DOI 10.31891/2307-5732-2024-345-6-25
УДК 621.391 160164

НІЧЕПОРУК АНДРІЙ
Хмельницький національний університет
https://orcid.org/0000-0002-7230-9475
e-mail: andrey.nicheporuk@gmail.com
ДАНЧУК СЕРГІЙ
Хмельницький національний університет
https://orcid.org/0000-0001-7854-4556
e-mail: sergey.danchuk.p@gmail.com
ІВАНЧЕНКО ОЛЕГ
Національний технічний університет «Дніпровська політехніка»
https://orcid.org/0000-0002-5921-5757
e-mail: vmsu12@gmail.com
ПОСОНСЬКИЙ СЕРГІЙ
Хмельницький національний університет
https://orcid.org/0000-0002-4697-7699
e-mail: p.s.f@ukr.net
АНДРЄЄВ ВЛАДИСЛАВ
Хмельницький національний університет
https://orcid.org/0009-0001-8302-6657
e-mail: vladvasabi@gmail.com

МЕТОД ВИЯВЛЕННЯ КІБЕРАТАК НА КАНАЛИ ЗВ'ЯЗКУ ІНФОРМАЦІЙНИХ СИСТЕМ

У статті розглядається новий підхід до виявлення кібератак у мережах TCP/IP, заснований на використанні спектральної кластеризації та технологій машинного навчання. Метод передбачає кілька етапів: збір та попередня обробка даних, виконання кластеризації для виділення груп схожих об'єктів, навчання алгоритму класифікації та його подальше тестування. Спектральна кластеризація застосовується для виявлення декількох типів атак, використовуючи різноманітні параметри мережевого трафіку для побудови матриці подібності. В ході дослідження було обрано набір ознак із мережевого трафіку, що включали в себе: кількість запитів до сервера за певний період, загальний обсяг переданого трафіку, унікальні IP-адреси, середній час відповіді сервера та кількість невдалих спроб автентифікації або з'єднання. Запропонований метод поєднує кластеризацію з алгоритмами машинного навчання, такими як Random Forest, J48 та Naive Bayes.

Ключові слова: кібератаки, комунікаційний канал, інформаційна система

NICHEPORUK ANDRII, DANCHUK SERHII
Khmelnytskyi National University
IVANCHENKO OLEG
Dnipro University of Technology
POSONSKIY SERHII, ANDREYEV VLADYSLAV
Khmelnytskyi National University

THE METHOD OF DETECTING CYBER ATTACKS ON COMMUNICATION CHANNELS OF INFORMATION SYSTEMS

In the modern digital age, where communication networks are integral to nearly every aspect of life and activity, the significance of detecting cyberattacks on these channels has become paramountAs a result, a new method for detecting cyberattacks in TCP/IP networks has been proposed, based on the use of spectral clustering and machine learning technologies. The method involves several stages: data collection and pre-processing, performing clustering to select groups of similar objects, training the classification algorithm and its subsequent testing. Spectral clustering is applied to detect DDoS attacks by using various network traffic parameters to construct a similarity matrix. Key metrics include: number of server requests over a period, total traffic transferred, unique IP addresses, average server response time, and number of failed authentication or connection attempts.

The technique combines clustering with machine learning algorithms such as Random Forest, J48 and Naive Bayes. During the training process, the data is divided into groups using spectral clustering, after which a separate classifier is created for each cluster. During anomaly detection, the test data is first classified using spectral clustering, which determines which cluster the sample belongs to, after which the Random Forest algorithm evaluates whether it is normal or abnormal. Experimental results show that the semi-supervised learning model proposed in this article achieves a fairly high accuracy rate. The effectiveness of the proposed approach is tested on new data sets that have not been used for training before. The proposed method shows significant potential for accurate detection of DDoS attacks and can be effectively applied in various cyber security scenarios to protect communication channels from unwanted interference.

Keywords: cyberattack, communication channel, information system

Вступ

В сучасному цифровому світі, де мережі зв'язку є необхідним складником практично всіх аспектів життя та діяльності, важливість виявлення кібератак на канали зв'язку набуває критичного значення. Зростання залежності від технологій і поширення Інтернету призводять до збільшення обсягів цифрового обміну інформацією, але водночас і до збільшення кількості загроз та атак, спрямованих на порушення цілісності, конфіденційності та доступності цих каналів.

