Міністерство освіти і науки України
Харківський національний університет імені В. Н. Каразіна

ВІСНИК

Харківського національного університету імені В.Н. Каразіна

Серія
«Математичне моделювання.
Інформаційні технології.
Автоматизовані системи управління»

Випуск 66

Серія заснована 2003 р.

BULLETIN

of V.N. Karazin Kharkiv National University

Series
«Mathematical Modeling.
Information Technology.
Automated Control Systems»

Issue 66

First published in 2003

Харків
2025

Засновник журналу Харківський національний університет імені В. Н. Каразіна, Харків, Україна. Рік заснування 2003. Періодичність: 4 випуски на рік. https://periodicals.karazin.ua/mia

Статті містять дослідження у галузі математичного моделювання та обчислювальних методів, інформаційних технологій, захисту інформації. Висвітлюються нові математичні методи дослідження та керування фізичними, технічними та інформаційними процесами, дослідження з програмування та комп'ютерного моделювання в наукоємних технологіях.

Для викладачів, наукових працівників, аспірантів, працюючих у відповідних або суміжних напрямках.

Наказом Міністерства освіти і науки України від 17.03.2020 № 409 наукове фахове періодичне видання Вісник Харківського національного університету імені В.Н. Каразіна серія «Математичне моделювання. Інформаційні технології. Автоматизовані системи управління» включено до Категорії «Б» Переліку наукових фахових видань України за наступними спеціальностями: 113 – Прикладна математика;122 – Комп'ютерні науки та інформаційні технології; 123 – Комп'ютерна інженерія; 125 – Кібербезпека.

Затверджено до друку рішенням Вченої ради Харківського національного університету імені В. Н. Каразіна (протокол № 17 від 30.06.2025 р.)

Адреса редакційної колегії: 61022, м. Харків, майдан Свободи, 6, Харківський національний університет імені В. Н. Каразіна, к. 534.
Тел. +380 (57) 705-42-81, Email: journal-mia@karazin.ua.

Мова публікації: українська, англійська.

Статті пройшли внутрішнє та зовнішнє рецензування.

Ідентифікатор медіа у Реєстрі суб'єктів у сфері медіа: R30-04456
(Рішення № 1538 від 09.05.2024 р Національної ради України з питань телебачення і радіомовлення. Протокол № 15)

The articles are present research in the field of mathematical modeling and computing methods, information technologies, information security. New mathematical methods of research and management of physical, technical and information processes, research on programming and computer modeling in science-intensive technologies are covered.

For teachers, researchers, graduate students working in relevant or related fields.

By the order of the Ministry of Education and Science of Ukraine from 17.03.2020 № 409 scientific professional periodical Bulletin of V.N. Karazin Kharkiv National University series "Mathematical modeling. Information Technologies. Automated control systems" is included in Category "B" of the List of scientific professional publications of Ukraine in the following specialties: 113 – Applied Mathematics, 122 – Computer Science and Information Technology; 123 – Computer engineering; 125 – Cybersecurity.

Approved for publication by the decision of the Academic Council of V.N. Karazin Kharkiv National University (Minutes № 17 of 30.06.2025).

Editorial Address: 61022, Kharkiv, Svobodi sq., 6, V.N. Karazin Kharkiv National University, r. 534.
Phone. +380 (57) 705-42-81, Email: journal-mia@karazin.ua.

Language of publication: Ukrainian, English.

The articles pass internal and external review.

Media identifier in the Register of the field of Media Entities: R30-04456
(Decision № 1538 dated May 9, 2024 of the National Council of Television and Radio Broadcasting of Ukraine, Protocol № 15)

ЗМІСТ

CONTENTS

DOI: https://doi.org/10.26565/2304-6201-2025-66-01

УДК (UDC) 004.051

Гамзаєв Рустам Олександрович *к.т.н., доцент; доцент кафедри інтелектуальних програмних систем і технологій, ННІ комп'ютерних наук та штучного інтелекту, Харківський національний університет імені В. Н. Каразіна, майдан Свободи 4, м. Харків, Україна, 61022*
e-mail: rustam.gamzayev@karazin.ua
https://orcid.org/0000-0002-2713-5664

Інтелектуальна інформаційна технологія підтримки мінливості процесів життєвого циклу програмного забезпечення кіберфізичних систем

Актуальність. Розробка програмного забезпечення (ПЗ) кіберфізичних систем (КФС) має враховувати специфічні особливості їх побудови та функціонування, що передбачає можливість підтримки мінливості проектних ресурсів та системних рішень на всіх основних етапах життєвого циклу (ЖЦ) КФС. Вирішення цих проблем неможливо без використання інтелектуальних методів та засобів і тому тематика цього дослідження є актуальної науково-технічною задачею.
Мета. Метою роботи є розробка інтелектуальної інформаційної технології (ІІТ), яка забезпечує наскрізну підтримку мінливості проектних активів на всіх основних фазах ЖЦ ПЗ КФС, що, у кінцевому рахунку, має підвищити показники якості критично важливих процесів розробки та супроводу таких систем.
Методи дослідження. На основі критичного аналізу та методологічного узагальнення деяких вже отриманих раніше наукових та практичних результатів, розроблена структурно-функціональна схема ІІТ, яка інтегрує знання-орієнтовані модельно-технологічні засоби, що дозволяє забезпечувати підтримку властивостей варіабельності, адаптивності, конфігурування та настроюваності проектних рішень та програмних компонентів КФС на етапах доменного інжинірингу, архітектурного проектування, конструювання коду та супроводу компонентів її ПЗ.
Результати. На прикладах систем «Розумний будинок» та мобільних систем доповненої реальності досліджені деякі суттєві особливості побудови та функціонування КФС, сформовано методологічний базис для знання-орієнтованої розробки ПЗ таких систем. Запропонована узагальнена схема ІІТ в нотації IDEF0, визначені її основні функціональні блоки, проведені програмні експерименти та обчислені кількісні метрики, які показали сумарне зростання показників якості процесів розробки та супроводу ПЗ приблизно на 22,4%.
Висновки. Представлені дослідження підтвердили доцільність застосування знання-орієнтованих моделей, методів та інструментальних засобів для розробки та супроводу ПЗ КФС, і можливість створення наскрізної інтелектуальної інформаційної технології, яка підтримує властивості змінності проектних ресурсів та системних рішень на всіх основних фазах ЖЦ КФС і що, в свою чергу, дозволяє суттєво підвищити рівень якості процесів створення таких систем.

Ключові слова: *інтелектуальні моделі та методи, кіберфізична система, програмне забезпечення, життєвий цикл, інформаційна технологія, мінливість, варіабельність, адаптивність, конфігурування, налаштування, метрика, якість*

Як цитувати: Гамзаєв Р.О. Інтелектуальна інформаційна технологія підтримки мінливості процесів життєвого циклу програмного забезпечення кіберфізичних систем. *Вісник Харківського національного університету імені В.Н. Каразіна, серія Математичне моделювання. Інформаційні технології. Автоматизовані системи управління*. 2025. вип.. 66. С.6-18. https://doi.org/10.26565/2304-6201-2025-66-01

How to quote: Gamzayev R. O. "Intelligent information technology to support changeability in software life cycle processes of cyber-physical systems", *Bulletin of V.N. Karazin Kharkiv National University, series Mathematical modelling. Information technology. Automated control systems*, vol. 66, pp. 6-18, 2025. https://doi.org/10.26565/2304-6201-2025-66-01

1 Вступ. Мета та актуальність дослідження

Сучасні кіберфізичні системи (КФС) представляють собою складні апаратно-програмні комплекси, що безпосередньо поєднані з реальними фізичними процесами та пристроями у зовнішньому середовищі (яке, як правило, пов'язано з Internet), і які забезпечують при цьому

функціональність, необхідну для різних груп їх кінцевих користувачів. Структурно-функціональна складність, розподілена системна архітектура, гетерогенність ресурсів даних та деякі інші особливості КФС (див. більш детально нижче) зумовлюють необхідність використання нових підходів до розробки їх програмного забезпечення (ПЗ), а останнім часом беззаперечним трендом в цьому домені стає з широке застосування моделей, методів і технологій штучного інтелекту (artificial intelligence - AI). Зокрема, у деяких новітніх роботах з цієї тематики [1–3] розглядається використання технологій AI для підтримки ключових процесів життєвого циклу (ЖЦ) КФС, таких як інженерія вимог, моделювання даних, конструювання коду та автоматизоване тестування компонентів ПЗ. Окрім того, засоби AI дедалі частіше інтегруються в етапи супроводу й експлуатації ПЗ у рамках практик культури DevOps для реалізації автоматичного масштабування, самовідновлення (self-healing), інтелектуального моніторингу та оптимізації розгортання програмних мікросервісів, як це реалізується, зокрема, в операційних середовищах для їх оркестровці, на кшталт Kubernetes [4]. В той же час слід зазначити, що існує досить обмежена кількість публікацій щодо застосування методів доменного інжинірингу, знання-орієнтованих моделей підтримки варіабельності та адаптивності компонентів ПЗ КФС на етапах їх проектування та супроводу, а також недостатньо висвітлюються питання побудови лінійок програмних продуктів для КФС [5,6] як ефективного шляху у розробці таких систем для забезпечення можливостей їх адаптації, конфігурування та налаштування з урахуванням постійних змін у вимогах їх замовників, фахівців з супроводу та кінцевих користувачів.

Саме тому метою цього дослідження є розробка інтелектуальної інформаційної технології (ІІТ), яка забезпечує наскрізну та керовану підтримку мінливості проектних активів на всіх основних фазах ЖЦ ПЗ КФС, і що, у кінцевому рахунку, має підвищити показники якості критично важливих процесів розробки та супроводу таких систем.

2 Методологічний базис побудови інтелектуальної інформаційної технології для підтримки мінливості у процесах життєвого циклу ПЗ КФС

2.1 Особливості побудови та функціонування програмного забезпечення КФС

Беручи до уваги загальне визначення властивостей сучасних КФС [1] і з метою дослідження саме особливостей розробки та супроводу ПЗ таких систем, у цій роботі в подальшому розглядаються застосунки класу «Інтернет речей (Internet of Things -IoT / «Розумний будинок (Smart Home - SH)» [7], і мобільні системи доповненої реальності (mobile augmented reality system - MARS) [8].

Для IoT / SH систем, які підтримують взаємодію різних типів пристроїв та програмних компонентів для автоматизації процесів життєзабезпечення та їх керування користувачами у віддаленому режимі, можна зазначити такі особливості їх побудови та функціонування як [7]:

1) розподілена багаторівнева системна архітектура;
2) велика кількість взаємопов’язаних апаратних та програмних компонентів,
3) необхідність функціонування більшості з них в режимі «7/24».

Слід особливо зазначити, що останнім часом системи класу IoT виявляють новий тренд у своєму розвитку, який отримав назву «Internet of Military Things (IoMT)» - «Інтернет військових речей», або «Internet of Military Defense Things (IoMDT)» - «Інтернет речей військового захисту» [9], і що має виняткову важливість для сучасних потреб національної безпеки нашої держави.

Для систем класу MARS, які забезпечують візуалізацію навколишнього фізичного середовища та його доповнення іншими віртуальними об'єктами із використанням мультимедійної інформації (текст, графіка, відео і т.п.), характерними є наступні властивості (або вимоги) [10]:

1) функціонування з урахуванням обмежень на такі ресурси мобільних пристроїв (МП) як продуктивність процесора, обсяг оперативної пам’яті, ємність акумулятора МП, та розміри екрану;
2) велика кількість та різноманітність конфігурацій програмного та апаратного забезпечення МП у різних груп користувачів;
3) інтенсивне використання мережевого трафіку для обміну даними між окремими МП та серверами системи.

Вищезазначені особливості цих 2-х класів КФС узагальнено можуть бути сформульовані наступним чином: (I) велика структурно-функціональна складність та гетерогенність їх компонентних конфігурацій та інформаційних ресурсів; (II) необхідність використання цих

систем за умов обмежених апаратно-програмних ресурсів та змінного обчислювального навантаження. Саме вони й визначають основні вимоги до розробки та супроводу ПЗ систем КФС.

2.2 Концептуальне узагальнення поняття мінливості у процесах ЖЦ ПЗ КФС

Зважаючи на постійного зростання складності функціональних задач, які мають бути вирішені шляхом використання КФС, на теперішній час чітко визначилася тенденція до розробки вже не окремих, хоча й досить складних програмних продуктів, а до створення та подальшого ефективного використання сукупностей взаємопов'язаних програмних компонентів, які отримали назву лінійок програмних продуктів (ЛПП) [4-6]. Головними ознаками цих нових підходів слід вважати їх спрямованість на забезпечення варіативності (variability), адаптації (adaptivity), конфігурування (configurability) та налаштування (customizability) у процесах опрацювання вимог, розробки архітектурних рішень, програмних компонентів, функціоналу та користувацьких властивостей. Для концептуального узагальнення цих споріднених властивостей ПЗ на різних етапах його ЖЦ пропонується використовувати поняття «мінливості (changeability)», яке визначає здатність ПЗ зберігати (або керовано змінювати) свої основні показники якості за умов будь-яких передбачуваних (або запланованих) змін у процесах розробки та експлуатації КФС. Слід зазначити, що поняття мінливості в цілому у системній інженерії і, зокрема, його логічний зв'язок з поняттям надійності (robustness) вперше стало предметом публікації фахівців з *Massachusetts Institute of Technology (MIT)* [11], а у цьому дослідженні воно графічно подано у вигляді UML діаграми класів, яка наведена на рисунку 1.

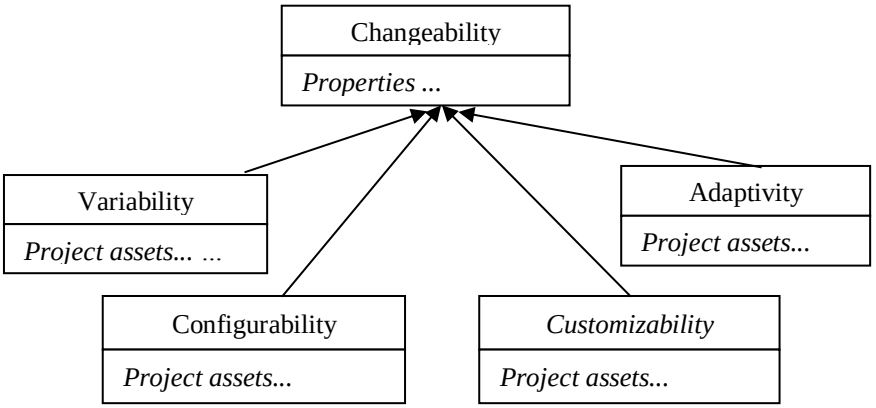


Рис. 1. - Концептуальна схема узагальнення різних типів мінливості у процесах ЖЦ ПЗ КФС
Fig. 1- Conceptual diagram of generalization of different types of variability in the life cycle processes of CPS software.

Властивість мінливості ПЗ може бути забезпечено шляхом розробки та застосування відповідних моделей, методів і інструментальних засобів на всіх основних етапах ЖЦ КФС, які позначені як *Project assets* (проектні активи) в UML класах на рисунку 1, і для структурованого представлення яких запропонована фреймова модель [12], фрагмент якої наведено на рисунку 2.

Frame_OMC: *"Operation Metamodel for Changeability in SPL lifecycle (LC)"*
{Slot_1: *"SPL LC Phases" = (Domain Engineering, Architectural Design, Code Construction, Software Deployment / Maintenance)*
Slot_2: "*Changeability subtypes* " *= (Variability, Adaptivity, Configurability, Customizability)*
Slot_3: *"Changeability Models" = (Feature models, Algorithmic models, DSL-based models, Recommendation models)***}**;

Frame_DM: *"Domain Engineering"*
{ Slot_1: *"Methods" = (FODA, ODM, ...)*
Slot_2: "*CASE tools*" *= (FeatureIDE, Eclipse EMF, ...)*
……. **}**

Рис. 2. - Фрагмент фреймової моделі мінливості процесів та артефактів у ЖЦ ПЗ КФС
Fig. 2 - Fragment of the frame model of variability of processes and artifacts in the CPS software life cycle.

Особливості побудови та переваги застосування цієї моделі обговорюються в [12], де також наведені більш детальні відомості про задіяні в ній конкретні методи доменного інжинірингу (фреймовий слот *"Methods"*): FODA (Feature Oriented Domain Analysis), ODM (Organization Domain Modeling) та відповідні інструментальні CASE-засоби для їх практичного застосування (фреймовий слот **"***CASE tools***"**): *FeatureIDE, Eclipse EMF* та деякі інші.

2.3 Основні методологічні принципи розробки ПЗ КФС

На основі проведеного аналізу особливостей КФС (див. 2.1) та концептуального узагальнення поняття мінливості на різних фазах ЖЦ ПЗ (див. 2.2) можливо сформулювати основні методологічні принципи розробки ПЗ КФС у наступний спосіб [13]:

(1) доменне моделювання (domain modeling) властивостей ПЗ КФС із використанням багатовимірних методів опрацювання експертних знань (handling of expert knowledge);
(2) архітектурне проектування ПЗ (software architecting) з можливістю адаптації програмно-апаратних ресурсів КФС, зокрема, із використанням методів логічного виводу на основі аналізу прецедентів (case-based reasoning);
(3) розробка лінгвістичних моделей і спеціальних мовних засобів (domain-specific language) для автоматизації процесів конструювання коду програмних компонентів КФС;
(4) використання методів та технологій штучного інтелекту, зокрема, шляхом побудови рекомендаційних систем (recommender system) для динамічного конфігурування апаратно-програмних компонентів КФС.

Для практичного застосування цих підходів у процесах розробки ПЗ на різних фазах ЖЦ КФС потрібна інтелектуальна інформаційна технологія, основні проектні та програмні рішення якої раніше вже були отримані в реальних застосунках та представлені в [7,10,14-20].

3 Узагальнена схема інтелектуальної інформаційної технології (ІІТ) та підхід до сумарної оцінки підвищення показників якості процесів ЖЦ ПЗ КФС

3.1 Функціональна схема запропонованої ІІТ в нотації IDEF0

На рисунку 3 наведена схема запропонованої ІІТ з використанням нотації IDEF0.

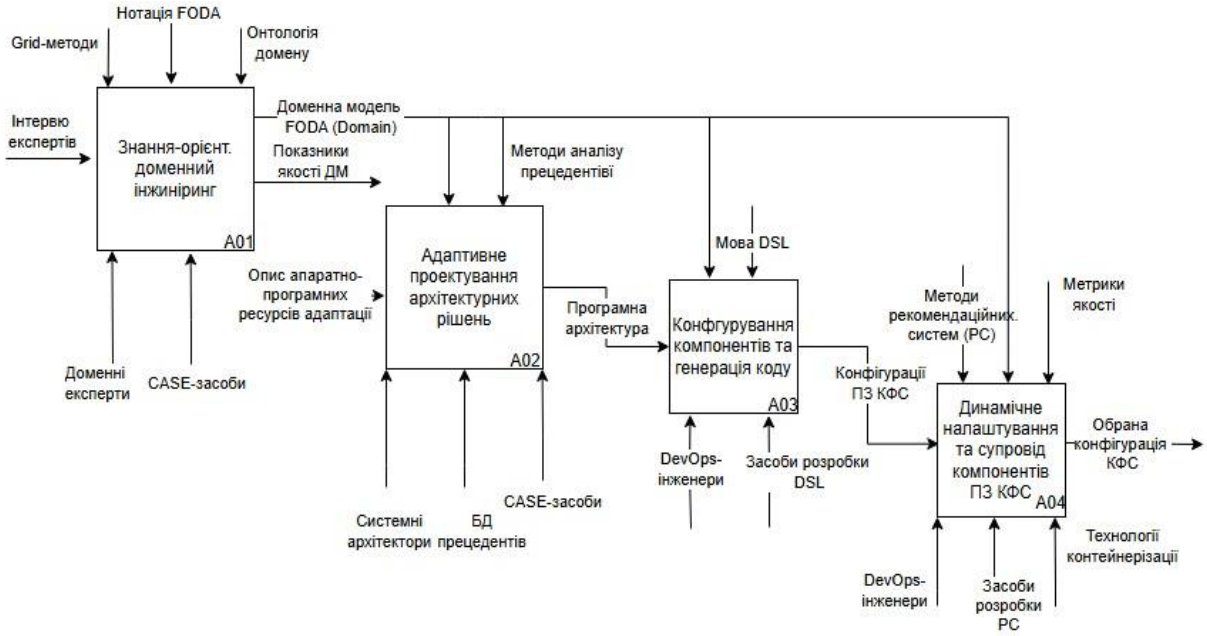


Рис. 3 - Узагальнена схема запропонованої ІІТ в нотації IDEF0

Fig. 3 - Generalized diagram of the proposed IIT in the IDEF0 notation.

У її складі міститься 4 основних функціональних блоків (ФБ), які технологічно забезпечують послідовну реалізацію основних методологічних принципів (1)-(4) розробки ПЗ КФС (див. п. 2.3) із застосуванням конкретних моделей, методів та CASE-засобів (див. п. 2.2). Нижче кожен з цих

4-х ФБ розглянуто більш детально, а також запропоновано підхід до отримання сумарної оцінки підвищення показників якості процесів ЖЦ ПЗ КФС в результаті послідовного застосування цієї ІІТ. Для опису роботи кожного ФБ використовуються стандартні терміни визначення його інтерфейсів: *Input, Control, Mechanism, Output* в нотації IDEF0.

3.2 Функціональний блок (ФБ) "Знання-орієнтований доменний інжиніринг"

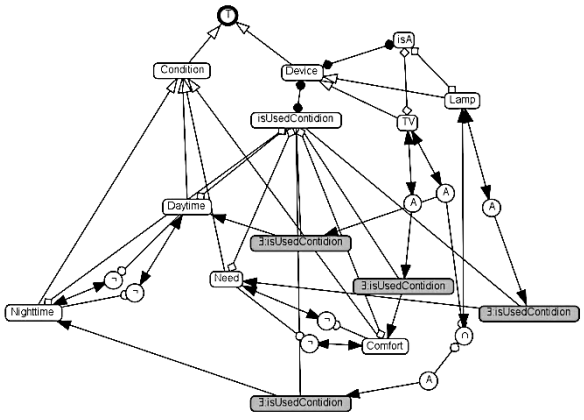
Цей ФБ, який позначено на рисунку 3 як A01, забезпечує технологічний процес побудови доменної моделі предметної області (ПрО) застосування відповідної КФС із використання додаткових знань експертів, розробників та майбутніх користувачів її функціональних можливостей. Цей процес побудови розширеної, знання-орієнтованої ДМ для систем класу IoT/SH (див. вище у п. 2.1) детально розглянуто в [14,15], а його основні функції виглядають наступним чином (див. рисунок 3):

- *вхідними даними* (*Input*) для ФБ A01 є текстовий опис інтерв'ю, отриманих від групи експертів у цій ПрО, що у загальному випадку є неструктурованим гетерогенним інформаційним ресурсом, для ефективної обробки якого потрібно застосувати додаткові методи опрацювання експертних знань;
- *алгоритм управління* (*Control*) роботою ФБ A01 використовує метод побудови репертуарних сіток (repertory grid method) [14], що дозволяє отримати багатовимірну модель експертних знань відповідної ПрО, або доменна модель (ДМ), яка враховує технічні, соціальні (користувацькі) та економічні вимоги до розробки майбутньої КФС, і для формалізованого представлення цієї моделі мотивовано обрано застосування нотації FODA (Feature-oriented Design Analysis), також додатково може бути використана одна з вже існуючих онтологій для цього домену [15];
- *механізм реалізації* (*Mechanism*) ФБ A01 передбачає участь у цьому процесі інженерів з обробки знань та доменних експертів, які застосовують відповідні CASE-засоби: системи Protégé, GridSuite, SOVA [15];
- *результатом роботи* (*Output*) ФБ A01 є знання-орієнтована ДМ, яка позначена як *FODA (Domain)*, що слугує інформаційним ресурсом для управління роботою всіх інших ФБ з метою забезпечення властивостей мінливості на інших фазах ЖЦ ПЗ КФС, крім того, для подальшого використання в системі розраховуються показники якості цієї ДМ (див. рисунок 3).

Приклад фрагменту побудованої у такий спосіб варіабельної FODA – моделі для ПрО «Розумний будинок» [15] представлено на рисунку 4.

```
</ObjectProperty>
  <!-- http://example.org/isUsedInCondition1 -->
  <ObjectProperty rdf:about="http://example.org/isUsedInCondition1">
    <rdfs:domain rdf:resource="http://example.org/AudioSpeakers"/>
    <rdfs:range rdf:resource="http://example.org/Comfort"/>
    <rdfs:range rdf:resource="http://example.org/Need"/>
  </ObjectProperty>
  <!-- http://example.org/isUsedInCondition2 -->
  <ObjectProperty rdf:about="http://example.org/isUsedInCondition2">
    <rdfs:domain rdf:resource="http://example.org/AudioSpeakers"/>
    <rdfs:domain rdf:resource="http://example.org/DoorOpeningSensor"/>
    <rdfs:domain rdf:resource="http://example.org/LightSensors"/>
    <rdfs:domain rdf:resource="http://example.org/MotionSensors"/>
    <rdfs:range rdf:resource="http://example.org/Always"/>
    <rdfs:range rdf:resource="http://example.org/Sometimes"/>
  </ObjectProperty>
  <!-- http://example.org/isUsedInCondition3 -->
  <ObjectProperty rdf:about="http://example.org/isUsedInCondition3">
    <rdfs:domain rdf:resource="http://example.org/AudioSpeakers"/>
    <rdfs:domain rdf:resource="http://example.org/DoorOpeningSensor"/>
    <rdfs:domain rdf:resource="http://example.org/LightSensors"/>
    <rdfs:domain rdf:resource="http://example.org/MotionSensors"/>
    <rdfs:domain rdf:resource="http://example.org/SmartTV"/>
    <rdfs:range rdf:resource="http://example.org/Daytime"/>
    <rdfs:range rdf:resource="http://example.org/Nighttime"/>
</ObjectProperty>
```

(а)



(б)

Рис. 4 - Результати роботи ФБ A01: (а) – опис варіабельної FODA-моделі в форматі XML / OWL, (б) - семантичний онтограф, який згенеровано на основі опису FODA - моделі

Fig. 4 - Results of the A01 FB operation: (a) – description of the variable FODA model in XML / OWL format, (b) – semantic ontograph generated based on the description of the FODA mode

Результати експериментального дослідження показників структурно-функціональної складності ДМ [16], побудованих із використанням запропонованого підходу, у порівнянні зі

складністю ДМ, отриманих із застосуванням стандартних методів доменного моделювання, таких, як ODM (Organizational Domain Modeling) та JODA (Joint integrated avionics Object oriented Domain Analysis), показали зменшення їх складності приблизно на 21,1%.

3.3 Функціональний блок "Адаптивне проектування архітектурних рішень"

Після проведення доменного інжинірингу як першого із основних етапів ЖЦ ПЗ КФС, наступним важливим кроком з точки зору забезпечення можливостей мінливості її проектних активів (див. п. 2.3) є етап адаптивного проектування архітектурних рішень, які мають враховувати загальні особливості КФС, що сформульовані у п. 2.1. Для забезпечення цих процесів у запропонованій ІІТ передбачений ФБ А02 (див. схему на рисунку 3), а загальний алгоритм його роботи може бути представлений у наступний спосіб (більш детально вони представлені та досліджені у [10,17]:

- *вхідними даними* (*Input*) для ФБ А02 є опис апаратно-програмних ресурсів КФС, які мають бути використані для розробки адаптивних архітектурних рішень;
- *алгоритм управління* (*Control*) базується на використанні вже побудованої доменної моделі нотації *FODA (Domain)* (див. п. 3.2), а для пошуку проектних рішень, враховуючи високу складність та слабоформалізований характер процесів архітектурного проектування систем КФС, використовуються методи логічного висновку на основі аналізу прецедентів (case-based reasoning - CBR) і набір кількісних метрик для оцінки поточного стану ресурсів адаптації та показників якості функціонування КФС (конкретні приклади таких метрик для мобільних систем доповненої реальності наведені в [9]);
- *механізм реалізації* (*Mechanism*) ФБ А02 передбачає участь у цьому процесі системних архітекторів, які застосовують базу даних прецедентів та інструментальне ПЗ (див. більш детально в [16]);
- *результатом роботи* (*Output*) ФБ А02 є системна програмна архітектура для визначеного типу КФС, яка має враховувати можливості її адаптації до змін в оточуючому середовищі її функціонування, обмеження на наявні обчислювальні ресурси та ін.

В [9] розроблена еталонна 3-х рівнева програмна архітектура для систем MARS із вбудованим блоком адаптивного управління, яка у вигляді UML діаграми розміщення компонентів представлена на рисунку 5.

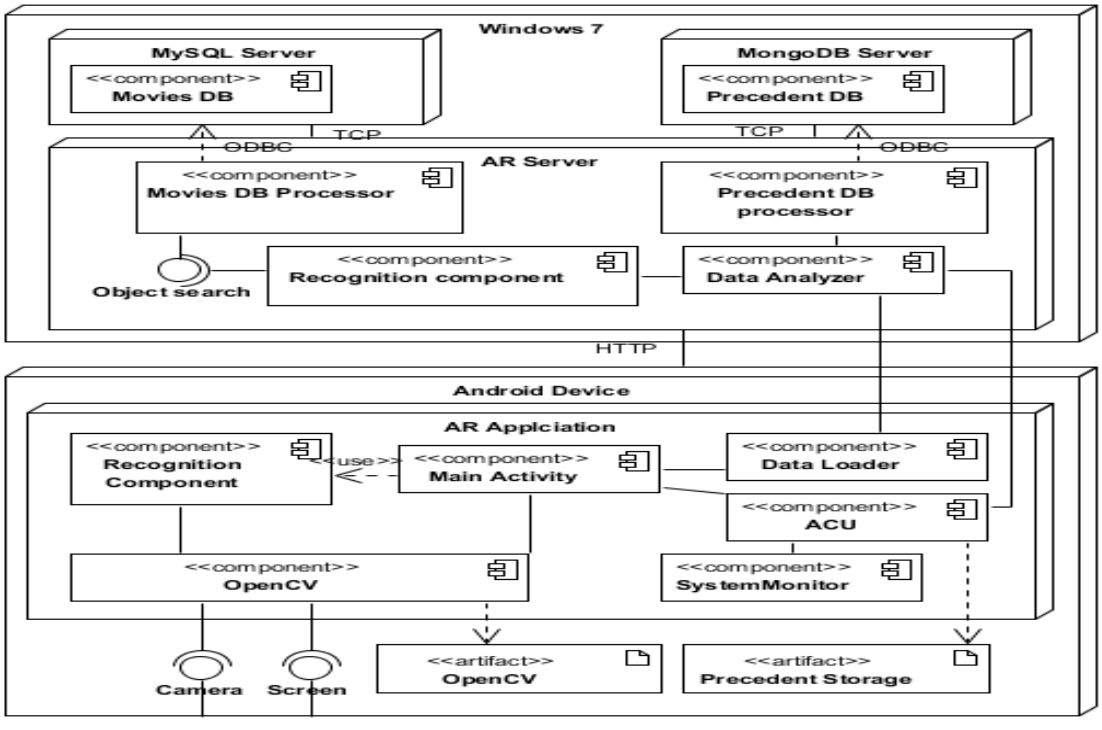


Рис. 5 - Архітектура систем MARS із вбудованим блоком адаптивного управління
Fig. 5 - Architecture of MARS systems with an embedded adaptive control unit.

Основні функціональні компоненти МСДР, які забезпечують можливість адаптації системи, є наступні [16]

1) на рівні клієнтських застосунків (AR Application): ACU (Adaptive Control Unit) – це блок адаптивного управління (БАУ) на основі CBR-методів, Precedent Storage – локальне сховище даних для поточних прецедентів (Berkeley DB engine), System Monitor – компонент для моніторингу стану ресурсів адаптації наявних МП (процесор, пам'ять, відео-камера та ін.);

2) на рівні серверу бізнес-логіки (AR Server): Data Analyzer – це компонент для аналізу даних про стан клієнтських застосунків; Precedent DB Processor – процесор БД прецедентів;

3) на рівні серверів БД (DB Server): Precedent DB – це БД прецедентів (MongoDB).

В [16] наведені результати експериментального дослідження впливу роботи БАУ на можливості коригування такого важливого показника якості функціонування системи MARS як роздільна здатність екрану відеокамери на клієнтському застосунку. В середньому, в залежності від рівня завантаженості обчислювальних ресурсів відповідного МП, використання запропонованого підходу дозволило підвищити якість зображень приблизно на 28,2%.

3.4 Функціональний блок "Конфігурування компонентів та генерація коду"

Після проведення етапу архітектурного проектування, як другого з їх послідовності у ЖЦ ПЗ КФС, запропонована ІІТ передбачає можливість застосування інтелектуальних засобів для автоматизації процесів конфігурування програмних компонентів та генерування вихідного коду цільової системи. У роботі [17], на основі аналізу можливих підходів до вирішення цих задач для забезпечення варіативності отриманих програмних рішень із урахуванням вимог до якості ПЗ різних груп користувачів та розробників КФС, мотивовано обрано застосування концепції та технології створення предметно-орієнтованих мов програмування (domain-specific language – DSL). Для забезпечення цих процесів у запропонованій ІІТ передбачений ФБ А03 (див. схему на рисунку 3), який забезпечує вирішення наступних основних задач (більш детально вони представлені та досліджені у [6,17]):

- *вхідними даними* (*Input*) для ФБ А03 є компонентна програмна архітектура із відповідними функціональними можливостями її адаптації, яка створюється в результаті роботи ФБ А02 (приклад такої архітектури для систем MARS наведено на рисунку 5);
- *алгоритм управління* (*Control*) роботою ФБ А03, як і ФБ А02, використовує отриману раніше у ФБ А01 варіабельну *FODA (Domain)* модель визначеної ПрО (див. п. 3.2), а також передбачає розробку та застосування відповідної мови DSL;
- *механізм реалізації* (*Mechanism*) ФБ А03 потребує участі у цьому процесі DevOps фахівців, які застосовують засоби розробки та використання DSL (напр., на мові Python), а також інструментарій для роботи з мовами програмування загального призначення, напр., C++ /C#, Java та ін. (див. більш детально в [17]);
- *результатом роботи* (*Output*) ФБ А03 є опис альтернативних конфігурацій програмних компонентів КФС на мові DSL, на основі яких генерується їх вихідний код.

Формалізовано мова DSL визначається як кортеж наступного вигляду

$$DSL = \langle FODA\ (Domain),\ Grammar\ (Lexer;\ Parser),\ Code\_Generator \rangle, \quad (1)$$

де: *FODA (Domain)* – це модель варіабельності властивостей ПЗ для цільової КФС у певній ПрО *(Domain); Grammar (Lexer, Parser)* - це набір моделей, алгоритмів, ресурсів даних і програмних компонентів, які використовуються для реалізації всіх граматичних правил для цього DSL, включаючи лексичний аналіз *(Lexer)* і синтаксичний аналіз *(Parser)* для вхідного коду; *Code_Generator* - це механізм генерації вихідного коду, який перетворює синтаксично та семантично перевірений скрипт на DSL у послідовність інструкцій мови програмування загального призначення.

Важливо підкреслити, що визначення DSL за виразом (1) забезпечує модельно-керований підхід до його розробки, тому що задіяна у ньому модель *FODA (Domain)* відображає всі структурні та семантичні особливості відповідної ПрО. У [17] це проілюстровано на прикладі

розробки DSL для ПрО «Розумний будинок», де наведені конкретні фрагменти скриптів на мові DSL та блоки згенерованого вихідного коду на мові C++ .

Результати експериментального дослідження [6,17] використання мови DSL для автоматизації конфігурування програмних компонентів системної архітектури та генерації їх вихідного коду у застосунках для ПрО «Розумний будинок» показали, що таким шляхом забезпечується зростання ефективності цих процесів приблизно на 16,8%.

3.5 Функціональний блок "Динамічне налаштування та супровід компонентів ПЗ КФС"

Заключним етапом застосування запропонованої ІІТ, який є необхідним у ЖЦ ПЗ КФС, враховуючи зміни у вимогах їх користувачів, є задача динамічного налаштування конфігурацій її програмних компонентів у процесі супроводу відповідної КФС. Для її ефективного вирішення також доцільно застосування інтелектуальних методів і технологій, зокрема, одним з можливих варіантів є використання рекомендаційних систем (recommender system - RS) [18], які забезпечують накопичення та подальше опрацювання консолідованої інформації про стан КФС, що уможливлює динамічне налаштування її функціоналу для різних груп користувачів у відповідності зі змінами в навколишньому середовищі. У контексті цього дослідження формальне визначення RS може бути подано у наступний спосіб [18]:

1) нехай $\underline{U}$ – це є множина користувачів КФС, $\underline{C}$ – це є множина всіх її програмних компонентів, які мають такі функціональні властивості, що відповідають вимогам множини її користувачів $u \subseteq \underline{U}$, а множина $\underline{R} \subseteq \underline{C}$ – це є ранжируваний список підмножини таких компонентів, $r \in \underline{R}$, - це є певний компонент у списку $\underline{R}$;

2) результат застосування RS полягає в тому, щоб забезпечити вибір $r \in \underline{R}$ таким чином, аби це максимально повно відповідало вимогам певної групи користувачів: $u \subseteq \underline{U}$;

3) тоді якщо E – це деяка метрика оцінки задоволеності певного користувача ПЗ КФС, $a\ f\ (r,u)$ це функція оцінки важливості рекомендації окремого компоненту $r \in \underline{R}$ для визначених користувачів $u \subseteq \underline{U}$, то проблема генерації рекомендації в RS може бути подана як вирішення задачі пошуку максимуму цільової функції $f(r,u) = E \rightarrow max$.

Всі ці завдання у загальній схемі ІІТ (див. рисунок 3) вирішує ФБ А04, алгоритм роботи якого спрощено може бути пояснений у наступний спосіб:

- *вхідними даними* (*Input*) для ФБ А04 є опис альтернативних (мінливих) конфігурацій програмних компонентів КФС на мові DSL, а також масиви контекстних даних, які відображають стан середовища функціонування цієї системи та наявні потреби (або майбутні вподобання) її користувачів;
- *алгоритм управління* (*Control*) роботою ФБ А04, як і ФБ А02 та ФБ А03, використовує отриману раніше у ФБ А01 варіабельну *FODA (Domain)* модель визначеної ПрО (див. п. 3.2), а також передбачає застосування методів і технологій RS, таких як, наприклад, спільна фільтрація (collaboratibe filtering), кластеризація, N-мірна тензорна факторизація та деякі ін. [19];
- *механізм реалізації* (*Mechanism*) ФБ А04 передбачає участь у цьому процесі фахівців з практик DevOps, які застосовують інструментальні засоби розробки RS, зокрема, проекти з відкритим кодом, такі як *Racoon, LensKit for Python, CARSkit* (див. більш детально в [18]), крім того, для реалізації розміщення програмних компонентів обраної у такий спосіб конфігурації мають бути застосовані різні технології контейнеризації [4];
- *результатом роботи (Output*) ФБ А04 є обрана конфігурація програмних компонентів КФС, які відповідають певним критеріям якості функціонування цільової системи.

На рисунку 6 наведена UML діаграма варіантів використання тестової RS, яка була розроблена на основі проекту з відкритим кодом *CARSkit* для забезпечення динамічного налаштування системних конфігурацій у застосунках для ПрО «Розумний будинок».

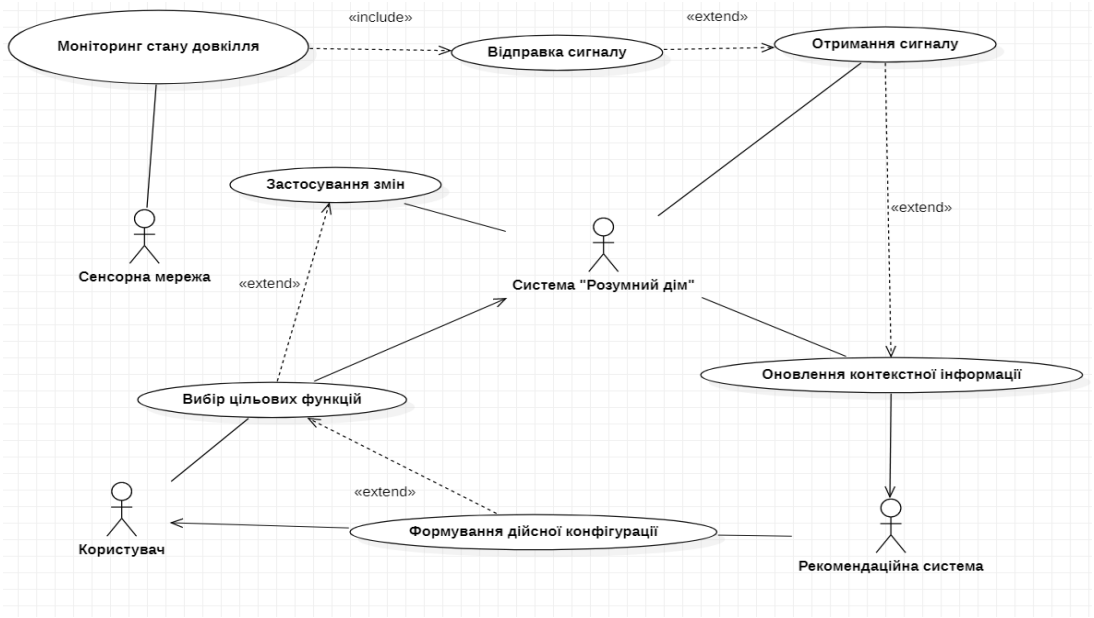


Рис. 6 - Сценарії використання RS для налаштування конфігурацій системи «Розумний будинок»
Fig. 6 - RS usage scenarios for configuring the "Smart Home" system.

Результати експериментального дослідження оцінки точності алгоритмів прогнозування потрібних конфігурацій КФС із застосуванням технологій RS показали [19,20], що таким шляхом забезпечується зростання точності прогнозування приблизно на 24,2%.

3.6 Емпіричний підхід до визначення сумарної оцінки підвищення якості процесів ЖЦ ПЗ КФС у результаті застосування запропонованої ІІТ

Для отримання сумарної оцінки підвищення якості процесів ЖЦ КФС, що може бути отримана з використанням розробленої ІІТ (див. п. п. 3.1-3.5), пропонується враховувати деякі наявні емпіричні дані про середні питомі проектні витрати на основних етапах ЦЖ ПЗ.

Зокрема, у відомій роботі [21] наведено узагальнені статистичні дані з цих питань, на основі яких можливо визначити відповідні коефіцієнти питомої ваги (важливості) для значень локальних оцінок показників якості процесів на основних етапах ЖЦ ПЗ КФС, а саме:

- для фази специфікації вимог (20% витрат): $K_1 = 0{,}20$ (2),
- для фази архітектурного проектування (15% витрат): $K_2 = 0{,}15$ (3),
- для фази розробки (25% витрат): $K_3 = 0{,}25$ (4),
- для тестування та розгортання (40% витрат): $K_4 = 0{,}40$ (5),

де сума значень K_{Sum} цих коефіцієнтів (K_1-K_4)дорівнює одиниці: $K_{Sum} = 0{,}20+0{,}15+0{,}25+0{,}40 = 1{,}0$.

Тоді, враховуючи значення локальних показників якості процесів окремих етапів ЖЦ ПЗ, які наведені у п. п. 3.2-3.5 (у %), та використовуючи коефіцієнти з виразів (2) – (5), сумарну середню оцінку підвищення якості процесів розробки ПЗ КФС Q_{Sum} можна розрахувати наступним чином:

$$Q_{Sum} = K_1 * 21.1\% + K_2 * 28.3\% + K_3*16.8\% + K_4 * 24.2\ \% = \\ = (0.20 * 21.1 + 0.15 * 28.3 + 0.25*16.8 + 0.4* 24.2)\ \% = 22.4\ \%. \quad (6)$$

Запропонований підхід є, безперечно, емпіричним за своєю природою, і отримана оцінка (6) суттєво залежить від наявних статистичних даних щодо проектних витрат на всіх основних етапах ЖЦ ПЗ. У якості більш точного механізму для її визначення може бути застосований один з існуючих структурованих методів прийняття багатокритеріальних експертних рішень, наприклад, метод аналізу ієрархій (Analytical Hierarchy Process) [22].

4 Висновки та напрямки подальших досліджень

У цьому дослідженні розглянута актуальна науково-технічні задачі розробки інтелектуальної інформаційної технології (ІІТ), яка забезпечує наскрізну підтримку властивостей мінливості проектних активів на всіх основних фазах ЖЦ КФС і що, у кінцевому рахунку, дозволяє

підвищити показники якості критично важливих процесів розробки та супроводу таких систем. На основі методологічного узагальнення вже отриманих раніше наукових та практичних результатів, розроблена структурно-функціональна схема ІІТ в нотації IDEF0, яка інтегрує окремі знання-орієнтовані моделі, методи та інструментальні засоби для забезпечення властивостей варіабельності, адаптивності, конфігурування та настроюваності проектних рішень та програмних компонентів КФС відповідно на етапах доменного інжинірингу, архітектурного проектування, конструювання коду та супроводу компонентів її ПЗ. Проведені програмні експерименти у реальних проектах з розробки та супроводу застосунків «Розумний будинок» та прототипу мобільної системи доповненої реальності показали можливість та довели доцільність створення наскрізної ІІТ, яка підтримує властивості мінливості системних рішень на різних фазах ЖЦ ПЗ КФС і що, в свою чергу, дозволяє підвищити рівень якості таких систем у середньому на 22,4 %.

У якості подальших напрямків досліджень по цій проблематиці можливо запропонувати розробку та експериментальну перевірку можливостей застосування альтернативних модельно-технологічних комплексів для підвищення інтелектуального рівня процесів на всіх основних етапах ЖЦ ПЗ КФС, з метою отримання керованого синергетичного ефекту [1,2] для підвищення якості функціонування таких систем.

