DOI 10.34185/1562-9945-5-142-2022-03
UDC 004.056.53:[004.7:004.032.26]

V. Pakhomova, Y.Mehelbei

DETECTION OF ATTACKS OF THE U2R CATEGORY BY MEANS OF THE SOM ON DATABASE NSL-KDD

Annotation. Creating an effective system for detecting network attacks requires the use of qualitatively new approaches to information processing, which should be based on adaptive algorithms capable of self-learning. The mathematical apparatus of the Kohonen self-organizing map (SOM) was used as a research method. Python language with a wide range of modern standard tools was used as a software implementation of the Kohonen SOM addition, this section compiles the Python software model «SOM_U2R» using a Kohonen SOM. Created «SOM_U2R» software model on database NSL-KDD an error research was performed for different number of epochs with different map sizes. On the «SOM_U2R» model the research of parameters of quality of detection of attacks is carried out. It is determined that on the «SOM_U2R» created software model the error of the second kind of detection of network classes of attacks Buffer_overflow and Rootkit is 6 %, and for the class Loadmodule reached 16 %. In addition, a survey of the F-measure was conducted for a different number of epochs of learning the Kohonen SOM. It is determined that for all network attack classes (except Buffer_overflow) the F-measure increases, reaching its maximum value at 50 epochs.
Keywords: category; class; NSL-KDD; SOM; Python; error; epoch; F-measure.

Formulation of the problem. Creating an effective system for detecting network attacks requires the use of qualitatively new approaches to information processing, which should be based on adaptive algorithms capable of self-learning. The most promising direction in the creation of such systems for detecting attacks on a computer network is the use of neural network technologies, which confirms the relevance of the topic of this work.

The aim of this article is to identify U2R network classes by means of a Kohonen SOM. The following tasks are set according to the purpose: 1) review neural networks to identify network attacks; 2) create a software model of a Kohonen SOM to identify network classes of U2R category; 3) determine the optimal parameters of the Kohonen SOM; 4) conduct research on network quality detection parameters.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

Analysis of recent research. At the present stage, the most promising direction in the creation of attack detection systems is the use of neural networks: Multi Layer Perceptron, MLP [10-11]; Radial Basis Function Network, RBF [4]; Self Organizing Maps, SOM [2, 5-6, 8]; Adaptive-Network-Based Fuzzy Inference System, ANFIS) [4] and based on a combination of computational intelligence methods [1, 4].

On the one hand, neural networks with different topologies can detect different attacks, but erroneous triggers also do not always occur on the same network packets when analyzed using different types of neural networks. In addition, each type of neural network has its advantages and disadvantages that need to be considered or additional research. For example, the RBF learns faster than the MLP, it is necessary to determine the number of radial elements, their location and deviation values, the RBF model requires slightly more elements, namely it will run slower and requires more memory than the MLP model.

On the other hand, attempts are being made to use neural networks at different levels. For example, in [11] the structure of a hypothetical complex is considered, and consists of five neural networks (NNs) of the multilayer perceptron type. In [9] reviewed existing datasets, the most common of which is the NSL-KDD database, initiated by the US DARPA Agency based on the KDD'99 database [7]. It should be noted that today there are a number of scientific papers by various scientists and scholars on the definition of network attacks in the categories of DoS and Probe, but there is little work on the study of network classes for categories R2L and U2R. According to [6], existing intrusion detection systems based on SOM have difficulties due to the long computation time and low detection rate of U2R and R2L attacks. In [3], a research of two approaches to detecting network attacks using a single neural network and a set of neural networks based on the calculation of quality indicators for detecting attacks, among which errors of the first and second kind are important.

Setting task. The rapid development of computer networks and information technology causes a number of problems related to the security of network resources, which require effective approaches. The use of neural network technology is the most rational, because neural networks have the following advantages: solving problems with unknown patterns; resistance to input data noise; adaptation to changes in the environment; potential ultra-high speed. In this paper, it is necessary to identify network attack classes of the U2R category. U2R network attacks are system attacks in which a hacker starts a system with a normal user account and tries to abuse vulnerabilities in the system to gain superuser privileges. This type of attack is

divided into the following classes: Buffer_overflow, Loadmodule, Perl, Rootkit. The NSL-KDD database [7] presents a sufficient number of parameters for the network classes Buffer_overflow, Rootkit, Loadmodule. Because the database does not have enough parameters for the Perl class, it will not be used.

