FOUNDERS

Kyiv National University of Construction and Architecture (Ukraine)

Certificate of the State Registration KB № 24366-14206 P dated 13.02.2020. Decision of the National Council of Ukraine on Television and Radio Broadcasting No. 223 dated 01.02.2024.

ISSN 2415-8550 (print) ISSN 2415-8569 (online)

DOI: 10.32347/st.2025.3
It is published twice a year

INTERNATIONAL SCIENTIFIC JOURNAL

SMART TECHNOLOGIES

INDUSTRIAL AND CIVIL ENGINEERING

Founded in August 2015

Issue 3(16), 2025

INFLUENCE OF WATER ON AN ENVIRONMENT AND INNOVATIVE TECHNOLOGIES

Natural sciences
Mathematics and statistics
Information technologies
Mechanical and electric engineering
Automation and instrument-making
Production and technologies
Architecture and building

Scientific professional publication of Ukraine (Category “Б”)
Orders of the Ministry of Education and Science
30.11.2021 № 1290, 07.04.2022 № 320, 01.02.2024 № 223 (specialties – F2, F3, F7, F5, F6)

Approved by Academic Council of Kyiv National University of Construction and Architecture
Mart 28, 2025, Protocol No.31

Languages of the publications English, Ukrainian

Content

Дослідження сучасного стану технологій квантового розподілу секретних ключів

Сергій Лєнков[1], Володимир Джулій[2], Ігор Муляр[3], Євген Лєнков[4]

[14]Військовий інститут Київського національного університету імені Тараса Шевченка
Юлії Здановської, 81, Київ, Україна, 03189
[1] lenkov_s@ukr.net https://orcid.org/0000-0001-7689-239
[2]Хмельницький національний університет
вул. Інститутська, 11, Хмельницький, Україна, 29000
[2]dzhuliivm@khmnu.edu.ua, http://orcid.org/0000-0003-1878-4301,
[3]muliariv@khmnu.edu.ua, 0000-0002-6659-605X,
[4]Міністерство енергетики України
вул. Хрещатик, 30, Київ, Україна, 01001
[4]lenkov85es@gmail.com, https://orcid.org/0000-0001-5819-2656

Received 03.02.2025, accepted 29.03.2025
https://doi.org/10.32347/st.2025.3.1201

Анотація. Робота присвячена дослідженню задач використання технології квантового розподілу секретних ключів, а саме задачам використання та формування нових секретних ключів, ґрунтуючись на виробленні квантових ключів та розподілу нових ключів у мережа складних топологій.

В системах захищеного зв'язку інформація передається мережами загального користування, і таким чином, доступна потенційному порушнику для проведення різного роду атак. Потенційний порушник здатний зберігати шифровані дані, що передаються по каналах мережі, а спробу дешифрування робити пізніше, коли технічні засоби та атаки на алгоритми захисту інформації дозволять, за прийнятний час, провести дешифрування. У цьому полягає основний принцип "Store now, decrypt later", що є однією з проблем систем захисту інформації [1, 2,19].

Для блокових шифрів існує параметр - навантаженням на ключ, який обмежує об’єм інформації, яку може обробити один ключ із використанням алгоритму захисту даних. NIST рекомендує часові рамки використання секретних ключів, що обчислюються роками [3,6].

У сфері інформаційної безпеки є суттєвіші обмеження навантаження на ключ, пов'язані з числом блоків, які можна на одному ключі обробити, обумовлені атаками на блокові шифри. При даних обмеженнях виникає необхідність регулярної зміни використовуваного секретного ключа, та пошук варіантів генерації та доставки ключів між двома вузлами мережі, що організовують захищений канал передачі інформації.

Зміна секретного ключа на новий ключ, незалежний від попереднього, дозволяє захистити передачу конфіденційної інформації при компрометації використовуваного ключа.

Сергій Лєнков
д.т.н. професор, головний науковий співробітник

Володимир Джулій
Доцент кафедри кібербезпеки

Ігор Муляр
к.т.н., доцент, ст. викладач кафедри кібербезпеки

Євген Лєнков
Провідний спеціаліст

Технологія створення квантового комп'ютера дозволяє атакувати ефективно відомі алгоритми генерації секретних ключів, що базуються на факторизації великих чисел, чи складності обчислення дискретного логарифму. Квантовий алгоритм Шора [3-5,15] дозволяє вирішувати дані задачі за поліноміальний час, тому необхідно шукати альтернативні варіанти вирішення задачі регулярної зміни в каналі мережі секретного ключа в парі вузлів.

Важливо враховувати необхідність забезпечення якості Perfect forward secrecy, при компрометації майстер-ключа, всі наступні секретні ключі повинні бути не скомпрометованими. Для досягнення такої властивості сучасних алгоритмів шифрування є періодичне оновлення майстер-ключів.

