Anomaly Detection in Wireless Sensor Network of the "Smart Home" System

Anton Kanev, Aleksandr Nasteka, Catherine Bessonova, Denis Nevmerzhitsky, Aleksei Silaev, Aleksandr Efremov, Kseniia Nikiforova
ITMO University
Saint Petersburg, Russia
kanev.a.n@mail.ru, nasteka.av@gmail.com, merom812@gmail.com, dennevmer@mail.ru, sila3v@gmail.com, alexandrovefim@mail.ru, nikiforova.k.a@yandex.ru

***Abstract*—Subject. The paper reviews the problem of anomaly detection in home automation systems. Authors define specificities of the existing security networks and accentuate the need of the detection of informational and physical impact on sensors. Characteristics of the transmitted information and physical impacts on automation devices are analysed and used as metrics for the anomalous behavior detection. Various machine learning algorithms for anomaly detection are compared and reviewed. Methods. The paper reviews the anomaly detection method that includes artificial neural networks as a detection tool. In this method characteristics of the security network devices are analysed to detect an anomalous behaviour, and exactly this type of data should be used to train the artificial neural network. This paper describes tools that can be used to implement the offered anomaly detection method. Main results. As an experiment the scenario has been created so that the model of the "Smart home" system produces the data of network information streams and the artificial neural network decides from this data. As a result the training and testing sets has been produced. The configuration of the artificial neural network has been defined as a result of tests. The experiment shows the potential of described method due to the fact that the area under ROC curve is 0.9689, which is better than basic machine learning algorithms performance. Practical importance. The offered method can be used at the development stage while implementation of the information and security systems requiring monitoring of the connected devices. Anomaly detection technology excludes the possibility of the inconspicuous violation of the information's confidentiality and integrity.**

I. Introduction

Nowadays one of the main guarantee of property safety in houses and apartments is set of different alarm systems connected with control rooms of private security companies. Law enforcement agencies use embedded hardware to create security networks consisting of nodes and sensors that are designed for warning and prevent possible illegal actions [1]. These devices successfully indicate physical intrusion or other traceable action (opened doors or windows, movements, etc.) [2].

However, when network nodes or sensors become targets of the harmful impact, system becomes unable to timely react on the external digital or physical attacks which are not included in the standard model of the attacker's behavior.

Due to the home automation ("Smart home") systems development security devices are being included into the existing infrastructure, so they become ones of the most vulnerable network elements and, moreover, a potential object of the standard distributed denial of service (DDoS) attacks [3]. These attacks lead to the system anomaly formation.

At the same time with the "Smart home" introduction and first homemade automated systems creation the first problems and researches of their vulnerabilities appeared [4], [5], [6], [7], [8]. The national recommendation for "Smart home" system construction has no necessary information about security mechanisms [9], [10]. In 2015 B.B. Mario, W. Candid showed in their work vulnerabilities to basic attacks found in the existing serial commercial systems [11]. However, attackers are able to bypass the security mechanisms, or attempt to influence the automated device physically.

Raja Jurdak together with co-authors examined a similar scenario and offered a theory that allowed to detect anomalies in those wireless sensor networks (WSN) that showed signs of the abnormal behavior. One significant drawback of this work was lacking practical application of the research [12].

In reference [13] A.V. Starikovskii shows vulnerabilities of the "Smart home" system and possible ways of its infrastructure intrusion. System elements that should be protected are defined in detail. In addition the functionality of the theoretical anti-malware and self-protection software is described.

Practical results were reached in the School of Information and Communication Technology Gautam Buddha University by the team of Girik Pachauri investigating anomalies in the medical equipment (pressure sensors, oxygen sensors etc.) [14]. These devices are integrated in the WSN, which means that may also break down or be hacked. This case is completely unacceptable as long as these devices are the ones to deal with people's lives and timely response to an anomalous behavior becomes crucial. To track anomalies authors proposed several basic methods of machine learning, the best of which was a Random Forests algorithm with is under receiver operating characteristic (ROC) curve of 0.9654. However, the conditions of the experiment conduction and the input data used by the authors, remain unclear.

