Data-driven failure analysis for the cyber physical infrastructures

Viacheslav Belenko
Branch office of LG Electronics Inc. (Korea) in St.Petersburg
St.Petersburg, Russia
viacheslav.belenko@lge.com

Valery Chernenko
Branch office of LG Electronics Inc. (Korea) in St.Petersburg
St.Petersburg, Russia
valery.chernenko@lge.com

Vasiliy Krundyshev
Peter the Great St.Petersburg Polytechnic University
St.Petersburg, Russia
vmk@ibks.spbstu.ru

Maxim Kalinin
Peter the Great St.Petersburg Polytechnic University
St.Petersburg, Russia
max@ibks.spbstu.ru

***Abstract*—Digital transformation is a main driver of a modern approach to providing cyber safety and ecurity in the environment of smart systems: smart home, IIoT, smart building, smart megapolis, VANET, FANET, WSN, etc. For traditional computer networks, security was to ensure confidentiality, integrity, and availability of data. The last decade, with the advent of dynamic machine-to-machine infrastructures, the goal is forced to be ensured in safety and reliability of physical processes at cyber system. Due to the increased mobility of topology and the growing amount of data (Big Data) undergoing the processing, traditional methods of system analysis become ineffective, so the researchers are faced with the task of creating new methods for ensuring cyber security that meet new challenges. This paper outlines the essence of the approach to the detection of weaknesses caused by failures of the components of cyber physical infrastructure. The paper discusses a method of data-driven analysis for failure detection and prediction. The proposed technique is based on the modified 'k' method of nearest neighbors (KNN) extended with application of Dempster-Shafer (DS) theory and our suggestion to estimate the spatial-temporal correlation of the connected devices in the cyber physical environment. Our method shows above 99%-level of effectiveness comparing to common fault management approaches.**

Keywords—cyber physical system, Dempster-Shafer, security, data-driven analysis, fault, failure, internet of things, spatial-temporal correlation, smart building, IIoT, IoT, KNN

I. Introduction

Last decade, due to the active development of dynamic machine-to-machine digital infrastructures (for example, IoT [1], IIoT [2], WSN [3], MANET [4], VANET [5], a concept of smart homes, buildings, and cities), the object of protection acquires a new glance as an element of the cyber physical space, in which traditional read/write operations have real physical consequences. The advent of Internet of Things (IoT) has led to a technological shift that could potentially change the scope of our social and economic landscape. The emergence of IoT will affect urban planning and design just as it will affect network computing services, as well as security services. The Internet of Things is a core of the smart buildings and smart cities in which we live. Technology forecasts suggest that about 30.7 billion IoT devices will be installed by 2020. Many of these devices will be deployed in critical infrastructures of cyber physical systems [6].

For convenience of classification, Rob Van Kranenburg introduced 4 levels of Internet of Things - by "coverage" [7]: BAN (Body Area Network), LAN (Local Area Network), WAN (Wide Area Network), VWAN (Very Wide Area Network). On each level of propagation, IoT is a cyber space equipped with integrated technological systems. Traditionally, these systems were installed separately. Due to the interoperability and interdependence of data between these systems, the concept of fully integrated intelligent infrastructure has emerged [8]. This infrastructure is aimed at improving energy efficiency and reducing operating costs [9].

While cyber physical infrastructure that fully integrates your preferences and actions can offer almost limitless possibilities, there is also a risk of cyberattacks targeting both cyber system itself and all systems associated with it. Cybercriminals who previously concentrated on corporate networks and online services are now increasingly targeting industrial control systems (ICS) [10]. Overcoming a retail website is no longer the target of a criminal online presence. Critically important infrastructure such as e-hospitals, e-government offices and automated power plants can now be their targets. Many cyberattacks are aimed at disabling various important devices and meta-structures of devices. All this can lead to large economic losses or increase the risk of emergency. Below there are some kinds of security threats that can be detected in cyber physical infrastructures and possible consequences of their activity:

- system hardware failure: system malfunction;
- power outages: system disruption;
- impact of viruses and trojans on the system: malfunctions in the system software, disruption of work or disabling of the system;
- user mistakes: possible system failures due to improper use of equipment;
- interception of information transmitted over wired and wireless communication channels: violation of the confidentiality of information transmitted over the channel, it is possible to seize control of the system;
- presence of inner intruders (security guards, service men, cleaners, etc.): system malfunctions due to improper maintenance of equipment, the level of danger depends on the degree of insider access to the system.