The structure of the Kohonen SOM. Kohonen neural networks – class of neural networks used to solve classification problems. It is divided into many types according to the methods of adjusting the scales. This work used a Kohonen SOM, structure of the SOM is presented in Fig. 1. The initial parameters are 41 parameters of network traffic. As the resulting data: Y1 – there was an attack of the Rootkit class; Y2 – there was an attack of the Loadmodule class; Y3 – there was an attack of the Buffer overflow class; Y4 (normal) – there was no attack.

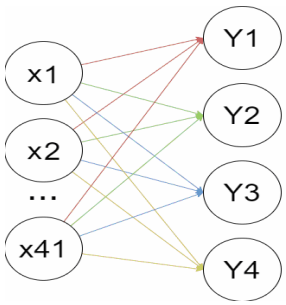

Figure 1 – The structure of the Kohonen SOM

Algorithm of functioning of the Kohonen SOM can be represented:

Step 1. Arrange the weight vectors of the node in random order on the map.

Step 2. Randomly select the input vector.

Step 3. Bypass each node on the map.

Step 4. To find the similarity between the input vector and the weight vector of the map node, you must use the euclidean distance.

Step 5: Remember the node that has the shortest distance as the Best Matching Unit (BMU).

Step 6. Update the weight vectors of the nodes near the BMU inclusive, by approaching the input vector according to the following formula:

$$W_v^{new} = W_v^{old} + \theta(u,v,s) * a(s) * (D(t) - W_v^{old}),$$

where W_v^{new} – new node weight vector; W_v^{old} – previous node weight vector; $\theta(u,v,s)$ – proximity function; $a(s)$ – learning rate; $D(t)$ – vector target input; s – current iteration; u – index of the best matching node on the map; v – node index on the map.

Step 7. Increase s and repeat until $s < \lambda$, where λ – limit of iterations.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

Characteristics of the created «SOM_U2R» software model. To detect U2R attacks, the «SOM_U2R» software model was developed, which is based on the algorithm of the Kohonen SOM. The structure of «SOM_U2R» model is shown Fig. 2. 41 parameters of network traffic were fed to the input «SOM_U2R», the result of execution is a two-dimensional map with distribution on it of network classes of attacks. The software model is waiting for the input clearly 41 parameters, otherwise it will not work properly. Connect the libraries needed to implement the required functions, namely: Numpy, Matplotlib, MiniSom.

Neural network training and testing. A training sample of 55 selections is presented at the input of the neural network. The number of epochs of study was 10; the dimension of the card was 20*20. A sample containing 40 vectors (examples) was used to test the neural network. Neural network testing has been conducted for 10 epochs. The result of the «SOM_U2R» software model is presented in Fig. 3; Rootkit class (red circle); Loadmodule class (green square); Buffer_overflow class (blue cross); Normal (yellow cross).