Along with the basic machine learning methods an an artificial neural network should be considered. An artificial neural network has much more flexible underlying algorithms and higher resistance to noise in the input data. Its application

in practice with the equal accuracy, allowing much faster to adapt the mechanism for solving new problems (changes in network topology, changing the signs). In this paper the authors propose to consider the use of artificial neural network based on Kohonen network and multilayered perceptron (MLP) to identify anomalies in the WSN with area under ROC curve not less than 0.9654.

II. Anomaly detection using an artificial neural network

A. "Smart home" anomaly detection theory

The general scheme of the possible impact on the "Smart home" system devices consists of two sides: the attacker (subject) and the device under attack (object) (see Fig. 1).

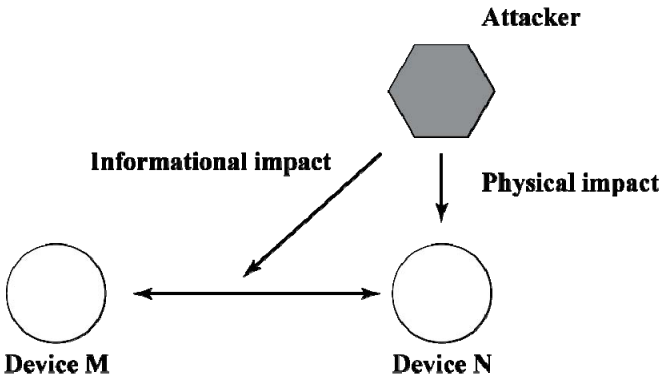

Fig. 1 . Structure of the impact on "Smart Home" devices

The attacker has the ability of the physical intervention in the operation of the device N (turn it on or off). In this case, the device M can detect anomaly in the form of inaccessible device N. The main informational impact here is directed on information flow between the device N and M. If device N controls network activity, in particular the network activity of the device M, the basic type of attack Man-in-the-middle, replay-attack will also lead to the formation of anomalies that an attacker won't be able to hide.

Each of the above described impact influences on the "Smart home" system's network characteristics in one way or another. To solve the tasks set firstly it is necessary to define a set of characteristics which would be analysed by anomaly detection mechanism, in other words - to define metrics.

These metrics might differ depending on the investigated anomaly [15], [2]. In this paper the most common ones were identified:

1) Number of incoming / outgoing packets per unit of time.
2) Packet loss / error per unit time.
3) The power of the outgoing signal.
4) Energy consumption per unit of time.

Due to the requirement of the various types of anomalies identification the metric of neighboring nodes should be taken into account. Also, the values of the metrics are required to be stored for a certain time period in order to monitor their changes. Thus, each "Smart home" system's node can be represented as a set of metrics that are distributed in time.

The above set of metrics is impermanent and individual for the specific implementation of the "Smart home" system. Taking these factors into account, it is proposed to use machine learning as a mechanism for network anomaly detection.

B. Machine learning

Machine learning is an extensive section, which includes a variety of algorithms and methods used for data analysis. To solve the problem of detection of the anomalies in the "Smart home" systems it is necessary to denote each state as abnormal or normal according to its available values.

Taking into account the above described experience of the medical equipment anomaly detection [14], the following popular machine learning methods can be highlighted:

1) k-Nearest Neighbours: One of the simplest and most effective methods used to solve classification problems is the k-Nearest Neighbours. Each element is considered related to a class based on the distance from the other elements of this class.

$$a(u) = \arg\max \sum_{i=1}^{m} [x_{i;u} = y] w(i,u)$$
$$y \in Y$$

where $w(i,u)$ – a function that evaluates weight of neighbor i, for the classification of the object u.

The following pros can be listed: high accuracy; high resistance to errors, no requirements of the training.

The following cons can be listed: large amount of computational power required; a good distance search function.

2) Tree decisions: This is a hierarchical model also known as classification tree or regression tree, consisting of "leaves" and edges. In order to classify each new state, you must go through the decision tree to the final result.

The following pros can be listed: simple interpretation, no processing required.