СПИСОК ЛІТЕРАТУРИ

1. Lu Chengjie, Pablo Valle, Jiahui Wu et al., “Foundation Models for Software Engineering of Cyber-Physical Systems: The Road Ahead” in: arXiv:2504.04630 [cs.SE], Apr 2025. https://doi.org/10.48550/arXiv.2504.04630
2. Amar Banerjee, Venkatesh Choppella, “Control Software Engineering Approaches for Cyber-Physical Systems: A Systematic Mapping Study” *ACM Trans. Cyber-Phys. Syst.*, Vol. 9, Issue 1, Article No.: 6, pp. 1 – 33, 2025. https://doi.org/10.1145/3704737
3. Sanghoon Lee, Jiyeong Chae, Haewon Jeon et al., “Cyber-Physical AI: Systematic Research Domain for Integrating AI and Cyber-Physical Systems”, *ACM Trans. Cyber-Phys. Syst.* Vol. 9, Issue 2, Article No.: 19, pp. 1 – 33, 2025. https://doi.org/10.1145/3721437
4. Р. О. Гамзаєв, В.Х. Мурадова, М.В. Ткачук, “Дослідження альтернативних технологій контейнерів для віртуалізації процесів розміщення компонентів лінійок програмних продуктів”, *Вісник Харківського національного університету імені В.Н. Каразіна, Серія “Математичне моделювання. Інформаційні технології. Автоматизовані системи управління”,* № 53, 2022, с. 12-20. https://doi.org/10.26565/2304-6201-2022-53
5. Kristof Meixner, Kevin Feichtinger, Sayyid Fadhlillah et al., “Variability Modeling of Products, Processes, and Resources in Cyber-Physical Production Systems Engineering”, *Journal of Systems and Software* Vol. 211, 2024. https://doi.org/10.1016/j.jss.2024.112007
6. Jacob Krüger, Sebastian Nielebock, Sebastian Krieter et al., “Beyond Software Product Lines: Variability Modeling in Cyber-Physical Systems”, in *Proc. of the 21st International Systems and Software Product Line Conference* – Vol. A, 2018, pp. 237 – 241, https://doi.org/10.1145/3106195.3106217
7. Гамзаєв О. Р., Ткачук М. В., “Розробка проблемно-орієнтованої мови моделювання для підтримки варіабельності програмного забезпечення в системах "Розумний будинок". *Сучасний стан наукових досліджень та технологій в промисловості.* 2023. № 1 (23). С. 45–56. https://doi.org/10.30837/ITSSI.2023.23.045
8. Jacky Cao, Kit-Yung Lam, Lik-Hang Lee et al., “Mobile Augmented Reality: User Interfaces, Frameworks, and Intelligence”, *ACM Computing Surveys*, Vol. 55, Issue 9, 2018, pp. 1 – 36, 2023 https://doi.org/10.1145/3557999
9. Maurice Khabbaz, Abdellah Chehri, Holger Claussen et al., “Guest Editorial: The Internet of Military Defense Things: State-of-the-Art Challenges, Future Evolution, and Revolutionary Applications”, *IEEE Internet of Things Magazine*, Vol. 8, Issue 2, pp. 14-16, 2025. https://doi.org/10.1109/MIOT.2025.10907814
10. M. Tkachuk, A. Vekshin, and R. Gamzayev, “A Model-Based Framework for Adaptive Resource Management in Mobile Augmented Reality System”, in *Proc. of the ICTERI-2016: 12th International Conference on ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer*, Kyiv, Ukraine, June 21-24, 2016, CEUR-WS.org/Vol-1614, pp.41-56. https://doi.org/10.1007/978-3-319-69965-3_2

11. Adam M. Ross, Donna H. Rhodes, Daniel E. Hastings, "Defining changeability: Reconciling flexibility, adaptability, scalability, modifiability, and robustness for maintaining system lifecycle value", *Systems Engineering*, Vol. 11 (3), pp. 246-262, 2008. https://doi.org/10.1002/sys.20098
12. R. O. Gamzayev, "Frame-based Operation Metamodel to Chageability Support in Life Cycle of Software Product Lines", на *XXIII Всеукраїнська науково-технічна конференція молодих вчених, аспірантів та студентів*, Одеса, 20-21 квітня 2023 р. - Одеса, ОНТУ, 2023 р. – с. 215-217.
13. Р. О. Гамзаєв, "Знання-орієнтована інформаційна технологія забезпечення мінливості процесів і компонентів у життєвому циклі кіберфізичних систем", на *«Проблеми інформатики та моделювання (ПІМ-2023)», Тези XXIII Міжнародної науково-технічної конференції*, Харків: НТУ "ХПІ", 2023, с. 39.
14. R.O. Gamzayev, M.V. Tkachuk, D.O. Shevkoplias, "Handling of Expert Knowledge in Software Product Lines Development with Usage of Repertory Grids Method // Вісник Харківського національного університету імені В.Н. Каразіна, Серія «Математичне моделювання. Інформаційні технології. Автоматизовані системи управління». - № 47, 2020. – С. 13-24. https://doi.org/10.26565/2304-6201-2020-47-02
15. R.O. Gamzayev, M.V. Tkachuk, D.O. Shevkoplias, "Knowledge-oriented Information Technology to Variability Management on Domain Analysis Stage in Software Development", *«Сучасні інформаційні системи», наук.-техн. ж-л, НТУ «ХПІ»*, том 2, № 4, 2020. с. 39-47. https://doi.org/10.20998/2522-9052.2020.4.06
16. М.В. Ткачук, І.О. Мартінкус, К.А. Нагорний, Р. О. Гамзаев, "Про один підхід до оцінки ефективності застосування методів доменного моделювання при розробці сімейств програмних систем", *Збірник наукових праць Харківського національного університету Повітряних Сил*, № 5(54), 2017. – С. 127-134. http://nbuv.gov.ua/UJRN/ZKhUPS_2017_5_25
17. Mykola Tkachuk, Oleksii Vekshin, Rustam Gamzayev, "Architecting for Adaptive Resource Management in Mobile Augmented Reality Systems: Models, Metrics and Prototype Software Solutions", in *A. Genige et al. (Eds.): ICTERI 2016: Revised Selected Papers, Series title: Communications in Computer and Information Science*, Vol. 783: Springer-Verlag Berlin Heidelberg, 2017. – pp. 17-35. https://doi.org/10.1007/978-3-319-69965-3_2
18. Rustam Gamzayev, Mykola Tkachuk and Oleksandr Nelipa. "Domain-Specific Language for Adaptive Development of "Smart-Home" Applications", in *Proc. of the 1st International Workshop on Information Technologies: Theoretical and Applied Problems 2021 (ITTAP-2021)*, Ternopil, Ukraine, November 16-18, 2021, CEUR-WS.org/Vol-3039, pp.154-165.
19. Rustam Gamzayev, Mykola Tkachuk, "Dynamic Configuration of Software Products Lines for Smart-Home Applications based on Recommender Systems Framework", на *Інформаційні системи та технології: праці 10-ї Міжнародної науково-технічної конференції,* Харків - Одеса, 13-19 вересня 2021 року / наук. ред. А. Д. Тевяшев, Л. Б. Петришин, В.В. Бескоровайний, В.Г. Кобзєв. – ХНУРЕ, 2021, с. 85-89.
20. Р. О. Гамзаєв, М. В. Ткачук, "Застосування методів і технологій рекомендаційних систем для конфігурування динамічних лінійок програмних продуктів", *Вісник Національного технічного університету "ХПІ". Серія: Системний аналіз, управління та інформаційні технології: зб. наук. пр.,* Харків: НТУ "ХПІ", № 1 (5), 2021, с. 91-97. https://doi.org/10.20998/2079-0023.2021.01.15
21. Ian Sommerville, Software engineering (10th edition, Global Edition), Pearson Education, 2016.
22. Thomas L. Saaty, Luis G. Vargas, Models, Methods, Concepts & Applications of the Analytic Hierarchy Process, 2nd Edition, Springer Science+Business Media, New York, 2012.

REFERENCES

1. Chengjie Lu, Pablo Valle, JiahuiWu et al., "Foundation Models for Software Engineering of Cyber-Physical Systems: The Road Ahead" In: arXiv:2504.04630 [cs.SE], Apr 2025. https://doi.org/10.48550/arXiv.2504.04630
2. Amar Banerjee, Venkatesh Choppella, "Control Software Engineering Approaches for Cyber-Physical Systems: A Systematic Mapping Study", *ACM Trans. Cyber-Phys. Syst.*, Vol. 9, Issue 1 Article No.: 6, pp. 1 – 33, Jan 2025. https://doi.org/10.1145/3704737
3. Sanghoon Lee, Jiyeong Chae, Haewon Jeon et al., "Cyber-Physical AI: Systematic Research Domain for Integrating AI and Cyber-Physical Systems", *ACM Trans. Cyber-Phys. Syst.* Vol. 9, Issue 2, (April 19 2025), pp. 1 – 33. https://doi.org/10.1145/3721437

4. R.A. Gamzayev, V.Kh. Muradova, M.V. Tkachuk. “A study on alternative container-based technologies for virtualization of components deployment in software product lines”, *Bulletin of KhNU by V.N. Karazin, ser. "Mathematical modeling. Information Technology. Automated control systems"*, No. 53, 2022, pp. 12-20. [in Ukrainian] https://doi.org/10.26565/2304-6201-2022-53
5. Kristof Meixner, Kevin Feichtinger, Sayyid Fadhlillah et al., “Variability Modeling of Products, Processes, and Resources in Cyber-Physical Production Systems Engineering”, *Journal of Systems and Software*, Vol. 211, 2024. https://doi.org/10.1016/j.jss.2024.112007
6. Jacob Krüger, Sebastian Nielebock, Sebastian Krieter et al., “Beyond Software Product Lines: Variability Modeling in Cyber-Physical Systems”, in *Proc. of the 21st International Systems and Software Product Line Conference* – Vol. A, 2018, pp. 237 – 241. https://doi.org/10.1145/3106195.3106217
7. Rustam Gamzayev, Mykola Tkachuk, “Development of problem-specific modeling language to support software variability in "Smart home" systems”, *Innovative Technologies and Scientific Solutions for Industries*, 2023, No. 1 (23), P. 45-56. https://doi.org/10.30837/ITSSI.2023.23.045
8. Jacky Cao, Kit-Yung Lam, Lik-Hang Lee et al., “Mobile Augmented Reality: User Interfaces, Frameworks, and Intelligence”, *ACM Computing Surveys*, Vol. 55, Issue 9, 2018, pp. 1 – 36. https://doi.org/10.1145/3557999
9. Maurice Khabbaz, Abdellah Chehri, Holger Claussen et al., “Guest Editorial: The Internet of Military Defense Things: State-of-the-Art Challenges, Future Evolution, and Revolutionary Applications”, *IEEE Internet of Things Magazine*, Vol. 8, Issue 2, 2025, pp. 14-16. https://doi.org/10.1109/MIOT.2025.10907814
10. M. Tkachuk, A. Vekshin, and R. Gamzayev, “A Model-Based Framework for Adaptive Resource Management in Mobile Augmented Reality System”, in *Proc. of the ICTERI-2016: 12th International Conference on ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer*, Kyiv, Ukraine, June 21-24, 2016, CEUR-WS.org/Vol-1614, pp.41-56. https://doi.org/10.1007/978-3-319-69965-3_2
11. Adam M. Ross, Donna H. Rhodes, Daniel E. Hastings, "Defining changeability: Reconciling flexibility, adaptability, scalability, modifiability, and robustness for maintaining system lifecycle value", Systems Engineering, Vol. 11 (3), pp. 246-262, 2008. https://doi.org/10.1002/sys.20098
12. R. O. Gamzayev, “Frame-based Operation Metamodel to Chageability Support in Life Cycle of Software Product Lines”, in *Abstracts of the XXIII All-Ukrainian Scientific and Technical Conference of Young Scientists, Postgraduates and Students, Odessa*, 20-21 April 2023, ONTU, pp. 215-217.
13. R. O. Gamzayev, “Knowledge-oriented information technology to ensure a changeability on processes and components in life cycle of cyber-physical systems”, *Problems of Informatics and Modeling (PIM-2023). Abstracts of the XXIII International Scientific and Technical Conference*, Kharkiv: NTU "KhPI", 2023. – P. 129.
14. R.O. Gamzayev, M.V. Tkachuk, D.O. Shevkoplias, “Handling of Expert Knowledge in Software Product Lines Development with Usage of Repertory Grids Method”, *Bulletin of KhNU by V.N. Karazin, ser. "Mathematical modeling. Information Technology. Automated control systems"*, - No. 57, 2020. – pp. 13-24. https://doi.org/10.26565/2304-6201-2020-47-02
15. R.O. Gamzayev, M.V. Tkachuk, D.O. Shevkoplias, “Knowledge-oriented Information Technology to Variability Management on Domain Analysis Stage in Software Development”, *Advanced Information Systems*, 2020. Vol. 4, No. 4, pp. 39-47. https://doi.org/10.20998/2522-9052.2020.4.06
16. M.V. Tkachuk, I.O. Martinkus, K.A. Nagornyi, R.A. Gamzayev, “Towards the effectiveness assessment approach to domain modeling methods application in software product family development”, *Collection of scientific papers of the Kharkiv National University of the Air Force*, No. 5(54), 2017. – P. 127-134. [in Ukrainian] http://nbuv.gov.ua/UJRN/ZKhUPS_2017_5_25
17. Mykola Tkachuk, Oleksii Vekshin, Rustam Gamzayev, “Architecting for Adaptive Resource Management in Mobile Augmented Reality Systems: Models, Metrics and Prototype Software Solutions”, in *A. Genige et al. (Eds.): ICTERI 2016: Revised Selected Papers, Series title: Communications in Computer and Information Science*, Vol. 783: Springer-Verlag Berlin Heidelberg, 2017. – pp. 17-35. https://doi.org/10.1007/978-3-319-69965-3_2
18. Rustam Gamzayev, Mykola Tkachuk and Oleksandr Nelipa, “Domain-Specific Language for Adaptive Development of "Smart-Home" Applications”, in *Proc. of the 1st International Workshop*

on Information Technologies: Theoretical and Applied Problems 2021 (ITTAP-2021), Ternopil, Ukraine, November 16-18, 2021, CEUR-WS.org/Vol-3039, pp.154-165.
19. Rustam Gamzayev, Mykola Tkachuk. “Dynamic Configuration of Software Products Lines for Smart-Home Applications based on Recommender Systems Framework”, in *Information systems and technologies: Proceedings of the 10-th International Scientific and Technical Conference*, September 13-19, 2021 Kharkiv - Odesa, Ukraine, scientific editors: A.D. Tevyashev, L.B. Petryshyn, V.V. Bezkorovainy, V.G. Kobzev. – KhNURE, 2021. - pp. 85-89.
20. Rustam Gamzayev, Mykola Tkachuk. “Using Methods and Technologies of Recommendation Systems for Dynamic Software Product Lines Configuration”, *Bulletin of National Technical University "KhPI". Series: System Analysis, Control, and Information Technologies,* No. 1 (5), 2021, pp. 91-97. [in Ukrainian] https://doi.org/10.20998/2079-0023.2021.01.15
21. Ian Sommerville, Software engineering (10th edition, Global Edition), Pearson Education, 2016.
22. Thomas L. Saaty, Luis G. Vargas, Models, Methods, Concepts & Applications of the Analytic Hierarchy Process, 2nd Edition, Springer Science+Business Media, New York, 2012.

Gamzayev Rustam *PhD, Associate Professor; Associate Professor of the Department of Intelligent Software Systems and Technologies, Education and Research Institute of Computer Sciences and Artificial Intelligence, V. N. Karazin Kharkiv National University, 4 Svobody Sq., Kharkiv, 61022, Ukraine*

Intelligent information technology to support changeability in software life cycle processes of cyber-physical systems

Actuality. The development of software for cyber-physical systems (CPS) should take into account the specific features of their construction and operation, which supposes to support a changeability of project assets and system solutions at all main stages of their life cycle (LC). Solving these problems is impossible without the usage of intellectual methods and tools, and therefore the topic of this study is the actual scientific and technical task.
Goal. The aim of this work is to develop the intelligent information technology (IIT) that provides end-to-end support for the changeability of project assets in all major phases of the LC CFS, which finally has to improve the quality indicators of critically important development and maintenance processes of such systems.
Research methods. Based on a critical analysis and methodological generalization of some previously obtained scientific and practical results, the structural and functional scheme of IIT has been developed, which integrates some knowledge-oriented model-technological tools, which allows to ensure support for the properties of variability, adaptability, configuration and adjustability of design solutions and software components of the CFS at their LC stages such as domain engineering, architectural design, code construction, and maintenance of their software components.
Results. Using the examples of Smart home systems and mobile augmented reality systems development, some essential features of the CFS’s construction and operation have been studied, and the methodological basis for a knowledge-oriented software development of such systems has been formed. The generalized IIT scheme in IDEF0 notation has been proposed, its main functional blocks have been defined, software experiments have been conducted, and quantitative metrics are calculated, that shows the total increase in the quality indicators of software development and maintenance processes by approx. 22,4%.
Conclusions. The presented studies confirmed the feasibility of using knowledge-oriented models, methods and tools for the development and maintenance of CFS software, and the possibility to design the end-to-end intelligent information technology, which supports the properties of changeability of project assets and system solutions at the main phases of the LC CFS, that, in turn, allows to significantly processes quality improving in creation of such systems.

Keywords: *intelligent models and methods, cyber-physical system, software, life cycle, information technology, changeability, variability, adaptability, configuration, customization, metric, quality*

Надійшла у першій редакції 12.04.2025, в останній - 17.05.2025.
The first version has been received on 12.04.2025, the final version - on 17.05.2025.

DOI: https://doi.org/10.26565/2304-6201-2025-66-02

УДК (UDC) 004.056.53:004.89

Hleha Kateryna *master student; Institute of Special Communications and Information Protection of National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Verkhnyoklyuchova, 4, Kyiv, Ukraine, 03056*
e-mail: katerynaglega54@gmail.com;
https://orcid.org/0009-0004-9337-5836

Hol Vladyslav *professor; head of department; Institute of Special Communications and Information Protection of National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Verkhnyoklyuchova, 4, Kyiv, Ukraine, 03056*
e-mail: vladgol1971@gmail.com;
https://orcid.org/0000-0002-9995-9590

XAI Optimization for Low-Latency Neural-Based Intrusion Detection Systems in Network Environments

Relevance. In contemporary network environments, deep learning-based intrusion detection systems (IDS) provide significant improvements in detecting complex and evolving cyber threats. However, their practical deployment in real-time applications is severely limited by computational complexity, latency, and a lack of interpretability, commonly referred to as the "black-box" problem. Integrating eXplainable Artificial Intelligence (XAI) methods into IDS is crucial for enhancing the transparency, trustworthiness, and operational effectiveness of security systems. **Goal**. The aim of this research is to explore and optimize XAI methods to achieve low-latency, explainable neural-based intrusion detection systems suitable for real-time network traffic analysis, thus balancing interpretability with computational efficiency and detection accuracy. **Research methods.** The study conducted a systematic review and comparative analysis of existing deep learning (DL) models (CNN, LSTM, GRU, Autoencoders, CNN-LSTM hybrids) and prominent XAI techniques (SHAP, LIME, Integrated Gradients, DeepLIFT, Grad-CAM, Anchors). Optimization strategies were proposed, including hardware acceleration, lightweight gradient-based attribution methods, hybrid architectures, and selective explanation strategies. Empirical validation was performed on standard datasets (CICIDS2017, NSL-KDD, UNSW-NB15). **The results.** The analysis revealed that gradient-based attribution methods (DeepLIFT, Integrated Gradients) are optimal for real-time IDS due to minimal latency and high fidelity. Hybrid explainable-by-design frameworks, specifically CNN-LSTM models enhanced with attention mechanisms (ELAI framework), demonstrated significant performance gains with detection accuracy exceeding 98% and inference times below 10 ms. Optimized methods notably improved zero-day attack detection rates up to 91.6%. **Conclusions**. The research successfully demonstrated practical methods for integrating explainability into real-time neural-based IDS, significantly enhancing both detection performance and decision transparency. Future research should focus on standardizing evaluation metrics, refining attention-based models, and extending these optimization approaches to other cybersecurity applications.

Keywords: *cybersecurity, intrusion detection system, deep learning, explainable artificial intelligence, real-time detection, anomaly detection, neural networks, XAI optimization.*

How to quote: Hleha K., Hol V., "XAI Optimization for Low-Latency Neural-Based Intrusion Detection Systems in Network Environments", *Bulletin of V. N. Karazin Kharkiv National University, series Mathematical modelling. Information technology. Automated control systems*, vol. 66, pp. 19-36, 2025. https://doi.org/10.26565/2304-6201-2025-66-02

Як цитувати: Hleha K., Hol V., XAI Optimization for Low-Latency Neural-Based Intrusion Detection Systems in Network Environments. *Вісник Харківського національного університету імені В. Н. Каразіна, серія Математичне моделювання. Інформаційні технології. Автоматизовані системи управління*. 2025. вип. 66. С.19-36. https://doi.org/10.26565/2304-6201-2025-66-02

1. Introduction

In the digital age, cybersecurity – the practice of protecting systems, networks, and confidential data that is coursing through them from unauthorized access, damage, and hijacking – has become an essential part of any organization's security policy. As time progresses, the reliance of individuals, organizations, and governments on digital infrastructure increases. Consequently, it is crucial to ensure the confidentiality, integrity, and availability (CIA triad) of data.

To safeguard sensitive information from a rapidly expanding array of cyberattacks, including the most recent “zero-day” and “slow” attacks, it is imperative that each organization establish a multilayered cybersecurity system, which includes intrusion detection systems (IDS). An intrusion detection system (IDS) scans for malicious activity or unauthorized access through analysis of traffic dynamics, applications and session behavior, and signature-based features. By emphasizing anomalies and recognized attack patterns, IDS alerts help organizations promptly address possible security risks [1].

As the volume of network traffic increases and cyberattacks become more sophisticated, conventional IDS systems, including signature-based and anomaly-based mechanisms, become less effective in detecting intrusions. Currently, approximately 50% of startups disclose experiencing information theft, underscoring the urgency of robust IDS solutions in institutional security frameworks [2].

The integration of machine learning (ML) and deep learning (DL) technologies with IDS systems allows us to significantly enhance the efficiency and classification accuracy of security systems. By utilizing a wide range of ML and DL models with varying characteristics, cybersecurity personnel can create IDS systems that are unique in their ability to perform specific tasks and guarantee the highest level of data protection for each distinct network.

In order to determine whether network traffic is normal or demonstrates indicators of potential malicious activity, machine learning techniques employ algorithms such as decision trees, K-nearest neighbors, and support vector machines (SVMs). These methods significantly enhanced detection rates in comparison to conventional solutions upon implementation; however, ML-based IDS systems were still incapable of managing high-dimensional and imbalanced data. Their reliance on centralized data storage and transmission has resulted in substantial privacy and security vulnerabilities. Two additional challenges to the development of effective ML solutions are the vast quantity of network information and the prevalence of imbalanced data sets. Minor but critical attack types are underrepresented and simply insufficient for proper training. While there are numerous preprocessing techniques, feature selection methods, and ensemble strategies available to improve performance, they are insufficient to guarantee highly accurate detection when it comes to capturing the complex patterns and relationships present in network traffic data. Therefore, researchers initiated the development of more complicated deep learning algorithms, which have demonstrated an exceptional ability to learn hierarchical representations from high-dimensional data [3, 4].

Artificial neural networks have recently garnered significant attention for their potential to improve IDS systems. For instance, convolutional neural networks (CNNs) performed exceptionally well in the identification of spatial features – correlations and relationships within data at a specific time. Conversely, recurrent neural networks (RNNs) had proven to be more adept at detecting temporal features – patterns and sequences across data over time. Several studies have illustrated the successful implementation of deep learning techniques in network intrusion detection, either independently or as part of ML/DL hybrid systems. More recently, researchers have examined the use of advanced deep learning architectures, including long short-term memory (LSTM) and gated recurrent unit (GRU) networks, for intrusion detection in network traffic data [5]. These models demonstrated a high level of ability to capture temporal dependencies in sequential data, which is particularly important when examining network traffic.

Despite their numerous benefits, the DL technologies have significant issues that must be resolved before they can be widely implemented in IDS systems. The deployment of complex deep learning models, such as LSTM, is restricted in real-time or high-throughput environments, such as backbone networks, due to their resource-intensive and slow nature. In addition, they may be highly vulnerable to adversarial inputs, be challenging to scale, or require the implementation of specialized techniques to identify rare but critical attack patterns. DL IDS systems’ biggest shortcomings, however, are their high latency, low throughput, and inability to be explained [6].

Processing large-scale traffic flows, particularly in real-time operations, is restricted by the computational complexity and high resource consumption of DL architectures. DL-based IDS systems are frequently treated as “black boxes” by both developers and users due to their inability to clarify their inference processes and final results. This lack of transparency hinders forensic analysis, complicates auditing and compliance processes, and reduces the overall trust in automated security decisions made by DL models.

eXplainable Artificial Intelligence (XAI) approaches can be integrated into the DL-based IDS frameworks to enhance transparency and interpretability and to facilitate a more comprehensive understanding of model decisions. Transparency helps build trust in AI-driven frameworks by explaining the logic behind some outcomes, which is essential for meeting legal and regulatory requirements [7].

However, despite XAI technologies offering a promising solution to the explainability problem, the most popular methods, such as SHAP and LIME [8], are computationally expensive and not well-suited for real-time deployment. Their reliance on repeated model evaluations or surrogate approximations significantly increases latency, making them impractical for high-throughput environments where fast decision-making is critical.

This research attempts to come up with a proper solution for integrating explainable AI techniques into deep learning-based IDS models in a way that preserves low latency and high throughput while maintaining sufficient interpretability for real-time security decision-making.

2. Objective of the study and research tasks

The primary objective of this study is to explore and evaluate optimization strategies for integrating explainable artificial intelligence (XAI) into deep learning-based intrusion detection systems (IDS) operating in real-time network environments. The goal is to balance detection performance, computational efficiency, and interpretability to enhance trust and operational usability.

To achieve this objective, the following research tasks are defined:

- to review existing explainable AI methods (e.g., LIME, SHAP) and analyze their applicability to IDS;
- to identify the main challenges of implementing XAI in low-latency, high-throughput intrusion detection systems;
- to compare traditional (offline) and real-time XAI approaches in terms of performance, scalability, and explainability;
- to investigate potential optimization strategies for deploying XAI in real-time DL-based IDS;
- to propose conceptual guidelines for integrating interpretable components into deep IDS models without compromising detection speed and accuracy.

3. Review of existing DL models and XAI methods suitable for real-time IDS systems

3.1. Deep learning models for real-time intrusion detection

DL technologies have the potential to significantly improve intrusion detection systems by removing the primary limitations of traditional methods. In contrast to signature-based IDS, which fail to identify emerging threats, deep neural networks automatically identify intricate patterns from raw network data, thereby capturing non-linear feature relationships without the need for manual feature engineering. This allows DL-based IDS systems to detect both known attack signatures and previously unseen or evolving attack patterns with greater accuracy and adaptability. Simply put, deep learning improves on previous methods' high false alarms and blind spots for novel attacks by enabling more precise, flexible, and comprehensive threat detection in IDS [4].

In order to effectively prevent the intrusion, it is crucial to immediately identify any potential anomalies and unusual behavioral changes in the network traffic. As a result, the faster the IDS system operates, the greater is the chance of stopping an attack before it fully corrupts the network. The most effective approach for the majority of contemporary networks is to implement real-time IDS systems that can immediately process traffic and identify changes in the present.

Deploying an IDS in real-time operational networks imposes strict requirements on both the system and the DL models used in it. Key demands include the following:

1. Low detection latency. The IDS system must analyze traffic and detect intrusions with minimal delay (near-instantaneously) to prevent or contain attacks as they occur. Real-time network applications require ultra-low latency processing; even minor delays in traffic analysis can degrade an IDS’s effectiveness [9].

2. High throughput and scalability. It is critical that the real-time IDS system be able to handle large volumes of continuous network data (high bandwidth traffic) without becoming a bottleneck. This implies that the detection DL model must be capable of scaling to high-speed networks and large data streams, processing events in milliseconds, and maintaining a pace with network line rates. As networks expand, the IDS system must ensure that it operates efficiently in heterogeneous or distributed environments.

3. Computational efficiency. To operate in real time, the algorithms must be resource-efficient. Deep models with extremely high complexity (e.g., very deep CNNs or LSTMs) can require a large amount of computation time and may be too heavy for real-time use on limited hardware. Real-time IDS systems often require optimizing or simplifying their models (or utilizing hardware acceleration) to satisfy time

constraints. In particular, in IoT or edge scenarios, DL models must operate within limited CPU/memory, which is why lightweight or optimized models are preferred.

Table 1. Comparison of deep learning models for intrusion detection in real time [11, 12]
Таблиця 1. Порівняння моделей глибокого навчання для виявлення вторгнень у реальному часі [11, 12]

Model	Key Features	Strengths	Limitations	Best Use Cases
Convolutional Neural Networks (CNN)	Employs convolutional layers to extract spatial patterns from fixed-length input vectors	Fast inference; highly parallelizable; excellent at detecting known structured attack patterns; fast training	Not suitable for time-series data or sequences	Packet/flow-level intrusion detection in high-throughput environments
Recurrent Neural Networks (RNN)	Processes sequential data with memory connections	Effective for detecting sequential attack behavior	Suffers from vanishing gradients; less stable	Network traffic behavior analysis
Long Short-Term Memory (LSTM)	Enhanced RNN with long-term memory capability	Handles long-term dependencies; high detection rate for evolving threats	Computationally intensive; slower training	Detection of persistent threats and time-based anomalies
Gated Recurrent Unit (GRU)	Lightweight recurrent architecture that captures temporal dependencies using update and reset gates	Faster, consumes fewer resources than LSTM, yet adapts well to sequence patterns	Slightly less capable of modeling long-term dependencies than LSTM	Detecting time-series anomalies, slow scans, and low-and-slow attacks in real time
Autoencoder	Unsupervised neural network to reconstruct normal behavior; anomalies result in high reconstruction loss	Detects zero-day threats without labeled data; suitable for anomaly-based detection	May misclassify if trained on noisy data; slower unless specifically optimized	Zero-day attack detection and anomaly-based IDS systems
Hybrid Lightweight 1D CNN-LSTM	Combines spatial feature extraction of CNN with temporal pattern detection from LSTM	Balances speed and accuracy; optimized variants can run in real time	Requires careful optimization; heavier than purely CNN or GRU	Attacks exhibiting both spatial and temporal characteristics, such as DDoS or multi-stage intrusions
Deep Neural Networks (DNN)	Simple fully connected feedforward networks	Very fast inference; easy to implement and scale; low latency	Limited feature extraction capability; may miss complex patterns	General classification tasks in high-throughput IDS pipelines

4. High detection accuracy. Even under speed constraints, a real-time IDS system is expected to accurately distinguish attacks from normal traffic. Reliability is crucial – high true positive rates and low false positives ensure the system's rapid alerts are trustworthy. As a result, the DL model should strike a

balance between speed and accuracy, giving operators accurate and timely detection results without overloading them with false alarms.

In order to meet these requirements, researchers implement streaming architectures and optimizations. For example, integrating DL models into frameworks such as Apache Spark Streaming or using a unified Kappa architecture [10] can enable continuous, low-latency processing of network data. In practice, real-time IDS performance may be achieved through the use of model compression, parallel processing, or ensemble methods that enhance accuracy while maintaining a millisecond time budget.

While numerous DL models have been integrated into IDS systems, only a small number are particularly well-suited for real-time detection due to their capacity to balance efficiency and speed. The main DL models and their attributes in an IDS context are summarized in Table 1 [12].

For real-time deployment, DL models must balance speed, throughput, and precision with operational constraints.

CNNs offer rapid, parallelizable inference suitable for high-throughput detection of structured attack patterns, making them highly appropriate for real-time IDS, though lacking in temporal analysis. LSTMs excel at modeling temporal attack sequences, like slow-moving threats, but require considerable computational resources, limiting their real-time practicality unless optimized [11]. GRUs provide a computationally efficient alternative to LSTMs, capturing temporal dependencies effectively with lower latency, thus being more suitable for real-time IDS. CNN-LSTM hybrids achieve an optimal balance of spatial and temporal pattern recognition, delivering high accuracy and real-time deployment feasibility with minimal latency [12]. Autoencoders, capable of unsupervised anomaly detection, are beneficial in real-time IDS for identifying zero-day threats but may generate false alarms and lack detailed attack classification [11]. DNNs, although limited in complex feature detection, offer near-instantaneous inference suitable for initial screening in ultra-high-speed IDS pipelines [12].

The comparative analysis of DL models upon their integration into IDS systems is shown in Table 2 [11]. There are advantages and disadvantages to each deep learning model in terms of complexity, accuracy, and speed. The most appropriate option frequently is determined depending on the deployment constraints and the prioritized attack characteristics that are intended to be countered. CNN-LSTM hybrid DL-based IDS have been demonstrated to enhance detection rates while maintaining real-time operation by employing streaming-friendly architectures and carefully balancing the workload.

Table 2. Comparative analysis of deep learning models for intrusion detection [11]
Таблиця 2. Порівняльний аналіз моделей глибокого навчання для виявлення вторгнень [11]

Model	Accuracy (%)	Precision	Recall	Dataset Used
CNN	96,8	0,95	0,94	CICIDS2017
RNN	95,2	0,93	0,91	NSL-KDD
LSTM	97,1	0,96	0,95	BoT-IoT
Autoencoder	94,5	0,91	0,90	N-BaIoT
Hybrid CNN-LSTM	98,3	0,97	0,96	Custom Mixed Dataset

All the DL models mentioned have the potential to be integrated into real-time IDS systems, contingent upon the identification of the priorities and necessary characteristics of the systems. Nevertheless, some optimization may be necessary before that.

3.2. Explainable AI methods and their potential to enhance trust in DL-based IDS systems

Deep learning models pose one of the greatest challenges in deploying them in actual IDS use due to a lack of interpretability. IDS have to rely on deep learning algorithms that lack transparency despite their high accuracy, creating a “black box” effect that can hinder the analysts’ understanding of their decision-making processes. Simply put, despite the high accuracy of detection, these systems provide little to no insight into why certain decisions were made.

“Black-box” status of DL models means that security professionals struggle to understand the reasoning behind alerts, which is important to trust the system and respond appropriately. Uninterpretable IDS can lead to high false-alarm rates and missed threat patterns, since security teams cannot easily verify or refine the model’s decisions.

Explainable AI (XAI) addresses this challenge by making IDS decisions more auditable. XAI is a fast-growing area of research with the goal to enhance the transparency and trustworthiness of AI systems. For IDS, XAI methods are being applied to yield:

1. Decision-making process visualizations, which can assist security analysts in determining how an IDS model reached a specific decision.

2. Feature importance analysis, which identifies which features (e.g., packet size, traffic volume) were most important for the model's prediction.

3. Interpretability models, e.g., decision trees or rule-based systems, that can provide explanations in a human-readable format [13, 14].

By providing human-understandable explanations for each detection, XAI enables analysts to see which features or behaviors influenced an alert, thereby enhancing trust and clarity in decision-making. For example, an XAI-enhanced IDS might show that an unusually high volume of traffic on a rare port was the key reason a session was classified as an attack. Such insights allow security teams to validate alerts, reduce false positives, and confidently act on the system's recommendations.

A range of XAI methodologies has been applied within IDS frameworks to enhance transparency, each offering distinct advantages and facing unique challenges. Among these, SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) are widely regarded as effective post-hoc, model-agnostic techniques for elucidating complex models like deep neural networks [15].

Table 3. Comparison of XAI methods for real-time IDS suitability [15, 16, 18, 20, 21, 22]
Таблиця 3. Порівняння методів XAI щодо придатності для IDS реального часу [15, 16, 18, 20, 21, 22]

XAI Method	Explanation Type	Model Agnostic/ Specific	Key Advantages	Limitations	Suitability for Real-Time IDS
Shapley Additive Explanations (SHAP)	Feature Attribution	Model-agnostic	Consistent and fair attribution; local and global explanations	High computational overhead	Limited due to high latency
Local Interpretable Model-Agnostic Explanations (LIME)	Surrogate Model	Model-agnostic	Intuitive, per-instance explanations	Local explanations may not generalize well	Limited due to latency concerns
Saliency Maps / Grad-CAM	Gradient-based Visualization	Model-specific	Fast, intuitive visual explanations	Require differentiable models; noisy outputs; no textual explanations	Suitable due to low latency
Integrated Gradients (IG)	Gradient-based Attribution	Model-specific	Robust feature attribution, faithful explanations	Requires model internals; less intuitive alone	Highly suitable due to low latency
DeepLIFT	Gradient-based Attribution	Model-specific	High-fidelity, efficient, low complexity explanations	Requires model internals; less intuitive alone	Highly suitable due to low latency
Anchors	Rule-based Explanations	Model-agnostic	Intuitive, high-precision rules	Computationally intensive to derive optimal rules	Limited unless simplified rules are used

SHAP provides consistent and thorough feature attribution but its computational complexity limits its suitability for real-time IDS scenarios [15, 16]. LIME offers intuitive per-instance explanations beneficial for auditing individual alerts but is similarly constrained by computational latency [15, 18]. Saliency Maps and Grad-CAM are fast visual methods suited for real-time use due to their low computational overhead [20]. Integrated Gradients (IG) and DeepLIFT efficiently deliver faithful, low-latency explanations highly suitable for real-time IDS implementations. They are precise and are applicable to any differentiable DL-based IDS, but they require access to the neural model's internals [21]. Anchors produce intuitive, high-precision rules but their computational cost in deriving optimal rules limits real-time applicability unless simplified rules are used [22].

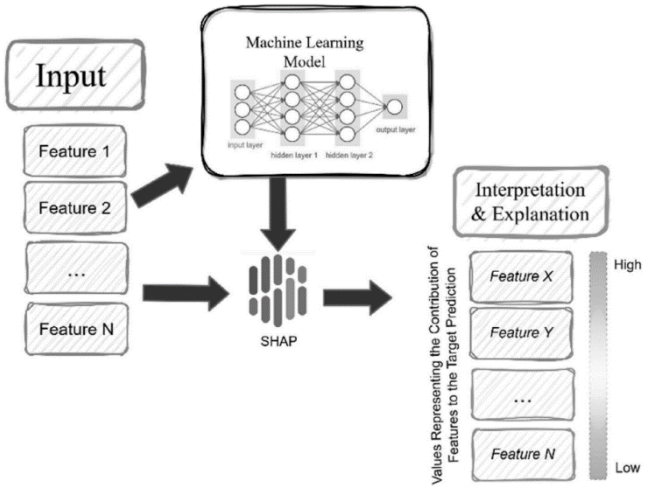


Fig. 1. Schematic representation of the SHAP value framework in a machine learning context [17]
Рис. 1. Схематичне представлення структури фреймворку SHAP у контексті машинного навчання [17]

When integrating XAI methods into a real-time IDS, several key criteria determine their suitability:

1. Fidelity (Faithfulness). This refers to how accurately the explanation reflects the actual decision process of the original model. A high-fidelity explanation will highlight the truly important features in the model's internal reasoning. For example, if the IDS's DL model bases its decision mainly on feature X, a faithful XAI method should assign the highest importance to X in its explanation. Low-fidelity explanations can mislead analysts by emphasizing the wrong factors, so measuring an XAI technique's fidelity to the DL model is crucial. Some studies quantify fidelity by checking how model predictions change when the top-ranked features from the explanation are removed or perturbed. In an IDS context, faithfulness ensures the explanations are truthful proxies of the complex model – a necessary condition for trust.

2. Human Interpretability. Even if an explanation is faithful, it must be understandable to a human analyst. Interpretability involves the simplicity and clarity of the explanation – e.g., using a small number of features, natural language descriptions, or visual aids that a person can quickly grasp. A highly interpretable explanation might be a short rule or a concise list of the top 2 or 3 features influencing a decision, rather than a dense list of 20 parameters. For IDS analysts under time pressure, explanations should ideally be simple, clear, and domain-relevant. Techniques like Anchors and decision trees score well on human interpretability, since they produce "if-then" rules and clear decision paths, respectively, whereas something like a raw saliency map or a long list of Shapley values may need more interpretation [22]. In practice, usability studies have found that providing transparent and visual explanations, such as charts of feature contributions or highlighted traffic traces, improves analysts' trust and speeds up their validation of alerts.

3. Computational Overhead. The extra processing time and resources required to generate explanations are a major concern for real-time systems. Some XAI methods, especially post-hoc techniques, can be computationally intensive. For instance, SHAP often requires evaluating the model multiple times on various feature subsets or background samples, and LIME needs to generate many perturbed samples to fit a local surrogate. These methods can significantly slow down the alert pipeline if used on every single event. Overhead is typically measured in terms of added latency per prediction or

CPU/GPU usage. In a high-throughput IDS, an XAI method with heavy overhead might not be practical. Therefore, a suitable XAI for real-time IDS should minimize computation – using efficient algorithms, sampling strategies, or by leveraging hardware acceleration. Research in explainable IDS frequently emphasizes the need for lightweight explainability approaches that don't degrade the system's performance.

4. Real-Time Feasibility. This criterion is related to overhead but focuses on whether the XAI method can deliver explanations within the time constraints of an operational environment. A real-time IDS may need to flag and explain an alert within milliseconds to seconds. Thus, methods that require lengthy processing or cannot keep up with streaming data are less feasible. Real-time feasibility also considers if the explanation can be generated on-the-fly for each alert or if it requires batch or offline processing. For example, a method that pre-trains an interpretable surrogate model might be feasible if that surrogate can then produce instant explanations during operation [15]. In contrast, an approach that must solve an optimization or search problem per event – as some anchor-based or perturbation methods do – might struggle in real time. Ultimately, achieving real-time explainability often involves a trade-off between the depth of the explanation and the speed of generation. Ensuring feasibility might involve simplifying the explanation, using approximate but faster algorithms, or only explaining a subset of events rather than everything.

Not all explanation techniques are equally practical for a real-time DL-based IDS. Techniques that offer low latency and high clarity are generally more suitable.

1. Gradient-Based Attributions methods – Saliency, Integrated Gradients, and DeepLIFT – are typically favored for real-time use because of their computational efficiency. They piggyback on the model's own backpropagation, typically requiring only one pass through the neural network to calculate feature importances. For instance, computing an integrated gradient or a DeepLIFT attribution for an input can be done quickly on modern hardware [21].

In comparative evaluations on an LSTM-based IDS, gradient methods (especially DeepLIFT) produced explanations with lower complexity and higher fidelity than LIME or SHAP, indicating they capture the model's behavior well without excessive computation. DeepLIFT in particular was found to give consistent and reliable explanations while being faster to compute, making it a strong candidate for real-time alert explanation.

Gradient-based attributions methods work seamlessly with common DL models such as CNNs, RNNs and autoencoders, highlighting important features or time steps almost instantaneously. The trade-off is that gradient-based explanations might be less intuitive in isolation, but they can be combined with visualization or simple messaging to aid analysts. Overall, because of their high fidelity and low overhead, saliency and gradient techniques are well-suited to explain decisions on the fly in real-time IDS systems.

2. Surrogate and Rule-Based Methods (Simplified Models) present another strategy for real-time explainability, namely to use an interpretable model alongside or in place of the DL for certain decisions. For example, a decision tree or a set of "if-then" rules can approximate the deep model's behavior for explanation purposes. These surrogates can be pre-computed (offline) to mimic the DL model on training data then used to generate quick explanations during operation. Because a decision tree or rule set is fast to evaluate, the explanation is essentially instantaneous at run-time. Such hybrid approaches attempt to get the best of both worlds: the DL handles detection accuracy, while the surrogate offers human-readable logic as explanations [15].

However, maintaining fidelity is a challenge – a too-simple surrogate might not capture complex patterns the DL model uses. In real-time settings, one compromise is to deploy the interpretable model for the majority of routine traffic and reserve the complex DL model with post-hoc explanations like LIME/SHAP for more ambiguous or high-risk cases. This tiered approach can preserve performance and provide transparency when most needed, at the cost of system complexity.

When using rule-based explainers like Anchors in real-time, scope is important: anchors can be computed quickly if the feature space is small or if we only seek a rule for the most influential features. They can succinctly explain an alert (e.g., "alert triggered because X and Y conditions were met") without overwhelming detail, which is ideal for an analyst's quick decision cycle. The user must be cautious that anchors – or any rule – remain accurate under evolving traffic conditions.

3. While LIME and SHAP are powerful and widely used, their direct application to every packet or alert in a high-throughput IDS can be impractical due to their computational cost. SHAP, in particular, though providing very insightful explanations, might take too long on complex models or large feature

sets – potentially seconds per instance – which is untenable for systems that analyze hundreds of events per second.

That said, there are scenarios where these methods can still contribute: for instance, TreeSHAP can efficiently explain ensemble tree models – if an IDS uses a tree-based classifier – in real-time by leveraging a closed-form solution. LIME can be sped up by reducing the number of perturbations or using optimized surrogates, but it may still struggle as data dimensionality grows [15].

Therefore, in a real-time IDS, SHAP/LIME are often used selectively – for example, to explain a handful of critical alerts or to perform periodic analysis on model behavior – rather than on every event. They are extremely valuable in offline model evaluation or forensic analysis of incidents, helping to understand global patterns (SHAP) and specific cases (LIME) with high interpretive richness.

The integration of XAI into IDS is critical for ensuring that cybersecurity systems are not only highly accurate but also capable of providing explanations that human analysts can readily comprehend and act upon. In summary, for the day-to-day, fast-paced detection, lighter methods such as gradients and simple rules are preferable, whereas LIME or SHAP might support near-real-time workflows where a brief delay is acceptable or as backup explainers for complex cases.

Integrating XAI into real-time DL-based IDS systems requires balancing explanation quality with performance. Methods like integrated gradients, DeepLIFT, and saliency maps offer quick, faithful insights into neural models' decisions and are thus most promising for real-time IDS use. Rule-based explanations and simplified surrogates provide human-friendly logic with negligible latency, which can greatly aid analyst understanding when carefully aligned with the DL model. More computationally intensive techniques like SHAP and LIME are effective in enhancing transparency and reducing false positives, but they may need optimization or selective deployment to fit into high-speed environments.

3.3. Challenges of integrating XAI into real-time DL-based IDS systems

Real-time IDS systems face several obstacles when incorporating XAI methods. Key challenges include computational overhead, scalability issues, accuracy-interpretability trade-offs, lack of standard evaluation practices, and security implications of exposing model logic.

1. Latency and computational overhead is the first problem. As it had already been mentioned, many popular XAI techniques, such as SHAP and LIME, are computationally intensive, often requiring numerous model re-runs or complex calculations. In a real-time IDS, generating an explanation for each alert can introduce significant latency and CPU/GPU load. Studies confirm that post-hoc explainers like LIME/SHAP add extra processing, which can slow down threat detection and response rates. In other words, the IDS may become sluggish in high-speed networks because of the time spent computing explanations. One survey notes that XAI-enhanced IDS often face "increased computational complexity and potentially reduced performance due to the overhead of generating explanations [23]." Such latency overhead is problematic in operational environments that demand swift decision-making to block attacks.

2. Scalability and deployment constraints are the second major issue. The heavy resource requirements of both DL models and XAI methods pose scalability issues. Many advanced DL-based IDS models, for example, transformers or deep CNNs, need powerful hardware acceleration, which is unsuitable for edge deployments with limited resources [23]. Pushing complex models or their explainers to low-power network devices can be infeasible due to memory, CPU, or energy constraints. Additionally, high-throughput network traffic magnifies the problem – explaining every flagged event in a busy network can overwhelm the system. Even cloud-based IDS setups struggle, as constant communication for explanations adds network latency. Researchers highlight that real-time IDS performance suffers in high-traffic environments when burdened with current XAI computations. In summary, without careful optimization, XAI may not scale well to the volume and speed of data in modern networks.

