УДК 004.056.53:[004.7:004.032.26]
DOI https://doi.org/10.32782/2663-5941/2022.5/19

Пахомова В.М.
Український державний університет науки і технологій

Маслак А.В.
Український державний університет науки і технологій

ВИЗНАЧЕННЯ АТАК КАТЕГОРІЇ PROBE З ВИКОРИСТАННЯМ БАЗИ ДАНИХ KDDCUP99 ТА НЕЙРОНЕЧІТКОЇ ТЕХНОЛОГІЇ

Сучасний світ неможливо уявити без комп'ютерних мереж: як локальних, так і глобальних; тому питання мережевої безпеки стає все більш злободенним. На сучасному етапі найчастіше пропонуються системи виявлення мережевих атак, що побудовані на основі наступних нейронних мереж: багатошарового персептрону, мережі Кохонена або самоорганізуючої карти, мережі радіально-базисних функцій. Наразі методики виявлення мережевих атак можна підсилити використанням нейронечіткої технології, що підтверджує актуальність теми. Метою дослідження є визначення параметрів якості виявлення мережевих атак на основі бази даних KDDCup99 та з використанням нейронечіткої технології. У якості методу дослідження використана адаптивна мережа нечіткого висновку ANFIS конфігурації 4-5-8-16-1 (де 4 – кількість вхідних нейронів; 5 – загальна кількість шарів; 8 – кількість нейронів першого прихованого шару; 16 – кількість нейронів другого прихованого шару; 1 – кількість результуючих нейронів), що створена за допомогою пакета Fuzzy Logic Toolbox системи MatLAB; за результуючу характеристику взято ступень впевненості, що атака відбулася з наступними термами: низький; середній; високий. На створеній ANFIS проведено дослідження похибки на вибірках різної довжини (60, 80, 100 прикладів) за різними методами оптимізації: Backprop та Hybrid. Визначено, що найменше значення похибки ANFIS склало за методом Hybrid (40 епох), при цьому достатньо мати вибірку із 60 прикладів. Досліджено, що на ANFIS оптимальної структури значення помилки другого роду склали 0,5 %; 1,5 %; 3 %; 4 % для класів Ipsweep; Satan; Portsweep і Nmap відповідно.

Ключові слова: *атака, ANFIS, похибка, вибірка, оптимізація, якість.*

Постановка проблеми. Створення ефективної системи виявлення мережевих атак вимагає застосування якісно нових підходів до обробки інформації, які повинні ґрунтуватися на адаптивних алгоритмах здатних до самонавчання. Найбільш перспективним напрямком у створенні подібних систем виявлення мережевих атак є застосування нейронечіткої технології.

Аналіз останніх досліджень і публікацій. Відомо, що для виявлення мережевих атак можливе використання наступних нейронних мереж (НМ): багатошарового персептрону (Multi Layer Perceptron, MLP) [3, 13, 15-17]; мережі Кохонена або самоорганізуючої карти (Self Organizing Map, SOM) [4-5, 10]; радіально-базисної мережі (Radial Basis Function Network, RBF) [7, 9]. НМ з різними топологіями (MLP, RBF, SOM) можуть визначати різні атаки, але помилкові спрацьовування також відбуваються не завжди на одних і тих самих мережевих пакетах при аналізі за допомогою різних типів НМ. На сучасному етапі з'являються роботи на основі комбінованого підходу [1, 5-9. 11, 14], в основі одного із них: 1) використання одного із типів НМ (MLP, RBF або SOM); 2) використання нейронечіткої мережі (ННМ) для підвищення точності виявлення, зменшення кількості помилкових спрацьовувань та забезпечення більш високого рівня виявлення для нечастих атак. Однак, разом з тим важливим недоліком таких методик є відсутність універсальності їх застосування при визначенні мережевих атак різних категорій (DoS, PROBE, R2L, U2R), проте значна кількість джерел присвячена дослідженню DoS-атак [2, 6, 8, 15]; крім того, необхідно провести додаткові дослідження стосовно оптимальної структури ННМ (кількості термів, функції приналежності, методу оптимізації, довжини вибірки).