Для запобігання накопиченню інформації, для успішного здійснення атак порушником та дотримання вимоги навантаження на секретний ключ необхідно вирішити задачу регулярної зміни секрета, та регулярної генерації/доставки ключів у вузли мережі.

З урахуванням нагальних потреб, що на сьогоднішній день інтенсивно розвиваються, у захищеній та швидкісній передачі інформації, та не менш, стрімкому зростанні технічних можливостей, існує потреба розробляти заздалегідь варіанти, які забезпечуть при передачі інформації її захищеність. Регулярна зміна, у географічно рознесених системах захисту даних, загальних секретних ключів, у разі забезпечення незалежності секретних ключів дозволяє вирішити успішно поставлену задачу.

Ключові слова: інформаційна безпека, вразливості, атаки, квантовий секретний ключ, протокол квантового розподілу.

ВСТУП

Серед сучасних технологій генерації та доставки секретних ключів можна виділити два підходи: розподіл секретних ключів з використанням технології квантового розподілу ключів; генерація секретних ключів між двома учасниками мережі з використанням пост-квантових алгоритмів.

Другий варіант полягає в складних математичних обчисленнях. Сучасний рівень опрацювання пост-квантових алгоритмів недостатній для їхнього впровадження в існуючі системи. Існує небезпека, що після створення ефективного квантового комп'ютера будуть запропоновані квантові алгоритми, що дозволяють ефективно вирішувати задачі, використовувані в пост-квантових алгоритмах, створять нові вектори атак на пост-квантові алгоритми та вимагатимуть нові підходи регулярної доставки секретних ключів.

Проблема розподілу секретних ключів між парами вузлів мережі є актуальною науковою проблемою, яка потребує вирішення в умовах розробки нових загроз, пов'язаних зі створенням квантового комп'ютера.

Існуючі на сьогодні рішення розроблялися до розвитку квантових обчислень і комунікацій, не враховують із застосуванням квантових пристроїв нові вектори атак.

Варіант розподілу секретних ключів із застосуванням технології квантового розподілу є найперспективнішим для вирішення поставленої проблеми. Генерація ідентичних квантових секретних ключів на двох кінцях вузлів мережі квантового каналу ґрунтується на інших фізичних принципах, що надає можливість запобіганню проведення ефективних атак (загроз) із застосуванням квантового комп'ютера, також не дозволяє проводити атаки із накопиченням об’єму захищеної інформації.

Технологія квантового розподілу ключів має істотні обмеження, які необхідно враховувати під час проектування систем генерації/доставки секретних ключів з урахуванням квантового розподілу. Технологія квантового розподілу ключів здійснює розподіл ключів на обмеженій відстані (визначається протоколом квантового розподілу ключів) та якістю квантового каналу.

Для протоколів на InGaAs/InP граничним є відстань близько 100 км [5,6,13,14].

Найпростішим прикладом із застосуванням квантового розподілу ключів є використання двох екземплярів квантової апаратури, з'єднаних квантовим каналом. Для вирішення задачі обмеження довжини квантового каналу апаратура квантового розподілу ключів об'єднується, у так звані,

мережі квантового розподілу ключів. Для отримання дійсно квантового ключа між двома вузлами мережі, які не пов'язані квантовим каналом, необхідно використання технології квантових повторювачів, які на сьогодні далеко від практичної реалізації.

Реалізований на сьогодні підхід, полягає у побудові мереж квантового розподілу ключів з урахуванням довірених проміжних вузлів мережі. При такому підході секретні квантові ключі виробляються тільки між вузлами мережі з'єднаними безпосередньо квантовим каналом.

На інші вузли квантової мережі квантові секретні ключі передаються під захистом інших згенерованих секретних квантових ключів.

У мережах квантового розподілу ключів необхідно передавати не тільки ключову інформацію із застосуванням секретних квантових ключів по ланцюжку вузлів квантової мережі, але також здійснювати дані процеси в мережах зв'язку змішаної топології загального користування. Змішана топологія - топологія зв'язків вузлів квантовими каналами, так і топологія класичних зв'язків між вузлами мережі (мережа зв'язків загального користування), також захищені канали взаємодії між вузлами квантової мережі, при цьому граф, що відображає зв'язки квантовими каналами, має бути зв'язковим.

З урахуванням нагальних потреб, що на сьогоднішній день інтенсивно розвиваються, у захищеній та швидкісній передачі інформації, та не менш стрімкому зростанні технічних можливостей, існує потреба розробляти заздалегідь варіанти, які забезпечуть при передачі інформації зберегти її захищеність.

Регулярна зміна, у географічно рознесених системах захисту даних, загальних секретних ключів, у разі забезпечення незалежності секретних ключів дозволяє вирішити успішно поставлену задачу.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ ТА ПОСТАНОВКА ЗАДАЧІ

Технологія квантового розподілу ключів-отримання ідентичних випадкових послідовностей абонентами, отриманих із використанням передачі між абонентами деякої інформації, з використанням квантових частинок. Абоненти використовують квантові пристрої, що реалізують протокол квантового розподілу ключів.

