УДК 004.056.53:[004.7:004.032.26]
DOI https://doi.org/10.32782/2663-5941/2022.6/20

Пахомова В.М.
Український державний університет науки і технологій

Мотиленко В.А.
Український державний університет науки і технологій

ДОСЛІДЖЕННЯ МОЖЛИВОСТІ ВИКОРИСТАННЯ RBF ДЛЯ ВИЗНАЧЕННЯ SMURF АТАК НА ОСНОВІ БАЗИ ДАНИХ KDDCUP

Для виявлення мережевих атак у режимі реального часу використовуються системи виявлення вторгнень. Одним із найбільш ефективних підходів у класифікуванні великого обсягу даних є застосування нейромережевої технології, що дозволяє виявляти не тільки відомі мережеві атаки, але й виявляти нові. На сьогодні відомо, що помилкові спрацьовування *відбуваються не завжди на одних і тих самих мережевих пакетах при аналізі за допомогою різних типів нейронних мереж: багатошарового персептрону; мережі Кохонена або самоорганізуючої карти; радіально-базисної мережі; нейронечіткої мережі, а також їх комбінацій. Для визначення мережевих атак категорії DoS з використанням бази даних KDDCup створено мовою Rust програму «RBF_DoS», в основу якої покладена мережа RBF конфігурації N-M-K, де N – кількість вхідних нейронів (параметри мережевого трафіку); M – кількість нейронів прихованого шару (кількість базисних функцій); K – кількість результуючих нейронів (мережеві класи атак) за методом стохастичного градієнтного спуску, у якості функції належності прихованих нейронів взято Гаусовську функцію. Проведено апробацію програми «RBF_DoS» на основі RBF конфігурації 29-50-3 для визначення кластерів (Normal; Smurf; Another_attack) з використанням наступних вибірок: навчальної, що складалася із 200000 прикладів на кожний кластер; тестової, що складалася із 10000 прикладів на кожний кластер; контрольної, що складалася із 100 прикладів на кожний кластер. На створеній програмі «RBF_DoS» проведено дослідження точності та середньоквадратичної похибки RBF за наступними параметрами: епохами навчання; довжиною навчальної вибірки; кількістю прихованих нейронів. Визначено, що при виявленні атак мережевого класу Smurf найменше значення середньоквадратичної похибки RBF досягалося за 10 епох навчання з використанням 101 прихованих нейронів, при цьому достатньо мати навчальну вибірку із 2408 прикладів; точність визначення атаки Smurf склало 0,99.*

Ключові слова: *атака, Smurf, RBF, конфігурація, Гаусовська функція, точність, похибка, вибірка, епоха.*

Постановка проблеми. Для виявлення мережевих атак у режимі реального часу використовуються системи виявлення вторгнень. Одним із найбільш ефективних підходів у класифікуванні великого обсягу даних є застосування нейромережевої технології, що дозволяє виявляти не тільки відомі мережеві атаки, але й виявляти нові.

Аналіз останніх досліджень і публікацій. На сучасному етапі відомо про існування наступних категорій атак: DoS; R2L; U2R; Probe, зокрема значна кількість джерел присвячена дослідженню DoS-атак [6, 8, 13, 15, 18]. У загалі для виявлення атак на комп'ютерну мережу можливе використання наступних нейронних мереж (НМ): багатошарового персептрону (Multi Layer Perceptron, MLP) [3-4, 7, 18]; мережі Кохонена або самоорганізуючої карти (Self Organizing Map, SOM) [7, 13]; радіально-базисної мережі (Radial Basis Function Network, RBF) [7, 9] та використання нейронечіткої технології (Adaptive-Network-Based Fuzzy Inference System, ANFIS) [5-6, 9], проте значна кількість джерел присвячена дослідженню атак на основі використання MLP. На сьогодні відомо, що помилкові спрацьовування також відбуваються не завжди на одних і тих самих мережевих пакетах при аналізі за допомогою різних типів НМ. Так, зокрема, до основних переваг використання RBF мережі можна віднести [1]: спрощену структуру мережі (наявність лише одного прихованого шару); високу швидкість навчання; здатність навчатися на неоднорідній вибірці даних; здатність моделювати випадкові процеси. Але проблемою RBF-мережі є вибір кількості радіально-базисних функцій [10-11, 16-17]. У [10] зазначено, що число необхідних радіально-базисних функцій росте експоненціально із зростанням числа вхідних змінних. Тому для визначення DoS атак доцільно

провести дослідження можливості використання RBF мережі.

