КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (33), 2026
ISSN 2663 – 4023
DOI 10.28925/2663-4023.2026.33.1167
УДК 004.056

Шулімова Дар'я Денисівна
асистент кафедри Систем та технологій кібербезпеки,
Державний університет інформаційно-комунікаційних технологій, Київ, Україна
ORCID: 0009-0002-9557-990X
d.shulimova@duikt.edu.ua

ЗАСТОСУВАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ВИЯВЛЕННЯ АТАК У КОРПОРАТИВНІЙ МЕРЕЖІ НА ОСНОВІ FLOW-ОЗНАК

Анотація. Виявлення шкідливої мережевої активності в корпоративних інформаційних ресурсах на основі статистичних характеристик потоків трафіку є практично значущим завданням, оскільки ефективність детектування визначається не лише загальною точністю, а й співвідношенням хибних спрацювань і пропусків атак, що безпосередньо впливає на навантаження на операторів безпеки та рівень залишкового ризику для організації. У статті подано підхід до виявлення атак у потоці мережевих з'єднань за flow-ознаками з використанням деревоподібних методів машинного навчання та аналізом їх поведінки для різних класів загроз у межах єдиного відтворюваного експериментального протоколу. Для експериментального дослідження використано набір даних CSE-CIC-IDS2018 з ознаками, сформованими CICFlowMeter, та побудовано бінарну постановку benign проти attack для трьох сценаріїв шкідливої активності, що охоплюють botnet-активність, volumetric-атаки типу DDoS (HOIC, LOIC-UDP) і web-атаки (Brute Force-Web, Brute Force-XSS, SQL Injection). Реалізовано порівняння моделей Decision Tree та Random Forest із балансуванням класів та фіксованими параметрами розбиття даних на навчальну і тестову вибірки, що забезпечує коректне зіставлення якості на різних типах атак. Оцінювання виконано за матрицею помилок і похідними показниками для класу атаки, включно з precision і recall, а також аналізом абсолютних значень FP і FN, які є найбільш інформативними у випадку рідкісних атак. Отримані результати демонструють майже повне відокремлення benign і attack для Bot та DDoS, що узгоджується з наявністю виражених патернів у потоці трафіку та високою відокремлюваністю класів у просторі ознак. Для web-атак виявляється принципово різний профіль помилок: Decision Tree забезпечує вищу повноту виявлення за рахунок збільшення кількості хибних тривог і зниження точності спрацювань, тоді як Random Forest формує істотно точніші спрацювання при збільшенні числа пропусків. Показано, що вибір методу детектування доцільно здійснювати з урахуванням типу атаки, дисбалансу класів і допустимого балансу між хибними спрацюваннями та пропусками, а інтерпретація результатів має спиратися на показники, які відображають експлуатаційні наслідки для систем моніторингу корпоративної мережі.

Ключові слова: виявлення атак, flow-ознаки, Decision Tree, Random Forest, DDoS, botnet, web-атаки, матриця помилок, precision.

ВСТУП

Виявлення шкідливої активності в корпоративних мережах залишається одним із ключових завдань кіберзахисту, оскільки сучасні інформаційні ресурси працюють у режимі постійної мережевої взаємодії, а характер трафіку є динамічним і неоднорідним. Зростання кількості сервісів, віддалених підключень і автоматизованих обмінів даними підвищує складність моніторингу та ускладнює відокремлення нормальної активності від атак. У практичній експлуатації критичне значення має не лише факт виявлення інциденту, а й керованість потоку спрацювань: надлишок хибних тривог створює операційний шум і знижує ефективність реагування, тоді як пропуски атак формують прямі ризики компрометації. Саме тому задачі детектування доцільно формулювати з явним контролем похибок двох типів, які мають різну ціну для корпоративного середовища.

Машинне навчання є одним із базових підходів до автоматизації детектування, оскільки дозволяє будувати класифікатори на основі статистичних закономірностей у мережевих даних. У багатьох системах моніторингу зручним компромісом між деталізацією та обчислювальною вартістю виступає подання мережевої активності у форматі потоків із набором числових flow-ознак. Таке подання

переводить аналіз трафіку у простір ознак і робить можливим навчання моделей, які працюють у режимі класифікації benign проти attack. Водночас різні типи атак формують різні патерни у flow-ознаках: volumetric-атаки на кшталт DDoS часто мають виражені відмінності від нормального трафіку, botnet-активність може проявлятися у повторюваних шаблонах взаємодії, а web-атаки нерідко є рідкісними та ближчими до benign за статистичними характеристиками. Через це одна й та сама модель може демонструвати різний профіль помилок залежно від сценарію атаки, а загальна точність без аналізу FP і FN не відображає реальної експлуатаційної придатності.

