УДК 004.056:004.8

МЕТОДИ ВИЯВЛЕННЯ АНОМАЛЬНОЇ МЕРЕЖЕВОЇ АКТИВНОСТІ В КОРПОРАТИВНИХ МЕРЕЖАХ НА ОСНОВІ МАШИННОГО НАВЧАННЯ БЕЗ ВЧИТЕЛЯ

Д. Д. Невмержицька[1,а], Ю. В. Наконечна[1,б]

[1] Навчально-науковий Фізико-технічний інститут

Анотація

Об'єм мережевого трафіку, постійна загроза атаки та постійний розвиток арсеналу нападників знижує ефективність традиційних методів виявлення аномального трафіку в мережі, таких як сигнатурний метод. Запропоновано підхід на базі машинного навчання без вчителя використовуючи алгоритм $IsolationForest$ і протестовано на датасеті $UNSW-NB15$. У результаті аналізу з 82332 подій виявлено 4117 аномалій, включаючи потенційно невідомі типи атак. Отримана система орієнтована на моніторинг корпоративних мереж і раннє виявлення кіберінцидентів.

Ключові слова: машинне навчання, навчання без учителя, IDS-система, аномалії, корпоративна мережа, мережевий трафік, кібербезпека.

Вступ

Постійне збільшення кількості користувачів, активних задач і підключених пристроїв в корпоративних мережах ускладнює убезпечення мережі від небажанного впливу. Особливо гостро стоїть проблема виявлення аномалій, що супроводжують зловмисні дії та/або технічні збої.

Традиційні підходи на базі сигнатурного методу малоефективні проти атак невідомого типу, що зумовлює пошук більш гнучких і автономних рішень. Методи машинного навчання без учителя не потребують попереднього розмічення данних та дозволяють динамічно адаптуватись для пошуку нових аномалій. Задача роботи формулюється як детекція аномалій у мережевому трафіку корпоративних мереж.

Основною метою є ідентифікація підозрілої мережевої активності, що може свідчити про кіберзагрози або порушення в роботі системи. В даній роботі використовується підхід навчання без учителя, що дозволяє виявляти аномалії без наявності розмічених даних.

1. Аномалії в мережевому трафіку

В контексті даної проблеми аномальна мережева активність може набувати вигляду нестандартного обсягу трафіку, нетипової поведінки вузлів або активність у нетипові часові проміжки. Аномалії вказують зловмисні дії або на випадкові. Зловмисні дії виникають у наслідок здійснення кібератак зловмисником, до цього відновиться DDoS-атаки, сканування портів, злам облікових записів тощо. Випадкові — це аномалії, які виникли з технічних причин, наприклад помилка конфігурації, збій в системі тощо. Головною складністю виявлення невідомих аномалій полягає в тому, що через відсутність їх шаблонних сигнатур традиційні методи можуть не ідентифікувати їх.

2. Застосування машинного навчання без учителя

Машинне навчання без учителя використовує алгоритми машинного навчання для аналізу та кластеризації наборів даних без міток [1]. Головною метою цього методу в кібербезпеці є виявлення прихованих закономірностей та відхилення в даних без використання попередньо розмічених даних [2].

Провідний німецький незалежний дослідницький інститут з безпеки $ITAV-TEST$ реєструє щодня понад 450000 нових загроз [3]. Таким чином можна зробити висновок, що даний метод для виявлення аномальної мережевої активності є критично важливим, оскільки сучасні атаки здебільшого нові або модифіковані, що унеможливлює їх вчасне внесення до навчальних вибірок [4]. Існує кілька основних методів машинного навчання без учителя:

1. Кластеризація — це поєднання схожих даних мережевого трафіку у кластери, що дозволяє виявляти аномалію, оскільки вона не належить жодному сформованому кластеру.
2. Метод виявлення аномалій — використовується для даних, які не були попередньо розміченими.

Алгоритми методів машинного навчання без учителя навчаються на нормальній поведінці даних мережевого трафіку і будь-яке різке відхилення від цього стану вважається аномалією, яка може виявитися

[а]dashadom78@gmail.com
[б]yulnak-ipt@lll.kpi.ua

Рис. 1. Архітектура системи виявлення аномалій

простим порушення роботи мережі або кібератакою.