The following cons can be listed: optimal decisions that are taken at each node may not lead to an optimal global solution.

3) Random forest: Random forest - an algorithm which includes the construction of a plurality of decision trees.

For the task of classification the results of construction will be determined by a majority vote.

So called "Chart of algorithm" that is used consists of the following steps:

- Selecting sub-sample from the training set to construct the tree (unique for each tree).
- A predetermined number of random signs considered for splitting.
- Selecting the best attribute and splitting based on it.

The following pros can be listed: good accuracy with a large number of the input attributes and a small set of the training data.

The following cons can be listed: The final model is large and difficult to understand (may contain thousands of trees).

4) Artificial neural network: Artificial neural network is a mathematical model based on the example of biological neural networks. By connecting the relatively simple algorithms together and constructing an optimal connection between them, the technology can detect the complex relationship between the input parameters, even if they are initially missing in the training set.

This allows the algorithm to be flexible in dealing with different types of tasks.

The following pros can be listed: high accuracy and flexibility of the algorithm.

The following cons can be listed: requires a large amount of computational power.

In our research we use a hybrid neural network which is a combination of the two models of artificial neural networks: self-organizing network with competitive learning (Kohonen layer) and MLP [16]. The structure of hybrid neural network is shown in Fig. 2. There are capabilities to choose another neural network width different structure to search for the best ROC curve, however this is a separate research and is the next task for the authors.

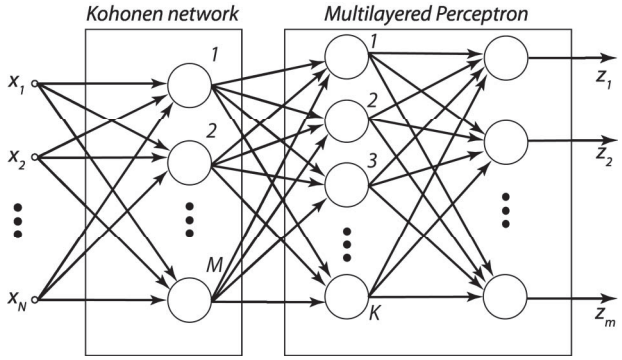

Fig. 2. Structure of Hybrid Neural Network

The main advantage of the Kohonen layer is a high speed of learning in comparison to neural networks with the teacher. The given structure will allow to select the most important input data (the isolation property). Then, the resulting vector is fed to the input of the MLP, the function of which is to determine whether the vector is anomalous or not. In this case, a property of perceptron network's approximation is used.

Training of hybrid network consists of several stages: the first one - the Kohonen layer training, the second one – MLP training, in this case training samples are fed through a the Kohonen layer. The method of the Kohonen layer training is a Winners Take All (WTA) method with conscience mechanism. The method of perceptron training is the method of backpropagation.

The model described above uses the following mathematical apparatus [16].

The value of each neuron in the layer of Kohonen:

$$u_i = \sum_j w_{ji} x_j$$

where u_i - the value of the neuron i, w_{ji} - connection weight of i -th neuron to the j -th input, x_j -j -th input.

In the layer "winner" is selected:

$$u_{\max} = \max\{u_i\}$$

where $u_{\max}$ - the "winner". It uses a conscience mechanism to activate "dead" neurons.

Kohonen layer output:

$$y_i = \exp\left(-\frac{|u_{\max} - u_i|^2}{a^2}\right)$$

where y_i – i -th output, a– picked value .

During the training, "winner" Kohonen layer weights are adjusted :

$$w_{ji} = w_{ji} + a\left(x_j - w_{ji}\right)$$

where α is the speed of training .

The value of neurons in MLP:

$$z_i^{(k)} = \sum_j w_{ji}^{(k)} y_j^{(k-1)}$$

where $z_i^{(k)}$ – the value of i-th neuron in the k-th layer, $w_{ji}^{(k)}$ – weight of connection between i-th neuron of k-th layer and j-th neuron of *(k-1)* -th layer, $y_j^{(k-1)}$ - the value of j-th neuron in the *(k-1)* -th layer, $k = 0$ – input.