Постановка проблеми. Виявлення шкідливої активності в корпоративній мережі природно формулюється як задача класифікації мережевих спостережень на нормальні та аномальні. Для того щоб перейти від сирих мережевих пакетів до компактного опису, трафік агрегується у мережеві потоки (flows), кожен з яких відповідає групі взаємопов'язаних пакетів між вузлами мережі протягом певного інтервалу часу. Для кожного потоку обчислюється набір числових ознак, що узагальнюють поведінку обміну: тривалість з'єднання, кількість пакетів і байтів у напрямках forward/backward, статистичні характеристики інтервалів між пакетами, співвідношення обсягів трафіку, ознаки активності та інтенсивності. Таке подання є придатним для побудови моделей машинного навчання, оскільки переводить мережеву активність у простір ознак $\mathbb{R}^d$, де d – кількість доступних показників потоку.

Нехай сформовано вибірку $D = \{(x_i, y_i)\}_{i=1}^{N}$, де $x_i \in \mathbb{R}^d$ – вектор ознак -го потоку, а y – його мітка. У бінарній постановці $y_i = 0$ відповідає benign-трафіку, а $y_i = 1$ відповідає атаці або шкідливій активності. Потрібно побудувати класифікатор $f(x; \theta)$, який для нового потоку x формує прогноз $\hat{y} \in \{0,1\}$ або оцінку належності до класу атаки $\hat{p}(y = 1|x) \in [0,1]$. Практичний сенс задачі полягає не лише у максимізації загальної точності, а у керованості помилок двох типів: хибні спрацювання FP, коли нормальний трафік помилково позначається як атака, та пропуски атак FN, коли шкідлива активність не виявляється. Саме ці помилки визначають експлуатаційну придатність методу: надлишок FP перевантажує операторів безпеки та знижує довіру до системи, а надлишок FN підвищує ризик успішної реалізації атаки. Тому оцінювання якості доцільно прив'язувати до матриці помилок $\{TP, TN, FP, FN\}$ та похідних метрик, зокрема $Precision = \frac{TP}{TP+FP}$, $Recall = \frac{TP}{TP+FN}$, а також показників $FPR = \frac{FP}{FP+TN}$ і $FNR = \frac{FN}{FN+TP}$, які напряму відображають частоту хибних тривог і частоту пропусків.

Окремою особливістю задачі в корпоративному середовищі є неоднорідність сценаріїв атак: одна й та сама ознакова база може містити різні типи шкідливої активності, що формують різні патерни у просторі ознак. Для volumetric-атак на кшталт DDoS зазвичай характерні різкі відмінності інтенсивності та структури потоків від нормального трафіку, тоді як web-атаки (Brute Force-Web, Brute Force-XSS, SQL Injection) часто проявляються тонше, а їх частка у даних може бути вкрай малою. Для botnet-активності можливі виражені повторювані шаблони обміну зараженого вузла з мережею керування, що також може робити клас атаки добре відокремлюваним. Через це одна й та сама модель може демонструвати різний профіль помилок на різних типах атак, і саме ця відмінність є змістовною для порівняння методів: метод може бути майже безпомилковим на DDoS, але давати або шумні тривоги, або пропуски на web-атаках. Таким чином, задача формулюється як побудова та порівняння моделей $f(x; \theta)$ на декількох сценаріях шкідливої активності з аналізом того, як змінюються $FP, FN, Precision$ і $Recall$ залежно від типу атаки та обмежень експлуатації.

Аналіз останніх досліджень і публікацій. У науковій літературі зберігається стійкий інтерес до виявлення шкідливої активності в корпоративних мережах на основі мережевих потоків і статистичних ознак трафіку, оскільки такий підхід дозволяє перейти від пакетного аналізу до компактного опису поведінки з'єднань і застосовувати методи машинного навчання для детектування аномалій. В якості поширеного еталонного середовища експериментів використовуються відкриті набори даних, зокрема CSE-CIC-IDS2018 [5],[6], який містить кілька типів атак (brute-force, botnet, DoS/DDoS, web attacks, infiltration тощо) та набір flow-ознак, сформованих засобом CICFlowMeter.

У роботі [1] виявлення аномалій у мережевому трафіку організацій пов'язується з архітектурою збору статистики на основі NetFlow/IPFIX та подальшою побудовою моделі машинного навчання; підкреслюється практична цінність матриці плутанини для оцінювання похибок алгоритмів і порівняння підходів, серед яких згадуються Random Forest та інші класичні методи. Це формує приклад інженерної постановки задачі, де модель є частиною конвеєра моніторингу.

У статті [2] фокус зміщується в бік глибокого навчання та інтеграції моделі у середовище моніторингу: використовуються архітектури CNN, LSTM та автоенкодери для виявлення аномалій трафіку DoS, а також описується інтеграція з Node-RED і підхід цифрового двійника як платформи для візуалізації, симуляції та реакції на події. Така лінія робіт показує актуальність поєднання моделей детектування з практичними механізмами експлуатації та автоматизації реагування.

Питання якості й коректності бенчмарків для IDS традиційно пов'язується з формуванням репрезентативних датасетів і ознакових описів. У статті [7] пропонують підхід до створення набору даних для задач IDS і виконують характеризацію вторгнень, акцентуючи на необхідності різноманітності трафіку та оцінюванні наборів ознак і алгоритмів на спільній основі. Для CSE-CIC-IDS2018 офіційний опис підкреслює наявність декількох сценаріїв атак та використання flow-ознак, екстрагованих CICFlowMeter, що робить цей набір даних зручним для порівняння моделей на різних типах загроз.

