УДК 004.942
DOI: https://doi.org/10.53920/ITS-2022-2-4

Ольга Миколаївна ТКАЧЕНКО,
д-р техн., проф.,
Київський національний університет імені Тараса Шевченка
ORCID ID: 0000-0001-7983-9033
Владислав Олексійович СОСНОВИЙ,
асистент кафедри комп'ютерної інженерії
Державний університет телекомунікацій
ORCID ID: 0000-0002-3217-4537

МОДЕЛЬ ПРОГНОЗУВАННЯ БЕЗПЕКИ МЕРЕЖІ ЗА ДОПОМОГОЮ НЕЙРОННИХ МЕРЕЖ

В статті розглянуто чотири алгоритми, а саме алгоритм SVM, алгоритм нечіткої кластеризації, алгоритм кластеризації K-Means і алгоритм Apriori. Деталізуємо 4 різних кроки безпеки користувачів мережі та їх контролю доступу статті є розробка надійної моделі прогнозування безпеки мережі. Розроблена модель виявлення вторгнень, побудована з використанням нейронних мереж. Модель виявлення вторгнень виявляє аномалії та атаки на основі зловживання. Модель виявлення вторгнень також виконує три типи завдань класифікації. Завдання включають класифікацію між появою атаки чи звичайним випадком, класифікацією між різними типами атаки чи звичайним випадком . Модель виявлення вторгнень також показує точність класифікації, час виконання та обсяг використання пам'яті. Цілями моделі виявлення вторгнень є висока точність, малий час виконання та мінімальний обсяг використання пам'яті. Модель виявлення вторгнень, побудована за допомогою нейронних мереж, відповідає цілям високої точності, малого часу виконання та мінімального використання пам'яті.

Ключові слова: *алгоритм SVM, алгоритм кластеризації K-Means, алгоритм Apriori, Алгоритм нечіткої кластеризації, Нейронна мережа.*

Olha TKACHENKO
Doctor of technical sciences, professor
Taras Shevchenko National University of Kyiv
ORCID ID: 0000-0001-7983-9033
Vladyslav SOSNOVYY
assistant of the department of computer engineering
State University of Telecommunications
ORCID ID: 0000-0002-3217-4537

NETWORK SECURITY PREDICTION MODEL USED BY NEURAL NETWORKS

The article discusses four algorithms, namely the SVM algorithm, the fuzzy clustering algorithm, the K-Means clustering algorithm, and the Apriori algorithm. We detail the 4 different steps of network user security and their access control. The article is the development of a reliable network security prediction model. An intrusion detection model built using neural networks has been developed. The intrusion detection model detects anomalies and abuse-based attacks. The intrusion detection model also performs three types of classification tasks. Tasks include classification between the occurrence of an attack or a normal case, classification between different types of attack or a normal case. The intrusion detection model also shows classification accuracy, execution time, and memory usage. The goals of the intrusion detection model are high accuracy, low execution time, and minimal memory usage. An intrusion detection model built using neural networks meets the goals of high accuracy, low execution time, and minimal memory usage.

In today's world, networks are becoming increasingly complex, interconnected, and widely used. Today, network traffic is growing almost exponentially. Networks also become more vulnerable to attack by hackers or anyone with malicious intent to disrupt network systems. Vulnerable networks are at risk of a blow to the economy and the destruction of confidential information. Thus, there is a need to improve network vulnerability detection mechanisms and improve network security prediction. The network security prediction model also aims to reduce memory consumption and improve the detection of different types of attacks in terms of timing and accuracy. In the network security prediction model, the memory consumption was low, and the time spent to detect attacks was also low. Attack detection accuracy is also high.

The above methods used to design the model are also easy to design. The above methods are also much more cost-effective since the use of neural networks is free. In addition, calculations are simplified by using this model. Therefore, using a neural network is also an effective way to develop a network security prediction model. Thus, the use of neural networks is recommended for the development of any type of network security prediction model. Future tasks are to develop models that will detect any intrusions even more accurately and quickly.