Протокол квантового розподілу ключів включає: підготовку та перетворення інформаційних квантових станів в пристрої, який повинен мати джерело квантових станів; спосіб передачі квантовим каналом інформаційних квантових станів; спосіб інтерпретації, реєстрації та перетворення результатів вимірювань на з'єднаному пристрої; спосіб обробки отриманої за результатами вимірювань, послідовності, із застосуванням відкритого автентифікованого каналу зв'язку. Результатом роботи протоколу квантового розподілу ключів є отримання секретного квантового ключа, ідентичного на обох вузлах мережі квантового каналу.

Квантовий комплекс, що реалізує протокол квантового розподілу ключів, є апаратура з двох пристроїв, з'єднаних квантовим каналом. Архітектура комплексу наведена на Рис. 1. Клієнт квантового розподілу ключів містить джерело (генератор) одиночних фотонів. Під'єднаний пристрій, що містить приймач (детектор) одиночних фотонів, називають

Сервер квантового розподілу ключів. Кожен із вузлів квантового каналу містить датчик випадкових чисел. Таким чином, можна отримати випадкову істинну послідовність, з якої в подальшому буде формуватиметься квантовий секретний ключ.

Сервер та Клієнт квантового розподілу ключів з'єднані двома логічними каналами: класичним та квантовим. Квантовий канал, зазвичай реалізується оптоволокном, призначений для передачі фотонів, квантових інформаційних станів. Існують системи квантового розподілу ключів, в

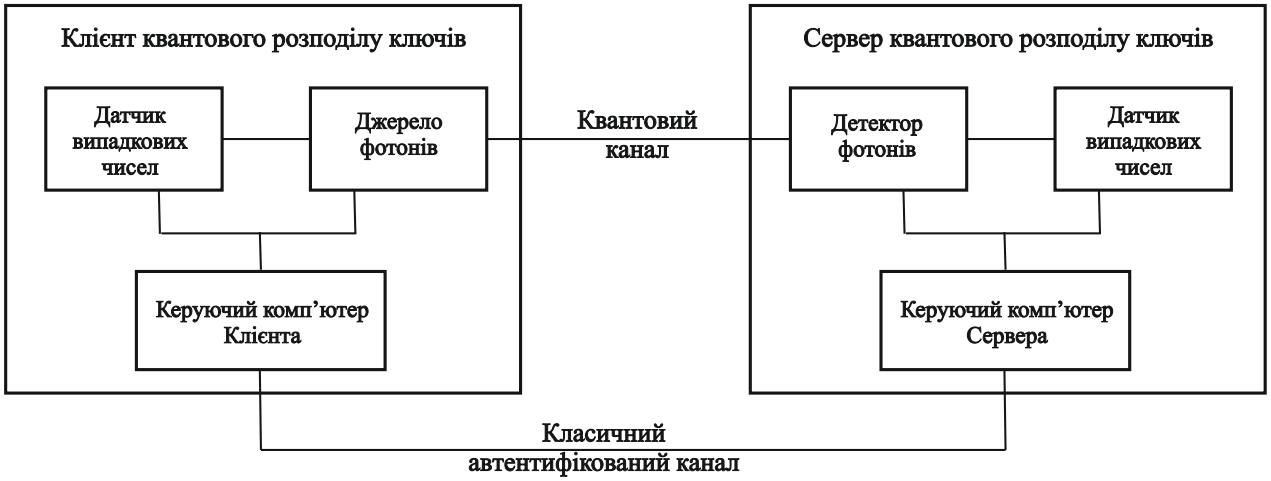


Рис. 1. Спрощена архітектура комплексу квантової апаратури

яких використовується повітряне середовище передачі, як квантовий канал, проте вони зна-ходяться ще в стадії лабораторних розробок [5,10,15,17,20].

Важливою особливістю технології квантового розподілу ключів є повна доступність потенційого порушника до квантового каналу, цей канал не захищається і не контролюється від доступу порушника до квантового каналу

Крім квантового каналу Клієнт і Сервер квантового розподілу ключів повинні з'єднуватися класичною лінією зв'язку мережі, в якій реалізується класичний логічний канал зв'язку - класичний автентифікований канал.

До автентифікованого каналу пред'являються вимоги щодо забезпечення автентифікації відправника інформації та цілісності даних.

Реальна система квантового розподілу ключів, також має логічний службовий канал, що з'єднує Сервер та Клієнт квантового розподілу ключів, по якому передаються дані моніторингу апаратури та управління, команди, які не пов'язані безпосередньо з протоколом квантового розподілу ключів. Залежно від конкретної реалізації системи квантового розподілу ключів, склад цих даних і команд може вимагати забезпечення не тільки цілісності інформації, також її конфіденційності. Для функціонування системи квантового розподілу ключів в апаратуру необхідно завантажити попередньо розподілені секретні ключі, які будуть використанні при побудові класичного автентифікованого каналу мережі до першої успішної генерації достатньої кількості квантових секретних ключів.