Окрему увагу приділяють саме інструментарію побудови flow-ознак. У документації CICFlowMeter [10] підкреслено, що інструмент генерує bi-flow та набір статистичних характеристик, включно з розділенням ознак за напрямками forward/backward і керуванням тайм-аутами потоків. Це важливо для інтерпретації результатів IDS, оскільки різні типи атак відбиваються в різних групах статистик потоку і можуть по-різному відокремлюватися в ознаковому просторі.

Проблематика узагальнюваності та пояснюваності ML-based IDS активно обговорюється в роботах, що порівнюють ознакові формати і підходи до оцінювання. У роботі [3] та [4] звертають увагу на те, що результати часто залежать від конкретного набору ознак, а також підкреслюють значення порівняння feature set між датасетами для отримання більш надійних висновків і підвищення довіри до моделей у прикладних впровадженнях. Додатково, роботи з аналізу релевантності ознак для CSE-CIC-IDS2018 демонструють, що вибір підмножини характеристик може суттєво впливати на якість та стабільність детектування для різних категорій атак.

Мета статті. Побудова та експериментальне порівняння деревоподібних методів детектування шкідливої мережевої активності на основі flow-ознак із аналізом відмінностей їх поведінки на різних типах атак. Для цього використано дані CSE-CIC-IDS2018 та сформовано три сценарії виявлення, що охоплюють botnet-активність, DDoS (HOIC, LOIC-UDP) і web-атаки (Brute Force-Web, Brute Force-XSS, SQL Injection). Як базові методи обрано Decision Tree та Random Forest, що дозволяє зіставити інтерпретовану модель правил і ансамблевий підхід із підвищеною стійкістю. Оцінювання виконується за матрицею помилок і похідними показниками для класу атаки, що дає змогу інтерпретувати результати в термінах хибних тривог і пропусків та прив'язати висновки до практичних вимог корпоративного моніторингу.

МЕТОДИКА ДОСЛІДЖЕННЯ

Виявлення шкідливої активності за flow-ознаками можна інтерпретувати як задачу побудови відображення $f: \mathbb{R}^d \to \{0,1\}$, яке для вектора ознак потоку x визначає належність до класу атаки. У практичній реалізації зручно розглядати скорингову функцію або ймовірнісну оцінку $\hat{p}(y = 1 \mid x) \in [0,1]$, після чого рішення приймається пороговим правилом:

$$\hat{y} = \mathbb{I}(\hat{p}(y = 1 \mid x) \geq \tau), \tag{1}$$

де $\tau \in (0,1)$ – поріг спрацювання, а $\mathbb{I}(\,)$ – індикатор. Саме параметр τ пов'язує модель із експлуатаційними вимогами: збільшення τ зазвичай зменшує кількість хибних спрацювань FP, але може збільшувати кількість пропусків атак FN; зменшення τ діє протилежно. Таким чином, навіть для фіксованої моделі можливе налаштування компромісу між шумом тривог і ризиком невиявлення, що є ключовим для корпоративних систем моніторингу.

Навчання моделі задається вибіркою $D = \{(x_i, y_i)\}_{i=1}^{N}$. У загальному вигляді параметри θ визначаються як мінімізатор емпіричного ризику:

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \ell(y_i, s(x_i; \theta)), \tag{2}$$

де $s(x; \theta)$ – скорингове значення моделі (або логіт/ймовірність), а $\ell(\cdot)$ – функція втрат. У задачах виявлення атак клас $y = 1$ часто є менш представленим, тому доцільним є використання зваженої постановки, яка підсилює внесок рідкісного класу в критерій навчання:

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} w_{y_i} \ell(y_i, s(x_i; \theta)), \tag{3}$$

де $w_1 > w_0$ відповідає більшим штрафам за помилки на класі атаки. На рівні реалізації для деревоподібних моделей це відповідає параметру балансування класів, який робить побудову розбиттів більш чутливою до атак навіть за малого їх числа.

Для обраних методів основою є деревоподібні моделі, де рішення формується через послідовність простих правил. У дереві рішень простір ознак $\mathbb{R}^d$ рекурсивно ділиться гіперплощинами виду $x_j \leq t$, де j – індекс ознаки, t – порогове значення. На кожному вузлі дерева обирається така пара (j, t), що максимізує "виграш" від поділу. Для бінарної класифікації типовою мірою неоднорідності є індекс Джині:

$$G(S) = 1 - \sum_{c \in \{0,1\}} p_c^2, \tag{4}$$

де p_c – частка класу c у множині прикладів S, які потрапили до вузла. Якщо поділ розбиває S на S_L і S_R, то зменшення неоднорідності визначається як:

$$\Delta G = G(S) - \frac{|S_L|}{|S|} G(S_L) - \frac{|S_R|}{|S|} G(S_R), \tag{5}$$

і обирається поділ з максимальним ΔG. У такий спосіб дерево "підбирає" пороги, які найкраще відокремлюють атаки від benign у поточному фрагменті даних. Для контролю складності вводяться обмеження, зокрема максимальна глибина D (параметр max_depth), що зменшує ризик перенавчання: надто глибокі дерева можуть підлаштовуватися під випадкові флуктуації та шум і, як наслідок, генерувати зайві FP на нових даних [9],[11].

