Machine learning based approach to detection of anomalous data from sensors in cyber-physical water supply systems

A V Meleshko¹,², V A Desnitsky¹,* and I V Kotenko¹

¹St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), 199178, 39, 14-th Linia, St. Petersburg, Russia
²ITMO University, 197101, 49, Kronverkskiy pr., St. Petersburg, Russia

*desnitsky@comsec.spb.ru

Abstract. The paper proposes an approach to detection of anomalous data from sensors in cyber-physical systems on an example of a water supply supply system. The approach is based on methods of machine learning and modeling of technical systems. The primary data for machine learning was obtained on the developed software/hardware prototype of the water supply system by using a number of microcontrollers, sensors and actuators. The experiments confirmed the feasibility of the proposed approach. Several machine learning methods from scikit-learn library of the Python programming language were tested. As a result of the experiments we identified a learning method and its parameters ensuring the highest accuracy of the abnormal situations recognition.

1. Introduction

The paper is aimed at solving the problem of detecting anomalous data in modern cyber-physical systems on an example of a water supply system. The significance of this task is due to the critical character of such systems, the high risks of compromising security of their operation for end users, government and commercial organizations. Provided high automation of management processes and water supply, a malicious modification by an intruder of operational data on the state of the water level in a tank can lead to disastrous consequences associated with overflow of the tank and flooding of adjacent territories or shallowing and interruptions in water supply to consumers.

The task of monitoring and timely detection of anomalous data from sensors in critical water supply systems is an extremely relevant and requires effective solutions. The complexity of organizing the effective detection of anomalous data from sensors is related to the need for joint (group) automated analysis and data processing from an undefined set of sensors by rules that cannot be formed previously, and taking into account the previous history of readings.

The search for anomalous situations can be carried out by several methods, for example, such as (i) the creation of strict rules that monitor anomalous situations or (ii) the use of artificial intelligence methods, such as machine learning [1]. The application of the rules involves the comparison of data received from sensors of the cyber-physical system with certain patterns by using the constructions 'if' and 'else'. This approach has the following limitations. The rules relate to the specific logic of the system and when it changes it will have to be rewritten. Besides, the development of rules requires a deep understanding of the decision-making process [1].
At the same time, application of artificial intelligence methods for solving the problem of searching for anomalous data implies the development of dependencies between the sensors of the cyber-physical system and further prediction of their anomalies based on the existing knowledge. Having some samples of anomalous and non-anomalous data, it is possible to construct a classifier that is able to recognize anomalous situations, using methods of machine learning.

In this paper proposes an approach and software and hardware implementations of components for detection of anomalous data from sensors of water supply control system by using k-nearest neighbors and naive Bayesian models. Training and test samples are formed on simulated system operation scenarios. The experiments performed on the software/hardware prototype of the water supply system showed a fairly high accuracy of detection of anomalies.

The rest of the paper is organized as follows. Section 2 provides an overview of the related work in the field. Section 3 reveals the essence of the proposed approach to the detection of anomalous data. Section 4 presents the results of experiments and discussion. The last section concludes the paper.

2. Related work
Consider an analysis of existing works in the field of detection anomalous data in systems with cyber-physical devices. Cao, et al. proposed an approach to detecting anomalies in the logs of web servers, using machine learning [2]. The log files can contain detailed information about operation of the system, and allow tracing attacks, including SQL injection attacks or cross-site scripting. A two-level machine learning algorithm is used. At the first level one selects normal log files, using a decision tree classifier. At the second level, a normal data set is modeled, using a hidden Markov model. Experiments on more than 4 million original log messages demonstrated a detection rate of 93.54%.

Che, et al. are describing a learning model for the analysis of a transport cyber-physical system [3]. The purpose of building such a model is to find out the features of malicious attacks on Android devices. The model includes two main components, a handling processor and a mechanism for detecting malicious attacks. Detection accuracy of the model exceeds four alternative algorithms. Specifically the resulting accuracy is 12.61% higher than Softmax Regression, 5.76% higher than Decision Trees, 3.20% higher than Supporting Vector method, and 2.61% higher than RandomForest method.

Yavanoglu and Aydos compare data sets that are often used in machine learning methods being the main tools for analyzing network traffic and detecting anomalies in it [4]. Henkel presents a series of papers that covers a wide range of research in the field of lightweight encryption for Internet of Things, modeling cyber-physical systems and detection of anomalies [5]. The issues of finding a compromise between energy consumption and reliability are regarded. In particular, the presented platform allows adapting the power consumption, performance or reliability, depending on the priority.