Keywords: *SVM algorithm, K-Means clustering algorithm, Apriori algorithm, Fuzzy clustering algorithm. Neural network.*

Постановка проблеми. У сучасному світі мережі стають все складнішими, взаємопов'язаними та широко використовуються. Сьогодні мережевий трафік зростає майже в геометричній прогресії. Мережі також стають більш вразливими до атак з боку хакерів або будь-кого зі злими намірами зруйнувати мережеві системи. Вразливі мережі знаходяться під загрозою удару по економіці та знищенні конфіденційної інформації. Таким чином, існує потреба у вдосконаленні механізмів виявлення вразливості мережі та покращенні прогнозування безпеки мереж. Модель прогнозування безпеки мережі також має на меті зменшити споживання пам'яті, а також покращити виявлення різних типів атак з точки зору часу та точності.

Аналіз останніх досліджень і публікацій. Було розглянуто моделі виявлення атак. Усі моделі мають спільну мету — точніше, ефективніше та швидше виявляти вразливості в мережі. Для досягнення мети були запропоновані різні алгоритми [1]. Запропонований метод виявлення вторгнень, повністю заснований на алгоритмі K-методі [2]. Був запропонований гібридний алгоритм виявлення вторгнень, заснований на K-методі та дереві вибору [3]. Для попередньої обробки 41- функції у статистичному наборі використовували моніторинг функцій даних [4].

Мета статті — опис створення файлів, які використовуються для виявлення аномальних атак. Описати деталі кількості атак або звичайних випадків для процесу виявлення аномалій або атак на основі неправильного використання. Описати спосіб видалення непотрібних або малокорисних непотрібних даних, щоб отримати оптимальну кількість даних для класифікацій. Використати нейронну мережу для виявлення різноманітних атак. Розрізнити різні

процеси класифікації, які відбуваються в атаках виявлення аномалій, атаках виявлення неправильного використання та окремих типів атак. Виявити аномалії за допомогою класифікації виникнення атаки або звичайного випадку.

Виклад основного матеріалу. Розглянемо чотири алгоритми, а саме алгоритм SVM, алгоритм нечіткої кластеризації, алгоритм кластеризації K- Means і алгоритм Apriori. Далі деталізуємо 4 різних кроки безпеки користувачів мережі та їх контролю доступу. Алгоритм SVM використовується для вирішення проблем класифікації. Базуючись на базовій конструкції статистичного принципу, до процесу обчислення додається функція ядра для відображення проблеми низької розмірності в просторі високої розмірності і отримує простір рішень високої розмірності. Це означає, що використання алгоритму SVM розблокує приховані шаблони у великій кількості даних, щоб виявити інформацію. Після завантаження інформації ,система може ідентифікувати часові ряди або тенденцію розвитку даних і робити точні висновки [5]. Було також використано метод автоматичного отримання найбільш оптимальних параметрів Гауса для отримання найкращої гіпосфери [6]. Була вдосконалена версія, у якій алгоритм K-методу змінено на комбінацію з Apriori для досягнення правильного значення виявлення Root to Learn і User to Root за інформацією БД KDDCUP99, встановленою на 98% і 79% [7]. Була запропонована ідея використання набору правил розмірності, змішаного з одно-класовою SVM [8].

Алгоритм нечіткої кластеризації виглядає наступним чином:

- Визначення функції подібності.
- Встановлення відповідної нечіткої матриці подібності відповідно до функції подібності.
- Обчислення нечіткого відношення та використання плоского методу. Також включає випередження при знаходженні транзитивного закриття.
- Класифікація відповідно до надзвичайних порогів і отримання специфічного динамічного ефекту кластеризації.
- Ступені групуються разом у набір серій.
- Алгоритм аналізу зразків використовується для виявлення нападу з можливої послідовністю нападу.
- Встановлення відповідної нечіткої матриці подібності, відповідно до функції подібності.

