DOI: https://doi.org/10.25140/2411-5363-2025-4(42)-266-276
УДК 004.85:004.056

Вадим Ігорович Штанько[1], Євгеній Васильович Нікітенко[2]

[1]аспірант кафедри комп'ютерних систем, мереж та кібербезпеки
Національний університет біоресурсів і природокористування України (Київ, Україна)
E-mail: vadym.shtanko@nubip.edu.ua. **ORCID:** https://orcid.org/0009-0001-4977-1450

[2]кандидат фізико-математичних наук, доцент кафедри комп'ютерних систем, мереж та кібербезпеки
Національний університет біоресурсів і природокористування України (Київ, Україна)
E-mail: ev.nikitenko@nubip.edu.ua. **ORCID:** https://orcid.org/0000-0002-9222-644X

ВИЗНАЧЕННЯ ЕФЕКТИВНОСТІ ДВОРІВНЕВОГО КЛАСИФІКАТОРА МЕРЕЖЕВИХ ПОТОКІВ

У статті розглянуто ефективність виявлення мережевих атак шляхом використання дворівневої класифікаційної моделі, яка спочатку визначає потенційно підозрілі потоки, а далі здійснює їх деталізовану багатокласову обробку. Такий підхід дозволяє покращити точність розпізнавання, зменшити кількість помилок і підвищити стабільність результатів. Проведено аналіз роботи моделі на різнорідних датасетах, включно з перевіркою її стійкості до зміни умов середовища та оцінкою обчислювальної продуктивності. Результати свідчать про доцільність впровадження дворівневого підходу у системи виявлення вторгнень.

Ключові слова: *класифікація трафіку; дворівнева модель; виявлення атак; Naïve Bayes; Random Forest; IDS; кібербезпека.*

Рис.: 2. Табл.: 3. Бібл.: 17.

Актуальність теми. У сучасних інформаційно-телекомунікаційних системах спостерігається зростання обсягів мережевого трафіку, що супроводжується ускладненням форм і механізмів кібератак. Серед них особливу загрозу становлять DDoS-атаки та інші методи навмисного перевантаження ресурсів, здатні призводити до відмови сервісів, втрати доступності, затримок у передаванні даних і фінансових збитків. З огляду на масштаби можливих наслідків, важливим завданням є своєчасне виявлення аномальної активності та визначення характеру загроз для зниження ризиків порушення стабільності функціонування мережевої інфраструктури. У цьому контексті системи виявлення вторгнень (IDS) залишаються ключовим засобом технічного реагування на подібні виклики та підтримки кіберзахисту [1].

Водночас сучасний мережевий трафік демонструє значну варіативність і динаміку, а його поведінкова структура стає помітно складнішою. Це створює суттєві труднощі для традиційних IDS-підходів, що спираються на сигнатури або жорстко визначені правила. У багатьох ситуаціях такі системи виявляються недостатньо гнучкими, особливо коли йдеться про атаки, які змінюють свої ознаки чи намагаються імітувати легітимний трафік. За цих умов виникає потреба у вивченні більш адаптивних, інтелектуально орієнтованих підходів до класифікації, здатних працювати майже в реальному часі та зберігати ефективність у середовищі з високою мінливістю. Особливе значення набуває розвиток рішень, що забезпечують точніше виявлення різних типів атак, зокрема DDoS, із можливістю коригувати роботу відповідно до специфіки конкретного трафіку [1].

Аналіз останніх досліджень і публікацій. Ідея дворівневої класифікації в IDS не є новою, і її результативність уже підтверджена низкою робіт. Наприклад, Pajouh та ін. запропонували дворівневу модель виявлення аномалій, у якій окремо обробляються «більшість» та «меншість» класів атак. Такий підхід дав змогу суттєво підвищити виявлення рідкісних атак і зменшити кількість помилок порівняно з одноетапними схемами. Експерименти на сучасних наборах даних, зокрема UNSW-NB15, продемонстрували, що дворівнева модель перевершує більшість актуальних методів [2]. Подібний напрям простежується і в роботі Azzaoui та Boukhamla (2020), де запропоновано двоетапну гібридну IDS: спершу бінарне розмежування Normal/Attack, а далі — визначення конкретного типу атаки. Система була протестована на наборах CICIDS2017 і NSL-KDD та показала високі показники точності й виявлення при мінімальному рівні хибних спрацювань [3], що додатково підтверджує ефективність дворівневих рішень для класифікації сучасних кіберзагроз.

Серед новіших прикладів варто згадати роботу Zhang та ін. (2023), де представлено дворівневу IDS-модель для IoT-мереж, побудовану на комбінації методів машинного і глибинного навчання. На першому етапі швидкий класифікатор LightGBM розмежовував нормальний і аномальний трафік, а на другому згорткова нейромережа (CNN) виконувала детальну багатокласову класифікацію виявлених атак. Гібридна модель продемонструвала дуже високу ефективність: точність другого етапу сягнула близько 99,9%. У цілому система показала кращу стійкість до дисбалансу вибірки та перевершила інші сучасні методи на великому наборі даних CSE-CIC-IDS2018 [1], доводячи переваги дворівневих підходів над окремими моделями машинного чи глибинного навчання.

На сьогодні вже існують дворівневі IDS, у яких на першому етапі Naïve Bayes розмежовує трафік на нормальний і підозрілий, а далі метод Elliptic Envelope аналізує саме потенційно загрозливі записи [4]. Після другого етапу така система змогла досягти точності 98,59% на наборі CIC-IDS2017 [4], що помітно перевищує результати однорівневих моделей на цьому ж наборі. Для порівняння, запропонований авторами двоетапний підхід продемонстрував дещо вищу ефективність, ніж інші класичні алгоритми (зокрема дерева рішень і випадковий ліс) на CIC-IDS2017 [5]. Загальний висновок цих робіт однозначний: дворівневі схеми нерідко мають перевагу над однорівневими за точністю виявлення мережевих атак [5].

