УДК 004.7 **doi: 10.26906/SUNZ.2024.1.127**

О. С. Ляшенко, І. А. Великодний, В. Г. Знайдюк, О. Д. Журило

Харківський національний університет радіоелектроніки, Харків, Україна

# МОДЕЛЬ ТА МЕТОДИ ВИЯВЛЕННЯ ШИРОКОМАСШТАБНОЇ АТАКИ В СЕРЕДОВИЩІ IoT

**Анотація.** Головною концепцією і предметом дослідження є виявлення різного типу обширних атак в інфраструктурі IoT, огляд представленої моделі, методів та існуючих передових систем виявлення вторгнень. **Метою** даної роботи є запропонування системи виявлення вторгнень в режимі реального часу, яка буде навчена на наборі з великим обсягом даних, за допомогою нейронної мережі з використанням ансамблевого **методу** машинного навчання. **Предметом дослідження** є огляд існуючих методів та моделей виявлення широкомасштабної атаки та запропонування власного рішення системи виявлення вторгнень, яка буде базуватися на методі виявлення аномалій та нейронної мережі. **Висновок.** Побудована система виявлення вторгнень, яка аналізує інтернет трафік, вилучає ознаки з пакету, обробляє їх та передбачує різні види атак, а також характеризує їх за типом. Загрозу безпеці можна вважати основною критичною проблемою для пристроїв IoT, тому використання таких систем зменшує ризики втрати даних.

**Ключові слова:** набір даних, нейронна мережа, машинне навчання, мережевий трафік, IDS, навчання, передбачення, виявлення аномалій, атака, Інтернет речей, система виявлення вторгнень.

## Вступ

Інтернет речей – концепція мережі, яка об'єднує фізичні пристрої з вбудованими датчиками, а також програмним забезпеченням, що забезпечує ефективну та спрощену взаємодію між фізичним світом і комп'ютерними системами, за допомогою, найчастіше, стандартних протоколів зв'язку. Протягом останніх років він стрімко зростав та продовжує зростати у різних галузях. Пристрої IoT функціонують у сферах освіти, охорони здоров'я, сільському господарстві, транспортних системах та промисловості. Кількість підключених пристроїв по всьому світу стрімко росте. Системи включають в себе масу датчиків, які дозволяють збирати дані в реальному часі. Отримані дані, це свого роду фундамент для створення інтелектуальних алгоритмів прийняття рішень. Ростуча кількість пристроїв, ціна і важливість інформації збільшує ризик кіберзагроз і викраденню інформації в корисних цілях. Виходячи з цього розробка інтелектуальних методів та систем виявлення вторгнень для пристроїв IoT стає необхідною для їх ефективного захисту. Тема безпеки інформаційного середовища стає дедалі актуальною і кібербезпека набуває життєвої важливості, з огляду на те що IoT є драйвером промислової революції та системою для збору живих даних [1]. Таким чином, система виявлення вторгнень є необхідною для виявлення і захисту мережі та пов'язаних систем від поточних і майбутніх кібератак.

## Системи виявлення вторгнень

Визначимо концепцію IDS (*Intrusion Detection System, або система виявлення вторгнень*). Це програмний або апаратний засіб, який виявляє або запобігає несанаційному доступу до комп'ютерної мережі чи системи. Головна мета IDS полягає в реагуванні на небезпечні події, потенційно небезпечні, або аномалії, що можуть вказувати на вторгнення чи інші безпекові порушення. З основних завдань системи виявлення вторгнень можна виділити: *виявлення аномалій* є функцію моніторингу системи чи мережі для виявлення незвичайних патернів, подій або некоректних дій, які можуть бути ознакою вторгнення чи іншої загрози безпеці; *виявлення вторгнень* – розпізнавання несанкційного доступу, спроб атак на інформаційні системи, вірусів, троянських програм та іншого шкідливого коду; *відслідковування і реагування* – забезпечення можливості вжиття заходів до виявлених загроз, включаючи блокування доступу, відключення систем які на думку є найбільш вразливі або відправлення сповіщень адміністраторам, котрі відповідають за безпеку.