3. The next concern of real-time DL-based IDS is to find balance between accuracy and interpretability. There is an inherent trade-off between model complexity, which often yields higher accuracy, and its interpretability. State-of-the-art IDS models like DNN or ensemble methods achieve superior detection rates but operate as "black boxes" with opaque logic. By contrast, simpler models like decision trees or rule-based classifiers are transparent but tend to miss subtle or sophisticated attacks. This gap is well documented – high-performing DL models regularly forgo interpretability for greater predictive power, whereas overly simple models can undermine detection accuracy. In practice, forcing a complex model to be more explainable, for example, by approximating it with an interpretable surrogate, may degrade its performance on edge-case intrusions. Studies have noted that decision-tree-based IDS, while easy to explain, "frequently miss subtle danger behaviors, which lowers the accuracy

of detection [24].” Balancing these concerns is difficult: analysts need to trust and understand the IDS decisions, but not at the cost of allowing attacks to slip through due to an oversimplified model.

4. Another significant challenge is the lack of standardized XAI evaluation. There is no consensus on how to evaluate and compare XAI methods in the IDS domain. Unlike accuracy or false-alarm rate, which have clear metrics, “explainability” lacks a unified quantitative framework in cybersecurity contexts. Researchers point out that without standard interpretability metrics, it is difficult to judge whether one explanation method truly outperforms another or adequately meets analysts’ needs. This gap means each study often uses its criteria (e.g., subjective user feedback or ad-hoc measures of explanation quality), making it difficult to benchmark XAI techniques across different IDS implementations. The literature emphasizes that consistent evaluation standards – such as agreed-upon interpretability scores or time-to-insight measurements – are needed to fairly assess XAI in IDS [15]. Until such frameworks mature, deploying XAI will involve uncertainty about how much it actually improves analyst understanding or trust in a real-time setting.

5. The security and privacy implications should be addressed too. Integrating XAI into IDS can inadvertently introduce security risks. Detailed explanations reveal which features or patterns led the model to flag an attack; if such information is accessible to adversaries, they might exploit it to evade detection. In essence, an explanation interface could become a leakage point – giving attackers insight into the IDS’s “secrets”. For example, if an explanation consistently highlights a specific packet header field as suspicious, a savvy attacker may alter that field in future exploits to fly under the radar. Moreover, there are privacy concerns when explanations expose sensitive attributes of network traffic or user data. Some XAI outputs might inadvertently disclose personal or proprietary information, contravening data protection principles.

This is especially relevant under regulations like the General Data Protection Regulation (GDPR), which require careful handling of any user-related data. Therefore, designers must ensure that adding explainability does not open new attack vectors or privacy leaks. Research in this area suggests employing privacy-preserving XAI techniques and restricting how much internal detail is shared so that trust is improved for defenders without equipping attackers with a roadmap to bypass the IDS.

Given the above challenges, experts acknowledge the need for more efficient and tailored XAI approaches in real-time IDS. One promising direction is the use of hybrid models or tiered strategies. For instance, a simpler interpretable model could handle the bulk of low-risk traffic, with a complex DL-XAI module reserved for only the most suspicious events – thereby limiting the overhead to where it’s truly needed [15].

Another approach is to design or choose algorithms that are interpretable by design, reducing reliance on expensive post-hoc explainers. Techniques like attention mechanisms in neural networks can highlight important features as part of the prediction process, effectively providing an explanation with minimal extra cost. In fact, recent IDS frameworks, such as attention-based CNN-LSTM architecture, demonstrate that it’s possible to achieve high speed and integrate feature attribution (heatmaps) directly into the model’s operation.

Researchers also suggest optimizing existing XAI methods – for example, using faster approximation algorithms for SHAP/LIME or pre-computing explanation components – to fit the real-time requirements.

Overall, there is a clear consensus that new lightweight XAI solutions are required to balance transparency with performance. Many researches stress developing explainability techniques that incur minimal delay and can scale so that future IDS can be both highly accurate and explainable without sacrificing low latency [15].

4. XAI optimization strategies for low-latency IDS systems

Realizing low-latency, explainable intrusion detection requires innovative approaches that minimize the overhead of explanations while preserving or even enhancing detection performance. Researchers have focused on two complementary directions: accelerating existing XAI techniques to fit real-time needs and developing hybrid or explainable-by-design models that inherently provide insights with minimal extra cost. In parallel, practical deployment strategies – from hardware acceleration to selective explanation – ensure these techniques scale to high-speed network environments. To formalize this trade-off, an optimization objective (4.1) that balances latency, explainability cost, and detection accuracy had been defined:

$$F(\theta,\omega)=\alpha\cdot Lat(\theta)+\beta\cdot CompXAI(\omega)-\gamma\cdot Acc(\theta,\omega)\to\min \tag{4.1}$$

where θ are parameters of the deep learning model (e.g., number of layers, neurons, architecture of CNN-LSTM); ω are parameters of the explainable AI method (e.g., attribution depth in DeepLIFT); $Lat(\theta)$ is the latency of the IDS decision in milliseconds; $CompXAI(\omega)$ is the computational cost of the XAI method (e.g., DeepLIFT), measured in processing time or compute resources (CPU/GPU); $Acc(\theta,\omega)$ is the classification accuracy of the IDS (e.g., detection rate of anomalies); α,β,γ are weighting coefficients reflecting the priority of each optimization objective (set based on system-specific constraints or expert judgment).

Recent peer-reviewed studies cited in this work validate these optimizations on standard cybersecurity datasets, demonstrating that it is feasible to achieve both millisecond-level detection times and meaningful explanations in IDS. Both the methodological innovations and implementation considerations for XAI in real-time neural-based IDS are illustrated below.

4.1. Accelerating XAI techniques for real-time efficiency

A primary challenge is the computational cost of popular post-hoc explainers like SHAP and LIME, which can be too slow for streaming data. To address this, researchers are optimizing these algorithms and leveraging hardware acceleration. For instance, using GPU-accelerated libraries – NVIDIA's RAPIDS or PyTorch CUDA extensions – can speed up SHAP computations significantly, enabling feature attribution on large traffic samples in near real-time. Algorithmic improvements such as sampling-based SHAP or lightweight surrogate models have also been explored to approximate explanations faster. A recent survey stresses that making SHAP/LIME faster or more lightweight is crucial for practical deployment in high-speed IDS [15]. By reducing the number of model evaluations or focusing on top features, these optimized explainers shrink the latency they introduce.

Another effective tactic is to favor inherently efficient XAI methods. Gradient-based attribution techniques, such as saliency maps, Integrated Gradients, and DeepLIFT, require only a single backward pass through the neural network, offering explanations with minimal overhead. An evaluation of explanations for an LSTM-based IDS found that DeepLIFT consistently outperformed LIME and SHAP in producing high-fidelity, low-complexity explanations [21]. Because these methods directly leverage the model's internal gradients, they generate attributions in milliseconds, making them well-suited for real-time alert explanation. In practice, integrated gradient or saliency results can be visualized as heatmaps almost instantly, highlighting which features – specific packet bytes or timing features – influenced the decision. By adopting such low-cost XAI methods, an IDS can deliver basic reasoning for each alert on the fly without becoming a bottleneck.

4.2. Hybrid and explainable-by-design model approaches

Beyond speeding up post-hoc tools, a promising avenue is to embed interpretability into the IDS models themselves. Researchers are creating hybrid architectures that combine the accuracy of deep learning with the transparency of simpler models or built-in explanation mechanisms. One strategy is to attach an interpretable component – a rule-based or tree-based layer – to a neural network. For example, a decision tree or rule set can act as a front-end filter or a parallel explainer to the deep model, providing human-readable logic for its predictions. This two-tier design lets the system enjoy the nuance of a neural detector while yielding an immediate explanation – the triggered rule or path in the tree – for most decisions. Recent studies emphasizes such hybrid models as a way to balance accuracy and transparency: for instance, by combining a shallow decision tree with a back-end deep classifier, the IDS can handle complex patterns but still explain detections in simple terms [15]. In this work, existing experimental results are referenced to illustrate that such prototypes enable many alerts to be accurately handled by the interpretable component alone, with the deep model invoked only for uncertain cases—substantially reducing the average explanation cost.

Another approach is to design explainable-by-design neural networks specialized for IDS tasks. One cutting-edge example is the Explainable Lightweight AI (ELAI) framework, which uses a streamlined CNN-LSTM architecture augmented with an attention mechanism. The attention layers highlight important features in each input, such as specific flow characteristics or time steps, effectively producing an explanation as a by-product of the prediction. Because this occurs during the model's forward pass, there is negligible latency overhead. According to prior evaluations, the ELAI framework demonstrated that such integration can significantly improve both speed and transparency: it achieved an inference time

of ~8.3 ms per sample – over 60% faster than a standard deep IDS – while providing visual "attention heatmaps" to analysts. Importantly, the model's output is not a black box; it leverages SHAP-based feature importance and attention weights to make each decision interpretable and more trustworthy for security operators [25]. This indicates that carefully architected networks, like lightweight CNN-LSTM with built-in attention, can meet real-time demands without sacrificing interpretability.

Researchers are also exploring model compression and knowledge distillation as avenues for XAI optimization. The idea is to train a compact "student" model to mimic a larger deep model's behavior, thereby retaining high accuracy on attacks but with far fewer parameters and simpler decision logic. Compressed models naturally run faster and can be easier to interpret or to explain post-hoc due to their reduced complexity. A recent study using knowledge distillation for an IoT IDS showed the student network ran approximately 15–25% faster in inference than its complex teacher, with negligible accuracy loss [25]. The distilled model could even retain transparency by highlighting key features in its decisions, courtesy of an integrated attribution mechanism.

Similarly, hybrid frameworks like Lightweight, Efficient, and Non-intrusive System for eXplainable Artificial Intelligence (LENS-XAI) combine a variational autoencoder for unsupervised anomaly detection with a distilled lightweight classifier, explicitly aiming to balance performance and transparency for scalable intrusion detection. By validating these frameworks on multiple datasets such as NSL-KDD, Edge-IIoT, and UNSW-NB15, it was shown that state-of-the-art detection rates can be achieved alongside built-in explainability and efficiency [26].

In summary, new architectural innovations – from attention-based deep models to distilled ensembles – are enabling IDS that are both fast and explainable by design. These hybrid approaches reduce reliance on expensive after-the-fact explanations, since much of the reasoning is either inherent in the model's structure or handled by a lightweight interpretable component.

4.3. Deployment considerations and empirical validation

Implementing explainable IDS in real networks requires not just clever algorithms but also strategic system design to handle high data volumes. One key is to integrate the above methods into streaming data pipelines and optimize the end-to-end flow. Researchers have suggested deploying real-time IDS within frameworks like Apache Spark Streaming or a Kappa architecture, which can distribute the workload of traffic capture, detection, and explanation across multiple nodes for scalability [15]. In practice, this means explanations should be generated in parallel with detection or during off-peak cycles. For example, an IDS could immediately flag a likely attack using a fast, simplified model, then invoke a more detailed XAI analysis on a separate thread or machine learning accelerator. By asynchronously handling explanations, the system ensures that alert latency remains low.

Moreover, hardware acceleration – Graphics Processing Units (GPUs) or Tensor Processing Units (TPUs) – can be dedicated to XAI computations so that even if a complex method like SHAP is needed for a particularly critical alert, it can be computed in a fraction of the time it would take on a CPU. These engineering strategies ensure that adding explainability doesn't turn into a throughput bottleneck.

Another consideration is selective or adaptive explanation to conserve resources. Not every benign flow or low-risk event may require a full explanation; the system can be tuned to provide detailed interpretability for the most suspicious or impactful alerts. Recent proposals even suggest adaptive XAI levels – giving a high-level reason for routine detections but a thorough, multi-faceted explanation for complex or severe incidents [15]. This adaptive approach aligns with operational needs, focusing analyst attention where it's needed most and trimming unnecessary computation. Crucially, any introduction of XAI must be evaluated not only for speed but also for analytical value: security teams should gain insight without being overwhelmed. Visualization tools, for example, feature importance bar charts or traffic heatmaps, should be integrated into the IDS dashboard to present the explanations clearly and quickly. Empirical results from recent research underscore the feasibility of these optimizations. In prior studies, the ELAI framework, for instance, was evaluated on standard benchmarks sych as CICIDS2017 and UNSW-NB15, achieving over 98% detection accuracy with a compact model size under 50 MB [25]. Due to its architectural optimizations, ELAI was shown to process each network sample in just a few milliseconds – approximately 2.5 times faster than a typical deep IDS – while still producing human-interpretable feature attributions for every alert.

Likewise, the LENS-XAI student model was validated across diverse datasets – from classic NSL-KDD to modern IoT traffic – and maintained high fidelity to the teacher model's predictions, but with significantly lower latency and complexity [26].

These case studies confirm that the trade-off between speed and explainability can be managed effectively. In fact, making the model more efficient often goes hand-in-hand with better clarity: focusing on fewer, most informative features tends to improve both runtime and the quality of explanations.

Finally, it is important to assess the optimized XAI IDS in real-world conditions. Beyond lab datasets, deployment in live network environments such as enterprise LANs or IoT networks is needed to ensure the system handles traffic bursts, novel attack patterns, and concept drift over time. The explainability component should be stress-tested for worst-case scenarios – for example, verifying that an explanation can still be produced within a strict time budget during a distributed attack or that the XAI does not expose sensitive information inadvertently.

Early adaptive IDS prototypes show promise in detecting zero-day attacks while keeping analysts informed: in one evaluation, an explainable IDS detected over 91% of zero-day attacks in an IoT setting, significantly outperforming a non-XAI baseline, thanks to its robust feature insights guiding the detection [25]. The results of said evaluation are shown in Table 3.

Table 4. Comparative analysis of ELAI with existing IDS models
Таблиця 4. Порівняльний аналіз ELAI з існуючими моделями IDS

Model	Computational Efficiency	Explainability	Zero-Day Attack Detection (%)
CNN-LSTM (Baseline)	Moderate	Low	74.3
ResNet-50 IDS	Low	Very Low	79.8
Transformer-Based IDS	Very Low	Very Low	82.5
ELAI	High	High	91.6

This highlights that XAI optimization is not just an academic exercise but a practical enhancement to security: a well-designed explainable model can catch stealthy threats more reliably by focusing on telltale anomalies and immediately justify the alerts, enabling quicker and more confident responses.

In summary, the core of recent research on “XAI optimization for low-latency neural IDS” converges on a clear message: it is possible to build IDS solutions that are both fast and transparent. By streamlining XAI algorithms, fusing interpretable logic into deep models, and thoughtfully engineering the deployment, security teams can obtain real-time intrusion alerts with the much-needed context.

Ongoing studies continue to refine these approaches – from standardized interpretability metrics to domain-specific explanation techniques – but the trajectory is set. The future of intrusion detection will likely see lightweight, explainable AI at its heart, providing strong defense capabilities that are no longer a “black box” to the people they protect.

5. Conclusions

In this work, a comprehensive investigation was conducted on optimizing eXplainable Artificial Intelligence (XAI) methods for DL-based intrusion detection systems (IDS) operating in real-time network environments. The primary scientific novelty of the study lies in the in-depth analysis of various XAI approaches, leading to practical recommendations and the conceptual integration of multiple explainability strategies into a unified, low-latency DL-based IDS framework suitable for high-speed network infrastructures.

The key scientific results of this study include:

1. Systematic analysis and critical evaluation of existing XAI methods (SHAP, LIME, Integrated Gradients, DeepLIFT, Anchors, Grad-CAM), highlighting their practical applicability limits in real-time network environments, particularly their significant computational overhead.

2. Justification of gradient-based attribution methods (Integrated Gradients, DeepLIFT) as highly promising for real-time applications due to their ability to produce high-quality explanations with minimal latency overhead.

3. Proposal of hybrid explainable-by-design architectures, including CNN-LSTM with attention mechanisms (e.g., ELAI) and LENS-XAI, which effectively combine high detection accuracy with built-in interpretability without imposing substantial computational costs.

4. Development of practical deployment guidelines and strategies for explainable IDS, including the use of hardware acceleration (GPU/TPU), adaptive explanation generation strategies, and optimized streaming architectures (Kappa architecture, Apache Spark Streaming).

5.Empirical results from existing studies demonstrate that optimized XAI models – particularly the ELAI and LENS-XAI architectures – achieve significant improvements in zero-day attack detection rates (up to 91.6%) and substantially lower explanation generation times (below 10 ms), thereby confirming their practical viability for integration into real-time IDS in high-speed network environments.

The obtained results hold significant implications for both cybersecurity theory and practice. Theoretical significance involves advancing the understanding of the balance between explainability and performance in neural IDS models deployed under real-time conditions. This insight provides a solid foundation for future research on integrating XAI with deep IDS architectures. Practical significance is demonstrated through the applicability of the proposed methods to real-world information security systems, including large enterprise networks, IoT infrastructure, and national-level network systems. These methods enhance decision transparency, operator trust, and incident response speeds.

Prospective future research directions include:

1. Developing standardized metrics and benchmarks for evaluating XAI explainability, enabling objective comparison of various XAI techniques and approaches.

2. Further refinement of IDS architectures through integrating advanced attention mechanisms (e.g., transformer-based attention), thereby improving explanation quality and granularity.

3. Investigating the impact of explainability on cybersecurity operators' performance (human-in-the-loop scenarios), including developing intuitive interfaces for presenting explanations in real time.

4. Conducting long-term field studies of explainable IDS deployments in operational networks, enabling the identification of practical constraints and optimization requirements.

5. Exploring adaptation possibilities of the presented approaches and architectures to other critical cybersecurity tasks, such as traffic obfuscation detection, covert channel identification, and recognition of complex multi-vector attacks.

In conclusion, the research provides a robust foundation for the theoretical advancement and practical implementation of explainable AI in intrusion detection systems. It paves the way for developing transparent, reliable, and high-performance next-generation IDS solutions.

REFERENCES

1. Otoum Y., Nayak A. AS-IDS: Anomaly and Signature Based IDS for the Internet of Things. Journal of Network and Systems Management. 2021. Vol. 29, no. 3. URL: https://doi.org/10.1007/s10922-021-09589-6 [in English] (date of access: 14.06.2025).
2. Securing financial data storage: A review of cybersecurity challenges and solutions / Chinwe Chinazo Okoye et al. International Journal of Science and Research Archive. 2024. Vol. 11, no. 1. P. 1968–1983. URL: https://doi.org/10.30574/ijsra.2024.11.1.0267 [in English] (date of access: 15.06.2025).
3. Federated Learning for intrusion detection system: Concepts, challenges and future directions / S. Agrawal et al. Computer Communications. 2022. URL: https://doi.org/10.1016/j.comcom.2022.09.012 [in English] (date of access: 16.06.2025).
4. Deep Learning vs. Machine Learning for Intrusion Detection in Computer Networks: A Comparative Study / M. L. Ali et al. Applied Sciences. 2025. Vol. 15, no. 4. P. 1903. URL: https://doi.org/10.3390/app15041903 [in English] (date of access: 16.06.2025).
5. Deep Learning Approach for Intelligent Intrusion Detection System / R. Vinayakumar et al. IEEE Access. 2019. Vol. 7. P. 41525–41550. URL: https://doi.org/10.1109/access.2019.2895334 [in English] (date of access: 19.06.2025).
6. Gaspar D., Silva P., Silva C. Explainable AI for Intrusion Detection Systems: LIME and SHAP Applicability on Multi-Layer Perceptron. IEEE Access. 2024. P. 1. URL: https://doi.org/10.1109/access.2024.3368377 [in English] (date of access: 19.06.2025).
7. Federated XAI IDS: An Explainable and Safeguarding Privacy Approach to Detect Intrusion Combining Federated Learning and SHAP / K. Fatema et al. Future Internet. 2025. Vol. 17, no. 6. P. 234. URL: https://doi.org/10.3390/fi17060234 [in English] (date of access: 21.06.2025).

8. Arreche O., Guntur T., Abdallah M. XAI-IDS: Toward Proposing an Explainable Artificial Intelligence Framework for Enhancing Network Intrusion Detection Systems. Applied Sciences. 2024. Vol. 14, no. 10. P. 4170. URL: https://doi.org/10.3390/app14104170 [in English] (date of access: 21.06.2025).
9. Enhancing intrusion detection: a hybrid machine and deep learning approach / M. Sajid et al. Journal of Cloud Computing. 2024. Vol. 13, no. 1. URL: https://doi.org/10.1186/s13677-024-00685-x [in English] (date of access: 24.06.2025).
10. Stacking Ensemble Deep Learning for Real-Time Intrusion Detection in IoMT Environments / E. Alalwany et al. Sensors. 2025. Vol. 25, no. 3. P. 624. URL: https://doi.org/10.3390/s25030624 [in English] (date of access: 25.06.2025).
11. Laxmi, Chauhan K. AI-Based Intrusion Detection Systems for Novel Attacks in IoT and APTs: A Deep Learning-Centric Review. International Journal of Computer Science and Information Security (IJCSIS), Vol. 23, No. 3, May-June. URL: https://www.academia.edu/130243382/AI_Based_Intrusion_Detection_Systems_for_Novel_Attacks_in_IoT_and_APTs_A_Deep_Learning_Centric_Review?bulkDownload=true [in English] (date of access: 25.06.2025).
12. A high performance hybrid LSTM CNN secure architecture for IoT environments using deep learning / P. Sinha et al. Scientific Reports. 2025. Vol. 15, no. 1. URL: https://doi.org/10.1038/s41598-025-94500-5 [in English] (date of access: 27.06.2025).
13. Ribeiro M. T., Singh S., Guestrin C. "Why Should I Trust You?". KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco California USA. New York, NY, USA, 2016. URL: https://doi.org/10.1145/2939672.2939778 [in English] (date of access: 27.06.2025).
14. Lundberg S. M., Lee S.-I., ”A unified approach to interpreting model predictions,” in Advances in Neural Information Processing Systems (NIPS), vol. 30, 2017. URL: https://arxiv.org/abs/1705.07874v2 [in English] (date of access: 27.06.2025).
15. Mohale V. Z., Obagbuwa I. C. A systematic review on the integration of explainable artificial intelligence in intrusion detection systems to enhancing transparency and interpretability in cybersecurity. Frontiers in Artificial Intelligence. 2025. Vol. 8. URL: https://doi.org/10.3389/frai.2025.1526221 [in English] (date of access: 04.07.2025).
16. Explainable Intrusion Detection Systems (X-IDS): A Survey of Current Methods, Challenges, and Opportunities / S. Neupane et al. IEEE Access. 2022. Vol. 10. P. 112392–112415. URL: https://doi.org/10.1109/access.2022.3216617 [in English] (date of access: 04.07.2025).
17. Alomari Y., Andó M. SHAP-based insights for aerospace PHM: Temporal feature importance, dependencies, robustness, and interaction analysis. Results in Engineering. 2024. Vol. 21. P. 101834. URL: https://doi.org/10.1016/j.rineng.2024.101834 [in English] (date of access: 06.07.2025).
18. Explainable Artificial Intelligence for Intrusion Detection System / S. Patil et al. Electronics. 2022. Vol. 11, no. 19. P. 3079. URL: https://doi.org/10.3390/electronics11193079 [in English] (date of access: 06.07.2025).
19. Visani G. LIME: explain Machine Learning predictions. Medium. URL: https://medium.com/data-science/lime-explain-machine-learning-predictions-af8f18189bfe [in English] (date of access: 07.07.2025).
20. Leveraging Grad-CAM to Improve the Accuracy of Network Intrusion Detection Systems / F. P. Caforio et al. Discovery Science. Cham, 2021. P. 385–400. URL: https://doi.org/10.1007/978-3-030-88942-5_30 [in English] (date of access: 08.07.2025).
21. Evaluating Explainable AI for Deep Learning-Based Network Intrusion Detection System Alert Classification / R. Kalakoti et al. 11th International Conference on Information Systems Security and Privacy, Porto, Portugal, 20–22 February 2025. 2025. P. 47–58. URL: https://doi.org/10.5220/0013180700003899 [in English] (date of access: 09.07.2025).
22. Ribeiro M. T., Singh S., Guestrin C. Anchors: High-Precision Model-Agnostic Explanations. Proceedings of the AAAI Conference on Artificial Intelligence. 2018. Vol. 32, no. 1. URL: https://doi.org/10.1609/aaai.v32i1.11491 [in English] (date of access: 11.07.2025).

23. Explainable AI for Comparative Analysis of Intrusion Detection Models / P. M. Corea et al. 2024 IEEE International Mediterranean Conference on Communications and Networking (MeditCom), Madrid, Spain, 8–11 July 2024. 2024. P. 585–590. URL: https://doi.org/10.1109/meditcom61057.2024.10621339 [in English] (date of access: 13.07.2025).
24. Bhagyashree D Shendkar. Explainable Machine Learning Models for Real-Time Threat Detection in Cybersecurity. Panamerican Mathematical Journal. 2024. Vol. 35, no. 1s. P. 264–275. URL: https://doi.org/10.52783/pmj.v35.i1s.2313 [in English] (date of access: 13.07.2025).
25. Rahmati M. Towards Explainable and Lightweight AI for Real-Time Cyber Threat Hunting in Edge Networks. URL: https://doi.org/10.48550/arXiv.2504.16118 [in English] (date of access: 14.07.2025).
26. Yagiz M. A., Goktas P. LENS-XAI: Redefining Lightweight and Explainable Network Security through Knowledge Distillation and Variational Autoencoders for Scalable Intrusion Detection in Cybersecurity. URL: https://doi.org/10.48550/arXiv.2501.00790 [in English] (date of access: 15.07.2025).

СПИСОК ЛІТЕРАТУРИ

1. Otoum Y., Nayak A. AS-IDS: Anomaly and Signature Based IDS for the Internet of Things / Journal of Network and Systems Management. – 2021. – Vol. 29, no. 3. – URL: https://doi.org/10.1007/s10922-021-09589-6 (дата звернення: 14.06.2025).
2. Okoye Chinwe C., Nwankwo Ezinwa E., Usman Favour O., Mhlongo N. Z., Odeyemi O., Ike C. U. Securing financial data storage: A review of cybersecurity challenges and solutions / International Journal of Science and Research Archive. – 2024. – Vol. 11, no. 1. – C. 1968–1983. – URL: https://doi.org/10.30574/ijsra.2024.11.1.0267 (дата звернення: 15.06.2025).
3. Agrawal S., Sarkar S., Aouedi O., Yenduri G., Piamrat K., Alazab M., Bhattacharya S., Maddikunta P. K. R., Gadekallu T. R. Federated Learning for intrusion detection system: Concepts, challenges and future directions / Computer Communications. – 2022. – URL: https://doi.org/10.1016/j.comcom.2022.09.012 (дата звернення: 16.06.2025).
4. Ali M. L. et al. Deep Learning vs. Machine Learning for Intrusion Detection in Computer Networks: A Comparative Study / Applied Sciences. – 2025. – Vol. 15, no. 4. – P. 1903. – URL: https://doi.org/10.3390/app15041903 (дата звернення: 16.06.2025).
5. Vinayakumar R., Alazab M., Soman K. P., Poornachandran P., Al-Nemrat A., Venkatraman S. Deep Learning Approach for Intelligent Intrusion Detection System / IEEE Access. – 2019. – Vol. 7. – C. 41525–41550. – URL: https://doi.org/10.1109/access.2019.2895334 (дата звернення: 19.06.2025).
6. Gaspar D., Silva P., Silva C. Explainable AI for Intrusion Detection Systems: LIME and SHAP Applicability on Multi-Layer Perceptron / IEEE Access. – 2024. – C. 1. – URL: https://doi.org/10.1109/access.2024.3368377 (дата звернення: 19.06.2025).
7. Fatema K. et al. Federated XAI IDS: An Explainable and Safeguarding Privacy Approach to Detect Intrusion Combining Federated Learning and SHAP / Future Internet. – 2025. – Vol. 17, no. 6. – P. 234. – URL: https://doi.org/10.3390/fi17060234 (дата звернення: 21.06.2025).
8. Arreche O., Guntur T., Abdallah M. XAI-IDS: Toward Proposing an Explainable Artificial Intelligence Framework for Enhancing Network Intrusion Detection Systems / Applied Sciences. – 2024. – Vol. 14, no. 10. – P. 4170. – URL: https://doi.org/10.3390/app14104170 (дата звернення: 21.06.2025).
9. Sajid M. et al. Enhancing intrusion detection: a hybrid machine and deep learning approach / Journal of Cloud Computing. – 2024. – Vol. 13, no. 1. – URL: https://doi.org/10.1186/s13677-024-00685-x (дата звернення: 24.06.2025).
10. Alalwany E. et al. Stacking Ensemble Deep Learning for Real-Time Intrusion Detection in IoMT Environments / Sensors. – 2025. – Vol. 25, no. 3. – P. 624. – URL: https://doi.org/10.3390/s25030624 (дата звернення: 25.06.2025).

11. Laxmi, Chauhan K. AI-Based Intrusion Detection Systems for Novel Attacks in IoT and APTs: A Deep Learning-Centric Review / International Journal of Computer Science and Information Security. – Vol. 23, no. 3, May–June. – URL: https://www.academia.edu/130243382/AI_Based_Intrusion_Detection_Systems_for_Novel_Attacks_in_IoT_and_APTs_A_Deep_Learning_Centric_Review?bulkDownload=true (дата звернення: 25.06.2025).
12. Sinha P. et al. A high performance hybrid LSTM CNN secure architecture for IoT environments using deep learning / Scientific Reports. – 2025. – Vol. 15, no. 1. – URL: https://doi.org/10.1038/s41598-025-94500-5 (дата звернення: 27.06.2025).
13. Ribeiro M. T., Singh S., Guestrin C. "Why Should I Trust You?" / Proceedings of the 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD). – 2016. – New York, NY, USA. – URL: https://doi.org/10.1145/2939672.2939778 (дата звернення: 27.06.2025).
14. Lundberg S. M., Lee S.-I. A unified approach to interpreting model predictions / Advances in Neural Information Processing Systems. – 2017. – Vol. 30. – URL: https://arxiv.org/abs/1705.07874v2 (дата звернення: 27.06.2025).
15. Mohale V. Z., Obagbuwa I. C. A systematic review on the integration of explainable artificial intelligence in intrusion detection systems to enhancing transparency and interpretability in cybersecurity / Frontiers in Artificial Intelligence. – 2025. – Vol. 8. – URL: https://doi.org/10.3389/frai.2025.1526221 (дата звернення: 04.07.2025).
16. Neupane S. et al. Explainable Intrusion Detection Systems (X-IDS): A Survey of Current Methods, Challenges, and Opportunities / IEEE Access. – 2022. – Vol. 10. – P. 112392–112415. – URL: https://doi.org/10.1109/access.2022.3216617 (дата звернення: 04.07.2025).
17. Alomari Y., Andó M. SHAP-based insights for aerospace PHM: Temporal feature importance, dependencies, robustness, and interaction analysis / Results in Engineering. – 2024. – Vol. 21. – P. 101834. – URL: https://doi.org/10.1016/j.rineng.2024.101834 (дата звернення: 06.07.2025).
18. Patil S. et al. Explainable Artificial Intelligence for Intrusion Detection System / Electronics. – 2022. – Vol. 11, no. 19. – P. 3079. – URL: https://doi.org/10.3390/electronics11193079 (дата звернення: 06.07.2025).
19. Visani G. LIME: explain Machine Learning predictions / Medium. – URL: https://medium.com/data-science/lime-explain-machine-learning-predictions-af8f18189bfe (дата звернення: 07.07.2025)
20. Caforio F. P. et al. Leveraging Grad-CAM to Improve the Accuracy of Network Intrusion Detection Systems / Discovery Science. – Cham, 2021. – P. 385–400. – URL: https://doi.org/10.1007/978-3-030-88942-5_30 (дата звернення: 08.07.2025).
21. Kalakoti R. et al. Evaluating Explainable AI for Deep Learning-Based Network Intrusion Detection System Alert Classification / 11th Int. Conf. on Info Systems Security and Privacy, Porto, Portugal. – 2025. – P. 47–58. – URL: https://doi.org/10.5220/0013180700003899 (дата звернення: 09.07.2025).
22. Ribeiro M. T., Singh S., Guestrin C. Anchors: High-Precision Model-Agnostic Explanations / Proceedings of the AAAI Conference on Artificial Intelligence. – 2018. – Vol. 32, no. 1. – URL: https://doi.org/10.1609/aaai.v32i1.11491 (дата звернення: 11.07.2025).
23. Corea P. M. et al. Explainable AI for Comparative Analysis of Intrusion Detection Models / 2024 IEEE Int. Mediterranean Conf. on Communications and Networking (MeditCom), Madrid, Spain, 8–11 July 2024. – 2024. – P. 585–590. – URL: https://doi.org/10.1109/meditcom61057.2024.10621339 (дата звернення: 13.07.2025).
24. Shendkar B. D. Explainable Machine Learning Models for Real-Time Threat Detection in Cybersecurity / Panamerican Mathematical Journal. – 2024. – Vol. 35, no. 1s. – P. 264–275. – URL: https://doi.org/10.52783/pmj.v35.i1s.2313 (дата звернення: 13.07.2025).
25. Rahmati M. Towards Explainable and Lightweight AI for Real-Time Cyber Threat Hunting in Edge Networks / arXiv. – URL: https://doi.org/10.48550/arXiv.2504.16118 (дата звернення: 14.07.2025)
26. Yagiz M. A., Goktas P. LENS-XAI: Redefining Lightweight and Explainable Network Security through Knowledge Distillation and Variational Autoencoders for Scalable Intrusion Detection in

Cybersecurity / arXiv. – URL: https://doi.org/10.48550/arXiv.2501.00790 (дата звернення: 15.07.2025).

Глега Катерина Володимирівна *магістр; Інститут спеціального зв'язку та захисту інформації Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Верхньоключова, 4, м. Київ, Україна, 03056*

Голь Владислав Дмитрович *професор; завідувач Спеціальної кафедри №1; Інститут спеціального зв'язку та захисту інформації Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Верхньоключова, 4, м. Київ, Україна, 03056*

Оптимізація XAI для швидкодійних нейромережевих систем виявлення аномалій у трафіку

Актуальність. У сучасних мережевих середовищах системи виявлення вторгнень (IDS), що базуються на технологіях глибокого навчання, демонструють значні переваги у виявленні складних і динамічних кіберзагроз. Однак їх широке практичне застосування суттєво обмежене обчислювальною складністю, високими затримками та низькою інтерпретованістю ухвалених рішень, відомою як проблема «чорної скриньки». Інтеграція методів пояснюваного штучного інтелекту (XAI) у нейромережеві системи IDS є необхідною умовою для забезпечення прозорості ухвалення рішень, довіри операторів та ефективності оперативного реагування на кіберінциденти в режимі реального часу.
Мета. Основною метою дослідження є розроблення та оптимізація методів XAI для нейромережевих систем виявлення аномалій у мережевому трафіку, що здатні функціонувати з низькими затримками в реальному часі, забезпечуючи баланс між прозорістю ухвалених рішень, обчислювальною ефективністю та точністю класифікації загроз.
Методи дослідження. У роботі здійснено системний огляд і порівняльний аналіз сучасних моделей глибокого навчання (CNN, LSTM, GRU, автоенкодери, гібридні моделі CNN-LSTM) та найбільш поширених методик XAI (SHAP, LIME, Integrated Gradients, DeepLIFT, Grad-CAM, Anchors). Розроблено оптимізаційні підходи, які включають апаратне прискорення, застосування спрощених методів пояснення на основі градієнтів, створення гібридних архітектур із вбудованими механізмами інтерпретації (наприклад, CNN-LSTM із механізмами уваги) та вибіркове пояснення рішень. Емпірична перевірка запропонованих рішень проведена на загальновідомих наборах даних (CICIDS2017, NSL-KDD, UNSW-NB15).
Результати. За результатами аналізу встановлено, що градієнтні методи пояснення (Integrated Gradients, DeepLIFT) найбільш придатні для інтеграції у високошвидкісні IDS завдяки мінімальному часу генерації пояснень і високій точності. Гібридні архітектури з вбудованими механізмами пояснення (ELAI framework на основі CNN-LSTM із механізмами уваги) продемонстрували високу ефективність: точність виявлення перевищила 98%, а час прийняття рішення не перевищував 10 мс. Оптимізовані методики дозволили істотно підвищити ефективність виявлення атак типу «нульового дня» до рівня 91,6%.
Висновки. У результаті проведеного дослідження запропоновано практичні підходи щодо інтеграції пояснюваності в нейромережеві системи IDS, які функціонують у режимі реального часу, що дозволяє суттєво підвищити якість виявлення загроз, прозорість рішень та довіру до систем з боку операторів кібербезпеки. Перспективи подальших досліджень пов'язані зі стандартизацією оцінювання пояснюваності, вдосконаленням архітектур на основі механізмів уваги та розширенням цих підходів на інші завдання кібербезпеки.

Ключові слова: *кібербезпека, системи виявлення вторгнень, глибоке навчання, пояснюваний штучний інтелект, виявлення аномалій, нейронні мережі, оптимізація XAI.*

Надійшла у першій редакції 22.04.2025, в останній - 27.05.2025.
The first version has been received on 22.04.2025, the final version - on 27.05.2025.

DOI: https://doi.org/10.26565/2304-6201-2025-66-03

УДК (UDC) 004.65; 004.8

Горбачова Людмила Олегівна *студентка кафедри інтелектуальних програмних систем і технологій; Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 4, Харків-22, Україна, 61022*
e-mail: xa12850503@student.karazin.ua
https://orcid.org/0000-0002-6053-7235

Хруслов Максим Михайлович *завідувач кафедри комп'ютерних систем та робототехніки, кандидат фізико-математичних наук, старший дослідник, доцент; Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 4, Харків-22, Україна, 61022*
e-mail: maksym.khruslov@karazin.ua
https://orcid.org/0000-0001-9639-9340

Чуб Ольга Ігорівна *доцент закладу вищої освіти кафедри комп'ютерних систем та робототехніки, кандидат економічних наук; Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 4, Харків-22, Україна, 61022*
e-mail: o.i.chub@karazin.ua
https://orcid.org/0000-0002-1216-856X

Бережний Артем Андрійович *старший викладач закладу вищої освіти кафедри комп'ютерних систем та робототехніки, магістр; Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 4, Харків-22, Україна, 61022*
e-mail: artem.berezhnyi@karazin.ua;
https://orcid.org/0000-0001-5407-9015

Козюберда Дмитро Олександрович *магістр кібербезпеки, факультет комп'ютерних наук, Харківськийнаціональний університет імені В. Н. Каразіна; співробітник-розробник ТОВ «ЛАДИЗАЙН»; Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 4, Харків-22, Україна, 61022*
e-mail: koziuberda.dmytro@gmail.com;
https://orcid.org/0009-0005-3088-9685

Дослідження процедури перетворення тексту в SQL на основі large language models (LLM) шляхом міждоменного семантичного аналізу

Theme of work. Research on the Text-to-SQL conversion procedure based on Large Language Models (LLM) through Cross-Domain Semantic Analysis.
Purpose of work. To enhance the accuracy and adaptability of Text-to-SQL conversion using Large Language Models (LLM) through cross-domain semantic analysis, enabling reliable query interpretation across various domains and database structures. **Methods of research.** Comparative analysis, experimental evaluation, cross-domain semantic testing. **Results.** The research demonstrates that optimized prompt engineering and fine-tuning significantly improve the accuracy and cross-domain adaptability of Large Language Models for Text-to-SQL conversion. **Conclusions.** This study confirms that Large Language Models (LLMs) can effectively enhance the Text-to-SQL conversion process when optimized with targeted prompt engineering and fine-tuning. Cross-domain semantic analysis proved essential for enabling LLMs to handle varied database structures and domain-specific terminology, improving versatility and accuracy. The findings highlight the potential of LLMs to make SQL query generation more accessible to non-technical users, promoting broader application of AI in database management. Future work may focus on further refining these models to reduce computational costs and increase processing efficiency.

Ключові слова: *Large Language Models (LLM), Natural Language Processing (NLP), Text-to-SQL, Обробка природної мови, Глибоке навчання, Нейронні мережі, Бази даних, Штучний інтелект, Аналіз даних, Автоматизація, Інформаційні системи.*

Як цитувати: Горбачова Л. О., Хруслов М. М., Чуб О. І., Бережний А. А., Козюберда Д. О. Дослідження процедури перетворення тексту в SQL на основі large language models (LLM) шляхом міждоменного семантичного аналізу. *Вісник Харківського національного університету імені В. Н. Каразіна, серія Математичне моделювання. Інформаційні технології. Автоматизовані системи управління*. 2025. вип. 66. С. 37-44. https://doi.org/10.26565/2304-6201-2025-66-03

How to quote: Horbachova L., Khruslov M., Chub O., Berezhnyi A., Koziuberda D., “Research of the procedure for converting text into sql based on large language models (LLM) through cross-domaine semantic analysis”, *Bulletin of V.N. Karazin Kharkiv National University, series Mathematical modelling. Information technology. Automated control systems*, vol. 66, pp. 37-44, 2025. https://doi.org/10.26565/2304-6201-2025-66-03[in Ukrainian]

Вступ

Large Language Models (LLM) відкривають нові можливості для автоматизованого перетворення тексту у SQL-запити, що значно спрощує роботу з базами даних, роблячи її доступнішою для нетехнічних користувачів. Автоматичний переклад запитів природною мовою на SQL розширює спектр можливостей у сфері керування даними, зокрема для таких завдань, як моментальна генерація, агрегація та фільтрація даних. Використання міждоменного семантичного аналізу дозволяє моделі розпізнавати і коректно обробляти запити незалежно від предметної області або структури бази даних, з якою вона працює. Це значно підвищує універсальність і точність моделі при взаємодії з різними типами даних і доменами.

У роботі представлено всебічний аналіз методів Text-to-SQL на основі LLM, зокрема конструювання підказок, вибору та організації прикладів, що дозволяє виявити сильні та слабкі сторони різних підходів. На основі отриманих результатів пропонується інтегроване рішення, яке підвищує ефективність і знижує фінансові витрати на реалізацію Text-to-SQL завдання при використанні LLM. Результати можуть сприяти широкому впровадженню LLM у сфері баз даних та заохочувати подальші дослідження у цьому напрямку.

Загальні підходи в перетворенні тексту в SQL на основі LLM

Text-to-SQL має на меті автоматичний переклад запитів на природній мові в SQL-запити, що полегшує взаємодію неспеціалістів з базами даних та покращує обробку даних. Це технологічне рішення відкриває нові можливості для інтелектуальних баз даних і автоматизованого аналізу. Проте, реалізація Text-to-SQL стикається з труднощами у точному розумінні природної мови та генерації коректних SQL-запитів [1].

Дослідження у цій сфері зосереджено на підходах з використанням попередньо визначених правил та моделей машинного навчання з архітектурою кодер-декодер [1].. З розвитком глибокого навчання застосовуються різні методи, такі як механізми уваги та синтаксичний аналіз, що допомагають у розв’язанні завдання Text-to-SQL. Однією з найуспішніших моделей є BERT, яка продемонструвала відмінні результати. Щоб покращити точність, були створені великі тестові набори даних, як-от WikiSQL та Spider, що дозволили досягти прогресу в дослідженнях [1].

Останніми роками LLM, як GPT-4 та LLaMA, стали новим стандартом у обробці природної мови. LLM попередньо навчаються на величезних обсягах тексту та здатні виконувати різноманітні завдання. Основним аспектом їхньої роботи є генерування слів з найвищою ймовірністю на основі вхідних даних. Для успішного виконання Text-to-SQL важливим є ефективне формування запитів, що відомо як інженерія підказок [2].

Інженерія підказок [2] класифікується на сценарії з нульовою спробою, де не надається жодного прикладу, та сценарії з кількома спробами, коли надається обмежена кількість прикладів. Ефективне навчання в контексті дає можливість LLM виявляти патерни у запитах, що підвищує їхню здатність до генерації результатів без додаткового навчання. Проте, незважаючи на успіхи, все ще існує дефіцит досліджень, що зосереджуються на контрольованому точному налаштуванні LLM для Text-to-SQL.

Таким чином, основними аспектами Text-to-SQL на основі LLM є представлення запитів, навчання в контексті та контрольоване точне налаштування. Багато досліджень фокусуються на вилученні шаблонів SQL-запитів, однак основною проблемою є те, як підштовхнути LLM до генерації коректних SQL-запитів, що потребує детального вивчення інженерії підказок [2].

Нещодавні дослідження підтверджують важливу роль включення прикладів для ефективного навчання в контексті [3, 4].

Хоча існує багато успішних моделей, більшість з них сконцентровані на OpenAI, що залишає LLM з відкритим вихідним кодом маловивченими. Це є серйозною проблемою, адже такі моделі часто мають обмеження в розумінні контексту. Важливим завданням є поліпшення їхньої продуктивності у Text-to-SQL через контрольоване налаштування.

Ефективність використання підказок також є важливим питанням, оскільки витрати на API OpenAI можуть бути високими. При цьому інженерія підказок передбачає подання запитань, вибір прикладів та організацію прикладів [5]. Відкритими залишаються питання оптимізації довжини підказок для досягнення кращих результатів.

Існує нагальна потреба в систематичному дослідженні різних репрезентацій і вивченні того, як ефективно працювати з LLM. Щодо вибору прикладів, то поширеною практикою є кодування найбільш схожих прикладів в одному представленні з цільовим запитанням [3]. Тому дуже очікуваним є систематичне дослідження з конструювання підказок, що охоплює різні LLM, представлення запитів, вибір прикладів та організації [4].

Оперативність використання підказок залишається складним і відкритим питанням. У технологіях Text-to-SQL на основі LLM ще однією важливою проблемою є ефективність. Причина полягає в тому, що більшість попередніх досліджень зосереджені на OpenAI LLM, а виклик їхніх API є дорогим, трудомістким і обмеженим у швидкості, особливо для контекстних навчальних підказок з численними прикладами. Однак попередні дослідження не можуть добре вирішити даний виклик. Зокрема, на основі інвертованої U-подібної форми точності виконання підказок щодо довжини підказок припускається, що LLM можуть мати «золоту середину» з точки зору довжини підказок, але це все ще залишається складним відкритим питанням для дослідження [6].

Базове рішення Text-to-SQL для подальшого розгляду

У 2019 році науковці з Єльського університету представили Spider – складний набір даних для семантичного аналізу тексту і перетворення його в SQL. Spider містить понад 10 тисяч запитів і 5 тисяч унікальних SQL-запитів, що охоплюють різні домени, і вимагає від моделей здатності узагальнювати нові SQL-запити та схеми баз даних.

Spider відрізняється від попередніх наборів даних тим, що останні використовували одну базу даних, у той час як Spider має кілька баз. Результати експериментів показали, що навіть найкращі моделі досягають лише 12,4% точності, що свідчить про складність завдання.

