

**Давлетова Аліна Ярославівна**

викладач кафедри кібербезпеки

Західноукраїнський національний університет, Тернопіль, Україна

ORCID: 0000-0002-1192-2532

a7davletova@gmail.com

АДАПТИВНА МОДЕЛЬ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ГІБРИДНИХ СИСТЕМ ВИЯВЛЕННЯ ВТОРГНЕНЬ

Анотація. Зростання складності кіберзагроз, а також обмеження традиційних систем виявлення вторгнень, заснованих виключно на сигнатурному аналізі або ймовірнісних моделях машинного навчання, зумовлюють необхідність розробки комплексних адаптивних моделей прийняття рішень. Незважаючи на те, що методи машинного навчання можуть забезпечувати високу чутливість до аномалій у мережевому трафіку, їх застосування може супроводжуватися хибними спрацюваннями та недостатньою інтерпретованістю результатів. Окремою проблемою є узгоджене об'єднання різнорідних індикаторів безпеки, які мають різну природу, рівень достовірності та критичності, зокрема криптографічних індикаторів, часових і поведінкових характеристик та історичних оцінок ризику. У роботі запропоновано адаптивну модель прийняття рішень для гібридної системи виявлення вторгнень, що забезпечує об'єднання різнорідних ознак безпеки в єдиний механізм оцінювання ризику. Перший рівень моделі реалізовано на основі детермінованого криптографічного контролю, що забезпечує пріоритетне виявлення критичних порушень. У разі виявлення критичного криптографічного порушення система формує рішення про вторгнення незалежно від результатів інших компонентів. За відсутності такого порушення виконується адаптивне зважене об'єднання нормалізованої криптографічної оцінки, ймовірнісної оцінки модуля машинного навчання на основі Random Forest та накопиченого історичного ризику з подальшим пороговим формуванням підсумкового рішення. Розроблено програмний прототип та проведено серію експериментальних досліджень у розподіленому середовищі на вибірці з 1099 тестових сеансів, що охоплювали нормальні, підозрілі, атакуючі, комбіновані та раніше не представлені в навчанні сценарії. Отримані результати засвідчили, що інтегрована модель забезпечує макросереднє значення площі під ROC-кривою Macro AUC = 0,963, загальну точність класифікації 88,81 %, а значення Macro Precision, Macro Recall та Macro F1-score становлять відповідно 0,89, 0,88 та 0,88. При цьому жоден атакуючий сценарій не був класифікований як нормальний стан. Запропонована модель розширює можливості традиційних систем виявлення вторгнень шляхом поєднання детермінованого криптографічного контролю з адаптивним оцінюванням ризику на основі машинного навчання та історії поведінки вузлів. Такий підхід дає змогу враховувати як критичні порушення цілісності, так і слабкі накопичувані аномалії, що визначає можливість застосування запропонованої моделі у розподілених інформаційно-комунікаційних та кіберфізичних системах із підвищеними вимогами до безпеки.

Ключові слова: системи виявлення вторгнень, гібридна система виявлення вторгнень, адаптивна модель прийняття рішень, криптографічні індикатори, машинне навчання, історичний ризик, оцінювання ризику.

ВСТУП

Постановка проблеми. Сучасні системи виявлення вторгнень (IDS) функціонують в умовах постійного зростання складності кіберзагроз, що супроводжується використанням комбінованих атак, прихованих механізмів компрометації та багаторівневих сценаріїв втручання. Традиційні підходи до виявлення вторгнень, засновані виключно на сигнатурному аналізі, статистичних характеристиках трафіку або методах машинного навчання, часто демонструють недостатню ефективність під час роботи з неоднорідними джерелами ознак безпеки. Особливо це проявляється у розподілених системах, де атака може впливати не лише на вміст переданих даних, а й на часові характеристики маршруту, процеси повторного шифрування, поведінку вузлів або накопичення аномалій у часі.



Сучасні гібридні IDS використовують як механізми прийняття рішень на основі правил, так і ймовірнісні моделі машинного навчання та їх комбінування. У першому випадку забезпечується висока надійність виявлення критичних порушень цілісності, однак система має обмежену здатність адаптуватися до складних або частково прихованих атак. У другому випадку використання виключно методів машинного навчання дає змогу виявляти аномалії поведінки системи, проте може супроводжуватися підвищеною кількістю хибних спрацювань та нестабільністю результатів у разі зміни характеристик середовища.

Особливу складність становить необхідність об'єднання різномірних індикаторів безпеки, які мають різну природу, рівень достовірності та критичності. Зокрема, результати криптографічного контролю можуть мати детермінований характер і безпосередньо вказувати на порушення цілісності повідомлення, тоді як результати машинного навчання мають ймовірнісну природу, а часові та історичні характеристики потребують додаткового оцінювання й узгодженого поєднання.

Тому розроблення адаптивної моделі прийняття рішень, здатної поєднувати різномірні індикатори безпеки в єдиний механізм оцінювання ризику та формування підсумкового рішення про стан системи, є актуальним завданням.

Аналіз останніх досліджень і публікацій.

Для обґрунтування актуальності та наукової новизни запропонованої адаптивної моделі прийняття рішень проведено аналіз наукових праць у галузі систем виявлення вторгнень, зокрема досліджень із застосуванням методів машинного та глибокого навчання, гібридних моделей, пояснюваного штучного інтелекту та криптографічних механізмів захисту.

У роботі [1] проведено систематичний аналіз методів машинного та глибокого навчання, що застосовуються для побудови мережевих систем виявлення вторгнень. Авторами систематизовано основні підходи до використання машинного та глибокого навчання, проаналізовано їх переваги й обмеження, застосовувані набори даних та показники оцінювання. Окрему увагу приділено проблемам підвищення точності виявлення вторгнень, зменшення кількості хибних спрацювань і виявлення раніше невідомих атак.

У роботі [2] запропоновано набір даних NSL-KDD як удосконалену версію KDD Cup 99 для дослідження систем виявлення вторгнень. Авторами усунуто дубльовані записи та сформовано навчальні й тестові вибірки, що дають змогу коректніше оцінювати ефективність різних методів класифікації. NSL-KDD може використовуватися як еталонний набір даних для порівняння алгоритмів виявлення вторгнень та оцінювання моделей машинного навчання.

У роботі [3] досліджено застосування методів штучного інтелекту для виявлення вторгнень у середовищах Інтернету речей з урахуванням співвідношення між якістю виявлення та обчислювальними витратами. Авторами проведено порівняльне оцінювання гібридної моделі CNN-BiLSTM, Random Forest та XGBoost на наборах даних NSL-KDD і UNSW-NB15. Встановлено, що CNN-BiLSTM забезпечує найвищі показники виявлення, однак характеризується більшими часовими витратами на класифікацію. XGBoost і Random Forest забезпечують значно менші витрати часу на класифікацію, що робить їх придатнішими для систем з обмеженими обчислювальними ресурсами. Продемонстровано наявність компромісу між якістю виявлення атак і обчислювальною ефективністю.