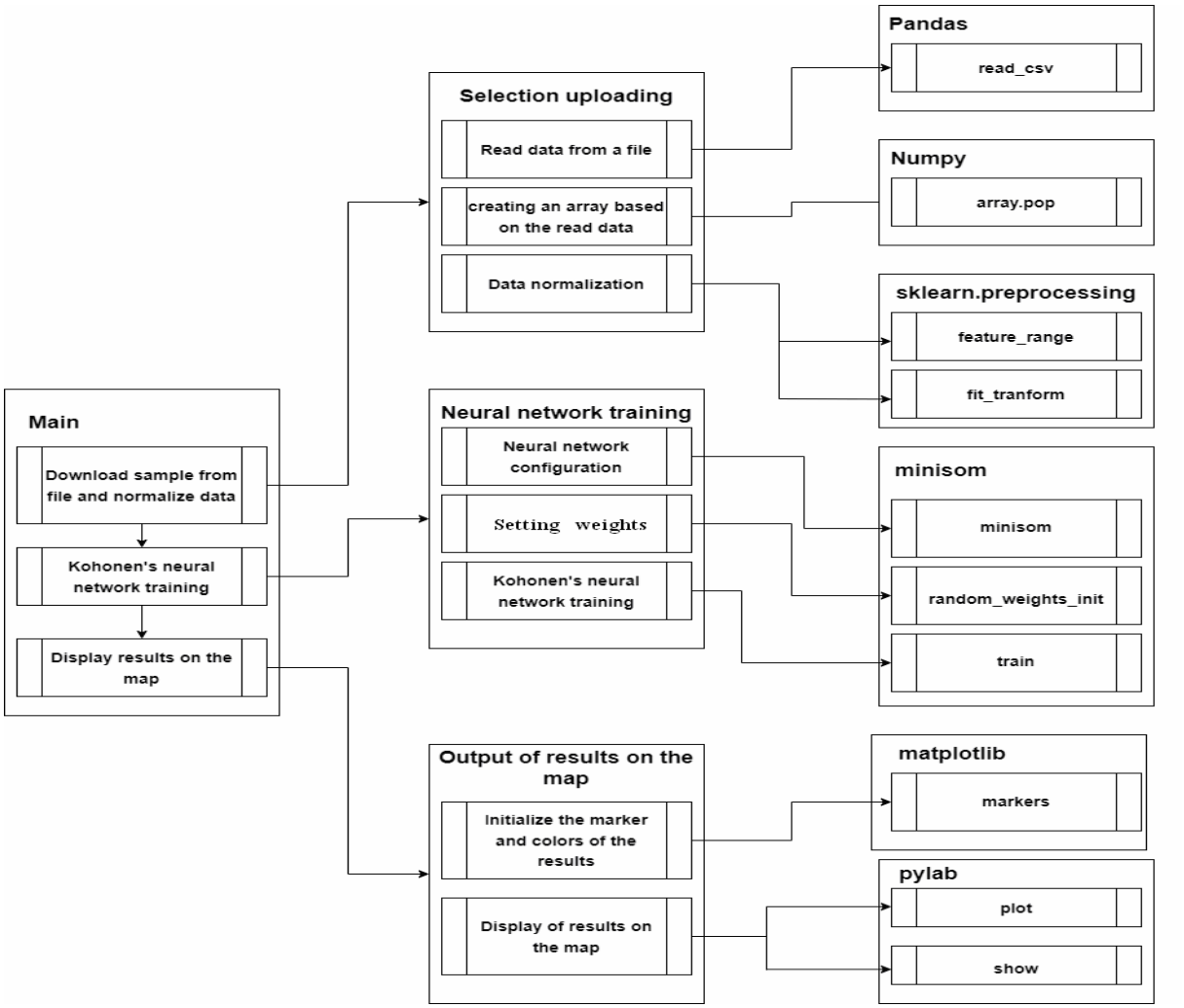

Figure 2 – The structure of the «SOM_U2R» created software model

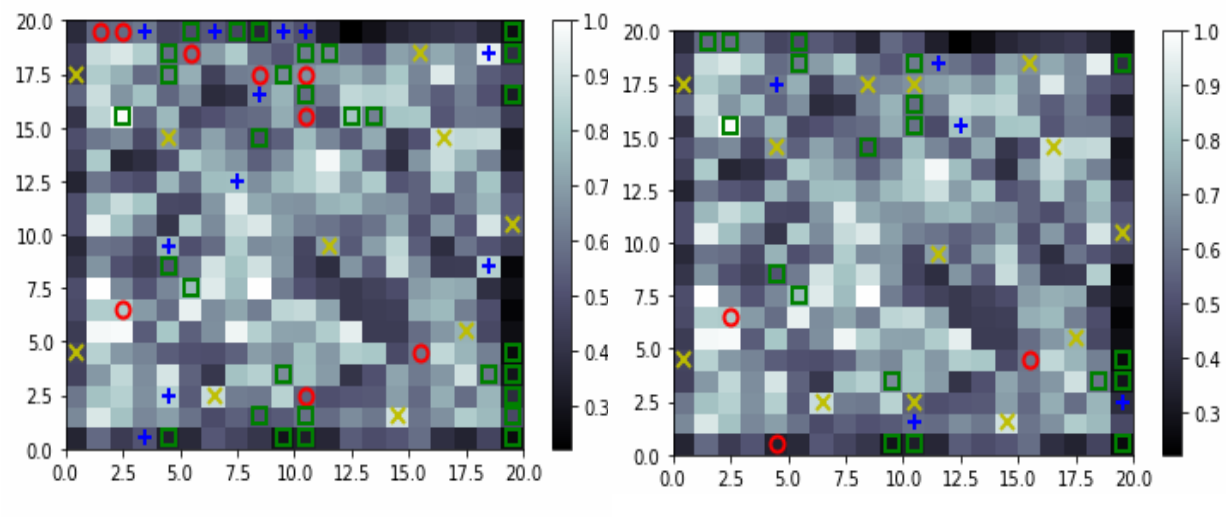

Figure 3– The result of the «SOM_U2R» model for training and testing

Determining the optimal parameters of the Kohonen SOM. Conducted on the created «SOM_U2R» model errors for different number of epochs (10, 20 and 50) for different map sizes: 5*5; 10*10; 20*20. The table shows that the smallest value of error is achieved for Kohonen SOM, the size of which is 20*20, with 10 epochs.

Research of parameters of quality of detection of network attacks. The assessment of the quality of detection of network attacks on the «SOM_U2R» model is performed according to the following parameters: TP (True Positive) – the classifier correctly assigned the object to the class under consideration; TN (True Negative) – the classifier correctly states that the object does not belong to the class under consideration; FP (False Positive) – the classifier incorrectly assigned the object to the class under consideration; FN (False Negative) – the classifier incorrectly states that the object does not belong to the class under consideration. One of the main ones is the second kind of error; the results obtained are summarized in table 1.

Table 1

The results of research of different network classes on «SOM_U2R»

Rootkit				Loadmodule				Buffer_overflow			
TP	FP	TP, %	FP, %	TP	FP	TP, %	FP, %	TP	FP	TP, %	FP, %
6	0	11	0	18	0	33	0	8	0	14	0
