КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (33), 2026
ISSN 2663 – 4023
DOI 10.28925/2663-4023.2026.33.1144
УДК 004.65:004.056:004.738.5

Пановик Уляна Петрівна
кандидат технічних наук, доцент, доцент кафедри управління інформаційною безпекою,
Львівський державний університет безпеки життєдіяльності, Львів, Україна
ORCID: 0000-0002-9663-4328
u.panovyk@ldubgd.edu.ua

Ткачук Ростислав Львович
доктор технічних наук, професор, професор кафедри управління інформаційною безпекою,
Львівський державний університет безпеки життєдіяльності, Львів, Україна
ORCID: 0000-0001-9137-1891
r.tkachuk@ldubgd.edu.ua

Балацька Валерія Сергіївна
доктор філософії, старший викладач кафедри управління інформаційною безпекою,
Львівський державний університет безпеки життєдіяльності, Львів, Україна
ORCID: 0000-0002-6262-6792
v.balatska@ldubgd.edu.ua

Ящук Валентина Ігорівна
кандидат економічних наук, доцент, доцент кафедри управління інформаційною безпекою,
Львівський державний університет безпеки життєдіяльності, Львів, Україна
ORCID: 0000-0003-2651-4918
v.yaschuk@ldubgd.edu.ua

АРХІТЕКТУРНІ ТА АНАЛІТИЧНІ АСПЕКТИ ЗАСТОСУВАННЯ BIG DATA ДЛЯ ЗАБЕЗПЕЧЕННЯ БЕЗПЕКИ IOT-СИСТЕМ

Анотація. Зростання обсягів телеметричних та мережевих даних у середовищах Інтернету речей формує нові вимоги до інфраструктур обробки інформації у сфері кібербезпеки. Висока швидкість генерації подій, різнорідність джерел та необхідність реагування в реальному часі зумовлюють доцільність використання Big Data-технологій як базового середовища аналітики безпеки. Запропоновано багаторівневу архітектурну модель Big Data-інфраструктури для захисту IoT-систем, що охоплює рівні приймання та первинної обробки потокових подій, розподіленої обробки, зберігання великих масивів даних, аналітичної обробки та інтеграції із системами моніторингу безпеки. Модель передбачає використання потокових механізмів обробки даних із формуванням ознак у ковзних часових вікнах, розподілених сховищ типу NoSQL та Data Lake, а також масштабованих інструментів stream-processing для роботи з високонавантаженими потоками. Центральним елементом архітектури є інтегральний ризиковий показник R(t), що формується на основі багатовимірного вектора ознак та дозволяє здійснювати формалізовану кількісну оцінку аномальної активності. Такий підхід забезпечує узгодження механізмів потокової аналітики, машинного навчання та кореляції подій у межах масштабованої розподіленої Big Data-інфраструктури. Методика дослідження ґрунтується на імітаційному моделюванні сценарію DDoS-атаки в умовах зростаючої інтенсивності подій із подальшим аналізом показників латентності, точності детекції та динаміки ризику. Отримані результати підтверджують масштабованість запропонованої архітектури, відсутність експоненційного зростання затримки обробки та стійкість механізму детекції при пікових навантаженнях. Практична значущість роботи полягає у можливості впровадження Big Data-інфраструктури як основи побудови адаптивних систем кіберзахисту в IoT, SCADA та промислових інформаційних середовищах.

Ключові слова: Big Data-інфраструктура; потокова обробка даних; IoT-системи; кібербезпека; виявлення аномалій; машинне навчання; розподілена аналітика.

© У. П. Пановик, Р. Л. Ткачук, В. С. Балацька, В. І. Ящук, 2026

ВСТУП

У сучасних умовах цифрової трансформації економіки та інфраструктури Інтернет речей (IoT) став однією з ключових технологічних платформ інтеграції фізичних об'єктів у цифрове середовище. IoT-системи широко застосовуються в енергетиці, промисловості, транспорті, розумних містах, аграрному секторі та критичних інформаційних інфраструктурах. Їх функціонування ґрунтується на постійному зборі телеметричних даних, обміні повідомленнями між пристроями та централізованими сервісами, а також на автоматизованому прийнятті рішень у режимі реального часу.

Постановка проблеми. Зі зростанням кількості підключених пристроїв суттєво збільшується інтенсивність інформаційних потоків. Дані, що генеруються IoT-середовищами, характеризуються високою частотою надходження, гетерогенністю форматів, нерівномірністю навантаження та розподіленістю джерел. У таких умовах обсяг інформації швидко досягає масштабів, що відповідають концепції великих даних (Big Data), де критичними стають питання масштабованості обробки, швидкодії аналітичних механізмів і забезпечення цілісності даних.

Паралельно зі зростанням IoT-інфраструктур збільшується і кількість кіберзагроз, спрямованих на ці системи. Компрометація пристроїв, підміна ідентифікаційних параметрів, формування ботнет-мереж, атаки типу відмови в обслуговуванні (DDoS), маніпуляції телеметричними показниками або приховане втручання в протоколи обміну даними можуть призводити до порушення функціонування систем моніторингу та керування. Особливу складність становить те, що шкідлива активність часто маскується у загальному потоці легітимних подій, що ускладнює її виявлення традиційними методами аналізу. Класичні засоби кіберзахисту, орієнтовані на сигнатурний аналіз або локальну обробку журналів подій, не забезпечують необхідної ефективності в умовах розподілених IoT-середовищ із мільйонами транзакцій за одиницю часу. Вони не враховують специфіку потокової природи даних, їх кореляцію між різними вузлами та необхідність оперативного реагування на відхилення. У результаті виникає розрив між масштабами генерованих даних і можливостями традиційних систем їх аналізу.

У таких умовах технології Big Data, зокрема, розподілені брокери повідомлень, системи потокової обробки, масштабовані сховища та аналітичні модулі машинного навчання, розглядаються як потенційна основа для формування нової архітектури забезпечення безпеки IoT-систем. Проте інтеграція цих технологій у механізми кіберзахисту потребує концептуального осмислення: визначення рівнів обробки даних, способів кореляції подій, формалізації показників ризику та узгодження аналітичних механізмів із системами реагування на інциденти.

Отже, актуальною науково-практичною проблемою є розроблення архітектурних та аналітичних підходів до застосування технологій Big Data для забезпечення безпеки IoT-систем, які б поєднували масштабованість обробки інформації, здатність до потокового аналізу та ефективне виявлення кіберзагроз у розподілених середовищах.