У роботі [4] запропоновано гібридний підхід до виявлення вторгнень, що поєднує метод опорних векторів, кластеризацію та метод головних компонент. Експериментальні результати показали доцільність комбінованого використання різних методів аналізу для підвищення ефективності виявлення вторгнень.

У роботі [5] проведено дослідження моделей виявлення вторгнень із використанням методів попередньої обробки даних та алгоритмів класифікації. Для експериментального оцінювання застосовано Decision Tree, Random Forest, XGBoost і Naive Bayes, а також набори даних NSL-KDD та CICIDS2017. Досліджено вплив попередньої обробки даних із застосуванням методу головних компонент (PCA) на результати класифікації. Встановлено, що Random Forest забезпечує високі показники точності, тоді як застосування PCA-орієнтованої схеми попередньої обробки не в усіх випадках сприяє покращенню результатів.

У роботі [6] запропоновано гібридний підхід до виявлення вторгнень, що поєднує методи машинного та глибокого навчання. У дослідженні використовуються моделі CNN-LSTM та XGBoost-LSTM, а експериментальне оцінювання здійснюється на наборах даних CICIDS2017, UNSW-NB15, NSL-KDD та WSN-DS. Для скорочення простору ознак використовуються CNN та XGBoost. Отримані результати показали високу ефективність запропонованих гібридних моделей для двокласової та багатокласової класифікації мережевого трафіку. Водночас автори відзначають вплив дисбалансу класів і складність виявлення окремих рідкісних типів атак.

У роботі [7] досліджено застосування методів LIME та SHAP для пояснення результатів роботи багатоваріантного перцептрона у задачах виявлення вторгнень. Запропонований підхід дає змогу визначати



внесок окремих ознак у рішення моделі та формувати пояснення для результатів класифікації мережевого трафіку. Використання таких методів сприяє підвищенню прозорості функціонування систем виявлення вторгнень.

У роботі [8] запропоновано концепцію інтерпретованої системи виявлення вторгнень на основі машинного навчання. Основну увагу приділено проблемі пояснення рішень складних моделей, оскільки відсутність зрозумілого обґрунтування результатів класифікації ускладнює їх практичне використання фахівцями з кібербезпеки. Запропонований підхід забезпечує можливість аналізу причин, на підставі яких модель відносить мережевий трафік до певного класу.

У роботі [9] досліджено застосування пояснюваного штучного інтелекту в системах виявлення вторгнень. Розглянуто проблему непрозорості моделей машинного навчання та можливість використання методів пояснення для визначення впливу окремих ознак на результати класифікації. Показано, що застосування методів пояснюваного штучного інтелекту дає змогу зробити результати роботи систем виявлення вторгнень зрозумілішими для фахівців з кібербезпеки.

У роботі [10] систематизовано підходи до пояснення результатів роботи моделей штучного інтелекту, проаналізовано їх переваги та обмеження, а також визначено проблеми, пов'язані з інтерпретованістю, надійністю та довірою до таких систем. Окрему увагу приділено перспективам розвитку пояснюваних систем виявлення вторгнень і питанням їх практичного застосування у сфері кібербезпеки.

У роботі [11] досліджено мережеву систему виявлення вторгнень на основі Random Forest з використанням методу перестановочної важливості ознак. Експериментальне оцінювання виконано на наборі даних CICIDS2017. Проведено відбір найбільш інформативних характеристик мережевого трафіку та проаналізовано їх вплив на результати класифікації. Показано, що скорочення кількості ознак із збереженням найбільш значущих характеристик дає змогу забезпечити високі показники виявлення атак і водночас підвищити зрозумілість роботи моделі.

Проведений аналіз наукових праць показує, що сучасні системи виявлення вторгнень активно розвиваються в напрямі застосування методів машинного та глибокого навчання, гібридних моделей та комбінованого використання різних методів аналізу [1, 3–6]. Для їх порівняльного оцінювання використовуються еталонні набори даних, зокрема NSL-KDD [2]. Окремим напрямом є відбір найбільш інформативних ознак мережевого трафіку, що дає змогу зменшити складність моделей і підвищити обґрунтованість рішень [11]. Водночас актуальною залишається проблема інтерпретованості результатів роботи складних моделей. Методи пояснюваного штучного інтелекту дають змогу оцінювати вплив окремих ознак на результати класифікації та підвищувати прозорість систем виявлення вторгнень [7–10].

Більшість проаналізованих досліджень зосереджена на підвищенні точності виявлення атак, удосконаленні класифікаційних моделей, відборі ознак та інтерпретованості рішень. Водночас недостатньо дослідженим залишається узгоджене поєднання результатів детермінованого криптографічного контролю, ймовірнісних оцінок машинного навчання та накопиченого історичного ризику в межах єдиного адаптивного механізму прийняття рішень.

Мета статті. Розробка та дослідження адаптивної моделі прийняття рішень для гібридної системи виявлення вторгнень, що забезпечує об'єднання криптографічних, часових, поведінкових та історичних ознак безпеки в єдиний механізм оцінювання ризику.

АРХІТЕКТУРА ТА МАТЕМАТИЧНА МОДЕЛЬ ПРИЙНЯТТЯ РІШЕНЬ

У попередніх дослідженнях [12, 13] було сформовано концепцію гібридної системи захисту, що поєднує криптографічні механізми контролю з методами штучного інтелекту для виявлення мережевих атак. Подальший розвиток цієї концепції потребує узгодженого механізму прийняття рішень на основі різномірних індикаторів безпеки, які відрізняються за природою, рівнем достовірності та критичності.

У запропонованій моделі джерела інформації розподілено за їх функціональною роллю у процесі оцінювання стану системи. Криптографічні індикатори характеризують результати контролю цілісності та коректності опрацювання повідомлення; модуль машинного навчання формує ймовірнісну оцінку стану на основі сукупності поточних ознак сеансу; історичний компонент відображає накопичені відхилення, зафіксовані у попередніх сеансах та поведінці вузлів. Модель використовує три основні групи інформаційних ознак:

- криптографічні індикатори C ;
- результати оцінювання методами машинного навчання M ;
- історичні характеристики поведінки вузлів H .

Загальну архітектуру запропонованої моделі наведено на рис. 1.

Архітектура побудована за ієрархічним принципом. На початковому етапі виконується криптографічний контроль, результати якого використовуються двома способами. Критичні порушення, пов'язані з неможливістю коректного відновлення повідомлення, помилками дешифрування або іншими невірними станами, розглядаються як підстава для формування пріоритетного детермінованого рішення про вторгнення. Водночас не критичні криптографічні ознаки формують окрему оцінку стану, яка може бути використана на наступному рівні аналізу.