Крім того, Pajouh та ін. у своїй дворівневій моделі відзначили суттєве поліпшення коефіцієнта виявлення атак за одночасного зниження рівня хибних спрацювань у порівнянні з попередніми методами [6]. Їхній підхід особливо вдало працював із рідкісними типами вторгнень, які раніше часто залишалися непоміченими (наприклад, атаки класів U2R та R2L) [6]. Подібний фокус на виявлення малорепрезентованих атак простежується і в інших роботах: наприклад, Huang та ін. використали генеративну модель Imbalanced GAN для синтезу недостатньо представлених атак, а подальша класифікація за допомогою CNN суттєво підвищила рівень їхнього виявлення на різних наборах даних [1]. Запропоновано й окремі методи ресемплінгу. Зокрема, Zhang та ін. створили комбінацію SMOTE з Gaussian Mixture (підхід SGM), і в поєднанні з CNN він показав відчутне покращення виявлення атак, що трапляються порівняно рідко [1]. У підсумку дворівневі методи аналізу трафіку можна вважати дієвим інструментом для IDS: поєднання швидкого фільтрування з точною багатокласовою моделлю дає змогу краще врівноважити повноту виявлення атак і рівень хибних спрацьовувань. У багатьох роботах наголошується, що такі системи знижують False Positive Rate і водночас підвищують Detection Rate для нетипових атак [3].

Такий дворівневий підхід зараз набуває особливої ваги, зважаючи на стрімкий розвиток Інтернету речей (IoT) та промислового Інтернету (IIoT), де мережевий трафік зростає не просто швидко, а фактично безперервно й нерівномірно. У сучасних IoT-середовищах формується колосальний обсяг даних, і ці дані потрібно обробляти досить оперативно, причому з вимогою до точності, яка часто виходить за межі можливостей традиційних IDS. За таких умов дворівневі моделі виявляються практичнішими: вони краще витримують розбалансованість вибірок, менш чутливі до коливань у структурі трафіку та здатні масштабуватися, коли обсяги даних починають зростати стрибками. Показовим є приклад дворівневої IoT-IDS на основі глибокої рекурентної нейромережі (RNN), яка демонструвала виявлення атак – зокрема DoS – майже в реальному часі, з доволі високою чутливістю [1]. Загалом такі системи стабільно перевершують одноетапні рішення в великих і різнорідних середовищах, забезпечуючи надійніший рівень захисту в умовах, коли кількість кіберзагроз не просто зростає, а й постійно змінює свої форми.

Виділення недосліджених частин загальної проблеми. Попри активний розвиток напрямку дворівневих систем виявлення вторгнень (IDS), більшість доступних досліджень і далі зосереджується переважно на тестуванні моделей у межах статичних, наперед анотованих наборів даних, таких як CICIDS2017, UNSW-NB15 чи CSE-CIC-IDS2018. Такі на-

бори дійсно зручні для порівняння алгоритмів між собою, однак мають низку суттєвих обмежень: вони не відображають повною мірою динамічний характер реального трафіку, зберігають фіксований класовий розподіл і часто мають структуру, певною мірою вже адаптовану під навчання моделей. Через це виникає ризик переоцінити ефективність запропонованих підходів: класифікатори, що демонструють високу точність на таких наборах, можуть поводитися нестабільно в реальних мережах, особливо там, де присутній доменний зсув або з'являються нові типи трафіку [7; 8].

Ще однією помітною проблемою є відсутність послідовного порівняння дворівневої архітектури з її окремими етапами. У значній частині публікацій подаються лише підсумкові результати, однак не аналізується, як саме кожен компонент — скажімо, бінарний детектор або багатокласовий класифікатор — впливає на загальний показник якості. Подібне порівняння могло б окреслити внесок кожного рівня окремо, але його часто або ігнорують, або згадують лише побіжно [9; 10].

До цього додається ще одна системна прогалина: більшість робіт не включає тестування дворівневих IDS на потоковому чи неанотованому трафіку. Навіть ті дослідження, що позиціонують свої рішення як придатні для роботи в реальному часі або для IoT-середовищ, зазвичай покладаються на штучно згенеровані дані або обмежуються відтворенням заздалегідь зафіксованих сесій. У результаті між моделлю та її потенційним середовищем застосування формується відчутний розрив [11; 12].

Постановка завдання. Попри помітну увагу до розробки засобів виявлення вторгнень, задача побудови багатокласової класифікації атак в один етап і далі лишається непростою. Моделі, які мають не лише зафіксувати сам факт загрози, а й одразу визначити її тип, стикаються з цілою низкою обмежень: мережевий трафік надто неоднорідний, ознаки атак помітно змінюються залежно від контексту, а сама поведінка атак часто коригується або маскується під впливом середовища. У реальних умовах функціонування мереж додатково присутні численні фонові процеси, паралельна робота сервісів та постійні коливання параметрів потоків. Усе це істотно знижує ефективність загальних підходів і ускладнює створення стійких багатокласових класифікаторів [2].

За таких умов доцільним видається застосування підходів, які дають змогу зменшити складність задачі шляхом її поетапного розв'язання. Дворівнева модель у цьому контексті постає як конструктивна альтернатива: на першому рівні відбувається виокремлення підозрілих потоків за допомогою бінарної класифікації, що формує основу для більш стабільного і спеціалізованого аналізу на наступному етапі. На другому рівні можуть застосовуватися методи, орієнтовані на розмежування конкретних типів атак в умовах меншої варіативності. Така структурна організація відкриває можливість для підвищення точності класифікації, зменшення кількості хибних рішень та покращення адаптивності системи до особливостей мережевого середовища [2].