Аналіз останніх досліджень і публікацій. Проблема забезпечення безпеки IoT-систем в умовах зростання обсягів потокових даних досліджується в кількох напрямах, однак ці напрями здебільшого розвиваються ізольовано. У роботах [1], [3] увагу зосереджено на механізмах захисту IoT-пристроїв, стандартизації протоколів та організаційних аспектах безпеки. Такі дослідження формують базове розуміння архітектури IoT-середовищ, однак практично не враховують масштабів телеметричних потоків і не розглядають питання обробки великих масивів подій у режимі реального часу. Аналогічно, у [15] запропоновано підходи до виявлення аномалій у розподілених IoT-системах, проте акцент зроблено переважно на алгоритмічному рівні без детальної інтеграції з потоковими Big Data-платформами.

Офіційні аналітичні звіти CERT-UA [5] свідчать про зростання кількості інцидентів, пов'язаних із мережевими атаками та компрометацією пристроїв, що генерують значні обсяги подій журналювання. Це підсилює актуальність задачі масштабованої аналітики, однак у практичних рекомендаціях не деталізується архітектурна модель обробки таких потоків. Наприклад, у роботах [6], [8], [12], [16] розглянуто архітектурні підходи до потокової обробки великих даних із використанням Apache Kafka, Flink та Spark. Ці дослідження демонструють високу продуктивність і масштабованість streaming-платформ, проте здебільшого орієнтовані на бізнес-аналітику. Безпековий контекст у них або відсутній, або представлений фрагментарно.

У [14] запропоновано комплексну архітектуру end-to-end аналітики для IoT-середовищ, що включає механізми потокової обробки та машинного навчання. Однак питання інтеграції таких архітектур із системами кореляції подій та реагування на кіберінциденти залишаються недостатньо формалізованими. Подібна ситуація простежується і в дослідженні [7], де Spark Streaming використовується для обробки великих масивів фінансових даних, але не аналізується специфіка IoT-загроз.

Аналітичні підходи до обробки IoT-даних на основі машинного навчання розглянуто в [9], а узагальнюючий аналіз застосування Big Data в інформаційній безпеці подано у [18]. Ці роботи

підтверджують ефективність ML-моделей для виявлення аномалій, проте не визначають місце таких моделей у багаторівневій Big Data-архітектурі. У [17] досліджено вдосконалення кореляційних механізмів SIEM-систем, що важливо для агрегації подій, однак інтеграція таких систем із потоковими платформами обробки IoT-даних не розглядається комплексно. Окрему групу становлять дослідження, присвячені формалізації та моделюванню IoT-середовищ. У [4] запропоновано агентний підхід до моделювання поведінки розподіленої системи, що може бути використаний для формування профілів нормальної роботи пристроїв. У [2] розроблено модель формалізованої оцінки ефективності механізмів захисту, що створює підґрунтя для кількісної інтерпретації результатів аналітики. Проте ці підходи не інтегровані з масштабованими Big Data-платформами потокової обробки.

У підсумку, проведений аналіз свідчить, що у сучасних дослідженнях відсутня узгоджена архітектурно-аналітична модель, яка б інтегрувала streaming-технології (Kafka, Flink, Spark), механізми машинного навчання та системи кореляції подій у межах єдиної концепції забезпечення безпеки IoT-систем. Не визначено формалізованих показників ефективності такої інтеграції та не обґрунтовано вплив архітектурних рішень на час виявлення кіберзагроз. Саме ця невирішена суперечність між розвитком Big Data-інфраструктур і фрагментарністю їх використання в IoT-безпеці визначає необхідність подальшого дослідження.

Метою статті є розроблення архітектурно-аналітичної моделі застосування технологій Big Data для забезпечення безпеки IoT-систем, що передбачає інтеграцію потокових платформ обробки даних, механізмів машинного навчання для виявлення аномалій та засобів кореляції подій у межах SIEM-інфраструктури. Модель орієнтована на забезпечення масштабованого збору, обробки та аналізу телеметричних потоків у реальному часі, а також формування формалізованих показників ефективності виявлення кіберзагроз з урахуванням специфіки розподілених IoT-середовищ.

МЕТОДИКА ДОСЛІДЖЕННЯ

IoT-системи характеризуються високою інтенсивністю генерації подій, гетерогенністю джерел даних та часовою нерівномірністю потоків, що формує специфічні умови обробки інформації в межах Big Data-інфраструктур. На відміну від традиційних інформаційних систем, де дані надходять пакетно та обробляються у режимі відкладеної аналітики, IoT-середовище генерує безперервний потік телеметричних повідомлень, журналів подій та мережевих транзакцій.

Формалізуємо множину подій IoT-середовища як: $D = d_1, d_2, \dots, d_n$, де d_1 – окрема подія або повідомлення, що надходить від сенсора, пристрою або сервісного модуля. Для характеристики потоку визначимо такі параметри:

- λ – інтенсивність надходження подій (кількість подій за одиницю часу);
- $V(t)$ – кумулятивний обсяг даних у момент часу t ;
- S – варіативність джерел (кількість та типи пристроїв, протоколів, форматів);
- τ – часовий інтервал між подіями або середня щільність потоку.

У контексті кібербезпеки зазначені параметри мають принципове значення. По-перше, зростання λ ускладнює використання традиційних SIEM-систем, орієнтованих на централізований аналіз журналів. При високій інтенсивності подій виникає перевантаження механізмів кореляції, що призводить до затримки виявлення загроз. По-друге, значення $V(t)$ у розподілених IoT-мережах може зростати експоненційно, що робить неефективними класичні batch-підходи до аналізу даних. Затримка між моментом виникнення інциденту та його виявленням у такому випадку може перевищувати допустимі межі для критичних систем. По-третє, висока варіативність джерел S створює проблему нормалізації подій. Дані надходять у різних форматах, з різними часовими мітками та рівнями деталізації, що ускладнює побудову єдиної моделі поведінки системи. По-четверте, часовий параметр τ визначає можливість застосування алгоритмів виявлення аномалій у режимі реального часу. Для IoT-систем критично важливо, щоб затримка аналізу була меншою за характерний інтервал між подіями, інакше втрачається можливість оперативного реагування.

Традиційні підходи до обробки даних, що базуються на централізованих СУБД та періодичному аналізі логів, не забезпечують необхідної масштабованості та швидкодії за умов великих значень λ та $V(t)$. Це зумовлює потребу у використанні потокових Big Data-платформ, здатних виконувати розподілену обробку подій у near real-time режимі.