Кібератаки на канали зв'язку можуть призводити до серйозних наслідків, включаючи втрати конфіденційної інформації, порушення приватності, фінансові збитки та навіть загрозу національній безпеці. Злоумисники, використовуючи різноманітні методи та атаки, намагаються використовувати слабкі місця в інфраструктурі зв'язку, щоб отримати несанкціонований доступ до систем та данних. Тому виявлення кібератак на канали зв'язку є необхідним елементом цифрової безпеки. Використання передових технологій та методів аналізу мережевої активності дозволяє вчасно розпізнавати та відвертати потенційні загрози, щоб забезпечити надійність та безпеку інфраструктури зв'язку. Важливість цього процесу стає особливою у контексті постійно зростаючої складності кіберзагроз та появи нових векторів атак, що вимагає постійного вдосконалення заходів безпеки та ефективних методів виявлення.

Огляд попередніх досліджень

Проблема кібератак на канали зв’язку інформаційних систем є серйозною загрозою для безпеки даних та функціонування будь-якої інформаційної системи. Кібератаки можуть впливати на різні типи комунікаційних каналів, такі як провідні, безпровідні, супутникові та інші. Тому на сьогоднішній день сучасні методи виявлення кібератак на канали зв'язку інформаційних систем включають в себе різні підходи для виявлення незвичайних або шкідливих активностей в мережі. Вони орієнтовані на виявлення аномалій, атак, вторгнень або вразливостей. Деякі з підходів включають системи виявлення вторгнень (IDS), що моніторять трафік на предмет незвичайних патернів або аномалій, методи використання сигнатур для виявлення відомих атак, використання алгоритмів машинного навчання для аналізу та виявлення аномалій, аналіз журналів подій для виявлення несподіваних змін чи неавторизованого доступу, а також використання інтелектуальних систем для розпізнавання незвичайної чи шкідливої поведінки у мережі. Розглянемо детальніше деякі методи виявлення кібератак на канали зв’язку.

Автори дослідження [1] запропонували ForkDec, систему виявлення майнінг атак на основі повної зв’язаної нейронної мережі з метою ефективного стримування зловмисників. Нейронна мережа містить загалом 100 нейронів (10 прихованих шарів і 10 нейронів на шар), вивчених на навчальному наборі, що містить близько 200 000 зразків форків. Набір даних, який використовується для навчання моделі, генерується симулятором майнінгу біткойнів, який був попередньо створений. Експеримент оцінки показує, що ForkDec має певну прикладну цінність і дослідницькі перспективи. Таким чином перевагами представленого рішення є висока точність завдяки використанню штучних нейронних мереж та великому набору даних для навчання. Однак, можливість перевантаження системи при роботі з великою кількістю нейронів та об'ємними даними може бути негативним аспектом.

У роботі [2] пропонується система виявлення VGA, заснована на вдосконаленому машинному навчанні. Зокрема, використовується методологія напівконтрольованого навчання, яка використовує гібридну комбінацію алгоритмів. Зокрема, використовується евристичний метод кластеризації, заснований на лінійній фрагментації групових класів. Методологія ELM використовується як алгоритм для отримання прихованих змінних за допомогою опуклої оптимізації. Проте слід відзначити, що використання таких складних методів може призвести до високого рівня обчислювальної складності та вимог до ресурсів системи.

Ще одним підходом автори [3] запропонували повною мірою використати переваги глибокого навчання з підкріпленням у прийнятті рішень і створили систему навчання зі здатністю до постійного навчання. Зокрема, для розробки унікального механізму винагороди та навчання застосували промислову мережу управління та характеристики навчання глибокого підкріплення. Крім того, побудована промислова система виявлення аномалій управління на основі глибокого навчання з підкріпленням. Автори досліджували алгоритм на наборі даних промислового керування газопроводом Університету штату Міссісіпі. Експериментальні результати показали, що швидкість збіжності цієї моделі значно вища, ніж у традиційних методів глибокого навчання. Однак, висока складність реалізації таких систем та їхній обмежений застосунок може бути негативним фактором.