Random Forest формалізує ідею ансамблю: замість одного дерева будується множина дерев $\{f_m(x)\}_{m=1}^M$, а підсумкове рішення формується агрегуванням [8]. Для класу атаки зручно розглядати середню оцінку ймовірності:

$$\hat{p}(y = 1|x) = \frac{1}{M} \sum_{m=1}^{M} \hat{p}_m (y = 1|x), \tag{6}$$

після чого застосовується поріг τ. Різноманітність дерев забезпечується двома механізмами: випадковим вибором підвибірок об'єктів (bootstrap) та випадковим вибором підмножин ознак при пошуку розбиття. Це зменшує кореляцію між деревами і знижує дисперсію моделі: у порівнянні з одиничним деревом ансамбль менш чутливий до випадкових особливостей навчальної вибірки та частіше демонструє більш стабільний профіль помилок. Саме ця властивість зазвичай приводить до зменшення FP на складних сценаріях, хоча ціною може бути зростання FN, якщо ансамбль стає більш консервативним і потребує сильніших "доказів" атаки у просторі ознак для формування позитивного рішення.

Формально якість моделі на тестовій множині виражається через матрицю помилок (TP, TN, FP, FN). З неї випливають основні показники, що безпосередньо пов'язані з експлуатаційними вимогами: частота хибних спрацювань:

$$FPR = \frac{FP}{FP + TN}, \tag{7}$$

та частота пропусків

$$FNR = \frac{FN}{FN + TP}. \tag{8}$$

У сценаріях із сильним дисбалансом, коли кількість benign-потоків значно перевищує кількість атак, метрика accuracy може залишатися високою навіть при неприйнятній кількості FN або надмірному FP, тому саме $Precision, Recall, FPR$ і FNR є більш інформативними для оцінювання системи виявлення вторгнень. Таким чином, математична модель включає не лише побудову скорингового класифікатора на основі деревоподібних правил, але й явний механізм керування рішенням через поріг τ і аналіз помилок як функції типу атаки, що дозволяє інтерпретувати різницю в поведінці Decision Tree та Random Forest на DDoS, botnet-активності та web-атаках.

Для моделювання виявлення шкідливої активності використовується набір мережевих спостережень у форматі flow-ознак, отриманих шляхом агрегування пакетного трафіку в потоки [10]. Потік у такому поданні є узагальненням мережевої взаємодії між вузлами, яке формується за правилами групування з'єднань і тайм-аутів, а результатом агрегування є один запис із числовими характеристиками. Це дозволяє представити мережеву активність як матрицю ознак $X \in \mathbb{R}^{N \times d}$, де N – кількість потоків, а d – кількість обчислених показників, та вектор міток $y \in \{0,1\}^N$. У якості джерела

даних використано CSE-CIC-IDS2018, що містить трафік різних сценаріїв атак і нормальної активності, а експерименти виконано на трьох окремих файлах, які відповідають різним типам загроз: botnet-активність, volumetric-атака типу DDoS та web-атаки. Такий вибір дозволяє порівнювати поведінку моделей у ситуаціях з різною природою аномалій: від різко виражених відмінностей у трафіку до сценаріїв, де атака проявляється тонко і є рідкісною [5].

Ознаковий простір формується з груп статистичних характеристик потоку, які узагальнюють інтенсивність, структуру та часову динаміку обміну. До типових груп належать показники тривалості потоку та активності, кількість пакетів і байтів у напрямках forward/backward, середні значення та дисперсії розмірів пакетів, статистики інтервалів між пакетами і похідні від них міри варіативності, співвідношення обсягів у напрямках, а також індикатори поведінкових особливостей, які відображають характер взаємодії. У такому поданні різні типи атак залишають різні сліди:

– DDoS зазвичай породжує багато потоків з високою інтенсивністю або нетиповими співвідношеннями обсягів, що робить клас атаки добре відокремлюваним;

– Botnet-активність часто демонструє повторювані шаблони взаємодії заражених вузлів з мережею керування або сервісами, які відбиваються у часових і статистичних характеристиках потоків;

– Web-атаки, навпаки, можуть бути близькими до нормального веб-трафіку, а їх прояв може концентруватися в невеликій частині ознак, через що відокремлення класів стає складнішим і більш чутливим до вибору моделі та налаштувань прийняття рішення.

Підготовка даних виконується з урахуванням вимог алгоритмів машинного навчання та можливих артефактів обчислення ознак. Спочатку забезпечується коректність числових значень шляхом заміни нескінченних значень на пропуски та видалення рядків із некоректними записами, оскільки наявність NaN або $\pm\infty$ призводить до нестабільної роботи процедур навчання та оцінювання. Далі мітки класів приводяться до бінарної постановки: benign інтерпретується як 0, а будь-який тип атаки – як 1. Формально це відповідає перетворенню багатокласової мітки на індикатор шкідливої активності, що узгоджується з практичною логікою IDS, де ключовим є факт наявності підозрілої поведінки. Після цього ознакова матриця формується лише з числових колонок, що виключає проблеми кодування нечислових атрибутів і забезпечує однаковий формат для всіх сценаріїв. Розбиття на навчальну і тестову вибірки виконується зі збереженням частки класів у кожній частині за допомогою стратифікації, що особливо важливо для web-атак, де атака є рідкісною і випадкове розбиття може дати некоректне уявлення про якість.