Формулювання цілей статті. Проведені дослідження ставили за мету розвиток методики визначення атак категорії PROBE. Для досягнення поставленої мети вирішувалися наступні задачі: розробити методику виявлення мережевих атак з використання нейронечіткої технології; при виконанні машинного навчанні виявити оптимальні параметри ННМ, що забезпечить достатньо високий рівень достовірності виявлення

вторгнень в комп'ютерну мережу; оцінити помилки першого та другого роду при виявленні мережевих атак на створеній ННМ.

Виклад основного матеріалу дослідження. Атаки PROBE полягають в скануванні мережевих портів з метою отримання конфіденційної інформації. Відомі наступні класи мережевих атак відповідно до категорії PROBE: Portsweep, Ipsweep, Satan, Nmap. У якості початкових даних використана відкрита база даних KDDCup99 [12]. У якості методу дослідження використана адаптивна мережа нечіткого висновку (Adaptive-Network-Based Fuzzy Inference System, ANFIS), що поєднує методи штучної НМ та системи нечіткої логіки виводу Такагі-Сугено, структура якої подана на рис. 1.

Перший шар (input) має наступні нейрони: X1 (dst_host_srv_count) – сума підключень до того самого номера порту призначення; X2 (dst_host_srv_diff_host_rate) – відсоток підключень до різних машин призначення серед з'єднань, агрегованих у dst_host_srv_count; X3 (srv_diff_ host_rate) – відсоток підключень до різних машин призначення серед з'єднань, агрегованих у srv_count; X4 (diff_srv_rate) – відсоток підключень до різних служб.

Другий шар (inputmf) має 4*2 = 8 вузлів; кожному нейрону відповідає 2 терми.

Третій шар (rule) має 2^4=6 правил, які мають наступний вигляд:

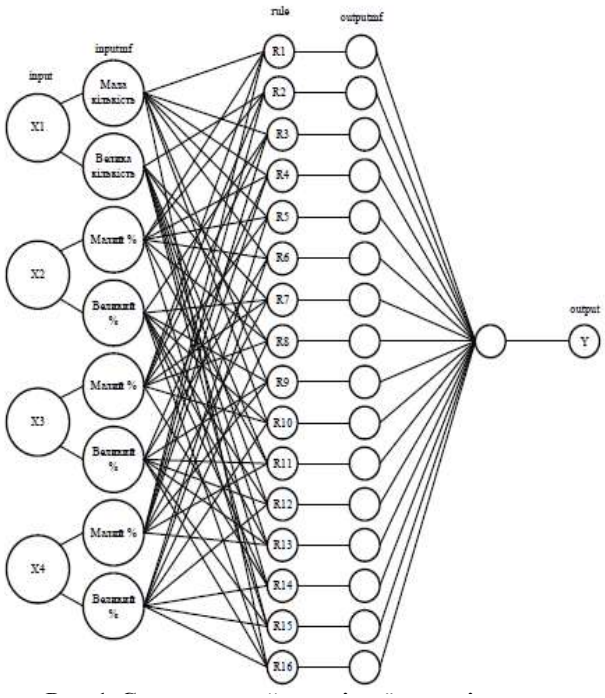

Рис. 1. Структура нейронечіткої мережі

1) якщо X1 = мінімальне значення І X2 = мінімальне значення І X3 = мінімальне значення І X4 = мінімальне значення, тоді низький ступень впевненості;

2) якщо X1 = максимальне значення І X2 = мінімальне значення І X3 = мінімальне значення І X4 = мінімальне значення, тоді низький ступень впевненості;

3) якщо X1 = мінімальне значення І X2 = максимальне значення І X3 = мінімальне значення І X4 = мінімальне значення, тоді низький ступень впевненості

4) якщо X1 = мінімальне значення І X2 = мінімальне значення І X3 = максимальне значення І X4 = мінімальне значення, тоді низький ступень впевненості;