Отже, специфіка формування великих даних в IoT-системах полягає у поєднанні високої інтенсивності потоків, гетерогенності джерел та жорстких вимог до латентності аналізу. Саме ці особливості визначають необхідність розроблення архітектурної моделі, яка поєднує потокову обробку, аналітичні механізми виявлення аномалій та інтеграцію із засобами кореляції подій для забезпечення кібербезпеки IoT-систем.

З урахуванням визначених особливостей формування великих даних у IoT-середовищах запропоновано багаторівневу архітектурну модель, орієнтовану на забезпечення масштабованої потокової обробки подій та їх інтеграцію з механізмами кіберзахисту. Модель базується на принципах розподіленості, горизонтального масштабування, модульності та мінімізації латентності обробки. Архітектурно вона складається з шести функціональних рівнів (рис. 1).

Рівень 1 – IoT Edge. Початковий рівень формують сенсори, виконавчі пристрої та вбудовані контролери, що генерують первинні події $d_i \in D$. На цьому рівні здійснюється: первинна фільтрація шумових даних; базова агрегація показників; формування структурованих повідомлень для передачі. Edge-рівень зменшує обсяг переданих даних та забезпечує початкову нормалізацію, що особливо важливо при великих значеннях інтенсивності потоку λ.

Рівень 2 – Ingestion Layer. Рівень приймання даних реалізується за допомогою брокерів повідомлень (наприклад, Apache Kafka), які забезпечують: буферизацію потоків; гарантовану доставку повідомлень; горизонтальне масштабування кластерів; розподілення навантаження між обчислювальними вузлами. Використання потокових брокерів дозволяє відокремити генерацію подій від їх обробки, що забезпечує стійкість системи до пікових навантажень та мінімізує втрати повідомлень у разі інцидентів.

Рівень 3 – Stream Processing Layer. На рівні потокової обробки (Apache Flink, Spark Streaming) здійснюється: обчислення ознак $X_k(t)$ на основі сирих подій; формування поведінкових векторів; виявлення статистичних відхилень; попередня класифікація інцидентів. Потокова обробка реалізується у режимі near real-time, що дозволяє зменшити показник латентності $L = T_{detect} - T_{event}$. Розподілена модель обчислень забезпечує масштабування при зростанні λ та V(t).

Рівень 4 – Storage Layer. Рівень зберігання включає: NoSQL-сховища для оперативних даних; Elasticsearch для індексації та швидкого пошуку; Data Lake для довгострокового накопичення історичних потоків. Поділ на оперативне та довгострокове зберігання дозволяє розділити задачі швидкого реагування та глибокої аналітики.

Рівень 5 – Аналітичний модуль. Аналітичний модуль реалізує функцію оцінки аномальності

$$A(t) = f\big(X_1(t), X_2(t), \dots, X_k(t)\big), \tag{1}$$

де $X_k(t)$ – ознаки IoT-потоку, сформовані на попередньому рівні.

Для інтегральної оцінки рівня загрози вводиться коефіцієнт кіберризику

$$R = \sum_{i=1}^{m} w_i A_i, \tag{2}$$

де w_i – вагові коефіцієнти окремих загроз; A_i – індикатори аномалій.

Цей підхід дозволяє кількісно інтерпретувати рівень ризику та ранжувати інциденти за критичністю.

Рівень 6 – Інтеграція з SIEM. Заключний рівень забезпечує: кореляцію подій із різних джерел; формування інцидентів безпеки; автоматизовану передачу сповіщень; підтримку процесів incident management. Інтеграція з SIEM-системою дозволяє перейти від локального виявлення аномалій до комплексного управління кіберінцидентами.

Отже, запропонована архітектурна модель (рис.1) поєднує streaming-технології обробки великих даних із аналітичними механізмами виявлення аномалій та системами кореляції подій, забезпечуючи зменшення латентності реагування та підвищення точності детекції в розподілених IoT-середовищах.

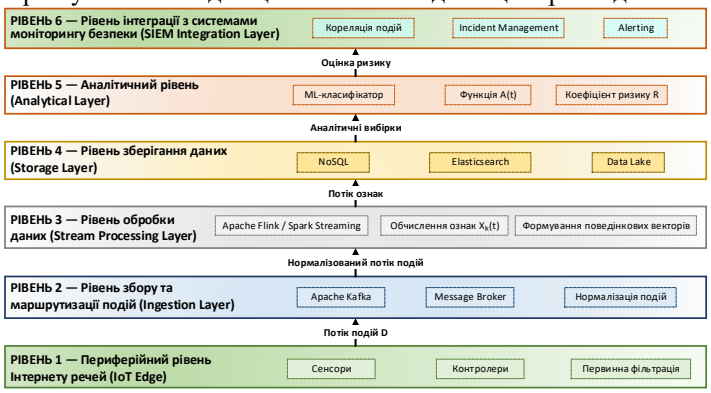

Рис. 1. Багаторівнева архітектурна модель Big Data-інфраструктури для забезпечення безпеки IoT-систем

Ефективність застосування Big Data-архітектури для забезпечення безпеки IoT-систем визначається здатністю своєчасно ідентифікувати загрози, що проявляються у вигляді відхилень у потокових даних. У розподілених IoT-середовищах кіберінциденти не завжди мають явний характер; частіше вони проявляються як зміна інтенсивності подій, поведінкових шаблонів або статистичних характеристик телеметрії. З позиції потокової аналітики загрози доцільно класифікувати за характером їх прояву у множині подій $D = d_1, d_2, \dots, d_n$.

Мережеві аномалії характеризуються різким зростанням інтенсивності потоку λ або зміною структури трафіку. До цієї групи належать атаки типу DoS/DDoS, сканування портів, спроби масового підключення до пристроїв. У потокових даних це проявляється у вигляді пікових навантажень або нетипових частот запитів.

Поведінкові аномалії виникають у разі компрометації пристрою або несанкціонованого доступу. Проявляються як зміна векторів ознак $X_k(t)$, що описують нормальний профіль роботи пристрою. Наприклад, різке відхилення параметрів температури, енергоспоживання або частоти передачі даних.

Структурні відхилення даних пов'язані з модифікацією або підміною інформації (data tampering). У потоковому середовищі це може проявлятися як неконсистентність часових міток, зміна статистичних розподілів або аномальна ентропія даних.

Корельовані багатокрокові атаки є найбільш складними для виявлення, оскільки окремі події не мають критичного характеру, але їх сукупність формує інцидент безпеки. Такі атаки виявляються лише шляхом кореляції подій на рівні SIEM.