Системи виявлення вторгнень можуть використовувати різні методи, такі як сигнатурний аналіз, виявлення аномалій, використання інтелектуальних технологій, включаючи *машинне навчання* та евристичний аналіз. Ефективний захист включає в себе інтеграцію системи виявлення вторгнень з іншими методами безпеки, та поєднання цих методів, для створення комплексного захисту інформаційного стеку. Зазвичай системи виявлення вторгнень використовують два основні підходи для виявлення потенційних загроз: сигнатурний аналіз та виявлення аномалій. Розглянемо ці методи більш детально.

Сигнатурний аналіз – метод який грунтується на використанні визначених сигнатур або патернів для ідентифікації або розпізнавання конкретних відомих загроз. Сигнатури можуть представляти з себе конкретні приклади або вирази в шкідливому програмному коді, унікальні характеристики того чи іншого вірусу чи способу вторгнення, які раніше вже були визначені або вивчені [2]. Спеціалісти з безпеки аналізують атаки і розробляють сигнатури для кожного виду. Зазвичай це може бути характеристика конкретних строк коду, значень в певних полях або якийсь інший ідентифікатор, який буде унікальним для деяких типів атак. Система виявлення вторгнень застосовує ці сигнатури для пошуку вхідних даних чи активності в мережі, які відповідають зазначеним сигнатурам. Якщо є збіг, система дає сповіщення про потенційне вторгнення. Сигнатурний аналіз ефективний і має високу точність проти відомих векторів атак та відомих загроз, але не ефективний проти нових та невідомих загроз. Він потребує постійного оновлення

 

бази сигнатур для визначення нових загроз, більш того, зловмисники можуть уникати виявлення, шляхом зміни або шифрування свого коду. В сучасному середовищі сигнатурний аналіз залишається надійним засобом для виявлення вторгнень, але йому важко справлятися векторами атак, що постійно змінюються, які все частіше використовують нові техніки та методи, тому в сучасних *IDS* його часто доповнюють інші методи, тобто використовується комбінація різних методів для комплексного захисту, такі як виявлення аномалій, для більшої ефективності виявлення нових атак.

Виявлення аномалій – метод який базується на аналізі звичайної поведінки мережі, системи, користувачів чи інших об'єктів. Система методу будує модель так званої "норми" на основі історичних даних, фокусується на виявленні незвичайностей, відхиленню від цієї норми, яке може бути ознакою нових загроз або підозр [3]. До підходів виявлення можемо віднести: *статистичні методи,* які використовуються для аналізу величин, таких як середнє значення, середне відхилення, тощо. Відхилення від норми цих величин може вказувати на присутність аномалії; *методи машиного навчання* – створення моделей за допомогою алгоритмів машинного навчання, які можуть визначати незвичайні патерни в даних та вказувати на аномалію, наприклад, за допомогою алгоритмів кластеризації або нейронної мережі; *методи порівняння зразків* – ґрунтуються на порівнянні поточної поведінки з історичними даними, якщо виявляється відхилення від звичайної моделі, це може бути зафіксовано як аномалія.

Виявлення аномалій може проводитися на основі патернів мережевого трафіку, неправильних адрес або портів, незвичних об'ємів даних, аналіз лог файлів, що містять інформацію про дію та поведінку системи чи користувачів, надто часті або великі запити, невластиві часові рамки, тощо. Застосування методу виявлення аномалій допомагає виявляти атаки, які можуть бути невідомими(нуль-день) та непередбаченими, що робить його ефективним і корисним для захисту від нових атак та загроз [3].

Сигнатурний аналіз і виявлення аномалій часто використовують в комплексі, як частина більших систем виявлення вторгнень. Комбінація цих методів дозволяє створити більш ефективну систему виявлення вторгнень, здатну протидіяти різноманітним загрозам безпеки.

## Машинне навчання в системах виявлення вторгнень

Машинне навчання відіграє важливу роль у покращені ефективності та адаптивності в системах виявлення вторгнень. Воно дозволяє системам аналізувати дані, навчатися на їх основі, та виявляти нові невідомі загрози, класифікувати події як безпечні чи підозрілі. Навчання моделі на основі історичних даних та поведінки допомагає автоматично розпізнати нові атаки чи загрози. Щоб адаптуватися до змін у поведінці системи чи користувачів системи, виявлення вторгнень можуть використовувати онлайн навчання. Це дозволить системі навчатися в реальному часі, а також підтримувати актуальність моделей. Машинне навчання ефективно працює з великими обсягами даних, що дозволяє виявляти складні патерни та взаємодії, які може бути важко виявити за допомогою традиційних методів [4–7]. Застосування машинного навчання дозволяє створювати інтелектуальні *IDS*, які можуть взаємодіяти та розпізнавати атаки на високому рівні. Використання машинного навчання в *IDS* є ключовим елементом для підвищення рівня захисту від сучасних загроз та забезпечення реактивності на нові типи атак. Розглянемо дві основні парадигми, які використовуються для розв'язання різних задач в машинному навчанні:

*Supervised Learning (Навчання з вчителем)* — спрямоване на розуміння зв'язку між вхідними та вихідними даними. Алгоритм, після встановлення цього зв'язку, може передбачити вихід для нових вхідних даних на основі того, що він дізнавався і зосереджується на методах класифікації та регресії. Групи класифікацій розбивають точки даних на різні класи. Цей підхід знаходить найкращій спосіб відокремити точки даних і призначити їх певним класам. Регресія відрізняється від класифікацій тим, що вона виводить число замість присвоєння точок даних класам. Класифікація фокусується на виведенні класу, тоді як регресія дає числовий вихід. Методи навчання з вчителем використовуються для виявлення відомих загроз і класифікації нових загроз за категоріями, як спам, фішинг та зловмисне програмне забезпечення [8].

*Unsupervised learning (Навчання без вчителя)* — набір даних містить лише вхідні дані та має справу з даними без, так званих, міток. Метою його є виявлення закономірностей або подібностей у наборі даних. Після отримання характеристик він групує дані на основі подібностей. Різниця від навчання з вчителем полягає в тому, що навчальний процес унікальний, оскільки алгоритм навчається на власному досвіді, а не на попередньо визначеному наборі вхідних даних із встановленим зв'язком. Методи навчання без вчителя використовуються для виявлення невідомих загроз і аномалій, які не належать до категорій відомих загроз [9, 10].

Оскільки кількість и складність кіберзагроз зростає, ці типи машинного навчання особливо корисні для виявлення загроз, тому що вони можуть ідентифікувати аномалії та закономірності, які можуть бути виявленні не відразу.

## Концепція IDS для IoT

На сьогоднішній день концепція *IDS* застосована до IoT не є чимось новим. Було розроблено і запропоновано багато рішень і систем які використовують різні підходи та технології. Відзначимо деякі системи виявлення вторгнень для IoT:

- *Cisco IoT Threat Defense:*

Запропоноване рішення від Cisco, яке використовує аналіз трафіку, машинне навчання та інтелектуальні алгоритми для виявлення аномалій в мережі. Також вони акцентують на захисті від різноманітних атак, включаючи ті, в яких використовуються віруси та зловмисні програми.

- *Darktrace Industrial:*

Спеціалізується на застосуванні технологій штучного інтелекту для виявлення відхилень від

звичайного патерну поведінки пристроїв. Їх система враховує контекст і адаптується до змін в мережі.

- *Bastille Networks:*

Спеціалізується на безпеці радіочастотного спектру для IoT пристроїв, таких як бездротові сенсори. Вони аналізують радіохвилі для виявлення аномалій та загроз.

- *Check Point IoT Protect:*

Пропонує рішення, яке включає виявлення вторгнень для IoT пристроїв. Вони використовують технології штучного інтелекту та аналіз трафіку.

- *ARM mbed OS Sexurity:*

Виходячи з назви, надає захист на рівні ОС для IoT пристроїв через свою платформу. Вони включають заходи безпеки, такі як аутентифікація та шифрування. Важливо відзначити, що ефективність кожної системи може залежати від конкретних задач та вимог використання. Для того щоб вибрати найкращу систему для конкретного випадку, потрібно ретельно ознайомитись з можливостями та різними рішеннями.

## Реалізація рішення

Після детального розгляду методів систем виявлення вторгнень було вирішено обрати *метод виявлення аномалій з* використанням технології машинного навчання. В комплексі ця система буде більш адаптивною та здатною реагувати як на старі, так і на нові, раніше невідомі, загрози.

Для навчання моделі було обрано набір даних від Канадського інституту кібербезпеки – CIC(*Canadian Institute of Cybersecurity*) IoT 2023, який був зібраний у реальному часі для масштабних атак у середовищі IoT. Це достатньо новий і розширений набір даних про атаки в IoT для сприяння розробці додатків, аналітик і безпеки.

В наборі відзначені дані з 33 атак, розділених на 7 класів (рис.1).