Автори [4] для боротьби з DDoS на канали звязку запропонували CyDDoS, інтегровану систему виявлення вторгнень (IDS), яка поєднує в собі набір алгоритмів розробки функцій із глибокою нейронною мережею. Вибір функцій ансамблю базується на п’яти класифікаторах машинного навчання, які використовуються для ідентифікації та виділення найбільш релевантних функцій, які використовуються в прогнозній моделі. Цей підхід покращує продуктивність моделі, обробляючи лише підмножину відповідних функцій, одночасно зменшуючи вимоги до обчислень. Оцінюємо продуктивність моделі на основі CICDDoS2019 набору даних, що складається зі звичайного трафіку та трафіку DDoS-атаки. Було розглянуто різні показники перевірки, такі як точність, F1-Score та збереження, щоб аргументувати ефективність запропонованої структури проти найсучасніших IDS. Перевагою запропонованого авторами рішення є висока продуктивність при обробці обмеженого набору функцій. Однак, обмежена гнучкість у виборі функцій та залежність від якості набору даних можуть обмежити застосування в реальних умовах.

Автори досліджень [5] запропонували XNBAD, нову систему виявлення аномалій поведінки мережі без нагляду. XNBAD інтегрує своєчасні стани хостів високого порядку в контексті динамічної взаємодії з моделями розмов між хостами для представлення поведінки. Стани високого порядку можуть краще узагальнити шаблони латентної взаємодії, але їх важко отримати безпосередньо. Таким чином, XNBAD використовує графову нейронну мережу (GNN) для автоматичного генерування ознак високого порядку із серій витягнутих базових. Ми оцінили ефективність виявлення XNBAD у загальнодоступному наборі даних ISCX-2012. Щоб повідомити про детальні та точні експериментальні результати, ретельно уточнили набір

даних перед оцінкою. Результати показують, що XNBAD ефективніше виявляв різні поведінки атак і значно перевершував існуючі репрезентативні методи, принаймні відносно покращуючи з точки зору загальної зваженої AUC. Однак, її залежність від точності побудови графових моделей та великі обчислювальні ресурси можуть бути негативними факторами.

Автори [7] запропонували захисний механізм безпеки в мережі для вирішення проблеми колапсу мережі, який може бути викликаний DDoS-атаками. Зокрема, на основі формулювання стохастичної динаміки черги зі стрибковим шумом представлено механізм, що характеризує поведінку черги на маршрутизаторах для стабілізації довжини черги при DDoS-атаках із постійною швидкістю. Застосовуючи теорію стохастичного керування для аналізу продуктивності динаміки черги під час DDoS-атак із постійною швидкістю, встановлюються деякі чіткі умови, за яких миттєва довжина черги збігається з будь-якою заданою ціллю на маршруті. Результати моделювання демонструють задоволення запропонованого механізму захисту з різким контрастом із сучасними схемами активного керування чергами (AQM). Механізм безпеки в мережі для стабілізації довжини черги при DDoS-атаках заснований на стохастичній динаміці черги. Він вирізняється здатністю до стабілізації черги під час DDoS-атак, але його ефективність може бути обмеженою в залежності від умов мережі та інших параметрів.

Таким чином проведений огляд відомих методів виявлення кібератак на канали зв'язку підтверджує, що сучасні методи виявлення кібератак на каналах зв'язку володіють значною ефективністю. Зокрема, вони здатні з досить високим рівнем ефеткивності виявляти потенційно небезпечні аномалії та атаки. Проте, наявність визначених недоліків, таких як обмежена гнучкість у роботі з різноманітними сценаріями та великі обчислювальні витрати, ставить під сумнів універсальність застосування цих методів. Таким чином, розробка нових методів та підходів до виявлення кібератак на канали зв'язку є досить актуальним завданням.

Метод виявлення кібератак на канали зв'язку на основі поєднання спектральної кластеризації та методів класифікації даних

Запропонований метод виявлення кібератак на канали зв'зку в TCP/IP мережах на основі спектральної кластеризації та методів машинного навчання включає в себе наступні кроки:

- збір і нормалізація даних;
- застосування спектральної кластеризації. Отримання кластерів;
- навчання класифікатора;
- тестування класифікатора.

Метод виявлення кібератак на канали зв'язку включає в себе аналіз каналів зв'язку а саме TCP/IP з'єднання. Для аналізу використовується метод кластерезації. Для виявлення DDoS атак [8] за допомогою спектральної кластеризації, важливо мати різноманітні дані мережевого трафіку, які використовувати для побудови матриці яка підходить для подальшої обробки за допомогою алгоритму кластеризації. У даному дослідженні для створення матриці схожості було відібрано такі ознаки: кількість запитів до сервера за певний час, загальний обсяг даних переданий усіма пакетами мережі, кількість унікальних IP-адрес, що взаємодіють з сервером, середній час, який сервер витрачає на обробку запитів, кількість невдалих спроб з'єднання чи авторизації.