Формулювання цілей статті. Проведені дослідження ставили за мету розвиток методики визначення мережевих атак категорії DoS. Для досягнення поставленої мети вирішувалися наступні задачі: розробити методику виявлення мережевих атак засобами мережі RBF; при виконанні машинного навчанні виявити оптимальні параметри НМ, що забезпечить високий рівень достовірності виявлення вторгнень в комп'ютерну мережу.

Виклад основного матеріалу дослідження. Категорія DoS характеризується генерацією великого обсягу трафіку, що призводить до перевантаження та блокування сервера. Відомі наступні класи мережевих атак відповідно до категорії DoS: Back, Land, Neptune, Pod, Smurf, Teardrop. У якості початкових даних використана відкрита база даних KDDCup [12].

Проведений аналіз бази даних KDDCup показав, що до її складу надходить: 2203 записів для мережевого класу Back; 21 – для Land; 1072017 – для Neptune; 264 – для Pod; 2807886 – для Smurf; 979 – для Teardrop; 972781 – для Normal (не має атаки), тому для подальшого дослідження сфокусовано увагу на численніших кластерах (Smurf і Normal). Smurf – це розподілена атака відмови в обслуговуванні, в якій велика кількість пакетів Internet Control Message Protocol (ICMP) з підробленою вихідною IP-адресою передбачуваної жертви транслюється в комп'ютерну мережу з використанням широкомовної IP-адреси. Більшість пристроїв у мережі за про мовчанням дадуть відповідь на вихідну IP-адресу. Якщо кількість машин у мережі, які отримують та відповідають на ці пакети, дуже велика, комп'ютер жертви буде переповнений трафіком. Це може сповільнити роботу комп'ютера жертви настільки, що з ним буде неможливо працювати.

Математичний апарат. У якості математичного апарату використана RBF мережа, структура якої подана на рис. 1.

RBF-мережа по своїй структурі відноситься до двошарової мережі, в який використовується єдиний прихований шар (радіально-базисний) з фіксованим нелінійним перетворенням вектора входу з постійними ваговими коефіцієнтами. Нейрони прихованого шару діють за принципом центрування на елементах навчальної вибірки. Навколо кожного центру існую область, яка зветься радіусом. Специфіка RBF-мережі полягає у тому, що в них налагоджуються тільки вагові коефіцієнти лінійного вихідного шару, що в свою чергу, сприяє швидкому процесу навчання мережі.

У якості вхідних змінних НМ використані із бази даних KDDCup [12] параметри мережевого трафіку $x_1 \ldots x_n$ (n=29 [2]; табл. 1).

У якості результуючих нейронів $y_1 \ldots y_K$ (k=3), де y_1 відповідає Normal (атаки не було); y_2 – Smurf; y_3 – Another_attack. Вихід НМ є лінійною комбінацією набору базових функцій:

$$y_k(x) = \sum_{j=1}^{M} w_{jk} \cdot \Phi_j(x),$$

де w_{jk} – вагові коефіцієнти; $\Phi_j(x)$ – базисні функції, що визначаються як:

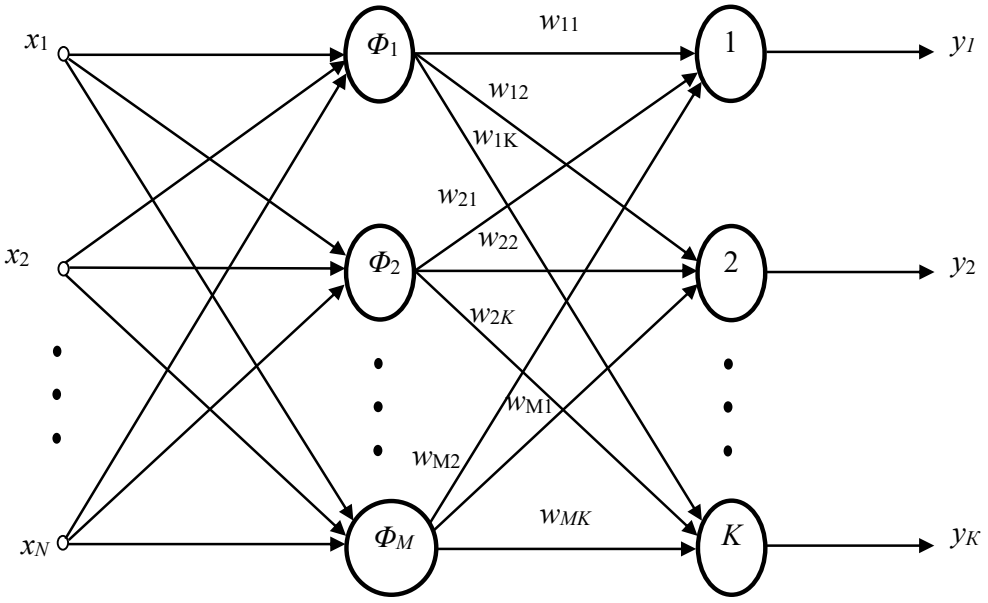

Рис. 1. Структура мережі RBF

Таблиця 1

Співвідношення нейронів і параметрів мережевого трафіку

Нейрон	Параметр	Нейрон	Параметр
x1	duraction	x16	srv_rerror_rate
x2	src_bytes	x17	same_srv_rate
x3	dst_bytes	x18	diff_srv_rate
x4	urgent	x19	srv_diff_host_rate
x5	hot	x20	dst_host_count
x6	su_attempted	x21	dst_host_srv_count
x7	root_shell	x22	dst_host_same_srv_rate
x8	num_failed_logins	x23	dst_host_diff_srv_rate
x9	is_host_login	x24	dst_host_same_src_port_rate
x10	is_guest_login	x25	dst_host_srv_diff_host_rate
x11	count	x26	dst_host_serror_rate
x12	srv_count	x27	dst_host_srv_serror_rate
x13	serror_rate	x28	dst_host_rerror_rate
x14	srv_serror_rate	x29	dst_host_srv_rerror_rate
x15	rerror_rate	y	label

$$\Phi_j(x) = e^{-\frac{|x-\mu_j|^2}{\sigma_j^2}},$$

де μ_j – координата центру j-ой функції RBF; σ_j – радіус *j*-ой функції RBF.

Формування вибірок. Згенеровано наступні вибірки: 1) навчальну, що складалася із 200000 прикладів на кожний кластер; 2) тестову, що складалася із 10000 прикладів на кожний кластер; 3) контрольну, що складалася із 100 прикладів на кожний кластер. У якості прикладу наведений фрагмент контрольної вибірки:

0,1032,0,0,0,0,0,0,0.0,0.0,511,511,0.0,0.0,0.0,0.0,
1.0,0.0,0.0,255.0,255.0,1.0,0.0,1.0,0.0,0.0,0.0,0.0,
0.0,0.0,smurf
0,0,0,0,0,0,0,0,0.0,0.0,1,1,0.0,0.0,1.0,1.0,1.0,0.0,
0.0,12.0,239.0,1.0,0.0,0.08,0.12,0.0,0.0,1.0,1.0,
normal
0,0,0,0,0,0,0,0,0.0,0.0,241,10,1.0,1.0,0.0,0.0,0.04,
0.06,0.0,255.0,9.0,0.04,0.05,0.0,0.0,1.0,1.0,0.0,
0.0,neptune

Алгоритм навчання НМ: 1) зчитати вхідні параметри; 2) завантажити масив даних з файлу для тренування НМ; 3) створити об'єкт НМ, вибравши необхідну кількість випадкових записів – це центри базисних функцій (кожен центр відповідає одному прихованому нейрону); 4) передати масив даних для навчання НМ з використанням методу стохастичного градієнтного спуску; 5) перемішати масив даних; 6) повторити п.п. 4–5 для заданої кількості епох; 7) зберегти НМ з найменшим значенням помилки; 8) повторити п.п. 3–7 для заданої кількості ітерацій. Алгоритм тестування НМ: 1) зчитати вхідні параметри; 2) зчитати коефіцієнти для налаштування НМ з файлу моделі; 3) завантажити коефіцієнти в об'єкт НМ; 4) завантажити масив даних з файлу; 5) передати елемент у функцію підрахунку результату в НМ; 6) підрахувати середньоквадратичне відхилення та точність НМ.