FN	TN	FN, %	TN, %	FN	TN	FN, %	TN, %	FN	TN	FN, %	TN, %
2	46	6	83	9	28	16	51	3	44	6	80

The table shows that the largest value of the error of the second kind of 16 % is achieved when detecting the network class Loadmodule. The obtained values of

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

other parameters are summarized in table 2, where TPR (True Positive Rate) – shows the proportion of found objects in the class; FPR (False Positive Rate) – shows the proportion of incorrect classifier triggers to the total number of objects outside the class; accuracy – shows the share of correct classifications; precision – shows the share of class objects among the objects selected by the classifier; recall – shows the proportion of class objects found in the total number of class objects.

Table 2

Parameters for assessing the quality of attack detection on «SOM_U2R»

Indicator	TP	FP	FN	TN	TPR	FPR	Accu-racy	Preci-sion	Re-call
Buffer_overflow	8	0	3	44	0,73	0	0,95	1	0,73
Loadmodule	18	0	9	28	0,67	0	0,84	1	0,67
Rootkit	6	0	3	46	0,67	0	0,95	1	0,67
Normal	8	0	2	45	0,80	0	0,96	1	0,80

The created «SOM_U2R» software model well detects network attacks such as Buffer_overflow and Rootkit with an accuracy of 0,95; but errors may occur when determining the Loadmodule network class (accuracy was 0,84).

Research of F-measures on different number of epochs. A research of the F-measure of detecting network classes of attacks on the «SOM_U2R» model for a different number of learning epochs (Fig. 4).