Якщо критичного криптографічного порушення не зафіксовано, система переходить до інтегрованого оцінювання. На цьому етапі криптографічна оцінка, ймовірнісний результат модуля машинного навчання та накопичений історичний ризик приводяться до узгодженого подання та об'єднуються з використанням вагових коефіцієнтів. Внесок окремих компонентів може змінюватися залежно від поточного стану системи та накопиченої інформації про попередні сеанси.

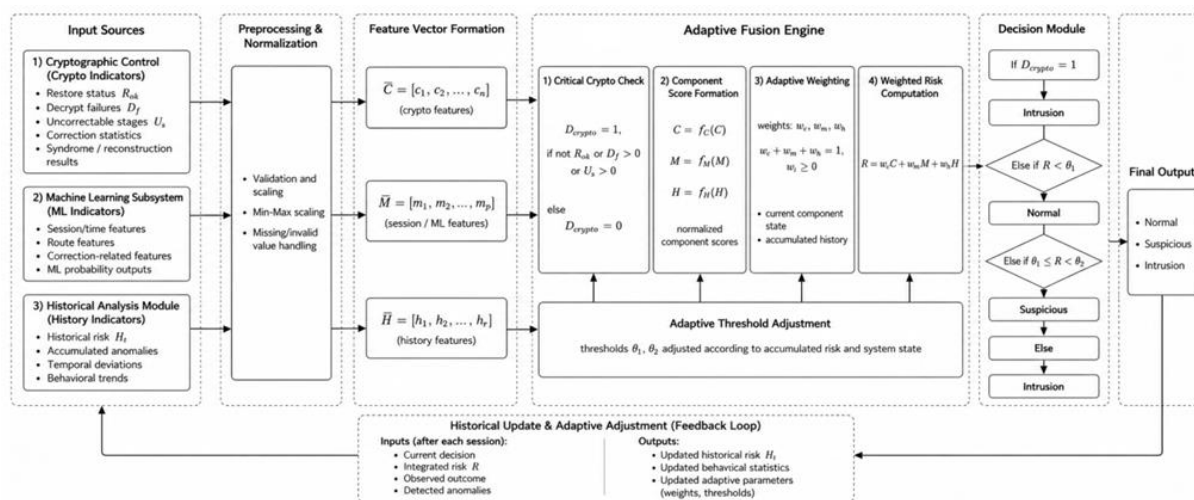


Рисунок 1 – Архітектура адаптивної моделі прийняття рішень

Часові та поведінкові характеристики не формують самостійних складових інтегрованого ризику. Вони використовуються під час побудови ознак поточного сеансу та історичного аналізу, впливаючи відповідно на ймовірнісну оцінку модуля машинного навчання та накопичену оцінку ризику.

Після формування інтегрованого показника ризику виконується його порівняння з пороговими значеннями, на основі чого система визначає один із трьох станів: нормальний, підозрілий або вторгнення. Результати поточного сеансу використовуються для оновлення історичної складової, що забезпечує врахування повторюваних аномалій у подальших циклах прийняття рішення.

Під час кожного сеансу передавання даних криптографічний модуль формує сукупність показників, що характеризують результати опрацювання та відновлення повідомлення. До основних показників належать успішність відновлення повідомлення, кількість помилок дешифрування та наявність невірних стадій опрацювання. Додатково враховуються кількість виправлених помилок, результати аналізу синдрому та характеристики корекції.

Сукупність зазначених показників подається у вигляді вектора криптографічних ознак:

$$\bar{C} = [c_1, c_2, \dots, c_n],$$

де c_i – окремий криптографічний показник поточного сеансу.

Для використання криптографічних ознак у подальшому механізмі об'єднання формується скалярна оцінка криптографічного ризику

$$C = f_C(\bar{C}), \quad C \in [0,1],$$

де $f_C(\cdot)$ – функція агрегування та нормалізації криптографічних ознак. Значення C характеризує ступінь відхилення результатів криптографічного опрацювання від нормального стану та використовується як одна зі складових інтегрованої оцінки ризику.

Для формалізації критичного криптографічного стану введено бінарний індикатор:



$$D_{crypto} = \begin{cases} 1, \neg R_{ok} \vee D_f > 0 \vee U_s > 0, \\ 0, \text{інакше.} \end{cases}$$

де R_{ok} – ознака успішного відновлення повідомлення; D_f – кількість помилок дешифрування; U_s – кількість невірних стадій опрацювання.

Значення $D_{crypto} = 1$ відповідає критичному криптографічному порушенню. У цьому випадку система формує рішення про вторгнення без виконання подальшого імовірнісного об'єднання результатів інших компонентів:

$$D_{crypto} = 1 \Rightarrow \text{Decision} = \text{Intrusion.}$$

Такий механізм реалізує принцип пріоритетності критичних криптографічних індикаторів під час формування підсумкового рішення.

Якщо критичного криптографічного порушення не виявлено, $D_{crypto} = 0$ система переходить до наступного рівня аналізу. Криптографічна оцінка C передається до модуля адаптивного зваженого об'єднання.

Модуль машинного навчання виконує ймовірнісне оцінювання стану поточного сеансу за сукупністю ознак, що характеризують процес передавання та опрацювання даних. У запропонованій моделі для цього використовується Random Forest, навчений на множині сценаріїв функціонування системи.

Вхідні дані подаються у вигляді вектора ознак:

$$\bar{M} = [m_1, m_2, \dots, m_p],$$

де m_i – окрема характеристика поточного сеансу. До складу $\bar{M}$ входять часові характеристики передавання та опрацювання даних, статистичні показники змін і корекції символів, характеристики маршруту, а також показники відхилення фактичних затримок від установлених базових значень.

На основі вектора $\bar{M}$ модель формує ймовірнісні оцінки належності поточного сеансу до трьох можливих станів:

$$P_{ML} = (P_N, P_S, P_I) = M_{RF}(\bar{M}),$$

де M_{RF} – навчена модель Random Forest; P_N – ймовірність нормального стану; P_S – ймовірність підозрілого стану; P_I – ймовірність вторгнення.

Для використання результатів машинного навчання в механізмі адаптивного об'єднання формується узагальнена оцінка:

$$M = \lambda_N P_N + \lambda_S P_S + \lambda_I P_I,$$

де $\lambda_N, \lambda_S, \lambda_I$ – вагові коефіцієнти, що визначають внесок відповідно ймовірності вторгнення та підозрілого стану в підсумкову оцінку модуля машинного навчання, причому

$$0 \leq \lambda_N < \lambda_S < \lambda_I \leq 1.$$

Для нормального стану може бути прийнято $\lambda_N = 0$, оскільки він не збільшує оцінку ризику.