5) якщо X1 = мінімальне значення І X2 = мінімальне значення І X3 = мінімальне значення І X4 = максимальне значення, тоді низький ступень впевненості;

6) якщо X1 = мінімальне значення І X2 = мінімальне значення І X3 = максимальне значення І X4 = максимальне значення, тоді середній ступень впевненості;

7) якщо X1 = максимального значення І X2 = максимального значення І X3 = мінімальне значення І X4 = мінімальне значення, тоді середній ступень впевненості;

8) якщо X1 = максимального значення І X2 = мінімальне значення І X3 = мінімальне значення І X4 = максимального значення, тоді середній ступень впевненості;

9) якщо X1 = мінімальне значення І X2 = максимального значення І X3 = максимального значення І X4 = мінімальне значення, тоді середній ступень впевненості;

10) якщо X1 = мінімальне значення І X2 = максимального значення І X3 = мінімальне значення І X4 = максимального значення, тоді середній ступень впевненості;

11) якщо X1 = максимального значення І X2 = мінімальне значення І X3 = максимального значення І X4 = мінімальне значення, тоді середній ступень впевненості;

12) якщо X1 = максимального значення І X2 = максимального значення І X3 = максимального значення І X4 = максимального значення, тоді високий ступень впевненості;

13) якщо X1 = максимального значення І X2 = максимального значення І X3 = максимального значення І X4 = мінімальне значення, тоді високий ступень впевненості;

14) якщо X1 = максимального значення І X2 = максимального значення І X3 = мінімальне значення І X4 = максимального значення, тоді високий ступень впевненості;

Таблиця 1

Фрагмент навчальної вибірки (25 із 100 прикладів)

X1	X2	X3	X4	Y	X1	X2	X3	X4	Y
26	0,67	0	0	1	255	0,09	0,14	0	1
36	0,50	0	0	1	255	0,09	0	0	1
119	0,18	0	0	0	227	0,09	0,14	0	1
255	0,17	0	0,67	1	197	0,09	0	0	1
180	0,17	0	0	1	187	0,09	1	0	1
255	0,13	0,40	0	1	177	0,09	0,33	0	1
127	0,13	0	0	1	57	0,09	1	0	1
16	0,13	0	0	0	56	0,09	1	0	1
255	0,12	0	1	1	47	0,09	1	0	1
15	0,11	0,40	0	0	17	0,09	0,67	0	1
237	0,10	1	0	1	7	0,09	1	0	1
117	0,10	0,50	0	1	6	0,09	0,67	0	1
50	0,10	0,12	0	0	…	…	…	…	…

15) якщо X1 = максимального значення І X2 = мінімальне значення І X3 = максимального значення І X4 = максимального значення, тоді високий ступень впевненості;

16) якщо X1 = мінімальне значення І X2 = максимального значення І X3 = максимального значення І X4 = максимального значення, тоді високий ступень впевненості.

Четвертий шар (outputmf) складається з функцій приналежності для кожного правила нечіткого виводу; кількість вузлів цього шару відповідає кількості правил 2^4=6.

П'ятий шар (output) – результуючий шар: Y – ступень впевненості, що атака відбулася, який має три терми: низький; середній; високий.

Формування вибірки. Навчальна вибірка складалася із 100 прикладів, фрагмент якої представлений в табл. 1, тестувальна вибірка – із 70 прикладів.

Створення ННМ. За допомогою пакета Fuzzy Logic Toolbox в MatLAB створено ННМ конфігурації 4-5-8-16-1; у якості функції приналежності нейронів взято Гаусовську функцію. Згенеровану в MatLAB структуру ННМ показано на рис. 2.

