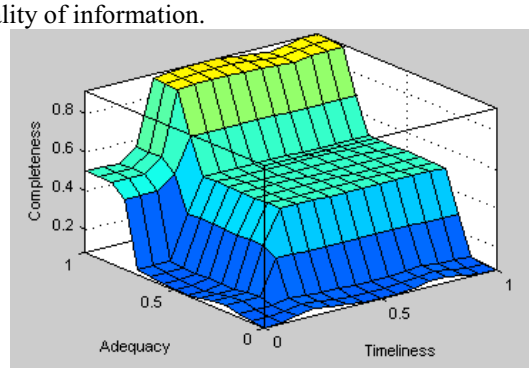


**Fig. 3.** Fuzzy system of data quality assessment.

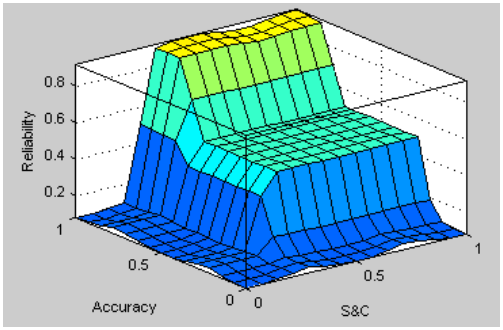


**Fig. 4.** The dependence of the information reliability on the data accuracy, sequence and consistency.

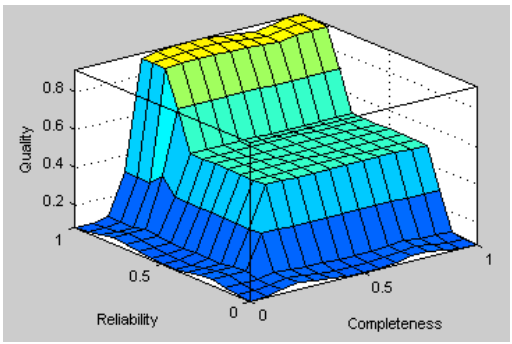


**Fig. 5.** Quality of measurements.

As can be seen from the graphs, the low level of any of the input factors caused by cyberattacks affects the quality of information (blue color), This especially affects the reliability of measurements.

This justifies the need to analyze the data when solving the state estimation problem, taking into account additional factors as a preliminary stage of data processing.

## 6 Conclusion

The EPS properties associated with the creation of a cyber-physical system are analyzed. The studies have shown an increased vulnerability of such systems to cyber attacks on information and communication infrastructure.

The interdependence between the cybersecurity properties of the SCADA system and WAMS and the quality of measurements is shown. An algorithm is proposed for assessing the data quality given violations of cybersecurity properties as a preliminary stage in EPS state estimation.

## Acknowledgement

This study is supported by the Siberian Branch of the Russian Academy of Sciences (Project III.17.4.2) of the Federal Program of Scientific Research (No. AAAAA-A17-117030310438-1) and partial by RFBR grant №19-07-00351 A

## References

1. L.V. Massel, Energy Policy **5**, 30-42 (2018)
2. S. Sridhar, A. Hahn and M. Govindarasu, *ISGT* (Washington, DC, 2012)
3. S. Sridhar, A. Hanh, M. Govindarasu, Proceeding of the IEEE. **100**, 210-224. (2012)
4. I. Kolosok, L. Gurina, *Methodological Problems in Reliability Study of Large Energy Systems* (Tashkent, Uzbekistan, 2019)
5. Kobec B.B., Volkova I.O. Innovative development of the electric power industry based on the Smart Grid concept, 208, 2010
6. N.V. Savina, L.A. Gurina, *Energy: Management, Quality and Efficiency of Energy Use* (Blagoveshchensk, Russia, 2003)
7. I. Kolosok, L. Gurina, *Industrial Engineering, Applications and Manufacturing* (Moscow, Russia, 2018)
8. I.N. Kolosok, L.A. Gurina, Information and Mathematical Technologies in Science and Management **2** (14), 40-51 (2019)
9. I.N. Kolosok, L.A. Gurina, Electricity **1**, 21-27 (2014)
10. Hua. Lin, Yi Deng, Sandeep Shukla, James Thorp, Lamine Mili, *Smart Grid Communications* (2012)
11. K. Gai, M. Qiu, Z. Ming, H. Zhao, L. Qiu, IEEE Transactions on Smart Grid **8** (5), 2431-2439 (2017)
12. Longfei Wei, Luis Puche Rondon, Amir Moghadasi, Arif I. Sarwat, *T&D* (2018)
13. Mohd Rihan, Mukhtar Ahmad, M. Salim Beg, *Smart Grid and Renewable Energy* (2013)
14. S. Sridhar, A. Hahn and M. Govindarasu, *ISGT* (Washington, DC, 2012)
15. Yao Liu, Peng Ning, Michael K. Reiter, *CCS'09* (Chicago, Illinois, USA, 2009)
16. Zanouz S., Rogers K. M., Berthier R., Bobba R.B., Sanders W.H, Overbye T.J., IEEE Transactions on Smart Grid **3**, 1790-1799 (2012)