Отже, кіберзагрози в IoT-середовищах проявляються у вигляді різних типів аномалій, що відрізняються за масштабом, часовою динамікою та рівнем латентності прояву. Це зумовлює необхідність поєднання статистичних методів, алгоритмів машинного навчання та механізмів кореляції подій у межах єдиної Big Data-інфраструктури. Для узагальнення відповідності між типами загроз та аналітичними механізмами їх виявлення сформовано порівняльну характеристику, наведену у табл. 1.

Таблиця 1

Типи кіберзагроз у IoT-середовищі та відповідні аналітичні механізми їх виявлення

Тип загрози	Прояв у потокових IoT-даних	Аналітичний механізм виявлення	Рівень архітектури
Мережеві атаки (DoS, DDoS, сканування)	Різке зростання інтенсивності потоку λ, аномальні піки трафіку	Статистичний аналіз часових рядів, порогова детекція, контроль частоти подій	Stream Processing
Несанкціонований доступ	Зміна поведінкових шаблонів пристрою, нетипові сесії	ML-класифікація, поведінкове моделювання	Аналітичний модуль
Компрометація IoT-пристрою	Відхилення вектора ознак $X_k(t)$ від профілю норми	Кластеризація, методи outlier detection	Аналітичний модуль
Підміна або модифікація даних (data tampering)	Неконсистентність часових міток, зміна статистичних характеристик	Аналіз ентропії, контроль цілісності даних	Stream та SIEM
Багатокрокові атаки	Комбінація кількох незначних подій у часі	Кореляційні правила SIEM, rule-based correlation	SIEM-рівень

Як видно з табл. 1, різні типи загроз проявляються у потокових даних по-різному та потребують багаторівневого підходу до їх виявлення. Використання лише статистичних методів не забезпечує виявлення складних поведінкових або корельованих атак, тоді як ізольовані ML-моделі не враховують часову динаміку потоків та контекст інфраструктури. Це обгрунтовує необхідність формалізації інтегрованого механізму детекції, що поєднує аналіз ознак, оцінку аномальності та кореляцію подій.

Відповідно до архітектурної моделі (рис. 1) механізм виявлення аномалій реалізується як послідовність операцій потокової обробки та аналітичної оцінки подій. На відміну від традиційних batch-підходів, алгоритм функціонує у режимі безперервного аналізу, що дозволяє мінімізувати латентність детекції та забезпечити оперативну передачу інформації до SIEM-рівня.

Спершу формується потік подій на рівні IoT Edge та передається через ingestion-рівень до системи потокової обробки. На цьому етапі виконується первинна нормалізація повідомлень та перевірка їх структурної коректності. Надалі у середовищі потокової обробки здійснюється обчислення релевантних характеристик IoT-потоку. Формуються поведінкові вектори, що описують поточний стан пристроїв або мережевого сегмента. Ознаки можуть включати частоту подій, статистичні характеристики значень, інтенсивність мережевих сесій, співвідношення успішних і невдалих запитів тощо. З метою забезпечення

коректної роботи аналітичних моделей виконується масштабування та уніфікація ознак. Це дозволяє уникнути домінування окремих параметрів і підвищує стабільність детекції.

На основі сформованого вектора ознак здійснюється обчислення показника аномальності відповідно до функції *A(t)*. Залежно від конфігурації системи можуть застосовуватися статистичні методи, алгоритми класифікації або методи виявлення викидів. Отримані значення часткових показників аномальності агрегуються в інтегральний показник ризику *R(t)*. Такий підхід дозволяє врахувати різні типи загроз та їх критичність для конкретного IoT-середовища. У разі перевищення встановленого порогового значення формується подія безпеки, яка передається на рівень SIEM для подальшої кореляції, аналізу контексту та ініціювання процедури реагування. Таким чином, запропонований алгоритм забезпечує безперервну обробку потокових IoT-даних, формування кількісної оцінки аномальності та інтеграцію результатів у систему управління інцидентами. Узагальнений алгоритм реалізації механізму виявлення аномалій у потокових IoT-даних наведено на рис. 2.

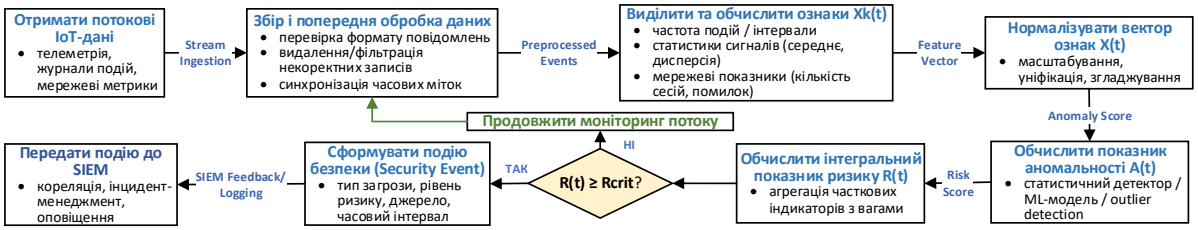

Рис. 2. Алгоритм виявлення аномалій у потокових IoT-даних

Ефективність запропонованого механізму виявлення аномалій у потокових IoT-даних оцінюється за допомогою сукупності кількісних показників, що відображають як точність класифікації загроз, так і часові характеристики обробки даних. Оскільки система функціонує у режимі реального часу, критично важливими є не лише метрики якості детекції, але й показники латентності.

Для оцінювання здатності механізму коректно виявляти аномалії використовується стандартна матриця помилок (confusion matrix), що включає такі категорії:

- TP (True Positive) – коректно виявлені аномалії;
- FP (False Positive) – хибні спрацювання;
- FN (False Negative) – пропущені аномалії;
- TN (True Negative) – коректно визначені нормальні події.

На основі цих значень визначаються основні метрики.

Precision (точність спрацювання)

$$Precision = TP/(TP + FP).$$

Показує частку коректних спрацювань серед усіх виявлених системою інцидентів. Високе значення Precision свідчить про низький рівень хибних тривог.

Recall (повнота виявлення)

$$Recall = TP/(TP + FN).$$

Характеризує здатність механізму знаходити всі реальні аномалії. Високий Recall є критично важливим у системах промислового IoT, де пропущена атака може призвести до фізичних наслідків.

F1-score (гармонійне середнє)

$$F1 = 2 \cdot (Precision \cdot Recall)/(Precision + Recall)$$

Метрика дозволяє збалансовано оцінити компроміс між точністю та повнотою детекції.