Thus, a new mechanism is needed that would allow timely monitoring of failures and predicting their possible occurrence in the future. Today there are many methods for detecting failures in cyber physical environment of connected devices, e.g. [11, 12, 13]. However, it is important not only timely detection of safety failures with high level of accuracy, but also the ability to predict the further occurrence of device faults. The goal of this paper is to develop a highly effective method for analysis and predicting of cyber physical infrastructure failures using machine learning. To present our work, the paper is structures as follows: Section 2 analyzes the known methods for detecting and predicting failures in cyber physical infrastructures, Section 3 describes the proposed failure analysis method that provides our solution, Section 4 describes our experimental results, and Section 5 makes the work conclusion.

II. The Related Works

The overview of the existing apparatuses of the fault management allows the following background:

K-nearest neighbors (KNN) method [14]:

- mathematical structure at the heart of the method – calculate distance function;
- fault detection rate – 81%;
- used to detect malfunctions of an oil-filled power transformer.

Modified KNN (KNN+DS method) [15]:

- calculating the distance function with application of Dempster-Shafer (DS) theory;
- fault detection rate – 95%;
- used to detect malfunctions of an oil-filled power transformer.

Support vector machine (SVM) + KNN [16]:

- construction of a hyperplane for a set of points, finding the distance function;
- fault detection rate – 75.3%;
- utilized to detect failures in various machines. For detection, the SVM method is used, and for its improvement, the KNN method is utilized.

Artificial neural networks (ANN) [17, 18].

- Mahalanobis distance calculation.
- ability to predict the deterioration of the details considered in the article environment.

SVM + back propagation neural network (BPNN) [19].

- search for a hyperplane for a set of points;
- fault detection rate – for SVM 91.93%; for BPNN 85.57%;
- detection of failures at a thermal power plant. SVM and BPNN are used for classification.

Fuzzy neural network (FNN) [20].

- counting fuzzy Gaussian function of membership;
- fault detection rate – 93.37%;
- applied in wastewater treatment systems. Based on the predicted sensor validity index (SVI) values and sensor outputs. To predict SVI values, FNN is entered with multiple inputs and one output (MISO). On the basis of various water quality sensors, water quality is assessed and, based on this, a possible failure in the treatment system is predicted.

The considered methods are quite diverse, however, none of them have a sufficiently high accuracy of fault detection, except KNN+DS method. The further section presents our proposal to increase accuracy of KNN+DS method. Our method is based on the data-driven analysis received from the devices of the digital infrastructure and further building the Spatial-Correlation Consistency Regions (SCCR) for the neighbor devices. We named our method as KNN+DS+SCCR.

III. The Proposed Method for Data-Driven Failure Analysis

The proposed method for data-driven failure analysis consists of a module for calculating regions of data consistency and modules for predicting the failures (Fig. 1).

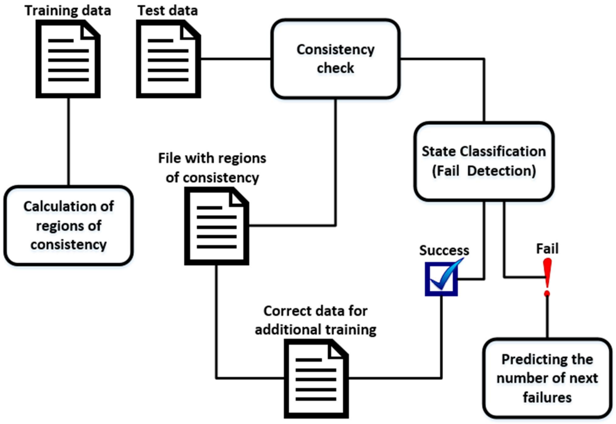
Training data
Test data
Consistency check
State Classification (Fail Detection)
File with regions of consistency
Calculation of regions of consistency
Success
Fail
Correct data for additional training
Predicting the number of next failures

Fig. 1. The scheme of the data-driven failure analysis.