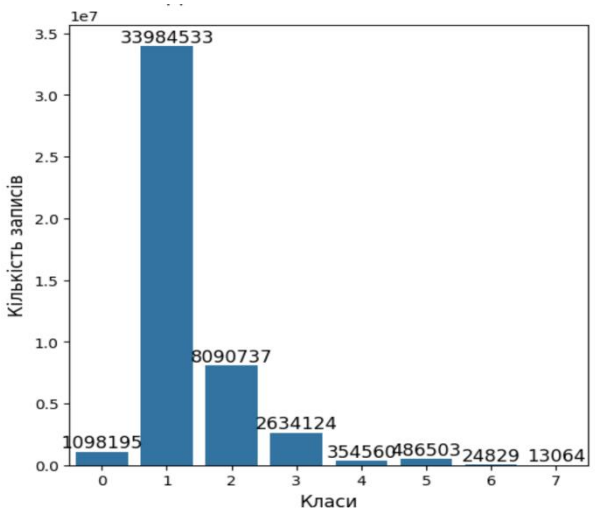


**Рис. 1** Класи атак

За допомогою нього буде навчена нейронна мережа, яка буде виконуватися в системі виявлення вторгнень, яка є метою цього документу, щоб класифікувати та виявляти мережевий трафік IoT, як зловмисний або безпечний.

Робота з підготовки даних, навчання та тестування буде проводитись у середовищі *Jupyter Notebook* на мові *Python,* версії 3.11. Набір даних розділено на піднабори, тож напочатку роботи об'єднаємо їх в один великий, це зменшить продуктивність системи з точки зору пам'яті, але дасть нам мобільності при виконанні тих чи інших операцій в процесі навчання чи підготовки до навчання. Набір даних містить 46686579 записів, 46 ознак для навчання і клас-ознаку для класифікації (рис. 2).

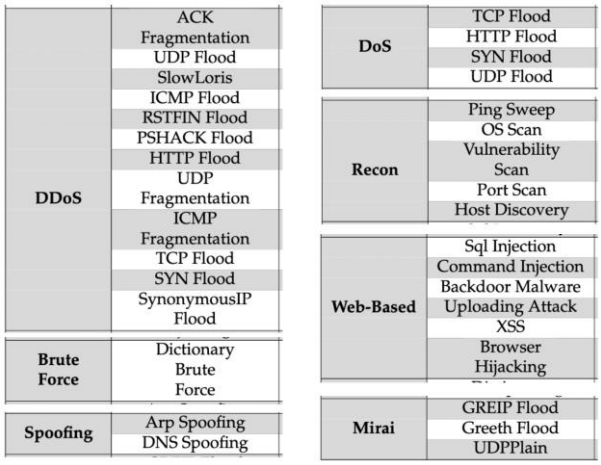


**Рис. 2.** Розподіл класів за кількістю записів:
0 – звичайний трафік, 1 – DDoS; 2 – DoS; 3 – Mirai;
4 – Recon; 5 – Spoofing; 6 – Web-Based, 7 – BruteForce

Переходимо до фази *підготовки даних,* яка є найважливішою у процесі машинного навчання, бо якість та обсяг даних безпосередньо впливають на результати роботи моделі. Очищаємо дані, видаляємо відсутні значення, скидаємо індекс нашого фрейму даних і використовуємо замість нього стандартний, видаляємо рядки, які повторюються.

Визначаються важливі ознаки(певні характеристики з набору даних), які далі будуть використовуватися для навчання моделі, нормалізуються дані, для забезпечення стабільності та швидкості навчання, розділяються на тренувальні та тестові набори. Правильна підготовка даних є критичним етапом, який може визначити невдачу чи успіх моделі в подальшому навчанні та роботою з реальними даними.

## Вибір моделі ансамблевого методу і оптимізація параметрів

В ході досліджень було вирішено використовувати модель ансамблевого методу *RandomForest.* Це метод машинного навчання, який використовується для класифікації та регресії. Він є типом і відповідає ряду класифікаторів дерева рішень на різних підвибірках набору даних. Використовує техніку випадковості і усереднення для підвищення точності прогнозування, покращення продуктивності та стабільності моделі. *RandomForest* включає в себе кілька дерев рішень, кожне з яких навчається на випадковій підмножині даних та ознак. Коли треба прийняти рішення, модель об'єднує прогнози всіх дерев, зазвичай за допомогою класифікації або середнього значення для регресії. Модель має властивість стійкості до пере-

навчання, оскільки кожне дерево навчається на випадковій підмножині даних та ознак. Це дозволяє ансамблю підтримувати генералізацію на нових, раніше не бачених ознак [10].