3. Розробка та дослідження

Під час дослідження було розроблено систему виявлення аномалій в корпоративній мережі використовуючи методи машинного навчання без учителя.

3.1. Архітектура розробленої системи

Архітектура розробленої системи має модульну архітектуру (рис. 1), яка забезпечує поетапну обробку даних. До фундаментальної складової системи входить модуль попередньої обробки даних, модуль формування ознак, модуль машинного навчання та візуалізація результатів.

3.2. Попередня обробка даних

Важливим етапом системи є обробка вхідних даних для подальшого аналізу, оскільки неправильна обробка даних може дати хибні результати. Вхідний набір даних очищається від пропущених значень та здійснюється відбір числових ознак, які придатні для машинного навчання. Наступний етап — нормалізація даних, що дозволяє привести ознаки до єдиного масштабу. Для нормалізації використовувалась формула $z - score$:

$$Z = \frac{\chi - \mu}{\sigma}$$

де χ — початкове значення ознаки, μ — середнє значення по стовпчику, σ — стандартне відхилення.

Згідно формули нормалізація виконується шляхом стандартизації значень відносно середнього та стандартного відхилення, що дозволяє підвищити стабільність та точність роботи алгоритму машинного навчання. У процесі обробки даних використовується повний набір числових ознак як вектор ознак для моделі. Дані подаються у вигляді батчів, що дозволяє здійснювати потокову обробку мережевого трафіку.

Таким чином, загальний пайплайн обробки даних включає етапи збору даних, попередньої обробки, формування ознак, застосування моделі $IsolationForest$ та прийняття рішення щодо належності об'єкта до аномального або нормального класу.

3.3. Реалізація моделі машинного навчання без учителя

Для виявлення аномалій у мережевому трафіку застосовано $IsolationForest$. Даний метод базується на принципі ізоляції аномальних об'єктів шляхом випадкового розбиття простору ознак [5]. Особливістю підходу є те, що аномалії ізолюються значно швидше, ніж нормальні, оскільки вони значно вирізняються з основної маси даних. Таким чином, це дозволяє ефективно виявляти аномалії без використання розмічення даних.

Алгоритм $IsolationForest$ базується на принципі ізоляції аномальних спостережень у просторі ознак шляхом побудови ансамблю випадкових дерев рішень. Об'єкти, які потребують меншої кількості розбиттів для ізоляції, розглядаються як аномальні. Такий підхід дозволяє ефективно виявляти відхилення без використання попередньо розмічених даних.

3.4. Результати дослідження

Під час дослідження провели експериментальну перевірку розробленої системи на даних мережевого трафіку. Для експерименту було обрано датасет $UNSW - NB15$ [6], оскільки він є сучасним та визнаним у сфері виявлення мережевих кібератак. Цей датасет був створений дослідницькою групою $CyberRangeLab$ університету $UNSW, UniversityofNewSouthWales$, для моделювання реального мережевого середовища та сучасних типів кібератак [7, 8].

Рис. 2. Виявлено 4117 аномалій із 82332 подій

Результати показали, що реалізована система виявлення аномалій ефективно виявила аномальні події без попередньо розмічених даних (рис. 2). У результаті аналізу з 82332 подій було виявлено 4117 аномалій, що становить приблизно 5% від загального обсягу даних.

Отримані показники демонструють коректну роботу алгоритму $IsolationForest$, яка виділяє підозрілу активність та візуалізує статистику щодо кількості аномалій у всьому потоці даних (рис. 3 та рис. 4).

Оцінювання якості моделі показало наступні результати: $F1 - score = 0.0852$ та $Recall = 0.0465$. Низькі значення показників пояснюються високим дисбалансом класів у датасеті $UNSW - NB15$, а також тим, що алгоритм $IsolationForest$ не використовує інформацію про мітки під час навчання.

Незважаючи на це, модель демонструє здатність виявляти нетипові та підозрілі події, включаючи потенційно невідомі типи атак.