Settanni, et al. describe how the detection of anomalies allows timely detection of critical threats on the base of certain security metrics [6]. This allows creating a process of self-adaptation, taking into account the detected attack, and also allows mitigating the impact of the attack on the cyber-physical system. Settanni, et al. tested their methods on a test-bed that simulates a cyber-physical production system. Also they modeled an attack aimed at remotely connecting to a programmable logic controller (PLC) and changing its operation algorithm. By analyzing the diagnostic buffer messages, it was revealed that the IP address of one of the clients accessing the PLC is not in the allowed range, and the anomaly detection system issued a warning.

Casas, et al. present an analysis environment for big data that is capable of analyzing and storing a large number of structured and unstructured heterogeneous data with their streaming and packet based processing [7]. Several algorithms have been implemented to detect network anomalies arisen as a result of DDoS attacks and Ping flood attacks by using methods of machine learning without a teacher.

Hosic, et al. proposed a method for the evolution of decision trees by applying methods of genetic programming [8]. This method is aimed at detecting network anomalies, such as disabling the device. According to the approach one extracts more than a dozen functions from network captures of packets
and streams, normalizes them and links them in decision trees, using fuzzy logic operators. Decision Tree apparatus was tested experimentally on a large-scale virtualized network. Five of the nine experiments resulting in 100% average accuracy, while the other two reached more than 98% average accuracy.

Ferragut, et al. proposed a detection mechanism for attacks of replacement of data in cyber-physical systems [9]. For example, when attacking Pacific Gas & Electric's Metcalf substation in northern California, attackers manipulated the sensors of the system, so the control center did not notice it. The successful operation of complex cyber-physical system depends on reliable control, that is a mechanism that receives data from the sensors as input data and produces a solution. It is the integrity of the data from the sensors that may be at risk, as they are usually the least protected [9]. The described detection mechanism is aimed at protecting the operation of data transmission and distribution in the power supply subsystem of the system. Basic physical relationships are derived for detection, using cross-sensory analytics in order to detect sensor failures, replay attacks and other data integrity problems in real time. The mechanism works on the basis of a neural network, acting on the voltage and current readings. The testing was conducted on a standard IEEE 30-bus power bus.

An approach is proposed to identify incidents of violation of the integrity of structures of hydraulic installations, including deformation, corrosion, etc., using microcontrollers and sensors, and the apparatus of correlation rules for events recorded using these sensors [10]. Analysis of the current literature in the field of the research showed that, the task of identifying anomalous data in water supply systems is stated in the given formulation for the first time. At the same time, the features of this paper, which distinguish it from existing works, include (i) the prototyping of the water supply system and the acquisition of primary data from sensors taken on a prototype of the developed cyber-physical system and (ii) the results of experimental studies on anomalies, using various machine learning methods.

3. Approach to identify anomalous data
The cyber-physical water supply system includes a number of hardware sensors distributed in the respective locations and responsible for collecting data on characteristics of water and environmental flows, such as pressure, water temperature, pressure, flow rate, etc. The proposed approach allows detecting anomalous data sets of sensors from sensors, which may be a result of both technical failures and deliberate actions of an intruder of cyber-physical security [11]. A distinctive feature of the approach is that it allows analyzing not only the readings of single sensors individually, but also take into account the deviations in the system behavior expressed by group characteristics based on the values for a certain set of sensors. For example, a water pressure sensor at one geographic location may correlate with data from a pressure sensor at some other location, taking into account possible temporal deviations. A correlation between data from a water flow sensor and a water level sensor as well as correlation between data from a water level sensor and a pressure sensor are also possible.

Figure 1 shows the general scheme underlying in the proposed approach to identifying anomalous data.

On a constructed prototype the original data from sensors are collected. Binary classifiers are built on them. The testing procedure is used to determine the accuracy of the classifiers. Further, the classifiers with the highest accuracy are used in dynamics to classify new sets of incoming data from sensors of the system on normal or anomalous ones. At the same time, the belongingness of some data set to the anomalous ones will denote a possible attack mounted on the system.

Collecting data records from sensors in the guarantied absence of failures and attacking effects on sensors allows empirically obtaining a sample characterizing the possible interdependencies of the correct (not anomalous) operation of the sensors [12]. In a similar way a sample of anomalous data is constructed by simulating specific attack effects on the sensors of the implemented model of the water supply system. In particular attacks of changing water level values to false ones are modeled, including attacks with the aim of possible overflow of the system’s tanks, flooding of the nearby space and bringing the system into an emergency state.
Further, on the basis of the obtained data samples, the classifiers are built. Each classifier is based on a certain machine learning method and allows classifying new (not previously analyzed) data samples into their normality/anomaly with certain accuracy. As a result the proposed approach allows for a large sample of the input data (i.e. records of readings of all sensors available in the system) to detect some correlations between them according to previously undefined laws and under conditions the immutability of the nature of the system work.