Перед навчанням подбаємо про гіперпараметри та оптимізуємо їх за допомогою бібліотеки *optuna.* Для моделі *RandomForest* нас цікавлять:

*max_depth* – максимальна глибина кожного дерева в ансамблі, яка визначає кількість рівнів у дереві рішень.

*max_features* – визначає максимальну кількість ознак, які випадково обираються для розгляду при побудові кожного дерева в "лісі".

*n_estimators* – гіперпараметр який вказує кількість дерев, які мають бути побудовані в ансамблі.

Коли гіперпараметри визначені і оптимізацію завершено, починається навчання, яке буде займати деякий час. Для оцінки ефективності моделі використовуємо метрики: *Accuracy*, *Precision, Recall, F score,* які враховують різні показники результатів класифікацій та дозволяють отримати більш повну картинку продуктивності моделі. Розберемо більш детально кожний з них:

*Accuracy (правильність)* – частка прогнозів, яку наша модель отримала правильно. Математично це співвідношення між кількістю правильних пронозів до загальної кількості прогнозів. Це корисно, коли всі класи мають однакову важливість, як у нашому випадку але є недолік зі сторони незбалансованого набіру даних.

*Precision (точність)* - це співвідношення $\frac{TP}{TP+FP}$, де *TP* – кількість справжніх спрацьовувань, а *FP* – кількість хибних спрацьовувань [11]. Точність – це інтуїтивно зрозуміла здатність класифікатора не позначати негативний зразок як позитивний.

*Recall (запам'ятовування)* – це відношення $\frac{TP}{TP+FN}$, де *TP* – кількість справжніх позитивних результатів, а *FN* –помилкових негативних результатів.

З*апам'ятовування* – це інтуїтивно зрозуміла здатність класифікатора знаходити всі позитивні зразки [11].

F-оцінка -може бути інтерпретована як зважене гармонічне середнє значення точності та запам'ятовування, досягає найкращого значення при 1, а найгіршого при 0. Визначається як:

$$F_1 = \frac{2}{\frac{1}{recall} \times \frac{1}{precision}} = 2 \times \frac{precision \times recall}{precision + recall}.$$

За підсумком навчання маємо результати, наведені в табл. 1, 2 та рис. 3.

*Таблиця 1* – **Кількісна оцінка якості**

| | **Precision** | **Recall** | **F1-Score** | **Support** |
|---|---|---|---|---|
| 0 | 0.91 | 0.98 | 0.94 | 878773 |
| 1 | 1.00 | 1.00 | 1.00 | 27188627 |
| 2 | 1.00 | 1.00 | 1.00 | 6470759 |
| 3 | 1.00 | 1.00 | 1.00 | 2107532 |
| 4 | 0.91 | 0.84 | 0.87 | 283896 |
| 5 | 0.92 | 0.86 | 0.89 | 389289 |
| 6 | 0.98 | 0.54 | 0.70 | 19893 |
| 7 | 0.99 | 0.57 | 0.72 | 10467 |
| | | | | |
| accuracy | | | 1.00 | 37349236 |
| macro avg | 0.96 | 0.85 | 0.89 | 37349236 |
| weighted avg | 1.00 | 1.00 | 1.00 | 37349236 |

*Таблиця 2* **– Метрики ефективності моделі**

| **Train Score** | **Test Score** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
|---|---|---|---|---|---|
| 0.9974162791442 | 0.9963210760187 | 0.9963210760187 | 0.9646470264108 | 0.8473565625810 | 0.8903067510023 |