C. Practical results

Two separate modules have been developed for practical implementation, both were later united into a single system to identify anomalies in the "Smart home" system.

The first module is the software implementation of the artificial neural network written in C++ language. It repeats the model provided by the artificial neural network. In accordance with the described model a training of the artificial neural network is performed:

1) training of Kohonen layer;
2) training of MLP network.

After training the artificial neural network is done the network itself is capable to make decisions regarding the affiliation of the current state of the node to abnormal on the basis of incoming data.

The second module is the implementation of the system model, “Smart home” system in a special simulation environment (see. Fig. 3). For the development of this module an integrated development environment (IDE) OMNeT++ was used. OMNeT++ is an object-oriented modular framework used for the simulation of the network events. OMNeT++ allows to simulate the work of the wireless and wired networks, the protocols of their work, and also has the ability to embed custom modules in C++ [17]. With its help, a network with sensor devices is simulated. Each node (device sensor) generated a certain traffic flow per unit time.

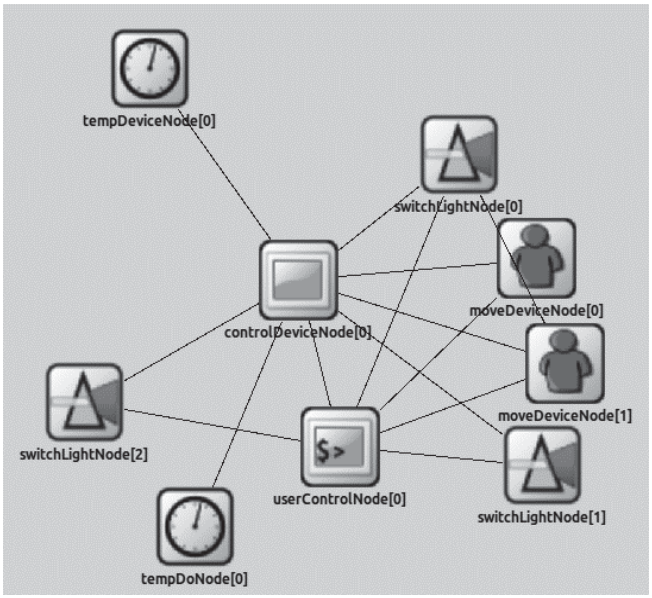

Fig. 3. A simplified model of “Smart home” system in the IDE OMNeT++

The experiment has been conducted to find out the performance and efficiency of the artificial neural network for the purpose of the described problem.

For the experiment the script was created, in which the model “Smart home” system yielded data on information flows in a network, and the artificial neural network makes decision on the basis of the data provided.

The basis of the experimental model is a real-world scenario which includes:

1) living quarters or other automated room with network of general purpose sensors (light sensors, humidity sensors, temperature sensors, etc.) and critical devices (fire and invasion alarm sensors, motion sensors, electronic locks, etc.);
2) an attacker that has sufficient knowledge and tools for an attack;
3) a device with the artificial neural network (“controlDevice[0]”).

At some point the attacker makes a connection to the temperature sensor (“tempDeviceNode[0]”), and his actions create additional informational impact, which should not happen at that moment. Later when the data passes through “controlDevice[0]” device the artificial neural network detects the traffic increase and marks a current state as abnormal (anomaly).

“Smart home” model uses two key devices (according to Fig. 3): “controlDevice[0]” (device with anomaly detector) and “tempDeviceNode[0]” (the source of the abnormal traffic). A situation has been designed where the network traffic includes alternate data streams that are not in normal mode of operation.

For example, if a survey of all devices with their data (send request for package) runs every 5 minutes, “tempDeviceNode[0]” sends messages to recipients every minute at random. The obtained data allows an analysis of incoming packets rate on “controlDevice[0]” using the artificial neural network.

The simulation of simple DDoS attack has been produced in the experiment. So two main metrics has been chosen to detect this attack. Thus input data consists of following metrics:

- the number of incoming packets per unit of time;
- the number of outbound packets per unit of time.