Створення програми. У якості мови програмування обрано Rust [14], яка орієнтована на безпеку та забезпечує високий паралелізм виконання завдань. Однією із особливостей Rust є використання trait, що підтримують об'єктно-орієнтоване програмування, і позначають спільну поведінку різних типів, що дозволяє уникати помилок на етапі компіляції. Загальна структура складеної програми «RBF_DoS» показана на рис. 2.

Дослідження параметрів НМ. На програмі «RBF_DoS» проведено ряд дослідів: дослідження точності та середньоквадратичної похибки за різною кількістю епох навчання НМ (рис. 3), за різною довжиною навчальної вибірки (рис. 4) та за різною кількістю прихованих нейронів (рис. 3). Із рис. 3 видно, що найменше значення похибки НМ досягнуто протягом 10 епох навчання; при цьому точність визначення атаки Smurf склало 0,99. Із рис. 4 видно, що найменше значення середньоквадратичної похибки НМ досягалося на навчальних вибірках, довжина яких 2408 і 2806 прикладів. Із рис. 5 видно, що найменше значення середньоквадратичної похибки навчання досягалося на НМ з використанням 101 прихованих нейронів (базисних функцій).

Висновки. Для визначення мережевих атак категорії DoS на основі використання бази даних KDDCup99 створена мовою Rust програма «RBF_DoS» на основі реалізації мережі RBF