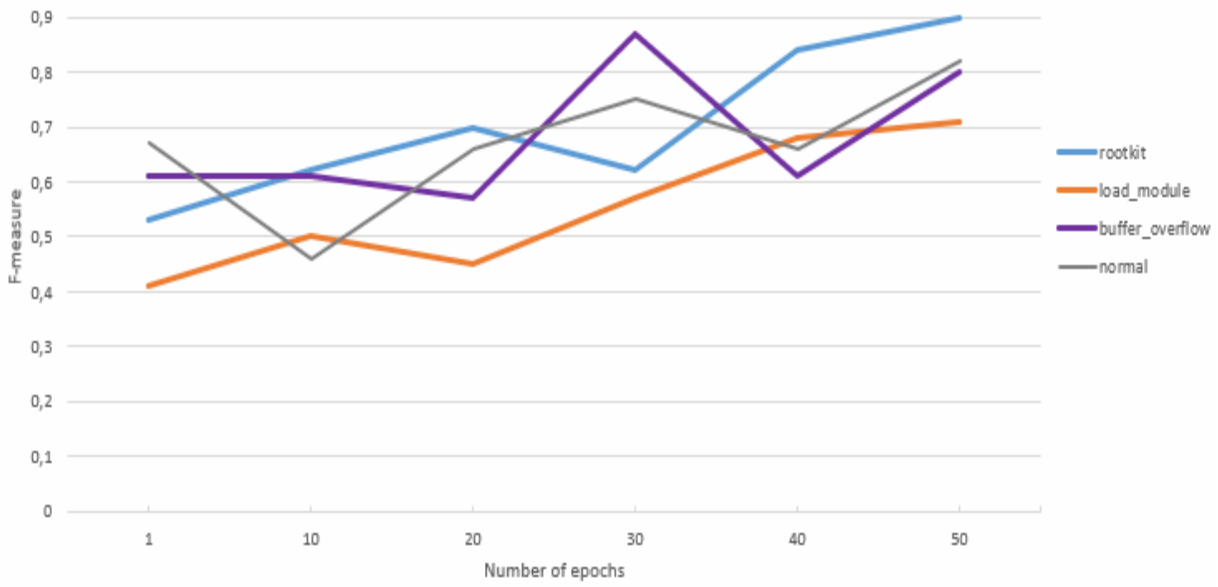

Figure 4 – The value of the F-measure by the number of epochs of learning

The F-measure is the average harmonic value between accuracy and completeness. The figure shows that for all types of attacks, except Buffer_overflow, the F-measure increases, reaching its maximum value

at around 50 epochs. For Buffer_overflow, the maximum value of the F-measure is observed in 30 epochs, but to ensure maximum efficiency in detecting network attacks, the neural network must be stopped at 50 epochs. As the Rootkit and Loadmodule network attack classes show the best results at the 50 epoch mark, thus sacrificing the growth of the Buffer_overflow network class.

Conclusions

- Based on the results of the survey, the following neural networks may be used to determine attacks: MLP; RBF; SOM; fuzzy network. To detect U2R network classes: Buffer_overflow; Load modules; Rootkit using NSL-KDD open database and further research selected Kohonen SOM.
- To identify U2R network classes, the «SOM_U2R» software model was created on Python using a Kohonen SOM, the input of which was supplied with 41 network traffic parameters. On the «SOM_U2R» software model error studies on the number of epochs (10, 20 and 50) with different map sizes: 5*5; 10*10; 20*20. It is determined that the smallest value of error is achieved on the map 20*20 at 10 epochs.
- On the «SOM_U2R» model researches of parameters of quality of detection of network classes of the U2R category are carried out. It is determined that on the «SOM_U2R» model the error of the second kind of detection of network classes of attacks Buffer_overflow and Rootkit made 6 %, and for the Loadmodule class reached 16 %.
- On the «SOM_U2R» created software model the research of F-measure on various quantity of epochs of training of the Kohonen SOM is carried out. It is determined that using the «SOM_U2R» model for all classes of attacks of the U2R category (except Buffer_overflow) the F-measure increases, reaching its maximum value at the level of 50 epochs.

ЛІТЕРАТУРА

1. Браницкий А.А. Обнаружение аномальных сетевых соединений на основе гибридизации методов вычислительного интеллекта : автореф. дис. ... канд. техн. наук. Санкт-Петербург, 2018. 18 с.