- Обчислення нечіткого відношення та використання плоского методу. Також включає випередження при знаходженні транзитивного закриття.
- Класифікація відповідно до надзвичайних порогів і отримання специфічного динамічного ефекту кластеризації.
- Ступені групуються разом у набір кандидатських серій.
- Алгоритм аналізу зразків використовується для виявлення режиму нападу з можливою послідовністю.

Новий підхід до генерації нечітких правил став пропозицією, в якій кластери в шаблоні навчання встановлюються відповідно до техніки кластеризації нечіткого С-методу, відповідно до характеристик кожного шаблону та кластера [9].

Алгоритм кластеризації K-means припускає, що необхідні значення кластеризації відомі, однак фактично в аналізі безпеки значення k зазвичай невідомі. І вибір початкового центру кластеризації К-ознак набору правил є важливим. Міні-серія К-ознак використовується для розділення звичайного набору даних і набору даних атаки на кластери однакового розміру окремо, а центр кожного кластера використовується як індекс кластера. Існує метод вибору репрезентативних екземплярів з кожного кластера. Репрезентативність елемента пов'язана як з щільністю, так і з відстанню. Вища репрезентативність збільшує ймовірність того, що елемент буде обраний як репрезентативний . Після відбору кожному репрезентативному елементу присвоюється вага. Цей крок не тільки зменшує розмір вихідних даних, але й зберігає максимальну кількість інформації [1].

Алгоритм апріорі використовується для аналізу внутрішніх асоціацій правил безпеки фактів, оскільки він має значущість високої якості. Проблемою цього алгоритму є часте сканування бази даних транзакцій і непомірний набір додаткових параметрів очікування [10, 11]. Значення мінімальної підтримки та мінімальної довіри мають величезний вплив на результати виявлення. Алгоритм Апріорі був вперше запропонований [10].

Необхідно переконатися, що конфіденційні дані не витікають до тих, хто має зловмисні наміри. Через це існують конкретні цілі контролю доступу для різних користувачів. Існує 4 таких основних ролі користувача. Вони є постачальниками даних, збирачами даних, майнерами даних і особами, які приймають рішення [12]. Крім того, вкрай важливо переконатися, що конфіденційні

факти не потрапили до тих, хто має злі наміри. Завдяки цьому існують цілі контролю доступу для надзвичайних ролей споживачів безпеки мережі. Існує 4 таких основних ролі користувача. Вони є постачальниками інформації, збирачами інформації, добувачами інформації та особами, які приймають рішення [12]. Для постачальників даних мета контролю доступу полягає в тому, щоб ефективно контролювати кількість конфіденційних даних, які розкриваються іншим. Для досягнення цієї мети можна використовувати інструменти захисту, щоб обмежити доступ інших до їхньої інформації, просувати дані на аукціоні, щоб отримати достатню компенсацію за втрату конфіденційності, або фальсифікувати інформацію, щоб приховати свою справжню особу Для збирачів даних мета контролю доступу полягає в тому, щоб запустити корисні факти для майнерів фактів, не розкриваючи особи постачальників записів і конфіденційну статистику щодо них. Для досягнення цієї мети необхідно розробити правильні моделі конфіденційності для кількісної оцінки можливої втрати контролю доступу під час виняткових атак, а також застосувати стратегії анонімізації статистики [12]. Для майнерів даних мета збереження конфіденційності полягає в тому, щоб отримати правильні результати видобутку записів, зберігаючи при цьому конфіденційну статистику нерозкритою ні в рамках методу видобутку записів, ні в результатах видобутку. Для досягнення цієї мети може бути обраний правильний підхід для регулювання інформації до виконання алгоритмів позитивного видобутку. Крім того, протоколи стабільного обчислення можуть бути використані для забезпечення безпеки приватних даних і конфіденційної статистики, що міститься в навченій моделі. Для осіб, які приймають рішення, мета контролю доступу полягає в тому, щоб зробити правильний висновок, наближаючи достовірність результатів аналізу фактів, які вони отримують. Щоб досягти цієї мети, можна використовувати методи походження, щоб підказати повернуті записи отриманих фактів або створити класифікатор [12].