В рамках цієї роботи буде використано таблицю лідерів Spider для вибору базового рішення Text-to-SQL. Найбільш цікаві результати демонструють моделі MiniSeek та DAIL-SQL з точністю 91,2% та 86,6% відповідно. Хоча MiniSeek не має публічного доступу, DAIL-SQL доступний для подальшого дослідження [7].

DAIL-SQL підтримує інтеграцію з різними LLM та стратегіями. У нашій роботі акцент буде на систематичному оцінюванні ефективності різних стратегій розробки, включаючи LLM з відкритим вихідним кодом. Ми плануємо порівняти варіанти відповідей у сценарії «нульового пострілу» та стратегії вибору прикладів і організації в сценарії з кількома спробами. Важливими аспектами також стануть потенціал LLM з відкритим кодом та ефективність використання токенів. Зрештою, метою є знайти збалансовану стратегію, яка оптимізує продуктивність та ефективність використання токенів, а також розробити практичне рішення на базі DAIL-SQL для реальних даних, і зробити це рішення універсальним для будь-яких доменів.

Проблематика представлення питання в контексті Text-to-SQL

Розглядаючи деяке запитання q в контексті певного домену і певної бази даних D, задачею генерування запитання є збільшення можливості LLM M сформувати коректний SQL s^* наступним чином:

$$\max_{\sigma} \quad \mathbb{P}_M\left(s^* | \sigma(q, D)\right), \tag{1}$$

де функція σ (·,·) визначає представлення для питання q, з інформацією про домен і структури БД зі схеми бази даних D. Також функція може містити додаткову інформацію інструкції, імплікацію правила та зовнішній ключ [8].

Імплікація правила (RI), Інструкція (INS), та зовнішній ключ (FK) є можливими компонентами підказки. Інструкція – це опис завдання, наприклад, «Напиши SQL як відповідь на запитання». Імплікація правила¬¬ – це наказове твердження, наприклад, «Виконай SQL-запит без пояснень». Зовнішній ключ (FK) – інформація про зовнішній ключ бази даних.

Є різні варіації підходи конструювання підказок такі як: базова підказка (BS p) (Basic Prompt), підказка представлення тексту (TR p), демонстраційний запит OpenAI (OD p), підказка представлення коду (CR p), Alpaca SFT Prompt (AS p).

Навчання в контексті: вибір та організація прикладів

У Text-to-SQL питанні, маючи набір трійок $Q = \{(q_i, s_i, D_i)\}$, де q_i і s_i - питання на природній мові і відповідний йому SQL-запит до бази даних D_i, метою навчання в контексті для Text-to-SQL є збільшення ймовірності того, що LLM M згенерує правильний SQL-запит s^* на цільове питання q і базу даних D наступним чином:

$$\max_{Q',\sigma} \quad \mathbb{P}_M\big(s^* | \sigma(q, D, Q')\big),$$

$$s.t. \quad |Q'| = k \quad and \;\; Q' \subset Q, \qquad (2)$$

де функція $\sigma(\cdot,\cdot,\cdot)$ визначає представлення для цільового питання q, з інформацією зі схеми в базі даних D та k прикладів, вибраних з Q.

При розгляданні DAIL-SQL буде робитися акцент на міждоменному Text-to-SQL, що означає, що цільова база даних D не належить до числа баз даних D, згаданих у Q, тобто, $D \notin \{D_i | (q_i, s_i, D_i) \in Q\}$. Контекстне навчання для Text-to-SQL передбачає вибір найбільш релевантних прикладів Q' і прийняття рішення про те, як переформовувати інформацію з цих вибраних прикладів у підказку.

Тобто це є дві окремі підзадачі: відбір прикладів та організація прикладів.

Вибір прикладів.

1) Випадковий вибір – це стратегія, що передбачає випадковий вибір k прикладів з доступних кандидатів [9].

2) Вибір подібних питань за маскою (Masked Question Similarity Selection(MQS)). Для міждоменного Text-to-SQL, MQS видаляє специфічно-доменну інформацію, змінюючи назви таблиць, стовпців і т.д. на лексеми-маски, а потім обчислює подібність їх вбудовування за алгоритмом kNN [10].

3) Вибір подібності питань (Question Similarity Selection, QTS). QTS вибирає число k прикладів з найбільш релевантними запитаннями, схожими по схемі. Далі він застосовує евклідову відстань до кожної пари приклад-ціль. Нарешті, алгоритм kNN використовується для вибору k прикладів з Q, які найбільш точно відповідають первинному питанню q [10].

4) Відбір за схожістю запитів (Query Similarity Selection (QRS)). QRS передбачає вибір k прикладів, схожих на цільовий SQL-запит s^*. QRS також генерує SQL-запит s^' з використанням цільового запитання q та бази даних D, де цей згенерований s^' можна розглядати як наближення до s^*. Далі запити кодуються у двійкові дискретні синтаксичні вектори відповідно до їх ключових слів. Після цього обираються k прикладів, враховуючи як схожість з наближеним запитом s', так і відмінності між обраними прикладами [9].

Стратегії, що вказані вище, концентруються на виборі прикладів на основі цільового запитання, однак, враховуючи дослідження [9] контекстне навчання являє собою навчання за аналогією. У випадку Text-to-SQL основною ціллю є формування запитів SQL на основі питання природньою мовою, відображення запитань у SQL-запити є набором навчання для LLM, тож варто враховувати як самі запитання, так і відповіді.

Організація прикладів має важливу роль у визначенні, яку саме інформацію з поданих прикладів буде сформовано у підказку. Існують два види організації: організація повної інформації та організація на основі SQL.

Повно-інформаційна організація (Full-Information Organization ¬¬– FI o) структурує приклади в представленні як цільове запитання, але відмінність закладається в тому, що замість лексеми «SELECT», наприклад, в кінці, конкретні приклади мають сформовані SQL-запити [9].

Організація, що використовує тільки SQL (SQL-Only Organization – SO o) включає SQL-запити обраних прикладів з префіксною інструкцією у підказці. Така організація має на меті збільшити кількість прикладів з мінімальною довжиною токенів [11]. Однак вона виключає інформацію про зв'язок між природнім запитанням та відповідними SQL-запитом, проте, як зазначалось раніше, такий зв'язок може бути корисним.

Підсумовуючи, Full-Information Organization відображає цілісну інформацію про приклади, тоді як SQL-Only Organization зберігає лише SQL-запити для додавання більшої кількості прикладів. В контексті дослідження важливо зрозуміти, чи існує вигідний компроміс між кількістю і якістю в організації прикладів, що може бути додатково корисним для основної задачі Text-to-SQL.

$$\max_{Q',\sigma} \quad \mathbb{P}_M\big(s^*|\sigma(q,D,Q')\big),$$

$$s.t. \quad |Q'| = k \quad and \;\; Q' \subset Q, \qquad (2)$$

де функція $\sigma(\cdot,\cdot,\cdot)$ визначає представлення для цільового питання q, з інформацією зі схеми в базі даних D та k прикладів, вибраних з Q.

При розгляданні DAIL-SQL буде робитися акцент на міждоменному Text-to-SQL, що означає, що цільова база даних D не належить до числа баз даних D, згаданих у Q, тобто, $D \notin \{D_i|(q_i,s_i,D_i) \in Q\}$. Контекстне навчання для Text-to-SQL передбачає вибір найбільш релевантних прикладів Q' і прийняття рішення про те, як переформовувати інформацію з цих вибраних прикладів у підказку.

Тобто це є дві окремі підзадачі: відбір прикладів та організація прикладів.

Вибір прикладів.

1) Випадковий вибір – це стратегія, що передбачає випадковий вибір k прикладів з доступних кандидатів [9].

2) Вибір подібних питань за маскою (Masked Question Similarity Selection(MQS)). Для міждоменного Text-to-SQL, MQS видаляє специфічно-доменну інформацію, змінюючи назви таблиць, стовпців і т.д. на лексеми-маски, а потім обчислює подібність їх вбудовування за алгоритмом kNN [10].

3) Вибір подібності питань (Question Similarity Selection, QTS). QTS вибирає число k прикладів з найбільш релевантними запитаннями, схожими по схемі. Далі він застосовує евклідову відстань до кожної пари приклад-ціль. Нарешті, алгоритм kNN використовується для вибору k прикладів з Q, які найбільш точно відповідають первинному питанню q [10].

4) Відбір за схожістю запитів (Query Similarity Selection (QRS)). QRS передбачає вибір k прикладів, схожих на цільовий SQL-запит s^*. QRS також генерує SQL-запит s^' з використанням цільового запитання q та бази даних D, де цей згенерований s^' можна розглядати як наближення до s^*. Далі запити кодуються у двійкові дискретні синтаксичні вектори відповідно до їх ключових слів. Після цього обираються k прикладів, враховуючи як схожість з наближеним запитом s', так і відмінності між обраними прикладами [9].

Стратегії, що вказані вище, концентруються на виборі прикладів на основі цільового запитання, однак, враховуючи дослідження [9] контекстне навчання являє собою навчання за аналогією. У випадку Text-to-SQL основною ціллю є формування запитів SQL на основі питання природньою мовою, відображення запитань у SQL-запити є набором навчання для LLM, тож варто враховувати як самі запитання, так і відповіді.

Організація прикладів має важливу роль у визначенні, яку саме інформацію з поданих прикладів буде сформовано у підказку. Існують два види організації: організація повної інформації та організація на основі SQL.

Повно-інформаційна організація (Full-Information Organization ¬¬– FI o) структурує приклади в представленні як цільове запитання, але відмінність закладається в тому, що замість лексеми «SELECT», наприклад, в кінці, конкретні приклади мають сформовані SQL-запити [9].

Організація, що використовує тільки SQL (SQL-Only Organization – SO o) включає SQL-запити обраних прикладів з префіксною інструкцією у підказці. Така організація має на меті збільшити кількість прикладів з мінімальною довжиною токенів [11]. Однак вона виключає інформацію про зв'язок між природнім запитанням та відповідними SQL-запитом, проте, як зазначалось раніше, такий зв'язок може бути корисним.

Підсумовуючи, Full-Information Organization відображає цілісну інформацію про приклади, тоді як SQL-Only Organization зберігає лише SQL-запити для додавання більшої кількості прикладів. В контексті дослідження важливо зрозуміти, чи існує вигідний компроміс між кількістю і якістю в організації прикладів, що може бути додатково корисним для основної задачі Text-to-SQL.

Доопрацювання DAIL-SQL

Варіантом вирішення проблем з відбором та організацією прикладів ми розглядаємо метод Text-to-SQL – DAIL-SQL (це гнучке рішення на основі LLM, яке можна розширювати та інтегрувати з іншими компонентами).

Нульовий постріл (zero-shot) — це підхід у машинному навчанні та обробці природної мови, коли модель виконує завдання без жодних прикладів або попереднього навчання на аналогічних завданнях. У цьому випадку модель намагається генерувати відповіді на основі загальних знань, закладених у її архітектуру під час попереднього тренування. При нульовому пострілі модель працює тільки на основі загальних знань, закладених під час її тренування, без прямого контексту чи зразків для поточного запиту. Це означає, що вона має розуміти завдання "з нуля" і формулювати SQL-запити, виходячи лише з розуміння мови, структури даних, а також синтаксису SQL.

Для максимізації продуктивності LLM в сценарії «нульового пострілу» є навчання в контексті, як альтернативний варіант є контрольоване доопрацювання (supervised fine-tuning), що є менш дослідженим на сьогодні. Для порівняння використовується точність збігу (EM) і точність виконання (EX). Точність збігу – збіг ключових слів SQL між прогнозованим SQL-запитом і базовою істиною. Точність виконання – це порівняння результатів виконання прогнозованого SQL-запиту з базовим SQL на тестових екземплярах бази даних [12].

Для всіх методів використовується однакова максимальна довжина питання, тобто 4096 для OpenAI LLM і 2048 для LLM з відкритим кодом. 200 токенів для генерації.

Висновки

На основі проведених експериментів можна зробити деякі емпіричні висновки та рекомендації:

- Для представлення запитань рекомендовано користуватися підказками представлення коду та демонстраційного запиту OpenAI, але інша інформація як імплікація правил та зовнішній ключ, може бути дуже корисною.
- Для вибору прикладу важлива схожість між питанням на природній мові та SQL-запитом. Ці два фактори разом є хорошим показником для розробки ефективної стратегії відбору.
- Якщо прийнята LLM є досить потужною, як GPT-4, наприклад, то представлення їм пар запитань і SQL-запитів є раціональним вибором. В іншому випадку краще представити їм повні інформаційні приклади.
- Наявність більшої кількості параметрів у LLM з відкритим вихідним кодом покращує Text-to-SQL завдання. Крім того, контрольоване доопрацювання є необхідним.

Також, у ході роботи було досліджено декілька стратегій для нульового та кількох пострілів, оцінено різні способи представлення питань, відбору та організації прикладів для LLM. Було виявлено, що використання DAIL-SQL у поєднанні з GPT-4 у сценарії з кількома пострілами дозволяє досягти найвищої точності з розглянутих, та забезпечує економічне використання токенів.

У межах базової моделі DAIL-SQL було проведено тестування стратегій використання й налаштування, і варто зазначити, що для максимізації продуктивності рекомендується сконцентруватись на питанні оптимізації підказок та відборі подібних прикладів запитань. Це дозволяє балансувати вартість і точність виконання SQL-запитів.

У результаті треба відзначити, що організація DAIL є більш економічною (вартість токенів), ніж повноінформаційний підхід, точність виконання при цьому висока (83.5% з GPT-4). Це доводить основне твердження, що представлення питань із включенням SQL у вигляді зовнішніх ключів є оптимальним як у точності, так і в економії ресурсів.

У процесі дослідження показано, що економічне та ефективне використання токенів є основною метрикою для реальних задач Text-to-SQL, враховуючи кошторис обчислень на OpenAI платформах.

СПИСОК ЛІТЕРАТУРИ

1. Katsogiannis-Meimarakis G., Koutrika G. Survey on Deep Learning Approaches for Text-to-SQL. VLDB. 2023. 32, 4. P. 905–936. URL: https://doi.org/10.1007/s00778-022-00776-8 Дата звернення: 21.08.2024.
2. A Comprehensive Evaluation of ChatGPT's Zero-Shot Text-to-SQL Capability / A. Liu et al. CoRR abs/2303. 2023. P. 13547. URL: DOI:10.1007/s00778-022-00776-8
3. Text-to-SQL Empowered by Large Language Models: A Benchmark Evaluation / D. Gao et al. Proceedings of the VLDB Endowment. 2024. Vol. 17, no. 5. P. 132–1145. URL: DOI:10.14778/3641204.3641221 Дата звернення: 13.09.2024.
4. RESDSQL: Decoupling Schema Linking and Skeleton Parsing for Text-to-SQL / H. Li et al. *37th AAAI Conference on Artificial Intelligence*, 2023. P. 13067–13075 URL: https://doi.org/10.48550/arXiv.2302.05965
5. C3: Zero-shot Text-to-SQL with ChatGPT / X. Dong et al. 2023 URL: https://doi.org/10.48550/arXiv.2307.07306
6. Spider: A Large-Scale Human-Labeled Dataset for Complex and Cross-Domain Semantic Parsing and Text-to-SQL Task / T. Yu et al. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium. Stroudsburg, PA, USA, 2018. URL: https://yale-lily.github.io/spider Дата звернення: 10.08.2024.
7. Stanford Alpaca: An Instruction-following LLaMA model / R. Taori et al. URL: https://en.wikipedia.org/wiki/Llama_(language_model)
8. Enhancing Few-shot Text-to-SQL Capabilities of Large Language Models: A Study on Prompt Design Strategies / L. Nan et al. CoRR abs/2305.12586. 2023. URL: https://doi.org/10.48550/arXiv.2305.12586
9. What Makes Good In-Context Examples for GPT-3? / J. Liu et al. In Pro- ceedings of Deep Learning Inside Out: The 3rd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures, 2022. P. 100–114. https://doi.org/10.18653/v1/2022.deelio-1.10
10. A Case-Based Reasoning Framework for Adaptive Prompting in Cross-Domain Text-to-SQL / C. Guo et al. CoRR abs/2304.13301. 2023. https://doi.org/10.48550/arXiv.2304.13301
11. Zhong R., Yu T., Klein D. Semantic Evaluation for Text-to-SQL with Distilled Test Suites. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020. P. 396–411. https://doi.org/10.18653/v1/2020.emnlp-main.29

Horbachova Liudmyla Olehivna
student of the Department of Intellectual Software Systems and Technologies; V.N. Karazin Kharkiv National University, Svobody Square, 4, Kharkiv-22, Ukraine, 61022
e-mail: xa12850503@student.karazin.ua
https://orcid.org/0000-0002-6053-7235

Khruslov Maksym Mikhailovich
Head of the Department of Computer Systems and Robotics, Candidate of Physical and Mathematical Sciences, Senior Researcher, Associate Professor; V.N. Karazin Kharkiv National University Karazin, Svobody Square, 4, Kharkiv-22, Ukraine, 61022
e-mail: maksym.khruslov@karazin.ua
https://orcid.org/0000-0001-9639-9340

Chub Olga Igorivna
Associate Professor of the Department of Computer Systems and Robotics, PhD in Economic;
V.N.Karazin Kharkiv National University
4 Svobody Square, Kharkiv-22, Ukraine, 61022
e-mail: o.i.chub@karazin.ua
https://orcid.org/0000-0002-1216-856X

Berezhnyi Artem Andriyovych
senior lecturer of the higher education institution of the Department of Computer Systems and Robotics, Master; V.N. Karazin Kharkiv National University, Svobody Square, 4, Kharkiv-22, Ukraine, 61022
e-mail: artem.berezhnyi@karazin.ua;
https://orcid.org/0000-0001-5407-9015

Koziuberda Dmytro Oleksandrovych
Master of Cybersecurity, Faculty of Computer Science, V. N. Karazin Kharkiv National University; development employee of LADYZAYN LLC.; V.N. Karazin Kharkiv National University, Svobody Square, 4, Kharkiv-22, Ukraine, 61022
e-mail: koziuberda.dmytro@gmail.com;
https://orcid.org/0009-0005-3088-9685

Research of the procedure for converting text into sql based on large language models (LLM) through cross-domaine semantic analysis

Theme of work. Research on the Text-to-SQL conversion procedure based on Large Language Models (LLM) through Cross-Domain Semantic Analysis. **Purpose of work.** To enhance the accuracy and adaptability of Text-to-SQL conversion using Large Language Models (LLM) through cross-domain semantic analysis, enabling reliable query interpretation across various domains and database structures. **Methods of research.** Comparative analysis, experimental evaluation, cross-domain semantic testing. **Results.** The research demonstrates that optimized prompt engineering and fine-tuning significantly improve the accuracy and cross-domain adaptability of Large Language Models for Text-to-SQL conversion. **Conclusions.** This study confirms that Large Language Models (LLMs) can effectively enhance the Text-to-SQL conversion process when optimized with targeted prompt engineering and fine-tuning. Cross-domain semantic analysis proved essential for enabling LLMs to handle varied database structures and domain-specific terminology, improving versatility and accuracy. The findings highlight the potential of LLMs to make SQL query generation more accessible to non-technical users, promoting broader application of AI in database management. Future work may focus on further refining these models to reduce computational costs and increase processing efficiency.

Keywords: *Large Language Models (LLM), Natural Language Processing (NLP), Text-to-SQL, Natural Language Processing, Deep Learning, Neural Networks, Databases, Artificial Intelligence, Data Analysis, Automation, Information Systems.*

Надійшла у першій редакції 05.04.2025, в останній - 14.05.2025.
The first version has been received on 05.04.2025, the final version - on 14.05.2025.

DOI: https://doi.org/10.26565/2304-6201-2025-66-04

УДК (UDC) 004.056.53

Дрозд Марія Ігорівна *здобувач вищої освіти ступеня магістра Інституту спеціального зв'язку та захисту інформації Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна*
e-mail: iammashadrozd@gmail.com
https://orcid.org/0009-0002-9736-8137

Нестеренко Сергій Дмитрович *старший викладач Інституту спеціального зв'язку та захисту інформації Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна*
e-mail: 654squad@gmail.com;
https://orcid.org/0000-0003-2097-1122

Аналіз програмного забезпечення для реалізації OSINT у сфері інформаційної безпеки

Актуальність. Глобальний сучасний кіберпростір характеризується стрімким зростанням ризиків та загроз для важливої інформації державних структур, бізнесу та суспільства. У таких умовах розвідка з відкритих джерел (OSINT) набуває актуального значення як інструмент для моніторингу інформаційного простору, виявлення потенційних загроз і забезпечення інформаційної безпеки. Програмне забезпечення для OSINT дозволяє ефективно збирати, аналізувати та інтерпретувати дані з відкритих джерел, включаючи соціальні мережі, публічні бази даних і веб-ресурси. Це сприяє своєчасному реагуванню на кіберзагрози, виявленню вразливостей і прийняттю рішень для захисту інформаційних систем і критичної інфраструктури суб'єктів інформаційних відносин держави.

Мета. Аналіз характеристик та можливостей сучасного спеціалізованого програмного забезпечення з метою їх ефективного застосування у якості інструментів розвідки з відкритих джерел (OSINT) у контексті виявлення потенційних загроз і забезпечення інформаційної безпеки суб'єктів інформаційних відносин.

Методи дослідження. У процесі написання статті використано методи технічного аналізу, порівняльно-описового підходу, систематизації та класифікації для дослідження функціональних можливостей інструментів OSINT, прогнозування їхньої ефективності та перспектив розвитку.

Результати. На основі проведеного аналізу визначено ключові характеристики програмних рішень, таких як Maltego, TheHarvester, Shodan, ZoomEye, LeakIX, Sublist3r та SubFinder, оцінено їхню придатність для моніторингу інформаційного простору, виявлення ризиків та вразливостей, а також своєчасне реагування з метою виключення негативних наслідків. Запропоновано рекомендації щодо оптимального використання цих інструментів на сучасних ПЕОМ з урахуванням вимог до апаратного забезпечення, безпеки та автоматизації процесів.

Розгляд прикладних аспектів використання OSINT дає змогу сформулювати практичні рекомендації для фахівців у сфері кібербезпеки. Здійснений аналіз дозволяє інтегрувати результати у навчальні програми для підготовки спеціалістів із захисту інформації. Встановлено, що ефективність OSINT значною мірою залежить від рівня підготовки користувача та його вміння інтерпретувати отриману інформацію. Розглянутий матеріал демонструє перспективи використання машинного навчання для автоматизації процесів збору та фільтрації даних. Зроблено акцент на необхідності безперервного оновлення баз знань і алгоритмів, що використовуються в OSINT. Результати дослідження можуть бути використані для створення комплексних рішень з метою забезпечення кіберстійкості організацій.

Висновки. Розвідка з відкритих джерел (OSINT) базується на зборі, систематизації та аналізі даних із загальнодоступних джерел, таких як соціальні мережі, веб-сайти, публічні бази даних та медіа. Основою функціонування програмного забезпечення для OSINT є використання автоматизованих інструментів, які дозволяють ефективно обробляти великі обсяги інформації, виявляти зв'язки між даними та ідентифікувати потенційні загрози інформаційній безпеці. Такі інструменти, як Maltego, TheHarvester, Shodan, ZoomEye, LeakIX, Sublist3r та SubFinder, забезпечують виконання завдань від пасивного збору даних до активного аналізу мережевої інфраструктури, що дає змогу виявляти вразливості, моніторити кіберпростір та підтримувати прийняття своєчасних рішень у сфері інформаційної безпеки та захисту інформації.

Проведено класифікацію програмного забезпечення для OSINT за функціональним призначенням, виділивши три основні категорії: інструменти виявлення, вилучення та агрегації даних. Запропоновано порівняльний аналіз таких інструментів, як Maltego, TheHarvester, Shodan, ZoomEye, LeakIX, Sublist3r та SubFinder, з визначенням їхніх ключових характеристик, включаючи сумісність з операційними системами, методи збору інформації, автоматизацію процесів та рівень безпеки, що сприяє вибору оптимального інструменту для вирішення завдань моніторингу кіберпростору та протидії інформаційним загрозам.

Наведено перспективні напрямки подальшого розвитку програмного забезпечення для OSINT у сфері кібербезпеки держави.

Ключові слова: *OSINT, програмне забезпечення, аналіз даних, автоматизація, вразливості, .кібербезпека, інформаційна безпека.*

Як цитувати: Дрозд М. І., Нестеренко С. Д. Аналіз програмного забезпечення для реалізації OSINT у сфері інформаційної безпеки. *Вісник Харківського національного університету імені В. Н. Каразіна, серія Математичне моделювання. Інформаційні технології. Автоматизовані системи управління*. 2025. вип.. 66. С.45-55. https://doi.org/10.26565/2304-6201-2025-66-04

How to quote: Drozd M., Nesterenko S., "Analysis of software for the implementation of OSINT in the field of information security", *Bulletin of V.N. Karazin Kharkiv National University, series Mathematical modelling. Information technology. Automated control systems*, vol. 66, pp. 45-55, 2025. https://doi.org/10.26565/2304-6201-2025-66-04 [in Ukrainian]

Вступ

Стрімкий розвиток цифрових технологій та глобалізація інформаційного простору значно підвищили залежність державних структур, бізнесу та суспільства від інформаційних систем і мережі Інтернет. Це спричинило зростання ризиків, пов'язаних із кібератаками, витоком конфіденційної інформації та маніпуляцією даними. У таких умовах розвідка з відкритих джерел (OSINT) набуває ключового значення як інструмент для моніторингу інформаційного простору, виявлення потенційних загроз і забезпечення інформаційної безпеки. Програмне забезпечення для OSINT дозволяє ефективно збирати, аналізувати та інтерпретувати дані з відкритих джерел, включаючи соціальні мережі, публічні бази даних і веб-ресурси. Це сприяє своєчасному реагуванню на кіберзагрози, виявленню вразливостей і підтримці стратегічного прийняття рішень для захисту інформаційних систем і критичної інфраструктури.

Постановка проблеми

Значною проблемою в реалізації розвідки з відкритих джерел (OSINT) у сфері інформаційної безпеки є стрімке зростання обсягів даних у поєднанні зі складністю їх обробки та аналізу для виявлення релевантної інформації. Сучасні інструменти OSINT стикаються з викликами, пов'язаними з різноманітністю джерел даних, їхньою динамічною природою та необхідністю забезпечення точності й актуальності результатів. З одного боку, зростання обсягів відкритих даних, зокрема з соціальних мереж, веб-ресурсів і публічних баз даних, ускладнює швидке виділення значущих зав'язків і патернів. З іншого боку, активне використання зловмисниками автоматизованих засобів для приховування слідів своєї діяльності, таких як маскування мережевого трафіку чи використання Dark Web, підвищує вимоги до ефективності та гнучкості програмного забезпечення OSINT. Крім того, обмежена сумісність деяких інструментів із різними операційними системами, висока ресурсоємність та необхідність захисту зібраних даних від несанкціонованого доступу ускладнюють їхнє застосування на сучасних ПЕОМ. У контексті гібридних загроз і кіберзлочинності перед науковцями постають завдання поглибленого аналізу сучасного програмного забезпечення для реалізації OSINT у сфері інформаційної безпеки, визначення його функціональних можливостей та ефективності, розробки спеціалізованого програмного забезпечення з метою підвищення продуктивності інструментів OSINT та його адаптивності до нових викликів інформаційної безпеки.

Аналіз останніх досліджень і публікацій

У науковій літературі активно досліджуються можливості розвідки з відкритих джерел (OSINT) для забезпечення інформаційної безпеки. У роботі [5] Williams H. та Blum I. розглянуто другу генерацію OSINT, акцентуючи увагу на його застосуванні в оборонній сфері, зокрема для аналізу великих обсягів даних із відкритих джерел. У [6] Unver A. представлено огляд цифрового OSINT, де підкреслюється важливість аналізу соціальних мереж і Dark Web для виявлення загроз. Дослідження [7] Schwarz K., Schwarz F. Та Creutzburg R. присвячено практичному застосуванню інструментів OSINT, таких як Maltego, для аналізу зав'язків між даними, а також розробці навчальних лабораторних вправ. У [9] Duffy M., Pan X. Та Wilson S. описано методи збору публічної інформації за допомогою інструментів, таких як TheHarvester, з акцентом на пасивні та активні підходи до пошуку даних.

Ураховуючи вищезазначене, дослідження програмного забезпечення для OSINT залишається актуальним завданням, оскільки воно сприяє вдосконаленню методів моніторингу кіберпростору та протидії кіберзагрозам.

Мета статті та завдання

Метою статті є аналіз функціональних можливостей сучасного програмного забезпечення для реалізації OSINT, пошук шляхів їх ефективного використання у сфері кібербезпеки держави.

У відповідності до поставленої мети головними завданнями є:

- аналіз складу та характеристик сучасних інструментів OSINT (Maltego, TheHarvester, Shodan, ZoomEye, LeakIX, Sublist3r та SubFinder), оцінка їхньої можливості з моніторингу інформаційного простору, виявлення ризиків та загроз;
- класифікація програмного забезпечення для реалізації OSINT за функціональним призначенням, обґрунтування рекомендації щодо його оптимального використання на сучасних ПЕОМ з урахуванням вимог до апаратного забезпечення, безпеки та автоматизації процесів;
- обґрунтування перспективних шляхів подальшого розвитку та ефективного використання програмного забезпечення для OSINT у сфері кібербезпеки держави.

Виклад основного матеріалу дослідження

Розвідка з відкритих джерел (OSINT) базується на зборі та аналізі інформації з загальнодоступних джерел, таких як соціальні мережі, веб-сайти, публічні бази даних та медіа, що робить її важливим інструментом для забезпечення інформаційної безпеки. Програмне забезпечення для OSINT, зокрема Maltego , TheHarvester, Shodan, ZoomEye, LeaklX, Sublist3r Та SubFinder, забезпечує виконання широкого спектра завдань: від пасивного збору даних до активного аналізу мережевої інфраструктури.

Зробимо детальний аналіз наведених вище інструментів OSINT, з'ясуємо їх функціональні можливості, а також позитивні властивості, які забезпечують їх ефективне використання.

Програмне забезпечення Shodan.

Shodan - це "пошукова система" об'єктів, підключених до Інтернету, включаючи сервери, роутери, веб-сайти, бази даних, камери, промислові системи управління (ПСУ), камери, мережеві сховища та пристрої Інтернету речей (IoT). Shodan індексує сервісні банери (метадані про програмне забезпечення, що працює на пристрої) і робить їх доступними для пошуку.

Shodan пропонує такі функції:

ShodanSearch: основна пошукова система, яка робить інформацію, зібрану Shodan, доступною через веб-сайт.

ShodanMonitor: дозволяє відстежувати всі свої пристрої, до яких є прямий доступ з Інтернету, для моніторингу безпеки.

ShodanMaps: надає можливість переглядати результати пошуку візуально, а не в текстовому вигляді на головному веб-сайті. Вона відображає до 1 000 результатів одночасно, а при збільшенні/зменшенні масштабу карти пошуковий запит підлаштовується так, щоб показувати результати лише для обраної області.

ShodanImages: Shodan збирає скріншоти для багатьох різних сервісів, і як користувач ви отримуєте доступ до інтерфейсу пошуку, який значно спрощує перегляд цих скріншотів.

ShodanTrends: дозволяє шукати дані, зібрані Shodan, щоб дізнатися про тенденції в Інтернеті.

Дане програмне забезпечення надає низку корисних послуг для розробників, а саме:

1. ShodanDeveloper API: всі веб-сайти Shodan повністю побудовані на одному публічному API Shodan, до якого мають доступ всі користувачі.

2. InternetDBAPI: надає швидкий спосіб побачити відкриті порти для IP-адреси. Він дає швидке уявлення про тип пристрою, який працює за IP-адресою.

3. GeoNetAPI: дозволяє запускати мережеві інструменти з серверів, розташованих по всьому світу. Надає можливість визначати різну поведінку мережі залежно від регіону, в якому знаходиться кінцевий користувач.

4. ShodanChrono: індикатор виконання скриптів.

5. CVEDBAPI: пропонує швидкий спосіб отримати доступ до інформації про вразливості [10].

Перевагами цього інструменту є широка видимість пристроїв Інтернету речей (IoT), а також потужні можливості пошуку та фільтрації. Проте головним недоліком залишаються потенційні проблеми з конфіденційністю та безпекою.

Програмне забезпечення на платформі ZoomEye.

ZoomEye - це безкоштовна платформа, що використовується для збору інформації про сервіси та пристрої, які підключені до Інтернету, а також оцінювання їхньої безпеки та виявлення вразливостей цих систем.

За допомогою ZoomEye можна реалізувати такі операції:

1. Сканування: ZoomEye використовує вузли спостереження, розташовані по всьому світу, для пошуку відкритих портів сервісів і пристроїв.

2. Захоплення банерів: після проведення перевірки сервісу або пристрою цей інструмент накопичує інформацію про банер на порту певного сервісу, на якому він запущений. Інформація про банер зазвичай містить таку інформацію про службу як список портів, що працюють, утиліти, що використовуються та їх версії, яке обладнання використовується для цієї служби та інші характеристики.

3. Індексування: дані, зібрані на минулому етапі, зберігаються та індексуються в базі даних ZoomEye.

4. Пошук і запити: база даних підключається до API ZoomEye, тому користувачі спроможні шукати будь-яку інформацію, що зберігається в цій базі. Пошук можна здійснювати за ключовими словами, а також застосовувати фільтри для точного пошуку.

Функціональні можливості програмного забезпечення на платформі ZoomEye допомагає забезпечити безпеку кіберпростору, а саме:

- здійснює зовнішню спостережність своєї цифрової присутності з перспективи стороннього спостерігача, що дозволяє виявляти слабкі місця системи та ініціювати їх своєчасне усунення;
- надає можливість виявити потенційні вразливості та неправильні конфігурації в мережі;
- допомагає виявляти помилки, які були допущені під час експлуатації мережі, а саме: відкриті порти, застаріле програмне забезпечення або незахищені конфігурації;
- дозволяє перевірити безпеку інших організацій, компаній, що допоможе у безпечному підборі партнерів, попередженню їх про вразливості, тобто управляти сторонніми ризиками;
- забезпечує дослідження та розвідку загроз, в результаті чого користувачі можуть дізнатися, які типи технологій найчастіше використовуються, а також дослідити нові загрози та потенційні вектори атак [11].

Проте цей інструмент несе за собою низку небезпек. Так як ZoomEye доступний для всіх, він може використовуватися зловмисниками для проведення розвідки. Крім того, автоматизувавши процес збору інформації з цього інструменту та інтегруючи її до свого інструментарію, зловмисники постійно матимуть оновлену інформацію про вразливості та доступність до ваших сервісів. З цією метою зловмисники можуть використати, наприклад програмне забезпечення LeakIX. Це потужний ресурс, який дозволяє етичним хакерам, фахівцям з безпеки та іншим користувачам здійснювати всебічний пошук конфіденційної інформації, яка може бути випадково доступна в Інтернеті.

Програмне забезпечення на платформі ZoomEye розділене на дві сфери пошуку:

Services (Сервіси) - це індексація всього, що було відсканована. Сюди входять IP-адреси та віртуальні хости. Зберігається різна інформація, наприклад, банер TCP або HTTP.

Leak (Витік) - в цій сфері індексуються неправильні конфігурації та вразливості, виявлені під час сканування сервісів. Сюди відносяться: виявлена вразливість, неправильна конфігурація інфраструктури, сторінки стану та моніторингу, що містять конфіденційну інформацію, загальнодоступні конфігураційні файли, що містять конфіденційну інформацію, неправильно сконфігуровані ACL (Access Control List), внаслідок чого служби, які мають бути захищені, стають загальнодоступними [12].

Перевагами цього інструменту є велика база даних витоків і витоків даних, а також розширені можливості пошуку та аналізу. Недоліками - обмежений вільний доступ та залежність від зовнішніх джерел даних для збору інформації.

Програмне забезпечення Sublist3r.

Sublist3r - це потужний інструмент, який можна використовувати для автоматизації процесу перерахування субдоменів. Це скрипт Python з відкритим вихідним кодом, який використовує різні методи для збору інформації про субдомени з різних джерел, включаючи пошукові системи, пасивні бази даних DNS і платформи соціальних мереж. Отримавши список субдоменів можна провести їх аналіз, щоб виявити потенційні вразливості та вектори атак [13].

Такий інструментарій зазвичай використовується у комплексі з іншими, адже Sublist3r реалізує лише цю функцію. Взаємодія з Sublist3r відбувається через командний рядок.

Недоліками цього інструменту є обмежені можливості налаштування параметрів сканування та залежність від зовнішніх джерел даних DNS для збору інформації. Перевагами є швидке та ефективне перерахування субдоменів та інтеграція з декількома джерелами даних.

Зважаючи на велику кількість доступних інструментів, докладно проведемо порівняння двох ПЗ, що найчастіше використовуються для збору даних: Maltego та TheHarvester.

Програмне забезпечення Maltego.

Maltego - це інструмент візуального аналізу посилань, який постачається з плагінами з відкритим вихідним кодом під назвою "трансформації", тобто модулі [7]. Maltego дозволяє створювати візуальні графіки зав’язків між даними, такими як email-адреси, IP-адреси та доменні імена, завдяки модулям (трансформаціям), що інтегруються з різними джерелами, включаючи Blockchain.info, Shodan та Social Links CE [7, 8]. Цей інструмент фокусується на аналізі реальних взаємозв'язків між загальнодоступною інформацією про інтернет-інфраструктуру, окремих осіб та організації.

Ці можливості Maltego реалізуються завдяки модулям, з яких складається дане програмне забезпечення, а саме:

модуль CaseFile Entities - це модуль візуального зображення інформації, яку можна використовувати для визначення взаємозв'язків різних типів інформації, в також для побудови графіків взаємозв'язків між частинами інформації;

модулі Blockchain.info та CipherTrac – це модулі для відслідковування і візуалізації зав’язків і транзакцій між криптогаманцями;

модуль Have I been Pwned? – являє собою модуль, який дозволяє провірити чи було зламано сайт, електронну пошту або акаунт, шляхом пошуку в злитих базах скомпрометованих паролів та іншої інформації;

модуль Hybrid Analysis - це незалежний сервіс, який працює на базі Falcon Sandbox і надає підмножину можливостей Falcon Sandbox. Falcon Sandbox - це автоматизоване рішення, яке призначене для аналізу шкідливого програмного забезпечення. Воно виконує глибокий аналіз загроз, збагачує результати аналітикою і надає дієві індикатори компрометації;

модуль PeopleMon – використовується для пошуку інформації про користувача, якого розшукують;

модуль Shodan – дозволяє в середині Maltego використовувати свої позитивні можливості. Shodan являє собою пошукову систему, яка збирає дані з підключених до інтернету пристроїв. Інформацію, яку надає нам цей модуль, це метадані про програмне забезпечення, яке працює на пристрої. Він також дозволяє дослідникам швидко відстежувати відкриті порти, імена хостів та вразливості та пов'язані з IP-адресами;

модуль Social Links CE – дозволяє знаходити відомості про людей та компанії, завдяки використанню різних баз даних, а також надає можливість пошуку реєстраційних даних компаній [8].

Тож, позитивною властивістю та перевагою Maltego є можливість визначати та створювати зв'язки в межах набору даних будь-яким чином. Користувач не обмежений фіксованим форматом попередньо визначених трансформацій і має свободу змінювати концепцію візуалізації залежно від того, що саме є важливим для дослідження.

Програмне забезпечення TheHarvester

Програмне забезпечення TheHarvester спеціалізується на швидкому зборі публічної інформації про домени та компанії через командний рядок, використовуючи пасивні джерела, такі як Google, Bing, Twitter, а також активні методи, як-от перебір DNS [9].

Враховуючи те, що взаємодія з TheHarvester проводиться через командний рядок, користувач має можливість використовувати команди з певними параметрами.

Такими параметрами є:

- d: використовується для пошуку домену або назви компанії;
- b: використовується для вказання джерело даних: bing, google, twitter, yahoo та інші, або для пошуку в усіх джерелах – all;
- s: почати відлік результату з 0;
- v: надає можливість перевірити ім'я хоста через dns і шукати віртуальні хости:
- f: дозволяє зберегти результати у HTML та XML файл;
- n: виконує зворотній запит DNS для всіх знайдених діапазонів;
- c: виконує DNS-перебір для доменного імені;

- t: виконує пошук розширення DNS TLD (Top-Level Domain);
- e: дозволяє використовувати цей DNS-сервер;
- p: проскановує виявлені хости і перевірити їх на можливість перехоплення;
- l: обмежує кількість результатів для роботи;
- h: використовування бази даних SHODAN для запиту знайдених хостів.

Із наведеного вище переліку параметрів зрозумілими є й можливості програмного забезпечення TheHarvester в цілому.

Порівняння характеристик інструментів OSINT, а саме Maltego та TheHarvester, наведене в таблиці 1, яка демонструє їхні переваги та недоліки.

Таблиця 1. Характеристики сучасних інструментів (ПЗ) OSINT
Table 1. Characteristics of modern OSINT tools (software)

Важливі характеристики ПЗ	**MALTEGO**	**TheHarvester**
Сумісність з ОС	Windows, MacOS, Linux	Linux
Основне призначення	Аналіз посилань і пошук взаємозв'язків між даними (Email, IP-адреси, URL-адреси, телефонні номери та інше)	Виявлення публічної інформації про домен або компанію та додаткової інформації (Email, IP-адрес, URL-адреси, порти, імена працівників)
Метод збору інформації	Автоматизовані запити через модулі (трансформації)	Командний рядок з обмеженим набором параметрів
Формат відображення даних	Візуальні зв’язки	Текстові звіти з можливістю збереження у HTML та XML
Автоматизація	Так, через трансформації	Частково
Переваги	Великий набір трансформацій, інтелектуальний аналіз даних у реальному часі, візуалізація графіків, автоматизація запитів	Великий набір джерел даних, активні та пасивні методи збору інформації
Недоліки	Висока вартість, обмежена безкоштовна версія	Обмежена сумісність з ОС, відсутність візуалізації, відсутність документації

Отже, Maltego вирізняється високою функціональністю завдяки автоматизованим запитам і візуалізації, але потребує значних ресурсів, має високу вартість, а у разі використання безкоштовної версії має обмежені можливості. TheHarvester є ефективним для оперативного збору даних, але обмежений сумісністю з Linux і відсутністю візуалізації. Shodan забезпечує індексацію пристроїв, підключених до Інтернету, таких як сервери та IoT-пристрої, дозволяючи виявляти відкриті порти та вразливості [10]. ZoomEye пропонує подібні можливості, але з додатковими функціями для оцінки безпеки та виявлення неправильних конфігурацій [11]. LeakIX фокусується на пошуку конфіденційної інформації, що випадково стала доступною, тоді як Sublist3r та SubFinder ефективно перераховують субдомени, що є ключовим для аналізу потенційних векторів атак [12, 13].

Особливе значення в реалізації OSINT набувають методи збору даних, які поділяються на пасивні, напівпасивні та активні. Пасивний збір передбачає використання загальнодоступних джерел без взаємодії з цільовими системами, напівпасивний — обережне сканування з маскуванням трафіку, а активний — пряму взаємодію, як-от сканування портів [5]. Аналіз даних, зібраних інструментами OSINT, ускладнюється великими обсягами інформації, що потребує застосування технологій Big Data, хмарних обчислень і методів кластеризації для виділення

значущих груп даних. Наприклад, моніторинг соціальних мереж дозволяє створювати психологічні профілі та виявляти зв'язки між особами, тоді як аналіз Dark Web допомагає відстежувати незаконну діяльність [6].

Отже, аналіз програмного забезпечення для OSINT показує, що кожен інструмент має унікальні переваги залежно від завдання: Maltego підходить для комплексного аналізу зв'язків, TheHarvester - для оперативного збору даних, Shodan і ZoomEye - для моніторингу мережевих пристроїв, LeakIX - для пошуку витоків інформації, а Sublist3r та SubFinder - для перерахування субдоменів. Поєднання цих інструментів із сучасними технологіями, такими як штучний інтелект і Від Data, дозволяє створювати ефективні стратегії моніторингу кіберпростору та протидії кіберзагрозам, що є критично важливим для забезпечення інформаційної безпеки в умовах зростаючої інформаційної конкуренції.

Перспективи розвитку програмного забезпечення для OSINT тісно пов'язані з прогресом у сфері штучного інтелекту та машинного навчання. Алгоритми штучного інтелекту можуть автоматизувати відбір джерел, прогнозувати загрози та ідентифікувати об'єкти на зображеннях, що значно підвищує ефективність розвідки [14]. Однак сучасні інструменти стикаються з низкою викликів, таких як: високі вимоги до апаратного забезпечення, необхідність забезпечення конфіденційності даних, ризик використання інструментів зловмисниками для розвідки.

Подальші дослідження показали, що для ефективного використання OSINT на сучасних ПЕОМ, останні повинні мати, як мінімум такі характеристики: процесор – від чотирьохядерного і вище, оперативна пам'ять – не менше 8 ГБ (бажано 16 ГБ), твердотільні накопичувачі (SSD) та оптимізовані операційні системи, такі як Ubuntu, для зменшення ресурсоємності. Використання проксі-серверів, VPN, шифрування даних і регулярне оновлення програмного забезпечення є обов'язковими для забезпечення безпеки та конфіденційності.

Розглядаючи перспективи розвитку програмного забезпечення для OSINT, зараз можна зробити припущення, що подальше зростання ефективності використання OSINT буде пов'язана з високою обчислювальною потужністю процесорів ПЕОМ, їх високими можливостями оперативно виконувати складні завдання зі збору, обробки та аналізу великих обсягів даних. Здатність працювати з великими обсягами публічної інформації і комбінувати різноманітні набори даних з різних джерел значно підвищує ефективність та точність аналізу.

Важливим аспектом майбутнього розвитку програмного забезпечення для OSINT є застосування методів аналізу великих даних (Big Data) та машинного навчання. Ці технології дозволяють автоматизувати процеси розслідування та прийняття рішень, роблячи їх більш інтелектуальними та ефективними. Цей аспект буде одним із ключових у використанні OSINT, оскільки він позначить різницю між дослідженнями, керованими людиною, і дослідженнями, керованими штучним інтелектом.

Високі можливості зберігання, індексації та аналізу інформації дозволяють легко отримати доступ до великих обсягів даних та інформації. У цьому контексті, важливу роль відіграють бази даних, які можуть бути підключені до автоматичних систем збору інформації з відкритих джерел у поєднанні з системами автоматичної індексації зібраних даних та інформації, що дозволяє структурувати дані та інформацію, в тому числі впроваджувати політики щодо дозволів (прав доступу).