Модель виявлення, яка пропонується у цій статті, ґрунтується на підході напівконтрольованого навчання, а саме спектральної кластеризації, яка в свою чергу і отримує нормалізовані дані [9]. Спектральна кластеризація використовує властивості спектральної теорії графів для групування точок даних, які взаємодіють між собою більше, ніж з іншими точками. Зібрані данні групуються в кластери для подальшого аналізу. До отриманого набору кластерів застосовується класифікатор який визначає чи є дані аномальними. Модель також включає в себе використання алгоритму випадкового лісу, J48, Naive Bayes. За допомогою даних класифікаторів здійснюється аналіз даних щодо їх аномальності.

У фазі навчання визначений набір даних S являє собою $(X i , Y i), i = 1, 2, ..., N$, де X і представляє N - вимірну матрицю, $Y i = \{0,1\}$, де 0 вказує на нормальний потік, а 1 - на аномальний потік. Під час навчання набір навчальних даних спочатку розділяється на k непересічних кластерів за допомогою спектральної кластеризації. Далі, для кожного кластера окремо, здійснюється навчання відповідного класифікатора. На етапі виявлення використовується метод спектральної кластеризації для визначення, до якого з k кластерів належить вибірка тестових даних. Далі, використовуючи відповідний класифікатор, що відповідає обраному кластеру, визначається, чи є вибірка тестових даних нормальною (тобто відповідає звичайному потоку даних) чи аномальною.

Після цього проводиться тестування на частині даних які раніше не використовувались в навчанні. Загальний алгоритм навчання класифікаторів представлено на рис. 1. Розглянемо детальніше етапи запропонованого методу.

Перш ніж будувати матрицю схожості для виконання спектральної кластиризації даних, важливо нормалізувати дані для того, щоб усунути можливий вплив різниці в шкалі ознак. Для цього було залучено міні-макс нормалізацію [7]. Цей метод перетворює значення таким чином, щоб вони потрапляли в діапазон між 0 та 1. Це може бути корисно, коли важливо зберегти відносні відстані між значеннями. Проте, цей метод може бути чутливим до викидів у даних.

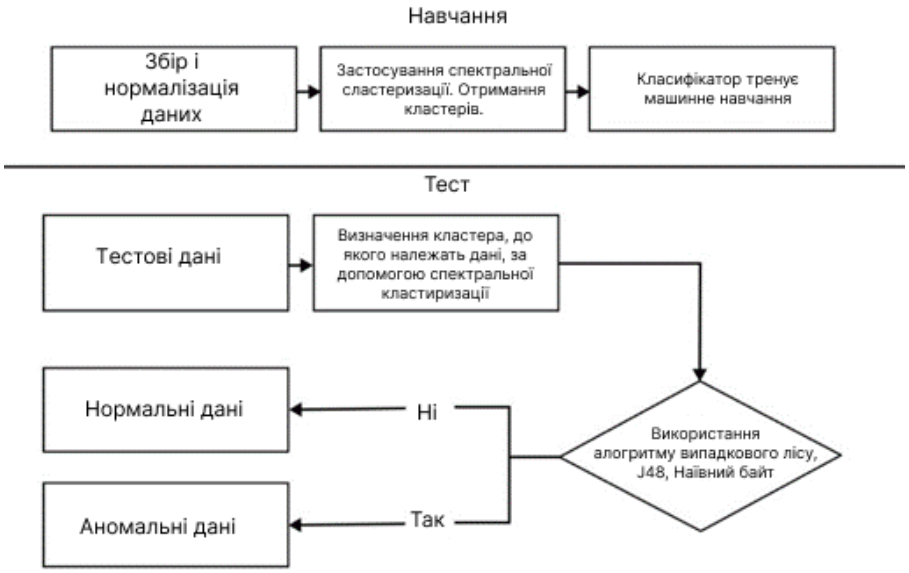

Рисунок 1. Алгоритм навчання класифікатора і його тестування

Міні-макс нормалізація – це метод масштабування даних до певного діапазону, зазвичай від 0 до 1. Процес нормалізації відбувається за допомогою наступних кроків:

– Визначення мінімального (min) та максимального (max) значень для кожного стовпця даних.

– Віднімання мінімального значення від кожного значення в стовпці (це центрує дані навколо 0).

– Розділення отриманого значення на різницю між максимальним та мінімальним значеннями в стовпці (це масштабує дані, щоб вони потрапляли в діапазон від 0 до 1).