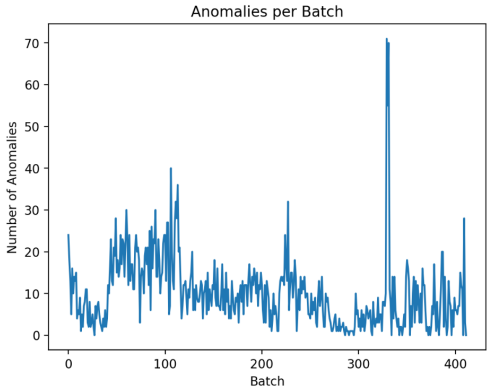

Рис. 3. Графік залежності кількості аномалій від номера пакету даних

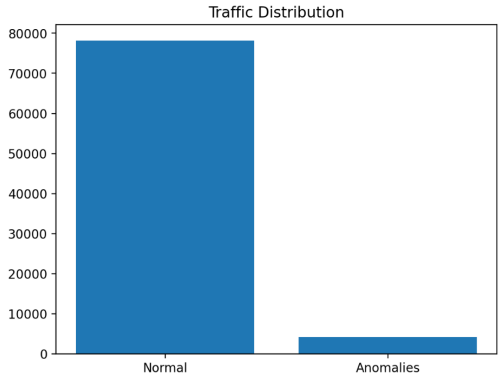

Рис. 4. Статистика оброблених даних

Висновки

В ході дослідження система виявлення аномальної мережевої активності в корпоративних мережах на основі методу машинного навчання без учителя продемонструвала ефективність в умовах постійної появи нових, невідомих та модифікованих кіберзагроз.

Застосування алгоритму *IsolationForest* забезпечило системі ефективне виявлення відхилень у мережевій поведінці шляхом побудови моделі нормального трафіку та ідентифікацій аномалій без попередньої розмітки даних. В ході експерименту було виявлено 4117 аномалій серед 82332 подій, що підтверджує здатність системи виявляти підозрілу активність та формувати статичні оцінки стану мережі.

Отримані результати свідчать про доцільність використання методів машинного навчання без учителя для задач моніторингу та аналізу виявлення загроз в корпоративних мережах. Розроблений підхід може бути використаний як основа для більш складних систем виявлення аномалій в кібербезпеці.

Перелік використаних джерел

1. *IBM.* What is unsupervised learning? — 05/03/2026. — URL: https://www.ibm.com/think/topics/unsupervised-learning.
2. *Rehman H.M.R.U. Liaquat S. G. M. e. a.* A systematic literature study of machine learning techniques based intrusion detection: datasets, models, challenges, and future directions. — 05/03/2026. — doi: 10.1186/s40537-025-01323-2. — URL: https://link.springer.com/article/10.1186/s40537-025-01323-2.
3. *AV-Test.* Malware. — URL: https://www.av-test.org/en/statistics/malware.
4. *Mohamed N.* Artificial intelligence and machine learning in cybersecurity: a deep dive into state-of-the-art techniques and future paradigms. — 05/03/2026. — doi: 10.1007/s10115-025-02429-y. — URL: https://link.springer.com/article/10.1007/s10115-025-02429-y.
5. *scikit-learn.* StandardScaler. — 05/03/2026. — URL: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html.
6. *Sydney U.* Dataset UNSW-NB15. — 05/03/2026. — URL: https://unsw-my.sharepoint.com/personal/z5025758_ad_unsw_edu_au/_layouts/15/onedrive.aspx?id=%2Fpersonal%2Fz5025758%5Fad%5Funsw%5Fedu%5Fau%2FDocuments%2FUNSW%2DNB15%20dataset%2FCSV%20Files&viewid=f8d1dec5%2Dcd5f%2D42ae%2D8b06%2D2fece580c74a.
7. *Hozouri A. M. A., M. E.* A comprehensive survey on intrusion detection systems with advances in machine learning, deep learning and emerging cybersecurity challenges. — 05/03/2026. — doi: 10.1007/s44163-025-00578-1. — URL: https://link.springer.com/article/10.1007/s44163-025-00578-1.
8. *Kareem L.* Survey of Machine Learning Techniques for Anomaly Detection in Cybersecurity. — 05/03/2026. — URL: https://www.researchgate.net/publication/390920977_Survey_of_Machine_Learning_Techniques_for_Anomaly_Detection_in_Cybersecurity.