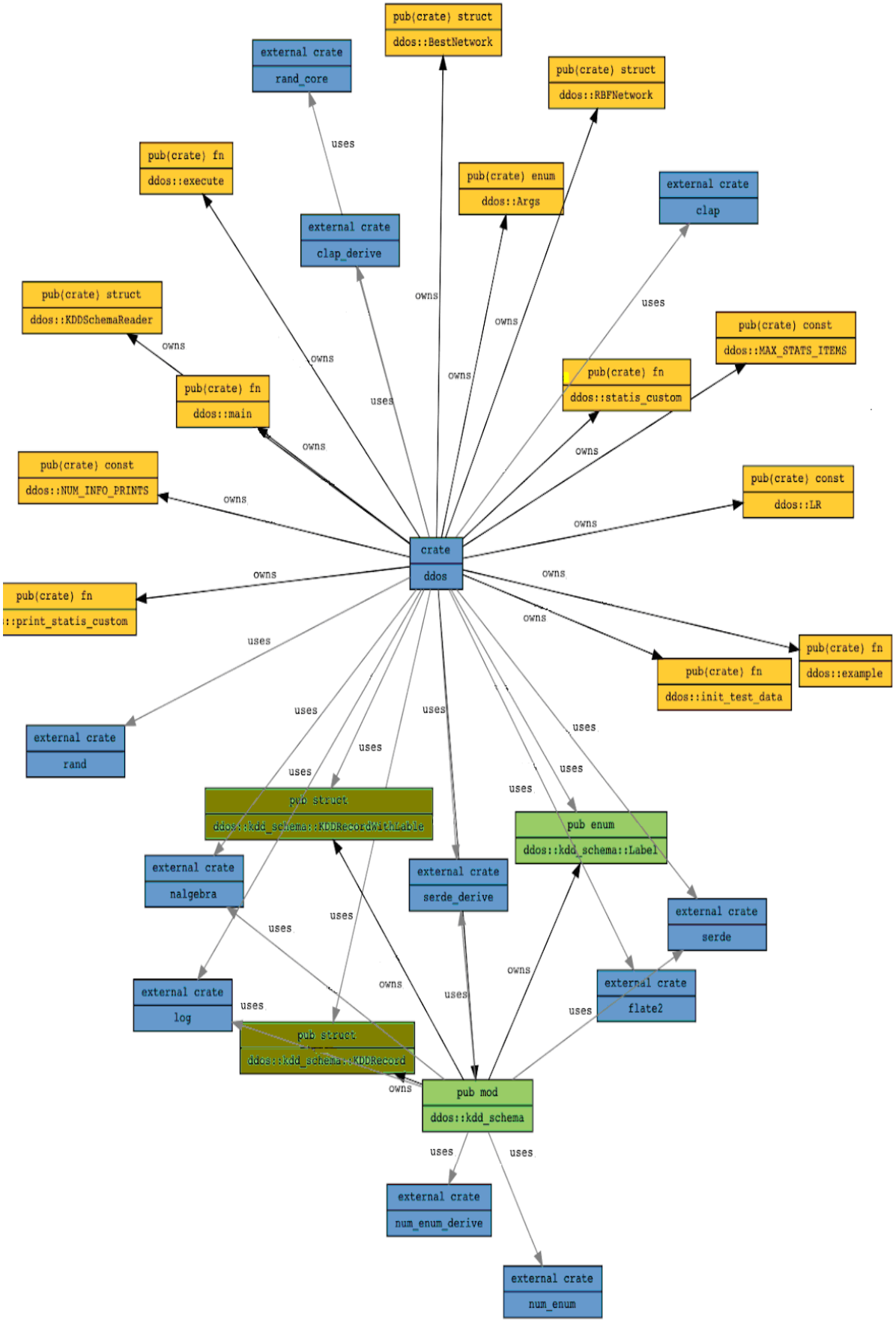

Рис. 2. Загальна структура складеної програми “RBF_DoS”