У цьому підході важливим фактором є дисбаланс класів, тобто ситуація, коли benign суттєво переважає або, навпаки, в окремих сценаріях частка атаки є домінуючою. Дисбаланс не є просто статистичною деталлю, він прямо впливає на те, які помилки буде робити модель і як інтерпретувати результати: при рідкісних атаках навіть невелика абсолютна кількість хибних спрацювань може створювати значний шум у тривогах, а при домінуванні атаки важливо контролювати пропуски benign як потенційне джерело некоректних політик. Саме тому вибір трьох різних сценаріїв дозволяє показати, що поведінка моделей визначається не лише типом алгоритму, а й структурою даних і тим, як ознаки відображають конкретний механізм атаки.

Експериментальна перевірка методів організовується так, щоб порівняння моделей відбувалося в однакових умовах для різних типів атак, а результати можна було повторити на іншому середовищі без зміни логіки обробки. Для кожного з обраних сценаріїв використовується окремий CSV-файл, який містить потоки benign і потоки з відповідним типом шкідливої активності, після чого застосовується єдина процедура підготовки: очищення даних від некоректних значень, бінаризація міток і формування матриці числових ознак. Такий підхід дозволяє трактувати кожен сценарій як окрему задачу виявлення аномалії певної природи при збереженні незмінного інструментарію, що робить відмінності в результатах наслідком властивостей даних та алгоритмів, а не різних налаштувань обробки.

Після формування X та y виконується розбиття на навчальну та тестову частини за схемою hold-out. Частка тестових даних фіксується на рівні 20 відсотків, а відтворюваність забезпечується фіксацією псевдовипадкового зерна генератора, що гарантує однаковий поділ при повторному запуску. Додатково використовується стратифікація за міткою, тобто у навчальній і тестовій вибірках зберігається близьке співвідношення benign та attack. Це критично для рідкісних атак у веб-сценаріях, де відсутність стратифікації може призвести до ситуації, коли у тесті буде надто мало позитивних прикладів і оцінка якості стане нестабільною. Окремо контролюється базова статистика даних після розбиття, зокрема кількість об’єктів кожного класу в train і test, що дозволяє відразу виявити перекоси у частках класів до навчання моделей.

Для побудови класифікаторів використовуються два алгоритми, які належать до класу деревоподібних методів, але відрізняються підходом до узагальнення: Decision Tree як одинична

інтерпретована модель правил і Random Forest як ансамбль дерев. Щоб коректно працювати з дисбалансом класів, для обох алгоритмів задається балансування класів, що зменшує тенденцію моделі ігнорувати рідкісні атаки. Для дерева додатково вводиться обмеження глибини, яке виступає регуляризацією та знижує ризик перенавчання, особливо на сценаріях із тонкими відмінностями між benign і атакою. Для Random Forest фіксується кількість дерев в ансамблі, що забезпечує стабільність результатів та зменшує чутливість до випадкових флуктуацій у даних.

Оцінювання моделей виконується на тестовій множині, яка не використовується під час навчання. Базовим об'єктом інтерпретації результатів є матриця помилок, яка однозначно відображає кількість правильних рішень та помилок двох типів. На її основі розраховуються показники, що мають прямий експлуатаційний зміст у системах моніторингу: precision для класу атаки як частка істинних тривог серед усіх тривог, recall як частка виявлених атак серед усіх атак, а також похідні характеристики, що відображають баланс помилок. У випадку рідкісних атак акцент переноситься на аналіз FP і FN у абсолютних значеннях, оскільки навіть невеликі відносні відхилення можуть призводити до значних операційних наслідків, наприклад перевантаження аналітиків через надлишок хибних спрацювань або неприпустимий ризик через пропуски. Для сценаріїв з добре відокремлюваними класами високі значення precision і recall доповнюються аналізом одиничних помилок, що дозволяє оцінити стабільність моделі та її потенційну придатність як базового компонента IDS.

Відтворюваність забезпечується не лише фіксацією єдиного протоколу розбиття, але й однаковістю коду для всіх сценаріїв: змінюється лише вхідний файл, тоді як процедура підготовки даних, набір ознак і налаштування моделей залишаються незмінними. Така організація дозволяє коректно порівнювати профілі помилок моделей на різних типах атак і робити узагальнення щодо того, які алгоритми є більш придатними для певного класу загроз. У підсумку протокол формує відтворюваний цикл, який можна розширювати додатковими методами або налаштуваннями без зміни базової структури експериментів:

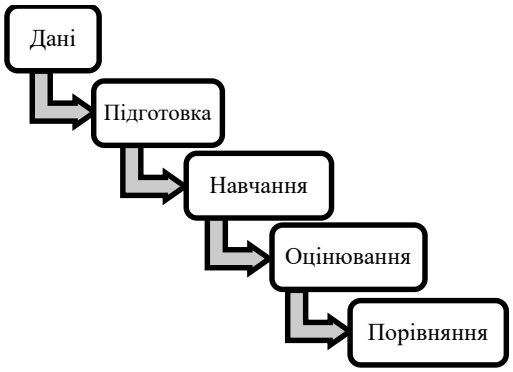

Рис.1. Цикл побудови, оцінювання та порівняння моделей виявлення атак

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Порівняння Decision Tree та Random Forest на трьох сценаріях шкідливої активності демонструє, що профіль помилок моделі визначається не лише вибором алгоритму, а й природою атаки та тим, як вона відображається у просторі flow-ознак.