Метою статті є експериментальне підтвердження переваг дворівневого класифікатора над традиційними однорівневими підходами у задачі класифікації мережевого трафіку. Передусім йдеться про демонстрацію того, що послідовне виконання двох етапів — бінарного виявлення аномалій із подальшою багатокласовою класифікацією — дає вищі показники якості, ніж робота кожного з цих етапів окремо. Для досягнення мети необхідно:

– здійснити збір і класифікацію трафіку за допомогою дворівневої моделі, де на першому рівні використовується Naïve Bayes, а на другому – Random Forest;

– виміряти основні метрики (accuracy, precision, recall, F1) як для кожного рівня окремо, так і для системи загалом на єдиному наборі даних;

– порівняти результати та оцінити, наскільки саме дворівнева схема підвищує загальну ефективність класифікації, а також визначити, у чому полягає виграш (наприклад, у зменшенні кількості хибнопозитивних спрацювань чи у кращому розпізнаванні окремих типів атак).

У підсумку мета полягає в обґрунтуванні доцільності застосування дворівневого підходу шляхом кількісного аналізу його ефективності на прикладі реальних даних. Очікується, що отримані результати покажуть: дворівнева архітектура не просто поєднує дві моделі, а й забезпечує відчутне покращення виявлення загроз у трафіку порівняно з використанням лише одного класифікаційного рівня.

Методика дослідження. Для експериментальної перевірки було побудовано дворівневу систему класифікації мережевих потоків. На першому рівні (бінарному класифікаторі) кожен потік відноситься або до класу «нормативний трафік», або до класу «потенційна загроза». Для цього етапу застосовано модель машинного навчання Naïve Bayes, оскільки саме вона показує стабільно високі результати у подібних задачах бінарного поділу [13]. Другий рівень — багатокласовий класифікатор — працює вже тільки з потоками, позначеними першим рівнем як «загроза», і розподіляє їх за конкретними типами атак. У межах цього дослідження враховувалися чотири види DoS-атак: Hulk, Slowhttptest, Slowloris, TCP-Flood, а також клас «нормативний». Додавання нормального трафіку на другий рівень дає змогу скоригувати можливі помилки першого етапу: якщо деякі безпечні потоки були хибно віднесені до підозрілих, другий класифікатор може повернути їх у нормальний клас.

Щоб формалізувати загальну логіку класифікації, запишемо її у вигляді математичних позначень.

Нехай $\boldsymbol{x} \in \boldsymbol{R^d}$ — вектор ознак мережевого потоку. Тоді функція, яка виявляє, чи x є атакою (бінарна класифікація), може приймати наступні значення:

$$\boldsymbol{f_b}(\boldsymbol{x}) \in \{\boldsymbol{C}_{\text{normative}}, \boldsymbol{C_{threat}}\}, \tag{1}$$

де $\boldsymbol{C}_{\text{normative}}$ — клас нормальний трафік, а $\boldsymbol{C}_{\text{threat}}$ — зловмисний.

Позначимо сукупність відомих класів даних як вектор:

$$\mathrm{c}_{\boldsymbol{known}} = \{\boldsymbol{C}_{\text{normative}}\} \cup \{\boldsymbol{C}_{\text{hulk}}, \boldsymbol{C}_{\text{slowhttptest}}, \boldsymbol{C}_{\text{slowloris}}, \boldsymbol{C}_{\text{tcpFlood}}\} \tag{2}$$

де $\boldsymbol{C}_{\text{normative}}$ — клас «нормальний трафік», $\boldsymbol{C}_{\text{hulk}}$, $\boldsymbol{C}_{\text{slowhttptest}}$, $\boldsymbol{C}_{\text{slowloris}}$, $\boldsymbol{C}_{\text{tcpFlood}}$ — класи окремих атак, а $\boldsymbol{C}_{\text{unknown}}$ — клас, який описує випадок, коли модель не може визначити жоден з відомих класів.

Також приймемо, що функції багатокласової класифікації $\boldsymbol{f_{rf}}(\boldsymbol{x})$ та визначення загального результату $\boldsymbol{f_r}(\boldsymbol{x})$ будуть приймати наступні значення:

$$\boldsymbol{f_{rf}}(\boldsymbol{x}) \in \mathrm{c}_{\boldsymbol{known}} \cup \{\boldsymbol{C}_{\text{unknown}}\},\ \boldsymbol{f_r}(\boldsymbol{x}) \in \mathrm{c}_{\boldsymbol{known}} \cup \{\boldsymbol{C}_{\text{unknown}}\}. \tag{3}$$

Опишемо функцію, яка здійснює рішення при виявленні атак серед мережевих потоків.

$$\boldsymbol{f_b}(\boldsymbol{x}) = \begin{cases} \boldsymbol{C_{threat}},\ \boldsymbol{P}(\boldsymbol{C_{threat}}|\boldsymbol{x}) \geq \boldsymbol{T_b} \vee \boldsymbol{P}(\boldsymbol{C_{normative}}|\boldsymbol{x}) < \boldsymbol{T_b}, \\ \boldsymbol{C_{normative}},\ \boldsymbol{P}(\boldsymbol{C_{threat}}|\boldsymbol{x}) < \boldsymbol{T_b} \vee \boldsymbol{P}(\boldsymbol{C_{normative}}|\boldsymbol{x}) \geq \boldsymbol{T_b} \end{cases} \tag{4}$$

де $\boldsymbol{P}(\boldsymbol{C_i}|\boldsymbol{x})$ — ймовірність належності потоку до класу $\boldsymbol{C_i}$, а $\boldsymbol{T_b}$ — поріг спрацьовування бінарного класифікатора

Нехай $\boldsymbol{K} \in \boldsymbol{c_{known}}$, та $\boldsymbol{K} = \boldsymbol{arg\ max_k P(k \mid x)}$, де $P(k|x)$ – ймовірність належності x до класу k, тоді запишемо функцію багатокласової класифікації як

$$\boldsymbol{f_{rf}}(\boldsymbol{x}) = \begin{cases} \boldsymbol{K},\ \boldsymbol{P(K \mid x)} \geq \boldsymbol{T_{rf}} \\ \boldsymbol{C_{unknown}},\ \text{інакше} \end{cases} \tag{5}$$