Було створено дві системи одна для виявлення нападів на основі аномалій, а інша для виявлення нападів на основі зловживання. Ці системи мали приблизно 4500 записів. Вхідні дані поділялися на набір даних (75%) для навчання нейронної мережі та тестовий набір даних (25%) для навченої нейронної мережі. Першою метою методу було спрощення фактів для обробки. Спрощення

передбачає відмову від познак, які є менш корисними. Перевага полягає в тому, що позбавлення від ознак зменшує розміри даних, що обробляються для підвищення продуктивності нейронної мережі. Недоліком може бути той факт, що якщо ключові атрибути будуть випадково видалені, точність виявлення вторгнень знизиться. Усі постійні ознаки було видалено, а ознаки, які передають найбільший відсоток дисперсії, було видалено додатково. Сценарії «R» також перевіряли на задовільні характеристики на основі дисперсії всередині зразків. Не чіпали атрибути , які не вносять навіть 1 відсотка сукупної варіації в набір фактів. Для зношування виявлення вторгнень для атак, які переважно базуються на аномаліях, і атак, які переважно базуються на неправильному використанні, було створено два файли: набір даних аномалії та набір даних неправильного використання. У наборі інформації про виявлення аномалії клас або змінна передбачення були або нормальними, що представляло повсякденний випадок, або атаку. Набір фактів виявлення неправильного використання мав змінну категорії «Звичайна» або «Назва нападу», яка представляє певний вид нападу, наприклад Smurf, NMap, Rootkit тощо.

Очищення даних було досягнуто для файлів, що складаються з аномалії набору даних і неправильного використання набору даних. Використання Weka для отримання вибору атрибутів аномалії набору даних і вибору атрибутів неправильного використання набору даних означало набагато менші атрибути, які сприяли прискоренню NN. Пакет «neuralnet» доступний у R і має відкритий код. Його почали використовувати для IDS та аналізу на основі нейронної мережі. Пакетна угода надала можливості як для створення нейронної мережі, так і для проведення класифікації.

В цій роботі було розглянуто понад 4500 випадків атак. Було обрано 10 типів атак, включаючи атаки Neptune, NMap, PortSweep, Satan, Smurf, BufferOverflow, FTPWrite, GuessPassword, Back і Rootkit. В процесі виявлення атак було реалізовано виявлення вторгнень на основі аномалій. Для виявлення атак реалізовано виявлення вторгнень на основі зловживання, щоб запропонувати матрицю плутанини, точність класифікації, час, витрачений на впровадження, і споживання ресурсів. Було створено класифікацію серед 10 нападів і звичайного випадку. Для деяких атак система виконувала виявлення зловживань.

Атака з виявленням аномалій У наборі результатів, наведених у таблиці 1, є деталі точності під час виявлення атаки з аномаліями, час виконання та обсяг споживання пам'яті [13]. Існує класифікація виникнення атаки або звичайного випадку в цьому процесі. Значення, зазначені в таблиці 1, вказують на кількість записів. Значення координат (Attack, Attack) і значення (Attack, Normal) координати дорівнюють 389 і 6 відповідно. Значення координати (Атака, Атака) вище, ніж координата (Атака, Норм), як показано в таблиці 1. Це означає наявність атаки. Точність класифікації висока – 99,57 відсотка. Мінімальний час виконання становить 3,9979 секунди. Використання пам'яті є мінімальним і становить 2191,311 Kbit, як показано в таблиці 2.