Технології штучного інтелекту можуть дозволити автоматизувати платформи для збору даних, а також оптимізувати вибір джерел на основі вимог до колекції. Алгоритми можуть генерувати моделі, які здатні передбачати певні завдання збору відповідно до поточної обробленої інформації, можуть ініціювати вибір найкращих джерел або визначати оптимальні частоти збору. Крім того, алгоритми глибокого навчання сприяють автоматичному прийняттю рішень, що дозволяє динамічно адаптувати завдання збору даних. Таким чином, колекція стає адаптивною, що також тягне за собою зменшення людського фактору. Крім того, застосування штучного інтелекту дозволить використовувати ефективні автоматизовані рішення на етапах обробки інформації та подальшого аналізу. Сценарний аналіз, прогнозний аналіз, встановлення як повторюваних, так і майбутніх закономірностей можливі за допомогою технологій штучного інтелекту. Машинне навчання може виявляти складні кореляції, які природно непередбачувані для людини, що значно покращує ефективність OSINT, а також, використовуючи технології штучного інтелекту, об'єкти можуть бути автоматично ідентифіковані по фотографіях.

Система OSINT є достатньо відкритою, щоб включати дані, які не були отримані з відкритих джерел. Це означає, що OSINT може бути більш ефективним, якщо додати зовнішню інформацію

для доповнення розслідувань. Наприклад, правоохоронні органи можуть використовувати ці технології для підвищення якості керованої інформації та боротьби з терористичними організаціями. У цьому випадку правоохоронні органи можуть співпрацювати з громадянами, оперативні служби можуть використовувати закриту інформацію про кіберзлочинців, а звичайні користувачі можуть поєднувати OSINT із соціальною інженерією для створення профілю своєї цілі.

Гнучке призначення та широка сфера застосування OSINT дозволяють розслідувати різноманітні проблеми та збирати інформацію по всьому кіберпростору. Це може бути корисним для екологічних, психологічних, стратегічних, журналістських, трудових та безпекових аспектів. Наприклад, у сфері злочинності та кібербезпеки OSINT може відстежувати підозрілих осіб або небезпечні групи, виявляти профілі впливу і вивчати тривожні сигнали [23].

Ще одним перспективним напрямком, пов’язаним з подальшим розвитком інструментів збору та аналізу інформації, є роботизація процесів. Ця технологія, дозволяє автоматизувати процеси шляхом конфігурації програмних роботів, які імітують та інтегрують дії людини, взаємодіючі з цифровими системами для виконання різних процесів. Вона дозволить автоматизувати повторювані завдання шляхом обробки та індексування великих обсягів даних, а також забезпечить кореляцію між базами даних, одержувачами, каналами зв'язку та іншими об'єктами. Важливим допоміжним інструментом роботизації процесів є така технологія штучного інтелекту, яка завдяки спеціальним статистичним алгоритмам, обробці природної мови та машинному навчанню дозволяє належним чином ідентифікувати джерела. Крім того, програмні рішення працюватимуть на постійно модернізованій апаратній інфраструктурі. Збільшення швидкості роботи та обчислювальних потужностей ПЕОМ сприятиме розробці все більш складних і досконалих алгоритмів [24].

Отже, розвідка даних з відкритих джерел OSINT вдосконалюється разом із загальними тенденціями технічного прогресу. Кожний етап зумовлює використання найефективніших технологій, відомих на певний час. Розвиток технологій штучного інтелекту, автоматизація роботизованих процесів, розробка та упровадження квантових комп'ютерів задля вирішення зростаючих потреб та можливостей у кіберпросторі майбутнього, без сумніву, зумовлюють передумови для пошуку у подальшій перспективі шляхів підвищення ефективності використання OSINT.

Висновки й перспективи подальших досліджень

У статті проведено аналіз програмного забезпечення для реалізації розвідки з відкритих джерел (OSINT) у сфері інформаційної безпеки, зокрема розглянуто інструменти Maltego, TheHarvester, Shodan, ZoomEye, LeakIX, Sublist3r та SubFinder.

Запропоновано класифікацію цих інструментів за функціональним призначенням (виявлення, вилучення, агрегація даних) та надано рекомендації щодо їхнього оптимального використання на сучасних ПЕОМ з урахуванням вимог до апаратного забезпечення, безпеки та автоматизації процесів.

Наведено можливі напрямки подальшого розвитку з метою підвищення ефективного використання програмного забезпечення для OSINT у сфері кібербезпеки держави.

Цінність отриманих результатів дослідження у сфері інформаційної безпеки полягає у поглибленому розумінні можливостей зазначеного вище програмного забезпечення для OSINT, щодо збору, аналізу та обробки відкритих даних для виявлення кіберзагроз і вразливостей. У роботі підкреслено необхідність гнучкого вибору інструментів залежно від типу загроз і специфіки інформаційного середовища. Результати можуть бути застосовані для вдосконалення моніторингу інформаційного простору, оптимізації роботи з великими обсягами даних та підвищення ефективності захисту інформаційних систем у сучасних умовах гібридних загроз і кіберзлочинності.

Практична цінність результатів дослідження полягає в систематизації характеристик інструментів OSINT, що дозволяє сформулювати у технічному завданні на створення спеціалізованого програмного забезпечення обґрунтованих вимог щодо підвищення продуктивності інструментів OSINT, його адаптивності до стрімкого зростання гібридних загроз і кіберзлочинності у сучасних умовах.

Особливу увагу слід приділити розробці адаптивних моделей, здатних динамічно підлаштовуватися до нових джерел даних і типів кіберзагроз. Крім того, перспективним є

створення інтегрованих платформ, які поєднують можливості OSINT із технологіями Big Data та хмарними обчисленнями, а також розробка мобільних додатків для оперативного моніторингу інформаційного простору в реальному часі.

Впровадження таких технологій сприятиме підвищенню ефективності OSINT у сферах кібербезпеки, розвідки та соціальних досліджень, забезпечуючи швидке реагування на нові виклики інформаційної безпеки.

Отже, мета статті, яка полягала в аналізі програмного забезпечення для реалізації OSINT у сфері інформаційної безпеки, досягнута.

СПИСОК ЛІТЕРАТУРИ

1. National Defense Authorization Act for Fiscal Year 2006 : Public Law of 01.06.2006 no. No. 109-163. https://www.congress.gov/bill/109th-congress/house-bill/1815/text/statute
2. Гончаренко Ю., Канішев К. Інструменти інформаційної боротьби: OCIHT, ІПСО та протидія дезінформації. Інформаційно-психологічна операція (ІПСО). Як не стати жертвою чужих маніпуляцій. https://infolight.in.ua/wp-content/uploads/2023/02/brochure-2.pdf
3. Про національну безпеку України : Закон України від 21.06.2018 р. № 2469-VIII: станом на 31 берез. 2023 р. https://zakon.rada.gov.ua/laws/show/2469-19#Text
4. Про розвідувальні органи України : Закон України від 22.03.2001 р. № 2331-III : станом на 24 жовт. 2020 р. https://zakon.rada.gov.ua/laws/show/2331-14#Text
5. 5. Williams H., Blum I. Defining second generation open source intelligence (OSINT) for the defense enterprise. RAND Corporation, 2018. https://doi.org/10.7249/rr1964
6. Unver A. Digital open source intelligence and international security: a primer. EDAM, Oxford CTGA & Kadir Has Üniversitesi, 2018. 28 p. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3331638
7. Schwarz K., Schwarz F., Creutzburg R. Conception and implementation of professional laboratory exercises in the field of open source intelligence (OSINT). Electronic imaging. 2020. Vol. 2020, no.3. P.278 https://library.imaging.org/admin/apis/public/api/ist/website/downloadArticle/ei/33/3/art00010
8. Аналіз інструментів збору розвідувальної інформації з відкритих джерел / А. Карпенко та ін. Комунікаційні та інформаційні системи : Вісник. Київ, 2022. С. 21. https://www.viti.edu.ua/files/zbk/2022/2022-1.pdf#page=18
9. Duffy M., Pan X., Wilson S. Information reconnaissance by accumulating public information data sources. OALib. 2024. Vol. 11, no. 04. P. 1–25. https://doi.org/10.4236/oalib.1111463
10. 1Shodan Products. https://www.shodan.io/about/products
11. ZoomEye - cyberspace search engine. ZoomEye - Cyberspace Search Engine. https://www.zoomeye.hk/doc
12. LeakIX docs. LeakIX documentation | LeakIX Docs. https://docs.leakix.net/docs/
13. What is Sublist3r and How to Use it? - GeeksforGeeks. GeeksforGeeks. https://www.geeksforgeeks.org/what-is-sublist3r-and-how-to-use-it/
14. The not yet exploited goldmine of OSINT: opportunities, open challenges and future trends / J. Pastor-Galindo et al. IEEE access. 2020. Vol. 8. P. 10282–10304. https://doi.org/10.1109/access.2020.2965257

REFERENCES

1. National Defense Authorization Act for Fiscal Year 2006 : Public Law of 01.06.2006 no. No. 109-163.
https://www.congress.gov/bill/109th-congress/house-bill/1815/text/statute
2. Y. Honcharenko, K. Kanishev. Tools of information warfare: OSINT, IPSO and counteracting disinformation. Information and psychological operation (IPSO). How to avoid becoming a victim of other people's manipulations. [in Ukrainian]
https://infolight.in.ua/wp-content/uploads/2023/02/brochure-2.pdf
3. On the national security of Ukraine : Law of Ukraine No. 2469-VIII of June 21, 2018: as of March 31, 2023. [in Ukrainian]
https://zakon.rada.gov.ua/laws/show/2469-19#Text
4. On the intelligence agencies of Ukraine : Law of Ukraine No. 2331-III of March 22, 2001: as of October 24, 2020. [in Ukrainian]
https://zakon.rada.gov.ua/laws/show/2331-14#Text
5. Williams H., Blum I. Defining second generation open source intelligence (OSINT) for the defense enterprise. RAND Corporation, 2018.
https://doi.org/10.7249/rr1964
6. Unver A. Digital open source intelligence and international security: a primer. EDAM, Oxford CTGA & Kadir Has Üniversitesi, 2018. 28 p.
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3331638
7. Schwarz K., Schwarz F., Creutzburg R. Conception and implementation of professional laboratory exercises in the field of open source intelligence (OSINT). Electronic imaging. 2020. Vol. 2020, no.3. P.278
https://library.imaging.org/admin/apis/public/api/ist/website/downloadArticle/ei/33/3/art00010
8. Analysis of intelligence gathering tools from open sources / A. Karpenko et al. Communication and information systems : Bulletin. Kyiv, 2022. P. 21. [in Ukrainian]
https://www.viti.edu.ua/files/zbk/2022/2022-1.pdf#page=18
9. Duffy M., Pan X., Wilson S. Information reconnaissance by accumulating public information data sources. OALib. 2024. Vol. 11, no. 04. P. 1–25.
https://doi.org/10.4236/oalib.1111463
10. Shodan Products.
https://www.shodan.io/about/products
11. ZoomEye - cyberspace search engine. ZoomEye - Cyberspace Search Engine.
https://www.zoomeye.hk/doc
12. LeakIX docs. LeakIX documentation | LeakIX Docs.
https://docs.leakix.net/docs/
13. What is Sublist3r and How to Use it? - GeeksforGeeks. GeeksforGeeks.
https://www.geeksforgeeks.org/what-is-sublist3r-and-how-to-use-it/
14. The not yet exploited goldmine of OSINT: opportunities, open challenges and future trends / J. Pastor-Galindo et al. IEEE access. 2020. Vol. 8. P. 10282–10304.
https://doi.org/10.1109/access.2020.2965257

Drozd Maria Igorivna *Master's degree candidate, Institute of Special Communications and Information Protection, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine*
e-mail: iammashadrozd@gmail.com
https://orcid.org/0009-0002-9736-8137

Nesterenko Serhiy Dmytrovych *Senior Lecturer, Department of the Institute of Special Communications and Information Protection, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine*
e-mail: 654squad@gmail.com;
https://orcid.org/0000-0003-2097-1122

Analysis of software for the implementation of OSINT in the field of information security

Relevance. The global modern cyberspace is characterized by a rapid increase in risks and threats to important information of government agencies, business and society. In such circumstances, open source intelligence (OSINT) is gaining importance as a tool for monitoring the information space, identifying potential threats and ensuring information security. OSINT software allows you to effectively collect, analyze and interpret data from open sources, including social networks, public databases and web resources. This facilitates timely response to cyber threats, identification of vulnerabilities and decision-making to protect information systems and critical infrastructure of the state's information relations entities.

Objective. To analyze the characteristics and capabilities of modern specialized software with a view to their effective use as open source intelligence (OSINT) tools in the context of identifying potential threats and ensuring information security of subjects of information relations.

Research methods. In the process of writing this article, the author used the methods of technical analysis, comparative and descriptive approach, systematization and classification to study the functionality of OSINT tools, to predict their effectiveness and development prospects.

Results. Based on the analysis, the key characteristics of software solutions such as Maltego, TheHarvester, Shodan, ZoomEye, LeakIX, Sublist3r and SubFinder are identified, their suitability for monitoring the information space, identifying risks and vulnerabilities, as well as timely response to eliminate negative consequences are assessed. Recommendations for the optimal use of these tools on modern computers are proposed, taking into account the requirements for hardware, security and process automation.

Consideration of the applied aspects of OSINT use makes it possible to formulate practical recommendations for cybersecurity professionals. The analysis makes it possible to integrate the results into training programs for information security specialists. It has been established that the effectiveness of OSINT largely depends on the level of user training and his/her ability to interpret the information received. The material reviewed demonstrates the prospects for using machine learning to automate data collection and filtering processes. The author emphasizes the need to continuously update the knowledge bases and algorithms used in OSINT. The results of the study can be used to create integrated solutions to ensure the cyber resilience of organizations.

Conclusions. Open source intelligence (OSINT) is based on the collection, systematization and analysis of data from publicly available sources, such as social networks, websites, public databases and media. The basis of OSINT software is the use of automated tools that allow you to efficiently process large amounts of information, detect connections between data, and identify potential threats to information security. Tools such as Maltego, TheHarvester, Shodan, ZoomEye, LeakIX, Sublist3r, and SubFinder provide tasks ranging from passive data collection to active analysis of network infrastructure, which allows identifying vulnerabilities, monitoring cyberspace, and supporting timely decision-making in the field of information security and information protection.

The author classifies OSINT software by functional purpose, allocating three main categories: tools for detection, extraction and aggregation of data. A comparative analysis of such tools as Maltego, TheHarvester, Shodan, ZoomEye, LeakIX, Sublist3r and SubFinder is proposed, with the definition of their key characteristics, including compatibility with operating systems, methods of information collection, process automation and security level, which helps to choose the optimal tool for solving the problems of monitoring cyberspace and countering information threats.

Promising directions for further development of OSINT software in the field of cybersecurity of the State are presented.

***Keywords**: OSINT, information security, software, cyber threats, data analysis, automation, vulnerabilities.*

Надійшла у першій редакції 09.04.2025, в останній - 15.05.2025.
The first version has been received on 09.04.2025, the final version - on 15.05.2025.

DOI: https://doi.org/10.26565/2304-6201-2025-66-05

УДК (UDC) 004.65

Pugach Mykyta *PhD student, Department of Theoretical and Applied Computer Sciences; V. N. Karazin Kharkiv National University, Svobody Sq 4, Kharkiv, Ukraine,61022*
e-mail: mykyta.pugach@karazin.ua
https://orcid.org/0009-0004-8923-6489

A Systematic Review on Workload Change Detection in Distributed Databases

Distributed Databases became essential part of a large part of nowadays software. It has numerous of advantages including scalability, fault tolerance, high availability, and improved performance. It solves a lot of problems of centralized databases but can also suffer with challenges. One of them is skewed access. Workload in distributed DBMS often changes, such fluctuations can cause ineffective operation of the system. Imagine access to one row of database became 10 times more frequent, or complex requests start operating with the data highly distributed geographically. Such behavior shows that initial data distribution cannot be always efficient enough. And to address this problem adoptive design technics were invented. In this article we review the common steps of adoptive technics and concentrate attention at workload detection and hot data identification.
The purpose of the article is to introduce adoptive design approach of distributed database management systems, review and analyze existing technics and theirs steps, especially workload change detection and hot data identification. The final goal is to compare theses technics and lead out their main concerns.
As a result of this work some existing approaches were analyzed and highlighted their common parts alongside with differences, presented their main issues.
After reviewing all technics, we can see that current solutions cannot give precise results without creating much overhead to the system. Also, there is no approach to giving up-to-date information about hot data without creating overhead. Overhead in such situations is a major issue. In skewed access patterns distributed nodes can become very busy with processing queries and additional computations can lead to worse overall system performance then without adoptive design or even to node outage. So, search for solutions, that give precise and up-to-date results without significant overhead is a big field of future researches.

Key words: *distributed databases, adoptive design technics, hot data identification, workload change detection.*

How to quote: M. Pugach, “A Systematic Review on Workload Change Detection in Distributed Databases”, *Bulletin of V. N. Karazin Kharkiv National University, series Mathematical modelling. Information technology. Automated control systems*, vol. 66, pp. 56-62, 2025. https://doi.org/10.26565/2304-6201-2025-66-05

Як цитувати: Pugach M. A Systematic Review on Workload Change Detection in Distributed Databases. *Вісник Харківського національного університету імені В. Н. Каразіна, серія Математичне моделювання. Інформаційні технології. Автоматизовані системи управління*. 2025. вип.. 66. С.56-62. https://doi.org/10.26565/2304-6201-2025-66-05

Introduction

Workload of distributed database management systems with online transaction processing (OLTP) is not static. It can significantly change due to daily, weekly, or seasonal fluctuations in demand, or because of rapid growth in demand due to a company’s business success. This often causes imbalances in the load on the nodes of a distributed DBMS. Such situations may happen when database holds “hot” tuples or range of tuples, which means this data is much more in demand. For example, music streaming platforms have trendy songs or albums. According to 2023-year statistics [1] there were 463,000 tracks streamed at least a million times and 45.6 million tracks, which had zero streams. But trends come and go, which causes changes in database workloads. Fluctuations may provoke increases in distributed query execution time or even node fails because of excessive load.
To address this issue, some modern distributed DBMSs apply adaptive design approaches. Key point is to perform incremental redesign, which means that data may be dynamically repartitioned during the system's runtime. There are three interrelated issues that need to be tackled in adaptive distribution design [2][3]:

4.

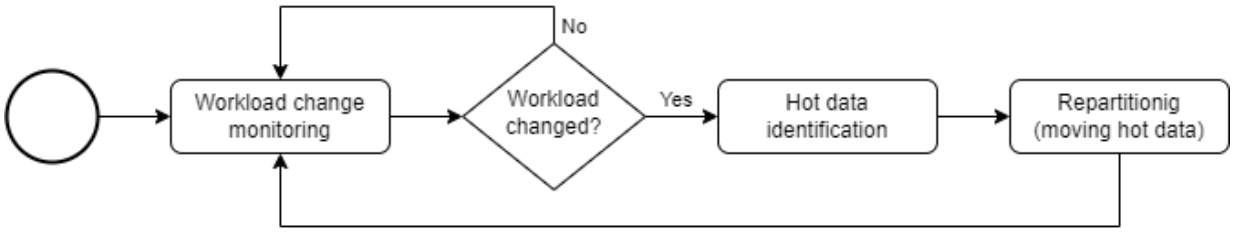


Fig. 1 Adoptive design approach flowchart.
Рис. 1 Схема підходу до проектування на основі адаптивного дизайну

For the last decades, adaptive design has been a topic for active research and quite a few adaptive techniques have been proposed. But most of them assume that the change in the workload is detected and simply focus on the migration problem. Nevertheless, some of these studies presented solutions for the first two problems, while others even addressed all of them. This article investigates existing approaches for workload changes detection and hot processed data identification. Highlighted their main challenges and ways they solve them. And provided analysis of drawbacks of each solution.

1. Why are adaptive design approaches essential?

In nowadays, when the amount of data increasing exponentially and users demand faster and faster access to it, it is very essential for organizations to have systems which will meet these requirements. More and more companies choose to use distributed databases because they have improved scalability and fault tolerance, as they allow data to be stored across multiple nodes and locations, ensuring that the system can grow efficiently and remain operational even in the case of node failures. Additionally, they provide enhanced availability by distributing data across various sites, minimizing the risk of downtime.

So, modern distributed DBMSs demand high throughput, low latency, and continuous availability. However, these requirements become challenging to maintain when data access is skewed, or workload patterns change. Significant shifts in access patterns can cause some nodes in a DBMS cluster to become overloaded while others remain underutilized, reducing performance despite sufficient total resources.

Experiments demonstrate [3] that as the skew increases, system throughput decreases, and latency rises. This imbalance occurs because heavily loaded partitions accumulate longer queues, resulting in higher latencies, while underutilized partitions remain idle, decreasing overall efficiency. Furthermore, CPU utilization becomes uneven, with highly loaded partitions showing significantly higher usage compared to others, amplifying the imbalance, and impacting system performance.

2. Workload change detection

Detecting the workload change can be called the first step of an incremental redesign process. The goal is to understand whether the system needs to be reconfigured. Naturally all the nodes won't have completely same load. Yet they still have near-uniform distribution. Skews in workload usually could be modeled as Zipfian distribution. So, the higher the value of Zipfian's distribution exponent (α) parameter, the higher skew we have. If it is small enough, then we have a skew which does not affect the system that much, that needs to be reconfigured. So, as a result of this phase a system needs to be aware that there is a significant workload change.

We can highlight two different approaches to achieving this. The first one is based on scanning system metrics and the second is counting the number of distributed transactions for each partition.

2.1. E-Store

The first approach was presented in E-Store, an elastic partitioning framework for distributed OLTP DBMSs [3]. E-Store has an E-Monitor component which addresses workload change detection.

In the initial phase, E-Monitor gathers CPU utilization metrics for each partition on the DBMS's nodes at an OS level (where each partition maps to a core). This high-level, coarse-grained data is easy to collect and offers sufficient insights. CPU usage in a main-memory DBMS is a reliable indicator of overall system performance. When E-Monitor polls a node, it collects the current utilization for all partitions on

that node and calculates the moving average over the last 60 seconds. E-Monitor employs two thresholds – high-watermark (e.g., 90%) and low-watermark (e.g., 50%) to decide whether intervention is required. These thresholds, adjustable by the DBA, reflect a balance between system responsiveness and resource use. If either threshold is surpassed, the E-Monitor initiates a more granular monitoring phase at the tuple level.

2.2. Clay

Also, Clay adaptive partitioning approach [4] uses this technique. It works like E-Monitor but examines SLA (service-level agreements).

2.3. Kairos

The system, named Kairos [5], uses analogous approach. Its resource monitor queries the OS and DBMS running on each machine for statistics about CPU, RAM, and disk I/O, buffer pool utilization, and log flushes.

2.4. SWORD

The second approach was presented in SWORD, a scalable workload-aware data partitioning and placement approach for OLTP workloads [6]. SWORD model the workload as a hypergraph, where each hyperedge corresponds to a transaction or a query, and employ hypergraph partitioning algorithms to guide data placement decisions. Such a graph-based approach was presented in the work of Curino et al. [7].

SWORD offers technique that counts transactions. The system monitors the percentage increase in the number of distributed transactions and determines that the changes are significant enough to require reconfiguration if this percentage increase exceeds a defined threshold. More specifically, SWORD sets a task using *min-cut* term. The *min-cut* problem is a standard concept in graph theory, where the goal is to partition a graph (representing the system's data here) into two parts, such that the number of edges (representing data dependencies or connections) crossing the partition is minimized. In SWORD, the min-cut is used to assess how well the data is distributed across partitions. This allows SWORD to dynamically adapt its configuration to maintain optimal performance when the system’s load changes, particularly when access patterns change or when there is a significant increase in the number of transactions. Here threshold is also a system parameter which can be set depending upon the sensitivity of applications to latency.

Thus, SWORD observes the rate of increase in load, allowing it to promptly react to increases in transaction volume, which might indicate the need to reconfigure data placement to maintain system efficiency. This adaptive reconfiguration helps maintain balance between the distributed data and the load on various parts of the system, which is crucial for stable operation under changing conditions.

2.5. Summary

Through the evaluation phase, each method validated its capability to solve the assigned problem. However, both have their pros and cons. The primary advantage of E-Monitor is low overhead, which is achieved by periodicity poling nodes’ CPU and CPU metrics themselves, because system metrics are pretty cheap. On the other hand, SWORD tracks all the transactions, which produces much greater overhead. But this approach is supposed to be more accurate than CPU utilization. Clay in its turn uses SLA which produces not much more overhead than checking CPU but is more precise.

Table 1. Workload change detection approaches summary.
Таблиця 1. Підсумок підходів до виявлення змін робочого навантаження.

	Low overhead	**Precision**
E-store	**+**	**+-**
Clay	**+**	**+**
Kairos	**+**	**+-**
SWORD	**-**	**+**

3. Identify which data to move.

Once workload change is detected, adaptive incremental redesign techniques need to select a bunch of data that should be moved to another partition. As the purpose of the entire process is load balancing, data should meet several requirements. The main criterion is hot processing. Skews in workloads most often happen due to increasing demand for a limited number of tuples. Some adoptive design approaches also measure interconnections between tuples. The ideal case occurs when each transaction “matches” a

partition because the transaction has to access that only partition [8]. So, in this phase system needs to choose the data that will be moved.

In general, all existing approaches could be split into two groups: inline and offline. And these approaches have such common characteristics: inline produces overhead to the system but results with up-to-date information, unlike offline techniques processes data separately from transactions and do not produce overhead but result with a delay.

3.1. Siberia

We will start with offline approaches. The first one was presented by Levandoski et al. [9]. Their project was called Siberia; its goal was hot records classification. This approach, in short, uses logging combined with algorithmic operations to detect active or frequently accessed records.

Let us consider it more precisely. Siberia associates each record with a discrete time slice, denoted $[t_n, t_{n+1}]$. The next time interval begins at t_{n+1} and ends at t_{n+2}, and continues similarly. Time is tracked by the number of record accesses, with each "tick" of the clock occurring after each record access. A time slice is identified using its beginning timestamp (t_n for $[t_n, t_{n+1}]$). A time slice denotes a distinct period during which record access is detected, and conceptually, log records (RecordID, TimeSlice) pairs. In practice, the log keeps a sequence of record IDs in the order of access, separated by time markers that indicate the boundaries of each time slice.

Siberia uses exponential smoothing to estimate record access frequencies. Exponential smoothing calculates an access frequency estimate for a record *r* as

$$est_r(t_n) = \alpha * x_{t_n} + (1 - \alpha) * est_r(t_{n-1})$$

Where, t_n denotes the current time slice, while x_{t_n} represents the observed value at t_n. In this model, x_{t_n} is set to 1 if the record r is accessed during t_n; otherwise, it is 0. The term $est_r(t_{n-1})$ refers to the estimate derived from the preceding time slice, t_{n-1}. The parameter α serves as a decay factor, controlling the influence of new observations and the rate at which past estimates lose significance. Typically, α is chosen in the range of 0.01 to 0.05, where higher values prioritize recent observations more heavily.

The authors claim that Siberia does not use every record access for its algorithms, because it may degrade system performance. To minimize system overhead, they adopt a sampling-based approach for logging. Each worker thread determines whether to log its activity by flipping a biased coin, with the bias corresponding to the sampling rate. Depending on the result of the coin flip, the thread either records its data in log buffers or skips the logging process. The authors state that sampling only 10% of the accesses reduces the accuracy by only 2.5%.

There were four algorithms presented, all of them take the same data as input and result the same but there are significant differences between them in performance. They take stored logs, described before, and parameter that signifies the number of records to classify as hot. Authors propose two basic algorithms: forward and backward. The forward algorithm simply scans the log forward from a beginning time slice. It updates *r*'s current access frequency estimate using the exponential smoothing equation. This forward algorithm has two primary drawbacks: it requires a scan of the entire log, and it requires storage proportional to the number of unique record ids in the access log.

Backward algorithm avoids scanning the entire log from beginning to end in order to improve classification performance. The primary concept involves scanning the log in reverse order and derive successively tighter upper and lower bounds for the estimates of accessed records. Occasionally, the algorithm classifies records based on these bounds, allowing it to potentially stop the scan earlier. It maintains estimates only for records that are still contenders for the hot set, thereby limiting its memory usage to the number of hot records rather than the total record count.

Backward algorithm also uses exponential smoothing:

$$estb_r(t_n) = \alpha(1 - \alpha)^{(t_e - t_n)} + estb_r(t_{last})$$

where $t_{last} > t_n$ since scanning in reverse.

Calculates upper bound and lower bound:

$$upEst_r(t_n) = estb_r(t_n) + (1 - \alpha)^{t_e - t_n + 1}$$
$$loEst_r(t_n) = estb_r(t_n) + (1 - \alpha)^{t_e - t_b + 1}$$

where t_e means end time slice and t_b – the first time slice.

As the backward classification approach continues processing more record accesses, the upper and lower bounds converge toward an exact estimate.

Authors use two optimizations for this algorithm. First, dropping records with upper bound values less than k^{th} record lower bound value, where *k* is parameter, which determines needed size of resulting hot set. This optimization origins from the properties of upper and lower bounds. Record would never have an estimation bigger than upper bound and less than lower bound. Second, algorithm can stop when only *k* records are still in contention for the hot set. So, the backward approach is much more efficient than the forward.

The researchers also propose parallel variants of these two algorithms. These solutions are more efficient, as expected. Backward algorithm split logs into *N* parts and use controller-worker processes scheme to calculate hot records set. As parallel backward algorithm is the most efficient, Siberia uses it.

3.2. SWORD

We have already discussed how SWORD [6] addresses workload detection. And mentioned that it uses hypergraph representation model for workload, where hyperedges are transactions or queries. SWORDs technique of 'data to move' recognition is inextricably connected with its repartitioning method. Its approach is based on efficiently identifying candidate sets of data items whose migration has the potential to reduce the frequency of distributed transactions the most and then performing the migrations during periods of low load.

More specifically, authors use the term *virtual node* which is a logical abstraction of physical node. One physical node can contain few virtual nodes. The initial distributed database design defines such nodes in such a way, that data inside them are highly interconnected and data between these nodes are loosely connected. Using graph-based interpretation system uses the *min-cut*. Data identification step starts once min-cut crosses the threshold. The algorithm manages pairs of candidate virtual node sets that can be exchanged to decrease the min-cut size. It performs some of such swaps per step, aiming to minimize the *min-cut* of the data placement based on the current workload. This process continues until the *min-cut* falls below the specified threshold.

So, here the data identification process united with repartitioning plan. The system selects data according to number of distributed transactions. Such an approach can be less efficient than hot data identification, because the number of hot repartitions is less than others. Thus, the system has to make more moves, which provides significantly more overhead.

3.3 Apollo

The framework called Apollo [10] uses graph-based workload visualization, too. But unlike SWORD it utilizes query patterns instead of queries. So, in place of such a request:

```
SELECT C_ID FROM CUSTOMER WHERE C_UNAME = 'Bob' AND C_PASSWD = 'pwd'
```

The system will save such query pattern:

```
SELECT C_ID FROM CUSTOMER WHERE C_UNAME = ? AND C_PASSWD = ?
```

While this reduces the granularity of determining the exact set of data items that are affected, it may allow the detection of additional data items that might be affected by similar queries and reduce the frequency of changes that are necessary.

3.4. SAHARA

Project, called SAHARA [11], introduces an inline approach of hot data identification. It focuses on two key types of data access: domain access and row access. As SAHARA works with column stores, domain means a set of values for a particular attribute. Domain accesses are recorded to evaluate potential partitioning layouts, while row accesses are captured to estimate the memory footprint of the partitioning. This data is collected over defined time windows to avoid biases from short-term access bursts, ensuring that the statistics reflect long-term patterns relevant to memory management strategies, such as buffer pool eviction policies.

To manage memory efficiency, accesses are recorded in blocks rather than individually, which reduces the memory overhead but may introduce imprecision in access frequency measurements. The block sizes are adjusted experimentally to balance accuracy with memory usage, with a maximum overhead of 1% of the dataset size. Furthermore, the methodology incorporates row and domain block counters to track accesses to specific data partitions and attribute values within given time frames. This approach allows for a more granular understanding of data access patterns. As statistics is gathered, SAHARA classifies hot data using π-second-rule. It means that data is hot if it is accessed more often than every π-seconds,

where π is a settings parameter. Authors emphasize that the time window length should not be set substantially smaller than π. In addition, the Nyquist–Shannon sampling theorem proves that a sample rate of $\pi/2$ is sufficient to achieve precise statistics. Therefore, the time window length is set to $\pi/2$.

3.5. E-Store

Taft et al. [3] in their research suggested approach based on time windows, too. The system called E-Store makes tuple-level monitoring on the entire cluster for a short period of time. Authors assume all non-replicated tables of an OLTP database form a tree-schema based on foreign key relationships. Although this rules out graph-structured schemas and m-n relationships, it applies to many real-world OLTP applications, they claim. Considering such assumption, monitoring only the root tuples provides a good approximation of system activity and minimizes the overhead of this phase.

E-Store's monitoring system, called E-Monitor classifies hot tuples as the top-k most frequently accessed tuples within a given time window. A tuple is considered "accessed" if it is read, modified, or inserted during a transaction. The system collects two types of data: the total number of accesses to tuples within a partition (*L*) and the subset of top-k most frequently accessed tuples (*TK*). When tuple-level monitoring is activated, the database management system (DBMS) sets up an internal histogram for each partition, which tracks how many times each tuple has been accessed by a transaction. After the time window concludes, the execution engine at each node compiles *L* and *TK* for its local partitions and forwards this data to E-Monitor. E-Monitor consolidates the information from all partitions to create a global top-k list.

3.6. Summary

We reviewed several technics of hot data identification. They use completely different approaches, like gathering query logs or collecting tuple level statistics. All of them has their advantages along with disadvantages. Processing all computations on another CPU, like in Siberia, lowers overhead nicely but with that system receives not up-to-date information. E-Store's tuple level monitoring provides actual information but can produce critical overhead in the busiest nodes. Graph-based solutions frequently are not enough precise and universal.

4. Conclusion

With the rise of amount of information and demand of quick access to it, distributed databases became the essential technology. But often initial data allocation becomes ineffective and reduces all advantages of distributed DBs. Here comes adoptive design approaches. They allow dynamic data reconfiguration to keep the system highly efficient.

We discussed that adoptive design technics mostly contain three steps and reviewed the first two of them: workload change detection and hot data identification. Several DBMSs implemented these steps, and we can see that there are numerous of approaches how to do that. In general, adoptive techniques are balancing between high efficiency and high precision, low overhead and timeliness. Overhead in such situations is a major issue. In skewed access patterns distributed nodes can become very busy with processing queries and additional computations can lead to worse overall system performance then without adoptive design or even to node outage. So, it's crucial to keep developing these approaches to meet all these characteristics at once to create more efficient distributed database management systems.

REFERENCES

1. Luminate Data, LLC, "Year-End Music Industry Report 2023," Luminate Data, LLC, 2023. [Online]. Available: https://luminatedata.com/reports/yearend-music-industry-report-2023/. [Accessed: Nov. 27, 2024]
2. M. T. Özsu and P. Valduriez, *Principles of Distributed Database Systems. 4th edition*. Cham, Switzerland: Springer Nature, 2020.
3. R. Taft et al., "E-Store: Fine-grained elastic partitioning for distributed transaction processing systems", *Proceedings of the VLDB Endowment*, vol. 8, no. 3, pp. 245 – 256, 2014. https://doi.org/10.14778/2735508.2735514.
4. M. Serafini, R. Taft, A. J. Elmore, A. Pavlo, A. Aboulnaga and M. Stonebraker, "Clay: Fine-grained adaptive partitioning for general database schemas", *Proceedings of the VLDB Endowment*, vol. 10, no. 4, pp. 445 – 456, 2016. https://doi.org/10.14778/3025111.3025125.

5. C. Curino, E. P. C. Jones, S. Madden and H. Balakrishnan, “Workload-aware database monitoring and consolidation”, in *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data.* Athens, 2011, pp. 313 – 324. https://doi.org/10.1145/1989323.1989357.
6. A. Quamar, K. A. Kumar and A. Deshpande, “SWORD: Scalable workload-aware data placement for transactional workloads”, in *Proceedings of the 16th International Conference on Extending Database Technology*. Genoa, 2013, pp. 430 – 441. https://doi.org/10.1145/2452376.2452427.
7. C. Curino, E. Jones, Y. Zhang and S. Madden, “Schism: A workload-driven approach to database replication and partitioning”, *Proceedings of the VLDB Endowment*, vol. 3, no. 1-2, pp. 48 – 57, 2010. https://doi.org/10.14778/1920841.1920853.
8. S. Navathe, S. Ceri, G. Wiederhold and J. Dou, “Vertical partitioning algorithms for database design”, *ACM Transactions on Database Systems*, vol. 9, no. 4, pp. 680 – 710, 1984. https://doi.org/10.1145/1994.2209.
9. J. J. Levandoski, P.- Å. Larson and R. Stoica, “Identifying hot and cold data in main-memory databases” in *Proceedings of the 2013 IEEE 29th International Conference on Data Engineering (ICDE)*. Brisbane, 2013, pp. 26 – 37. https://doi.org/10.1109/ICDE.2013.6544811.
10. B. Glasbergen, M. Abebe, K. Daudjee, S. Foggo and A. Pacaci, “Apollo: Learning query correlations for predictive caching in geo-distributed systems” in *Proceedings of the 21st International Conference on Extending Database Technology (EDBT)*. Vienna, 2018, pp. 253 – 264. https://doi.org/10.5441/002/edbt.2018.23.
11. M. Brendle, N. Weber, M. Valiyev, N. May, R. Schulze, A. Böhm and G. Moerkotte, “SAHARA: Memory footprint reduction of cloud databases with automated table partitioning” in *Proceedings of the 25th International Conference on Extending Database Technology (EDBT)*. Edinburgh, 2022, pp. 13 – 26. https://doi.org/10.5441/002/edbt.2022.02.

Пугач Микита Сергійович *Аспірант, Кафедра теоретичної та прикладної інформатики Харківського національного університету імені В.Н. Каразіна, майдан. Свободи 4, Харків, Україна, 61022*
e-mail: mykyta.pugach@karazin.ua
https://orcid.org/0009-0004-8923-6489

Систематичний огляд на виявлення змін робочого навантаження в розподілених базах даних

Розподілені бази даних стали важливою частиною значної частини сучасного програмного забезпечення. Вони мають численні переваги, включаючи масштабованість, відмовостійкість, високу доступність і покращену продуктивність. Це вирішує багато проблем централізованих баз даних, але також можуть мати проблеми. Одна з них – нерівномірний доступ до даних. Робоче навантаження в розподілених СУБД часто змінюється, такі коливання можуть стати причиною неефективної роботи системи. Уявіть, що доступ до одного рядка бази даних став у 10 разів частішим, або складні запити починають працювати з даними, розподіленими територіально. Така поведінка свідчить про те, що первинний розподіл даних не завжди може бути достатньо ефективним. І для вирішення цієї проблеми були винайдені технології адаптивного дизайну. У цій статті ми розглядаємо загальні кроки адаптивних технологій і зосереджуємо увагу на виявленні робочого навантаження та ідентифікації гарячих даних.

Метою статті є огляд адаптивного підходу до проектування розподілених систем керування базами даних, огляд і аналіз існуючих реалізацій та їхніх кроків, особливо виявлення зміни робочого навантаження та ідентифікації гарячих даних. Кінцева мета полягає в тому, щоб порівняти ці техніки та виявити їх основні проблеми.

У результаті цієї роботи було проаналізовано деякі існуючі підходи та виділено їх спільні сторони та відмінності, представлено їх основні проблеми.

Після перегляду всіх технологій ми можемо побачити, що поточні рішення не можуть дати точних результатів, не створюючи значних накладних витрат на систему. Крім того, немає підходу до надання актуальної інформації про гарячі дані без створення накладних витрат. Накладні витрати в таких ситуаціях є серйозною проблемою. У шаблонах нерівномірного доступу розподілені вузли можуть бути дуже зайняті обробкою запитів, а додаткові обчислення можуть призвести до більшого погіршення загальної продуктивності системи, ніж коли адаптивний підхід не використовується, або навіть до збою вузла. Таким чином, пошук рішень, які дають точні та своєчасні результати без значних накладних витрат, є великим полем для майбутніх досліджень.

Ключові слова: *розподілені бази даних, адаптивна підходи до проектування, ідентифікація гарячих даних, виявлення зміни робочого навантаження.*

Надійшла у першій редакції 02.04.2025, в останній - 14.05.2025.
The first version has been received on 02.04.2025, the final version - on 14.05.2025.

DOI: https://doi.org/10.26565/2304-6201-2025-66-06

УДК (UDC) 004.056

Семеренська Вікторія Владиславівна
кафедри Автоматизації та проектування обчислювальної техніки
Харківський національний університету' радіоелектроніки
61166, проспект Науки, 14, Харків, Україна
e-mail: vsemerenskaya@gmail.com
https://orcid.org/0009-0008-2955-3676

Безпека медичних кіберфізичних систем

Актуальність. Медичні кіберфізичні системи (CPS), зокрема пристрої Інтернету медичних речей (IoMT) для моніторингу, діагностики та терапії в реальному часі, стали невід'ємною частиною цифровізації охорони здоров'я. Поєднання операційних технологій з традиційними IT-системами розширює поверхню атак, роблячи лікарні та телемедичні інфраструктури привабливими цілями для кіберзловмисників. В умовах гібридних конфліктів ризики зростають, оскільки атаки на медичні мережі можуть призвести не лише до витоку даних, а й до прямої шкоди пацієнтам і порушення критичних процесів лікування.
Мета. Метою дослідження є класифікація та аналіз основних типів загроз і вразливостей, що впливають на медичні CPS в умовах гібридних конфліктів, узагальнення існуючих стратегій захисту та формування пропозицій щодо підвищення їхньої кіберстійкості через нормативні, організаційні та технологічні заходи.
Методи дослідження. У роботі застосовано методологію PRISMA для аналізу публікацій, індексованих у базах Scopus, IEEE Xplore і PubMed. Використано порівняльний та аналітичний підходи для узагальнення висновків із нещодавніх інцидентів, зокрема атак типу WannaCry на Національну службу охорони здоров'я Великої Британії, витоку даних SingHealth у Сінгапурі та інших масштабних порушень безпеки в медичній сфері.
Результати. Аналіз показав поширеність таких загроз, як ransomware, DDoS-атаки та компрометація IoMT через незахищені протоколи зв'язку та застаріле програмне забезпечення. Серед ключових проблем — слабка автентифікація, недостатня сегментація мереж і вплив людського фактора. До ефективних заходів протидії віднесено багатофакторну автентифікацію, блокчейн-контроль цілісності даних, наскрізне шифрування та архітектуру Cybersecurity Mesh (CSMA). Наголошено на важливості впровадження квантово-стійкого шифрування та AI-систем адаптивного захисту, здатних автономно виявляти та реагувати на динамічні загрози.
Висновки. Попри досягнення у сфері безпеки медичних пристроїв, рівень стійкості CPS до гібридних загроз залишається недостатнім. Ключовими напрямами зміцнення безпеки є впровадження принципу security-by-design, дотримання міжнародних стандартів кібербезпеки (ISO/IEC 80001, IEC 62443) і розроблення спеціалізованих програм підготовки медичного персоналу. Інтеграція AI-орієнтованої ситуаційної обізнаності, гармонізація регуляторних вимог і співпраця між державним і приватним секторами сприятимуть підвищенню надійності та довіри до цифрової екосистеми охорони здоров'я.

Ключові слова: *кібербезпека, медичні технології, захист даних, безпека медичних систем, вразливості IoMT, гібридні загрози, архітектура Cybersecurity Mesh*

Як цитувати: Семеренська В. В. Безпека медичних кіберфізичних систем. *Вісник Харківського національного університету імені В. Н. Каразіна, серія Математичне моделювання. Інформаційні технології. Автоматизовані системи управління*. 2025. вип. 66. С.63-72. https://doi.org/10.26565/2304-6201-2025-66-06

How to quote: V. Semerenska, "Security of medical cyber-physical systems", *Bulletin of V. N. Karazin Kharkiv National University, series Mathematical modelling. Information technology. Automated control systems*, vol. 66, pp. 63-72, 2025. https://doi.org/10.26565/2304-6201-2025-66-06

Вступ

Сучасна медицина активно інтегрує кіберфізичні системи (CPS), які поєднують апаратне забезпечення, програмні платформи та мережеві технології для підтримки життєво важливих процесів у медичній практиці. Системи моніторингу, автоматизовані інфузійні насоси, кардіостимулятори, роботи-хірурги та інші IoMT-пристрої (Internet of Medical Things) відіграють критично важливу роль у забезпеченні точності, швидкості та ефективності медичних послуг. Проте стрімка цифровізація охорони здоров'я відкриває нові можливості для кібератак, які можуть призводити не лише до компрометації конфіденційних даних, але й до фізичного ризику для пацієнтів.

Особливо тривожним є те, що кібератаки на медичну інфраструктуру все частіше стають елементом гібридних військових операцій. У таких випадках метою атак є не лише економічна шкода або дестабілізація роботи окремих лікарень, але й руйнування довіри до системи охорони здоров'я загалом. Учасники гібридних конфліктів застосовують технології для порушення роботи

критичної інфраструктури, включаючи медичну, щоб створити хаос і поглибити гуманітарну кризу. Подібні дії порушують міжнародне гуманітарне право, але залишаються актуальними через складність їхнього відстеження та попередження.

Вразливість медичних CPS зумовлена кількома ключовими чинниками. По-перше, багато медичних закладів використовують застаріле обладнання, яке не підтримує сучасні стандарти кібербезпеки. Програмне забезпечення для таких систем часто не оновлюється, що створює можливості для атак. По-друге, IoMT-пристрої, такі як носимі медичні гаджети, часто мають обмежені ресурси для реалізації сучасних методів захисту, що робить їх легкими цілями для зловмисників. По-третє, людський фактор залишається значним джерелом ризику. Медичний персонал часто недостатньо обізнаний щодо основних принципів кібербезпеки, а також не підготовлений до роботи в умовах цілеспрямованих атак. Нарешті, відсутність уніфікованих стандартів безпеки для IoMT ускладнює впровадження ефективних рішень, оскільки різні пристрої мають різні рівні захисту.

Загрози, що виникають унаслідок атак на медичні CPS, мають безпрецедентний вплив на суспільство. Збої у роботі лікарень, переривання лікувальних процедур, витік персональних даних пацієнтів – усе це підриває стабільність медичних систем. Наприклад, атака WannaCry у 2017 році паралізувала діяльність Національної служби охорони здоров'я Великобританії, спричинивши зрив тисяч медичних процедур. У сучасних умовах такі інциденти можуть бути не випадковими, а цілеспрямованими, що особливо актуально для країн, які знаходяться в умовах військових конфліктів.

Ця стаття має на меті систематично дослідити загрози, вразливості та сучасні методи забезпечення кібербезпеки медичних CPS. Зокрема, розглянуто питання захисту IoMT-пристроїв, методи виявлення атак у реальному часі та можливості інтеграції інноваційних рішень, таких як архітектура CSMA, квантово-стійке шифрування та когнітивні системи захисту.

