



УДК 004.056.53:[004.7:004.032.26]

RESEAECH OF PARAMETERS OF QUALITY OF DEFINITION OF NETWORK ATTACKS OF THE PROBE CATEGORY WITH USE OF THE SELF ORGANIZING MAP**ДОСЛІДЖЕННЯ ПАРАМЕТРІВ ЯКОСТІ ВИЗНАЧЕННЯ МЕРЕЖЕВИХ АТАК КАТЕГОРІЇ PROBE З ВИКОРИСТАННЯМ САМООРГАНІЗУЮЧОЇ КАРТИ****Pakhomova V.M. / Пахомова В.М.***s.t.s., as.prof. / к.т.н., доц.*

ORCID: 0000-0002-0022-099X

Pavlenko I.I. / Павленко І.І.*master / magіstr*

ORCID: 0000-0003-4941-6755

*Ukrainian State University of Science and Technology, Dnipro, Lazaryan St., 2, 49010
Український державний університет науки і технологій, Дніпро, вул. Лазаряна, 2, 49010*

Анотація. Для визначення мережесих атак категорії Probe створена в Python програма «SOM_Probe» на основі реалізації самоорганізуючої карти Кохонена 30*30, структура якої 15-5 (де 15 – кількість вхідних параметрів мережевого трафіку, 5 – кількість результуючих нейронів), та з використанням відкритої бази KDDCup. Визначені параметри якості (Precision, Recall, F-мірка) виявлення наступних класів атак: Portsweeper, Ipsweeper, Satan, Nmap, а також проведені дослідження F-мірки за різною кількістю епох навчання нейронної мережі.

Ключові слова: категорія, Probe, клас, атака, SOM, точність, повнота, F-мірка.

Вступ

Постановка проблеми. Створення ефективної системи виявлення мережесих 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 мережесих атак є застосування нейромережної технології.

Аналіз останніх досліджень. Відомо, що для виявлення мережесих атак можливе використання наступних нейронних мереж (НМ): багатошарового перцептронну (Multi Layer Perceptron, MLP); мережі Кохонена або самоорганізуючої карти (Self Organizing Map, SOM); радіально-базисної мережі (Radial Basis Function Network, RBF), серед яких для кластеризації найбільш підходить SOM [3]. Так, наприклад, Palomo A. та Esteban J. провели дослідження, використовуючи зростаючу ієрархічну самоорганізуючу карту при використанні бази KDD99. У результаті визначено майже максимальний рівень виявлення атаки, що складає 99,99 %, але при цьому є суттєвий недолік, який полягає в великій ймовірності помилки 5,44 % в зрівнянні з іншими НМ. Ortis Andres також проводила дослідження із зростаючою ієрархічною самоорганізуючою картою. У результаті створена НМ з підвищенням швидкості виявлення атак, яка при навчанні використовувала метод маркування ймовірностей. Для даної НМ використовувалася база NSL-KDD, досягнуто найвищу частоту виявлення атак 99,68 % з найменшим помилковим спрацьовуванням 0,02 %. У [1] доведено, що якість класифікації залежить від кількості еталонів окремих класів (Dos, Probe, U2R, R2L) у навчальній вибірці.



Дана робота виконана відповідно до НДР [2]. *Метою статті* є дослідження параметрів якості визначення мережевих класів атак категорії Probe з використанням самоорганізуючої карти.

1. Постановка задачі та вибір методу

Атаки Probe полягають в скануванні мережевих портів з метою отримання конфіденційної інформації. Відомі наступні класи мережевих атак відповідно до категорії Probe: Portswweep, Ipsweep, Satan, Nmap. У якості початкових даних використана відкрита база даних KDDCup [5]. У якості математичного засобу взята самоорганізуюча карта Кохонена, що навчається без вчителя. Структура SOM має єдиний шар нейронів (шар Кохонена) без коефіцієнтів зсуву та показана на рис. 1.