2. Пахомова В.М., Павленко І. І. Дослідження параметрів якості визначення мережевих атак категорії PROBE з використанням самоорганізуючої карти. SworldJournal. 2022. Issue 11. Part 1. pp.100-104. DOI: 10.30888/2663-5712.2022-11-01-022

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

3. Пахомова В.М., Коннов М.С. Дослідження двох підходів до виявлення мережних атак із використанням нейромережної технології. Наука та прогрес транспорту. 2020. № 3 (87). pp. 81-93. DOI: https://doi.org/10.15802/stp2020/208233

4. Amini M., Rezaeenour J., Hadavandi E. A Neural Network Ensemble Classifier for Effective Intrusion Detection using Fuzzy Clustering and Radial Basis Function Networks. International Journal on Artificial Intelligence Tools. 2016. Vol. 25. Iss. 02. pp. 1-32. DOI: https://doi.org/10.1142/s0218213015500335

5. Gunes K., Zincir-Heywood A., Malcolm I. H. A hierarchical SOM-based intrusion detection system. Engineering Applications of Artificial Intelligence. 2007. pp. 439-451.

6. Kruti C., Bhavin S., Ompriya K. Improving user-to-root and remote-to-local attacks usinggrowing hierarchical self organizing map. International journal of engineering sciences and research technology. 2015. том 4. № 6. URL: http://paper.researchbib.com/view/paper/45808

7. NSL-KDD dataset. URL: https://www.unb.ca/cic/datasets/nsl.html

8. Ortiz A. Improving Network Intrusion Detection with Growing Hierarchical Self-Organizing Maps. University of De La Plata Argentin. 2011. URL: https://www.semanticscholar.org/paper/Improving-Network-Intrusion-Detection-with-Growing-Ortiz-Ortega/f3fbcf7dfd84d9f2f2ace73580c32eb7c469b6e7

9. Ring M., Wunderlich S., Scheuring D., Landes D., Hotho A. A Survey of Network-based Intrusion Detection Data Sets. Компьютер и безопасность. 2019. DOI: 10.1016 / j.cose.2019.06.005

10. Zhukovyts'kyy I. V., Pakhomova V. M. Identifying threats in computer network based on multilayer neural network. Science and Transport Progress. 2018. № 2 (74). pp. 114-123. DOI: https://doi.org/10.15802/stp2018/130797

11. Zhukovyts'kyy I. V., Pakhomova V. M., Ostapets D. O., Tsyhanok O. I. Detection of attacks on a computer network based on the use of neural network complex. Science and Transport Progress. 2020. № 5(89). pp. 68-79. URL: https://doi.org/10.15802/stp2020/218318

REFERENCES

1. Branitskiy, A.A. (2018). Obnaruzhenie anomalnykh setevykh soedineniy na osnove gibridizatsii metodov vychislitelnogo intellekta (Extended abstract of PhD dissertation). St. Petersburg, Russia. (in Russian)

2. Pakhomova, V.M., & Pavlenko, I.I. (2022). Research of parameters of quality of definition of network attacks of the PROBE category with use of the self organizing map. SworldJournal, 11-1, 100-104. DOI: 10.30888/2663-5712.2022-11-01-022 (in Ukrainian)