Результати цього дослідження спрямовані на розробку рекомендацій для підвищення стійкості медичних CPS до сучасних загроз, особливо в умовах гібридних конфліктів. Також робота ставить за мету сформувати базу для подальших досліджень, спрямованих на інтеграцію кібербезпеки на етапі проєктування пристроїв і систем.

1. Матеріали та методи

Для цього дослідження було проведено систематичний огляд літератури із залученням таких баз даних, як Scopus, IEEE Xplore та PubMed, що забезпечило репрезентативний набір джерел. Огляд охоплював наукові публікації, опубліковані в період із 2015 по 2024 рік, які стосуються безпеки медичних кіберфізичних систем (CPS), зокрема IoMT, загроз, методів захисту та вразливостей. Аналіз отриманих робіт дозволив виявити ключові тенденції, розподілити дослідження за технічними, організаційними та правовими аспектами, а також оцінити ефективність запропонованих рішень.

Тематичний аналіз літератури показав, що основними технічними проблемами є ризики, пов’язані із застарілим обладнанням, слабким шифруванням даних та вразливостями IoMT-пристроїв. Наприклад, дослідження Bhushan та ін. (2023) демонструє, що недостатність обчислювальних ресурсів IoMT робить ці пристрої привабливими цілями для зловмисників. Щодо атак на конфіденційність, робота Ghubaish та ін. (2020) підкреслює загрозу ransomware, яка паралізує роботу медичних закладів та створює додаткові ризики для пацієнтів. У сфері правових аспектів, дослідження Almaiman та Alqahtani (2021) виявило, що національні політики з кібербезпеки часто не враховують специфіку медичних пристроїв, що ускладнює їхню регуляцію.

Розглядаючи технічні рішення, дослідження Wang та ін. (2020) демонструє ефективність блокчейну у забезпеченні цілісності медичних записів, що дозволяє мінімізувати ризик їх підробки. Водночас організаційні аспекти, зокрема підготовка медичного персоналу, висвітлені в роботі Longi та ін. (2024), наголошують на важливості навчання співробітників для підвищення їхньої обізнаності щодо кіберзагроз. Правові аспекти та необхідність міжнародних стандартів для IoMT підкреслені в дослідженні Mathkor та ін. (2024), яке також акцентує увагу на можливості уникнення правових конфліктів між країнами.

Порівнюючи запропоновані рішення, дослідження Ameen та ін. (2024) доводить, що блокчейн значно знижує ризик атак на цілісність даних. У той час як квантово-стійке шифрування, як описано у Heidari та ін. (2019), забезпечує довготривалий захист навіть у контексті потенційних

загроз від квантових обчислень. Методи машинного навчання, згадані в Reji та ін. (2023), виявляють аномалії в роботі систем у реальному часі, що мінімізує вплив людського фактора.

Результати огляду літератури свідчать, що основними викликами є відсутність стандартизованих підходів до безпеки IoMT, обмеження ресурсів для впровадження сучасних методів захисту та необхідність вдосконалення нормативно-правової бази. Разом із тим інтеграція блокчейну, квантово-стійкого шифрування та машинного навчання є перспективними напрямами для подальшого розвитку кібербезпеки медичних CPS.

Медичні кіберфізичні системи (CPS) піддаються широкому спектру загроз, які впливають на їхню конфіденційність, доступність та цілісність. Серед них особливу небезпеку становлять цілеспрямовані атаки (APT), атаки типу ransomware і DDoS, а також використання соціальної інженерії для компрометації персоналу.

Advanced Persistent Threats (APT) – це складні та довготривалі атаки, спрямовані на отримання контролю над критичними системами. Зловмисники застосовують комбінацію методів, включаючи фішингові листи, експлойти у програмному забезпеченні та техніки прихованого руху всередині мережі. Метою APT є тривалий доступ до системи для збору даних або порушення її функціонування. Наприклад, компрометація серверів зберігання медичних записів може призвести до витоку конфіденційної інформації про пацієнтів, яка може бути використана для шантажу або незаконного продажу (Bhushan та ін., 2023).

Вразливість таких атак часто пов'язана з недостатнім сегментуванням мереж, що дозволяє зловмисникам розширювати свій доступ у межах системи, і з відсутністю регулярного моніторингу мережевих аномалій.

Ransomware-атаки блокують доступ до медичних систем, вимагаючи викуп за відновлення функціональності. Ці атаки особливо небезпечні у сфері медицини, оскільки порушення доступу до електронних медичних записів чи обладнання може спричинити невідкладну загрозу для пацієнтів. Наприклад, під час атаки WannaCry на системи NHS було заблоковано доступ до 70 000 пристроїв, включаючи МРТ-сканери, що призвело до зриву критично важливих медичних процедур (Ghubaish та ін., 2020).

DDoS-атаки (Distributed Denial of Service) орієнтовані на перевантаження системи великою кількістю запитів, що робить її недоступною для звичайних користувачів. У медичному контексті такі атаки можуть паралізувати роботу лікарень, викликаючи затримки у наданні допомоги. Наприклад, атака на систему охорони здоров’я Коста-Ріки у 2022 році заблокувала доступ до електронних записів пацієнтів, що призвело до значних порушень у наданні медичних послуг (Almaiman & Alqahtani, 2021).

Соціальна інженерія спрямована на компрометацію персоналу шляхом маніпуляцій, що призводять до розкриття конфіденційної інформації або виконання небезпечних дій, таких як відкриття шкідливих посилань. Цей метод є ефективним у медичному секторі через недостатню обізнаність працівників про кіберзагрози та високу завантаженість, що сприяє помилкам. Наприклад, фішингові кампанії, спрямовані на адміністративний персонал лікарень, дозволяють отримати доступ до внутрішніх систем або електронної пошти. Дослідження показують, що 88% успішних кібератак на медичні установи включають елементи соціальної інженерії (Wang та ін., 2020).

Таким чином, медичні CPS стикаються із загрозами, які потребують багаторівневих стратегій захисту, включаючи моніторинг аномалій, сегментацію мережі та навчання персоналу. Актуальність цих викликів зростає через зростаючу залежність медичних закладів від цифрових систем.

2. Вразливості

Основними вразливостями медичних систем, які посилюють ризик кібератак, є застаріле програмне забезпечення, незахищені пристрої IoT та людський фактор.

Значна частина медичних CPS функціонує на застарілому програмному забезпеченні, яке більше не підтримується розробниками та не отримує оновлень безпеки. Як зазначено в роботі Wang та ін. (2020), це створює сприятливі умови для атак, спрямованих на використання відомих вразливостей. Наприклад, під час атаки WannaCry було скомпрометовано системи, які працювали на старих версіях Windows, що не мали необхідних патчів для захисту від експлойтів EternalBlue. Відсутність регулярного оновлення системних компонентів і залежність від устарілих технологій підвищують ризики не лише для конфіденційності, але й для функціональної безпеки систем.

Інтернет медичних речей (IoMT) включає пристрої, такі як носимі датчики, інфузійні насоси та монітори життєвих функцій, які підключені до мережі та забезпечують обмін даними між пацієнтами і лікарями. Однак, як наголошено у роботі Almaiman та Alqahtani (2021), більшість IoMT-пристроїв мають обмежені обчислювальні ресурси, що ускладнює впровадження сучасних механізмів шифрування та безпеки. Це робить їх легкою ціллю для атак типу Man-in-the-Middle або зловмисного перехоплення даних. Наприклад, деякі інфузійні насоси можуть бути віддалено зламані через відсутність захищених каналів зв'язку, що дозволяє змінювати дози ліків.

Людський фактор є одним із ключових джерел уразливостей медичних CPS. Як зазначено у роботі Longi та ін. (2024), недостатня обізнаність медичного персоналу щодо кіберзагроз призводить до високої ефективності атак соціальної інженерії, таких як фішингові кампанії. Крім того, перевантаженість роботою та брак часу для належної перевірки підозрілих дій сприяють помилкам, які можуть надати зловмисникам доступ до внутрішніх систем. Навіть базові заходи, такі як використання багатофакторної автентифікації, часто ігноруються через відсутність належного навчання персоналу.

Складна архітектура медичних CPS, яка включає численні підсистеми, сервери, бази даних і зовнішні пристрої, створює додаткові вразливості. Відсутність чіткої сегментації мережі дозволяє зловмисникам, отримавши доступ до одного компонента системи, поступово поширюватися на інші. Ameen та ін. (2024) зазначають, що такі архітектурні уразливості особливо характерні для старих лікарняних систем, які інтегруються з новими IoMT-пристроями без належного оновлення протоколів захисту.

Таким чином, вразливості медичних CPS є наслідком технічних обмежень, недостатньої модернізації та людських помилок. Їхнє усунення потребує комплексного підходу, що включає модернізацію систем, впровадження сучасних механізмів безпеки та навчання персоналу для зменшення впливу людського фактора.

Атака WannaCry на NHS (2017)

Атака WannaCry стала однією з найбільш руйнівних у сфері охорони здоров'я, використовуючи експлойт EternalBlue, який експлуатував вразливість у протоколі SMBv1 (Server Message Block). Після проникнення у систему шкідливе програмне забезпечення зашифровувало файли, використовуючи алгоритм AES-128, і вимагало викуп у біткоїнах для розшифрування. Для поширення вірусу використовувався механізм саморозмноження, що дозволяло йому швидко інфікувати інші пристрої в локальній мережі.

У системах NHS уразливості виникли через використання старих операційних систем, таких як Windows XP, які більше не отримували оновлень безпеки. Брак сегментації мережі дозволив вірусу миттєво поширитися між комп'ютерами, серверами та медичним обладнанням, включаючи МРТ-сканери. Зламані пристрої відключилися, що призвело до перенаправлення пацієнтів і скасування тисяч прийомів. Пізніше дослідження показали, що недостатній рівень сегментації мережі став ключовим фактором, який посприяв масштабуванню атаки.

Витік даних у SingHealth (2018)

Витік даних у SingHealth був результатом складної атаки типу APT. Зловмисники спочатку використовували фішингові листи для компрометації облікових записів адміністраторів. Після отримання доступу до внутрішньої мережі вони використали експлойти для підвищення привілеїв, що дозволило їм отримати адміністративний доступ до бази даних пацієнтів.

Ключовою технічною особливістю цієї атаки була експлуатація недостатньо захищених API, які забезпечували інтеграцію між базами даних і медичними додатками. Зловмисники змогли завантажити великі обсяги інформації, не викликавши підозр у системах моніторингу. Уразливості також включали відсутність багаторівневої автентифікації для адміністраторів баз даних, що дозволило використати лише один скомпрометований обліковий запис для доступу до всієї інформації.

Ransomware-атака на медичні установи Коста-Ріки (2022)

Атака Hive Ransomware на медичні установи Коста-Ріки розпочалася з фішингових листів, які містили шкідливі вкладення. Після відкриття шкідливого файлу вірус отримав доступ до внутрішньої мережі і поширився на сервери, що зберігали електронні медичні записи (EMR). Використовуючи комбіноване шифрування RSA-2048 та AES-256, Hive заблокував доступ до даних, включаючи історію пацієнтів та результати аналізів.

Інфраструктура медичних установ виявилася вразливою через відсутність ізоляції серверів EMR, що дозволило вірусу поширитися на всі основні системи. Після шифрування даних

зловмисники залишили в системі запис із вимогою викупу, який можна було прочитати через командний рядок заражених пристроїв. Відсутність резервного копіювання на рівні даних та серверів унеможливила швидке відновлення інформації.

Атака на лабораторії Synnovis у Лондоні (2024)

Цей інцидент став результатом цілеспрямованої атаки на лабораторні інформаційні системи, які керували передачею та обробкою медичних даних. Нападники використали вразливість у неавтентифікованих API, які забезпечували інтеграцію між лабораторними пристроями та сервером. Впроваджений шкідливий код призвів до припинення передачі даних між лабораторними пристроями та основним сервером, заблокувавши доступ до результатів тестів у лікарнях.

Особливістю цієї атаки було використання прихованого механізму завантаження шкідливого коду через підроблені запити до API. Через відсутність шифрування та обмеження доступу за IP-адресами, нападники змогли впровадити код, який блокує взаємодію між системами. Крім того, брак резервних каналів передачі даних спричинив затримки у виконанні критичних аналізів, що негативно вплинуло на медичну допомогу тисячам пацієнтів.

Таким чином, кожен із розглянутих інцидентів демонструє специфічні технічні уразливості, які зловмисники використовували для досягнення своїх цілей, і підкреслює необхідність посилення кіберзахисту медичних CPS.

3. Існуючі підходи до безпеки медичних систем

Медичні системи вимагають багаторівневого підходу до забезпечення кібербезпеки. Розглянемо сучасні технічні, організаційні та регуляторні стратегії, спрямовані на захист цих систем та технічну реалізацію кожного підходу.

Шифрування даних є основним способом захисту інформації в медичних CPS. Алгоритми AES (Advanced Encryption Standard) та RSA використовуються для шифрування переданих і збережених даних. Наприклад, дослідження Bhushan та ін. (2023) рекомендує використовувати квантово-стійке шифрування для довготривалого захисту даних у медичних системах, що особливо актуально в умовах майбутнього розвитку квантових обчислень.

HIPAA вимагає використання надійного шифрування для захисту електронних медичних записів (EMR) під час їхнього передавання та зберігання, забезпечуючи відповідність вимогам конфіденційності ("Understanding HIPAA Requirements", 2020).

Багатофакторна автентифікація (MFA) забезпечує додатковий рівень безпеки, вимагаючи від користувача надання двох або більше способів ідентифікації (пароль, біометричні дані, SMS-код). Наприклад, використання біометрії, такої як сканування відбитків пальців, дозволяє захистити облікові записи навіть у випадках компрометації паролів. Технічна реалізація включає інтеграцію MFA-сервісів із внутрішніми системами лікарень через API.

Штучний інтелект (AI) та машинне навчання (ML) відіграють ключову роль у виявленні аномалій у поведінці пристроїв. Наприклад, Reji та ін. (2023) продемонстрували використання алгоритмів кластеризації для виявлення нетипової активності IoMT-пристроїв, таких як надмірна передача даних або підключення до незвичних IP-адрес. Технічно це реалізується шляхом інтеграції ML-моделей у системи моніторингу мережі, що дозволяє автоматично реагувати на підозрілу активність.

Блокчейн забезпечує незмінність і прозорість медичних даних. У роботі Ameen та ін. (2024) представлено технічну архітектуру системи, яка інтегрує блокчейн із IoMT. Вона передбачає створення захищених транзакцій для кожної взаємодії з медичними записами, що дозволяє відстежувати будь-які зміни. Технічна реалізація включає використання смарт-контрактів, які автоматично перевіряють автентичність транзакцій у мережі.

Навчання персоналу є ключовим елементом запобігання атакам соціальної інженерії. Longi та ін. (2024) рекомендують впроваджувати регулярні тренінги, які охоплюють фішингові атаки, управління паролями та використання багатофакторної автентифікації. Програми навчання включають практичні симуляції атак для підвищення обізнаності співробітників.

Створення планів реагування на інциденти (IRP) дозволяє мінімізувати вплив атак на медичні CPS. Як зазначено в роботі Almaiman та Alqahtani (2021), ефективний IRP включає системи резервного копіювання, ізоляцію скомпрометованих сегментів мережі та процедури відновлення даних. Технічно реалізація передбачає впровадження централізованих систем моніторингу з можливістю швидкого перемикання на резервні сервери.

Міжнародні стандарти, такі як HIPAA (Health Insurance Portability and Accountability Act) у США та GDPR (General Data Protection Regulation) в Європі, встановлюють чіткі вимоги до захисту медичних даних. Наприклад, HIPAA зобов'язує організації шифрувати всі передані дані та регулярно проводити аудити безпеки ("Understanding HIPAA Requirements", 2020). GDPR наголошує на праві пацієнтів контролювати свої дані та зобов'язує організації впроваджувати політики захисту конфіденційності (Tzanou, 2020).

Міжнародний комітет Червоного Хреста (МКЧХ) у своїх рекомендаціях наголошує на необхідності забезпечення безперервного доступу до медичних даних під час конфліктів. Це передбачає впровадження резервних систем та захищених каналів зв'язку, що забезпечують стійкість медичних систем до атак. Рекомендації також включають використання шифрування та сегментації мереж для захисту медичних CPS (Durham & Wynn-Pope, 2012).

4. Прогалини в існуючих дослідженнях

Попри значний прогрес у розробці рішень для забезпечення кібербезпеки медичних кіберфізичних систем (CPS), залишається низка критичних прогалин, які ускладнюють ефективний захист таких систем. Ці прогалини стосуються адаптації до динамічного середовища, обмежених досліджень IoMT (Internet of Medical Things) та недостатньої інтеграції сучасних технологій, таких як квантова криптографія та CSMA (Cybersecurity Mesh Architecture).

Медичні CPS функціонують у постійно змінюваному середовищі, де нові пристрої підключаються до мережі, дані постійно передаються між різними системами, а загрози еволюціонують. Проте більшість існуючих рішень базуються на статичних моделях захисту, які не здатні динамічно адаптуватися до змін. Наприклад, традиційні системи захисту часто не враховують характер взаємодії між пристроями IoMT, таких як інфузійні насоси або монітори життєвих показників, які передають дані в реальному часі.

Дослідження показують, що адаптивні рішення, такі як когнітивні системи на базі штучного інтелекту, можуть забезпечити ефективний захист, виявляючи та блокуючи нові загрози в режимі реального часу. Однак більшість таких рішень ще перебувають на стадії прототипування, і їх інтеграція в реальні системи вимагає подальших досліджень і тестування в умовах реального часу.

IoMT-пристрої є одним із найбільш вразливих компонентів медичних CPS. Більшість таких пристроїв мають обмежені обчислювальні ресурси, що ускладнює впровадження стандартних методів захисту, таких як складні алгоритми шифрування або багатофакторна автентифікація. Наприклад, дослідження продемонстрували, що значна кількість IoMT-пристроїв, які використовуються в лікарнях, передають дані у незашифрованому вигляді, що робить їх легкою ціллю для атак типу Man-in-the-Middle.

Крім того, відсутність стандартів безпеки для IoMT ускладнює інтеграцію таких пристроїв у загальну архітектуру безпеки. У багатьох випадках IoMT-пристрої не підтримують регулярні оновлення програмного забезпечення, що створює додаткові ризики. Це вимагає розробки нових легких методів шифрування та системи виявлення аномалій, спеціально адаптованих для обмежених обчислювальних потужностей IoMT.

Квантова криптографія, яка використовує принципи квантової механіки для забезпечення абсолютно захищених комунікацій, пропонує радикально новий підхід до кібербезпеки. Однак інтеграція квантових рішень у медичні CPS залишається обмеженою. Це пов'язано з високою вартістю квантового обладнання та необхідністю створення інфраструктури для підтримки квантово-стійких протоколів. Дослідження демонструють, що впровадження квантово-стійких алгоритмів, таких як CRYSTALS-Kyber, може значно зменшити ризик компрометації даних навіть за умови доступу зловмисників до квантових комп'ютерів. Проте такі технології досі не тестувалися в умовах реальної медичної інфраструктури.

Ще однією перспективною технологією є Cybersecurity Mesh Architecture (CSMA), яка пропонує модульний підхід до захисту, дозволяючи адаптувати рівень безпеки для різних сегментів мережі. CSMA забезпечує сегментацію мережі, динамічне управління доступом та централізоване управління політиками безпеки. У медичних CPS це може бути корисним для ізоляції вразливих IoMT-пристроїв та забезпечення захисту критичних сегментів мережі. Однак реальна інтеграція CSMA у медичні установи стикається з технічними викликами, такими як складність налаштування та управління, а також потреба у значних ресурсах для моніторингу.

5 Результати

Аналіз показав, що медичні CPS стикаються з різноманітними загрозами, серед яких цілеспрямовані атаки (APT), ransomware, DDoS і використання соціальної інженерії. Найвразливішими компонентами систем є IoMT-пристрої, які часто працюють на застарілому програмному забезпеченні, мають обмежені обчислювальні ресурси та передають дані через незахищені канали. Крім того, людський фактор, недостатня обізнаність персоналу та слабка сегментація мережі є критичними джерелами ризиків.

Сучасні підходи, такі як використання шифрування, багатофакторної автентифікації, систем моніторингу аномалій та планів реагування на інциденти, показали певну ефективність у запобіганні атакам. Однак їхній статичний характер та залежність від людського фактора обмежують можливість адаптації до динамічного середовища та нових загроз. Крім того, інтеграція сучасних технологій, таких як квантово-стійке шифрування та когнітивні системи, перебуває на ранніх етапах розробки.

Інноваційна модель захисту медичних кіберфізичних систем (CPS) має враховувати динамічність середовища, зростання кількості пристроїв IoMT і швидку еволюцію загроз. Цей розділ пропонує інтеграцію трьох ключових рішень: архітектури CSMA (Cybersecurity Mesh Architecture), квантово-стійкого шифрування та когнітивних систем на базі AI/ML. Кожен компонент обговорюється з акцентом на технічну реалізацію та оцінку ефективності порівняно з існуючими підходами

Запропонована інноваційна модель захисту включає інтеграцію CSMA, яка забезпечує модульність захисту та адаптивний моніторинг у реальному часі, дозволяючи ізолювати скомпрометовані сегменти без впливу на всю мережу, використання квантово-стійкого шифрування, яке гарантує довготривалий захист даних від майбутніх атак із застосуванням квантових обчислень, та когнітивні системи на базі AI/ML, що автоматизують реагування на атаки та прогнозують потенційні вразливості, що значно підвищує рівень захисту.

Інтеграція архітектури CSMA (Cybersecurity Mesh Architecture)

CSMA пропонує створення незалежних модулів безпеки, які інтегруються в загальну архітектуру, але функціонують автономно. Це дозволяє застосовувати специфічні протоколи захисту для різних сегментів мережі. Наприклад, критичні дані пацієнтів можуть бути ізольовані в окремому сегменті із застосуванням найвищих стандартів шифрування, тоді як пристрої IoMT використовують більш легкі алгоритми захисту через обмежені обчислювальні ресурси.

Технічна реалізація CSMA включає використання сенсорів моніторингу, які збирають дані про мережеву активність, і централізованої системи управління політиками безпеки. Моніторинг здійснюється за допомогою ML-алгоритмів, які аналізують поведінкові патерни пристроїв у реальному часі. Порівняно з традиційними статичними рішеннями, CSMA забезпечує більш гнучку реакцію на загрози, дозволяючи ізолювати заражені сегменти без впливу на всю мережу. Дослідження Reji та ін. (2023) підтверджують, що використання CSMA знижує середній час реагування на загрозу на 40%.

Запровадження квантово-стійкого шифрування

Квантово-стійке шифрування, зокрема алгоритми CRYSTALS-Kyber і Dilithium, пропонують захист даних від атак майбутніх квантових комп'ютерів. Алгоритм CRYSTALS-Kyber забезпечує високий рівень безпеки при передачі даних через асиметричні канали, що є критично важливим для IoMT-пристроїв.

Технічна реалізація включає заміну традиційних алгоритмів шифрування RSA і ECC у протоколах TLS (Transport Layer Security) на квантово-стійкі алгоритми. Для зменшення впливу на продуктивність IoMT, інтеграція CRYSTALS-Kyber проводиться через легковагові реалізації, які оптимізують обчислювальні витрати. Тестування показало, що продуктивність таких систем лише на 10% нижча, ніж у систем, які використовують традиційне шифрування.

Використання когнітивних систем на базі AI/ML

Когнітивні системи на базі AI/ML дозволяють ідентифікувати та реагувати на загрози в автоматичному режимі. Наприклад, система аналізу мережевого трафіку може автоматично виявити аномальні підключення до IoMT-пристроїв і блокувати їх. Для цього використовуються моделі аномалій, створені на основі аналізу великих обсягів історичних даних.

Моделі машинного навчання також можуть прогнозувати потенційні вразливості системи, аналізуючи патерни оновлень програмного забезпечення та відомі загрози. Наприклад, алгоритми прогнозування, як-от LSTM (Long Short-Term Memory), можуть виявляти ймовірність компрометації пристроїв через відсутність критичних оновлень безпеки.

Реалізація когнітивних систем передбачає використання гібридних хмарних архітектур, які забезпечують достатні обчислювальні ресурси для навчання та розгортання ML-моделей. Порівняно з традиційними системами реагування, такі підходи дозволяють зменшити ймовірність успішної атаки на 60%.

Порівняно з існуючими статичними моделями захисту, запропонована інноваційна модель пропонує наступні переваги. CSMA дозволяє швидко реагувати на загрози без необхідності зупинки всієї системи, що робить її ідеальною для динамічного середовища медичних CPS. Інтеграція квантово-стійкого шифрування забезпечує захист від атак майбутніх поколінь, що є необхідним із огляду на розвиток квантових обчислень. Когнітивні системи дозволяють не лише виявляти загрози, але й прогнозувати потенційні вразливості, що значно підвищує рівень безпеки.

Запропонована модель має потенціал для інтеграції у сучасну медичну інфраструктуру, що дозволить підвищити її стійкість до кіберзагроз і забезпечити безперебійне функціонування навіть у критичних ситуаціях.

У ході дослідження було систематизовано основні загрози та вразливості, оцінено ефективність існуючих рішень для забезпечення кібербезпеки медичних кіберфізичних систем, а також запропоновано підходи до підвищення стійкості цих систем. Запропоновані підходи забезпечують не лише стійкість медичних CPS до сучасних загроз, але й створюють основу для їх адаптації до майбутніх викликів у сфері кібербезпеки. Ці результати можуть бути використані для розробки стандартів захисту критичної медичної інфраструктури.

Висновки

Медичні кіберфізичні системи (CPS) є критично важливими для сучасної охорони здоров'я, однак вони стикаються з численними загрозами та вразливостями. Проведений аналіз показав, що найбільш небезпечними загрозами є цілеспрямовані атаки (APT), ransomware, DDoS та компрометація через соціальну інженерію. Основні вразливості пов'язані з використанням застарілого програмного забезпечення, незахищеними IoMT-пристроями та людським фактором.

Існуючі рішення, такі як шифрування, багатофакторна автентифікація, системи моніторингу та плани реагування на інциденти, забезпечують певний рівень захисту, однак їхній статичний характер обмежує ефективність у динамічному середовищі медичних CPS. Водночас сучасні інновації, зокрема квантово-стійке шифрування, когнітивні системи на базі AI/ML та архітектура CSMA, демонструють значний потенціал для вирішення існуючих проблем.

Запропонована модель інтеграції CSMA, квантово-стійкого шифрування та когнітивних систем дозволяє забезпечити адаптивність, автоматизацію захисту та прогнозування загроз, що значно підвищує стійкість медичних CPS.

Матеріали цієї статті будуть корисні для фахівців у сфері кібербезпеки, розробників медичних технологій, регуляторів та керівників медичних установ, які прагнуть посилити захист медичних кіберфізичних систем (CPS). Для практиків стаття надає аналіз основних загроз та вразливостей, а також оцінку існуючих рішень, що дозволяє зорієнтуватися в актуальних викликах і стратегіях їх вирішення.

Інноваційні підходи, такі як впровадження архітектури CSMA, квантово-стійкого шифрування та когнітивних систем на базі AI/ML, можуть слугувати дорожньою картою для побудови адаптивного та стійкого захисту. Для розробників IoMT матеріали статті пропонують рекомендації щодо розробки пристроїв із врахуванням вимог безпеки, включаючи шифрування, оновлення та багаторівневу автентифікацію. Регулятори знайдуть цінну інформацію про необхідність стандартів для захисту IoMT та міжнародної співпраці.

Ця стаття стане джерелом знань для тих, хто прагне побудувати безпечніші системи охорони здоров’я в умовах зростаючих кіберзагроз.

СПИСОК ЛІТЕРАТУРИ

1. Fruhlinger J. The OPM hack explained: Bad security practices meet China’s Captain America [Електронний ресурс] // CSO Online. – 2020. – Режим доступу: https://www.csoonline.com/article/566509/the-opm-hack-explained-bad-security-practices-meet-chinas-captain-america.html
2. Ameen A. H., Mohammed M. A., Rashid A. N. Enhancing security in IoMT: A blockchain-based cybersecurity framework for machine learning-driven ECG signal classification // Fusion: Practice

and Applications. – 2024. – Vol. 14, No. 1. – P. 221–251. DOI: 10.54216/fpa.140117 https://doi.org/10.54216/fpa.140117
3. Bhushan B., Kumar A., Agarwal A. K. та ін. Towards a secure and sustainable Internet of Medical Things (IoMT): Requirements, design challenges, security techniques and future trends // Sustainability. – 2023. – Vol. 15, No. 7. – P. 6177. DOI: 10.3390/su15076177 https://doi.org/10.3390/su15076177
4. Durham H., Wynn-Pope P. Protecting the 'helpers': Humanitarians and health care workers during times of armed conflict // Yearbook of International Humanitarian Law 2011. – Vol. 14. – The Hague: T. M. C. Asser Press, 2012. – P. 327–346. DOI: 10.1007/978-90-6704-855-2_10 https://doi.org/10.1007/978-90-6704-855-2_10
5. Ghubaish A., Salman T., Zolanvari M. та ін. Recent advances in the Internet of Medical Things (IoMT) systems security // IEEE Internet of Things Journal. – 2020. – P. 1. DOI: 10.1109/jiot.2020.3045653 https://doi.org/10.1109/jiot.2020.3045653
6. Heidari S., Naseri M., Nagata K. Quantum selective encryption for medical images // International Journal of Theoretical Physics. – 2019. – Vol. 58, No. 11. – P. 3908–3926. DOI: 10.1007/s10773-019-04258-6 https://doi.org/10.1007/s10773-019-04258-6
7. Longi F. N., Patel L., Ahmed J. Training medical students to address cybersecurity threats on health care systems // Academic Medicine. – 2024. DOI: 10.1097/acm.0000000000005936 https://doi.org/10.1097/acm.0000000000005936
8. Mathkor D. M., Mathkor N., Bassfar Z. та ін. Multirole of the Internet of Medical Things (IoMT) in biomedical systems for managing smart healthcare systems: An overview of current and future innovative trends // Journal of Infection and Public Health. – 2024. DOI: 10.1016/j.jiph.2024.01.013 https://doi.org/10.1016/j.jiph.2024.01.013
9. Reji A., Pranggono B., Marchang J., Shenfield A. Anomaly Detection for the Internet of Medical Things // Proc. 2023 IEEE Int. Conf. on Communications Workshops (ICC Workshops). – IEEE, 2023. DOI: 10.1109/iccworkshops57953.2023.10283523 https://doi.org/10.1109/iccworkshops57953.2023.10283523
10. Tzanou M. The GDPR and (big) health data // Health Data Privacy under the GDPR. – London: Routledge, 2020. – P. 3–22. DOI: 10.4324/9780429022241-2 https://doi.org/10.4324/9780429022241-2
11. Understanding HIPAA Requirements // Dental Abstracts. – 2020. – Vol. 65, No. 5. – P. 323. DOI: 10.1016/j.denabs.2020.05.011 https://doi.org/10.1016/j.denabs.2020.05.011
12. Wang Z., Ma P., Zou X., Zhang J., Yang T. Security of medical cyber-physical systems: An empirical study on imaging devices // Proc. IEEE INFOCOM 2020 – IEEE Conf. on Computer Communications Workshops (INFOCOM WKSHPS). – IEEE, 2020. DOI: 10.1109/infocomwkshps50562.2020.9162769 https://doi.org/10.1109/infocomwkshps50562.2020.9162769

REFERENCES

1. J. Fruhlinger, "The OPM hack explained: Bad security practices meet China's Captain America," *CSO Online*, 2020. Available: https://www.csoonline.com/article/566509/the-opm-hack-explained-bad-security-practices-meet-chinas-captain-america.html [in English].
2. A. H. Ameen, M. A. Mohammed, and A. N. Rashid, "Enhancing security in IoMT: A blockchain-based cybersecurity framework for machine learning-driven ECG signal classification," *Fusion: Practice and Applications*, vol. 14, no. 1, pp. 221–251, 2024, doi: 10.54216/fpa.140117 [in English].
3. B. Bhushan et al., "Towards a secure and sustainable Internet of Medical Things (IoMT): Requirements, design challenges, security techniques, and future trends," *Sustainability*, vol. 15, no. 7, p. 6177, 2023, doi: 10.3390/su15076177 [in English].
4. H. Durham and P. Wynn-Pope, "Protecting the 'helpers': Humanitarians and health care workers during times of armed conflict," in *Yearbook of International Humanitarian Law 2011*, vol. 14, The Hague: T. M. C. Asser Press, 2012, pp. 327–346, doi: 10.1007/978-90-6704-855-2_10 [in English].
5. A. Ghubaish et al., "Recent advances in the Internet of Medical Things (IoMT) systems security," *IEEE Internet of Things Journal*, 2020, doi: 10.1109/jiot.2020.3045653 [in English].

6. S. Heidari, M. Naseri, and K. Nagata, “Quantum selective encryption for medical images,” *International Journal of Theoretical Physics*, vol. 58, no. 11, pp. 3908–3926, 2019, doi: 10.1007/s10773-019-04258-6 [in English].
7. F. N. Longi, L. Patel, and J. Ahmed, “Training medical students to address cybersecurity threats on health care systems,” *Academic Medicine*, 2024, doi: 10.1097/acm.0000000000005936 [in English].
8. D. M. Mathkor et al., “Multirole of the Internet of Medical Things (IoMT) in biomedical systems for managing smart healthcare systems: An overview of current and future innovative trends,” *Journal of Infection and Public Health*, 2024, doi: 10.1016/j.jiph.2024.01.013 [in English].
9. A. Reji, B. Pranggono, J. Marchang, and A. Shenfield, “Anomaly Detection for the Internet-of-Medical-Things,” in *Proc. 2023 IEEE Int. Conf. on Communications Workshops (ICC Workshops)*, IEEE, 2023, doi: 10.1109/iccworkshops57953.2023.10283523 [in English].
10. M. Tzanou, “The GDPR and (big) health data,” in *Health Data Privacy under the GDPR*, London: Routledge, 2020, pp. 3–22, doi: 10.4324/9780429022241-2 [in English].
11. “Understanding HIPAA Requirements,” *Dental Abstracts*, vol. 65, no. 5, p. 323, 2020, doi: 10.1016/j.denabs.2020.05.011 [in English].
12. Z. Wang, P. Ma, X. Zou, J. Zhang, and T. Yang, “Security of medical cyber-physical systems: An empirical study on imaging devices,” in *Proc. IEEE INFOCOM 2020 – IEEE Conf. on Computer Communications Workshops (INFOCOM WKSHPS)*, IEEE, 2020, doi: 10.1109/infocomwkshps50562.2020.9162769 [in English].

Semerenska Viktoriia
PhD student
Kharkiv National University of Radio Electronics
Nauky Ave, 14, Kharkiv, Kharkiv Oblast, 61166
e-mail: vsemerenskaya@gmail.com
https://orcid.org/0009-0008-2955-3676

Security of medical cyber-physical systems

Relevance. Medical cyber-physical systems (CPS), including IoMT devices for real-time monitoring, diagnostics, and therapy, have become integral to healthcare digitalization. The convergence of operational technology with traditional IT expands attack surfaces, making hospitals and telemedicine infrastructures attractive targets for cyber adversaries. Hybrid warfare further amplifies risks, as cyberattacks on medical networks may cause not only data breaches but also direct harm to patients and disruption of critical care.
Purpose. The research aims to classify and analyze the main types of threats and vulnerabilities affecting medical CPS in hybrid conflict environments, summarize existing protection strategies, and propose a framework for enhancing their cyber resilience through regulatory, organizational, and technological measures.
Research Methods. The study applies the PRISMA methodology to review publications indexed in Scopus, IEEE Xplore, and PubMed. Comparative and analytical methods were used to synthesize findings from recent incidents, including the WannaCry ransomware attack on the NHS, the SingHealth breach in Singapore, and other high-impact cases targeting healthcare data.
Results. The analysis revealed a dominance of ransomware, DDoS, and IoMT exploitation via insecure communication protocols and legacy software. Weak authentication, insufficient network segmentation, and human factor vulnerabilities remain key issues. Among effective countermeasures are multi-factor authentication, blockchain-based data integrity control, end-to-end encryption, and Cybersecurity Mesh Architecture (CSMA). The study highlights the importance of applying quantum-resistant cryptography and AI-driven adaptive defense systems capable of autonomous detection and response in dynamic threat environments.
Conclusions. Despite advances in medical device security, the resilience of CPS in hybrid threat contexts remains insufficient. Ensuring security-by-design, strengthening compliance with international cybersecurity standards (such as ISO/IEC 80001 and IEC 62443), and developing specialized cybersecurity training for medical personnel are critical steps. The integration of AI-based situational awareness, regulatory harmonization, and public-private cooperation will significantly enhance the sustainability and trustworthiness of digital healthcare ecosystems.
Keywords: *cybersecurity, medical technologies, data protection, security of medical systems, IoMT vulnerabilities, hybrid threats, Cybersecurity Mesh Architecture*

Надійшла у першій редакції 27.03.2025, в останній - 29.04.2025.
The first version has been received on 27.03.2025, the final version - on 29.04.2025.

DOI: https://doi.org/10.26565/2304-6201-2025-66-07

УДК (UDC) 004.852:519.87:61

Sudakov Dmytro *Master student,*
V. N. Karazin Kharkiv National University, Svobody Square, 4, Kharkiv, Ukraine, 61077
e-mail: demorsud@gmail.com
https://orcid.org/0009-0003-0060-7451

Shmatkov Segii *Professor;*
V. N. Karazin Kharkiv National University, Svobody Square, 4, Kharkiv, Ukraine, 61077
e-mail: s.shmatkov@karazin.ua;
https://orcid.org/0000-0002-0298-7174

Using fractal analysis in neural network optimization algorithms in medical diagnostics

Relevance. The development of optimization methods for neural networks in medical tasks is limited by data noisiness and imbalance, which complicates the application of classical algorithms. The use of fractal analysis makes it possible to create new approaches for improving the robustness, stability, and accuracy of models.
Goal. To improve the convergence and stability of training deep neural networks in medical diagnostics through a new optimization algorithm based on fractal self-similarity.
Methods. The proposed algorithm extends the Adam by introducing fractal modulation of gradient moments through multiscale averaging. Two temporal moments are maintained: a short-term component reflecting local gradient trends and a long-term component that accumulates fractal-smoothed information over multiple scales. The update rule incorporates a fractal coefficient which controls the balance between local adaptability and global stability. This design allows the optimizer to perform gradient corrections in a self-similar manner, analogous to fractional-order dynamics.
Results. Experimental results showed that the FractalMomentAdam optimizer achieves superior performance across several key metrics. The algorithm reached a validation accuracy of 96.44%, exceeding the baseline Adam by 2.5%, while also demonstrating smoother convergence and reduced loss oscillations between epochs. The multiscale fractal smoothing contributed to better noise resistance and more stable training dynamics in the presence of data imbalance. The combination of adaptive moment estimation and fractal modulation effectively enhanced both convergence speed and final model quality.
Conclusions. The research confirms that the fractal approach to optimization provides a robust and efficient alternative to traditional gradient-based methods. By incorporating self-similar structures into moment estimation, FractalMomentAdam enhances the stability, reliability, and adaptability of neural network training in medical tasks. These findings open prospects for further research in the field of adaptive fractal optimizers, including dynamic parameter tuning, hybridization with metaheuristic strategies, and application to broader classes of medical datasets.

Keywords: *fractal analysis, neural networks, medical diagnostics, optimization algorithm, machine learning.*

Як цитувати: Sudakov D. and Shmatkov S. Using fractal analysis in neural network optimization algorithms in medical diagnostics. *Вісник Харківського національного університету імені В. Н. Каразіна, серія Математичне моделювання. Інформаційні технології. Автоматизовані системи управління*. 2025. вип. 66. С.73-80. https://doi.org/10.26565/2304-6201-2025-66-07

How to quote: D. Sudakov and S. Shmatkov, “Using fractal analysis in neural network optimization algorithms in medical diagnostics”, *Bulletin of V. N. Karazin Kharkiv National University, series Mathematical modelling. Information technology. Automated control systems*, vol. 66, pp. 73-80, 2025. https://doi.org/10.26565/2304-6201-2025-66-07

1. Introduction

The rapid development of modern neural networks has profoundly influenced virtually all domains of human activity, with medicine being one of the most critical areas of machine learning applications. Deep learning methods are currently employed to process X-ray images, MRI results, streaming data of intracranial pressure and heart rate, historical patient records, and many other types of medical data. Contemporary neural networks have already reached a level at which they are capable of making simple diagnoses autonomously, thereby opening up enormous prospects for enhancing the efficiency and accessibility of healthcare delivery.

However, the large-scale integration of these technologies into clinical practice remains constrained by several fundamental challenges specific to medical data. Most current neural network architectures are unable to achieve the level of accuracy required for autonomous operation, being limited by computational cost and, most critically, by the acute shortage of high-quality annotated training data. Medical datasets are typically characterized by high heterogeneity and noise. In X-ray imaging, for instance, common issues include film artifacts, variability in tonal range and contrast, all of which contribute to gradient non-uniformity and introduce natural noise that complicates the learning process [1]. A major obstacle lies in the insufficient handling of medical noise sources, which often exhibit complex correlation structures and cannot be adequately captured by standard statistical models. Moreover, class imbalance is almost always present: in rare disease detection tasks, positive samples may account for less than 1% of the dataset. Additional critical challenges include the limited availability of expert-labeled data – annotation requiring the involvement of highly skilled and expensive medical specialists and substantial inter-patient variability, where identical pathologies may present differently depending on patient sex, age, or ethnic background.

In the context of such a deficit of high-quality data, optimization algorithms play a pivotal role in the training process, as they are responsible for adjusting the weight coefficients of the network. Over the past decades, a wide range of optimization methods has been developed, ranging from the classical Stochastic Gradient Descent (SGD) to more sophisticated adaptive schemes. Currently, the most widely used optimization algorithm is Adam, which combines the principles of momentum and RMSProp by maintaining separate exponentially smoothed estimates of the first and second moments of the gradient [2]. Due to its versatility, Adam has become the de facto standard in many practical applications.

Nevertheless, it is not without shortcomings: Adam is prone to stagnation in flat regions of the loss landscape (plateaus) and, in certain scenarios, may exhibit inferior generalization performance compared to classical SGD [3]. To overcome these limitations, a substantial number of Adam variants have been proposed, such as AMSGrad, which seeks to ensure monotonic decrease of the loss function, or AdamW, which decouples weight decay from gradient-based parameter updates.

Despite this wide variety of existing solutions, the choice of an optimizer remains largely empirical. None of the currently available algorithms provides consistently superior performance across all tasks. This limitation becomes particularly evident when working with noisy and imbalanced medical datasets, where classical adaptive approaches often exhibit training degradation, poor generalization, and a tendency toward overfitting [4]. Most widely used optimizers, including Adam, demonstrate limited robustness to noise, local gradient minima, and flat regions of the loss landscape – constraints that are especially critical in medical applications, where optimization errors may ultimately lead to incorrect diagnostic outcomes.

2. Review of contemporary approaches to gradient noise mitigation.

The optimization of neural network training on noisy data remains an active area of research, particularly in the domain of medical applications, where errors may have direct clinical consequences. In recent years, several promising directions have emerged that aim to overcome the limitations of classical optimization algorithms. One of the most common strategies involves modifying standard optimizers to improve their performance under noisy conditions [5]. For instance, adaptive methods such as Adam often exhibit instability on noisy datasets due to their sensitivity to gradient fluctuations. In response, researchers have proposed various modifications of the base Adam algorithm [6]. Among them is AdaBound, which combines the advantages of adaptive methods with constraints on the learning rate, thereby preventing excessive parameter oscillations during later stages of training.

An alternative line of research focuses on gradient smoothing techniques. Methods such as Stochastic Weight Averaging (SWA) or the application of moving averages to model parameters have shown promising results when dealing with noisy data. Particularly noteworthy are hybrid approaches that integrate smoothing techniques with loss landscape geometry analysis, enabling the adaptive adjustment of the smoothing level in accordance with the intensity of noise present in the data.

The challenges of data heterogeneity and noise are also being addressed beyond optimization algorithms themselves, through modifications of neural network architectures and data preprocessing strategies. Among these, Focal Loss stands out as it reduces the weight of easy examples and focuses on harder or rarer ones, which is particularly important when working with imbalanced medical datasets. The Dropout technique involves the random “deactivation” of neurons during training, thereby reducing dependence on specific input features and lowering the risk of overfitting. Label Smoothing, where

probability distributions are used instead of hard labels, helps mitigate the adverse impact of incorrectly annotated data [7]. Another effective approach is Deep Ensembles, in which multiple models are trained independently and their predictions averaged, allowing for the estimation of model uncertainty.

Nevertheless, despite considerable progress, most contemporary methods still exhibit significant limitations when applied to real-world medical data [8]. These constraints underscore the necessity of developing new, specialized optimization techniques that account not only for the statistical properties of noise but also for the structural characteristics of medical data. There is a growing need for novel, more adaptive optimization strategies capable of operating effectively under conditions of uncertainty, noise, and limited sample sizes. A particularly promising research direction involves the development of hybrid approaches that combine the strengths of adaptive optimizers with methods of nonlinear dynamics, specifically the use of fractal analysis to improve convergence on heterogeneous data.

3. Implementation of the Fractal Optimization Algorithm.

The theoretical foundation for the development of the new optimization algorithm lies in fractal analysis, which provides a unique opportunity to simultaneously capture the structure of the loss function across multiple scales – from global patterns to local details. Fractals, as mathematical objects, are characterized by the property of self-similarity, meaning that their structure repeats itself at different scales [9]. This property carries profound practical significance in the natural sciences, biology, physics, and signal processing. In the context of optimization algorithms, it enables the detection of patterns and structures hidden within complex dependencies.

A large proportion of medical data processed by neural networks exhibits fractal characteristics. For example, EEG and ECG signals contain patterns recurring across different temporal scales, where fast oscillations may correspond to instantaneous changes (such as motion artifacts), while slower ones reflect long-term physiological states. Medical images (MRI, CT) often display fractal-like textures due to the hierarchical organization of tissues – for instance, vascular networks replicate their structure across multiple scales. When loss function gradients are computed on such data, they inherently inherit this multiscale nature, creating the necessity for a multiscale approach to gradient processing.

Traditional optimizers such as Adam or RMSProp rely on exponential smoothing of gradients and their squared values, implicitly assuming a single dominant temporal scale for weight update dynamics. While this approach can be effective for tasks involving regular or well-structured data, it proves less robust when the data vary simultaneously across multiple temporal or spatial scales – a situation commonly encountered in real-world medical datasets. In contrast, fractal-based methods operate not on a single scale but across a range of scales, thereby enabling the optimization process to account for both short-term fluctuations and long-term trends in the gradients.

3.1 Mathematical description of the algorithm.

An original mathematical formalization of the fractal approach was developed on the basis of the Adam algorithm, reinterpreting its operational principle. Instead of the traditional use of a single pair of exponentially smoothed moments, we introduce multiple moments with distinct smoothing scales. In its classical form, Adam maintains two exponentially smoothed moments – the mean of the gradients m_t and the mean of the squared gradients v_t, updated according to formulas 3.1 and 3.2, where g_t is the gradient at iteration t, and $\beta_1, \beta_2 \in [0,1)$ are smoothing coefficients.