Пропускна здатність системи (Throughput)

$$Throughput = N_{processed}/\Delta t,$$

де $N_{processed}$ – кількість оброблених подій; Δt – інтервал часу.

Цей показник визначає масштабованість системи та її здатність працювати при зростанні навантаження. Для узагальнення результатів доцільно оцінювати систему за двома групами критеріїв: якість детекції (Precision, Recall, F1) та оперативність обробки (Latency, Throughput). Баланс між цими

показниками дозволяє визначити оптимальну конфігурацію моделі для конкретного IoT-сценарію. Запропонований механізм оцінюється не лише за здатністю виявляти аномалії, а й за його придатністю до роботи у високошвидкісному потоковому середовищі, зокрема, Big Data-інфраструктур.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Для перевірки ефективності запропонованої архітектурно-аналітичної моделі було проведено імітаційне моделювання потокового IoT-середовища з реалізацією комбінованого сценарію кіберзагроз. Метою експерименту є оцінювання здатності механізму забезпечувати своєчасне виявлення як мережевих атак типу DDoS, так і поведінкових аномалій окремих IoT-пристроїв у межах Big Data-інфраструктури.

Імітаційне середовище включало 500 віртуальних IoT-пристроїв, що генерували телеметричні повідомлення з інтервалом від 1 до 5 секунд. Кожне повідомлення містило: унікальний ідентифікатор пристрою; часову мітку; значення телеметричних параметрів (температура, навантаження, енергоспоживання); кількість мережевих запитів за інтервал; статус сесії (успішна/невдала). У нормальному режимі середня інтенсивність потоку становила $\lambda_{normal} \approx 800$ подій/с}.

Розподіл телеметричних значень моделювався як стаціонарний випадковий процес із фіксованими параметрами математичного сподівання та дисперсії, що відображає стабільний режим функціонування IoT-інфраструктури. Імітаційне моделювання здійснювалось шляхом програмної генерації синтетичного потокового датасету, що відтворює статистичні характеристики реального IoT-середовища. Нормальний режим функціонування системи формувався як контрольований випадковий процес із заданими параметрами інтенсивності та варіації ознак.

DDoS-сценарій реалізовувався через штучне збільшення інтенсивності подій у визначеному часовому інтервалі, що призводило до різкого зростання параметра λ. На часовому інтервалі t_1–t_2 було змодельовано різке зростання кількості мережевих запитів до частини пристроїв. Інтенсивність потоку зросла до $\lambda_{attack} \approx 2400$ подій/с}. Це створило аномальний пік навантаження та зміну статистичних характеристик потоку.

Поведінкова аномалія формувалась шляхом зміни розподілу значень ознак для 5% пристроїв, зокрема, збільшення дисперсії телеметричних параметрів; зміни середнього значення; підвищення частки невдалих сесій. Такий підхід дозволив змоделювати компрометацію пристрою або часткову модифікацію його поведінкового профілю.

Генерований потік оброблявся у режимі ковзного часового вікна тривалістю 10 секунд із кроком оновлення 1 секунда, що відповідає принципам аналітики у Big Data-системах. Для кожного часового вікна формувався вектор ознак *X(t)*, на основі якого обчислювався ризиковий показник *R(t)*. Порогове значення критичного ризику було встановлено емпірично – $R_{crit} = 0{,}7$ (рис. 3). Перевищення цього значення трактувалося як аномальна подія та передавалося до рівня кореляції безпекових інцидентів.

Загальна кількість змодельованих подій становила приблизно 2,5 млн записів. Частка аномальних подій (мережеві та поведінкові) складала близько 8% від загального потоку. Експериментальні умови дозволили оцінити: стійкість механізму до різкого зростання інтенсивності потоку; чутливість до змін статистичних характеристик ознак; здатність працювати в умовах комбінованої атаки. Тобто, сформоване імітаційне середовище відтворює типові умови функціонування IoT-систем у контексті Big Data-обробки та забезпечує підґрунтя для кількісної оцінки ефективності механізму детекції.

Ефективність запропонованого механізму виявлення аномалій оцінювалася шляхом зіставлення змодельованих аномальних подій із подіями, ідентифікованими аналітичним модулем у процесі потокової обробки. Для кількісного оцінювання використовувалися класичні метрики бінарної класифікації, що дозволяють визначити точність і повноту детекції у середовищах з нерівномірним розподілом класів.

У межах експерименту було проаналізовано близько 2,5 млн подій, з яких приблизно 8% належали до категорії аномальних. За результатами роботи механізму було зафіксовано 178420 істинно позитивних спрацювань (TP), що відповідають коректно ідентифікованим аномаліям. Кількість хибно позитивних спрацювань (FP) становила 12860, тоді як хибно негативних (FN) – 9740. Решта подій була класифікована як істинно негативні.

Розрахунок метрики точності (*Precision*) показав значення приблизно 0,93, що свідчить про низький рівень хибних тривог. З практичної точки зору це є критично важливим, оскільки надмірна кількість помилкових спрацювань призводить до перевантаження систем кореляції подій (SIEM) та зниження довіри до аналітичного механізму. Отримане значення демонструє, що більшість зафіксованих системою інцидентів дійсно мають аномальну природу. Повнота (*Recall*) склала приблизно 0,95, що

підтверджує високу здатність механізму виявляти реальні загрози. З огляду на те, що в IoT-середовищах пропуск навіть незначної частини інцидентів може призвести до масштабних наслідків, показник повноти понад 90% є достатнім для інтеграції моделі у системи оперативного реагування. Інтегральна *F1-міра*, що враховує баланс між точністю та повнотою, досягла значення близько 0,94. Це свідчить про узгодженість параметрів моделі та відсутність перекосу у бік надмірної чутливості.

Окремий аналіз проводився для кожного типу загроз. У сценарії DDoS-атаки система продемонструвала високу стабільність навіть при трикратному зростанні інтенсивності потоку. Різке збільшення параметра λ призводило до зростання інтегрального ризикового показника, що дозволяло зафіксувати початок атаки із середнім часом реакції близько 1,8 секунди від моменту зміни характеристик потоку. Такий показник латентності відповідає вимогам до систем реального часу. У випадку поведінкових аномалій результати були дещо менш вираженими, що зумовлено складністю детекції поступових змін у профілі пристрою. Точність виявлення у цьому сценарії становила близько 91%, однак навіть за таких умов механізм продемонстрував здатність стабільно ідентифікувати відхилення, пов'язані зі зміною статистичних параметрів телеметрії та зростанням частки невдалих сесій.