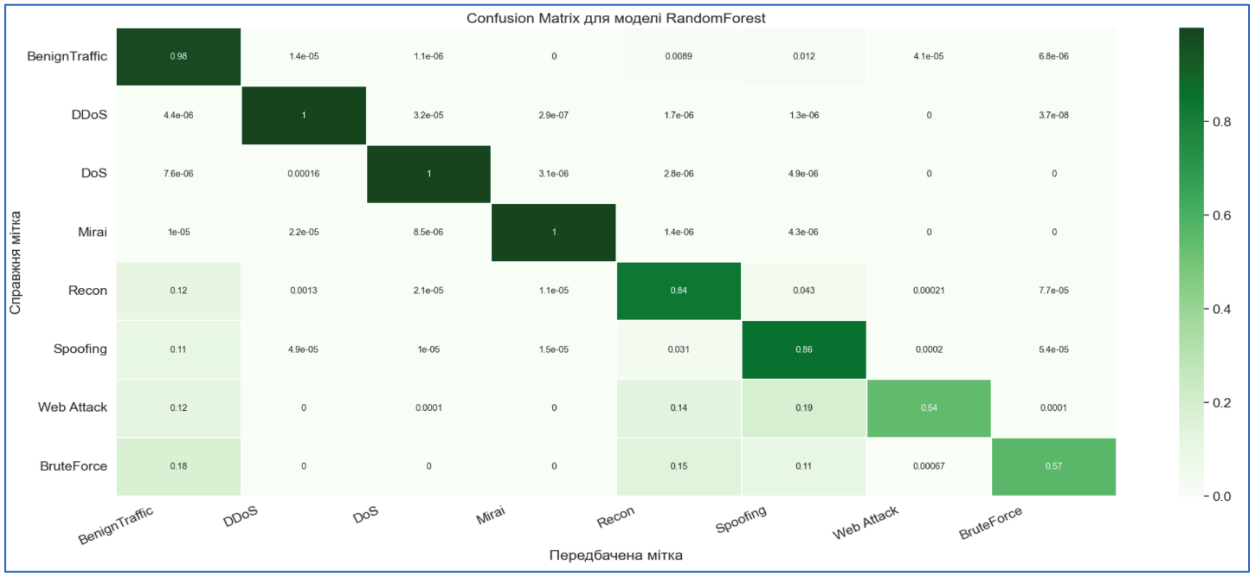


**Рис. 3.** Матриця помилок для моделі *Random Forest*

Після тестування модель зберігається локально за допомогою бібліотеки *pickle.* На виході модель *RandomForest,* це набір дерев-предикторів $\{t(x_{in}, \theta_n), n = 1, \dots\}$, які індивідуально роблять передбачення на заданому параметрі $x_{in}$. Кожен предиктор залежить від випадкового набору змінних $\{\theta_n\}$, які незалежно відбираються з однаковим розподілом (рис. 4).