Таблиця1

Профіль помилок моделей на різних типах атак

Сценарій	Модель	TN	FP	FN	TP	Precision (attack)
Bot (Friday-02-03)	Decision Tree	151665	2	4	57234	1.0000
Bot (Friday-02-03)	Random Forest	151667	0	4	57234	1.0000
DDoS (Wednesday-21)	Decision Tree	72166	1	0	137548	1.0000
DDoS (Wednesday-21)	Random Forest	72167	0	0	137548	1.0000
Web attacks (Friday-23)	Decision Tree	208130	331	9	104	0.2391
Web attacks (Friday-23)	Random Forest	208459	2	36	77	0.9747

Для сценарію botnet-активності (Friday-02-03) обидві моделі забезпечили майже безпомилкове відокремлення benign та Bot. Для Decision Tree на тестовій вибірці отримано $TN = 151665, FP = 2, FN = 4, TP = 57234$, що означає практично відсутні хибні спрацювання та одиничні пропуски атак. Random Forest у цьому сценарії показав $TN = 151667, FP = 0, FN = 4, TP = 57234$, тобто повністю прибрав

хибні тривоги, зберігши однаковий рівень пропусків. Така якість узгоджується з тим, що botnet-трафік часто формує характерні повторювані патерни взаємодії, які суттєво відрізняються від типових benign-потоків і тому добре відокремлюються навіть простими правилами [5]. На практиці це означає, що для подібних сценаріїв складність моделі не є визначальним фактором, а більш важливими стають стабільність і контроль одиничних помилок.

Для сценарію DDoS (Wednesday-21) також спостерігається майже повна відокремлюваність класів. Decision Tree дало $TN = 72166, FP = 1, FN = 0, TP = 137548$, а Random Forest – $TN = 72167, FP = 0, FN = 0, TP = 137548$. Таким чином, ансамблевий метод знову продемонстрував більш консервативну поведінку щодо benign-трафіку, повністю усунувши хибні спрацювання, тоді як пропуски атак у цьому сценарії відсутні для обох моделей. Змістовно це пояснюється тим, що volumetric-атаки, зокрема HOIC/LOIC-UDP, відображаються у flow-ознаках як виразні зміни інтенсивності та статистичних характеристик потоків. У такій ситуації навіть обмежене дерево рішень здатне побудувати правила, які майже ідеально відділяють DDoS-потоки від benign, а Random Forest додатково зменшує залишковий шум за рахунок усереднення по ансамблю.

Найбільш показовими є результати для web-атак (Friday-23), де атаки є рідкісними і їх ознакові прояви можуть бути близькими до нормального веб-трафіку [3]. У цьому сценарії Decision Tree сформувало матрицю помилок $TN = 208130, FP = 331, FN = 9, TP = 104$. Такий результат означає, що дерево виявляє більшість атак, оскільки кількість пропусків є малою відносно числа атак у тесті, проте ціною цього стає значний обсяг хибних тривог. Практично це проявляється як зниження точності спрацювань: значна частина потоків, позначених моделлю як атака, насправді належить до benign-класу. Для системи моніторингу це є критичним, оскільки велика кількість FP створює операційний шум і збільшує витрати на ручну верифікацію інцидентів. Random Forest у цьому ж сценарії продемонстрував інший профіль помилок: $TN = 208459, FP = 2, FN = 36, TP = 77$. Така поведінка характеризується майже повною відсутністю хибних спрацювань при одночасному зростанні кількості пропусків атак. З точки зору точності тривог це є перевагою, оскільки практично кожне спрацювання відповідає реальній атаці, однак частина атак залишається невиявленою, що підвищує ризик пропуску інциденту.

Різниця між моделями на web-атаках узгоджується з їхньою природою. Одиничне дерево рішень є високоваріативною моделлю: воно здатне побудувати набір локальних правил, які "підхоплюють" слабкі сигнали атаки, але при цьому може реагувати на випадкові коливання ознак і генерувати зайві спрацювання, особливо за умов дисбалансу класів. Random Forest, навпаки, зменшує дисперсію за рахунок ансамблювання, що робить рішення більш стабільним і, як наслідок, більш обережним: модель рідше позначає benign як атаку, але також може "відсікти" частину слабко виражених атак. У термінах порогового правила це означає, що ансамбль фактично формує більш концентровані оцінки $\hat{p}(y = 1 \mid x)$ для справді відокремлених атак і менш охоче підвищує скоринг для прикладів, що знаходяться поблизу межі класів.