With the information impact, there is a significant excess of traffic compared to the normal state. Number of packets per unit of time that is most revealing metric for detect the main attacks (DDoS, Man-in-the-middle, etc.).

Subsequently, the size of the input vector is 2 neurons.

In the course of the experiment the configuration tests have been conducted to choose the most suitable one. Root mean square error has been used as a criteria.

Fig. 4, Fig. 5 show the result of different neurons number implementation for Kohonen layer and MLP accordingly. Final configuration of the artificial neural network contains 27 neurons at Kohonen layer and 13 hidden neurons at MLP.

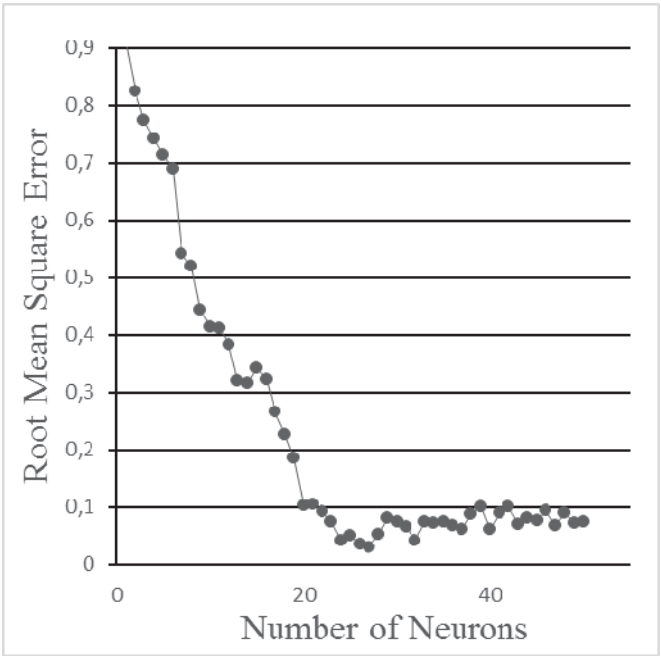

Fig. 4. Performance of the different Kohonen layer neurons number

Fig. 6, Fig. 7 show the result of different learning rate implementation for hybrid neural network. The optimal learning rate has been defined as 0.25 and 0.5 for Kohonen layer and MLP accordingly.

For testing purpose training and test samples have been created, size has been set as 10,000 each. The experiment simulates a situation in which periodically and briefly the data transmission is made from the source of the anomalous traffic "tempDeviceNode[0]". Test sample containes 5000 abnormal conditions.