Таблиця 1. Виявлення аномальних атак

	Axis2	
Axis1	Anomaly attack	Normal
Anomaly Attack	389	6
Normal	3	763

Таблиця 2. Результати виявлених атак

Точність	99,57%
Час виконання	3.9979 с
Використання пам'яті	2189.311 Kbs

Атака виявлення неправильного використання В наборі результатів, показаних у таблиці 3, є деталі точності виявлення атаки неправильного використання, час виконання та обсяг споживання пам'яті [13]. Існує класифікація між 10 типами атак і нормальними випадками. У таблиці 3, як у випадку виявлення аномалії, є 2 осі, тобто вісь 1 і вісь 2. Значення, згадані в таблиці 3, вказують на кількість записів. Значення (Назад, Назад), (Buffer Overflow, Buffer Overflow), (Guess Password, Guess Password), (Neptune, Neptune), (Nap, NMap), (Port Sweep, Port Sweep), (Satan, Satan) і (Smurf, Smurf) координати є найвищими серед усіх значень рядка (значення 67, 4, 12, 57, 75, 748, 63,

59 і 58 відповідно). Це вказує на легшу класифікацію та більшу ймовірність нападу. Координати (FTP Write, FTP Write), (Rootkit, Rootkit) значно низькі (значення 1 і 0 відповідно). Значення координат (FTP Write, Normal) і (Rootkit, Normal) дорівнюють 0 і 1 відповідно, що не є найвищим значенням рядка. Таким чином, нормальний випадок, як показано в таблиці 3, відсутній. Як показано в таблиці 4, точність класифікації висока і становить 98,1%. Час виконання 48,9282 секунди вищий, ніж у попередньому випадку, через більший обсяг інформації, однак все ще низький для обробленої інформації. Використання пам'яті 2988,14 Кб, збільшене завдяки більшій інформації, але не дуже високе.

Таблиця 3. Атака з виявленням неправомірного використання — відомості про записи

Axis1 / Axis2	Повернення	Bufer Overflow	FTP	Guess Pass.	Neptune	NMap	Normal	Port Sweep	Rootkit	Satan	Smurf
Повернення	67	0	0	0	0	0	0	0	0	0	0
Bufer Overflow	0	4	0	0	0	1	1	0	0	0	0
FTP	0	0	1	1	0	0	0	0	0	0	0
Guess Password	1	0	0	12	0	1	0	0	0	0	0
Neptune	0	0	0	0	57	0	0	1	0	0	3
NMap	0	0	0	0	0	75	0	0	0	0	0
Normal	0	0	0	0	0	0	748	0	1	0	0
Port Sweep	0	0	0	0	0	0	0	63	0	0	1
Rootkit	0	0	1	0	3	0	1	3	0	0	1
Satan	0	0	0	0	0	3	0	1	1	59	1
Smurf	0	0	0	0	0	0	0	0	0	0	58

Таблиця 4. Результати атаки виявлення неправильного використання

Точність	98.10%
Час використання	48.9282 с
Використання пам`яті	2988.14 Kbs

Висновки та пропозиції. У моделі прогнозування безпеки мережі споживання пам'яті було низьким, час, витрачений на виявлення атак, також низький. Також висока точність виявлення атак. Вищевказані методи, які використовуються для проектування моделі, також прості в проектуванні. Крім того, обчислення спрощуються завдяки використанню цієї моделі. Отже, використання нейронної мережі також є ефективним способом розробки моделі прогнозування безпеки мережі. Таким чином, використання нейронних мереж рекомендовано для розробки будь-якого типу моделі прогнозування безпеки мережі. Майбутні завдання полягають в розробці моделей, які виявлятимуть будь-які вторгнення ще точніше та швидше.

ЛІТЕРАТУРА

1. Qiuhua W Xiaoqin O Jiacheng Z 2019 A Classification algorithm based on data clustering and data reduction for intrusion detection system over big data KSII Trans. on internet and information systems 13 pp 3714-3732.

2. Jianliang M Haikun S Ling B 2009 The application on intrusion detection based on K-means cluster algorithm Int. Forum on Information Technology and Applications (Chengdu) pp. 150-152.

3. Aung Yi Y Myat M M 2018 Hybrid intrusion detection system using K-Means and classification and regression trees algorithms IEEE 16th Int. C. on Software Engineering Research, Management and Applications (SERA) (Kunming) pp 195-199.

4. Ravale U Marathe N Padiya P 2015 Feature selection based hybrid anomaly intrusion detection system using K-Means and RBF kernel function Procedia Computer Science pp 428-435.