Отримані результати дозволяють зробити узагальнення щодо залежності якості від типу атаки. Для Bot та DDoS flow-ознаки містять достатньо інформації для майже безпомилкового відокремлення класів, тому обидва методи забезпечують високу якість, а різниця проявляється переважно в одиничних FP. Для web-атак, де атака є рідкісною і більш схожою на benign, вибір методу визначає операційний компроміс: Decision Tree орієнтується на високу повноту виявлення при зростанні шуму тривог, а Random Forest забезпечує високу точність тривог при зростанні частки пропусків. Це створює підґрунтя для подальшого налаштування системи виявлення шляхом керування порогом τ або вагами помилок, що дозволяє адаптувати модель до конкретних вимог корпоративної експлуатації, наприклад зменшувати FP при фіксованому прийнятному рівні FN або навпаки.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Виявлення шкідливої мережевої активності за flow-ознаками зводиться до відокремлення нормальної поведінки трафіку від аномальної з контролем двох типів помилок, які мають різну "ціну" в експлуатації: хибні тривоги FP створюють шум і перевантажують обробку інцидентів, тоді як пропуски FN означають невиявлену атаку або компрометацію. Результати для трьох різних сценаріїв атак показують, що одна й та сама модель може демонструвати різну поведінку залежно від природи аномалії та її представленості в даних. Для botnet-активності та DDoS у просторі flow-ознак присутні виражені патерни, які суттєво відрізняють атаку від benign, тому деревоподібні моделі забезпечують майже повне розділення класів. У таких сценаріях рішення формується стабільно, а відмінності між Decision Tree та Random Forest проявляються здебільшого як різний рівень "обережності" щодо benign-трафіку, тобто в мінімізації одиничних хибних спрацювань.

Інша картина спостерігається для web-атак, де атаки є рідкісними, а частина потоків за своїми статистичними характеристиками може бути близькою до нормальної веб-активності. Саме тут проявляється змістовна різниця в профілі помилок моделей: Decision Tree демонструє високу повноту виявлення атак, але ціною помітної кількості хибних тривог, що знижує практичну цінність спрацювань та потребує додаткового фільтрування або кореляції в системі моніторингу. Random Forest, навпаки, формує значно "чистіші" тривоги з дуже малим числом FP, однак частина атак залишається невиявленою, що підвищує вимоги до компенсуючих механізмів, наприклад до кореляції з іншими джерелами подій або до політик повторного аналізу підозрілих сесій. Таким чином, на практиці порівняння методів набуває сенсу не як пошук єдиного найкращого алгоритму для всіх загроз, а як встановлення того, який режим роботи є прийнятним для конкретного класу атак: режим підвищеної чутливості з більшим шумом або режим мінімізації шуму з ризиком пропусків. Це узгоджується з експлуатаційною логікою корпоративної безпеки, де важливі не тільки метрики загальної точності, а й керованість потоку тривог і контроль ризику невиявлених інцидентів.

Перспективи подальших досліджень пов'язуються з підвищенням практичної придатності моделей у корпоративній мережі. Доцільним є налаштування порогу прийняття рішення τ для керування балансом FP та FN, особливо для web-атак, а також перевірка узагальнюваності на інших часових відрізках трафіку, щоб оцінити стійкість моделей до змін профілю мережевої активності [12]. Окремої уваги потребує аналіз інформативності та відбір ключових flow-ознак для підвищення інтерпретованості й зменшення обчислювальних витрат, а також розширення постановки до класифікації типів атак після первинного виявлення аномалії.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Haidur, H. I., Hakhov, S. O., Dmitriiev, V. Y., & Bondarenko, N. V. (2021). Detection of traffic anomalies in organizational information systems using machine learning methods based on categorical field prediction algorithms. *Telecommunications and Information Technologies*. https://tit.dut.edu.ua/index.php/telecommunication/article/view/2402
2. Savchenko, T. V., Lutska, N. M., Vlasenko, L. O., & Tomenko, N. D. (2025). Analysis of the effectiveness of network traffic anomaly detection based on machine learning models. *Cybersecurity: Education, Science, Technique*. https://csecurity.kubg.edu.ua/index.php/journal/article/view/898
3. Sarhan, M., Layeghy, S., & Portmann, M. (2022). Evaluating standard feature sets towards increased generalisability and explainability of ML-based network intrusion detection. *Array*. https://www.sciencedirect.com/science/article/abs/pii/S2214579622000533
4. Sarhan, M., Layeghy, S., & Portmann, M. (2021). Evaluating standard feature sets towards increased generalisability and explainability of ML-based network intrusion detection. *arXiv*. https://arxiv.org/abs/2104.07183
5. Canadian Institute for Cybersecurity. (2018). *CSE-CIC-IDS2018 dataset*. https://www.unb.ca/cic/datasets/ids-2018.html
6. Amazon Web Services. (n.d.). *A realistic cyber defense dataset (CSE-CIC-IDS2018)*. Registry of Open Data on AWS. https://registry.opendata.aws/cse-cic-ids2018/
7. Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. In *Proceedings of the International Conference on Information Systems Security and Privacy (ICISSP 2018)*. https://www.scitepress.org/papers/2018/66398/66398.pdf
8. Breiman, L. (2001). Random forests. *Machine Learning*. https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf
9. Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann. https://dl.acm.org/doi/abs/10.5555/152181
10. Lashkari, A. H. (n.d.). *CICFlowMeter* [Computer software]. GitHub. https://github.com/ahlashkari/CICFlowMeter
11. scikit-learn DecisionTreeClassifier documentation
12. scikit-learn RandomForestClassifier documentation