Значення M має імовірнісну природу та використовується як одна зі складових інтегрованого показника ризику. При цьому P_I відображає вираженість ознак вторгнення, а P_S дає змогу враховувати проміжні стани, для яких наявні ознаки відхилення від нормальної поведінки, але їх сукупність не є достатньою для безпосереднього формування рішення про вторгнення.

Модуль історичного аналізу призначений для врахування аномальних проявів, що повторюються в послідовності сеансів і можуть не мати достатньої вираженості для однозначного виявлення вторгнення в межах одного сеансу. Накопичення таких відхилень дає змогу виявляти стійкі зміни поведінки системи та підвищувати чутливість моделі до поступового розвитку аномальної активності.

Під час історичного аналізу враховуються повторювані порушення криптографічного опрацювання, часові відхилення, результати попередніх класифікацій та інші характеристики, накопичені за попередні сеанси. Сукупність таких показників для моменту t подається у вигляді вектора історичних ознак:



$$\bar{H} = [h_1(t), h_2(t), \dots, h_r(t)],$$

де $h_j(t)$ – окрема історична характеристика, сформована за результатами попередніх спостережень;

$$A_t = f_H(\bar{H}_t),$$

де $f_H(\cdot)$ – функція агрегування історичних ознак; A_t – поточна оцінка аномалії.

Для врахування попередньо накопиченого ризику використовується рекурсивна модель

$$H_t = \gamma H_{t-1} + (1 - \gamma) A_t,$$

де H_t – оцінка історичного ризику на поточному кроці; H_{t-1} – оцінка історичного ризику за попередній сеанс; $\gamma \in [0,1]$ – коефіцієнт пам'яті, що визначає внесок попередньої історії у формування поточної оцінки.

Зі збільшенням частоти та інтенсивності повторюваних аномалій значення H_t зростає, унаслідок чого посилюється внесок історичного компонента в подальше формування інтегрованої оцінки ризику. За відсутності стійких аномальних проявів вплив попередніх відхилень поступово зменшується.

Оскільки окремі групи ознак та компонентні оцінки мають різні масштаби значень, перед виконанням аналізу необхідна процедура нормалізації:

$$x_i^{norm} = \frac{x_i - x_i^{min}}{x_i^{max} - x_i^{min}}.$$

де x_i – поточне значення i -го показника; x_i^{min} та x_i^{max} – відповідно мінімальне та максимальне значення цього показника в установленому діапазоні спостережень; $x_i^{norm} \in [0,1]$ – нормалізоване значення.

Нормалізація забезпечує приведення різнорідних оцінок до порівнюваного масштабу, запобігає необґрунтованому домінуванню окремого компонента через особливості його числового діапазону та створює узгоджену основу для подальшого адаптивного зваженого об'єднання.

Фінальне рішення у системі формується за допомогою адаптивного зваженого об'єднання результати усіх модулів C , M та H . Алгоритм функціонування запропонованої адаптивної моделі прийняття рішень наведено на рис. 2.

Інтегрований показник ризику визначається як:

$$R = w_c C + w_m M + w_h H,$$

де R – інтегрована оцінка ризику; C, M, H – нормалізовані значення відповідних груп ознак; w_i – відповідні вагові коефіцієнти.

Для вагових коефіцієнтів виконуються умови:

$$w_c + w_m + w_h = 1, w_c + w_m + w_h \geq 0.$$

На відміну від статичного зважування, у запропонованій моделі співвідношення між ваговими коефіцієнтами може змінюватися відповідно до поточного стану компонентів та інформації, накопиченої за попередні сеанси. Це дає змогу змінювати внесок окремих джерел інформації у формування інтегрованої оцінки без зміни загальної структури моделі.

Для класифікації поточного стану системи S використовуються два адаптивні порогові значення θ_1 і θ_2 , які визначають межі між нормальним, підозрілим станом і вторгненням у межах нормалізованої шкали ризику. Їх початкові значення визначаються на етапі калібрування системи за еталонними сценаріями її нормального функціонування. Для порогових значень виконується умова:

$$0 \leq \theta_1 < \theta_2 \leq 1.$$

Порогові значення можуть коригуватися адаптивним механізмом відповідно до зміни стану системи та накопичених результатів попередніх сеансів. Остаточне рішення визначається з урахуванням результату детермінованого криптографічного контролю:

$$Decision = \begin{cases} \text{Normal}, & R < \theta_1 \\ \text{Suspicious}, & \theta_1 \leq R < \theta_2, \\ \text{Intrusion}, & R \geq \theta_2 \end{cases}$$

де Normal – нормальний режим; Suspicious – підозрілий стан; Intrusion – виявлене вторгнення.