Одна ітерація реалізації протоколу квантового розподілу ключів - сеанс квантового розподілу ключів. Сеанс квантового розподілу ключів складається з наступних етапів: підготовка квантового каналу; передача одиночних фотонів квантовим каналом мережі; пост-обробка переданої послідовності даних.

Перші два етапи задіють квантовий канал мережі. Результатом передачі квантовим каналом є сирий ключ у обох пристроїв.

Останній етап протоколу квантового розподілу ключів виконується з використанням класичного автентифікованого каналу без квантового каналу мережі.

Постобробка включає наступні три підетапи: узгодження базисів виміру на стороні Сервера з базисами шифрування квантових станів на стороні Клієнта. Під час узгодження позицій сирого ключа, у яких базиси не співпали, відкидаються, в результаті сирий ключ перетворюється на просіяний ключ; в отриманих просіяних ключах виправлення помилок з метою отримання ідентичні послідовності даних на Клієнті та Сервері квантового розподілу ключів. Результатом виправлення помилок є очищений ключ; посилення секретності, представляє стиск отриманих ідентичних послідовностей даних з метою зменшення

інформації про згенерований квантовий ключ, доступної порушнику.

Результатом посилення секретності очищений квантовий ключ перетворюється на секретний ключ.

Результатом сеансу квантового розподілу ключів є випадкова квантова гама, ідентична у абонентів квантового каналу, таким чином даний результат має змінну довжину, яка не завжди співпадає з довжиною секретних ключів, які використовуються в алгоритмах шифрування. Результат виконання протоколу квантового розподілу ключів істотно відрізняється з високими та низькими втратами для квантових каналів мережі, які впливають на величину помилок при передачі в квантовому каналі мережі, що в свою чергу впливає на величину квантової гамми, та доступна порушнику на етапі посилення секретності.

Для шифрування інформаційного стану протоколу квантового розподілу ключів необхідно щонайменше два біти інформації, один біт визначає базис вимірювання (шифрування), другий значення інформаційного біта 1 або 0. Більш складні протоколи квантового розподілу ключів [5,6,21] можуть мати 3 і більше біт для шифрування інформації в однофотонному стані. На основі експериментальних даних для одного сеансу квантового розподілу ключів необхідна випадкова послідовність даних довжиною не менше біт. Швидкість генерації датчика випадкових чисел для отримання 256 бітного ключа за секунду має бути не менше біт/с або 400 Мбіт/с. Якщо, при цьому, врахувати, що частина квантового секретного ключа використовується для аутентифікації наступного сеансу квантового розподілу ключів мережі, тому необхідно враховувати розмір ключа аутентифікації, що, в свою чергу, збільшить нижню оцінку швидкості датчика генерації випадкових чисел.

До датчиків генерації випадкових чисел в системах квантового розподілу ключів пред'являють відповідні вимоги, до природи випадковості та якості формуємої послідовності даних [15,17,23]. Швидкість фізичних датчиків, в основі яких лежать квантові процеси, обмежена, в результаті призводить до обмеження швидкостей генерації квантових ключів швидкостями датчиків випадкових послідовностей [6,17,21].

Для коректної побудови протоколу квантового розподілу ключів та отримання відповідної якості квантових секретних ключів необхідна передача інформації по квантовому каналу мережі, побудова автентифікованого класичного каналу та наявність датчика випадкових послідовностей відповідної якості для створюваних квантових ключів.

Проведене дослідження робіт, що описують фізичну сторону технології квантового розподілу ключів, не приділяють відповід-ної уваги побудові аутентифікованого класичного каналу мережі. Для створення квантового каналу зв’язку необхідно визначити підхід до забезпечення аутентифікації джерела даних та цілісності інформації. Відповідним рішенням є обчислення імітовставки від послідовностей, для забезпечення цілісності інформації.

Аутентифікація відправника послідовностей здійснюється за рахунок використання секретного загального ключа, відомого, при цьому, лише парі легітимних абонентів.

Квантові пристрої не використовуються самі по собі, а є джерелом загальних секретів легітимних абонентів для систем захисту, апаратури. Автентифікований класичний канал з’єднання, через різні причини, (для економії числа фізичних каналів), може бути реалізований з використанням системи захисту інформації.

У роботі квантових пристроїв необхідно враховувати, що можливі множинні та одноразові переривання сеансу квантового розподілу ключів, через порушення у автентифікованому класичному каналі цілісності даних у зв'язку з випадковими збоями чи діями порушника. Повторний запуск сеансу квантового розподілу ключів веде до втрат секретних ключів автентифікації.