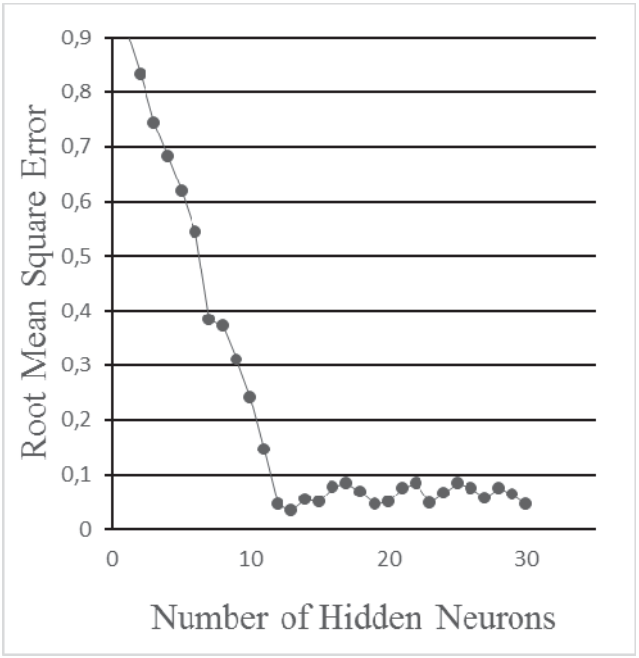

Fig. 5. Performance of the different MLP hidden neurons number

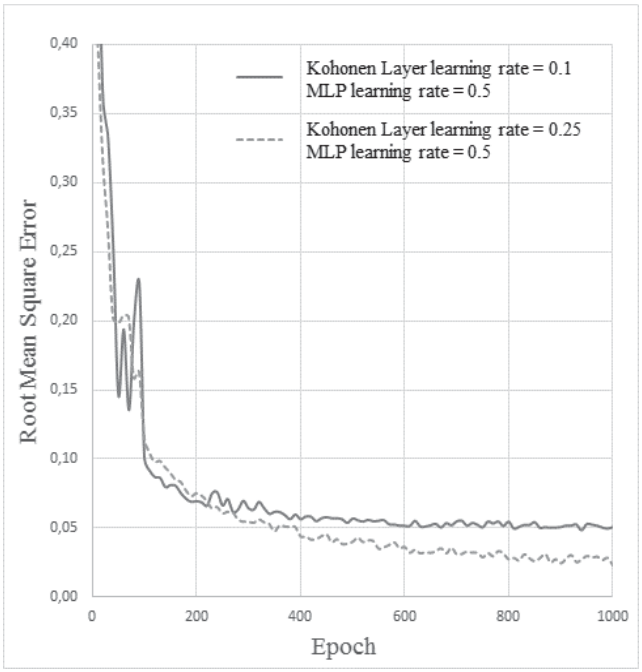

Fig. 6. Performance of the different learning rate implementation

The area under ROC curve measures discrimination, which is the measure of classification performance. The area under the ROC curve represents the probability that a random pair of normal and abnormal images will be correctly ranked as to their state [18], [19].

Fig. 8 shows the ROC curve of the neural network algorithm. The ROC curve illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve plots the true positive rate (also called sensitivity) against the false positive rate (also called specificity) at various thresholds [20]. The ROC curve is thus the sensitivity as a function of specificity. It can be seen that the area under ROC, which shows the overall performance of a classifier, for neural network algorithm is larger than for Random Forests algorithm.