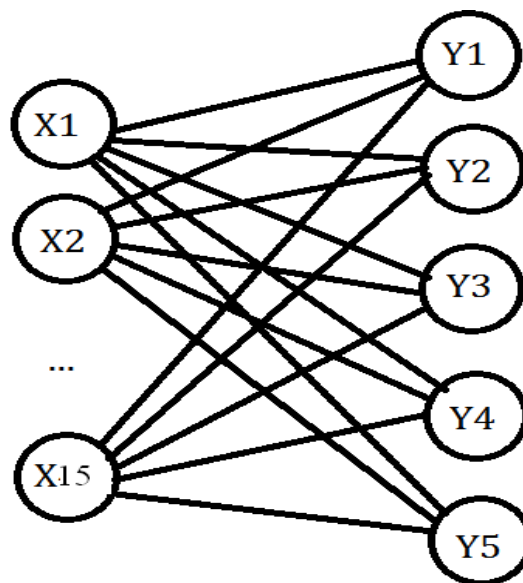


Рисунок 1 – Структура самоорганізуючої карти Кохонена

У якості вхідних змінних НМ параметри мережевого трафіку $x_1 \dots x_{15}$, де x_1 – Protocol type; x_2 – Service; x_3 – Flag; x_4 – Source bytes; x_5 – Destination bytes; x_6 – Hot; x_7 – Logged in; x_8 – Count compromised; x_9 – Serror rate; x_{10} – Error rate; x_{11} – Same srv rate; x_{12} – Diff srv rate; x_{13} – Srv Serror rate; x_{14} – Srv Error rate; x_{15} – Srv diff host rate. У якості результуючих нейронів $y_1 \dots y_5$, де y_1 відповідає Portswweep, y_2 – Ipsweep, y_3 – Satan, y_4 – Nmap, y_5 – атаки не було.

Ми вважаємо, що використання SOM [6] в якості математичного апарату є доцільним і достатнім. Хоча RBF і навчається швидше ніж MLP, але необхідно визначити кількість радіальних елементів, розташовування їх центрів і значення відхилення, модель RBF потребує декілька більшої кількості елементів, тобто буде працювати повільніше і потребує більше пам'яті, ніж модель MLP. Однак, при створенні MLP необхідно провести додаткові дослідження функцій активації нейронів, алгоритмів навчання та довжини вибірки.

За допомогою мови «Python» створена програмна модель «SOM_Probe», в основу якої покладена реалізація самоорганізуючої карти. У якості основного фреймворку для створення SOM використано MiniSom (мінімалістичну реалізацію самоорганізуючої карти на основі Numpy). Для відображення інформації обрана бібліотека Matplotlib, до складу якої входить модуль



Matplotlib.pyplot який містить в собі функції для графічного відображення інформації. У якості допоміжної бібліотеки обрано NumPy (бібліотеку з відкритим вихідним кодом, яка поєднує в собі багато математичних функцій). Для аналізу даних використаний модуль Metrics від Sklearn, що включає функції оцінки, метрики продуктивності та обчислення відстані.

2. Навчання та тестування нейронної мережі

Для навчання SOM подана вибірка із 400 прикладів, фрагмент якої наведено на рис. 2. Для тестування SOM використано вибірку із 205 прикладів. У результаті роботи програми «SOM_Probe» отримані карти 30*30, які приведені на рис. 3(а, б) при навчанні та тестуванні НМ відповідно. Розподілення атак на картах наступне: «красне кільце» – Portswweep, «зелений квадрат» – Ipsweep, «синій плюс» – Satan, «жовтий крестик» – Nmap.

3. Дослідження параметрів якості визначення мережевих класів атак

На створеній програмі «SOM_Probe» отримані параметри якості виявлення мережевих класів категорії Probe, деякі із яких зведені в табл. 1 (250000 epoch). У таблиці збалансована F-мірка – це гармонічне середнє між Precision та Recall.