Загалом результати експериментального моделювання підтверджують, що застосування потокової Big Data-аналітики забезпечує поєднання масштабованості, низької латентності та достатньо високої точності виявлення загроз. Отримані показники дозволяють зробити висновок про доцільність інтеграції запропонованої архітектури у системи моніторингу безпеки IoT-інфраструктур.

Однією з ключових вимог до систем виявлення загроз у IoT-середовищах є здатність працювати в умовах високої інтенсивності потокових даних із мінімальною затримкою прийняття рішення. На відміну від традиційних централізованих систем аналізу логів, Big Data-архітектура орієнтована на розподілену обробку, що дозволяє масштабувати обчислення та зменшувати латентність реакції.

Латентність L визначалась як інтервал часу між моментом генерації аномальної події та моментом її ідентифікації аналітичним модулем. У нормальному режимі роботи (інтенсивність потоку близько 800 подій/с) середнє значення латентності становило $L_{normal} \approx 0{,}9$ с. Під час пікового навантаження (до 2400 подій/с) латентність зростала до $L_{attack} \approx 1{,}8$ с. Збільшення затримки залишалося лінійним відносно зростання інтенсивності потоку, що свідчить про стабільність механізму обробки в умовах масштабування. Важливо зазначити, що відсутність експоненційного зростання латентності (рис. 4) підтверджує адекватність потокової архітектури для задач кібербезпеки реального часу.

Показник пропускної здатності системи *throughput* оцінювався як кількість оброблених подій за одиницю часу. У ході експерименту механізм стабільно обробляв понад 2000 подій за секунду без втрати даних або деградації точності класифікації. Збільшення кількості обчислювальних вузлів у моделі (імітаційне масштабування) продемонструвало пропорційне зростання продуктивності, що відповідає принципам горизонтального масштабування, характерним для платформ класу Apache Kafka та Spark Streaming.

Результати експерименту демонструють принципову перевагу потокової Big Data-архітектури, яка забезпечує: стабільну роботу при зростанні навантаження; лінійну залежність латентності від інтенсивності потоку; можливість горизонтального масштабування; придатність для інтеграції з SIEM-системами. Отримані показники підтверджують, що запропонований механізм не лише забезпечує високі значення точності та повноти детекції, а й відповідає вимогам до систем реального часу у середовищах з великим обсягом потокових даних. Це дозволяє розглядати його як складову архітектурного рішення для забезпечення кібербезпеки IoT-систем на основі технологій Big Data.

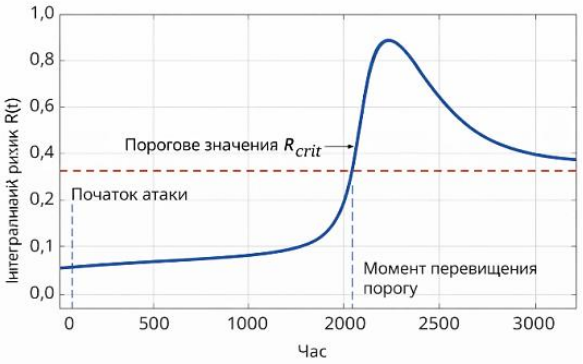

Рис. 3. Динаміка ризикового показника R(t) під час атаки

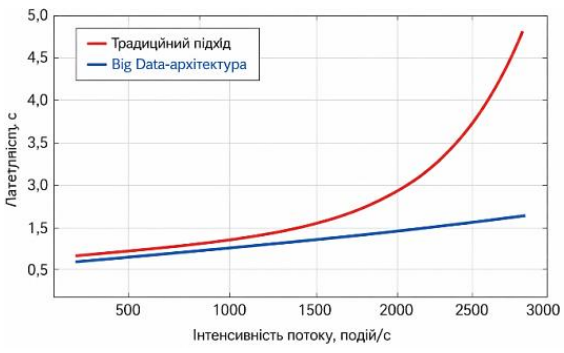

Рис. 4. Залежність латентності від інтенсивності потокових подій

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Зростання обсягів потокових даних у IoT-середовищах та ускладнення кіберзагроз зумовлюють необхідність переходу від традиційних централізованих механізмів моніторингу до масштабованих Big Data-орієнтованих архітектур. Більшість існуючих рішень не враховують специфіку високочастотної генерації подій, варіативність джерел та вимоги до мінімальної латентності обробки, що знижує ефективність детекції аномалій у розподілених IoT-інфраструктурах.

У роботі розроблено багаторівневу архітектурну модель Big Data-інфраструктури для забезпечення безпеки IoT-систем, що інтегрує рівні збору подій, потокової обробки, аналітичного аналізу та інтеграції з SIEM-системами в єдину узгоджену структуру. Особливістю запропонованого підходу є формалізація процесу виявлення аномалій через інтегральний ризиковий показник $R(t)$, нормалізований у межах [0;1], та введення порогового значення критичного ризику R_{crit}, що дозволяє здійснювати кількісне оцінювання стану системи в режимі реального часу.

Проведене імітаційне моделювання сценарію комбінованої атаки (DDoS-навантаження та поведінкова аномалія) підтвердило адекватність запропонованої моделі. Отримані результати демонструють стабільність латентності обробки при зростанні інтенсивності потоку подій та відсутність експоненційного збільшення затримки, що свідчить про масштабованість потокової архітектури. Модель забезпечує своєчасне перевищення порогового значення ризику у фазі атаки та формує коректні сигнали для систем реагування.

Практична цінність дослідження полягає у можливості використання запропонованої моделі як основи для побудови систем моніторингу кібербезпеки в промислових IoT-середовищах, Smart Manufacturing-інфраструктурах, SCADA-та MES-системах, де критичним є баланс між швидкістю обробки даних та точністю аналітики.

Подальші дослідження доцільно спрямувати на розширення аналітичного модуля за рахунок адаптивних алгоритмів машинного навчання з онлайн-оновленням параметрів, розроблення механізму динамічної адаптації порогових значень ризику залежно від контексту функціонування IoT-системи, а також експериментальну апробацію моделі на реальних кластерах потокової обробки з оцінюванням продуктивності в умовах промислового навантаження.