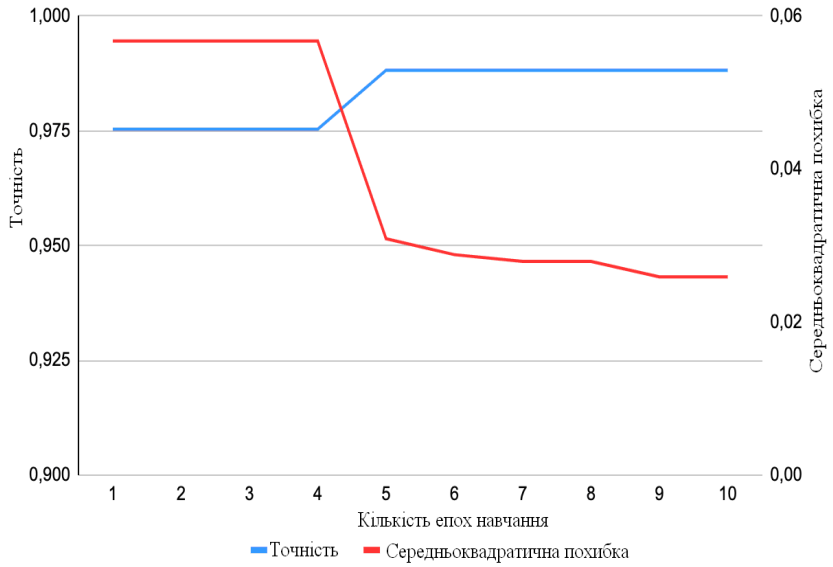

Рис. 3. Точність і похибка НМ за кількістю епох навчання

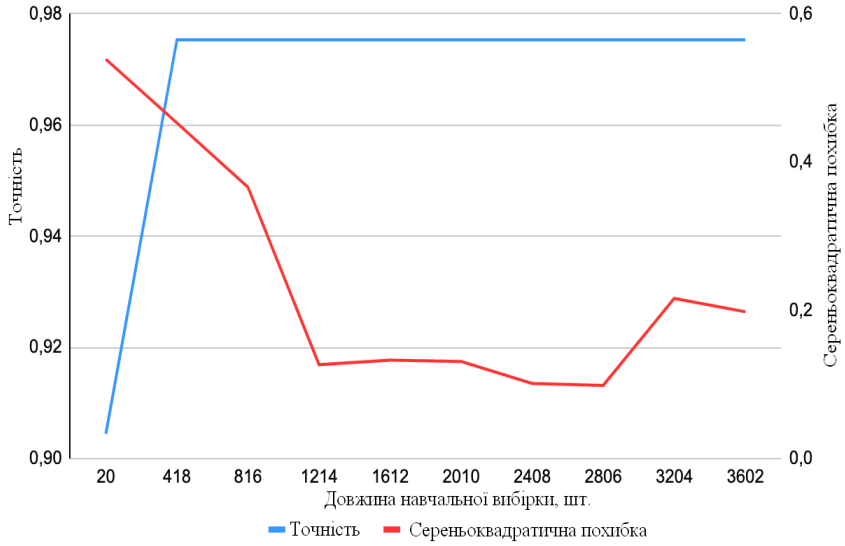

Рис. 4. Точність і похибка НМ за довжиною навчальної вибірки

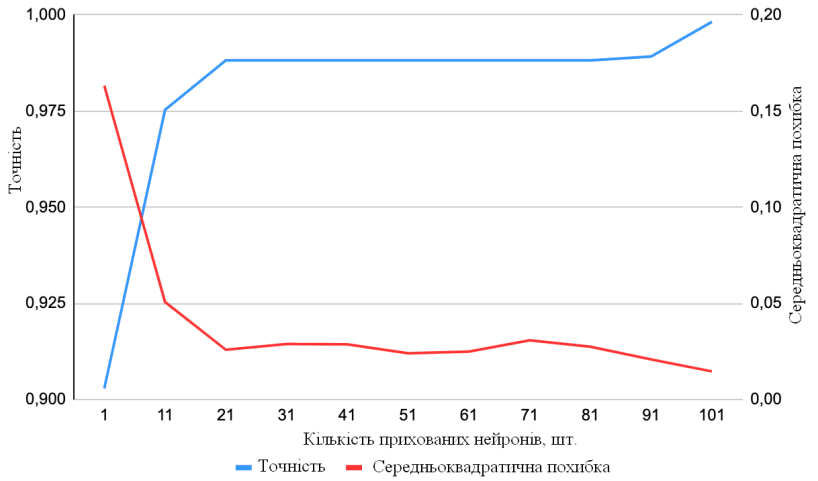

Рис. 5. Точність і похибка НМ за кількістю прихованих нейронів

конфігурації N-M-K, де N – кількість вхідних нейронів (параметри мережевого трафіку); M – кількість нейронів прихованого шару (кількість базисних функцій); K – кількість результуючих нейронів, у якості функції належності прихованих нейронів взято Гаусовську функцію (ширина розподілу 0,1). Проведена апробація програми «RBF_DoS», основу якої складала мережа RBF конфігурації 29-50-3 для визначення наступних кластерів: Normal; Smurf; Another_attack, робота якої залежить від кількості та положення радіально-базисних функцій.

На створеній програмі «RBF_DoS» проведено дослідження точності та середньоквадратичної похибки RBF. Визначено, що найменше значення середньоквадратичної похибки RBF досягалося за 10 епох навчання при використанні 101 прихованих нейронів (радіально-базисних функцій), при цьому достатньо мати навчальну вибірку із 2408 прикладів; точність визначення атаки Smurf склало 0,99.

У подальшому для виявлення мережевих атак доцільно провести дослідження деяких комбінованих варіантів з використанням мережі RBF.

Список літератури:

1. Бодянский Е. В., Руденко О. Г. Искусственные нейронные сети: архитектуры, обучение, применения. Харьков: ТЕЛЕТЕХ, 2004. 369 с.
2. Котов В. Д., Васильев В. И. Система обнаружения сетевых вторжений на основе механизмов иммунной модели. *Известия Южного федерального университета. Технические науки.* 2011. С. 180-189. URL: https://cyberleninka.ru/article/n/sistema-obnaruzheniya-setevyh-vtorzheniy-na-osnove-mehanizmov-immunnoy-modeli/viewer
3. Мустафаев А. Г. Нейросетевая система обнаружения компьютерных атак на основе анализа сетевого трафика. *Вопросы безопасности.* 2016. № 2. С. 1–7. DOI: 10.7256.2409-7543.2016.2.18834
4. Пахомова В. М., Коннов М. С. Дослідження двох підходів до виявлення мережних атак із використанням нейромережної технології. *Наука та прогрес транспорту.* 2020. № 3 (87). С. 81-93. URL: https://doi.org/10.15802/stp2020/208233
5. Пахомова В. М., Маслак А. В. Визначення атак категорії Probe з використанням бази даних KDDCup99 та нейронечіткої технології. *Вчені записки Таврійського національного університету імені В.І. Вернадського*. Серія: Технічні науки. Том 33 (72). № 5, 2022. С. 135–140. DOI: https://doi.org/10.32872/2663-5941/2022.5/19
6. Слеповичев И. И., Ирматов П. В., Комарова М. С., Бежин А. А. Обнаружение DDoS-атак нечеткой нейронной сетью. *Известия Саратовского университета*. Серия: «Математика. Механика. Информатика». 2017. № 3. С. 84–89.
7. Фролов П. В., Чухраев И. В., Гришанов К. М. Применение искусственных нейронных сетей в системах обнаружения вторжений. *Системный администратор*. 2018. № 9 (190). URL: http://samag.ru/archive/article/3724
8. Alguliyev R. M., Aliguliyev R. M., Imamverdiyev Y. N., Sukhostat L. V. An improved ensemble approach for DoS attacks detection. *Радіоелектроніка, інформатика, управління*. 2018. № 2. С. 73-82. DOI: 10.15588/1607-3274-2018-2-8
9. 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
10. Bajer D., Zoric B., Martinovic G. Automatic design of radial basis function networks through enhanced different evolution. *Hybrid artificial intelligent systems*: Springer. 2015. pp. 244–256.
11. Beheim L., Zitouni A. New RBF neural network classifier with optimized hidden neurons number. *CiteSeerX 10.1.1.497.5646*. Belloir, Fabien. January 2004.
12. KDDCup1999Data.URL:http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
13. Pakhomova V. Determination of network attacks using neural network technologies. Chapter 8. pp. 113-128. Prospektive globale wissenschaftliche trends: Innovative Technik, Transport, Sicherheit. *Monografische Reihe «Europaische Wissenschaft»*. Buch 7. Teil 8. Germany: Karlsruhe, 2021. 168 p.
14. Rust: A language empowering everyone to build reliable and efficient software. URL: https://www.rust-lang.org
15. Saied A., Overill R. E., Radzik T. Detection of known and unknown DDoS attacks using Artificial Neural Networks. *Neurocomputing*. 2016. Vol. 172. pp. 385–393. URL: https://doi.org/10.1016/j.neucom.2015.04.101
16. Schwenker F., Kestler H. A. Three learning phases for radial-basis-function networks. *Neural Networks 14*: Palm, Günther, 2001. pp. 439-458. DOI: 10.1016/s0893-6080(01)00027-2
17. Wu Y., Wang H., Zhang B., Du. Using radial basis function networks for function approximation and classification. *ISRN Appl Math*. 2012. pp. 1–34.

18. 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. *Наука та прогрес транспорту*. 2020. № 5 (89). pp. 68–79.

Pakhomova V.M., Motylenko V.A. STUDYING THE POSSIBILITY OF USING RBF FOR DETERMINING SMURF ATTACKS BASED ON THE KDDCup DATABASE

Intrusion detection systems are used to detect network attacks in real time. One of the most effective approaches in classifying a large volume of data is the use of neural network technology, which allows detecting not only known network attacks, but also new ones. Today, it is known that false positives do not always occur on the same network packets when analyzed using different types of neural networks: multilayer perceptron; Kohonen network or self-organizing map; radial base network; neurofuzzy network, as well as their combinations. To determine network attacks of the DoS category using the KDDCup database, the program «RBF_DoS» was created in the Rust language, based on the implementation of the RBF network of the N-M-K configuration, where N is the number of input neurons (network traffic parameters); M is the number of hidden layer neurons (the number of basis functions); K is the number of resulting neurons (network classes of attacks) by the method of stochastic gradient descent, the Gaussian function is taken as the membership function of hidden neurons. Approbation of the «RBF_DoS» program was carried out based on RBF configuration 29-50-3 to determine the following clusters (Normal; Smurfs; Another_attack) using the following samples: training, which consisted of 200,000 examples for each cluster; test, consisting of 10,000 examples for each cluster; control, consisting of 100 examples for each cluster. The created program «RBF_DoS» carried out a study of the accuracy and root mean square error of RBF according to the following parameters: training epochs; the length of the training sample; by the number of hidden neurons.It was determined that when detecting attacks of the Smurf network class, the smallest RBF value was achieved in 10 training epochs using 101 hidden neurons, while it is sufficient to have a training sample of 2408 examples; the accuracy of determining the Smurf attack was 0.99.

Key words: *attack, Smurf, RBF, configuration, Gaussian function, precision, error, sampling, epoch.*