The proposed system contains the following units:

- module for calculating the regions of consistency;
- module for consistency checking;
- databases with regions of consistency;
- fault detection module;
- failure prediction module.

The fault detection module is based on the modified KNN method and performs the classification of the sensor indicators coming from the devices. The result is a list of sensor states, based on which further prediction of failures is made. The fault prediction module is based on a method that uses information about the past number of failures. The result of this method is a number of failures that will occur for a particular pair of devices. Failure detection and prediction consists of the following steps:

1. The module for calculating the regions of consistency receives data from *N* sensors from each device. The data is a sequence of sensor readings for a certain period of time *T*.

2. Based on the training data $S_i(t,T)$ and $S_j(t, T)$ taken up to a predetermined time *T*, for each pair of counters that are neighbors, a region of consistency of spatial correlations is calculated.

3. The regions of consistency are constructed for each device, and these data are stored in a database.

4. To detect failures, the KNN method is extended with the Dempster-Shafer theory. Before applying this method, the incoming data is also split in pairs for each pair of devices and checked for consistency using saved regions of consistency. If the data is consistent, they are further classified, otherwise the data is not processed.

5. The result of the classification is a sequence of assessments of the state of each pair of devices ("Normal_state" or "Fault_state").

6. Next, prediction of failures occurs based on the number of "Fault_state" ratings for each pair of devices using the method based on the Weibull distribution.

A. Module for calculating the regions of consistency

In this module, correlation patterns are built for each pair of devices located in one correlation area, which is determined manually. There are several devices in the correlation area, while devices may belong to several different areas. The area of correlation, the Spatial-Correlation Consistency Regions (SCCR), ensures that the devices in it are spatially and temporaly correlated, which is required to solve the task.

The formula for the center of the SCCR ellipse is:

$$EWMA = a * p_t + (1 - a) * EWMA_{t-1},$$

where $EWMA$ – the value of the exponentially weighted moving average at the point *t* (the last value, in the case of a time series); $EWMA_{t\text{-}1}$ – the value of the exponentially weighted moving average at the point *t*-1 (the previous value in the case of a time series); p_t – the value of the original function at time t (the last value, in the case of a time series); a – the coefficient characterizing the rate of decrease in weights, takes a value from 0 to 1, the smaller its value, the greater the influence of the previous values on the current average value.

Main and minor axes of the SCCR ellipse are calculated using Principal Component Analysis (PCA). The orthogonal principal components $\vec{a}$ and $\vec{b}$ determine the angle of rotation of the main axis to the axis θ, and the lengths of these axes are set as three deviations $3\sigma_a$ and $3\sigma_b$ In general, the multidimensional case, the process of isolating the main components occurs as follows:

1. The center of the data cloud is searched, and a new origin is transferred there – this is the zero principal component (PC0).

2. The direction of the maximum data change is selected – this is the first main component (PC1).

3. If the data is not fully described (the noise is large), then another direction (PC2) is chosen – a perpendicular to the first, so as to describe the remaining change in the data, etc.

Fig. 2 presents the work algorithm of SCCR calculation.

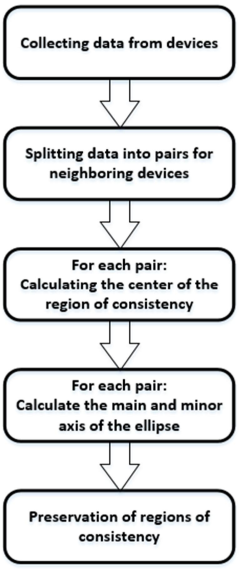
Collecting data from devices
Splitting data into pairs for neighboring devices
For each pair: Calculating the center of the region of consistency
For each pair: Calculate the main and minor axis of the ellipse
Preservation of regions of consistency

Fig. 2. Work algorithm of SCCR calculation.