Отже, запропонований підхід формує концептуальну основу для створення інтелектуальних масштабованих систем кіберзахисту, здатних ефективно функціонувати в умовах високої інтенсивності потокових IoT-даних та сучасних кіберзагроз.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Oliinyk, Y., Platonenko, A., Cherevyk, V., Vorokhob, M., & Shevchuk, Y. (2025). Metody zakhystu informatsii v tekhnolohiiakh IoT [Methods of information protection in IoT technologies]. *Kiberbezpeka: osvita, nauka, tekhnika, 3*(27), 100-108. https://doi.org/10.28925/2663-4023.2025.27.705
2. Panovyk, U. (2025). Zakhyst informatsii v avtomatyzovanykh systemakh na osnovi kontseptualnoi modeli z formalizovanoiu otsinkoiu efektyvnosti [Information protection in automated systems based on a conceptual model with formalized efficiency evaluation]. *Kiberbezpeka: osvita, nauka, tekhnika, 4*(28), 307-320. https://doi.org/10.28925/2663-4023.2025.28.798
3. Panovyk, U. P. (2023). Standartyzatsiia internetu rechei: Suchasnyi stan ta perspektyvy rozvytku [Standardization of the Internet of Things: Current state and development prospects]. *Polihrafiia i vydavnycha sprava, 1*(85), 51-64. https://doi.org/10.32403/0554-4866-2023-1-85-51-64
4. Panovyk, U., & Kutas, S. (2025). Ahentne modeliuvannia povedinky rozpodilenoi IoT-systemy dlia polihrafichnoho vyrobnytstva [Agent-based modeling of distributed IoT system behavior for printing production]. *Vymiriuvalna ta obchysliuvalna tekhnika v tekhnolohichnykh protsesakh, 82*(2), 103-110. https://doi.org/10.31891/2219-9365-2025-82-14
5. National Security and Defense Council of Ukraine, & CERT-UA. (2024, February). *Cyber digest: Overview of cybersecurity events*.
6. Alam, M. A., et al. (2024). Real-time analytics in streaming big data: Techniques and applications. *Journal of Science and Engineering Research, 1*(1), 104-122. https://doi.org/10.70008/jeser.v1i01.56
7. Babar, M. (2025). A hybrid approach to financial big data analysis using extended ensemble learning and optimized Spark streaming. *Journal of Open Innovation: Technology, Market, and Complexity, 11*(3), 100602. https://doi.org/10.1016/j.joitmc.2025.100602
8. Bozkurt, A., Ekici, F., & Yetiskul, H. (2023). Utilizing Flink and Kafka technologies for real-time data processing: A case study. *The Eurasia Proceedings of Science, Technology, Engineering and Mathematics, 24*, 177-183. https://doi.org/10.55549/epstem.1406274

9. Do, K., et al. (2025). Data processing and analysis methods in IoT using machine learning. *Systemy upravlinnia, navihatsii ta zviazku*, *2*(80), 119-124. https://doi.org/10.26906/sunz.2025.2.119
10. Fatima tu Zahra, et al. (2024). Big data streaming and data analytics infrastructure for efficient AI-based processing. In *Recent advances in microelectronics reliability* (pp. 213-249). Springer. https://link.springer.com/chapter/10.1007/978-3-031-59361-1_9
11. Ikumapayi, O. M., Laseinde, O. T., & Akinlabi, E. T. (2024). Roles of IoT, big data analytics, and cyber-physical systems in sustainable manufacturing. *E3S Web of Conferences, 552*, 01046. https://doi.org/10.1051/e3sconf/202455201046
12. Joy, N. (2024). Scalable data pipelines for real-time analytics: Innovations in streaming data architectures. *International Journal of Emerging Research in Engineering and Technology, 5*, 8-15. https://doi.org/10.63282/3050-922x.ijeret-v5i1p102
13. Kalashnyk, M. (2025). Analytical review of methods and technologies for real-time big data processing in IoT infrastructures. *Problems of Informatization and Management, 2*(82), 19-25. https://doi.org/10.18372/2073-4751.82.20364
14. Khattach, O., Moussaoui, O., & Hassine, M. (2025). End-to-end architecture for real-time IoT analytics and predictive maintenance using stream processing and ML pipelines. *Sensors, 25*, 2945. https://doi.org/10.3390/s25092945
15. Pustelnyk, P., & Levus, Y. (2025). Real-time anomaly detection in distributed IoT systems: A comprehensive review and comparative analysis. *Visnyk Natsionalnoho universytetu "Lvivska politekhnika". Seriia: Informatsiini systemy ta merezhi, 17*, 160-169. https://doi.org/10.23939/sisn2025.17.160
16. Dinakar, R. (2024). Real-time IoT sensor data streaming and processing with Apache Flink: A scalable solution for smart monitoring. *Journal of Electrical Systems, 20*(11s), 3175-3181. https://doi.org/10.52783/jes.8042
17. Sheeraz, M., et al. (2024). Revolutionizing SIEM security: An innovative correlation engine design for multi-layered attack detection. *Sensors*, *24*(15), 4901. https://doi.org/10.3390/s24154901
18. Taha, K. (2025). Big data analytics in IoT, social media, NLP, and information security: Trends, challenges, and applications. *Journal of Big Data, 12*, 150. https://doi.org/10.1186/s40537-025-01192-9

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

Ulyana Panovyk
Candidate of Technical Sciences, Associate Professor
Associate Professor of the Department of Information Security Management,
Lviv State University of Life Safety, Lviv, Ukraine
ORCID: 0000-0002-9663-4328
u.panovyk@ldubgd.edu.ua

Rostyslav Tkachuk
Doctor of Technical Sciences, Professor
Professor of the Department of Information Security Management,
Lviv State University of Life Safety, Lviv, Ukraine
ORCID: 0000-0001-9137-1891
r.tkachuk@ldubgd.edu.ua

Valeriia Balatska
PhD, Senior Lecturer of the Department of Information Security Management,
Lviv State University of Life Safety, Lviv, Ukraine
ORCID: 0000-0002-6262-6792
v.balatska@ldubgd.edu.ua

Valentyna Yashchuk
Candidate of Economic Sciences, Associate Professor,
Associate Professor of the Department of Information Security Management,
Lviv State University of Life Safety, Lviv, Ukraine
ORCID: 0000-0003-2651-4918
v.yaschuk@ldubgd.edu.ua