Тоді загальна логічна функція дворівневого класифікатора буде мати такий вигляд:

$$\boldsymbol{f_r}(\boldsymbol{x}) = \begin{cases} \boldsymbol{C_{normative}},\ \boldsymbol{f_b}(\boldsymbol{x}) = \boldsymbol{C_{normative}} \\ \boldsymbol{f_{rf}}(\boldsymbol{x}),\ \boldsymbol{f_b}(\boldsymbol{x}) = \boldsymbol{C_{threat}} \end{cases} \tag{6}$$

а процес класифікації для кожного рівня — $\hat{y}_i = f_j(x_i)$, де $f_j \in \{f_b, f_{rf}, f_r\}$, з яких f_bвикористовується на першому етапі для бінарної класифікації, f_{rf} – на другому етапі для багатокласової, f_r— загальний класифікатор, який включає використання попередніх функцій.

Після класифікації для оцінки ефективності моделі використаємо наступні показники, які характеризують правильність класифікації даних для кожного класу (як для етапу бінарної, так і етапу багатокласової класифікації, а також для результату дворівневої класифікації).

$$TP_k = \sum_{i=1}^{N} \mathbb{1}_{\{y_i = C_k\}} \cdot \mathbb{1}_{\{\hat{y}_i = C_k\}} \quad (7)$$

$\mathbf{TP}_k$ — кількість об'єктів класу C_k, які модель правильно класифікувала як C_k.

$$FP_k = \sum_{i=1}^{N} \mathbb{1}_{\{y_i \neq C_k\}} \cdot \mathbb{1}_{\{\hat{y}_i = C_k\}} \quad (8)$$

$\mathbf{FP}_k$ — кількість об'єктів інших класів, які модель помилково віднесла до C_k.

$$FN_k = \sum_{i=1}^{N} \mathbb{1}_{\{y_i = C_k\}} \cdot \mathbb{1}_{\{\hat{y}_i \neq C_k\}} \quad (9)$$

$\mathbf{FN}_k$ — кількість об'єктів класу C_k, які модель помилково віднесла до інших класів.

$$TN_k = \sum_{i=1}^{N} \mathbb{1}_{\{y_i \neq C_k\}} \cdot \mathbb{1}_{\{\hat{y}_i \neq C_k\}} \quad (10)$$

$\mathbf{TN}_k$ — кількість об'єктів, що не належать до C_k, і були правильно класифіковані як не C_k.

Для всіх вищезазначених формул: y_i- істинна мітка для *i*-го прикладу, $\hat{y}_i$- передбачена мітка для *i*-го прикладу, C_k- цільовий клас, N – кількість класів.

$$Accuracy = \frac{\sum_{k=1}^{K} TP_k}{\sum_{k=1}^{K} (TP_k + FP_k + FN_k)} \quad (11)$$

Accuracy — показує частку правильно класифікованих прикладів серед усіх, використовується для загальної оцінки точності моделі [11; 14; 15; 16].

$$Precision = \frac{1}{K} \sum_{k=1}^{K} \frac{TP_k}{TP_k + FP_k} \quad (12)$$

Precision — показує, яка частка передбачень для певного класу була дійсно правильною, важлива при контролі хибнопозитивних спрацювань. У такому разі вона розраховується з макроусередненням, тобто як середнє арифметичне цієї метрики для усіх класів [11; 14; 15; 16].

$$Recall = \frac{1}{K} \sum_{k=1}^{K} \frac{TP_k}{TP_k + FN_k} \quad (13)$$

Recall — відображає, яку частку прикладів певного класу модель змогла правильно знайти, критична при мінімізації пропусків (особливо атак). Також розраховується з макроусередненням [11; 13; 14; 15].

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (14)$$

F1-score — є гармонійним середнім Precision і Recall, застосовується для балансування між чутливістю та точністю, особливо у випадках із дисбалансом класів [11; 14; 15; 16].

Результати дослідження. Експериментальна перевірка ефективності моделі проводилася окремо для кожного з рівнів класифікації та для системи загалом. Далі подано динаміку метрик класифікації за різних типів трафіку. Було перевірено роботу двох варіантів моделі — одна була навчена на публічному наборі даних USB-IDS-1, інша на наборі даних, зібраному в тій же віртуальній мережі, де проводилось тестування моделі. Для цього було здійснено серію класифікаційних запусків, де трафік оброблявся вікнами по 60 секунд (для якісного виявлення повільних DDoS-атак). У табл. 1 описано загальні показники тестових записів. Для всього зловмисного трафіку у віртуальній мережі додавались референтні мітки (y_i), які дозволили здійснити розрахунок метрик та оцінити ефективність роботи моделі.

Таблиця 1 – Кількісний опис мережевих даних, оброблених класифікаційними моделями

Дані	Загальна кількість оброблених потоків	Середня кількість потоків у вікні	Кількість вікон	Середня тривалість обробки одного вікна (секунд)
USB-IDS-1	63 952	789.53	81	0.1860
Власні дані	51 967	626.11	83	0.2434

Джерело: розроблено авторами.

Захоплення трафіку та його подальша класифікація виконувалися у віртуальному середовищі [17] у вигляді груп вікон, у кожній з яких генерувався нормативний трафік, що слугував фоновим «шумом». Одна з груп містила лише нормативний трафік і використовувалася як еталон. Ще чотири групи, крім нормального трафіку, включали по одній атаці з набору Hulk, Slowhttptest, Slowloris та TCPFlood. Окрема група вікон містила всі чотири атаки одночасно разом із нормативним трафіком. На рис. 1 подано значення основних метрик для кожного вікна на кожному рівні класифікації для моделі, навченої на наборі USB-IDS-1. Зі структури цих результатів видно, що перший рівень класифікації добре впорається з розпізнаванням як нормального трафіку, так і атак за наявності Hulk, а також за умов змішаних сценаріїв. Водночас для нормального трафіку та атак типу Slowloris, Slowhttptest і TCPFlood ефективність бінарного класифікатора помітно нижча. Частину цих помилок компенсує другий рівень багатокласової класифікації, що додатково підтверджує доцільність включення класу «нормативний» у другий етап. Наприклад, для атаки TCPFlood значення F1 на першому рівні тримається поблизу 0,2, тоді як після другого рівня загальна оцінка підвищується до понад 0,4.