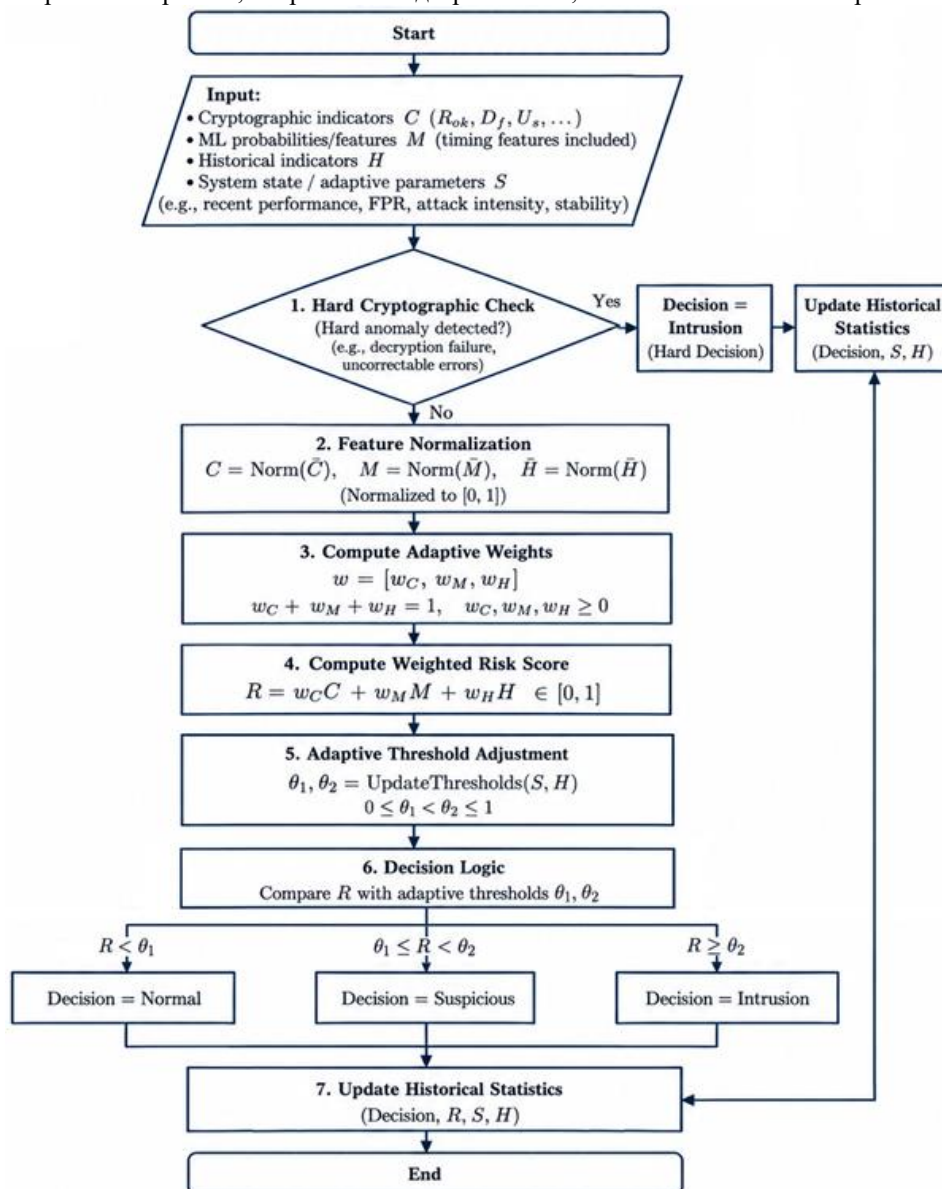


Рисунок 2 – Адаптивний алгоритм прийняття рішень

Після формування рішення результати поточного сеансу використовуються для оновлення історичної складової та параметрів адаптивного механізму, що впливають на наступний цикл оцінювання.

Запропонована адаптивна модель забезпечує багаторівневе прийняття рішень із використанням детермінованих та ймовірнісних індикаторів безпеки й адаптивної зміни порогів класифікації, що підвищує ефективність IDS при роботі з комбінованими та частково прихованими атаками.

ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ

Експериментальне дослідження проводилося на основі програмної реалізації запропонованої гібридної IDS. Метою дослідження була оцінка ефективності запропонованої адаптивної моделі прийняття рішень у задачі класифікації станів системи. Дослідження виконувалося у розподіленому середовищі передачі даних, що складалося з центрального вузла та чотирьох проміжних агентних вузлів



маршрутизації. Під час кожного тестового сеансу реєструвалися криптографічні показники, результати виявлення та корекції помилок, характеристики маршруту, часові параметри передачі даних, оцінки модуля машинного навчання та значення накопиченого історичного ризику. Експериментальна вибірка формувалася автоматично під час виконання тестових сеансів. Для відтворення різних умов функціонування використовувалися параметризовані профілі часових затримок, криптографічних порушень та їх комбінацій. Еталонний клас кожного сценарію задавався відповідно до умов його генерації незалежно від результату, сформованого адаптивною моделлю. Характеристики експериментальних сценаріїв наведено в табл. 1

Таблиця 1

Характеристика експериментальних сценаріїв гібридної IDS

Сценарій	Кількість сеансів	Частка, %	Характеристика
normal only	316	28,75	режим функціонування без аномалій
suspicious	323	29,39	помірні часові аномалії та поодинокі криптографічні помилки
intrusion	92	8,37	сценарії із явними ознаками вторгнення
intrusion stress	89	8,10	множинні криптографічні порушення, підвищені затримки передачі та накопичення аномальних подій
timing	82	7,46	часові атаки без суттєвих криптографічних порушень
mixed	106	9,64	комбіновані сценарії різних типів атак
unseen	91	8,28	випадкові умови, що імітують раніше невідомі сценарії

Загальний обсяг експериментальної вибірки склав 1099 тестових сеансів, що дозволило оцінити поведінку моделі як у типових умовах функціонування системи, так і за наявності аномалій.

На рис.3 наведено матрицю помилок фінального рішення запропонованої адаптивної моделі IDS.

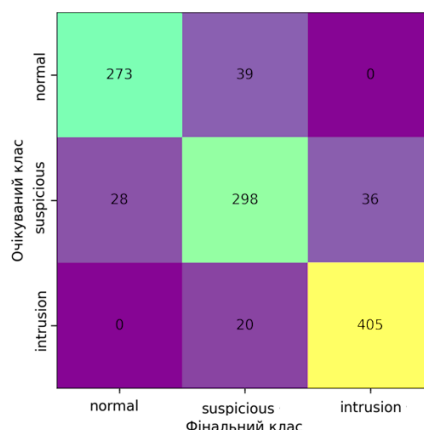


Рисунок 3 – Матриця помилок фінального рішення адаптивної моделі IDS

Найбільша кількість помилкових рішень спостерігається між класами Suspicious та Intrusion. Зокрема, 36 підозрілих сценаріїв було віднесено до класу Intrusion, тоді як 20 атакуючих сценаріїв класифіковано як Suspicious. Така поведінка пояснюється тим, що для частини комбінованих сценаріїв інтегральний показник ризику не перевищував верхній поріг прийняття рішення, хоча система вже фіксувала окремі ознаки аномальної активності. При цьому жоден сценарій із еталонним класом Intrusion не був класифікований як Normal. Такий результат є суттєвим з погляду зниження ймовірності пропуску атак у межах сформованої експериментальної вибірки.

В таблиці 2 наведено розподіл сеансів за очікуваними класами та кількісна оцінка результатів їх класифікації.

Таблиця 2

Результати класифікації за еталонними класами

Клас	Кількість сеансів	Частка, у вибірці%	Правильно класифіковано, %
Normal	312	28,4	87,5
Suspicious	362	32,9	82,3
Intrusion	425	38,7	95,3

Отримані результати підтверджують ефективність запропонованої адаптивної моделі та свідчать про її здатність забезпечувати надійне розмежування нормальних, підозрілих та атакуючих станів системи навіть в умовах наявності комбінованих аномалій та часткового перекриття ознак різних класів. Для оцінки здатності моделі класифікувати різні стани системи побудовано ROC-криві для кожного з трьох класів – Normal, Suspicious та Intrusion (рис. 4).