Figure 1. Approach to detecting anomalous data from sensors.

4. Implementation and experiments
For the practical implementation of the proposed approach a software/hardware prototype of the water supply system was developed by using a range of hardware sensors measuring physical characteristics. The prototype contains two tanks, water level and water pressure sensors, a controlled crane, a pump, a water flow sensor and control elements. The control elements include (i) an Arduino controller for processing data from sensors and to control a crane and pump; (ii) a power source; and (iii) a Raspberry Pi 3 single board computer to perform the monitoring functions. On Raspberry it is installed a program, that receives information on the status of the sensors and actuators from the Arduino and records them into a database. The monitoring program represents a Python web application developed by using Django framework. To enable remote viewing of the status of the prototype, the nginx and gunicorn servers are deployed on the Raspberry Pi, which allow connecting to the Raspberry Pi and using a web browser to view information from the database, as well as remotely influencing actuators (turning on the pump or opening the crane). The scheme of the test system prototype is shown in figure 2.

Under the gravity the water flows from tank 1 to tank 2, simulating the flow of a river. A controlled crane plays a role of a dam shutter. To monitor the water level in the tanks, there are several water level sensors (three ones in each tank). If the water level becomes critical, then the controlled crane is closed, and water from tank 2 is passed back to tank 1 by means of the pump. That is the real conditions are modeled – water on one side goes downstream, and on the other side new water comes.

A pressure sensor is also used as an alternative method for detecting water levels. If the level sensors indicate that the tank is full, and the pressure sensor shows low pressure, this is a sign of a possible attack.
The accumulation of a sufficient amount of different types of data, using the prototype, allowed application of machine learning methods, building classifiers and detecting normal and anomalous (attacking) data by means of them.

All values of water level sensors are interpreted as binary variables. The sensors are located at a fixed distance from each other, which allows for the readings of individual sensors to deduce on the degree of fullness of the tank. We get a four intervals, the triggering of each next sensor indicates that the tank was filled by another 25% (0-25% - no one sensor triggered; 26-50% - the first sensor (W1) triggered; 51-75% - the second sensor (W2) triggered; 76-100% - the third sensor (W3) triggered). For example, the first sensor triggered shows that the tank has been filled to at least 25%, and all three sensors triggered shows that at least 75%.

The pressure sensor includes a resistor, which changes its resistance depending on the force applied to it. The database does not store the resistance value, but contains the interpreted tank fullness in percents. The resistance without water and resistance with a full tank and was measured and calculated by a formula for converting to percentages. I.e. the current resistance readings are divided by the resistance at the full tank.

Therefore the format of data for training the model is given in the following form: 

\[
<1, 1, 0, 72.8611>
\]

Here the integers indicate values of the binary sensors (from the lower to the
upper, $W_1$, $W_2$, $W_3$, see figure 2), and the fractional value mean the actual value from the pressure sensor.

Consider an example of an anomalous dataset: $<1, 1, 0, 79.8261>$. According to the pressure sensor readings, the tank fullness is more than 75%, while the third water level sensor should have triggered, but it has not happened.

Solving the problem of finding anomalies on the test prototype was conducted according to the following algorithm:
- collect a training dataset consisting of 500 instances of events of each type;
- develop a classifier for each of the learning methods under consideration by using scikit-learn library for Python;
- collect new data from the prototype (750 copies in total), which the developed classifier was tested on;
- check recognition accuracy.

Table 1 presents the results of the testing of the implemented detection procedures on the base of the constructed datasets for each of the machine learning methods conducted by using corresponding scikit-learn classes.

| Method                      | Main parameters                                    | Accuracy on the built dataset |
|-----------------------------|----------------------------------------------------|-------------------------------|
| k-nearest neighbors method  | $n_{neighbors}$ is a number of neighbor samples used; $1, 2$ and $5$ values were used; $algorithm$ is an algorithm used to compute the nearest neighbors ($auto, ball\_tree, kd\_tree, brute$). The value used is $auto$ (automatic algorithm selection) | Accuracy is 83-87%           |
| (class $KNeighborsClassifier$) |                                                     | The number of neighbors has the greatest influence on the result; the larger the parameter value, the less accuracy. The best result was obtained with the number of neighbors equal to 1. But with $n_{neighbors} = 1$ and a large data set, the prediction time increases |                           |
| Naive Bayes model           | $alpha$ is an additive smoothing parameter ($0$ for no smoothing). The higher the value, the higher the degree of smoothing, which leads to the construction of less complex models. The values used were $alpha= 0, 0.8, 0.9, 1, 2$ | Accuracy is 54% ($alpha=0$, without smoothing). Accuracy decreases slightly with increasing alpha. Highest accuracy is at $alpha = 0$ |                           |
| (class $MultinomialNB$)     |                                                     |                               |                           |
| Naive Bayes model           | $alpha$ – additive smoothing parameter with the same conditions and interpretation as for class $MultinomialNB$ | Accuracy on the built dataset is 71% ($alpha=0$, without smoothing). Highest accuracy is at $alpha = 0$ |                           |
| (class $BernoulliNB$)       |                                                     |                               |                           |
| Naive Bayes model           |                                                      |                               |                           |
| (class $GaussianNB$)        |                                                      |                               |                           |