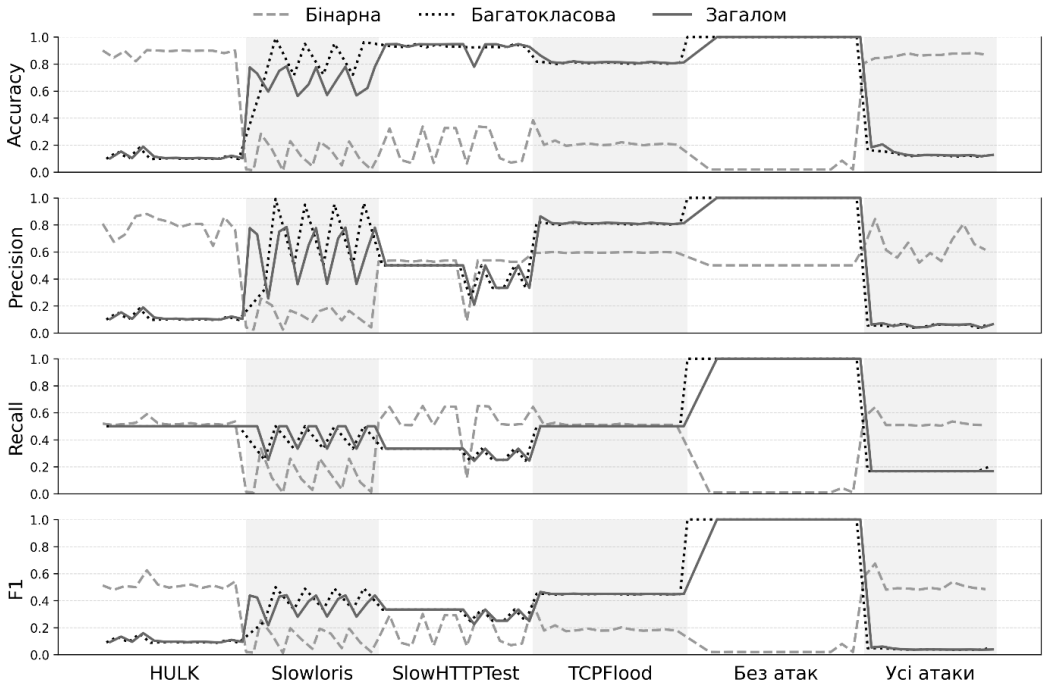

Рис. 1. Зміна метрик класифікації залежно від типу трафіку (модель навчена набором USB-IDS-1)

Джерело: розроблено авторами.

На рис. 2 наведені метрики для моделі, навченої вже на власноруч зібраних даних. З цих результатів можна бачити, що перший етап класифікації демонструє впевнене розпізнавання атак для всіх комбінацій трафіку, хоча оцінки дещо знижуються у випадку нормального трафіку та атак Slowhttptest і Slowloris.

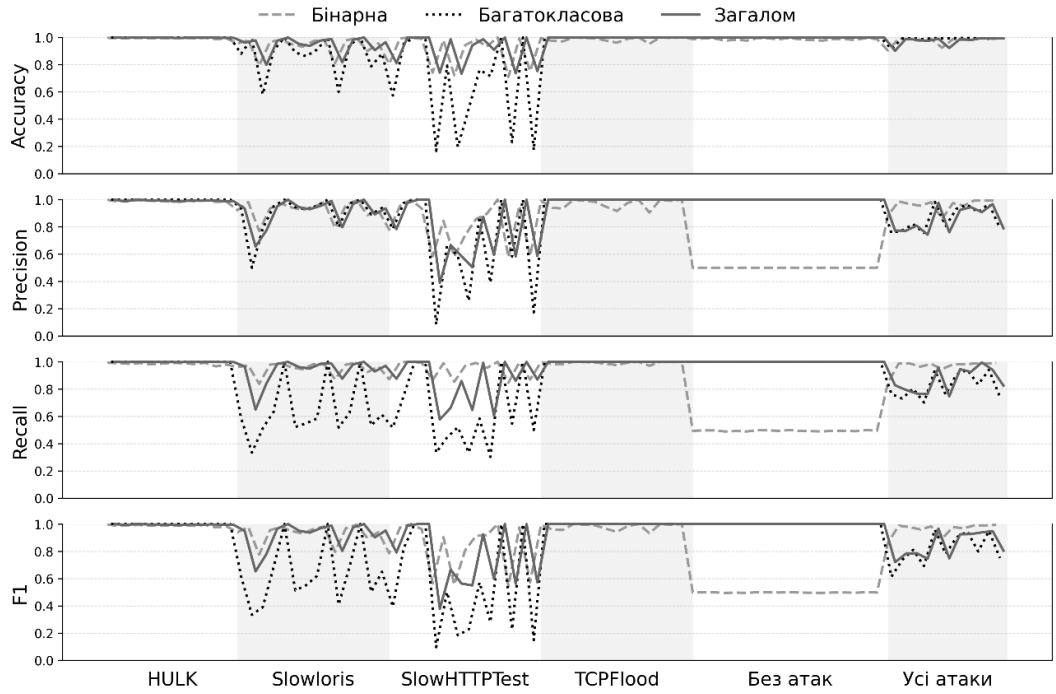

Рис. 2. Зміна метрик класифікації залежно від типу трафіку (модель навчена набором власних даних)

Джерело: розроблено авторами.