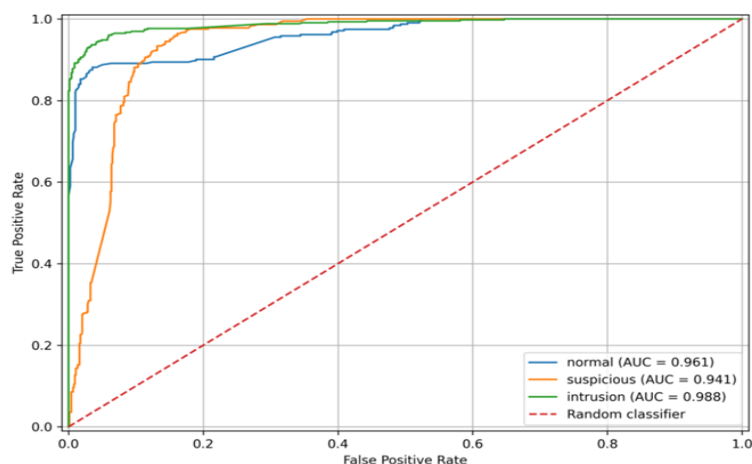


Рисунок 4 – ROC-криві фінального інтегрованого рішення

Як видно з рис.4, усі ROC-криві розташовані вище діагоналі випадкової класифікації, що свідчить про здатність моделі коректно ідентифікувати відповідні стани системи. Для додаткового оцінювання ефективності запропонованої адаптивної моделі було проаналізовано основні метрики класифікації (таблиця 3).

Таблиця 3

Оцінка якості багатокласової класифікації станів системи

Клас	Precision	Recall	F1
Normal	0,907	0,875	0,891
Suspicious	0,835	0,823	0,829
Intrusion	0,918	0,953	0,935

Отримані значення метрик свідчать про достатньо збалансовану якість класифікації для всіх трьох класів. Вони узгоджуються з даними матриці помилок та ROC-аналізу і підтверджують придатність запропонованої адаптивної моделі для багатокласової класифікації станів системи.

Для аналізу роботи запропонованої моделі було досліджено внесок окремих компонентів у формування інтегрованого ризику (рис. 5)

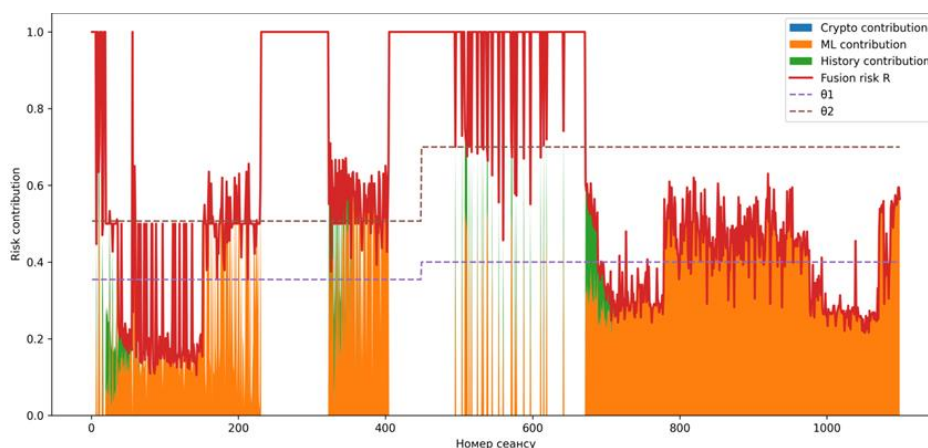


Рисунок 5 – Внесок компонентів у формування інтегрованого ризику



Для більшості некритичних сценаріїв найбільший внесок у формування інтегрованої оцінки ризику забезпечує компонент машинного навчання, тоді як історичний ризик враховує накопичену інформацію про аномальні прояви в попередніх сеансах. Криптографічний компонент, за відсутності критичного порушення, також бере участь у зваженому об'єднанні та відображає ступінь відхилення результатів криптографічного опрацювання від нормального стану. У разі виникнення критичних криптографічних порушень, зокрема неможливості коректного відновлення повідомлення або наявності невиправних стадій опрацювання, спрацьовує механізм D_{crypto} . Відповідно до алгоритму прийняття рішень такий стан безпосередньо класифікується як втручання без подальшого адаптивного зваженого об'єднання компонентів C , M та H . Отже, для таких сценаріїв криптографічні показники виконують не роль домінуючої складової інтегрованого ризику, а роль детермінованої підстави для формування рішення про вторгнення. Такий механізм забезпечує пріоритетне реагування на критичні криптографічні порушення без необхідності виконання подальшого зваженого оцінювання.

У таблиці 4 наведено середні значення криптографічного ризику, оцінки ML-модуля, історичного ризику та інтегрованого ризику для кожного класу.

Таблиця 4

Середні значення компонентних та інтегрованої оцінок ризику

Рішення	Crypto Risk	ML Risk	History Risk	Fusion Risk
Normal	0,138	0,422	0,029	0,257
Suspicious	0,174	0,624	0,084	0,497
Intrusion	0,795	0,822	0,425	0,917

Як видно з наведених даних, із зростанням рівня аномальності сценаріїв спостерігається послідовне збільшення значень усіх компонентів ризику. Найбільш виражене зростання характерне для криптографічного та інтегрованого ризиків, що узгоджується з принципами роботи запропонованої адаптивної моделі прийняття рішень та свідчить про її здатність відображати рівень небезпеки поточного стану.

Для порівняльної оцінки запропонованої адаптивної моделі побудовано ROC-криві окремих модулів та підсумкового рішення (рис. 6). Серед окремих компонентів найвище значення Macro AUC отримано для модуля машинного навчання – 0,838. Для криптографічного компонента цей показник становить 0,810, а для компонента історичного ризику – 0,615. Macro AUC підсумкового рішення адаптивної моделі становить 0,963, що перевищує відповідні значення кожного з компонентів при їх окремому використанні.

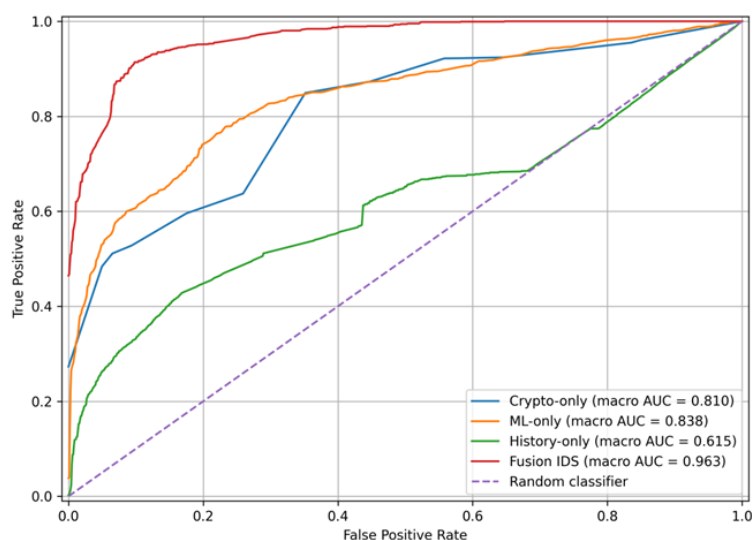


Рисунок 6 – ROC-криві окремих компонентів та підсумкового рішення адаптивної моделі

Вищий показник модуля машинного навчання серед окремих компонентів свідчить про інформативність використаних ним часових, маршрутних та інших ознак поточного сеансу. Криптографічний компонент має дещо нижче значення AUC при автономному використанні, однак у



запропонованій моделі виконує також окрему функцію детермінованого виявлення критичних криптографічних порушень за допомогою індикатора D_{crypto} . Найнижче значення AUC отримано для компонента історичного ризику, що узгоджується з його допоміжною роллю: він не призначений для самостійної класифікації, а використовується для врахування накопичених аномальних проявів у попередніх сеансах.