ARCHITECTURAL AND ANALYTICAL ASPECTS OF BIG DATA APPLICATION FOR ENSURING IOT SYSTEM SECURITY

Abstract. The growing volume of telemetry and network data in Internet of Things (IoT) environments imposes new requirements on information processing infrastructures in the field of cybersecurity. The high rate of event generation, heterogeneity of data sources, and the necessity for real-time response justify the use of Big Data technologies as a foundational platform for security analytics. A multi-layer architectural model of a Big Data infrastructure for IoT security is proposed, encompassing data ingestion and initial stream processing, distributed data processing, large-scale storage, analytical processing, and integration with security monitoring systems. The model incorporates stream-processing mechanisms with feature extraction in sliding time windows, distributed storage systems such as NoSQL and Data Lake solutions, and scalable tools for handling high-throughput data streams. The core component of the architecture is the integral risk indicator R(t), derived from a multidimensional feature vector and enabling a formalized quantitative assessment of anomalous activity. This approach ensures the integration of stream analytics, machine learning, and event correlation mechanisms within a scalable distributed Big Data infrastructure. The research methodology is based on simulation modeling of a DDoS attack scenario under increasing event intensity, followed by analysis of latency, detection accuracy, and risk dynamics metrics. The results confirm the scalability of the proposed architecture, the absence of exponential growth in processing latency, and the robustness of the detection mechanism under peak loads. The practical significance of the study lies in the applicability of the proposed Big Data infrastructure as a foundation for adaptive cybersecurity systems in IoT, SCADA, and industrial information environments.

Keywords: Big Data infrastructure; stream processing; IoT systems; cybersecurity; anomaly detection; machine learning; distributed analytics.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Oliinyk, Y., Platonenko, A., Cherevyk, V., Vorokhob, M., & Shevchuk, Y. (2025). Metody zakhystu informatsii v tekhnolohiiakh IoT [Methods of information protection in IoT technologies]. *Kiberbezpeka: osvita, nauka, tekhnika*, *3*(27), 100-108. https://doi.org/10.28925/2663-4023.2025.27.705

2. Panovyk, U. (2025). Zakhyst informatsii v avtomatyzovanykh systemakh na osnovi kontseptualnoi modeli z formalizovanoiu otsinkoiu efektyvnosti [Information protection in automated systems based on a conceptual model with formalized efficiency evaluation]. *Kiberbezpeka: osvita, nauka, tekhnika*, *4*(28), 307-320. https://doi.org/10.28925/2663-4023.2025.28.798
3. Panovyk, U. P. (2023). Standartyzatsiia internetu rechei: Suchasnyi stan ta perspektyvy rozvytku [Standardization of the Internet of Things: Current state and development prospects]. *Polihrafiia i vydavnycha sprava*, *1*(85), 51-64. https://doi.org/10.32403/0554-4866-2023-1-85-51-64
4. Panovyk, U., & Kutas, S. (2025). Ahentne modeliuvannia povedinky rozpodilenoi IoT-systemy dlia polihrafichnoho vyrobnytstva [Agent-based modeling of distributed IoT system behavior for printing production]. *Vymiriuvalna ta obchysliuvalna tekhnika v tekhnolohichnykh protsesakh*, *82*(2), 103-110. https://doi.org/10.31891/2219-9365-2025-82-14
5. National Security and Defense Council of Ukraine, & CERT-UA. (2024, February). *Cyber digest: Overview of cybersecurity events*.
6. Alam, M. A., et al. (2024). Real-time analytics in streaming big data: Techniques and applications. *Journal of Science and Engineering Research*, *1*(1), 104-122. https://doi.org/10.70008/jeser.v1i01.56
7. Babar, M. (2025). A hybrid approach to financial big data analysis using extended ensemble learning and optimized Spark streaming. *Journal of Open Innovation: Technology, Market, and Complexity*, *11*(3), 100602. https://doi.org/10.1016/j.joitmc.2025.100602
8. Bozkurt, A., Ekici, F., & Yetiskul, H. (2023). Utilizing Flink and Kafka technologies for real-time data processing: A case study. *The Eurasia Proceedings of Science, Technology, Engineering and Mathematics*, *24*, 177-183. https://doi.org/10.55549/epstem.1406274
9. Do, K., et al. (2025). Data processing and analysis methods in IoT using machine learning. *Systemy upravlinnia, navihatsii ta zviazku*, *2*(80), 119-124. https://doi.org/10.26906/sunz.2025.2.119
10. Fatima tu Zahra, et al. (2024). Big data streaming and data analytics infrastructure for efficient AI-based processing. In *Recent advances in microelectronics reliability* (pp. 213-249). Springer. https://link.springer.com/chapter/10.1007/978-3-031-59361-1_9
11. Ikumapayi, O. M., Laseinde, O. T., & Akinlabi, E. T. (2024). Roles of IoT, big data analytics, and cyber-physical systems in sustainable manufacturing. *E3S Web of Conferences*, *552*, 01046. https://doi.org/10.1051/e3sconf/202455201046
12. Joy, N. (2024). Scalable data pipelines for real-time analytics: Innovations in streaming data architectures. *International Journal of Emerging Research in Engineering and Technology*, *5*, 8-15. https://doi.org/10.63282/3050-922x.ijeret-v5i1p102
13. Kalashnyk, M. (2025). Analytical review of methods and technologies for real-time big data processing in IoT infrastructures. *Problems of Informatization and Management*, *2*(82), 19-25. https://doi.org/10.18372/2073-4751.82.20364
14. Khattach, O., Moussaoui, O., & Hassine, M. (2025). End-to-end architecture for real-time IoT analytics and predictive maintenance using stream processing and ML pipelines. *Sensors*, *25*, 2945. https://doi.org/10.3390/s25092945
15. Pustelnyk, P., & Levus, Y. (2025). Real-time anomaly detection in distributed IoT systems: A comprehensive review and comparative analysis. *Visnyk Natsionalnoho universytetu "Lvivska politekhnika". Seriia: Informatsiini systemy ta merezhi, 17*, 160-169. https://doi.org/10.23939/sisn2025.17.160
16. Dinakar, R. (2024). Real-time IoT sensor data streaming and processing with Apache Flink: A scalable solution for smart monitoring. *Journal of Electrical Systems*, *20*(11s), 3175-3181. https://doi.org/10.52783/jes.8042
17. Sheeraz, M., et al. (2024). Revolutionizing SIEM security: An innovative correlation engine design for multi-layered attack detection. *Sensors*, *24*(15), 4901. https://doi.org/10.3390/s24154901
18. Taha, K. (2025). Big data analytics in IoT, social media, NLP, and information security: Trends, challenges, and applications. *Journal of Big Data*, *12*, 150. https://doi.org/10.1186/s40537-025-01192-9

Отримано редакцією журналу / Received: 03.02.26
Прорецензовано / Revised: 16.02.26
Схвалено до друку / Accepted: 25.06.26

(cc) BY-NC-SA
This work is licensed under Creative Commons Attribution-noncommercial-sharealike 4.0 International License.