Загалом другий рівень мультикласової класифікації також підвищує ефективність моделі, однак його внесок тут менш виражений, ніж у випадку моделі, навченої на USB-IDS-1. Ймовірно, це пояснюється тим, що модель, тренована на власних даних, краще узгоджена з параметрами мережі, у якій відбувалося тестування, і тому перший рівень уже справляється з виявленням атак значно впевненіше. Відповідно, потреба в «корекції» помилок на другому рівні є меншою, ніж у попередньому випадку.

Для узагальнення результатів обчислено середні значення метрик за всіма вікнами класифікації, а також розраховано діапазон значень (min–max) у межах кожної серії. Розглянемо детально числові показники метрик (таблиця 2). В ній відображено середні значення метрик для усіх запусків, згруповані за моделлю (навченою на наборі USB-IDS-1 та наборі власних даних).

Таблиця 2 – Загальні значення метрик для дворівневої класифікаційної моделі

Метрики		**Accuracy**		**Precision**		**Recall**		**F1**	
Дані	**Етап класифікації**	**сер.**	**Діапазон**	**сер.**	**Діапазон**	**сер.**	**Діапазон**	**сер.**	**Діапазон**
Власні	Двійкова	0,963	0,71–1,00	0,834	0,50–1,00	0,865	0,49–1,00	0,83	0,49–1,00
	Багатокласова	0,917	0,17–1,00	0,909	0,08–1,00	0,854	0,30–1,00	0,835	0,10–1,00
	Загальні оцінки	0,970	0,73–1,00	0,922	0,39–1,00	0,946	0,58–1,00	0,923	0,38–1,00
USB-IDS-1	Двійкова	0,366	0,01–0,90	0,519	0,02–0,88	0,369	0,01–0,65	0,246	0,01–0,67
	Багатокласова	0,668	0,09–1,00	0,560	0,03–1,00	0,532	0,17–1,00	0,439	0,03–1,00
	Загальні оцінки	0,633	0,09–1,00	0,525	0,04–1,000	0,495	0,17–1,00	0,391	0,04–1,00

Джерело: розроблено авторами.

З цієї табл. 2 видно, що для моделі навченої набором USB-IDS-1 загальні оцінки ефективності перевищують оцінки першого рівня (в 1,3-1,7 раза для деяких оцінок, наприклад F1-score для цієї моделі зростає з 0,246 до 0,391 (у 1,6 раза). Для моделі навченої на власному наборі даних загальні оцінки також перевищують оцінки першого рівня, щоправда, не так

значно, як для попереднього випадку. Також з таблиці видно, що для даних характерних для мережі, у якій проводились випробування, діапазон значень оцінок є меншим, порівняно з моделлю навченою на нехарактерних даних із набору USB-IDS-1, а отже, пристосування класифікаційної моделі підвищує як ефективність, так і стабільність роботи моделі.

Таким чином, модель, навчена на специфічних для середовища даних, демонструє не лише вищу ефективність, але й більшу стабільність оцінок. Це вказує на переваги адаптації класифікаторів до конкретного типу трафіку. Отже, результати експериментів підтверджують основну гіпотезу: дворівневий класифікатор демонструє вищу ефективність, ніж кожен із його етапів окремо. Така архітектура поєднує високий рівень виявлення атак із досить точним розмежуванням їхніх типів, що загалом сприяє зниженню кількості хибних спрацювань. Дворівнева схема дозволяє використовувати сильні сторони кожного етапу, вирішуючи дві підзадачі послідовно й з окремо підібраними налаштуваннями. У результаті формується більш надійна та точна система моніторингу мережевого трафіку.

Висновки та перспективи подальшого дослідження. У статті було проаналізовано ефективність дворівневої класифікаційної архітектури для задач виявлення мережевих атак і уточнення їхнього типу. Запропонована система поєднує бінарну модель на основі Naïve Bayes, що виконує первинний відсів потенційно шкідливих потоків, із багатокласовим класифікатором Random Forest, який уточнює тип загрози. Такий поділ дозволяє окремо оптимізувати роботу кожного рівня: перший – на максимальне виявлення атак (високий recall), другий — на точність класифікації (високий precision). Проведений експериментальний аналіз на двох джерелах (USB-IDS-1 та власноруч зібраний трафік у віртуальному середовищі) засвідчив, що ця архітектура перевершує кожен етап окремо за точністю, F1-мірою та загальною збалансованістю показників.

Зокрема, макроусереднені метрики показали, що порівняно з бінарним класифікатором, повна система забезпечує підвищення F1-міри приблизно у 1,3–1,6 рази залежно від набору даних. Крім того, підтверджено, що другий рівень частково компенсує помилки першого – особливо тоді, коли перший рівень демонструє високу чутливість, але має недостатню специфічність.

Таким чином, усі задачі, сформульовані у межах поставленої мети, були виконані: проведено навчання класифікаторів на двох наборах, обчислено метрики окремо для кожного рівня та об'єднаної системи, проаналізовано приріст ефективності від дворівневої архітектури, а також підтверджено її стабільність на різних типах трафіку. Отримані результати обґрунтовують доцільність використання дворівневого підходу у задачах виявлення атак як з погляду точності, так і адаптивності.

Подальші дослідження можуть розвиватися в кількох напрямках. По-перше, варто розширити набір атак, які розпізнає другий рівень, і розглянути механізми роботи з невідомими типами загроз (zero-day), зокрема через методи open-set recognition. По-друге, перспективним виглядає вивчення інших комбінацій моделей на кожному рівні — передусім із залученням глибоких нейронних архітектур або адаптивних баєсівських методів. По-третє, необхідно розглядати застосування цієї архітектури в потокових, неанотованих середовищах і оцінювати її продуктивність у режимі реального часу: затримки класифікації, пропускну здатність, ресурсні витрати тощо. З урахуванням цих аспектів дворівнева модель може стати основою для побудови практичної, масштабованої системи виявлення атак у сучасних мережах.