Отримані результати показують, що інтегроване використання криптографічного, машинного та історичного компонентів забезпечує вищу здатність до розмежування класів, ніж використання кожного з них окремо. Це підтверджує доцільність багаторівневого адаптивного механізму прийняття рішень, у якому поєднуються детерміновані та ймовірнісні ознаки безпеки.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У роботі запропоновано адаптивну модель прийняття рішень для гібридної IDS, що поєднує детермінований криптографічний контроль, ймовірнісне оцінювання на основі машинного навчання та накопичений історичний ризик. Особливістю моделі є пріоритетне виявлення критичних криптографічних порушень із подальшим адаптивним зваженням об'єднанням некритичних криптографічних оцінок, результатів машинного навчання та історичного ризику.

Експериментальне дослідження на вибірці з 1099 тестових сеансів показало, що Macro AUC підсумкового рішення становить 0,963, загальна точність класифікації – 88,81 %, а Macro Precision, Macro Recall і Macro F1-score – відповідно 0,89, 0,88 та 0,88. При цьому жоден сценарій класу Intrusion не був класифікований як Normal. Значення Macro AUC підсумкового рішення перевищило відповідні показники окремих складових: 0,838 для модуля машинного навчання, 0,810 для криптографічного компонента та 0,615 для історичного ризику.

Отримані результати свідчать про доцільність інтегрованого використання детермінованих та ймовірнісних механізмів оцінювання в межах єдиної адаптивної моделі прийняття рішень. Модуль машинного навчання забезпечив найвищі результати ідентифікації станів серед окремих модулів, криптографічний компонент – виявлення критичних порушень передавання даних, а історичний ризик – урахування накопичуваних аномальних проявів.

Подальші дослідження буде спрямовано на розвиток запропонованої адаптивної моделі прийняття рішень та розширення функціональності гібридної IDS за рахунок механізмів автентифікації вузлів і інтеграції аналізу мережевого трафіку.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Ahmad, Z., Shahid Khan, A., Wai Shiang, C., Abdullah, J., & Ahmad. F. (2021). Network intrusion detection system: A systematic study of machine learning and deep learning approaches. *Trans Emerging Tel Tech.* 32:e4150. <https://doi.org/10.1002/ett.4150>
2. Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A.A. (2009). A detailed analysis of the KDD CUP 99 data set. *IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA)*, 1–6. <https://doi.org/10.1109/CISDA.2009.5356528>
3. Joshi, V.R., Assa-Agyei, K., Al-Hadhrani, T., & Qasem, S.N. (2025). Hybrid AI Intrusion Detection: Balancing Accuracy and Efficiency. *Sensors*, 25(24), 7564. <https://doi.org/10.3390/s25247564>
4. Толюпа, С. & Кулько, А., (2025). Реалізація гібридного підходу на основі методів інтелектуального аналізу даних в системах виявлення вторгнень. *Кібербезпека: освіта, наука, техніка*, 2(30) 238–258. <https://doi.org/10.28925/2663-4023.2025.30.966>
5. Mohammed, Widad K., Taha, Mohammed A., & Mohammed, Saleh M. (2024). A Novel Hybrid Fusion Model for Intrusion Detection Systems Using Benchmark Checklist Comparisons, *Mesopotamian Journal of CyberSecurity*, Vol. 4: Iss. 3, Article 13. <https://doi.org/10.58496/MJCS/2024/024>
6. Sajid, M., Malik, K.R., Almogren, A., Malik, T.S., Khan, A.H., Tanveer, J., & Rehman, A.U. (2024). Enhancing intrusion detection: a hybrid machine and deep learning approach. *Journal of Cloud Computing*, 13(1), 123. <https://doi.org/10.1186/s13677-024-00685-x>
7. Gaspar, D., Silva, P., & Silva, C. (2024). Explainable AI for Intrusion Detection Systems: LIME and SHAP Applicability on Multi-Layer Perceptron. *IEEE Access*, 12, 30164–30175. <https://doi.org/10.1109/ACCESS.2024.3368377>
8. Wang, M., Zheng, K., Yang, Y., & Wang, X. (2020). An Explainable Machine Learning Framework for Intrusion Detection Systems. *IEEE Access*, 8, 73127–73141. <https://doi.org/10.1109/ACCESS.2020.2988359>



9. Patil, S., Varadarajan, V., Mazhar, S.M., Sahibzada, A., Ahmed, N., Sinha, O., Kumar, S., Shaw, K., & Kotecha, K. (2022). Explainable Artificial Intelligence for Intrusion Detection System. *Electronics*, 11(19), 3079. <https://doi.org/10.3390/electronics11193079>
10. Neupane, A., Sharma, R., & Skillicorn, D.B. (2022). Explainable Intrusion Detection Systems (X-IDS): Survey of Current Methods, Challenges, and Opportunities. *IEEE Access*, 10, 112352–112379. <https://doi.org/10.1109/ACCESS.2022.3216639>.
11. Abdelaziz, M.T., Radwan, A., Mamdouh, H., Saad, A.S., Abuzaid, A.S., Abdelhakeem, A.A., Zakzouk, S., Moussa, K., & Darweesh, M.S. (2025). Enhancing Network Threat Detection with Random Forest-Based NIDS and Permutation Feature Importance. *Journal of Network and Systems Management*, 33(1), 2. <https://doi.org/10.1007/s10922-024-09874-0>
12. Давлетова А.Я., Пітух І.Р., & Івасьєв С.В. (2025). Гібридна система автентифікації та захисту IoT пристроїв від мережеских атак. *Збірник матеріалів проблемно-наукової міжнародної конференції «Інформаційні проблеми комп'ютерних систем, енергетики, моделювання та управління» (ICSM-2025)*, 38-45. <https://doi.org/10.33744/978-966-632-336-4-2025-2>
13. Давлетова, А.Я. (2026). Модель гібридної системи виявлення вторгнень на основі криптографічних перетворень та методів штучного інтелекту. *Матеріали XV Міжнародної науково-технічної конференції «ITSec: Безпека інформаційних технологій»*, 162–164.