Іншою проблемою, є передача згенерованого квантового ключа від

квантових пристроїв легітимним абонентам, (пристроям, які використовуватимуть квантовий ключ). На відміну від класичних систем, взаємодія квантових пристроїв із системою захисту відбувається у двох парах пристроїв одночасно:

Сервер квантового розподілу ключів – система захисту та Клієнт квантового розподілу ключів – система захисту інформації. Квантовий секретний ключ створюється в одній парі квантових пристроїв, а передаватися повинен в пару інших пристроїв, систему захисту. При цьому, небезпечно застосування алгоритмів шифрування, не стійких до загроз (атак) із застосуванням квантового комп'ютера.

Формування асиметричними методами загальних секретів небезпечне через загрози із використанням квантового комп'ютера. Отже, можна визначити ряд невирішених проблем у розподілі спільних секретів із застосуванням квантового розподілу ключів в мережі: спосіб організації автентифікованого класичного каналу апаратури квантового розподілу ключів, що включає підхід до формування імітовставки для забезпечення цілісності даних та джерело секретних ключів для формування імітовставки; організація захищеної передачі загальних секретів абонентів у системах захисту інформації, стійкою до загроз (атак) квантовим комп'ютером; необхідність синхронізації ключової гами, в умовах відсутності можливості взаємодії систем захисту до отримання первинних ключів.

Вирішення перерахованих проблем дозволить застосовувати технологію квантового розподілу ключів для систем топології «крапка-крапка» для розподілу загальних секретів абонентів в пристрої користувача.

ДОСЛІДЖЕННЯ МЕТОДІВ ПОБУДОВИ МЕРЕЖ НА ОСНОВІ ТЕХНОЛОГІЙ КВАНТОВОГО РОЗПОДІЛУ КЛЮЧІВ

Протоколи квантового розподілу ключів мають граничну довжину квантового каналу мережі, де можливо створення спільних квантових ключів. У середньому максимальна довжина квантового каналу мережі для волоконних систем квантового розподілу ключів становить 100 км [5,6,13]. Пристрої, на які необхідно передати квантові ключі, розташовуються довільно.

Таким чином, постає задача подолання граничної віддаленості пристроїв квантової апаратури мережі з метою генерації загальних секретів для пар легітивних абонентів з необмеженою відстанню між ними.

Вирішення проблеми максимальної віддаленості легітимних абонентів є використання одного з двох підходів: застосування мереж квантового розподілу ключів на основі недовірених проміжних вузлів мережі; застосування мереж квантового розподілу ключів на основі довірених проміжних вузлів мережі. У кожному випадку використовуються мережі змішаної топології, в яких необхідно розподіляти квантові ключі.

В основі мереж квантового розподілу ключів з недовіреними проміжними вузлами лежить застосування комірок квантової пам'яті, розташованих на проміжних вузлах [4,6,15,17].

На сьогодні квантова пам'ять є абстрактним об'єктом, не має фізичного втілення, мережі з недовіреними проміжними вузлами і є перспективним, але з не реалізованим підходом найближчим часом.

Основа мереж квантового розподілу ключів з довіреними проміжними вузлами мережі є послідовна передача загального секрета (квантового ключа), отриманого на сегменті мережі квантового розподілу ключів через проміжні вузли мережі на відповідні вузли, на які розподіляється загальний секрет [6,11,14].

На кожному вузлі квантової мережі відбувається перекодування квантового ключа (загального секрету), і він доступний у відкритому вигляді на проміжних вузлах квантової мережі.

Таким чином, вузли квантової мережі повинні бути довіреними та забезпечений неможливий доступ порушника до квантових ключів, які у відкритому доступі на вузлах квантової мережі. Як алгоритм шифрування, в технології квантового

розподілу ключів, вибирають One-Time Pad Encryption (одноразовий шифроблокнот) [16 – 18,21].

Одним із варіантів вирішення проблеми появи загального секрету у відкритому виді на проміжних вузлах квантової мережі є застосування класичних стандартизованих підходів для захищеного представлення загального секрету на проміжних вузлах квантової мережі. Загальна схема магістральної мережі квантового розподілення ключів, представлена на Рис. 2.