Заява про використання генеративного ШІ та технологій на основі ШІ в процесі написання тексту статті

Під час написання цього матеріалу автор(и) не використовували технологій на основі штучного інтелекту.

Список використаних джерел

1. Zhang, H., Zhang, B., Huang, L., Zhang, Z., & Huang, H. (2023). An efficient two-stage network intrusion detection system in the internet of things. *Information*, *14*(2), 77. https://doi.org/10.3390/info14020077.

2. Zong, W., Chow, Y.-W., & Susilo, W. (2018). A two-stage classifier approach for network intrusion detection. У *Information security practice and experience* (с. 329–340). Springer International Publishing. https://doi.org/10.1007/978-3-319-99807-7_20

3. Azzaoui, H., & Boukhamla, A. (2020). Two-Stages intrusion detection system based on hybrid methods. У *ICIST '20: 10th international conference on information systems and technologies*. ACM. https://doi.org/10.1145/3447568.3448512.

4. Dubey, A. K., Singh, V., Sadaf, S., Sowjanya, G., Dehariya, K., & Mangal, A. (2023). Network traffic classification methods: A survey. *International Journal of Applied Engineering & Technology*, *5*(4), 3731–3740.

5. Hewapathirana, I. U. (2025). A comparative study of two-stage intrusion detection using modern machine learning approaches on the CSE-CIC-IDS2018 dataset. *Knowledge*, *5*(1), 6. https://doi.org/10.3390/knowledge5010006

6. Pajouh, H. H., Dastghaibyfard, G., & Hashemi, S. (2015). Two-tier network anomaly detection model: A machine learning approach. *Journal of Intelligent Information Systems*, *48*(1), 61–74. https://doi.org/10.1007/s10844-015-0388-x

7. Ahamed, M. K. U., & Karim, A. (2025). Cascaded intrusion detection system using machine learning. *Systems and Soft Computing*, 200182. https://doi.org/10.1016/j.sasc.2024.200182

8. Goldschmidt, P., & Chudá, D. (2025). Network intrusion datasets: A survey, limitations, and recommendations. *Computers & Security*, 104510. https://doi.org/10.1016/j.cose.2025.104510

9. Alshammari, F., & Alsaleh, A. (2025). Smart intrusion detection model to identify unknown attacks for improved road safety and management. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-03604-5.

10. Khamees, M. K., Ismail, M. A., Yunan, U., & Kasim, S. (2018). Review on intrusion detection system based on the goal of the detection system. *International Journal of Integrated Engineering*, *10*(6). https://doi.org/10.30880/ijie.2018.10.06.028.

11. Bagui, S., Mink, D., Bagui, S., Subramaniam, S., & Wallace, D. (2023). Resampling imbalanced network intrusion datasets to identify rare attacks. *Future Internet*, *15*(4), 130. https://doi.org/10.3390/fi15040130.

12. Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: Techniques, datasets and challenges. *Cybersecurity*, *2*(1). https://doi.org/10.1186/s42400-019-0038-7.

13. Konyrbaev, N., Nikitenko, Y., Shtanko, V., Lakhno, V., Baishemirov, Z., Ibadulla, S., Galymzhankyzy, A., & Myrzabek, E. (2024). Evaluation and optimization of the naive bayes algorithm for intrusion detection systems using the USB-IDS-1 dataset. *Eastern-European Journal of Enterprise Technologies*, *6*(2 (132)), 74–82. https://doi.org/10.15587/1729-4061.2024.317471

14. Chelloug, S. A. (2024). A robust approach for multi classification-based intrusion detection through stacking deep learning models. *Computers, Materials & Continua*, 1–10. https://doi.org/10.32604/cmc.2024.051539.

15. Zayid, E. I. M., Isah, I., Humayed, A. A., & Adam, Y. A. (2025). Innovating intrusion detection classification analysis for an imbalanced data sample. *Information*, *16*(10), 883. https://doi.org/10.3390/info16100883.

16. Bagui, S., & Li, K. (2021). Resampling imbalanced data for network intrusion detection datasets. *Journal of Big Data*, *8*(1). https://doi.org/10.1186/s40537-020-00390-x.

17. Штанько, В., & Нікітенко, Е. (2025). Проєктування та реалізація віртуального середовища для аналізу мережевого трафіку. *Наука і техніка сьогодні*, (7(48)). https://doi.org/10.52058/2786-6025-2025-7(48)-2028-2045.

References

1. Zhang, H., Zhang, B., Huang, L., Zhang, Z., & Huang, H. (2023). An efficient two-stage network intrusion detection system in the internet of things. *Information*, *14*(2), 77. https://doi.org/10.3390/info14020077

2. Zong, W., Chow, Y.-W., & Susilo, W. (2018). *A two-stage classifier approach for network intrusion detection*. In *Information security practice and experience* (pp. 329–340). Springer International Publishing. https://doi.org/10.1007/978-3-319-99807-7_20.

3. Azzaoui, H., & Boukhamla, A. (2020). Two-Stages intrusion detection system based on hybrid methods. У *ICIST '20: 10th international conference on information systems and technologies*. ACM. https://doi.org/10.1145/3447568.3448512.

4. Dubey, A. K., Singh, V., Sadaf, S., Sowjanya, G., Dehariya, K., & Mangal, A. (2023). Network traffic classification methods: A survey. *International Journal of Applied Engineering & Technology*, *5*(4), 3731–3740.

5. Hewapathirana, I. U. (2025). A comparative study of two-stage intrusion detection using modern machine learning approaches on the CSE-CIC-IDS2018 dataset. *Knowledge*, *5*(1), 6. https://doi.org/10.3390/knowledge5010006.