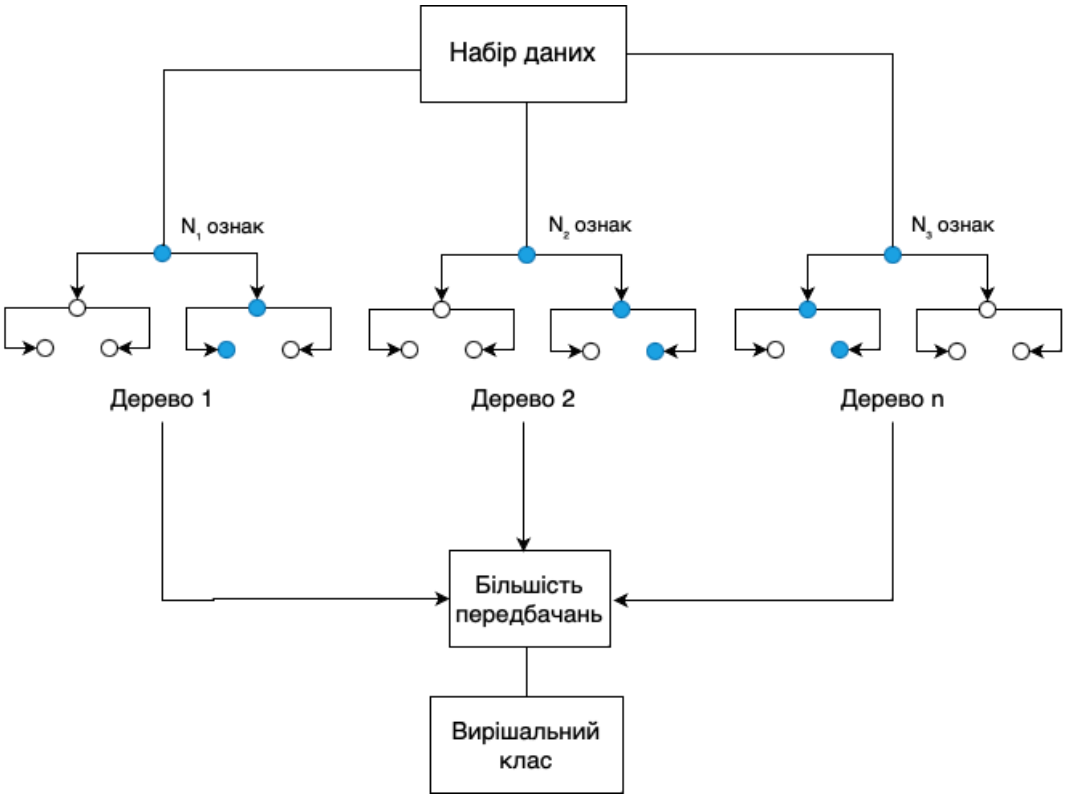


**Рис. 4** Архітектура моделі *Random Forest*

## Реалізація системи виявлення вторгнень з нейронною мережею (рис. 5)

Додаток буде в реальному часі зчитувати інтернет пакети або *.pcap* файли, діставати з них усі необхідні ознаки та відправляти моделі для отримання передбачення. За допомогою бібліотеки *pyshark,* для захоплення та аналізу інтернет-пакетів, витягуємо необхідні ознаки для подальшого використання. Після того як всі ознаки витягнуті відправляємо їх до попередньо навченої моделі для отримання передбачень. Обробляємо результати передбачень та приймаємо рішення щодо подальших дій, сповіщення або вжиття інших заходів безпеки. Було вирішено розробити інтерфейс командного рядка (*CLI – Command-line interface,* рис. 6). Користувач зможе встановити цей додаток за допомогою системи керування пакунками (*pip*) на операційну систему Windows або класу Linux. Він матиме змогу запускати його виконання з командного рядка або додати в автоматичний запуск за допомогою скриптів.


**Рис. 5** Архітектура IDS

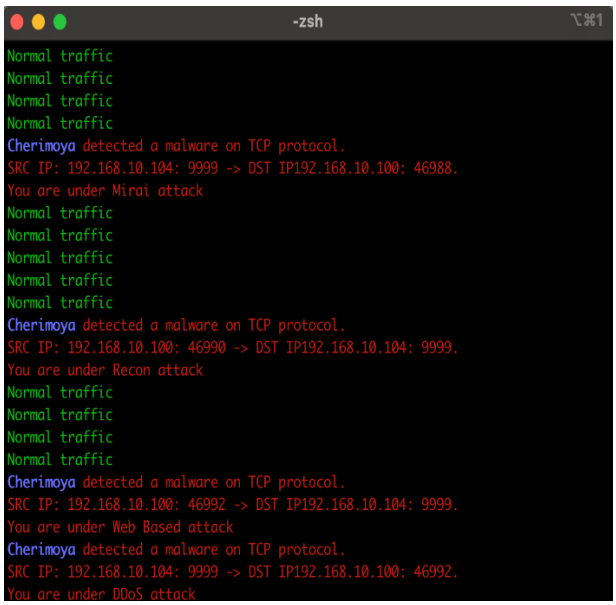


**Рис. 6** Інтерфейс командного рядка (*CLI)* застосунку Cherimoya (*назва системи виявлення вторгнень)*

## Висновки

У наслідок проведених досліджень і розглянутих концепцій передових систем та методів виявлення атак в інфраструктурі IoT було запропоновано власну систему виявлення вторгнень в реальному часі, яка базується на методі виявленні аномалій, та працює в комплексі з нейронної мережею ансамблевого методу *RandomForest,* яка була навчена на наборі з великим обсягом даних та має гарні показники:

*правильність – 0.996;*

*точність – 0.964;*

*запам'ятовування – 0.847;*

г*армонічне середнє значення точності та запам'ятовування – 0.89.*

Для зручності використання застосунку системи був розроблений інтерфейс командного рядка, котрий сповіщає користувача або іншу систему про вторгнення чи атаку. В майбутньому запропонована модель може бути використана для систем побудованих в поєднанні з концепцією *туманного обчислення* та Інтернету речей, на принципах Fog-IoT архітектури.