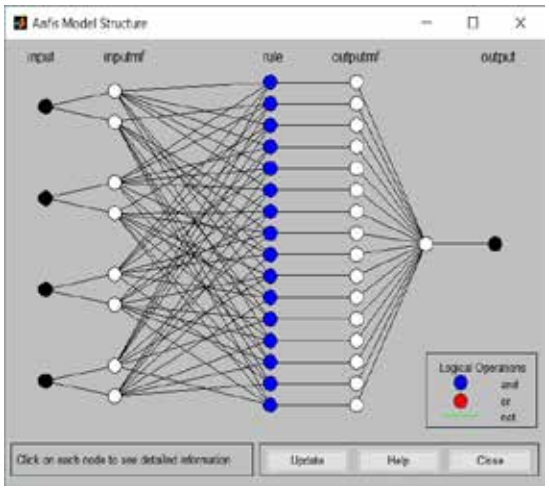

Рис. 2. Структура згенерованої ННМ в MatLAB

Апробація ННМ. Результати апробації ННМ зведено до таблиці 2

Таблиця 2

Результати апробації ННМ

KDDCup99		ANFIS		Ступень впевненості
мережевий клас	факт	[X1 X2 X3 X4]	відгук	
Nmap	0	[119 0,18 0 0]	0,2754	низький
Nmap	1	[255 0,17 0 0,67]	0,8013	високий
Nmap	1	[180 0,17 0 0]	0,9896	високий
Portsweep	1	[6 0,09 0,67 0]	0,9077	високий
Nmap	1	[127 0,13 0 0]	0,9021	високий
Ipsweep	0	[16 0,13 0 0]	0,0104	низький
Satan	1	[25 0,10 0 1]	0,8924	високий
Nmap	0	[15 0,11 0,4 0]	0,1211	низький
Satan	1	[237 0,1 1 0]	0,9361	високий
Satan	1	[117 0,1 0,5 0]	0,8221	високий

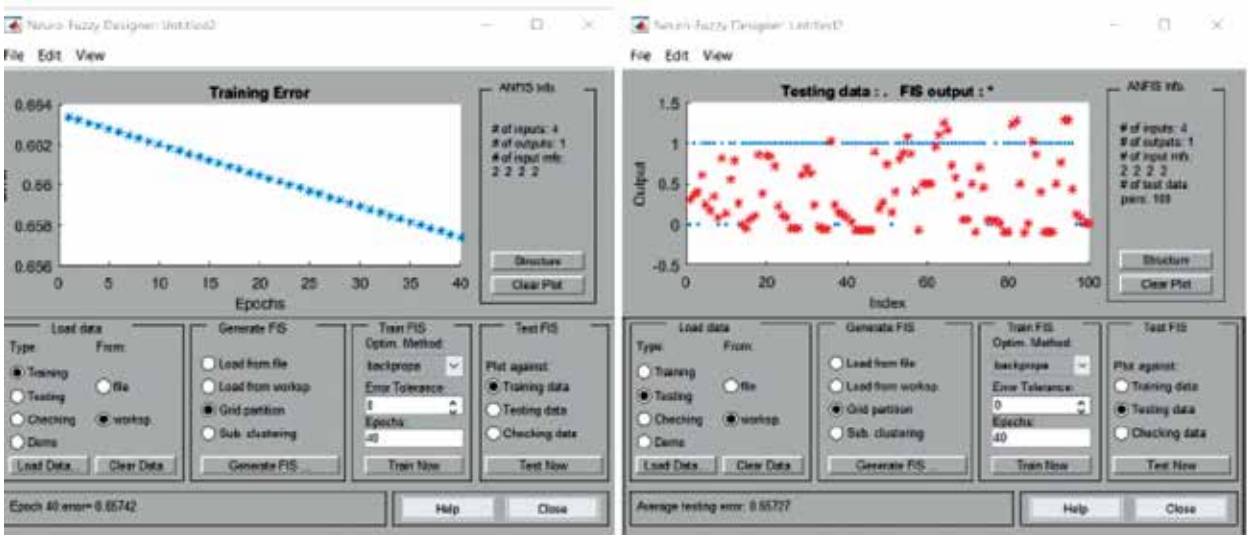

Рис. 3. Результати дослідження (вибірка із 100 прикладів)