Математично процес міні-макс нормалізації можна виразити наступним чином:

$$X_{\text{норм}} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

де $X_{\text{норм}}$ – нормалізоване значення, X – оригінальне значення, X_{min} – мінімальне значення в стовпці, X_{max} – максимальне значення в стовпці.

Цей процес забезпечує масштабування даних таким чином, щоб їх розподіл був у діапазоні від 0 до 1, зберігаючи відносні відстані між значеннями. Це дозволяє моделі легше збагатити особливості даних та покращити їх збіжність під час навчання.

Модель спектрального алгоритму кластеризації

В даній статті використовується алгоритм кластеризації на основі спектрального аналізу, який теоретично використовується для встановлення спектрів [10]. У порівнянні з традиційним алгоритмом кластеризації, спектральна кластеризація може краще розділити дані вибірки на кластери з високою схожістю незалежно від простору вибірки. Принцип роботи алгоритму спектральної кластеризації наступний. По-перше, дані набору вибірки перетворюються в матрицю схожості, яка відображає схожість між даними вибірки. Далі розв'язуються власні значення та власні вектори матриці. Завершально вибирається вектор ознак, який може досить добре кластеризувати дані. Цей алгоритм може збігатися до глобально оптимального рішення [11].

Дано x точок даних *z1,...,zn* і функцію подібності $f(||zi - zj||, \sigma)$, вагова матриця *H* визначається:

$$Hij = f(||zi - zj||, \sigma) \quad (2)$$

Схожість між двома точками даних залежить як від відстані між ними, так і від параметра масштабування σ; наприклад, гаусівська функція схожості:

$$f(||zi - zj||, \sigma) = e^{-||zi-zj||^2/2a^2} \quad (3)$$

Цей параметр масштабування часто використовується. Він визначає локальну структуру зв'язків між точками даних. Матриця ступенів D визначається наступним чином:

$$\text{Hii} = \sum_{j=1}^{n} \text{Lij} \quad (4)$$

та матриця Лапласа K визначається таким чином:

$$K = H - L \quad (5)$$

Власні значення та власні вектори матриці K потім використовуються для кластеризації даних; нормалізація Лапласа перед обчисленням спектрального розкладу призводить до більш збалансованих кластерів. Кількість кластерів k є обов'язковим вхідним параметром. Ми застосовуємо алгоритм спектральної кластеризації, запропонований]. Спочатку Лапласа нормалізується наступним чином:

$$K_{SYM} = C^{-1/2} K C^{1/2} \quad (6)$$

Нехай W – це матриця розміром n на k, стовпці якої є власними векторами, що відповідають k найменшим власним значенням Ksym. Потім рядки W нормалізуються, щоб отримати нову матрицю J:

$$J_{IJ} = \frac{W_{IJ}}{\sum_{J=1}^{N} W_{IJ}} \quad (7)$$

Тепер розглядаючи рядки J як колекцію n точок даних в R^k , застосовується алгоритм k-means для кластеризації даних.

Алгоритми випадкового лісу, J48, Naive Bayes

Випадковий ліс [12] ґрунтується на основній концепції ансамблювання для навчання серії дерев рішень та їх подальшого вдосконалення відповідно до характеристик кожного дерева. Під час навчання випадкового лісу відбувається випадковий вибір атрибутів для покращення відносної незалежності сформованих дерев рішень, що призводить до підвищення продуктивності. У випадку традиційного дерева рішень, коли кількість вузлів дорівнює n, вибір найкращого атрибута базується на всіх n атрибутах вузлів. У випадковому лісі кожен вузол дерева рішень базується на k випадково вибраних атрибутах, де k – це вирішальний параметр для ступеня випадковості.

З процесу навчання випадкових лісів видно, що даний алгоритм вносить лише невеликі модифікації до процесу ансамблювання, додаючи елемент випадковості до вибору атрибутів ознак на основі випадкових вибірок і узагальнює фінальну інтеграцію випадкових лісів. Це підвищення випадковості сприяє покращенню результатів. Таким чино, з огляду на високу продуктивність та відносно низьку обчислювальну складність, у даній роботі як алгоритм класифікатора, використано алгоритм випадкового лісу.