3. Pakhomova, V. M., & Konnov, M. S. (2020). Research of two approaches to detect network attacks using neural network technologies. Science and Transport Progress, 3(87), 81-93. DOI: https://doi.org/10.15802/stp2020/208233 (in Ukrainian)
4. Amini, M., Rezaeenour, J., & Hadavandi, E. (2016). A Neural Network Ensemble Classifier for Effective Intrusion Detection Using Fuzzy Clustering and Radial Basis Function Networks. International Journal on Artificial Intelligence Tools, 25(02), 1-32. DOI: https://doi.org/10.1142/s0218213015500335 (in English)
5. Gunes, K., Zincir-Heywood, A., & Malcolm, I. H. (2007). A hierarchical SOM-based intrusion detection system. Engineering Applications of Artificial Intelligence, 439-451.
6. Kruti, C., Bhavin, S., & Ompriya, K. (2015). Improving user-to-root and remote-to-local attacks usinggrowing hierarchical self organizing map. International journal of engineering sciences and research technology, 4(6),
URL: http://paper.researchbib.com/view/paper/45808 (in English)
7. NSL-KDD dataset. URL: https://www.unb.ca/cic/datasets/nsl.html (in English)
8. Ortiz, A. (2011). Improving Network Intrusion Detection with Growing Hierarchical Self-Organizing Maps. University of De La Plata Argentin. URL: https://www.semanticscholar.org/paper/Improving-Network-Intrusion-Detection-with-Growing-Ortiz-Ortega/f3fbcf7dfd84d9f2f2ace73580c32eb7c469b6e7 (in English)
9. Ring, M., Wunderlich, S., Scheuring, D., Landes, D., & Hotho, A. (2019). A Survey of Network-based Intrusion Detection Data Sets. Komp'yuter and bezopasnost. DOI: 10.1016 / j.cose.2019.06.005 (in English)
10. Zhukovyts'kyy, I. V., & Pakhomova, V. M. (2018). Identifying threats in computer network based on multilayer neural network. Science and Transport Progress, 2(74), 114-123. DOI: https://doi.org/10.15802/stp2018/130797 (in English)
11. Zhukovyts'kyy, I.V., Pakhomova, V.M., Ostapets, D.O., & Tsyhanok, O. I. (2020). Detection of attacks on a computer network based on the use of neural network complex. Science and Transport Progress, 5(89), 68-79. URL:
https://doi.org/10.15802/stp2020/218318 (in English)

Received 13.06.2022.
Accepted 17.06.2022.

Визначення атак категорії U2R засобами SOM на основі бази даних NSL-KDD

Створення ефективної системи виявлення мережевих атак вимагає застосування якісно нових підходів до обробки інформації, які повинні ґрунтуватися на адаптивних алгоритмах здатних до самонавчання. Найбільш

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

*перспективним напрямком у створенні подібних систем виявлення атак на комп'ютерну мережу є застосування нейромережних технологій, що підтверджує актуальність теми даної роботи. У якості методу дослідження використаний математичний апарат самоорганізуючої карти Кохонена 41-2-4, де 41 - кількість вхідних нейронів (параметри мережевого трафіку); 2 - кількість шарів; 4 - кількість результуючих нейронів (Rootkit, Loadmodule, Buffer_overflow та відсутність атаки). У якості програмної реалізації самоорганізуючої карти Кохонена використана мова Python з широким спектром сучасних стандартних засобів. На створеній програмній моделі «SOM_U2R» з використанням відкритої бази даних NSL-KDD проведено дослідження помилки за різною кількістю епох при різних розмірах карти: 5*5; 10*10; 20*20. Визначено, що найменше значення помилки досягається на карті 20*20. На створеній програмній моделі «SOM_U2R» проведено дослідження параметрів якості виявлення атак: True Positive; True Negative; False Positive; False Negative та інші. Визначено, що на програмній моделі «SOM_U2R» помилка другого роду склала 6 % для Buffer_overflow і Rootkit, 16 % для класу Loadmodule. Крім того, проведено дослідження F-мірки (середнєгармонічного значення між точністю та повнотою) за різною кількістю епох навчання самоорганізуючої карти Кохонена. Визначено, що для всіх атак (крім Buffer_overflow) F-мірка зростає, досягаючи свого максимального значення (50 епох).*

Пахомова Вікторія Миколаївна – к.т.н., доц. кафедри електронних обчислювальних машин Українського державного університету науки та технологій (Дніпро); ORCID 0000-0002-0022-099X

Pakhomova Victoria – PhD, Assoc. Department of Electronic Computers of the Ukrainian State University of Science and Technology (Dnipro); ORCID 0000-0002-0022-099X

Мегельбей Егор Олександрович – бакалавр спеціальності «Кібербезпека» кафедри електронних обчислювальних машин Українського державного університету науки та технологій (Дніпро)

Mehelbei Yehor – Bachelor of Cybersecurity in Department of Electronic Computers of the Ukrainian State University of Science and Technology (Dnipro)