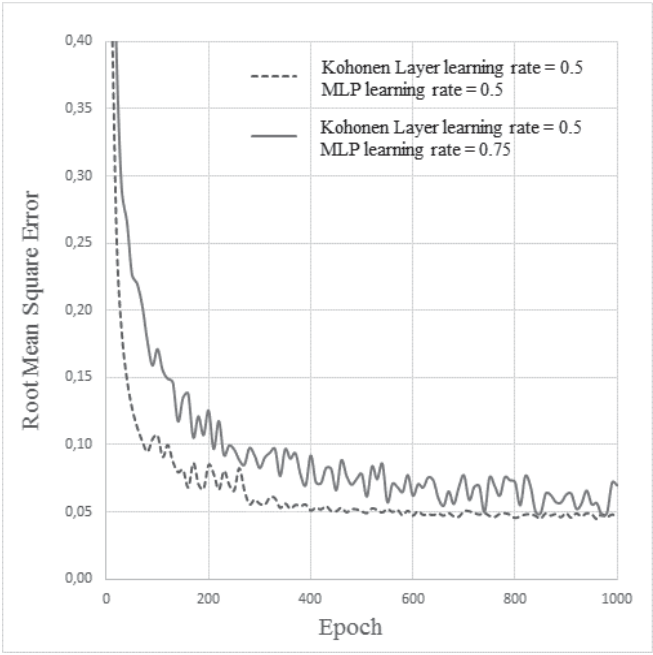

Fig. 7. Performance of the different learning rate implementation

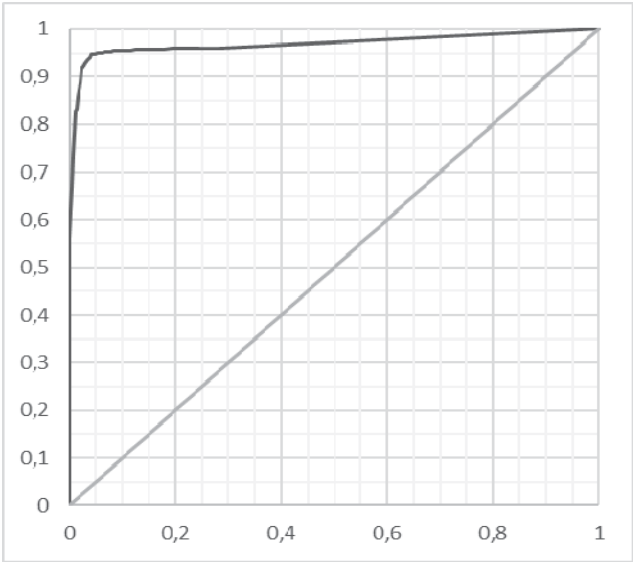

Fig. 8. Artificial neural networks ROC Curve

Fig. 9 shows the mean absolute error (MAE) for two classifier k-Nearest Neighbours and neural network. As another comparison metric for machine learning methods MAE metric has been chosen [21]:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| = \frac{1}{n}\sum_{i=1}^{n}|e_i|$$

This value represents the average of the absolute errors $e_i = |f_i - y_i|$, where f_i is the prediction and y_i the true value. MAE shows how different the predicted value and the actual value are.

We could see that the artificial neural network has good performance along with k-Nearest Neighbours and Random Forests classifiers, and the MAE value counted for the artificial neural network is slightly higher than for the k-Nearest Neighbours classifier. That means that on the same input data the k-Nearest Neighbours algorithm misclassifies much more instances than the artificial neural network classifier does [14].

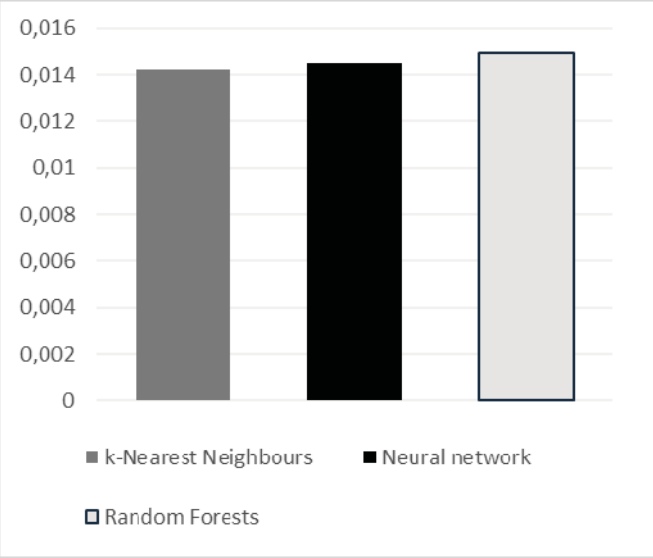

Fig. 9. MAE of Different Classifiers

III. Conclusion

This paper presents an algorithm based on the artificial neural network that can be used to detect anomalies in the network of the "Smart home" system. Scientific novelty lies in the application of the hybrid artificial neural network method, previously unused for the anomaly detection in “Smart home” or building automation systems.

The structure of the artificial neural network is defined as a hybrid network, which includes two layers:

1) Kohonen layer (with 27 neurons);
2) MLP (with 13 hidden neurons).

A different learning rates and different number of neurons for the Kohonen layer and MLP have been chosen experimentally.

Metrics for an abnormal state detection of transmitted information are considered. The experiment has been conducted using two of them:

- the number of the incoming packets per unit of time;
- the number of the outbound packets per unit of time.

With the IDE Omnet++ the model of WSN has been implemented including the following components:

1) general purpose sensors;
2) critical devices;
3) an attacked device;
4) a device with the artificial neural network.

Also with the IDE Omnet++ the training and test sets of data has been created. The testing dataset contained 10000 objects of normal and abnormal state, 5000 objects each.

Results of the experiment show that the artificial neural network has better performance than basic methods of machine learning with area under ROC curve of 0.9689.

The proposed method can be used while development of the information and monitoring systems, which have requirements for monitoring of individually connected devices. Anomaly detection technology allows to eliminate the possibility of undetected confidentiality and integrity violation of of transmitted information.