Алгоритм C4.5 [13] – це алгоритм класифікації, який створює дерева рішень на основі теорії інформації. Це розширення попереднього алгоритму ID3 Росса Квінлана, також відомого в Weka як J48 (де J означає Java). Дерева рішень, створені C4.5, використовуються для класифікації, і з цієї причини C4.5 часто називають статистичним класифікатором. Цей алгоритм будує дерева рішень на основі набору навчальних даних так само, як це робить алгоритм ID3, використовуючи концепцію інформаційної ентропії. Навчальні дані – це набір *S={s1, s2, ...}* уже класифікованих зразків. Кожна вибірка s_i складається з p-вимірного вектора *(x1, i, x2, i, …, xp, i)***,** де x_j представляє значення атрибутів або особливості відповідної вибірки, а також клас, до якого потрапляє вибірка. Щоб отримати найвищу точність класифікації, найкращий атрибут для розділення – це атрибут з найбільшою кількістю інформації.

Naive Bayes [14] це набір контрольованих алгоритмів навчання, заснованих на застосуванні теореми Байєса з «наївним» припущенням про умовну незалежність між кожною парою ознак, заданою значенням змінної класу. Класифікатори можуть бути надзвичайно швидкими порівняно з більш складними методами. Відокремлення розподілу умовних ознак класу означає, що кожен розподіл може бути незалежно оцінений як одновимірний розподіл. Це, у свою чергу, допомагає полегшити проблеми, що виникають через прокляття розмірності. Таким чином метою застосування цих класифікаторів є визначення чи дані в кластері аномальні.

Експериментальні дослідження

Експериментальна частина включала в себе дослідження ефективності запропнованого методу. Розглянемо детальніше набір даних, що використовувався в досліджені, досліджуванні атаки та метрики, за допомогою яких оцінювались результати експериментів.

У цьому експерименті використовується набір даних NSL-KDD[14]. NSL-KDD слугує ефективним еталоном для методів виявлення вторгнень. Завдяки розумній кількості записів у наборах тренування та тестування NSL-KDD використовуються для проведення експериментів на повному наборі даних без потреби випадкового відбору невеликої вибірки. Це забезпечує послідовні та порівнянні результати оцінки різних дослідницьких робіт. Для експерименту використовувались такі данні з набору: кількість запитів до сервера за певний час, загальний обсяг даних переданий усіма пакетами мережі, кількість унікальних IP-адрес, що взаємодіють з сервером, середній час, який сервер витрачає на обробку запитів, кількість невдалих спроб з'єднання чи авторизації.

У експерименті було досліджено 3 види атак: DDoS атака, Brute Force атака на авторизацію, Slowloris атака. DDoS атака характеризується кількість запитів до сервера за які за певний час зростають до великих значень, перевантажуючи сервер та спричиняючи зниження доступності. Загальний обсяг переданих даних в мережі також збільшується, оскільки зловмисники намагаються засмітити мережу шкідливим трафіком. Brute Force [15] атака збільшується кількість невдалих спроб з'єднання чи авторизації, оскільки зловмисники спробують надмірно велику кількість комбінацій для отримання доступу до системи. Кількість унікальних IP-адрес, що взаємодіють з сервером, може збільшитися, оскільки атакуючі можуть використовувати ботнети або різні проксі-сервери для приховування своєї ідентичності. Під час Slowloris атаки середній час, який сервер витрачає на обробку запитів, збільшується, оскільки зловмисники зберігають відкриті з'єднання, затримуючи їх завершення та перевантажуючи серверні ресурси. Кількість невдалих спроб з'єднання може також зростати, оскільки Slowloris спробує використовувати всі доступні з'єднання до сервера.

Навчання відбувалось за пирнципом k-cross validation, кожен експеримент використовує підмножину даних, що не використовувалась у попередніх експериментах. Ця модель використовується для тестування навченої моделі, де інші набори даних стають доступними. Навчання моделі проводиться на навчальному наборі, і потім модель тестується на k різних підвибірках цього набору. Результати кожного експерименту, тобто продуктивність моделі, усереднюються з усіх k експериментів.

Показники ефективності, які використовуються для оцінки [6] результатів експерименту, розраховувались на основі стандартної матриці плутанини (confusion matrix) та включали accuracy, TPR та FPR.

Таблиця 1

Результати дослідження

	DDoS			**Brute Force**			**Slowloris**		
	RF	**J48**	**NB**	**RF**	**J48**	**NB**	**RF**	**J48**	**NB**
Accuracy	95	89	92	94	90	91	94	90	94
TPR	92	90	88	91	88	89	93	88	90
FPR	2.3	2.8	2.2	2.0	2.3	2.4	2.5	2.4	2.5

У наведеній таблиці продуктивність методу спектральної кластеризації з алгоритмом випадкового лісу порівнюється з методами спектральної кластеризації з використанням алгоритмів J48 та Naive Bayes. Як показано в таблиці, алгоритм спектральної кластеризації, заснований на випадковому лісі, виявляється більш ефективним у порівнянні з J48 та Naive Bayes.