6. Pajouh, H. H., Dastghaibyfard, G., & Hashemi, S. (2015). Two-tier network anomaly detection model: A machine learning approach. *Journal of Intelligent Information Systems*, *48*(1), 61–74. https://doi.org/10.1007/s10844-015-0388-x.

7. Ahamed, M. K. U., & Karim, A. (2025). Cascaded intrusion detection system using machine learning. *Systems and Soft Computing*, 200182. https://doi.org/10.1016/j.sasc.2024.200182.

8. Goldschmidt, P., & Chudá, D. (2025). Network intrusion datasets: A survey, limitations, and recommendations. *Computers & Security*, 104510. https://doi.org/10.1016/j.cose.2025.104510.

9. Alshammari, F., & Alsaleh, A. (2025). Smart intrusion detection model to identify unknown attacks for improved road safety and management. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-03604-5.

10. Khamees, M. K., Ismail, M. A., Yunan, U., & Kasim, S. (2018). Review on intrusion detection system based on the goal of the detection system. *International Journal of Integrated Engineering*, *10*(6). https://doi.org/10.30880/ijie.2018.10.06.028.

11. Bagui, S., Mink, D., Bagui, S., Subramaniam, S., & Wallace, D. (2023). Resampling imbalanced network intrusion datasets to identify rare attacks. *Future Internet*, *15*(4), 130. https://doi.org/10.3390/fi15040130.

12. Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: Techniques, datasets and challenges. *Cybersecurity*, *2*(1). https://doi.org/10.1186/s42400-019-0038-7.

13. Konyrbaev, N., Nikitenko, Y., Shtanko, V., Lakhno, V., Baishemirov, Z., Ibadulla, S., Galymzhankyzy, A., & Myrzabek, E. (2024). Evaluation and optimization of the naive bayes algorithm for intrusion detection systems using the USB-IDS-1 dataset. *Eastern-European Journal of Enterprise Technologies*, *6*(2 (132)), 74–82. https://doi.org/10.15587/1729-4061.2024.317471.

14. Chelloug, S. A. (2024). A robust approach for multi classification-based intrusion detection through stacking deep learning models. *Computers, Materials & Continua*, 1–10. https://doi.org/10.32604/cmc.2024.051539

15. Zayid, E. I. M., Isah, I., Humayed, A. A., & Adam, Y. A. (2025). Innovating intrusion detection classification analysis for an imbalanced data sample. *Information*, *16*(10), 883. https://doi.org/10.3390/info16100883

16. Bagui, S., & Li, K. (2021). Resampling imbalanced data for network intrusion detection datasets. *Journal of Big Data*, *8*(1). https://doi.org/10.1186/s40537-020-00390-x.

17. Shtanko, V., & Nikitenko, Ye. (2025).). Proiektuvannia ta realizatsiia virtualnoho seredovyshcha dlia analizu merezhevoho trafiku. [Design and implementation of a virtual environment for network traffic analysis.] *Nauka i tekhnika sohodni - Science and technology today, (7(48)).* https://doi.org/10.52058/2786-6025-2025-7(48)-2028-2045.

18. Дата першого надходження статті до видання: 01.12.2025
Дата прийняття статті до друку після рецензування: 19.12.2025

UDC 004.85:004.056

Vadym Shtanko[1], Yevheniy Nikitenko[2],

[1]PhD Student, Department of Computer Systems, Networks and Cybersecurity
National University of Life and Environmental Sciences of Ukraine (Kyiv, Ukraine)
E-mail: vadym.shtanko@nubip.edu.ua. **ORCID** https://orcid.org/0009-0001-4977-1450

[2]PhD in Physical and Mathematical Sciences,
Associate Professor of the Department of Computer Systems, Networks and Cybersecurity
National University of Life and Environmental Sciences of Ukraine (Kyiv, Ukraine)
E-mail: ev.nikitenko@nubip.edu.ua. **ORCID** https://orcid.org/0000-0002-9222-644X

EVALUATING THE EFFECTIVENESS OF A TWO-LEVEL NETWORK FLOW CLASSIFIER

The growing complexity of cyber threats, especially multi-vector DDoS attacks, together with the rapid increase and heterogeneity of network traffic, creates a clear need for intrusion detection systems that can remain effective under highly dynamic network conditions. This work is motivated by the observation that conventional IDS approaches often fail to provide timely and reliable recognition of sophisticated attacks when they are hidden within large volumes of background traffic. For the field of information technology, it is therefore important to develop models that not only deliver high classification quality but can also be realistically applied in scenarios that are close to real time. At the same time, direct single-stage multiclass classification of network attacks remains a difficult task: traffic behaviour is highly variable, signatures of different threats can overlap with normal flows, and changes in the network environment directly affect the stability and consistency of the results. Addressing this challenge is essential for improving IDS reliability and for ensuring that attack types are correctly distinguished under changing traffic conditions. The aim of this study is to design and experimentally evaluate a two-level classification model that combines initial separation of suspicious flows with subsequent refinement of attack types, with the goal of achieving more stable and informative diagnostics. To reach this aim, machine-learning methods were employed, and dedicated tools for collecting, aggregating, and labeling traffic were developed and used. The proposed model was tested both on the public USB-IDS-1 dataset and on a custom traffic dataset generated in a controlled virtual environment. The main results of the study show improved consistency of classification, a reduction in misclassification events, and increased robustness of the model to changes in the operating environment. Overall, the two-level approach demonstrated practical effectiveness and potential for deployment in modern network-security monitoring systems, confirming that the stated research objectives have been achieved.

Keywords: *traffic classification; two-level model; attack detection; Naïve Bayes; Random Forest; cybersecurity.*

Fig.: 2. Table: 3. References: 17.

Штанько В. І., Нікітенко Є. В. Визначення ефективності дворівневого класифікатора мережевих потоків. *Технічні науки та технології*, 2025. № 4(42). С. 266-276. DOI: https://doi.org/10.25140/2411-5363-2025-4(42)-266-276.