Таблиця 3

Значення похибки ННМ за різними методами оптимізації

Кількість прикладів	Backpropa		Hybrid	
	Навчання	Тестування	Навчання	Тестування
60	0,67698	0,67681	0,024245	0,024245
80	0,67094	0,67078	0,030822	0,030822
100	0,65742	0,65727	0,039481	0,039481

Таблиця 4

Помилки першого та другого роду при визначенні PROBE атак

Ipsweep		Nmap		Portsweep		Satan	
TP	**FP**	**TP**	**FP**	**TP**	**FP**	**TP**	**FP**
11 %	4 %	2 %	1%	9 %	4 %	15 %	3 %
FN	**TN**	**FN**	**TN**	**FN**	**TN**	**FN**	**TN**
0,5 %	13 %	4 %	10 %	3 %	12 %	1,5 %	7 %

Дослідження параметрів ННМ. На створеній ANFIS проведено дослідження похибки на вибірках різної довжини: 60; 80; 100 прикладів за різними методами оптимізації: Backpropa (метод зворотного поширення помилки, заснований на ідеях методу найшвидшого спуску); Hybrid (гібридний метод, який об'єднує метод зворотного поширення помилки з методом найменших квадратів). У якості прикладу на рис. 3 подані результати дослідження за методом Backpropa **(**вибірка із 100 прикладів**)**.

Найменші значення похибки ННМ склали приблизно 0,02 та 0,68 відповідно за методами Hybrid і Backpropa (табл. 3); при цьому достатньо мати вибірку із 60 прикладів.

Оцінка параметрів якості. У ході проведення дослідження отримані на ННМ наступні результати: TP (True Positive); FP (False Positive); FN (False Negative); TN (True Negative), на основі яких на завершальному етапі залишилось дати оцінку якості рішень. Помилка першого роду – це кількість невірно виявлених атак (FP, False Positive), а помилка другого роду – це кількість пропусків атак (FN, False Negative); обчислені значення цих помилок при визначенні атак наступних мережевих класів: Ipsweep, Nmap, Portsweep і Satan зведені до таблиці 4. Із таблиці видно, що значення помилки першого роду склали 1%, 3 % і 4 % для класів Nmap; Satan і Ipsweep (Portsweep) відповідно, а значення помилки другого роду склали 0,5; 1,5 %; 3 %; 4 % для класів Ipsweep; Satan; Portsweep і Nmap відповідно.

Практична значимість. У [4] найменшу точність показала SOM при визначенні атак класу Nmap на основі бази даних KDDCup99, одночасне використання створеної ANFIS та SOM надає можливість підвищити точність виявлення атак цього класу.

Висновки. Для визначення ступеню впевненості здійснення атаки категорії PROBE на основі використання бази даних KDDCup99 створена за допомогою пакета Fuzzy Logic Toolbox системи MatLAB ННМ конфігурації 4-5-8-16-1 (де 4 – кількість вхідних

нейронів; 5 – загальна кількість шарів; 8 – кількість нейронів першого прихованого шару; 16 – кількість нейронів другого прихованого шару; 1 – кількість результуючих нейронів), у якості функції приналежності нейронів взято Гаусовську функцію. На створеній ANFIS проведено дослідження похибки на вибірках різної довжини (60; 80; 100 прикладів) за різними методами оптимізації: Backpropa і Hybrid. Найменше значення похибки ANFIS склало за методом Hybrid, при цьому достатньо мати вибірку із 60 прикладів. Визначено, що на створеній ANFIS значення помилки другого роду склали 0,5 %; 1,5 %; 3 %; 4 % для класів Ipsweep; Satan; Portsweep і Nmap відповідно. У подальшому для виявлення PROBE атак з використанням відповідних баз даних доречно провести дослідження декількох комбінованих варіантів, основу яких буде складати створена ННМ.