Експериментальні результати свідчать про те, що напівконтрольована модель навчання, запропонована в цій статті, досягає досить високого показника точності (accuracy), має низький рівень хибнопозитивних спрацювань. Цей метод краще підходить для виявлення атак на канали зв'язку, порівняно з іншими моделями виявлення.

Висновки

У ході проведеного дослідження була розроблена система навчання, що базується на методах спектральної кластеризації та алгоритмах випадкового лісу з метою підвищення ефективності виявлення DDoS-атак на каналах зв'язку. В роботі детально проаналізовано принципи функціонування алгоритму спектральної кластеризації та алгоритму випадкового лісу. Згідно з їхніми перевагами та принципом дії, вони були комбіновані з алгоритмами J48 та Naive Bayes для створення напівнавчальної моделі виявлення DDoS-атак.

Крім того, у роботі проведено порівняльний аналіз запропонованої напівнавчальної моделі з іншими існуючими методиками виявлення з метою перевірки її ефективності. Виявлено, що запропонована модель демонструє покращення в рівні виявлення DDoS-атак при зниженні рівня помилкових позитивних результатів. Таким чи*

Література

1. Zhaojie Wang, Qingzhe Lv, Zhaobo Lu, Yilei Wang, Shengjie Yue, ForkDec: Accurate Detection for Selfish Mining Attacks. Security and Communication Networks 5959698 (2021) 8 , doi:10.1155/2021/5959698

2. Wei Jiang, Machine Learning Methods to Detect Voltage Glitch Attacks on IoT/IIoT Infrastructures. Computational Intelligence and Neuroscience 6044071 (2022) 7, doi:10.1155/2022/6044071

3. Zhenze Liu, Chunyang Wang, Weiping Wang, Online Cyber-Attack Detection in the Industrial Control System: A Deep Reinforcement Learning Approach. Mathematical Problems in Engineering 2280871 (2022) 9, doi:10.1155/2022/2280871

4. Ivandro Ortet Lopes, Deqing Zou, Francis A Ruambo, Saeed Akbar, Bin Yuan, Towards Effective Detection of Recent DDoS Attacks: A Deep Learning Approach. Security and Communication Networks 5710028 (2021) 14, doi:10.1155/2021/5710028

5. Zhi-Quan Qin, Hong-Zuo Xu, Xing-Kong Ma, Yong-Jun Wang, Interaction Context-Aware Network Behavior Anomaly Detection for Discovering Unknown Attacks. Security and Communication Networks 3595304 (2022) 24, doi:10.1155/2022/3595304

6. H. Khoroshko V. Brailovskyi, M. Kapustian, M., Multi-criteria Assessment Of The Correctness Of Decision-making In Information Security Tasks, Computer Systems and Information Technologies, (2023) 81–86, doi.org:10.31891/csit-2023-4-11

7. Abdulwahid Al Abdulwahid, Detection of Middlebox-Based Attacks in Healthcare Internet of Things Using Multiple Machine Learning Models. Computational Intelligence and Neuroscience 2037954 (2022) 15, doi: 10.1155/2022/2037954

8. Oyewole, G.J., Thopil, G.A. Data clustering: application and trends. Artif Intell Rev 56 (2023) 6439–6475, doi:10.1007/s10462-022-10325-ytrends

9. Mittal, M., Kumar, K. & Behal, S. Deep learning approaches for detecting DDoS attacks: a systematic review. Soft Comput 27 (2023) 13039–13075 doi:10.1007/s00500-021-06608-1

10. Mizutani, T. Improved analysis of spectral algorithm for clustering. Optim Lett 15 (2021) 1303–1325 doi:10.1007/s11590-020-01639-3

11. S. Gopal Krishna Patro, Kishore Kumar Sahu, Normalization: A Preprocessing Stage, 4, 2015, doi:10.48550/arXiv.1503.06462

12. Lele Fu, Pengfei Lin, Athanasios V. Vasilakos, Shiping Wang, An overview of recent multi-view clustering. Neurocomputing 402 (2020) 148-161, doi:10.1016/j.neucom.2020.02.104