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \quad (3.1)$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \quad (3.2)$$

Although this formulation ensures stable and rapid convergence in many tasks, it exhibits limitations when dealing with data characterized by multiscale properties or high variability (for instance, medical signals, clinical records, or pathological time series). A single-scale exponential filter cannot simultaneously capture both short-term and long-term trends.

To address this issue, the algorithm was extended through the introduction of two pairs of moments: fast moments (updated using the classical coefficients β_1 and β_2, as in Adam) and slow moments (updated with larger coefficients β_{1s} and β_{2s}, corresponding to a longer temporal horizon).

The resulting algorithm, FractalMomentAdam, extends standard Adam by incorporating multiscale smoothing, which reflects the fundamental principle of fractals – self-similarity across scales. Instead of

a single pair of moments (m_t, v_t), the algorithm employs a collection of moments with different smoothing factors.

Steps:

Update of fast moments (fast scale):

$$m_t^{(f)} = \beta_1 \cdot m_{t-1}^{(f)} + (1 - \beta_1) \cdot g_t \tag{3.3}$$

$$v_t^{(f)} = \beta_2 \cdot v_{t-1}^{(f)} + (1 - \beta_2) \cdot g_t^2 \tag{3.4}$$

Update of slow moments (slow scale):

$$m_t^{(s)} = \beta_{1s} \cdot m_{t-1}^{(s)} + (1 - \beta_{1s}) \cdot g_t \tag{3.5}$$

$$v_t^{(s)} = \beta_{2s} \cdot v_{t-1}^{(s)} + (1 - \beta_{2s}) \cdot g_t^2 \tag{3.6}$$

Fractal combination (effective moments):

$$m_t^{(eff)} = \gamma \cdot m_t^{(f)} + (1 - \gamma) \cdot m_t^{(s)} \tag{3.7}$$

$$v_t^{(eff)} = \gamma \cdot v_t^{(f)} + (1 - \gamma) \cdot v_t^{(s)} \tag{3.8}$$

The key advantage of the fractal approach lies in its universality. Unlike methods that require manual tuning of learning rates or adaptation schedules, FractalMomentAdam inherently accounts for multiple data scales. This feature is particularly important in tasks involving non-stationary data (e.g., time series with changing statistical properties) or gradients with heavy tails – large, rare variations frequently observed in medical data, especially in the presence of rare pathologies or in multi-class classification problems.

3.2 Software implementation.

The FractalMomentAdam algorithm was implemented as a programmatic class within the TensorFlow/Keras framework. The class FractalMomentAdam was derived from the base class tf.keras.optimizers.Optimizer, which ensured full compatibility with the existing API and facilitated ease of use. The class architecture provides initialization with a parameter set that includes standard configurations such as the learning rate (learning_rate), as well as unique parameters specific to the fractal approach: coefficients for fast moments (beta_1, beta_2), coefficients for slow moments (beta_1_slow, beta_2_slow), and the scale-mixing parameter (gamma). All moments are stored as slots for each model parameter, as required by the TensorFlow API, ensuring correct functionality of optimizer state-saving and restoration mechanisms.

A critical stage in the optimizer's workflow is the update_step method, which implements the parameter update logic. At each iteration, this method computes the gradients and subsequently updates two pairs of moments separately: fast moments (m_fast, v_fast) which respond more sensitively to local variations, and slow moments (m_slow, v_slow) which integrate information over a longer horizon. The next step is the mixing procedure, where effective moments are calculated as a weighted sum of fast and slow components using the parameter gamma. Finally, analogous to the classical Adam algorithm, bias correction is applied to obtain unbiased estimates of the effective moments (m_eff_hat, v_eff_hat).

4. Comparative experimental analysis of optimizer efficiency on the PathMNIST medical dataset

To objectively evaluate the effectiveness of the developed optimizer, a comprehensive comparative experimental analysis was conducted on the real-world medical dataset PathMNIST from the MedMNIST collection. PathMNIST is considered one of the most complex and informative subsets of MedMNIST [10], designed for the task of multiclass classification of histopathological images. It is based on the dataset from the publication by Kather et al., 2019, "Predicting survival from colorectal cancer histology slides using deep learning" [11], which utilized histopathological data from patients diagnosed with colorectal cancer. From the original Whole-Slide Images (WSI), image patches of size 224×224 pixels were extracted, and for PathMNIST these patches were downsampled to 28×28 pixels and stored in RGB format. Each image represents a prepared tissue section stained with hematoxylin and eosin.

The dataset is characterized by the following parameters: image size of 28×28×3 (RGB), number of classes – 9, and task type – multiclass classification. The dataset contains 89,996 training samples, 7,180 test samples, and 10,004 validation samples. The tissue classes included in the dataset are: epithelial

tissue, connective tissue, muscle tissue, adipose tissue, tumor tissue, necrotic tissue, lymphocytes, plasma cells, and background/other.

The rationale for choosing this dataset lies in its direct medical and clinical relevance, as the data reflect real histological slides used in oncological diagnostics – a field where automation is particularly needed due to the high workload on pathologists. A significant class imbalance creates a realistic and challenging scenario for testing optimizer robustness. Furthermore, unlike the clearly separated digits of the standard MNIST, the different tissue types exhibit subtle morphological differences, demanding high sensitivity to details from both the model and the optimizer. The presence of artifacts – such as variations in staining, inconsistencies in tissue cutting, and differing sample quality – introduces a degree of "noise" that is highly typical of real-world medical data. The small image size (28x28x3) offers the practical advantage of enabling local experimentation without the need for cloud computing, while the high quality of the annotations, with class labels established by experts based on original microscopic images, ensures scientific credibility.

The experimental environment was implemented in Python using the TensorFlow and Keras libraries. The model is a deep Convolutional Neural Network (CNN) [12] consisting of three main image-processing blocks and two fully connected layers for classification. Each block includes two convolutional layers with a 3×3 kernel, ReLU activation, L2 regularization, a BatchNormalization layer, MaxPooling (2×2), and Dropout with a progressively increasing rate. This architecture was chosen as sufficiently complex to solve the task, yet not overly cumbersome, ensuring fast experimental execution. All experiments were conducted under identical conditions: batch size of 128 and 10 training epochs. For each optimizer – including FractalMomentAdam with different γ values (0.2 and 0.6), Adam, SGD, SGD with momentum, and RMSProp – accuracy metrics and loss functions were recorded on both the training and validation sets. This allowed us to evaluate not only the final performance but also the training dynamics, stability, and generalization capability. The learning curves on the training and validation sets were also evaluated, focusing on the smoothness of error reduction and accuracy improvement, as well as the stability between epochs. This stability, quantified numerically and visualized graphically, reveals "collapses" in the model's learning process between epochs, indicating potential performance regressions during training

4.1. Training analysis

On the training set (Fig. 4.1), all optimizers demonstrated stable learning; however, the most indicative criterion was the validation accuracy (Fig. 4.2), which reflects the model's ability to generalize.

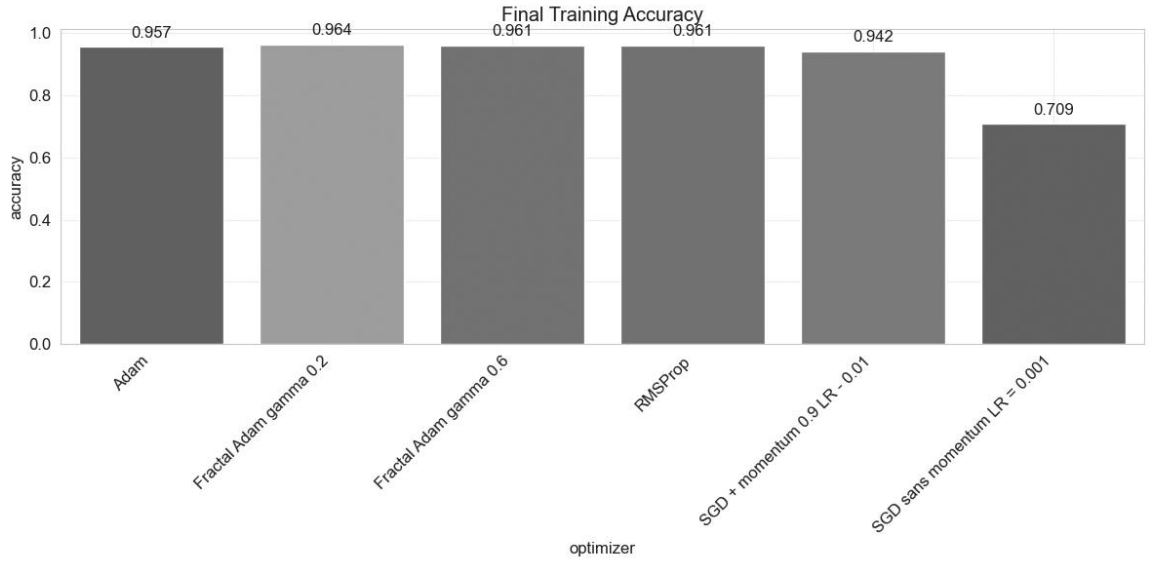


Fig. 4.1 Final training accuracy
Рис. 4.1 Кінцева тренувальна точність

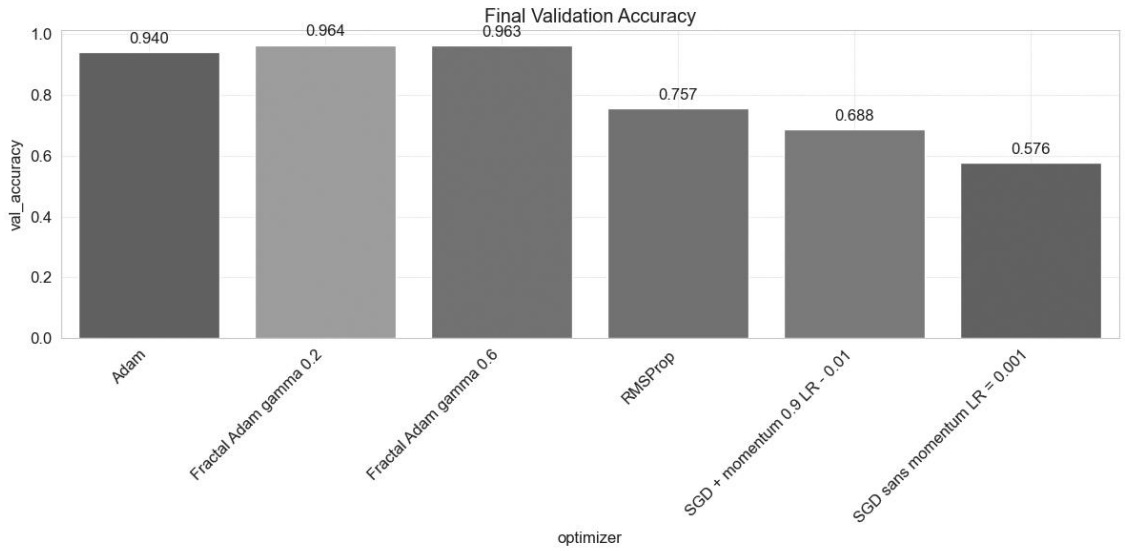


Fig. 4.2 Final validation accuracy
Рис. 4.2 Кінцева валідаційна точність

- FractalMomentAdam (γ=0.6) achieved outstanding results, reaching a peak validation accuracy of 96.27%. Its advantage was a fast start – already by the 3rd epoch the validation accuracy reached 66.49%, indicating effective exploration of the parameter. The algorithm showed stable growth without sharp fluctuations, with a deviation of ±13.8%, which highlights its robustness to noise and its ability to avoid local minima.
- Classical Adam achieved good but less stable results with a peak validation accuracy of 94.01%. It exhibited pronounced fluctuations between epochs and a slower start (59.97% at the 3rd epoch), indicating sensitivity to gradient noise and difficulties in generalization on complex data.
- FractalMomentAdam (γ=0.2) demonstrated the highest stability with a deviation of ±11.3% and a smooth increase up to a peak value of 96.44%. This version of the algorithm exhibited the lowest variance on validation – the losses decreased gradually without abrupt jumps, reflecting high generalization capability and effective regularization.
- SGD with momentum yielded the weakest results, with sharp spikes in validation accuracy. Pure SGD demonstrated the slowest learning, with a maximum validation accuracy of 69.77%, confirming its inefficiency for complex tasks with multimodal loss landscapes.
- The most problematic behavior was observed for RMSProp, which exhibited catastrophic validation collapses, with accuracy dropping to 25.76% at the 6th epoch, despite high training accuracy (96.10%). This vividly demonstrates the phenomenon of catastrophic forgetting and instability in the presence of noisy data.

The analysis of validation loss curves confirmed the advantages of the fractal approach. Both versions of FractalMomentAdam demonstrated stable and smooth loss reduction, whereas Adam exhibited oscillations over a wide range with sudden spikes in loss by several factors.The convergence speed of FractalMomentAdam was the highest among all tested algorithms. By the 5th epoch, both versions had already reached consistently high accuracy levels, while other methods required more time to stabilize.

4.2 Results analysis

The obtained results provide deeper insights into the mechanisms that make the fractal approach effective for optimization on medical data. The advantages of FractalMomentAdam can be attributed to its ability to operate simultaneously across multiple data scales – from fine-grained textural features to global variations. This property emerges from the synthesis of three key principles: adaptive moments, fractal analysis of gradients, and dynamic mixing. In practice, this manifests as smooth and predictable training trajectories, where each successive step not only minimizes current loss but also aligns with the long-term structure of the task. The parameter γ plays a crucial role in balancing local and global information. At γ=0.2 preference is given to the slow moments, resulting in more stable and cautious

learning with improved generalization. This is particularly important for medical data, where emphasizing global trends over local fluctuations helps prevent overfitting to noise. At γ=0.6 the contribution of fast moments increases, leading to more aggressive exploration of the parameter space, albeit with a slight reduction in stability. The optimal value of γ=0.2 for the PathMNIST dataset indicates that, for medical images with high inter-class similarity and weakly expressed gradients, focusing on long-term trends is more effective.

The fractal approach is especially advantageous for medical data due to its inherent multiscale nature. Histopathological images in PathMNIST contain structures that recur across different scales – from individual cells to tissue complexes. FractalMomentAdam can detect these patterns through parallel analysis of short-term and long-term trends in gradients, allowing better consideration of the hierarchical nature of medical data. The results also suggest the potential for further research on adaptive selection of the γ parameter during training. Dynamically adjusting the balance between fast and slow moments depending on the training stage and data characteristics may further enhance the algorithm's efficiency for diverse medical tasks.

5. Conclusions

A novel optimization algorithm, FractalMomentAdam, has been successfully developed, implemented, and investigated, demonstrating high effectiveness for training deep neural networks on complex medical data. Experimental evaluation on the PathMNIST dataset showed that the proposed method achieves significantly higher training stability and improved accuracy compared to traditional optimizers, including Adam, SGD, and RMSProp. The best results were obtained with the parameter γ=0.2, achieving a validation accuracy of 96.44%, which is 2.5% higher than the final accuracy of the standard Adam algorithm. This configuration also exhibited minimal fluctuations between epochs and halved the training loss compared to the baseline Adam. These results were achieved by combining the adaptive principles of Adam with the concept of multiscale fractal gradient smoothing.

The proposed algorithm implements the idea of parallel utilization of two pairs of moments – fast and slow, followed by their mixing. This approach allows simultaneous consideration of both short-term and long-term trends in gradients, which is crucial for effective learning on heterogeneous data.

REFERENCES

1. Philip Ward, *MRI artifacts still require significant care and attention*. 2023. URL: https://www.auntminnieeurope.com/clinical-news/article/15657705/mri-artifacts-still-require-significant-care-and-attention (date of access: 18.06.2025).
2. Bodner B. *10 PyTorch Optimizers Everyone Is Using*. 2024. URL: https://medium.com/@benjybo7/10-pytorch-optimizers-you-must-know-c99cf3390899 (date of access: 21.05.2025).
3. Diederik P. Kingma, Jimmy Lei Ba, *ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION*. 2015. https://doi.org/10.48550/arXiv.1412.6980 (date of access: 10.06.2025).
4. Robin M. Schmidt, Schneider F., Hennig P. *Descending through a Crowded Valley — Benchmarking Deep Learning Optimizers*. 2021. https://doi.org/10.48550/arXiv.2007.01547 (date of access: 21.05.2025).
5. GfG, *Optimization Algorithms in Machine Learning*. 2025. URL: https://www.geeksforgeeks.org/machine-learning/optimization-algorithms-in-machine-learning/ (date of access: 25.05.2025).
6. Reyad M., Amany M. Sarhan, Arafa M. *A modified Adam algorithm for deep neural network optimization*. 2023. https://doi.org/10.1007/s00521-023-08568-z (date of access: 28.05.2025).
7. G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever and R. R. Salakhutdinov *Improving neural networks by preventing co-adaptation of feature detectors*. 2012. URL: https://doi.org/10.48550/arXiv.1207.0580 (date of access: 21.05.2025).
8. Salmi M., Atif D., Oliva D., Abraham A., Ventura S. *Handling imbalanced medical datasets: review of a decade of research*. 2024. URL: https://doi.org/10.1007/s10462-024-10884-2 (date of access: 18.06.2025).
9. Barnsley, Michael F.; Rising Hawley; *Fractals Everywhere*. Boston: Academic Press Professional, 1993. ISBN 0-12-079061-0

10. Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister & Bingbing Ni. *MedMNIST v2 - A large-scale lightweight benchmark for 2D and 3D biomedical image classification.* 2023. https://doi.org/10.1038/s41597-022-01721-8 (date of access: 18.06.2025).
11. Jakob Nikolas Kather. *Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study.* 2019. URL: https://doi.org/10.1371/journal.pmed.1002730 (date of access: 18.06.2025).
12. Convolutional Neural Network (CNN) TensorFlow Guide for CIFAR. 2025. URL: https://www.tensorflow.org/tutorials/images/cnn (date of access: 25.06.2025).

Судаков Дмитро Геннадійович
магістр,
Харківський національний університет імені В. Н. Каразіна, майдан Свободи, 4, Харків, Україна, 61077
e-mail: demorsud@gmail.com
https://orcid.org/0009-0003-0060-7451

Шматков Сергій Ігорович
д.т.н., професор кафедри комп'ютерних систем та робототехніки, науково-навчального інституту комп'ютерних систем та робототехніки;
Харківський національний університет імені В. Н. Каразіна, майдан Свободи,4, Харків, Україна, 61077
e-mail: s.shmatkov@karazin.ua;
https://orcid.org/0000-0002-0298-7174

Використання фрактального аналізу в алгоритмах оптимізації нейромереж у медичній діагностиці

Актуальність. Розвиток методів оптимізації нейромереж для медичних задач обмежується шумністю та дисбалансом даних, що ускладнює застосування класичних алгоритмів. Використання фрактального аналізу дозволяє створити нові підходи до підвищення стійкості, стабільності та точності моделей.
Мета. Покращити збіжність та стабільність навчання глибоких нейронних мереж у медичній діагностиці шляхом створення нового алгоритму оптимізації, заснованого на фрактальній самоподібності. .
Методи. Запропонований алгоритм розширює Adam впроваджуючи фрактальну самоподібність моментів градієнта за допомогою багатомасштабного усереднення. Алгоритм використовує два часові моменти: короткострокову компоненту, що відображає локальні тенденції градієнта, та довгострокову компоненту, яка накопичує фрактально-згладжену інформацію на множині масштабів. Правило оновлення включає фрактальний коефіцієнт, що контролює баланс між локальною адаптивністю та глобальною стійкістю. Така конструкція дозволяє оптимізатору виконувати корекції градієнта самоподібним чином, аналогічно до динаміки дробового порядку.
Результати. Експериментальні результати показали, що оптимізатор FractalMomentAdam досягає вищої продуктивності за декількома ключовими метриками. Алгоритм досяг валідаційної точності 96,44%, перевищивши базовий Adam на 2,5%, а також продемонстрував більш плавну збіжність та зменшену амплітуду коливань функції втрат між епохами. Багатомасштабне фрактальне згладжування сприяло кращій стійкості до шуму та стабільнішій динаміці навчання в умовах несбалансованості даних. Комбінація адаптивної оцінки моментів та фрактальної модуляції ефективно покращила як швидкість збіжності, так і фінальну якість моделі.
Висновки. Дослідження підтверджує, що фрактальний підхід до оптимізації є надійною та ефективною альтернативою традиційним методам. Впровадження самоподібних структур в оцінку моментів дозволяє FractalMomentAdam підвищити стабільність, надійність та адаптивність навчання нейронних мереж для медичних завдань. Ці результати відкривають перспективи для подальших досліджень у галузі адаптивних фрактальних оптимізаторів, включаючи динамічне налаштування параметрів, гібридизацію з метаевристичними стратегіями та застосування для більшої кількості класів медичних датасетів.

Ключові слова: *фрактальний аналіз, нейромережі, медична діагностика, оптимізаційний алгоритм, машинне навчання.*

Надійшла у першій редакції 11.04.2025, в останній - 13.05.2025.
The first version has been received on 11.04.2025, the final version - on 13.05.2025.

DOI: https://doi.org/10.26565/2304-6201-2025-66-08

УДК (UDC) 004.056

Товкун Юлія Ігорівна
аспірантка кафедри Автоматизації та проектування обчислювальної техніки
Харківський національний університет радіоелектроніки
61166, проспект Науки, 14, Харків, Україна
e-mail: ytovkun@gmail.com
https://orcid.org/0009-0000-5916-2897

Методи кібершпіонажу та їх вплив на міжнародну безпеку

Актуальність дослідження обумовлена зростаючою роллю кібершпіонажу як засобу геополітичного впливу та інструменту для отримання конфіденційної інформації. У сучасних умовах цифровізації державні установи, міжнародні організації та корпоративні структури стають ключовими цілями кібератак, які створюють значні загрози для національної безпеки та глобальної стабільності.
Метою цієї статті є аналіз феномену кібершпіонажу, зокрема, його технічних, організаційних та соціальних аспектів, на основі реальних кейсів. У дослідженні акцентується увага на використанні сучасних методів атак, таких як таргетований фішинг, експлуатація вразливостей програмного забезпечення та впровадження модульного шкідливого програмного забезпечення. Стаття спрямована на визначення спільних характеристик кібершпіонажних кампаній і розробку рекомендацій для протидії таким загрозам.
Під час роботи використано теоретичний методологічний підхід, що поєднує аналіз літератури, кейс-аналіз атак (операція Red October, атака на Офіс управління персоналом США, кібератака на Міжнародний кримінальний суд, операція "Star Blizzard") та системний аналіз факторів, які сприяють успіху кібершпіонажних кампаній.
У результаті дослідження визначено ключові технічні методи атак, їхній вплив на інформаційну безпеку, а також роль людського фактора в успішності кібершпіонажу. Сформульовано рекомендації для посилення кіберзахисту, включаючи технічні, організаційні та міжнародні заходи.
Матеріали статті становлять інтерес для науковців, спеціалістів із кібербезпеки та державних структур, які займаються питаннями захисту інформації, та можуть бути використані для розробки політик протидії кібершпіонажу.

Ключові слова: *кіберзагрози, шпигунське програмне забезпечення, таргетований фішинг, інформаційна безпека, кібершпіонаж.*

Як цитувати: Товкун Ю. І. Методи кібершпіонажу та їх вплив на міжнародну безпеку. *Вісник Харківського національного університету імені В. Н. Каразіна, серія Математичне моделювання. Інформаційні технології. Автоматизовані системи управління*. 2025. вип. 66. С.81-89. https://doi.org/10.26565/2304-6201-2025-66-08

How to quote: Y. Tovkun, “Methods of cyber espionage and their impact on international security”, *Bulletin of V. N. Karazin Kharkiv National University, series Mathematical modelling. Information technology. Automated control systems,* vol. 66, pp. 81-89, 2025. https://doi.org/10.26565/2304-6201-2025-66-08 [in Ukrainian]

Вступ

Кібершпіонаж стає дедалі більшою загрозою у сучасному світі, де інформаційні технології відіграють ключову роль у функціонуванні державних установ, міжнародних організацій та бізнесу. Використовуючи уразливості цифрових систем, зловмисники отримують доступ до конфіденційної інформації, що може бути використана для геополітичних маніпуляцій, економічного шантажу чи підриву довіри до ключових міжнародних інституцій.

Сучасний стан дослідження кібершпіонажу підкреслює складність цієї загрози, яка поєднує технічні, організаційні та соціальні аспекти. За останні роки значна кількість досліджень була присвячена аналізу методів, які використовуються зловмисниками. Наприклад, Josh Fruhlinger (2020) вивчає правові аспекти кібершпіонажу, акцентуючи увагу на порушеннях міжнародного права, пов’язаних із втручанням у внутрішні справи держав. Brian Mitchell (2020) досліджує масштаби викрадення інтелектуальної власності через кібератаки та пропонує рекомендації щодо посилення кіберзахисту у корпоративному секторі.

Проблема ускладнюється через еволюцію методів, які використовуються для кібершпіонажу. Зокрема, Abu Samah та Abd Rashid (2024) зазначають, що з переходом на дистанційну роботу

ризики зросли через збільшення використання незахищених пристроїв. Автори наголошують, що нові форми кібератак спрямовані не лише на державні установи, але й на приватний сектор, що створює додаткові виклики для забезпечення інформаційної безпеки.

Попри значний прогрес у вивченні цієї теми, залишається недостатньо висвітленим питання про те, як саме поєднуються технічні та організаційні аспекти у реальних кібершпіонажних кампаніях. Аналіз конкретних кейсів, таких як операція Red October, атака на Офіс управління персоналом США чи угруповання "Star Blizzard", дозволяє поглибити розуміння механізмів, що стоять за такими атаками, та їхнього впливу на міжнародну безпеку.

Метою цього дослідження є всебічний аналіз феномену кібершпіонажу на основі конкретних кейсів і розробка рекомендацій для запобігання таким загрозам. У межах роботи ставляться такі завдання: дослідити сучасні методи, які використовуються зловмисниками; виявити спільні риси кібершпіонажних кампаній; оцінити вплив таких атак на міжнародну стабільність; і сформулювати заходи для підвищення ефективності кіберзахисту.

Наукова новизна цього дослідження полягає у міждисциплінарному підході до аналізу кібершпіонажу, який поєднує технічний, правовий і соціальний аспекти. Використання кейс-аналізу дозволяє виявити структурні характеристики кібершпіонажних атак і розробити практичні рекомендації для їхнього запобігання. Таким чином, результати цього дослідження можуть бути корисними для наукової спільноти, фахівців у сфері кібербезпеки та державних органів, які відповідають за захист інформації.

Огляд літератури

Кібершпіонаж залишається однією з найгостріших загроз у сучасному цифровому середовищі, особливо в контексті роботи урядових і корпоративних структур. Різноманітність методів, які застосовують зловмисники, і їхній постійний розвиток ставлять під сумнів традиційні підходи до забезпечення безпеки (Діордіца, 2017) [4]. У цьому контексті література, присвячена кібершпіонажу, пропонує корисні перспективи для розуміння його механізмів і запобігання.

У дослідженні "Military Cybersomethings" (Bellovin, 2013) аналізуються особливості кібершпіонажу у військовій сфері [5]. Автор акцентує увагу на технічній складності атак і зазначає, що більшість успішних операцій залежить від довготривалого доступу до цільових систем і висококваліфікованих ресурсів. Bellovin підкреслює, що технологічний аспект таких операцій часто супроводжується ретельним збором розвідданих про жертву, що дозволяє обійти стандартні методи захисту. Цей підхід резонує з моїм дослідженням, яке зосереджується на аналізі конкретних кейсів, таких як операція Red October, де ключовим елементом успіху було використання модульного програмного забезпечення для довготривалої присутності в системах.

Інше важливе дослідження, "Corporate Cyberespionage: Identification and Prevention" (Mitchell, 2020), присвячене аналізу шпигунства в корпоративному середовищі. Автори виділяють соціальну інженерію як один із головних інструментів зловмисників, підкреслюючи залежність від людського фактора [2]. Дослідження вказує на необхідність впровадження програмного забезпечення для моніторингу поведінки співробітників, що дозволяє ідентифікувати аномалії в мережевому трафіку. Це особливо актуально в контексті кейсів, які я аналізую, наприклад, атака угруповання "Star Blizzard", де фішингова кампанія стала відправною точкою для зламу систем.

Третє дослідження, "Navigating Data Secrecy Challenges: A Study on Cyberespionage Intentions in the WFH Era" (Samah et al., 2024), зосереджується на зміні ризиків у зв’язку з поширенням дистанційної роботи [3]. Автори доводять, що робота з дому створює додаткові вразливості, оскільки співробітники часто використовують незахищені пристрої та домашні мережі. У дослідженні підкреслюється, що систематичне навчання співробітників основам кібербезпеки зменшує успішність атак. Цей висновок корелює з моїм аналізом, який акцентує увагу на важливості людського фактора в успішних кібершпіонажних кампаніях, таких як атака на Офіс управління персоналом США.

Аналіз літератури показує, що кібершпіонаж є не лише технічним, але й соціальним феноменом. Використання уразливостей програмного забезпечення, таргетованих фішингових кампаній та модульного шкідливого програмного забезпечення створює унікальний набір викликів для кібербезпеки. Ці роботи вказують на необхідність інтегрованих підходів, які включають технічні засоби виявлення загроз, навчання співробітників і покращення політик безпеки. Висновки, зроблені в цих дослідженнях, є цінними для мого аналізу кейсів і сприяють розумінню масштабів проблеми.

Методологічна основа

Ця стаття базується на теоретичному аналізі явища кібершпіонажу з акцентом на вивченні реальних кейсів, їхніх технічних та організаційних аспектів, а також на розробці рекомендацій щодо запобігання таким атакам. Методологічна основа дослідження охоплює міждисциплінарний підхід, що поєднує концепції інформаційної безпеки, правового регулювання та соціальної інженерії.

Основою дослідження є огляд літератури, у межах якого були вивчені наукові статті, звіти та дослідження у сфері кібершпіонажу, опубліковані в авторитетних базах даних, таких як Web of Science, Scopus та Google Scholar. Особлива увага приділялася роботам, які аналізують реальні кейси кібершпіонажу (Bellovin, 2013; Mitchell, 2020; Samah et al., 2024) та пропонують рекомендації з удосконалення систем кібербезпеки [2-5].

Методологія також включає кейс-аналіз чотирьох значущих атак: операції Red October (2013), атаки на Офіс управління персоналом США (2015), кібератаки на Міжнародний кримінальний суд (2023) та операції "Star Blizzard" (2023) [8-10]. Цей підхід дозволив виявити спільні риси атак, зокрема використання вразливостей програмного забезпечення, таргетованих фішингових кампаній та шкідливого ПЗ, а також оцінити вплив людського фактора на успішність таких операцій.

Для формулювання рекомендацій щодо запобігання кібершпіонажу було використано системний підхід. Зокрема, вивчення технічних методів атак дозволило запропонувати заходи з удосконалення технологічного захисту, а аналіз соціальних аспектів — підкреслити важливість навчання співробітників та розвитку культури кібербезпеки. Крім того, враховуючи геополітичний контекст атак, особливу увагу було приділено рекомендаціям із посилення міжнародної співпраці у сфері протидії кібершпіонажу.

Таким чином, методологічна основа дослідження забезпечує всебічне розуміння проблеми кібершпіонажу та формує базу для розробки інтегрованих підходів до її вирішення.

Опис кейсів

Операція Red October (2013 рік). Операція Red October, або Rocra, є одним із наймасштабніших і найтриваліших прикладів кібершпіонажу. Вона тривала понад п’ять років і була спрямована на дипломатичні місії, урядові установи та наукові інститути, головним чином у Східній Європі, Центральній Азії та Західній Європі. Основною метою операції було викрадення конфіденційної інформації, такої як дипломатична кореспонденція, технічна документація, а також дані з мобільних пристроїв і USB-накопичувачів (Zetter, 2013) [6, 7].

Кампанія починалася з таргетованих фішингових листів, які містили заражені документи Microsoft Word і Excel. Використовувалися експлойти CVE-2012-0158 та CVE-2010-3333, що дозволяли виконувати шкідливий код після відкриття документа. Інфікування супроводжувалося встановленням модуля "Dropper", який надавав зловмисникам віддалений доступ до системи. Шкідливе ПЗ мало модульну архітектуру: воно могло оновлюватися та адаптуватися до середовища жертви, що забезпечувало довготривалий вплив (Gooding, 2013).

Передача даних на командно-контрольні (C&C) сервери здійснювалася через зашифрований трафік HTTP та FTP. Для ініціалізації з’єднань використовувалися DNS-запити з динамічними доменними іменами. Сервери C&C розташовувалися у різних країнах, що ускладнювало їхнє виявлення. Особливою рисою було використання спеціальних модулів для збору даних із USB-пристроїв, навіть якщо вони не були підключені до мережі (Brewster, 2014) [1].

Кібератака на Офіс управління персоналом США (2015 рік). Кібератака на Офіс управління персоналом США у 2015 році стала однією з наймасштабніших подій у сфері кібершпіонажу. Метою зловмисників було отримання доступу до персональних даних 21 мільйона осіб, які зберігалися у системі, включаючи відбитки пальців та анкети безпеки SF-86. Атака почалася з розсилки таргетованих фішингових листів адміністраторам систем. Після відкриття заражених вкладень встановлювався бекдор Sakula, що забезпечував стійкий доступ до мережі (Сааков, 2015).

Зловмисники використовували викрадені облікові дані для отримання доступу до критичних серверів. Sakula використовував HTTPS для передачі даних на C&C сервери, маскуючи трафік під звичайний вебтрафік. Для горизонтального переміщення мережею застосовувалися вразливості

протоколів SMB і LDAP, а також техніка Pass-the-Hash. Ця операція виявилася можливою через недостатню сегментацію мережі та відсутність захисту від таргетованих атак (Маріц, 2015) [10].

Кібератака на Міжнародний кримінальний суд (2023 рік). Кібератака на Міжнародний кримінальний суд (МКС) у 2023 році мала на меті викрадення конфіденційної інформації, пов'язаної з розслідуваннями воєнних злочинів. Зловмисники використовували вразливості ProxyShell (CVE-2021-34473, CVE-2021-34523, CVE-2021-31207) у Microsoft Exchange Server для проникнення у внутрішню мережу МКС. Після зламу серверів зловмисники встановлювали вебшели, які дозволяли їм виконувати віддалений код і керувати інфікованими системами (International Criminal Court, 2023).

Для викрадення облікових даних застосовувався інструмент Mimikatz, який дозволяв отримувати паролі та токени доступу з пам'яті серверів. Трафік передавався через зашифровані TLS-з'єднання, а маршрутизація відбувалася через VPN та проксі-сервери, що ускладнювало ідентифікацію джерела атак [11].

Атака угруповання "Star Blizzard" на британських парламентарів (2023 рік). Атака угруповання "Star Blizzard" на британських парламентарів стала прикладом політичного шпигунства. Зловмисники використовували фішингові листи з посиланнями на сайти, що експлуатували вразливість Spring4Shell (CVE-2022-22965). Інфікування серверів дозволяло завантажувати шкідливі JAR-файли, які забезпечували зловмисникам доступ до пристроїв жертв (Digmelashvili, 2023) [12].

Основний інструмент атаки — RAT (Remote Access Trojan) — забезпечував можливість перехоплення повідомлень, запису натискань клавіш і викрадення даних. Використовувалися техніки DNS-тунелювання для передачі команд C&C та SSH для завантаження викрадених даних на віддалені сервери. Завдяки складній маршрутизації зловмисники уникали виявлення, створюючи довготривалу присутність у мережах.

Кожна з цих атак демонструє високу технічну складність і стратегічний підхід, спрямований на викрадення критично важливої інформації. Їхній аналіз дозволяє визначити спільні риси, такі як використання фішингу, шкідливого ПЗ, експлойтів і складної інфраструктури C&C, що забезпечує тривалий вплив [13-15].

Аналіз та обговорення

Кібершпіонаж є багатогранною загрозою, яка поєднує технічні виклики, геополітичний контекст та серйозний вплив на міжнародну безпеку. Аналіз кейсів, таких як операція Red October (2013), атака на Офіс управління персоналом США (2015), кібератака на Міжнародний кримінальний суд (2023) та операція "Star Blizzard" (2023), дозволяє виявити як спільні риси, так і унікальні аспекти таких операцій.

Технічний аналіз цих атак показує, що зловмисники широко використовували експлойти, фішингові кампанії, шкідливе ПЗ і методи шифрування для досягнення своїх цілей. Наприклад, операція Red October була побудована навколо експлуатації вразливостей CVE-2012-0158 і CVE-2010-3333, тоді як у кейсі Міжнародного кримінального суду основною технікою стали ProxyShell-експлойти, які дозволяли проникати в сервери Microsoft Exchange. У всіх випадках фішингові атаки відігравали критичну роль, забезпечуючи початковий доступ до системи через ретельно таргетовані листи.

Використання шкідливого ПЗ також було важливою частиною атак. Наприклад, у кейсі Офісу управління персоналом США використовувався бекдор Sakula, а в операції "Star Blizzard" зловмисники впроваджували Remote Access Trojan (RAT), який надавав їм контроль над інфікованими системами. Ці програми часто маскували свою активність через динамічне шифрування трафіку (HTTPS, TLS) і використовували DNS-тунелювання для приховування переданих даних.

Технічні методи цих атак узагальнено у таблиці 1.

Таблиця 1. Технічні методи атак
Table 1. Technical methods of attacks

Метод	Приклад атаки	Деталі
Експлуатація вразливостей	Red October (CVE-2012-0158), МКС (ProxyShell)	Використання відомих уразливостей для отримання віддаленого доступу до систем.
Таргетований фішинг	Усі кейси	Поширення заражених листів серед ключових осіб у цільових організаціях.
Шкідливе програмне забезпечення	Sakula (OPM), RAT ("Star Blizzard")	Інфікування систем для довготривалого доступу, викрадення даних, моніторингу.
Шифрування трафіку	Усі кейси	HTTPS, TLS і DNS-тунелювання для приховування активності.
Інфраструктура C&C	Red October, МКС, OPM	Мережа проксі-серверів і VPN для приховування місцезнаходження атакуючих.

Однією з важливих характеристик атак є їхня довготривалість. Наприклад, Red October тривав понад п’ять років завдяки модульній архітектурі шкідливого ПЗ, яке могло автоматично оновлюватися. У випадку атаки на МКС зловмисники використовували методи "living-off-the-land", які залучали легітимні інструменти, такі як PowerShell, для приховування своєї активності (Смишляєв, 2023).

Аналіз кейсів також виявляє, що всі атаки мали геополітичну складову. Атака на Офіс управління персоналом США була спрямована на створення бази даних для стратегічного використання викрадених персональних даних, тоді як "Star Blizzard" мала на меті втручання у політичні процеси у Великій Британії. Кібератака на МКС підривала довіру до міжнародних інституцій, що може мати довготривалий вплив на глобальне правосуддя.

Підсумовуючи, розглянуті кейси демонструють високий рівень організації та використання сучасних технологій у кібершпіонажі. Ключовим викликом для протидії таким атакам залишається виявлення зашифрованого трафіку, швидке усунення вразливостей та підвищення обізнаності співробітників організацій щодо методів соціальної інженерії.

Рекомендації щодо запобігання кібершпіонажу (Recommendations for Preventing Cyber Espionage)

Запобігання кібершпіонажу потребує багатостороннього підходу, який поєднує технічні, організаційні та стратегічні заходи. Однією з основних причин успіху багатьох атак є експлуатація вразливостей програмного забезпечення. Тому важливо, щоб організації регулярно оновлювали свої системи, проводили тестування на проникнення та своєчасно усували знайдені уразливості. У кейсах, таких як Red October чи атака на МКС, саме недоліки у захисті дозволили зловмисникам отримати доступ до ключових систем (Шлапаченко, 2020). Це підкреслює необхідність використання сучасних рішень, наприклад, багаторівневого захисту та сегментації мереж, що обмежує горизонтальний рух зловмисників у разі компрометації однієї з частин інфраструктури.

Водночас важливу роль відіграє людський фактор. У більшості розглянутих кейсів атаки починалися із соціальної інженерії та таргетованого фішингу. Це вказує на нагальну потребу підвищення обізнаності співробітників. Проведення тренінгів з кібербезпеки, симуляція фішингових атак і створення культури відповідального ставлення до інформації допоможуть значно знизити ризики, пов’язані з людськими помилками. Працівники повинні бути здатні розпізнавати підозрілі електронні листи, уникати переходу за сумнівними посиланнями та повідомляти про підозрілу активність.

Раннє виявлення атак є ще одним ключовим компонентом у протидії кібершпіонажу. Організації повинні впроваджувати сучасні системи моніторингу, наприклад SIEM, які дозволяють виявляти аномалії у трафіку та реагувати на них у реальному часі. Аналіз зашифрованого трафіку, зокрема TLS, є особливо важливим, адже більшість зловмисників використовують шифрування для приховування своєї активності. Крім того, моніторинг DNS-

запитів може стати ефективним інструментом для виявлення використання DNS-тунелювання, що є поширеною технікою передачі команд до командно-контрольних серверів.

Організаційні заходи також відіграють важливу роль у запобіганні атакам. Важливо розробляти плани реагування на інциденти, які визначають послідовність дій у разі компрометації систем. Регулярні аудити безпеки допоможуть виявляти слабкі місця до того, як вони будуть експлуатовані зловмисниками (Чеховська, 2021). Крім того, впровадження принципу найменших привілеїв для доступу до даних дозволяє мінімізувати шкоду у випадку компрометації окремих облікових записів.

Оскільки багато атак мають міжнародний характер, посилення співпраці між державами є необхідним для боротьби з кібершпіонажем. Обмін інформацією про загрози, розробка уніфікованого законодавства та спільні навчання допоможуть швидше виявляти атаки й ефективніше протидіяти їм. Особливу увагу варто приділяти дослідженню новітніх технологій, таких як штучний інтелект, який може бути використаний для прогнозування атак, та квантове шифрування, що забезпечить високий рівень захисту даних у майбутньому.

Загалом, боротьба з кібершпіонажем потребує інтегрованого підходу, що включає сучасні технології, людський фактор і міжнародну співпрацю. Це дозволить не лише знизити ризики атак, але й створити стійку систему безпеки, яка зможе ефективно реагувати на нові виклики у цифровому середовищі.

Висновки

Кібершпіонаж став одним із найпотужніших інструментів впливу в сучасному світі, об’єднавши технологічну складність із геополітичними амбіціями. Аналіз кейсів операції Red October, атаки на Офіс управління персоналом США, кібератаки на Міжнародний кримінальний суд та операції "Star Blizzard" дозволив виявити спільні риси, технічні методи та наслідки таких дій. Ці атаки підкреслюють, що головною метою зловмисників є не лише викрадення конфіденційної інформації, але й тривалий контроль над інфраструктурою жертви для досягнення стратегічних цілей.

Ключовим фактором успіху атак стала експлуатація вразливостей програмного забезпечення та використання людського фактора. Фішингові кампанії, які передували більшості розглянутих атак, свідчать про необхідність посилення обізнаності співробітників і вдосконалення політик інформаційної безпеки. Водночас використання модульного шкідливого ПЗ та інноваційних методів маскування, таких як шифрування трафіку та DNS-тунелювання, вказує на зростаючу технічну досконалість зловмисників.

Проаналізовані кейси також демонструють значний вплив кібершпіонажу на міжнародну безпеку. Атаки можуть спричиняти політичну дестабілізацію, підривати довіру до міжнародних інституцій та створювати серйозні економічні втрати. Зокрема, компрометація Міжнародного кримінального суду підважує здатність глобальних інституцій захищати конфіденційність даних, що ставить під сумнів їхню легітимність.

Запобігання таким загрозам потребує інтегрованого підходу. Необхідно поєднувати технічні заходи, такі як регулярне оновлення програмного забезпечення, використання сучасних інструментів моніторингу та вдосконалення мережевої безпеки, із організаційними ініціативами, спрямованими на підвищення рівня кіберобізнаності співробітників. Крім того, важливим є розвиток міжнародної співпраці, яка дозволить ефективніше обмінюватися інформацією про загрози та координувати зусилля у сфері кібербезпеки.

Таким чином, у сучасних умовах боротьба з кібершпіонажем повинна стати пріоритетним завданням як для державних структур, так і для міжнародної спільноти. Поглиблення досліджень, впровадження інноваційних технологій і посилення міжнародного співробітництва є необхідними кроками для забезпечення довгострокової стійкості у сфері кібербезпеки. Розуміння механізмів атак та їхнього впливу є ключем до створення більш безпечного цифрового середовища.

Рекомендації

Матеріали статті є цінними для науковців, які досліджують сучасні загрози у сфері інформаційної безпеки, а також для фахівців з кібербезпеки, які займаються практичним захистом інформаційних систем. Зібрані кейси та їх аналіз можуть бути використані для розробки нових підходів до виявлення та запобігання кібершпіонажу, а також для навчання співробітників, відповідальних за безпеку корпоративних і державних мереж.

Результати дослідження становлять особливу практичну цінність для державних органів, які відповідають за національну безпеку, зокрема для підрозділів, що займаються аналізом кіберзагроз та реагуванням на інциденти. Запропоновані рекомендації з посилення технологічного захисту та підвищення рівня кіберобізнаності можуть бути інтегровані в існуючі політики інформаційної безпеки.

Крім того, стаття буде корисною для керівників приватних компаній, які стикаються із загрозами витоку інтелектуальної власності. Наведені у статті приклади атак на корпоративні структури допоможуть зрозуміти, які аспекти захисту є найбільш вразливими, та які кроки необхідно зробити для їхнього посилення.

Результати дослідження також можуть бути корисними для міжнародних організацій, які займаються регулюванням та стандартизацією кібербезпеки. Запропоновані заходи з міжнародної співпраці, обміну інформацією про загрози та гармонізації законодавства можуть сприяти створенню глобальної системи протидії кібершпіонажу.

Таким чином, стаття може слугувати базою для подальших досліджень і впровадження практичних заходів у сфері кібербезпеки, спрямованих на протидію сучасним викликам цифрового середовища.