Список літератури:

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. 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.
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). С. 81-93. URL: https://doi.org/10.15802/stp2020/208233.
4. 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.
5. Пахомова В. М., Видиш А. Д. Дослідження комбінованого варіанту визначення атак з використанням нейромережних технологій. *Системні технології*. Регіональний міжвузівський збірник наукових праць. № 3(140). 2022. С. 79-86. DOI: 10.34185/1562-9945-3-140-2022-08.
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 **(**дата звернення: 10.01.2022**).**
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. P. 1–32. DOI: https://doi.org/10.1142/s0218213015500335.
10. Esteban J. A New GHSOM Model applied to network security. *Artificial Neural Networks-ICANN 2008*. 2008. P. 680-689.
11. Hadi A. A. A. Performance Analysis of Big Data Intrusion Detection System over Random Forest Algorithm. *International Journal of Applied Engineering Research*. 2018. Vol. 13, No. 2. P. 1520-1527.
12. KDD Cup 1999 Data. URL: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html (дата звернення: 05.04.2022).
13. Pakhomova V. M., Bikovska D. G. Investigation of multilayer neural network parameters for determination of R2L category network attacks. *Modern engineering and innovatite technologies.* Germany, Karlsruhe: Sergeieva&Co, «ISE&E». 2021. № 18-02. pp. 39-43. DOI: 10.30890/2567-5273.2021-18-02-059.
14. Pakhomova V. N. Determination of network attacks using neural network technologies. Chapter 8. pp. 113-128. DOI: 10.30890/2709-2313.2021-07-08-003. Prospektive globale wissenschaftliche trends: Innovative Technik, Transport, Sicherheit. *Monografische Reihe «Europaische Wissenschaft»*. Buch 7. Teil 8. Germany: Karlsruhe, 2021. 168 p.
15. Saied A., Overill R. E., Radzik T. Detection of known and unknown DDoS attacks using Artificial Neural Networks. *Neurocomputing*. 2016. Vol. 172. P. 385-393. URL: https://doi.org/10.1016/j.neucom.2015.04.101**.**
16. Zhukovyts'kyy I. V., Pakhomova V. M. Identifying threats in computer network based on multilayer neural network. *Наука та прогрес транспорту*. 2018. № 2 (74). P. 114-123. DOI: https://doi.org/10.15802/stp2018/130797.
17. 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). P. 68-79. URL: https://doi.org/10.15802/stp2020/218318.

Pakhomova V.M., Maslak A.V. NETWORK ATTACK DETECTION USING KDDCup99 DATABASE AND NEURON FUZZY TECHNOLOGY

The modern world cannot be imagined without computer networks: both local and global; therefore, the issue of network security is becoming more and more urgent. At the current stage, network attack detection systems built on the basis of the following neural networks are most often proposed: a multilayer perceptron, a Kohonen network or a self-organizing map, a network of radial basis functions. Currently, methods of detecting network attacks can be strengthened using neurofuzzy technology, which confirms the relevance of the topic. The purpose of the study is to determine the parameters of the quality of detection of network attacks based on the KDDCup99 database and using neurofuzzy technology. As a research method, an adaptive fuzzy inference ANFIS network of the 4-5-8-16-1 configuration was used (where 4 is the number of input neurons; 5 is the total number of layers; 8 is the number of neurons of the first hidden layer; 16 is the number of neurons of the second hidden layer; 1 is the number of resulting neurons) created using the Fuzzy Logic Toolbox package of the MatLAB system; the resulting characteristic is the degree of confidence that the attack took place with the following terms: low; average; high. On the created ANFIS, an error study was carried out on samples of different lengths (60, 80, 100 examples) using different optimization methods: Backpropa and Hybrid. It was determined that the smallest error value of the ANFIS was obtained by the Hybrid method (40 epochs), while it is enough to have a sample of 60 examples. It was found that the second kind was 0,5 %; 1,5 %; 3 %; 4 % for Ipsweep; Satan; Portsweep і Nmap corresponding.

Key words: *attack, ANFIS, error, sampling, optimization, quality.*