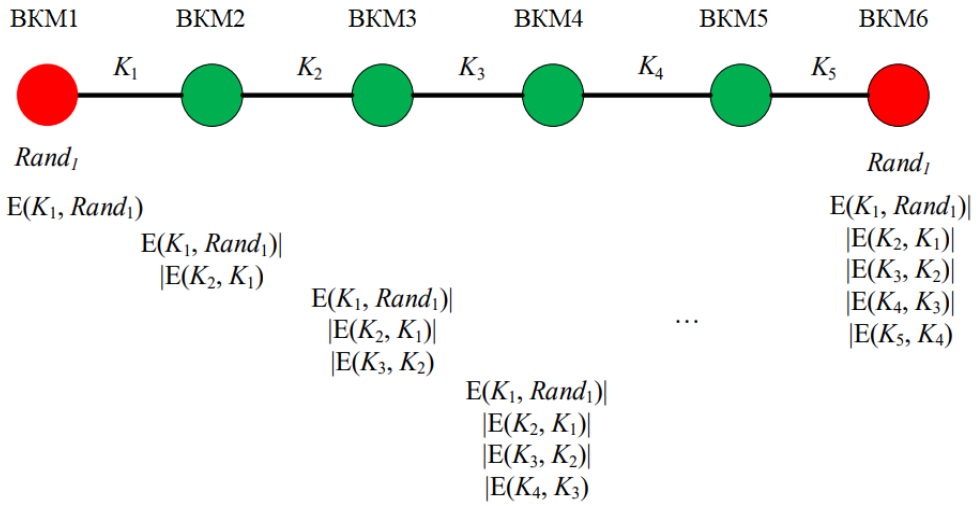


Рис.2. Метод формування квантового ключа «стеком»

Для захисту ключової інформації (загальний секрет) цільових вузлів квантової мережі, використовується загальний секрет першого сегмента мережі квантового розподілу ключів.

Формується зашифрований текст Зашифрований текст передається ланцюжком вузлів мережі до другого цільового вузла квантової мережі (ВКМ6). Для дешифрування на ВКМ6 необхідний квантовий ключ. Він передається від ВКМ2 мережі квантового розподілення ключів, із захистом на квантовому ключі. Передача та формування триває до тих пір, поки не буде переданий квантовий ключ зашифрований на ключі сегмента.

Тепер цільовий вузол мережі ВКМ6 може у зворотному порядку дешифрувати отримані зашифрованні тексти, в результаті отримати ключову інформацію.

Перевагами даного підходу є: відсутність перешифрування загальних секретів на проміжних вузлах квантової мережі, вирішує проблему появи загальних секретів на проміжних вузлах квантової мережі у відкритому виді; інформаційна стійкість захисту під час передачі квантових ключів; можливість скорочення зашифрованих текстів, при формуванні обчислювально стійких квантових ключів для магістральної підмережі квантового розподілу ключів з метою підвищення швидкості рознесення квантових ключів.

Недоліком даного підходу є пропорційне збільшення об’єму зашифрованих послідовностей, що передаються від сегментів в мережі квантового розподілу ключів.

Таким чином, визначено підхід до розподілу загального секрету та розглянуті спроби усунення недоліків даного підходу.

Вдосконалення розлянутого підходу до розподілу загального секрету на сьогодні є невирішеною задачею.

ДОСЛІДЖЕННЯ СТРУКТУРИ МЕРЕЖІ КВАНТОВОГО РОЗПОДІЛЕННЯ ЗАГАЛЬНОГО СЕКРЕТУ

На теперішній час ведуться роботи в галузі стандартизації мереж квантового розподілу ключів щодо способів їх функціонування та архітектури мереж. Група стандартизації ETSI в області квантового розподілу секретних ключів з 2010 року почала розробляти стандарти технології квантового розподілу ключів, базуючись на результатах

проекту SECOQC. Стандарти відносяться до різних областей застосування квантового розподілу ключів: реалізацій квантового розподілу ключів, секретності протоколів, методів вимірю-вань, інтерфейсів взаємодії.

Перший варіант мережі квантового розподілу ключів представлений в описі інтерфейсу взаємодії системи захисту інформації та квантової апаратури. Мережа квантового розподілу ключів побудована на основі довірених вузлів, наведених на рис. 3. У наведеному сценарії взаємодії відпо-відно до ETSI GS QKD беруть участь: апаратура квантового розподілення ключів – QKD; сервер керування ключами – KMS; сервер управління квантовими ключами мережі – KML; додатки користувача, клієнти мережі – App.

Структура мережі квантового розподілу побудована із незалежних частин. Довірений вузол квантової мережі містить кілька екземплярів квантових пристроїв, сервера керування ключами та сервери керування квантовими ключами. В довірений вузол мережі включається додаток користувача, для якого генеруються секретні ключі, проводиться аналіз об'єктів довіреного вузла мережі та його функції.

Базовим елементом мережі квантового розподілення ключів є квантова апаратура.

Пари комплектів квантової апаратури, пов'язані з автентифікованим класичним каналом та квантовим каналом, є самодостатнім елементом. Генерація квантових ключів відбувається незалежно від інших процесів у мережі квантового розподілення ключів. Квантова апаратура генерує випадкову послідовність нефіксованої довжини, а не квантовий ключ. Для формування із випадкових послідовностей квантових ключів використовуються KML (сервери керування квантовими ключами). В кожному вузлі квантової мережі використовується по серверу керування квантовими ключами. Сервер керування квантовими ключами містить сформовані квантові ключі з необхідною метаінформацією. Ідентичність квантових ключів, що генеруються в сусідніх вузлах квантової мережі, ґрунтується на довірі до протоколу квантового розподілу ключів, який гарантує ідентичність сгенерованих послідовностей.