13. Sekulić A, Kilibarda M, Heuvelink GBM, Nikolić M, Bajat B. Random Forest Spatial Interpolation. Remote Sensing, 12 (2020) 1687-1697, doi:10.3390/rs12101687

14. Nagesh Tambake, Bhagyesh Deshmukh, Abhishek Patange, Development of a low cost data acquisition system and training of J48 algorithm for classifying faults in cutting tool. Materials Today: Proceedings 72 Part 3 (2023) 1061-1067, doi:10.1016/j.matpr.2022.09.163tool

15. Andrew J Wilson, Ben S Lakeland, Tom J Wilson, Tim Naylor, A naive Bayes classifier for identifying Class II. YSO 10 (2023) 23-46, doi:10.1093/mnras/stad301

References

1. Zhaojie Wang, Qingzhe Lv, Zhaobo Lu, Yilei Wang, Shengjie Yue, ForkDec: Accurate Detection for Selfish Mining Attacks. Security and Communication Networks 5959698 (2021) 8 , doi:10.1155/2021/5959698

2. Wei Jiang, Machine Learning Methods to Detect Voltage Glitch Attacks on IoT/IIoT Infrastructures. Computational Intelligence and Neuroscience 6044071 (2022) 7, doi:10.1155/2022/6044071

3. Zhenze Liu, Chunyang Wang, Weiping Wang, Online Cyber-Attack Detection in the Industrial Control System: A Deep Reinforcement Learning Approach. Mathematical Problems in Engineering 2280871 (2022) 9, doi:10.1155/2022/2280871

4. Ivandro Ortet Lopes, Deqing Zou, Francis A Ruambo, Saeed Akbar, Bin Yuan, Towards Effective Detection of Recent DDoS Attacks: A Deep Learning Approach. Security and Communication Networks 5710028 (2021) 14, doi:10.1155/2021/5710028

5. Zhi-Quan Qin, Hong-Zuo Xu, Xing-Kong Ma, Yong-Jun Wang, Interaction Context-Aware Network Behavior Anomaly Detection for Discovering Unknown Attacks. Security and Communication Networks 3595304 (2022) 24, doi:10.1155/2022/3595304

6. H. Khoroshko V. Brailovskyi, M. Kapustian, M., Multi-criteria Assessment Of The Correctness Of Decision-making In Information Security Tasks, Computer Systems and Information Technologies, (2023) 81–86, doi.org:10.31891/csit-2023-4-11

7. Abdulwahid Al Abdulwahid, Detection of Middlebox-Based Attacks in Healthcare Internet of Things Using Multiple Machine Learning Models. Computational Intelligence and Neuroscience 2037954 (2022) 15, doi: 10.1155/2022/2037954

8. Oyewole, G.J., Thopil, G.A. Data clustering: application and trends. Artif Intell Rev 56 (2023) 6439–6475, doi:10.1007/s10462-022-10325-ytrends

9. Mittal, M., Kumar, K. & Behal, S. Deep learning approaches for detecting DDoS attacks: a systematic review. Soft Comput 27 (2023) 13039–13075 doi:10.1007/s00500-021-06608-1

10. Mizutani, T. Improved analysis of spectral algorithm for clustering. Optim Lett 15 (2021) 1303–1325 doi:10.1007/s11590-020-01639-3

11. S. Gopal Krishna Patro, Kishore Kumar Sahu, Normalization: A Preprocessing Stage, 4, 2015, doi:10.48550/arXiv.1503.06462

12. Lele Fu, Pengfei Lin, Athanasios V. Vasilakos, Shiping Wang, An overview of recent multi-view clustering. Neurocomputing 402 (2020) 148-161, doi:10.1016/j.neucom.2020.02.104

13. Sekulić A, Kilibarda M, Heuvelink GBM, Nikolić M, Bajat B. Random Forest Spatial Interpolation. Remote Sensing, 12 (2020) 1687-1697, doi:10.3390/rs12101687

14. Nagesh Tambake, Bhagyesh Deshmukh, Abhishek Patange, Development of a low cost data acquisition system and training of J48 algorithm for classifying faults in cutting tool. Materials Today: Proceedings 72 Part 3 (2023) 1061-1067, doi:10.1016/j.matpr.2022.09.163tool

15. Andrew J Wilson, Ben S Lakeland, Tom J Wilson, Tim Naylor, A naive Bayes classifier for identifying Class II. YSO 10 (2023) 23-46, doi:10.1093/mnras/stad301