Authors are going to continue described work. Other metrics will be considered and tested to achieve higher performance at anomaly detection.

Acknowledgment

We thank the Chair of Secure Information Technologies of the ITMO University for supporting and facilitating this research.

References

[1] A.V. Nasteka, C.E. Bessonova, "Authentication of automation devices at smart home system", Vestnik Policii, 2015, vol. 4, is. 2, pp. 68-74.
[2] Rischan Mafrur, Priagung Khusumanegara, Gi Hyun Bang, Do Kyeong Lee, I Gde Dharma Nugraha and Deokjai Choi, "Developing and evaluating mobile sensing for smart home control", International Journal of Smart Home, 2015, vol. 9, No. 3, pp. 215-230.
[3] Bezopasnost' ASUZ. Mozhno li vzlomat' Umnyi dom?, Web: http://www.cnews.ru/reviews/?2011/01/24/424494
[4] A.A. Nasteka, A.A. Efremov, V.V. Ovsyanikova, K.I. Salakhutdinova, A.A. Trofimov, “Protection of control signals in the "smart house" system”, Congress of Young Scientists, 2015.
[5] E.E Bessonova, A.A Efremov, A.V. Nasteka, V.V. Ovsyanikova, K.I. Salakhutdinova, A.A. Trofimov, “Analiz zashchishchennosti sistem "umnyi dom"”, Regional Information, 2014, p. 124.
[6] Yajing Pang, Sujuan Jia, "Wireless smart home system based on zigbee", International Journal of Smart Home, 2016, vol. 10, No. 4, pp. 209-220.
[7] Somia Belaidouni, Moeiz Miraoui, Chakib Tadj, "Towards an efficient smart space architecture", International Journal of Advanced Studies in Computer Science and Engineering, 2016, vol. 5, is. 1.
[8] A.A. Efremov, A.A. Nasteka, V.V. Ovsyanikova, K.I. Salakhutdinova, A.A. Trofimov, “Protecting the system "Smart House" from software failures”, Congress of Young Scientists, 2015.
[9] AVOK, “STO NP "AVOK" 8.1.2-2008 Standart AVOK. Avtomatizirovannye sistemy upravleniya zdaniyami. Chast' 2. Tekhnicheskie sredstva”, 2008.

[10] AVOK, "STO NP "AVOK" 8.1.3-2007 Standart AVOK. Avtomatizirovannye sistemy upravleniya zdaniyami. Chast' 3. Funktsii", 2007.
[11] B.B. Mario, W Candid, "Insecurity in the internet of things", Security response, 2015, pp. 9-14.
[12] Raja Jurdak, X. Rosalind Wang, Oliver Obst, Philip Valencia, "Wireless sensor network anomalies: diagnosis and detection strategies", Intelligence-Based Systems Engineering, 2011, pp. 309-325.
[13] A.V. Starikovskii, "Research vulnerabilities of smart home system", Special equipment and communication, 2012, is. 2, pp. 55-57.
[14] Girik Pachauri, Sandeep Sharma, "Anomaly detection in medical wireless sensor networks using machine learning algorithms", Procedia Computer Science, 2015, pp. 325 – 333.
[15] Sung-Yong Son, "Home electricity consumption monitoring enhancement using smart device status information", International Journal of Smart Home, 2015, vol. 9, No. 10, pp. 189-196.
[16] S. Osovsky, Neural networks for processing information, 2002.
[17] User Manual OMNeT++ version 4.6, Web: https://omnetpp.org/doc/omnetpp/manual/usman.html.
[18] James A. Hanley, Barbara J. McNeil, “The meaning and use of the area under a receiver operating characteristic (ROC) curve”, Radiology, 1982, vol. l43, is. l, pp 29-36.
[19] Thomas G. Tape, “The area under an ROC curve”, University of Nebraska Medical Center, 2005.
[20] K. Manning, P. Raghavan, H. Schutze, "Introduction to Information Retrieval". - Williams, 2011.
[21] 2.5 Evaluating forecast accuracy | OTexts, Web: https://www.otexts.org/fpp/2/5