5. Xiaoyi H 2020 Network security situation prediction based on grey relational analysis and support vector machine Algorithm Int. J. of network security 22 pp 177-182.

6. Xiao Y Wang H Xu W 2015 Parameter selection of gaussian kernel for one-class SVM IEEE Trans. on Cybernetics 45 pp 927-939.

7. Song C Ma K 2009 Design of intrusion detection system based on data mining algorithm Int. Conf. on signal processing systems (Singapore) pp 370-373.

8. Rochim A F Aziz M A Fauzi A 2019 Design log management system of computer network devices infrastructures based on ELK stack Int. C. on Electrical Engineering and Computer Science (ICECOS) (Batam Island) pp 338-342.

9. Jin C Ye Z Wang C, Yan L Wang R 2018 A network intrusion detection method based on hybrid rice optimization algorithm improved fuzzy C-means IEEE 4th Int. Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS) (Lviv) pp 47-52.

10. Han J W Kamber M P 2011 Data mining: concepts and techniques (3rd Edition. Elsevier Science).

11. Huang Z 2019 Research and implementation of intrusion detection based on host log". Proc. of the Int. Conf. on Big Data Engineering, pp 98–106.

12. Lei X Chunxiao J Jian W Jian Y, Yong R 2014 Information security in big data: privacy and data mining IEEE Access 2 pp 1149-1176.

13. Pratik M, Saunil D, Ravina D 2015 Intelligent Security Systems-Intrusion Detection System https://github.com/jgera/Network-Intrusion-Detection-System/blob/master/report.pdf.

REFERENCES

1. Qiuhua W Xiaoqin O Jiacheng Z 2019 A Classification algorithm based on data clustering and data reduction for intrusion detection system over big data KSII Trans. on internet and information systems 13 pp 3714-3732.

2. Jianliang M Haikun S Ling B 2009 The application on intrusion detection based on K-means cluster algorithm Int. Forum on Information Technology and Applications (Chengdu) pp. 150-152.

3. Aung Yi Y Myat M M 2018 Hybrid intrusion detection system using K-Means and classification and regression trees algorithms

IEEE 16th Int. C. on Software Engineering Research, Management and Applications (SERA) (Kunming) pp 195-199.

4. Ravale U Marathe N Padiya P 2015 Feature selection based hybrid anomaly intrusion detection system using K-Means and RBF kernel function Procedia Computer Science pp 428-435.

5. Xiaoyi H 2020 Network security situation prediction based on grey relational analysis and support vector machine Algorithm Int. J. of network security 22 pp 177-182.

6. Xiao Y Wang H Xu W 2015 Parameter selection of gaussian kernel for one-class SVM IEEE Trans. on Cybernetics 45 pp 927-939.

7. Song C Ma K 2009 Design of intrusion detection system based on data mining algorithm Int. Conf. on signal processing systems (Singapore) pp 370-373.

8. Rochim A F Aziz M A Fauzi A 2019 Design log management system of computer network devices infrastructures based on ELK stack Int. C. on Electrical Engineering and Computer Science (ICECOS) (Batam Island) pp 338-342.

9. Jin C Ye Z Wang C, Yan L Wang R 2018 A network intrusion detection method based on hybrid rice optimization algorithm improved fuzzy C-means IEEE 4th Int. Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS) (Lviv) pp 47-52.

10. Han J W Kamber M P 2011 Data mining: concepts and techniques (3rd Edition. Elsevier Science).

11. Huang Z 2019 Research and implementation of intrusion detection based on host log". Proc. of the Int. Conf. on Big Data Engineering, pp 98–106.

12. Lei X Chunxiao J Jian W Jian Y, Yong R 2014 Information security in big data: privacy and data mining IEEE Access 2 pp 1149-1176.

13. Pratik M, Saunil D, Ravina D 2015 Intelligent Security Systems-Intrusion Detection System https://github.com/jgera/Network-Intrusion-Detection-System/blob/master/report.pdf.

СТАТТЯ НАДІЙШЛА ДО РЕДАКЦІЇ 07.12.2022