**Alina Davletova**

Lecturer of the Department of Cybersecurity

West Ukrainian National University, Department of Cybersecurity, Ternopil, Ukraine

ORCID: 0000-0002-1192-2532

a.davletova@gmail.com

ADAPTIVE DECISION-MAKING MODEL FOR HYBRID INTRUSION DETECTION SYSTEMS

Abstract. The increasing complexity of cyber threats, together with the limitations of traditional intrusion detection systems based exclusively on signature analysis or probabilistic machine learning models, creates a need for comprehensive adaptive decision-making models. Although machine learning methods can provide high sensitivity to anomalies in network traffic, their application may be associated with false positives and insufficient interpretability of the obtained results. An additional challenge is the coordinated fusion of heterogeneous security indicators that differ in nature, reliability, and criticality, including cryptographic indicators, temporal and behavioral characteristics, and historical risk estimates. This paper proposes an adaptive decision-making model for a hybrid intrusion detection system that combines heterogeneous security features within a unified risk assessment mechanism. The first level of the model is based on deterministic cryptographic control, which provides prioritized detection of critical violations. When a critical cryptographic violation is detected, the system generates an intrusion decision independently of the results produced by the other components. In the absence of such a violation, adaptive weighted fusion of the normalized cryptographic score, the probabilistic estimate produced by a Random Forest-based machine learning module, and the accumulated historical risk is performed, followed by threshold-based formation of the final decision. A software prototype was developed and a series of experimental studies was conducted in a distributed environment using a dataset of 1,099 test sessions covering normal, suspicious, intrusion, combined, and previously unseen scenarios. The experimental results showed that the integrated model achieved a macro-averaged area under the ROC curve of Macro AUC = 0.963 and an overall classification accuracy of 88.81%, while Macro Precision, Macro Recall, and Macro F1-score reached 0.89, 0.88, and 0.88, respectively. None of the intrusion scenarios was classified as a normal state. The proposed model extends the capabilities of traditional intrusion detection systems by combining deterministic cryptographic control with adaptive risk assessment based on machine learning and node behavior history. This approach makes it possible to account for both critical integrity violations and weak accumulated anomalies, which indicates the potential applicability of the proposed model in distributed information and communication systems and cyber-physical systems with increased security requirements.

Keywords: intrusion detection systems, hybrid intrusion detection system, adaptive decision-making model, cryptographic indicators, machine learning, historical risk, risk assessment.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Ahmad, Z., Shahid Khan, A., Wai Shiang, C., Abdullah, J., & Ahmad, F. (2021). Network intrusion detection system: A systematic study of machine learning and deep learning approaches. *Trans Emerging Tel Tech.* 32:e4150. <https://doi.org/10.1002/ett.4150>
2. Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A.A. (2009). A detailed analysis of the KDD CUP 99 data set. *IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA)*, 1–6. <https://doi.org/10.1109/CISDA.2009.5356528>
3. Joshi, V.R., Assa-Agyei, K., Al-Hadhrani, T., & Qasem, S.N. (2025). Hybrid AI Intrusion Detection: Balancing Accuracy and Efficiency. *Sensors*, 25(24), 7564. <https://doi.org/10.3390/s25247564>
4. Toliupa, S., & Kulko, A. (2025). Implementation of a hybrid approach based on intelligent data analysis methods in intrusion detection systems. *Cybersecurity: Education, Science, Technique*, 2(30), 238–258. <https://doi.org/10.28925/2663-4023.2025.30.966>
5. Mohammed, Widad K., Taha, Mohammed A., & Mohammed, Saleh M. (2024). A Novel Hybrid Fusion Model for Intrusion Detection Systems Using Benchmark Checklist Comparisons, *Mesopotamian Journal of CyberSecurity*, Vol. 4: Iss. 3, Article 13. <https://doi.org/10.58496/MJCS/2024/024>
6. Sajid, M., Malik, K.R., Almogren, A., Malik, T.S., Khan, A.H., Tanveer, J., & Rehman, A.U. (2024). Enhancing intrusion detection: a hybrid machine and deep learning approach. *Journal of Cloud Computing*, 13(1), 123. <https://doi.org/10.1186/s13677-024-00685-x>



7. Gaspar, D., Silva, P., & Silva, C. (2024). Explainable AI for Intrusion Detection Systems: LIME and SHAP Applicability on Multi-Layer Perceptron. *IEEE Access*, 12, 30164–30175. <https://doi.org/10.1109/ACCESS.2024.3368377>
8. Wang, M., Zheng, K., Yang, Y., & Wang, X. (2020). An Explainable Machine Learning Framework for Intrusion Detection Systems. *IEEE Access*, 8, 73127–73141. <https://doi.org/10.1109/ACCESS.2020.2988359>
9. Patil, S., Varadarajan, V., Mazhar, S.M., Sahibzada, A., Ahmed, N., Sinha, O., Kumar, S., Shaw, K., & Kotecha, K. (2022). Explainable Artificial Intelligence for Intrusion Detection System. *Electronics*, 11(19), 3079. <https://doi.org/10.3390/electronics11193079>
10. Neupane, A., Sharma, R., & Skillicorn, D.B. (2022). Explainable Intrusion Detection Systems (X-IDS): Survey of Current Methods, Challenges, and Opportunities. *IEEE Access*, 10, 112352–112379. <https://doi.org/10.1109/ACCESS.2022.3216639>
11. Abdelaziz, M.T., Radwan, A., Mamdouh, H., Saad, A.S., Abuzaid, A.S., Abdelhakeem, A.A., Zakzouk, S., Moussa, K., & Darweesh, M.S. (2025). Enhancing Network Threat Detection with Random Forest-Based NIDS and Permutation Feature Importance. *Journal of Network and Systems Management*, 33(1), 2. <https://doi.org/10.1007/s10922-024-09874-0>
12. Davletova, A. Ya., Pitukh, I. R., & Ivasiev, S. V. (2025). Hybrid authentication and protection system for IoT devices against network attacks. *Proceedings of the Problem-Oriented Scientific International Conference “Information Problems of Computer Systems, Energy, Modeling and Management” (ICSM-2025)*, 38–45. <https://doi.org/10.33744/978-966-632-336-4-2025-2>
13. Davletova, A. (2026). Model of a hybrid intrusion detection system based on cryptographic transformations and artificial intelligence methods. *Proceedings of the 15th International Scientific and Technical Conference ITSec: Information Technology Security*, 27–29 May 2026, Ternopil, 162–164. TNTU.

Отримано редакцією журналу / Received: 26.07.26

Прорецензовано / Revised: 14.08.26

Схвалено до друку / Accepted: 26.10.26



This work is licensed under Creative Commons Attribution-noncommercial-sharealike 4.0 International License.