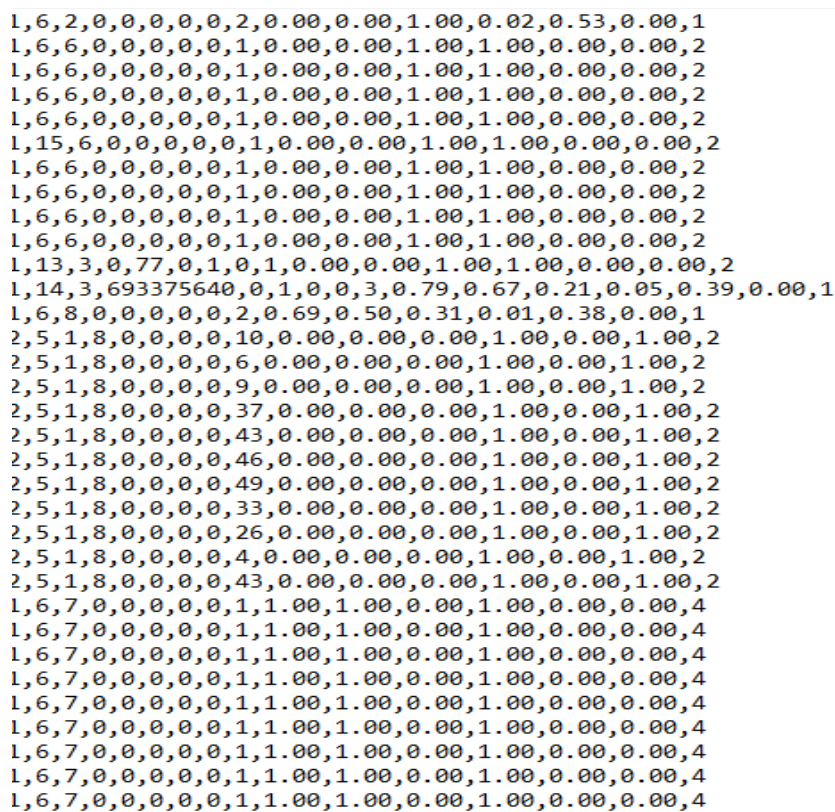


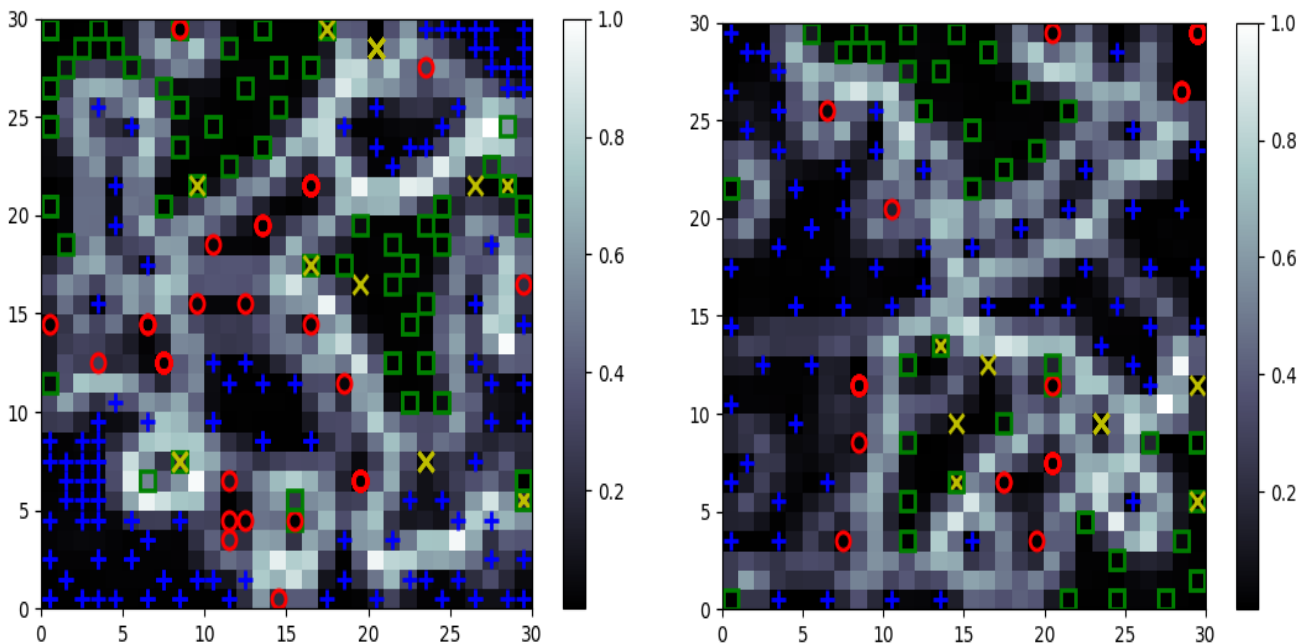
Рисунок 2 – Фрагмент навчальної вибірки

Таблиця 1 – Параметри якості, що отримані на програмі «SOM Probe»

Клас атаки	Навчання НМ			Тестування НМ		
	Precision	Recall	F-мірка	Precision	Recall	F-мірка
Portswweep	1,00	0,97	0,98	1,00	0,98	0,99
Ipsweep	0,93	0,99	0,93	0,97	0,97	0,97
Satan	1,00	1,00	1,00	1,00	1,00	1,00
Nmap	0,94	0,74	0,94	0,83	0,91	0,83



Із таблиці видно, що НМ добре визначає типи мережових атак. Найменшу точність НМ показала при визначенні атак типу Nmap. Тобто НМ добре визначає мережеві атаки типу Portsweep, Ipsweep та Satan, але при визначенні типу Nmap можуть виникати помилки.



**Рисунок 3 – Результат роботи програми «SOM_Probe»:
(а) - при навчанні НМ; (б) - при тестуванні НМ**

4. Дослідження F-мірки за різною кількістю епох навчання НМ

На створеній програмі «SOM_Probe» проведено дослідження F-мірки за різною кількістю епох навчання НМ (рис. 4), апробація результатів в [4].