Список літератури

1. T. Mazhar, D. B. Talpur, T. Al Shloul, Y. Y. Ghadi, I. Haq, I. Ullah, K. Ouahada, H. Hamam Analysis of IoT Security Challenges and Its Solutions Using Artificial Intelligence. 2023. DOI: https://doi.org/10.3390%2Fbrainsci13040683
2. A. Khraisat, I. Gondal, P. Vamplew, J. Kamruzzaman Survey of intrusion detection systems: techniques, datasets and challenges. 2019. URL: https://cybersecurity.springeropen.com/articles/10.1186/s42400-019-0038-7
3. Рубан І. В. Класифікація методів виявлення аномалій в інформаційних системах / І. В. Рубан, В. О. Мартовицький, С. О. Партика // Системи озброєння і військова техніка. — 2016. — № 3. — С. 100-105
4. Verma Abhisher, Virender Ranga Machine learning based intrusion detection systems for IoT applications. 2020. URL: https://link.springer.com/article/10.1023/A:1010933404324
5. Коваленко А. А., Кучук Г. А. Методи синтезу інформаційної та технічної структур системи управління об'єктом критичного застосування. *Сучасні інформаційні системи*. 2018. Т. 2, № 1. С. 22–27. DOI: https://doi.org/10.20998/2522-9052.2018.1.04
6. Ruban, I.V., Martovytskyi, V.O., Kovalenko, A.A. and Lukova-Chuiko, N.V. (2019), "Identification in Informative Systems on the Basis of Users' Behaviour", Proceedings of the International Conference on Advanced Optoelectronics and Lasers, CAOL 2019-September,9019446, pp. 574-577, DOI: https://doi.org/10.1109/CAOL46282.2019.9019446
7. Kovalenko, A., Kuchuk, H., Kuchuk, N. and Kostolny, J. (2021), "Horizontal scaling method for a hyperconverged network", *2021 International Conference on Information and Digital Technologies* (IDT), Zilina, Slovakia, doi: https://doi.org/10.1109/IDT52577.2021.9497534
8. J. Delua Supervised vs. Unsupervised learning. 2021. URL: https://www.ibm.com/blog/supervised-vs-unsupervised-learning/
9. 1. I. U. Khan, M. Ouaissa, M. Ouaissa, Z. A. El Houda, M. Fazal Cyber Security for Next-Generation Computing. 2024. DOI: https://doi.org/10.1201/9781003404361
10. . Журило, О., Ляшенко, О. і Аветісова, К. 2023. ОГЛЯД РІШЕНЬ З АПАРАТНОЇ БЕЗПЕКИ КІНЦЕВИХ ПРИСТРОЇВ ТУМАННИХ ОБЧИСЛЕНЬ У ІНТЕРНЕТІ РЕЧЕЙ. СУЧАСНИЙ СТАН НАУКОВИХ ДОСЛІДЖЕНЬ ТА ТЕХНОЛОГІЙ В ПРОМИСЛОВОСТІ. 1 (23) (Квіт 2023), 57–71. DOI: https://doi.org/10.30837/ITSSI.2023.23.057.
11. V. Martovytskyi, I. Ruban, H. Lahutin, I. Ilina, V. Rykun and V. Diachenko, "Method of Detecting FDI Attacks on Smart Grid," 2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T), Kharkiv, Ukraine, 2020, pp. 132-136, doi: 10.1109/PICST51311.2020.9468005

Received (Надійшла) 05.11.2023
Accepted for publication (Прийнята до друку) 24.01.2024

**Model and methods of detection of a large-scale attack in the IoT environment**

Oleksii Liashenko, Ihor Velykodnyi, Vasyl Znaidiuk, Oleh Zhurylo

**Abstract**. The main concept and subject of the study is the detection of various types of extensive attacks in the IoT infrastructure, an overview of the presented model, methods and existing advanced intrusion detection systems. **The purpose of this work** is to propose a real-time intrusion detection system that will be trained on a large data set using a neural network using an ensemble machine learning **method**. **The subject of the research** is an overview of existing methods and models for detecting a large-scale attack and proposing an intrusion detection system solution, which will be based on the method of detecting anomalies and a neural network. **Conclusion**. An intrusion detection system was built, which analyzes Internet traffic, extracts signs from the packet, processes them and predicts various types of attacks, as well as characterizes them by type. Security threat can be considered as the main critical issue for IoT devices, so the use of such systems reduces the risks of data loss.

**Keywords:** dataset, neural network, machine learning, network traffic, IDS, training, prediction, anomaly detection, attack, Internet of Things, IoT, intrusion detection system.