Для передачі квантових ключів у системи захисту інформації користувача, закріплених за вузлами квантової мережі, які не з'єднані напряму з квантовим каналом, використовується KMS (сервер керування ключами). Канали зв'язку (рис. 3) формування квантового ключа відокремлені від каналу апаратури квантової генерації

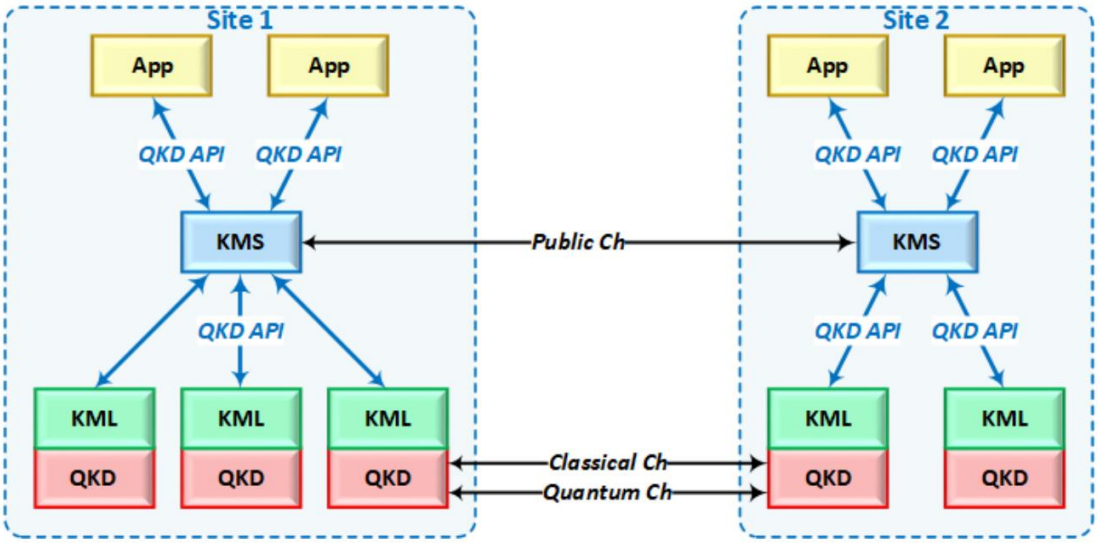


Рис. 3. Структура мережі квантового розподілу за версією ETSI GS QKD

ключів, забезпечення цілісності квантового ключа під час передачі не розглядається. Мережа квантового розподілення ключів допускає підключення декілька систем захисту інформації до одного вузла квантової мережі.

Перший варіант мережі квантового розподілення ключів за версією ETSI є чотирирівневою структурою. Два рівні квантової мережі оперують квантовими ключами, третій рівень - передача квантових ключів, згенерованих в сегменті мережі квантового розподілення ключів на інші сегменти квантової мережі. Четвертий рівень квантової мережі – рівень додатків користувача.

На сьогодні реалізований стандарт ETSI GS QKD V1.1, в якому внесено суттєві зміни до структури мережі квантового розподілення ключів. Схематично мережі квантового розподілення ключів представлено на рис. 4.

У сценарії роботи мережі квантового розподілення ключів беруть участь: апаратура квантового розподілення ключів – QKDE; клієнти мережі-SAE; пристрою керування ключами – KME.

В структурі мережі (рис.4) відбулося суттєве спрощення шляхом об'єднання сервера керування ключами та сервера керування квантовими ключами в єдиний об'єкт, систему захисту винесено за межі вузла квантової мережі, але для забезпечення безпечної передачі секретних ключів, повинна розташовуватися в межах контрольованої зони системи захисту. Довірений вузлом містить у собі клієнта мережі SAE та пристрій керування ключами KME. Квантова апаратура генерує квантові ключі нефіксованої довжини.

Згенерована квантова гамма передається на рівень керування секретними ключами, де з квантової гами формуються квантові ключі заданої, для мережі квантового розподілу ключів, довжини. До сгенерованих квантових ключів додаються метадані, дозволяють у мережі контролювати життєвий цикл квантових ключів (ідентифікатори вузлів, час створення ключа, довжину, час переміщення від квантової апаратури до рівня керування ключами).