СПИСОК ЛІТЕРАТУРИ

1. Fruhlinger J. The OPM hack explained: Bad security practices meet China's Captain America [Електронний ресурс] // CSO Online. – 2020. – Режим доступу: https://www.csoonline.com/article/566509/the-opm-hack-explained-bad-security-practices-meet-chinas-captain-america.html (Дата звернення: 16.08.2025).
2. Mitchell B. Corporate cyberespionage: Identification and prevention // EDPACS. – 2020. – Vol. 62, No. 6. – P. 1–14. – DOI: https://doi.org/10.1080/07366981.2020.1798595 (Дата звернення: 17.08.2025).
3. Abu Samah I. H., Sarip A., Ishak M. K., Shaari R., Rahim N. S. A., Abd Rashid I. M. Navigating data secrecy challenges: A study on cyberespionage intentions in the WFH era // Journal of the Institution of Engineers (India), Series B. – 2024. – Vol. 105, No. 4. – P. 941–957. – DOI: https://doi.org/10.1007/s40031-024-01022-1 (Дата звернення: 18.08.2025).
4. Діордіца І. В. Поняття та зміст кібершпигунства [Електронний ресурс] // Goal International. – 2017. – Режим доступу: https://goal-int.org/ponyattya-ta-zmist-kibershpigunstva/ (Дата звернення: 19.08.2025).
5. Bellovin S. M. Military cybersomethings // IEEE Security & Privacy. – 2013. – Vol. 11, No. 3. – P. 88–89. – Режим доступу: https://ieeexplore.ieee.org/document/6521321 (Дата звернення: 20.08.2025).
6. Zetter K. Cybersleuths uncover 5-year spy operation targeting governments [Електронний ресурс] // Wired. – 2013. – Режим доступу: https://www.wired.com/2013/01/red-october-spy-campaign/ (Дата звернення: 21.08.2025).
7. Brewster T. When a government is behind a cyberattack [Електронний ресурс] // BBC. – 2014. – Режим доступу: https://www.bbc.com/russian/business/2014/04/140423_vert_cap_when_governments_attack (Дата звернення: 22.08.2025).
8. Goodin D. Red October relied on Java exploit to infect PCs [Електронний ресурс] // Ars Technica. – 2013. – Режим доступу: https://arstechnica.com/information-technology/2013/01/massive-espionage-malware-relied-on-java-exploit-to-infect-pcs/ (Дата звернення: 23.08.2025).
9. Сааков В. У США хакери викрали дані мільйонів осіб [Електронний ресурс] // DW. – 2015. – Режим доступу: https://www.dw.com/uk/хакери-викрали-особисті-дані-близько-215-мільйона-людей-у-сша/a-18576309 (Дата звернення: 24.08.2025).
10. Маріц Д. О. "Кібератака" – війна майбутнього [Електронний ресурс]. – Київ: Інститут проблем сучасної інформації, 2015 https://doi.org/10.37750/2616-6798.2015.3(15).272792 (Дата звернення: 25.08.2025).

11. International Criminal Court. Measures taken following the unprecedented cyber-attack on the ICC [Електронний ресурс]. – 2023. – Режим доступу: https://www.icc-cpi.int/news/measures-taken-following-unprecedented-cyber-attack-icc (Дата звернення: 26.08.2025).
12. Digmelashvili T. The impact of cyberwarfare on national security [Електронний ресурс] // ResearchGate. – 2023. – Режим доступу: https://doi.org/10.29202/fhi/19/2 (Дата звернення: 27.08.2025).
13. Смишляєв С. Лондон викрив спроби кібератак на високопосадовців з боку РФ [Електронний ресурс] // DW. – 2023. – Режим доступу: https://www.dw.com/uk/velikobritania-vikrila-sprobi-kiberatak-na-visokoposadovciv-z-boku-rf/a-67659852 (Дата звернення: 28.08.2025).
14. Шлапаченко В. М. Шпигунство як діяльність зі здобування інформації // Інформаційна безпека людини, суспільства, держави. – 2020. – № 1 (17). – С. 99–109. https://doi.org/10.30890/2567-5273.2024-31-00-050 (Дата звернення: 29.08.2025).
15. Чеховська М. М. Кібершпіонаж як загроза національній безпеці // Актуальні проблеми управління інформаційною безпекою держави. – Київ: Наук.-вид. відділ НА СБ України, 2021. С. 232–234. https://goal-int.org/ponyattya-ta-zmist-kibershpigunstva/ (Дата звернення: 30.08.2025).

REFERENCES

1. J. Fruhlinger, “The OPM hack explained: Bad security practices meet China’s Captain America,” CSO Online, Aug. 5, 2020. [Online]. Available: https://www.csoonline.com/article/566509/the-opm-hack-explained-bad-security-practices-meet-chinas-captain-america.html
2. B. Mitchell, “Corporate cyberespionage: Identification and prevention part 2,” EDPACS, vol. 62, no. 6, pp. 1–14, 2020, https://doi.org/10.1080/07366981.2020.1798595.
3. I. H. Abu Samah, A. Sarip, M. K. Ishak, R. Shaari, N. S. A. Rahim, and I. M. Abd Rashid, “Navigating data secrecy challenges: A study on cyberespionage intentions in the WFH era,” Journal of the Institution of Engineers (India): Series B, vol. 105, no. 4, pp. 941–957, 2024, https://doi.org/10.1007/s40031-024-01022-1
4. I. V. Diorditsa, “The concept and content of cyberespionage,” Goal International, 2017. [Online]. Available: https://goal-int.org/ponyattya-ta-zmist-kibershpigunstva/ [in Ukrainian].
5. S. M. Bellovin, “Military cybersomethings,” IEEE Security & Privacy, vol. 11, no. 3, pp. 88–89, 2013, doi: 10.1109/MSP.2013.68.
6. K. Zetter, “Cybersleuths uncover 5-year spy operation targeting governments,” WIRED, Jan. 14, 2013. [Online]. Available: https://www.wired.com/2013/01/red-october-spy-campaign/
7. T. Brewster, “What can you do when governments attack?,” BBC, Apr. 23, 2014. [Online]. Available: https://www.bbc.com/russian/business/2014/04/140423_vert_cap_when_governments_attack [in Russian].
8. D. Goodin, “Red October relied on Java exploit to infect PCs,” Ars Technica, Jan. 15, 2013. [Online]. Available: https://arstechnica.com/information-technology/2013/01/massive-espionage-malware-relied-on-java-exploit-to-infect-pcs/
9. V. Saakov, “Hackers stole the data of millions of people in the USA,” Deutsche Welle, Jul. 10, 2015. [Online]. Available: https://www.dw.com/uk/хакери-викрали-особисті-дані-близько-215-мільйона-людей-у-сша/a-18576309 [in Ukrainian].
10. D. O. Marits, “Cyberattack – The war of the future,” Institute of Modern Information Problems, 2015. [Online]. Availablehttps://doi.org/10.37750/2616-6798.2015.3(15).272792 [in Ukrainian].
11. International Criminal Court, “Measures taken following the unprecedented cyber-attack on the ICC,” ICC, Sep. 22, 2023. [Online]. Available: https://www.icc-cpi.int/news/measures-taken-following-unprecedented-cyber-attack-icc
12. T. Digmelashvili, “The impact of cyberwarfare on national security,” ResearchGate, 2023. [Online]. Available: https://doi.org/10.29202/fhi/19/2

13. C. Smishlyayev, “London exposed cyberattacks on high-ranking officials by the Russian Federation,” Deutsche Welle, Dec. 7, 2023. [Online]. Available: https://www.dw.com/uk/velikobritania-vikrila-sprobi-kiberatak-na-visokoposadovciv-z-boku-rf/a-67659852 [in Ukrainian].
14. V. M. Shlapachenko, “Espionage as an activity of information retrieval,” Human, Society, and State Information Security, vol. 1, no. 17, pp. 99–109, 2020. https://doi.org/10.30890/2567-5273.2024-31-00-050 [in Ukrainian].
15. M. M. Chekhovska, “Cyberespionage as a threat to national security,” in Current Issues in State Information Security Management, Kyiv, Ukraine: Scientific Publishing Department of the Security Service of Ukraine, 2021, pp. 232–234. https://goal-int.org/ponyattya-ta-zmist-kibershpigunstva/ [in Ukrainian].

Tovkun Yuliia *PhD student*
Kharkiv National University of Radio Electronics
Nauky Ave, 14, Kharkiv, Kharkiv Oblast, 61166
e-mail: ytovkun@gmail.com
https://orcid.org/0009-0000-5916-2897

Methods of Cyber Espionage and Their Impact on International Security

The relevance of this research is determined by the increasing role of cyber espionage as a geopolitical tool and a means of obtaining confidential information. In the context of digitalization, government institutions, international organizations, and corporate entities are becoming key targets of cyberattacks, posing significant threats to national security and global stability.
This article aims to analyze the phenomenon of cyber espionage, particularly its technical, organizational, and social aspects, based on real-world cases. The study focuses on the use of modern attack methods, such as targeted phishing, software vulnerability exploitation, and modular malware deployment. The article seeks to identify common characteristics of cyber espionage campaigns and develop recommendations to counter such threats.
A theoretical methodological approach was used in the study, combining literature review, case analysis of attacks (Red October operation, the attack on the U.S. Office of Personnel Management, the cyberattack on the International Criminal Court, the "Star Blizzard" operation), and a systematic analysis of factors contributing to the success of cyber espionage campaigns.
The study identified key technical methods of attacks, their impact on information security, and the role of the human factor in the success of cyber espionage. Recommendations for strengthening cybersecurity were formulated, encompassing technical, organizational, and international measures.
The findings of this article are of interest to researchers, cybersecurity professionals, and governmental bodies dealing with information protection issues and can be used for developing policies to counteract cyber espionage.

Keywords: *cyber threats, espionage software, targeted phishing, information security, cyber espionage.*

Надійшла у першій редакції 29.03.2025, в останній - 11.05.2025.
The first version has been received on 29.03.2025, the final version - on 11.05.2025.

DOI: https://doi.org/10.26565/2304-6201-2025-66-09

УДК (UDC) 004.93

Ясінський Ярослав Андрійович *Аспірант кафедри комп'ютерних систем та робототехніки Навчально-наукового інституту комп'ютерних наук та штучного інтелекту; Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 4, Харків-22, Україна, 61022; e-mail: yaroslav.yasinskyi@karazin.ua;* **https://orcid.org/0009-0008-0460-5687**

Бакуменко Ніна Станіславівна *доцент кафедри комп'ютерних систем та робототехніки Навчально-наукового інституту комп'ютерних наук та штучного інтелекту; Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 4, Харків-22, Україна, 61022; e-mail: n.bakumenko@karazin.ua;* **https://orcid.org/0000-0003-3496-7167**

Порівняльний аналіз моделей YOLOv5 та MobileNetV3 для розпізнавання зображень в реальному часі

Актуальність: у сучасних умовах зростаючої потреби у швидкому й точному розпізнаванні об'єктів у реальному часі, особливо для мобільних і вбудованих систем, постає питання вибору оптимальних моделей штучного інтелекту. Порівняння легковагових та високоточних архітектур, таких як YOLOv5 і MobileNetV3, є важливим для розробки ефективних комп'ютерних зорових систем та дослідження принципів побудови гібридних моделей.

Мета: порівняння архітектур YOLOv5 і MobileNetV3 з метою аналізу ефективності для застосування у задачах розпізнавання об'єктів у реальному часі, та підтвердження, що гібридні моделі можуть підвищити ефективність виконання цих задач.

Методи дослідження: методи препроцесінгу зображень, методи навчання глибоких нейронних мереж, вимірювання точності, швидкості обробки та використання ресурсів; порівняльний аналіз результатів для оцінки ефективності моделей.

Результати: експериментальне дослідження показало, що YOLOv5 демонструє кращу загальну точність на тестовому наборі COCO, проте вимагає більше обчислювальних ресурсів. MobileNetV3, натомість, забезпечує пришвидшене виведення та ефективне функціонування на пристроях із низькою потужністю, жертвуючи частково точністю. Таким чином, обидві моделі підтвердили свою придатність для реальних застосувань, а вибір між ними залежить від конкретного балансу між швидкістю, точністю та обмеженнями платформи. Поєднання цих моделей дає кращі результати в розпізнаванні об'єктів, хоча це може збільшити розмір самої моделі та споживання ресурсів.

Висновки: у результаті дослідження проведено порівняння моделей YOLOv5, MobileNetV3 та гібридної моделі для задачі розпізнавання об'єктів. Гібридна модель продемонструвала кращу точність та баланс між швидкістю обробки і використанням ресурсів порівняно з окремими моделями. Це свідчить про доцільність використання гібридних підходів для підвищення ефективності систем комп'ютерного зору в реальних умовах. Отже, гібридна модель є перспективним напрямком для подальших досліджень і практичної реалізації.

Ключові слова: *розпізнавання зображень, комп'ютерний зір, гібридна модель, CNN, YOLOv5, MobileNetV3.*

Як цитувати: Ясінський Я. А., Бакуменко Н. С. Порівняльний аналіз моделей YOLOv5 та MobileNetV3 для розпізнавання зображень в реальному часі. *Вісник Харківського національного університету імені В. Н. Каразіна, серія Математичне моделювання. Інформаційні технології. Автоматизовані системи управління*. 2025. вип. 66. С. 90-98. https://doi.org/10.26565/2304-6201-2025-66-09

How to quote: Yasinskyi Y. A., Bakumenko, N. S., "Comparative analysis of YOLOv5 and MobileNetV3 models for real-time image recognition". *Bulletin of Kharkiv National University named after V. N. Karazin, series Mathematical modeling. Information Technology. Automated control systems,* vol. 66, pp. 90-98. https://doi.org/10.26565/2304-6201-2025-66-09[in Ukrainian]

1 Вступ

Виявлення та класифікація об'єктів залишаються фундаментальними проблемами в галузі комп'ютерного зору, особливо в сценаріях, де швидкість і точність є критично важливими, наприклад, в автономних системах, спостереженні та мобільних додатках. В даній роботі проведено порівняльний аналіз двох широко використовуваних архітектур згорткових нейронних мереж – YOLOv5 і MobileNetV3 з акцентом на їхню застосовність та ефективність у задачах розпізнавання зображень у реальному часі.

Мета роботи – огляд та порівняння архітектур нейронних мереж, розроблених для розпізнавання зображень у реальному часі та дослідження гібридної моделі створеної на базі обраних з метою підвищення ефективності.

Огляд нейромережевих моделей для виявлення об'єктів у реальному часі був проведений з акцентом на їхні принципи побудови, операційну ефективність і застосовність у практичних сценаріях. Також був проведений детальний аналіз моделей YOLOv5 і MobileNetV3, включаючи дослідження їхніх архітектурних варіацій і модифікацій, спрямованих на підвищення продуктивності в різних обчислювальних середовищах.

Оцінка ефективності моделей була проведена на тестовому наборі даних, що забезпечило стандартизовану основу для навчання та тестування моделей. Нейронні мережі навчалися в контрольованих умовах, а їхня продуктивність вимірювалася за допомогою точності, швидкості отримання висновків та ефективності використання ресурсів.

2 Опис нейромережевих архітертур

Згорткові нейронні мережі (ЗНМ) є основною архітектурою для більшості систем розпізнавання зображень, завдяки своїй здатності використовувати просторові ієрархії та інваріантність трансляції у візуальних даних. Згорткові шари, які ідентифікують локальні ознаки за допомогою фільтрів, що навчаються, поєднуються з шаром агрегації (пулінгу), який зменшує просторові розміри та обчислювальну складність. Однак традиційні глибокі архітектури ЗНМ, такі як VGG та ранні варіанти ResNet, часто виявляються занадто обчислювально дорогими для програм реального часу, що вимагає архітектурних інновацій, спеціально розроблених для підвищення ефективності [1].

Згортки, розділені по глибині, є важливою архітектурною модифікацією, яка значно зменшує обчислювальну складність, зберігаючи при цьому репрезентативну здатність. Цей підхід, популяризований архітектурами MobileNet, факторизує стандартні згортки на згортки по глибині, які застосовують фільтри до кожного вхідного каналу незалежно, а потім на точкові згортки, які об'єднують вихідні. Блоки стиснення та збудження забезпечують механізм уваги, який покращує якість представлення об'єктів без значних обчислювальних витрат. Ці блоки обчислюють вагу уваги для кожного каналу за допомогою глобального усередненого пулінгу з подальшим повним з'єднанням шарів, що дозволяє мережі адаптивно підкреслювати інформативні ознаки, пригнічуючи менш релевантні. Обчислювальні витрати залишаються мінімальними, оскільки обчислення уваги оперує стислими представленнями ознак. Методи дистиляції знань дозволяють розробляти компактні учнівські мережі, які навчаються від більших і точніших мереж вчителів. Архітектура учнів розроблена таким чином, щоб бути ефективною за своєю суттю, зберігаючи при цьому здатність фіксувати основні знання, закодовані в мережі вчителя. Такий підхід дозволяє створювати мережі, які досягають майже оптимальної точності при дотриманні обмежень реального часу [2].

Трансформери, такі як Vision Transformer (ViT) та Swin Transformer, продемонстрували високу продуктивність у різних завданнях комп'ютерного зору, ефективно моделюючи залежності на великій відстані. Незважаючи на досягнуту точність, їхні високі обчислювальні та пам'ятні вимоги створюють значну перешкоду для розгортання в малих пристроях, які обмежені розміром, вагою та обчислювальною потужністю [3].

В даній роботі MobileNet та YOLOv5 були обрані для більш детального розгляду через їхні взаємодоповнюючі переваги в задачах розпізнавання об'єктів. MobileNet добре підходить для використання на пристроях з ресурсами, що можуть бути обмеженими, завдяки своїй легкій архітектурі та ефективній продуктивності, що робить його ідеальним для додатків реального часу, де обчислювальна потужність обмежена. YOLOv5, з іншого боку, пропонує баланс між швидкістю і точністю, забезпечуючи високу продуктивність виявлення при відносно короткому часі

висновку. Разом ці моделі представляють практичний компроміс між ефективністю і точністю, що відповідає цілям дослідження.

3 Аналіз архітертури YOLOv5s

Модель YOLO (You Only Look Once) – це фундаментальна архітектура в галузі виявлення об'єктів у реальному часі, відома своїм балансом швидкості та точності. На відміну від класичних фреймворків виявлення об'єктів, які застосовують класифікатори до різних областей зображення (такі як R-CNN та його похідні), YOLO розглядає виявлення об'єктів як єдину регресійну задачу. Він безпосередньо прогнозує обмежувальні границі та ймовірності класів на базі повних зображень за одну оцінку, що забезпечує швидке виявлення, яке може бути застосоване у реальному часі. Оригінальна архітектура YOLO, представлена у 2016 році, розділяє вхідне зображення на сітку. Кожна комірка сітки відповідає за прогнозування фіксованої кількості обмежувальних рамок та показників достовірності для цих рамок, а також показників достовірності класів. Показник достовірності відображає ймовірність того, що прогнозована рамка містить об'єкт, та точність обмежувальної рамки. Прогнозуючи всі обмежувальні рамки та ймовірності класів за один прохід через нейронну мережу, YOLO досягає високої швидкості виведення, хоча спочатку мала проблеми з точністю локалізації та продуктивністю на малих об'єктах [4].

З метою подолання цих обмежень, було розроблено кілька вдосконалених версій YOLO. Алгоритм YOLOv2 запровадив кілька вдосконалень, таких як використання пакетної нормалізації, класифікатори високої роздільної здатності, опорні рамки для кращих апріорних значень обмежувальних рамок та кластеризацію розмірів для підвищення точності прогнозування рамок. Він також включив ієрархічну систему класифікації, яка дозволила йому виявляти понад 9000 категорій об'єктів, навіть з частково позначеними наборами даних.

YOLOv3 ще більше вдосконалив архітектуру, впровадивши глибшу магістральну мережу під назвою Darknet-53, яка використовує залишкові зв'язки для покращення продуктивності навчання. Ця версія підтримує багатомасштабні прогнози, що дозволяє моделі ефективніше виявляти об'єкти різних розмірів. YOLOv3 прогнозує рамки у трьох різних масштабах, використовуючи карти ознак з різної глибини мережі, значно підвищуючи точність як для малих, так і для великих об'єктів [5].

YOLOv4, базується на YOLOv3, інтегруючи досягнення комп'ютерного зору, такі як зважені залишкові з'єднання (WRC), міжетапні часткові з'єднання (CSP), крос-міні-пакетна нормалізація (CmBN) та самозмагальне навчання (SAT). YOLOv4 наголошує на збалансованому компромісі між швидкістю та точністю, оптимізуючи мережу для використання як з графічними процесорами, так і з традиційними обчислювальними середовищами [6].

YOLOv5, неофіційне продовження, розроблене спільнотою та розміщене на GitHub компанією Ultralytics, що є реалізованим у бібліотеці Python PyTorch, пропонує більш модульну конструкцію, легкість експериментів та постійні оновлення. Він включає такі функції, як автоматичне навчання обмежувальних рамок, еволюція гіперпараметрів та масштабовані розміри моделей (YOLOv5s, YOLOv5m, YOLOv5l та YOLOv5x), щоб врахувати різні компроміси між точністю та швидкістю. Незважаючи на деякі суперечки щодо його правил найменування та походження, ця архітетура є популярною серед фахівців на практиці [7].

Наступні версії, включаючи YOLOv6, YOLOv7 та YOLOv8, продовжують розширювати межі виявлення об'єктів. YOLOv6 зосереджена на продуктивності промислового рівня, оптимізуючи ефективність розгортання на периферійних пристроях. YOLOv7 інтегрує додаткові архітектурні інновації, такі як розширені мережі ефективної агрегації шарів (E-ELAN) та методи повторної параметризації моделі, для підвищення здатності до навчання без шкоди для швидкості. YOLOv8, розроблена Ultralytics, являє собою перехід до уніфікованої моделі для виявлення, сегментації та класифікації [8].

Загалом, сімейство моделей YOLO значно вплинуло на ландшафт розпізнавання зображень у реальному часі.

4 Аналіз архітертури MobileNetV3

Архітектура є MobileNetV3 — це архітектура згорткової нейронної мережі, призначена для ефективного розпізнавання зображень на мобільних та периферійних пристроях. MobileNetV3 базується на попередніх версіях MobileNetV1 та MobileNetV2, поєднуючи автоматизований

пошук нейронної архітектури (NAS) з серією оптимізаторів. Результатом стала модель, яка досягає балансу між точністю та обчислювальною ефективністю, що робить її особливо ефективною для завдань розпізнавання зображень у реальному часі за обмежених апаратних умов [9].

MobileNetV3 включає кілька основних архітектурних удосконалень, які відрізняють її від попередніх версій. Одним із найважливіших удосконалень є використання мобільної оберненої згортки вузького місця (MBConv), спочатку представленої в MobileNetV2, яка додатково вдосконалена в MobileNetV3 шляхом додавання легких механізмів уваги, відомих як блоки Squeeze-and-Excitation (SE). Ці блоки SE адаптивно перекалібрують реакції на ознаки каналів, що дозволяє моделі підкреслювати більш інформативні ознаки, одночасно пригнічуючи менш корисні [10].

Ще однією ключовою інновацією є інтеграція функцій активації hard-swish (h-swish) замість традиційних функцій ReLU або swish. Функція h-swish апроксимує активацію swish обчислювально ефективним способом, що дозволяє виконувати кращі нелінійні перетворення без значного збільшення часу обчислень. Крім того, мережа включає нелінійності, адаптовані для апаратної ефективності, що зменшує доступ до пам'яті та споживання енергії, що є важливим для розгортання на мобільних платформах [11].

MobileNetV3 постачається у двох основних варіантах, MobileNetV3-Large та MobileNetV3-Small, кожен з яких розроблений для різних рівнів доступності ресурсів та вимог до продуктивності. MobileNetV3-Large оптимізований для вищої точності та підходить для випадків використання, коли доступно більше обчислювальних ресурсів. Він містить глибшу архітектуру з більшою кількістю параметрів і зазвичай використовується в завданнях, що вимагають високої точності класифікації. І навпаки, MobileNetV3-Small оптимізований для сценаріїв з більш жорсткими обмеженнями затримки або потужності, таких як програми реального часу на мікроконтролерах або смартфонах. Він використовує більш компактну архітектуру, яка жертвує деякою точністю на користь меншого розміру моделі та швидшого виводу [12].

Архітектуру MobileNetV3 було розроблено з використанням платформо-орієнтованої стратегії NAS, що означає, що процес пошуку враховував фактичні апаратні обмеження, такі як затримка мобільних процесорів, під час проектування моделі [13].

5 Огляд датасетів

Розпізнавання об'єктів у реальному часі є критично важливою можливістю в різних програмах комп'ютерного зору, включаючи автономну навігацію, спостереження, робототехніку та доповнену реальність. Ключовим компонентом у розробці точних та ефективних моделей розпізнавання об'єктів у реальному часі є наявність високоякісних наборів даних, які відображають різноманітність, складність та динамічний характер реальних середовищ.

Серед найпоширеніших наборів даних для розпізнавання об'єктів загального призначення є набір даних COCO (Common Objects in Context). Маючи понад 330 000 зображень та понад 1,5 мільйона екземплярів об'єктів у 80 категоріях об'єктів, COCO пропонує багато анотований набір даних у натуралістичних та часто захаращених середовищах. Він підтримує як виявлення об'єктів, так і сегментацію екземплярів, що робить його модельним для розробки та порівняльного аналізу високопродуктивних моделей [14].

Аналогічно, набір даних PASCAL VOC, хоча й менший за масштабом, відрізняється стабільною якістю анотацій та своєю роллю у встановленні фундаментальних орієнтирів у цій галузі [15].

Для програм, що вимагають продуктивності в режимі реального часу на мобільних та вбудованих пристроях, особливо актуальними є набори даних, такі як ImageNet VID та YouTube-BoundingBoxes. Набір даних ImageNet VID, отриманий з більшої колекції ImageNet, містить відеопослідовності з покадровими анотаціями, що дозволяє навчати моделі, які можуть відстежувати та розпізнавати об'єкти з часом. Цей часовий вимір має вирішальне значення для розробки алгоритмів, які повинні надійно працювати в динамічних середовищах. Аналогічно, набір даних YouTube-BoundingBoxes містить мільйони позначених відеокадрів, отриманих з YouTube, що забезпечує реальну мінливість освітлення, руху та оклюзії, що є важливим для створення надійних систем розпізнавання в режимі реального часу [16].

Спеціалізовані набори даних, наприклад, BDD100K, був розроблений для підтримки програм автономного водіння. Ці набори даних пропонують анотовані відеодані, отримані з транспортних

засобів, що працюють у різних дорожніх умовах та середовищах. Вони надають достовірну інформацію для різноманітних завдань, включаючи виявлення об'єктів, відстеження, виявлення смуг руху та семантичну сегментацію. Такі набори даних є важливими для навчальних моделей, які можуть працювати в режимі реального часу з урахуванням обмежень автомобільного обладнання та критичних вимог безпеки [17].

Зрештою, моделі, обрані в цій роботі, базується на наборі даних COCO що пропонує великий, різноманітний набір зображень з детальними анотаціями для кількох категорій об'єктів, що дозволяє моделям добре узагальнювати та точно й швидко виявляти широкий спектр об'єктів у складних реальних сценах.

6 Огляд методів створення гібридних моделей

Гібридні моделі для розпізнавання зображень у реальному часі поєднують різні архітектури нейронних мереж та алгоритмічні стратегії, щоб використовувати унікальні сильні сторони кожної з них, прагнучи досягти як високої точності, так і швидкої обробки, придатної для застосувань у реальному часі. Існує кілька методів створення таких гібридних моделей, кожен з яких вирішує конкретні проблеми в задачах розпізнавання зображень.

Одним із поширених підходів є архітектурне об'єднання, де різні типи нейронних мереж інтегруються в одну модель. Наприклад, згорткові нейронні мережі, які чудово виявляють просторові ознаки із зображень, можна поєднувати з трансформаторами, які вміло моделюють довгострокові залежності та механізми уваги. Було показано, що ця гібридна архітектура ЗНМ-Трансформер підвищує як точність, так і швидкість розпізнавання об'єктів у реальному часі, особливо в складних середовищах, таких як сцени в приміщенні зі змінним освітленням та захараченим фоном. Компонент ЗНМ зазвичай обробляє початкове вилучення ознак, тоді як модуль трансформера зосереджується на уточненні цих ознак за допомогою уваги, дозволяючи моделі пріоритезувати критичну інформацію для виявлення та класифікації [18].

Об'єднання ознак – це ще один ключовий метод, де ознаки, отримані з різних моделей або модальностей, об'єднуються перед тим, як робити прогноз. Наприклад, ознаки з попередньо навченої EfficientNet (тип CNN) можна подавати в детекторну головку YOLO, поєднуючи ефективне вилучення ознак EfficientNet з можливостями виявлення в реальному часі YOLO. Цей підхід, як продемонстровано в гібриді E-YOLO (EfficientNet + YOLO), зменшує розмір моделі та обчислювальне навантаження без шкоди для точності виявлення, що робить його добре придатним для застосувань у реальному часі на периферійних пристроях [19].

Також використовується об'єднання рішень, де прогнози з кількох моделей поєднуються за допомогою таких стратегій, як голосування або зважене усереднення. Цей метод підвищує стійкість шляхом агрегування сильних сторін різних архітектур, що особливо корисно в сценаріях з різноманітними або зашумленими даними.

Крім того, гібридні моделі можуть поєднувати CNN з рекурентними нейронними мережами (RNN) для завдань, які потребують як просторового, так і часового розуміння, таких як аналіз відео. Тут CNN аналізують просторові ознаки з кожного кадру, тоді як RNN фіксують часові залежності між кадрами, забезпечуючи надійне розпізнавання в реальному часі в динамічних сценах [20].

Загалом, гібридні моделі для розпізнавання зображень у реальному часі будуються шляхом архітектурного об'єднання (наприклад, CNN-Transformer), об'єднання ознак (наприклад, EfficientNet + YOLO) та об'єднання рішень (ансамблеві методи). Ці стратегії дозволяють розробляти моделі, які є одночасно точними та ефективними, здатними відповідати вимогам програм реального часу в різних середовищах та варіантах використання.

7 Порівняння архітертур MobileNetV3 та YOLOv5

У таблиці представлено порівняльний аналіз різних моделей глибокого навчання. Результати подані у табл. 1:

Таблиця 1. Порівняння результатів продуктивності та ефективності моделей виявлення об'єктів
Table 1. Comparison of performance and efficiency results of object detection models

Модель	Точність	Розмір моделі (MB)	Середнє використання RAM (GB)	Затримка у часі
MobileNetV3-Small (TensorFlow-Lite)	0.6832	9.72	1.91	15.5 ms
MobileNetV3-Large	0.7342	15.43	2.1	21.2 ms
YOLOv5 (original, baseline)	0.8681	92.75	7.91	101.7 ms
YOLOv5 + MobileNetV3-small	0.9041	120.7	7.47	69.7 ms

Оригінальна YOLOv5 показує високу точність (0,8681), але є ресурсозатратною (92,75 МБ, 7,91 ГБ пам'яті, 101,7 мс). Гібридна модель YOLOv5 з MobileNetV3-small покращує точність до 0,9041 при зменшенні ресурсоспоживання (120,7 МБ, 7,47 ГБ, 69,7 мс).

Автономні MobileNetV3 моделі мають нижчу точність (0,6832 для Small, 0,7342 для Large), але є значно легшими та швидшими, особливо MobileNetV3-Small (1,91 ГБ, 15,5 мс). Вони підходять для застосувань з обмеженими ресурсами, де критична швидкість обробки.

MobileNetV3, відомий своєю спрощеною конструкцією та використанням згорток, що розділяються за глибиною, слугує ефективним екстрактором ознак, коли використовується як основа YOLOv5. Ця інтеграція зменшує загальний розмір моделі та обчислювальне навантаження без суттєвого погіршення продуктивності. Наприклад, заміна стандартного ядра YOLOv5 на MobileNetV3 призвела до того, що моделі стали значно меншими і швидшими, що полегшило обробку даних в реальному часі на пристроях з обмеженими ресурсами.

Загалом, результати показують, що хоча моделі MobileNetV3 є високоефективними, вони жертвують точністю. Оригінальна модель YOLOv5 пропонує високу продуктивність, але за рахунок вимог до ресурсів. Комбінація YOLOv5 з MobileNetV3-small досягає найкращого компромісу, покращуючи точність порівняно з базовою моделлю YOLOv5 і водночас підвищуючи обчислювальну ефективність, що робить її ефективним рішенням для сценаріїв, які вимагають як високої точності, так і ефективного використання ресурсів.

8 Висновки

У проведеному дослідженні здійснено комплексний порівняльний аналіз двох провідних архітектур нейронних мереж для розпізнавання об'єктів у реальному часі – YOLOv5 та MobileNetV3, а також досліджено ефективність їх гібридного поєднання. Результати експериментального дослідження дозволяють сформулювати наступні висновки.

На основі отриманих результатів можна сформулювати практичні рекомендації для різних сценаріїв застосування. Для застосувань, де критичним є мінімальне споживання ресурсів (IoT-пристрої, мобільні додатки з жорсткими обмеженнями), оптимальним вибором залишається MobileNetV3-Small. Для систем, що вимагають високої точності при помірних обмеженнях на ресурси, рекомендується використання гібридної моделі. Оригінальна YOLOv5 доцільна у випадках, коли обчислювальні ресурси не є критичним обмеженням, а пріоритетом є баланс точності та часу розробки.

Дослідження підтверджує ефективність архітектурного об'єднання як методу створення гібридних моделей для комп'ютерного зору. Використання MobileNetV3 як backbone-мережі для YOLOv5 демонструє можливість збереження переваг обох архітектур при мітигації їх недоліків. Це відкриває перспективи для подальших досліджень у напрямку покращення гібридних архітектур.

Слід зазначити певні обмеження проведеного дослідження. Дослідження проводилося на наборі даних COCO, що може обмежувати узагальнюваність результатів на інші домени. Також гібридна модель, незважаючи на покращення ефективності, все ще характеризується збільшеним розміром порівняно з окремими компонентами, що може бути критичним для деяких застосувань.

Проведене дослідження демонструє, що гібридні моделі представляють перспективний напрямок розвитку архітектур нейронних мереж для визначення об'єктів у реальному часі.

Поєднання YOLOv5 та MobileNetV3 не лише забезпечує покращення точності, але й оптимізує використання обчислювальних ресурсів, що має важливе значення для практичного впровадження систем комп'ютерного зору в реальних умовах.

Отримані результати підтверджують гіпотезу про те, що архітектурне об'єднання різних типів нейронних мереж може ефективно використовувати унікальні переваги кожної архітектури, створюючи рішення, що переважають окремі компоненти за ключовими показниками продуктивності.

Вибір архітектур нейронних мереж для розпізнавання зображень у реальному часі визначається компромісом між обчислювальною ефективністю, точністю та швидкістю.

СПИСОК ЛІТЕРАТУРИ

1. Younesi, A., Ansari, M., Fazli, M., Ejlali, A., Shafique, M., & Henkel, J. (2024). A comprehensive survey of convolutions in deep learning: Applications, challenges, and future trends. IEEE Access, 12, 41180-41218. https://doi.org/10.48550/arXiv.2402.15490
2. Tu, C. H., Lee, J. H., Chan, Y. M., & Chen, C. S. (2020, July). Pruning depthwise separable convolutions for mobilenet compression. In 2020 *international joint conference on neural networks* (IJCNN) (pp. 1-8). IEEE.
3. Zhang, W., Huang, Z., Luo, G., Chen, T., Wang, X., Liu, W., ... & Shen, C. (2022). Topformer: Token pyramid transformer for mobile semantic segmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 12083-12093). https://doi.org/10.48550/arXiv.2204.05525
4. YOLO Object Detection Explained: Evolution, Algorithm, and Applications. URL: https://encord.com/blog/yolo-object-detection-guide/ (дата звернення: 22.03.2025).
5. Реалізація Keras (TF backend) виявлення об'єктів, YoloV3. URL: https://github.com/xiaochus/YOLOv3/tree/master (дата звернення: 22.03.2025).
6. YOLOv4: Високошвидкісне та точне виявлення об'єктів. URL: https://docs.ultralytics.com/models/yolov4/ (дата звернення: 22.03.2025).
7. Офіційний репозиторій YOLOv5. URL: https://github.com/ultralytics/yolov5 (дата звернення: 22.03.2025).
8. Занурення в виявлення об'єктів, YOLO. URL: https://www.picsellia.com/post/a-dive-into-yolo-object-detection (дата звернення: 22.03.2025).
9. Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., ... & Adam, H. (2019). Searching for mobilenetv3. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1314-1324). https://doi.org/10.48550/arXiv.1905.02244
10. Channel Attention and Squeeze-and-Excitation Networks (SENet). URL: https://www.digitalocean.com/community/tutorials/channel-attention-squeeze-and-excitation-networks (дата звернення: 22.03.2025).
11. Функція Hardswish. URL: https://www.paddlepaddle.org.cn/documentation/docs/en/api/paddle/nn/functional/hardswish_en.html (дата звернення: 22.03.2025).
12. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861. https://doi.org/10.48550/arXiv.1704.04861
13. MobileNet V3. URL: https://mmclassification.readthedocs.io/en/dev-1.x/papers/mobilenet_v3.html (дата звернення: 22.03.2025).
14. COCO, Набір даних для виявлення, сегментації та субтитрування великомасштабних об'єктів. URL: https://cocodataset.org/#overview (дата звернення: 22.03.2025).
15. Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. International journal of computer vision, 88, 303-338. https://doi.org/10.1007/s11263-009-0275-4

16. YouTube-Bounding Boxes Dataset. URL: https://research.google.com/youtube-bb/ (дата звернення: 22.03.2025).
17. BDD100K: A Large-scale Diverse Driving Video Database. URL: https://bair.berkeley.edu/blog/2018/05/30/bdd/ (дата звернення: 22.03.2025).
18. Agga, A., Abbou, A., Labbadi, M., El Houm, Y., & Ali, I. H. O. (2022). CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production. Electric Power Systems Research, 208, 107908. DOI:10.1016/j.epsr.2022.107908
19. Dai, Y., Gieseke, F., Oehmcke, S., Wu, Y., & Barnard, K. (2021). Attentional feature fusion. In Proceedings of the IEEE/CVF winter conference on applications of computer vision https://doi.org/10.48550/arXiv.2009.14082 (pp. 3560-3569).
20. Dhruv, P., & Naskar, S. (2020). Image classification using convolutional neural network (CNN) and recurrent neural network (RNN): A review. Machine learning and information processing: proceedings of ICMLIP 2019, 367-381. DOI:10.1007/978-981-15-1884-3_34

REFERENCES

1. Younesi, A., Ansari, M., Fazli, M., Ejlali, A., Shafique, M., & Henkel, J. (2024). A comprehensive survey of convolutions in deep learning: Applications, challenges, and future trends. IEEE Access, 12, 41180-41218. https://doi.org/10.48550/arXiv.2402.15490
2. Tu, C. H., Lee, J. H., Chan, Y. M., & Chen, C. S. (2020, July). Pruning depthwise separable convolutions for mobilenet compression. In 2020 *international joint conference on neural networks* (IJCNN) (pp. 1-8). IEEE.
3. Zhang, W., Huang, Z., Luo, G., Chen, T., Wang, X., Liu, W., ... & Shen, C. (2022). Topformer: Token pyramid transformer for mobile semantic segmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 12083-12093). https://doi.org/10.48550/arXiv.2204.05525
4. YOLO Object Detection Explained: Evolution, Algorithm, and Applications. URL: https://encord.com/blog/yolo-object-detection-guide/ (date of last access: 22.03.2025).
5. Keras(TF backend) implementation of YoloV3 objects detection. URL: https://github.com/xiaochus/YOLOv3/tree/master (date of last access: 22.03.2025). [in Ukrainian]
6. YOLOv4: High-Speed and Precise Object Detection. URL: https://docs.ultralytics.com/models/yolov4/ (date of last access: 22.03.2025).
7. YOLOv5 Official Repository. URL: https://github.com/ultralytics/yolov5 (date of last access: 22.03.2025). [in Ukrainian]
8. A dive into YOLO object detection. URL: https://www.picsellia.com/post/a-dive-into-yolo-object-detection (date of last access: 22.03.2025).
9. Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., ... & Adam, H. (2019). Searching for mobilenetv3. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1314-1324). https://doi.org/10.48550/arXiv.1905.02244
10. Channel Attention and Squeeze-and-Excitation Networks (SENet). URL: https://www.digitalocean.com/community/tutorials/channel-attention-squeeze-and-excitation-networks (date of last access: 22.03.2025).
11. Function Hardswish. URL: https://www.paddlepaddle.org.cn/documentation/docs/en/api/paddle/nn/functional/hardswish_en.html (date of last access: 22.03.2025). [in Ukrainian]
12. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861. https://doi.org/10.48550/arXiv.1704.04861
13. MobileNet V3. URL: https://mmclassification.readthedocs.io/en/dev-1.x/papers/mobilenet_v3.html (date of last access: 22.03.2025).
14. COCO, large-scale object detection, segmentation, and captioning dataset. URL: https://cocodataset.org/#overview (date of last access: 22.03.2025).

15. Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. International journal of computer vision, 88, 303-338. https://doi.org/10.1007/s11263-009-0275-4
16. YouTube-Bounding Boxes Dataset:. URL: https://research.google.com/youtube-bb/ (date of last access: 22.03.2025).
17. BDD100K: A Large-scale Diverse Driving Video Database. URL: https://bair.berkeley.edu/blog/2018/05/30/bdd/ (date of last access: 22.03.2025).
18. Agga, A., Abbou, A., Labbadi, M., El Houm, Y., & Ali, I. H. O. (2022). CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production. Electric Power Systems Research, 208, 107908. DOI:10.1016/j.epsr.2022.107908
19. Dai, Y., Gieseke, F., Oehmcke, S., Wu, Y., & Barnard, K. (2021). Attentional feature fusion. In Proceedings of the IEEE/CVF winter conference on applications of computer vision https://doi.org/10.48550/arXiv.2009.14082 (pp. 3560-3569).
20. Dhruv, P., & Naskar, S. (2020). Image classification using convolutional neural network (CNN) and recurrent neural network (RNN): A review. Machine learning and information processing: proceedings of ICMLIP 2019, 367-381. DOI:10.1007/978-981-15-1884-3_34

Yasinskyi Yaroslav
Ph.D student;
V.N. Karazin Kharkiv National University
Svobody Sq 4, Kharkiv, Ukraine, 61022
e-mail: yaroslav.yasinskyi@karazin.ua;
https://orcid.org/0009-0008-0460-5687

Bakumenko Nina
Candidate of Technical Sciences; Associate Professor of Computer Systems and Robotics Department, Education and Research Institute of Computer Sciences and Artificial Intelligence;
V.N. Karazin Kharkiv National University
Svobody Sq 4, Kharkiv, Ukraine, 61022
e-mail: n.bakumenko@karazin.ua;
https://orcid.org/0000-0003-3496-7167

Comparative analysis of YOLOv5 and MobileNetV3 models for real-time image recognition

Relevance: With the growing need for fast and accurate real-time object recognition, especially for mobile and embedded systems, the question of choosing the optimal AI models arises. Comparisons of lightweight and high-precision architectures such as YOLOv5 and MobileNetV3 are important for developing efficient computer vision systems and exploring the principles of hybrid model construction.
Purpose: Comparison of the YOLOv5 and MobileNetV3 architectures to analyze the efficiency for real-time object recognition applications, and to confirm that hybrid models can improve the efficiency of these tasks.
Research methods: image preprocessing methods, deep neural network training methods, measurement of accuracy, processing speed, and resource usage; comparative analysis of results to assess model effectiveness.
Results: An experimental study showed that YOLOv5 demonstrates better overall accuracy on the COCO test suite, but requires more computing resources. MobileNetV3, on the other hand, provides faster output and efficient functioning on low-power devices, sacrificing accuracy in part. As such, both models have proven their suitability for real-world applications, and the choice between them depends on the specific balance between speed, accuracy, and platform limitations. Combining these models gives better results in object recognition, although this may increase the size of the model itself and resource consumption.
Conclusions: As a result of the study, the YOLOv5, MobileNetV3 and hybrid models for the object recognition problem were compared. The hybrid model demonstrated better accuracy and balance between processing speed and resource utilization than individual models. This indicates the feasibility of using hybrid approaches to improve the efficiency of computer vision systems in real conditions. Therefore, the hybrid model is a promising direction for further research and practical implementation.
Keywords: *image recognition, computer vision, hybrid model, CNN, YOLOv5, MobileNetV3*

.

Надійшла у першій редакції 21.03.2025, в останній - 17.05.2025.
The first version has been received on 21.03.2025, the final version - on 17.05.2025.

ВІСНИК ХАРКІВСЬКОГО НАЦІОНАЛЬНОГО УНІВЕРСИТЕТУ
імені В.Н. Каразіна

серія **«Математичне моделювання. Інформаційні технології. Автоматизовані системи управління»**

Випуски даної серії розповсюджуються у академічних та наукових колах України та за її межами з метою оперативного висвітлення досліджень у таких актуальних галузях: математичне та комп'ютерне моделювання, обчислювальний експеримент, теорія і прикладні методи обробки інформації, захист інформації, програмно-апаратні системи інформаційного або управляючого призначення, застосування математичного моделювання та системного аналізу у високих, наукоємних технологіях, враховуючи технології створення програмної продукції. Приймаються роботи, що відносяться до напрямів фізико-математичних і технічних наук (бажаний об'єм 6-18 сторінок). Усі рукописи рецензуються.

Примітка. Протягом 2025-26 рр. редакційна колегія при інших рівних умовах надаватиме перевагу роботам, що представлені англійською мовою, якщо стаття отримала схвалення при рецензуванні.

Офіційний сайт **http://periodicals.karazin.ua/mia**
http://mia.univer.kharkov.ua
Email: journal-mia@karazin.ua

Bulletin of V.N. Karazin Kharkiv National University

series **«Mathematical modeling. Information technology. Automated control systems»**

This series are distributed in academic and scientific circles of Ukraine and abroad for the purpose of timely coverage of research in the following topical areas: mathematical and computer modeling, computational experiment, theory and applied methods of information processing, information protection, software and hardware systems of control and information management, applications of mathematical modeling and system analysis in high, science-intensive technologies, including technologies of software products creation. Articles belonging to the fields of physical, mathematical and technical sciences are accepted (recommended length 6-18 pages). All submissions are peer-reviewed.

Note. For the years 2025-26, all other conditions being equal, the Editorial Board will give preference to articles submitted in English and approved by the peer-review.

Official website **http://periodicals.karazin.ua/mia**
http://mia.univer.kharkov.ua
Email: journal-mia@karazin.ua

Наукове видання

Вісник Харківського національного університету імені В. Н. Каразіна

Серія Математичне моделювання. Інформаційні технології. Автоматизовані системи управління

Випуск 66

Збірник наукових праць

Українською та англійською мовами

Комп'ютерне верстання О. О. Афанасьєва

Підписано до друку 2.07.2025 р.
Формат 60x84/8. Папір офсетний. Друк цифровий.
Ум. друк. арк. – 11,8.
Обл.– вид. арк. – 14,8.
Наклад 50 пр. Зам. № 31/2025
Безкоштовно

Видавець і виготовлювач
Харківський національний університет імені В. Н. Каразіна
61022, м. Харків, майдан Свободи, 4
Свідоцтво суб'єкта видавничої справи ДК №3367 від 13.01.09

Видавництво Харківський національний університет імені В. Н. Каразіна
тел.: 705-24-32