B. Fault detection module

To classify the state of the device, an improved "k" method of nearest neighbors is used using the Dempster–Shafer theory.

BPA is a basic probability assignment. For one neighbor it is calculated as $m_{s,i}(\{C_q\}) = \alpha_0 \phi_q(d^{s,i})$, where $\alpha_0 = 0.95$, $d^{s,i}$ is a distance between the state being classified and the nearest neighbor. The value of ϕ_q is calculated as $\phi_q(d^{s,i}) = e^{-\gamma_q d^\beta}$, where $\beta=2$, $\gamma_q=1/d_q^\beta$, d_q – average distance between the closest distances to the class C_q.

The total BPA for the set of all neighbors of the class C_q is calculated as $m_q^s(\{C_q\}) = 1 - \prod_{x_i \in \Phi_q^s}(1 - \alpha_0 \phi_q(d^{s,i}))$.

Fig. 3 presents fault detection algoritm. Fig. 4 shows a sample of the SCCR ellipse for 2 devices (the axes of coordinates are the values of sensor readings (voltage, temperature, etc.) at a certain time; sensors' data are marked with dark green dots).

C. Failure prediction module

To predict failures, the method described in [21] was used. This method is based on the use of the Weibull distribution to assess the reliability of various systems. Suppose that a set of classified states of a device of size N that was received in the period from 0 to t_c contains X failures. You need to know how many failures will be in the time interval $[t_c, t_w]$.

It is desirable to have one prediction $\hat{Y}$ for the number of failures Y in the future interval $[t_c, t_w]$. Considering the observed (non-zero) number of failures X for the period t_c, the predicted number of failures is calculated as $\hat{Y} = N * \hat{q}$, where $\hat{q}$ is calculated as $\hat{q} = \left[1 - \left(\frac{X}{N}\right)\right] - [1 - \left(\frac{X}{N}\right)]^{(\frac{t_w}{t_c})^\beta}$, where t_w– upper limit of the time interval, t_c – the lower limit of the time interval, β– form parameter (depending on the failure rate).

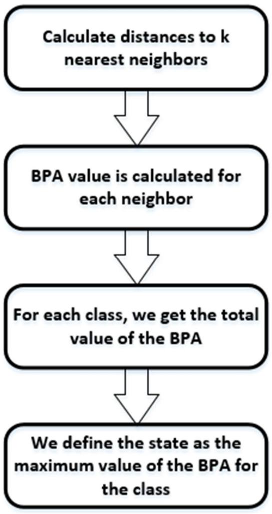
Calculate distances to k nearest neighbors
BPA value is calculated for each neighbor
For each class, we get the total value of the BPA
We define the state as the maximum value of the BPA for the class

Fig. 3. Fault detection algorithm.

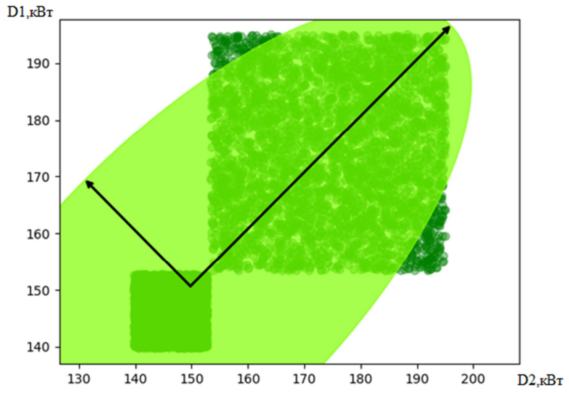

Fig. 4. An example of SCCR.