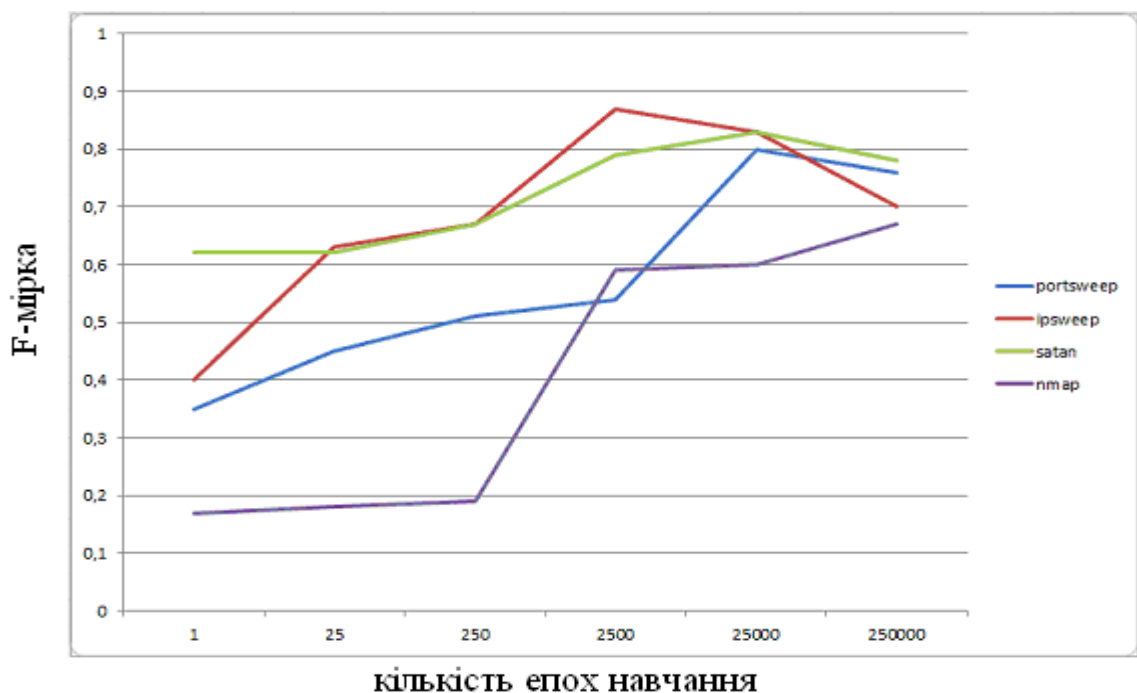


Рисунок 4 – F-мірка за різною кількістю епох навчання



Із рисунку видно, що для всіх типів атак (крім Nmap) F-мірка зростає, досягаючи свого максимуму на відмітці 25000 епох, після чого спостерігається невеликий спад мірки. Але для Nmap спостерігається постійний ріст, максимум якого випадає на 250000 епох. Незначний ріст значення F-мірки для Nmap та встановлення кількості епох навчання на 250000 буде означати втрати на визначенні інших типів атак і зниження загальної точності.

Висновки

На основі створеної в Python програми «SOM_Probe» з використанням самоорганізуючої карти Кохонена та бази даних KDDCup проведено дослідження параметрів якості визначення мережових класів атак категорії Probe. НМ добре визначає мережові атаки типів: PortswEEP; Ipsweep; Satan, але при визначенні типу Nmap можуть виникати помилки. Крім того, проведено дослідження F-мірки за різною кількістю епох навчання з використанням програми «SOM_Probe»: для всіх типів атак (крім Nmap) F-мірка зростає.

Література:

1. Емельянова Ю.Г., Талалаев А.А., Тищенко И.П., Фраленко В.П. Нейросетевая технология обнаружения сетевых атак на информационные ресурсы. Программные системы: теория и приложения. № 3(7). 2011. С. 3-15.
2. НДР «Дослідження механізмів визначення мережових атак з використанням методів штучного інтелекту»; наук. кер.: доц. Пахомова В.М. Державний реєстраційний номер: 0121U110676. 2021-2022 р.р.
3. Пахомова В.М., Коннов М.С. Дослідження двох підходів до виявлення мережних атак із використанням нейромережової технології // Наука та прогрес транспорту. 2020. № 3(87). С. 81-93. DOI: 10.15802/stp2020/208233.
4. Пахомова В.М., Павленко І.І. Дослідження на створеній мережі Кохонена виявлення PROBE атак. Сучасні інформаційні та комунікаційні технології на транспорті, в промисловості та освіті: Тези XV Міжнародної науково-практичної конференції (Дніпро, 16-17 грудня 2021 р.): УДУНТ, 2021. С. 198.
5. KDDCup1999Data.URL:<http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html>
6. Kohonen T. The self-organizing map. Proceedings of the IEEE. 1990. Vol. 78. No 9. pp.1464-1480.

Abstract. To determine network attacks of the Probe category, Python created the program "SOM Probe" based on the implementation of Kohonen Self Organizing Map 30*30, the structure of which is 15-5 (where 15 - the number of network traffic input parameters, 5 - the number of resulting neurons), and using an open database KDDCup. Quality parameters (Precision, Recall, F-measure) of detection of the following classes of attacks are defined: PortswEEP, Ipsweep, Satan, Nmap, and also researches of F-measure on various quantity of epochs of training of a neural network are carried out.

Keywords: category, Probe, class, attack, SOM, precision, recall, F-measure.