The results of the experimental study showed that the method of k-nearest neighbors gives the highest accuracy of recognition of anomalous situations. As stated before, the method of k-nearest neighbors to obtain a prediction for a particular situation finds the dataset closest to it in the training sets. The more data will be contained in the training set, the more accurate the predictions will be. However, the disadvantage of this method is that it stores the entire set of training data. Also, provided a large amount of data, the performance is low, as it takes a long time to find the nearest neighbor.

The accuracy of the rest methods compared to the k-nearest neighbor method is lower because they use statistics on features whose values are subject to significant changes and therefore cannot be used to conduct training.
Conclusions
The paper proposes an approach to identify anomalous data from sensors in cyber-physical systems on the example of a water supply system. The data of the normal and anomalous actions of the system obtained within the simulation allowed us to obtain datasets and use it for the generalization by means of machine learning methods. The constructed binary classifiers of normal and anomalous data confirmed the correctness of the proposed approach and its feasibility in practice with an average detection accuracy of about 83-87% under conditions of the k-nearest neighbor method.

In the further work is planned to solve the inverse problem, i.e. to determine the best locations of sensors within the water supply system in order to increase the reliability and security of the infrastructure of this system on the base of the classifiers built.

Acknowledgments
This work was partially supported by grants of RFBR (projects No. 18-07-01488, 18-37-20047, 18-29-22034, 19-07-00953), grant of the President of Russia (project No. MK5848.2018.9) and the budget theme No. 0060-2019-0010.

References
[1] Mueller A C and Guido S 2017 Introduction to Machine Learning with Python: A Guide for Data Scientists (O'Reilly)
[2] Cao Q, Qiao Y and Lyu Z 2017 Machine Learning to Detect Anomalies in Web Log Analysis 3rd IEEE International Conference on Computer and Communications pp 519-523
[3] Che Y, Chen F, Wu T, Hu W and Xu X 2018 A Deep Learning Model for Secure Cyber-Physical Transportation Systems IEEE Conference on Computer Communications Poster and Demo (INFOCOM'18 Poster/Demo)
[4] Yavanoğlu O and Aydos M 2017 A Review on Cyber Security Datasets for Machine Learning Algorithms IEEE International Conference on Big Data pp 2186-2193
[5] Henkel J 2017 Cyber–Physical Systems Security and Privacy p 4
[6] Settanni G, Skopik F, Karaj A, Wurzenberger M, Fiedler R 2018 Protecting Cyber Physical Production Systems using Anomaly Detection to enable Self-adaptation IEEE Industrial Cyber-Physical Systems (ICPS) pp 173-180
[7] Casas P, Soro F, Vanerio J, Settanni G, D’Alconzo A 2017 Network Security and Anomaly Detection with Big-DAMA, a Big Data Analytics Framework IEEE 6th International Conference on Cloud Networking (CloudNet) pp 1-7
[8] Hosic J, Lamps J and Hart D H 2015 Evolving Decision Trees to Detect Anomalies in Recurrent ICS Networks World Congress on Industrial Control Systems Security (WCICSS-2015) pp 50-57
[9] Ferragut E M, Laska J, Olama M M and Ozmen O 2017 Real-Time Cyber-Physical False Data Attack Detection in Smart Grids Using Neural Networks International Conference on Computational Science and Computational Intelligence pp 1-6
[10] Desnitsky V and Kotenko I 2019 Monitoring the State of Materials in Cyberphysical Systems: Water Supply Case Study Materials Today: Proceedings Vol 11 part 1 p 410-416
[11] Desnitsky V and Kotenko I 2017 Modeling and analysis of IoT energy resource exhaustion attacks Studies in Computational Intelligence Springer-Verlag vol 737 pp 263-270
[12] Desnitsky V and Kotenko I 2014 Expert knowledge based design and verification of secure systems with embedded devices Lecture Notes in Computer Science Springer-Verlag vol 8708 pp 194-210