IV. Experimental results of proposed method

For the experiment, three different situations of occurrence of failures in the cyber physical environment are selected [15]: random failures, periodic failures, noise. A dataset of 5000 values of indicators of various sensors for training and a test dataset of 1000 values were used. The experiment was conducted for 20 devices.

Figures 5 and 6 present the outputs of the suggested method for 20 devices for random and periodic failures. The figures show the average error of the developed method for the number of neighbors (1…40). The traditional KNN method copes with the task of classification worst of all (with a maximum number of neighbors k = 40, its accuracy is around 78%). The results of the KNN+DS method are better. The accuracy of 99.997% is reached by the suggested KNN+DS+SCCR method.

V. Conclusion

The potential of IoT and smart buildings, applications and prospects is growing rapidly. Nevertheless, there are several open problems for research, varying from protection against cyber attacks [2, 4, 6, 22] to failures detection [23, 24], make research in the IoT environment and smart buildings very attractive. Many important topics in the field of protection against cyber threats and disruptions are currently at the stage of intensive discussion.

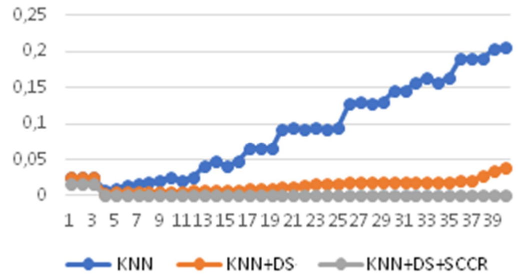

Fig. 5. Classification accuracy of random failures.

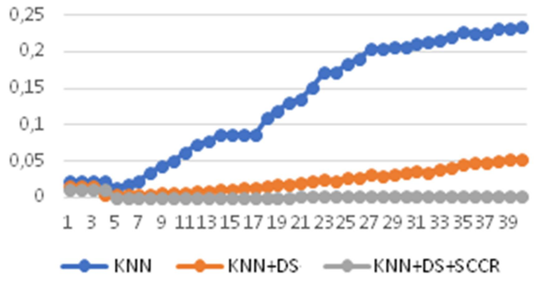

Fig. 6. Classification accuracy of periodic failures

The proposed data-driven method assembles the abilities of traditional KNN approach [14], Dempster-Shafer calculations [15] and data consistency checking. The obtained results have shown a high failure detection accuracy, above 99%. The further work is targeted for resources estimation and implementation of the suggested solution for the area of smart building management systems. Also, our work is concerning with a smart cyber physical infrastructure sustainability control.

Acknowledgment

Great appreciation to the Branch office of LG Electronics Inc. (Korea) in St.Petersburg (Russia) for assistance and support in synthetic modeling experiments of the smart building IoT environmant.

Funding

The research is funded by LG Electronics Inc.

References

[1] D. Lavrova, A. Pechenkin and V. Gluhov, “Applying correlation analysis methods to control flow violation detection in the internet of things,” Automatic Control and Computer Sciences. 2015. vol. 49, is. 8, pp. 735-740.

[2] E. Sisinni. et al., “Industrial Internet of Things: Challenges, Opportunities, and Directions,” IEEE Transactions on Industrial Informatics, 2018. vol. 14, no. 11, pp. 4724-4734.

[3] D. Parker, M. Stojanovic and C. Yu, “Exploiting temporal and spatial correlation in wireless sensor networks,” 2013 Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, 2013, pp. 442-446.

[4] Singh R. and Nand P., Literature review of routing attacks in MANET // 2016 International Conference on Computing, Communication and Automation (ICCCA). Noida. 2016. pp. 525-530.

[5] V. Belenko, V. Chernenko, M. Kalinin and V. Krundyshev, "Evaluation of GAN Applicability for Intrusion Detection in Self-Organizing Networks of Cyber Physical Systems," 2018 International Russian Automation Conference (RusAutoCon), Sochi, 2018, pp. 1-7.