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (33), 2026
ISSN 2663 – 4023

Daria Shulimova
Assistant, Department of Cybersecurity Systems and Technologies
State University of Information and Communication Technologies, Kyiv,Ukraine
ORCID: 0009-0002-9557-990X
d.shulimova@duikt.edu.ua

APPLYING MACHINE LEARNING METHODS TO DETECT ATTACKS IN A CORPORATE NETWORK BASED ON FLOW FEATURES

Abstract. Detecting malicious network activity in corporate information resources using statistical traffic flow characteristics is a practically important task, since detection effectiveness is determined not only by overall accuracy but also by the ratio of false alarms to missed attacks, which directly affects the workload of security operators and the level of residual risk for an organization. This paper presents an approach to attack detection in network connection streams based on flow features using tree-based machine learning methods and analyzes their behavior across different threat classes within a single reproducible experimental protocol. The experimental study employs the CSE-CIC-IDS2018 dataset with features extracted by CICFlowMeter and formulates a binary classification problem of benign versus attack for three malicious activity scenarios covering botnet activity, volumetric DDoS attacks (HOIC, LOIC-UDP), and web attacks (Brute Force-Web, Brute Force-XSS, SQL Injection). A comparison of Decision Tree and Random Forest models is implemented with class balancing and fixed train–test split parameters to ensure a consistent evaluation across different attack types. Performance is assessed using the confusion matrix and derived metrics for the attack class, including precision and recall, as well as an analysis of absolute FP and FN values, which are most informative in the presence of rare attacks. The obtained results demonstrate an almost complete separation between benign and attack classes for Bot and DDoS, which is consistent with the presence of pronounced traffic patterns and high class separability in the feature space. For web attacks, a fundamentally different error profile is observed: the Decision Tree achieves higher detection completeness at the cost of an increased number of false alarms and reduced alert precision, whereas the Random Forest produces substantially more precise alerts while increasing the number of missed attacks. It is shown that the choice of a detection method should account for the attack type, class imbalance, and the acceptable trade-off between false alarms and missed detections, and that result interpretation should rely on metrics that reflect operational consequences for corporate network monitoring systems.

Keywords: attack detection, flow features, Decision Tree, Random Forest, DDoS, botnet, web attacks, error matrix, precision.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Haidur, H. I., Hakhov, S. O., Dmitriiev, V. Y., & Bondarenko, N. V. (2021). Detection of traffic anomalies in organizational information systems using machine learning methods based on categorical field prediction algorithms. *Telecommunications and Information Technologies*. https://tit.dut.edu.ua/index.php/telecommunication/article/view/2402
2. Savchenko, T. V., Lutska, N. M., Vlasenko, L. O., & Tomenko, N. D. (2025). Analysis of the effectiveness of network traffic anomaly detection based on machine learning models. *Cybersecurity: Education, Science, Technique*. https://csecurity.kubg.edu.ua/index.php/journal/article/view/898
3. Sarhan, M., Layeghy, S., & Portmann, M. (2022). Evaluating standard feature sets towards increased generalisability and explainability of ML-based network intrusion detection. *Array*. https://www.sciencedirect.com/science/article/abs/pii/S2214579622000533
4. Sarhan, M., Layeghy, S., & Portmann, M. (2021). Evaluating standard feature sets towards increased generalisability and explainability of ML-based network intrusion detection. *arXiv*. https://arxiv.org/abs/2104.07183
5. Canadian Institute for Cybersecurity. (2018). *CSE-CIC-IDS2018 dataset*. https://www.unb.ca/cic/datasets/ids-2018.html

6. Amazon Web Services. (n.d.). *A realistic cyber defense dataset (CSE-CIC-IDS2018)*. Registry of Open Data on AWS. https://registry.opendata.aws/cse-cic-ids2018/
7. Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. In *Proceedings of the International Conference on Information Systems Security and Privacy (ICISSP 2018)*. https://www.scitepress.org/papers/2018/66398/66398.pdf
8. Breiman, L. (2001). Random forests. *Machine Learning*. https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf
9. Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann. https://dl.acm.org/doi/abs/10.5555/152181
10. Lashkari, A. H. (n.d.). *CICFlowMeter* [Computer software]. GitHub. https://github.com/ahlashkari/CICFlowMeter
11. scikit-learn DecisionTreeClassifier documentation
12. scikit-learn RandomForestClassifier documentation

Отримано редакцією журналу / Received: 07.02.26
Прорецензовано / Revised: 21.02.26
Схвалено до друку / Accepted: 25.06.26