[6] Protecting Smart Buildings from Cyber Attacks. https://medium.com/iot-security-institute/protecting-smart-buildings-from-cyber-attacks-b6a1ad2f4cd/

[7] IoT Interview Series: 5 questions with Rob van Kranenburg of the Internet of Things Council https://www.postscapes.com/iot-voices/interviews/iot-interview-series-5-questions-rob-van-kranenburg-internet-things-council/

[8] H. Chen, P. Chou, S. Duri, H. Lei and J. Reason, "The Design and Implementation of a Smart Building Control System," 2009 IEEE International Conference on e-Business Engineering, Macau, 2009, pp. 255-262.

[9] A. Crooks, K. Schechtner, A. K. Dey and A. Hudson-Smith, "Creating Smart Buildings and Cities," in IEEE Pervasive Computing, vol. 16, no. 2, pp. 23-25, April-June 2017.

[10] V. Belenko, V. Krundyshev and M. Kalinin, "Synthetic datasets generation for intrusion detection in VANET," Proceedings of the 11th International Conference on Security of Information and Networks, 2018.

[11] W. Kim and S. Katipamula, "A Review of Fault Detection and Diagnostics Methods for Building Systems. Science and Technology for the Built Environment," 2017. DOI: 10.1080/23744731.2017.1318008.

[12] R. C. Luo, K. L. Su and K. H. Tsai, "Fire detection and isolation for intelligent building system using adaptive sensory fusion method," Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292), Washington, DC, USA, 2002, pp. 1777-1781 vol.2.

[13] Y. Kim, R. Sharifi, Y. Cha and R. Langari, "Sensor fault diagnosis of smart buildings," 2010.

[14] F. Yu, J. Liu and D. Liu, "An approach for fault diagnosis based on an improved k-nearest neighbor algorithm," 2016 35th Chinese Control Conference (CCC), Chengdu, 2016, pp. 6521-6525.

[15] T. Denoeux, "A k-nearest neighbor classification rule based on Dempster-Shafer theory," in IEEE Transactions on Systems, Man, and Cybernetics, vol. 25, no. 5, pp. 804-813.

[16] A. Andre, E. Beltrame and J.Wainer, "A combination of support vector machine and k-nearest neighbors for machine fault detection," 2013, DOI: 10.1080/08839514.2013.747370.

[17] P. Bangalore and L. B. Tjernberg, "An Artificial Neural Network Approach for Early Fault Detection of Gearbox Bearings," in IEEE Transactions on Smart Grid, 2015, vol. 6, no. 2, pp. 980-987.

[18] S. Rajakarunakaran, et al. "Artificial neural network approach for fault detection in rotary system," Appl. Soft Comput, 2008, 8, pp. 740-748.

[19] K. Chen, L. Chen, M. Chen. and C. Lee, "Using SVM based method for equipment fault detection in a thermal power plant," Computers in Industry, 2011, 62, pp. 42-50.

[20] H. Han, Y. Li and J. Qiao, "A fuzzy neural network approach for online fault detection in waste water treatment process," Computers & Electrical Engineering, 2014, 40, pp. 2216-2226.

[21] D. Nordman and W. Meeker, "Weibull Prediction Intervals for a Future Number of Failures," 2000, Technometrics. 44. DOI: 10.1198/004017002753398191.

[22] D. S. Lavrova, "An Approach to Developing the SIEM System for the Internet of Things," Automatic Control and Computer Sciences, 2016. 50 (8), pp. 673-681.

[23] D. Lavrova, M. Poltavtseva, A. Shtyrkina, and P. Zegzhda, "Detection of cyber threats to network infrastructure of digital production based on the methods of Big Data and multifractal analysis of traffic", SHS Web of Conferences 44, 00007 (2018), CC-TESC2018WoS.

[24] E. Pavlenko, D. Zegzhda, "Sustainability of cyber-physical systems in the context of targeted destructive influences", 2018 IEEE Industrial Cyber-Physical Systems (ICPS), 2018, pp. 830-834.