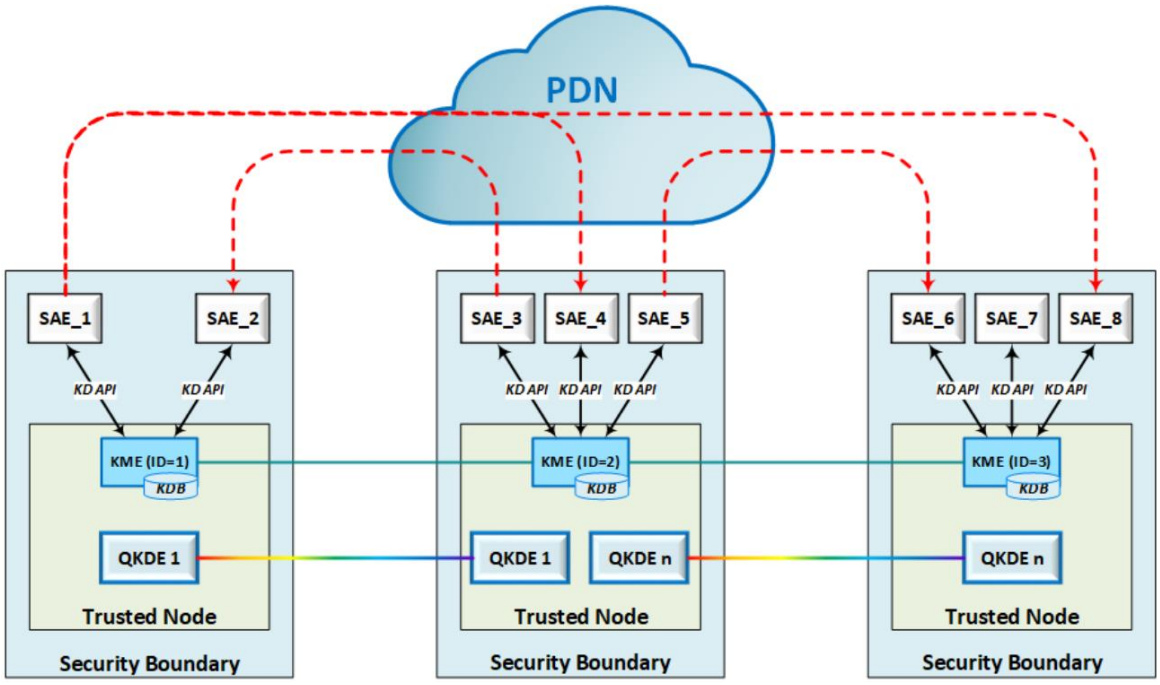


Рис. 4. Структура мережі квантового розподілення ключів ETSI-014

Мережі квантового розподілення ключів, мають багаторівневу структуру із централізованим управлінням, що описуються в європейських рекомендаціях. Взаємодія із системою захисту інформації, зовнішніми по відношенню до мережі квантового розподілу ключів, та розподілу загального секрету для пари вузлів квантової мережі відносяться до одного рівня мережі квантового розподілу ключів. Централізоване управління мережею квантового розподілу ключів, призведе до точки відмови у зв'язку з підвищеним навантаження на єдиний центр управління.

З 2019 року ведуться активні роботи зі створення рекомендацій щодо управління квантовими ключами в мережах КРК, включаючи наступні питання: зберігання згенерованих у мережі КРК ключів; розподіл ключів між вузлами у мережі КРК; передача ключів від апаратури КРК користувачам.

В даний час випущено документ [15], що містить короткий опис основних положень, що містяться в розроблених рекомендаціях, а також ряд документів, що конкретизує приватні аспекти, такі як управління ключами в мережі КРК, а саме процеси розподілу загальних секретів через ВКМ, функціональні вимоги до мереж КРК та їх елементів. Причому питання безпеки взаємодії ВКМ найчастіше винесені за рамки документів. На рис. 5 представлена загальна структура мережі КРК, що пропонується в рекомендаціях ITU-T.

Вузли мережі КРК (довірені вузли) складаються із двох логічних рівнів (Рис. 5):

1. Рівень квантових сигналів (Quantum layer) відповідає: за вироблення квантових ключів каналом квантового зв'язку між сусідніми вузлами; за передачу виробленого квантового ключа рівню управління ключами.

2. Рівень управління ключами (Key management layer) відповідає: за розподіл спільних секретів у мережі між несусідними вузлами; за видачу спільних секретів користувачу.

Розподіл спільних секретів між несусідними вузлами є складним процесом, що включає: керування розміром загального секрету; зміна формату розподіленого секрету та метаданих (ідентифікатор загального секрету, дата формування, довжина тощо) для управління секретами; зберігання спільних секретів; отримання характеристик квантового каналу QBER, швидкість генерації квантових ключів, статус квантового каналу зв'язку; безпечну (тобто стійку в теоретико-інформаційному сенсі) передачу спільних секретів між вузлами на рівні управління ключами; синхронізацію отриманих загальних секретів на вузлах; формування ключового контейнера та видачу загальних секретів користувачеві

Істотна зміна щодо попередніх пропозицій структури мережі КРК в описі архітектури ITU-T полягає у виділенні спеціалізованого рівня управління мережею КРК (QKD network management layer), що реалізується центром управління мережею, та рівня контролю за мережею КРК (QKD control layer).

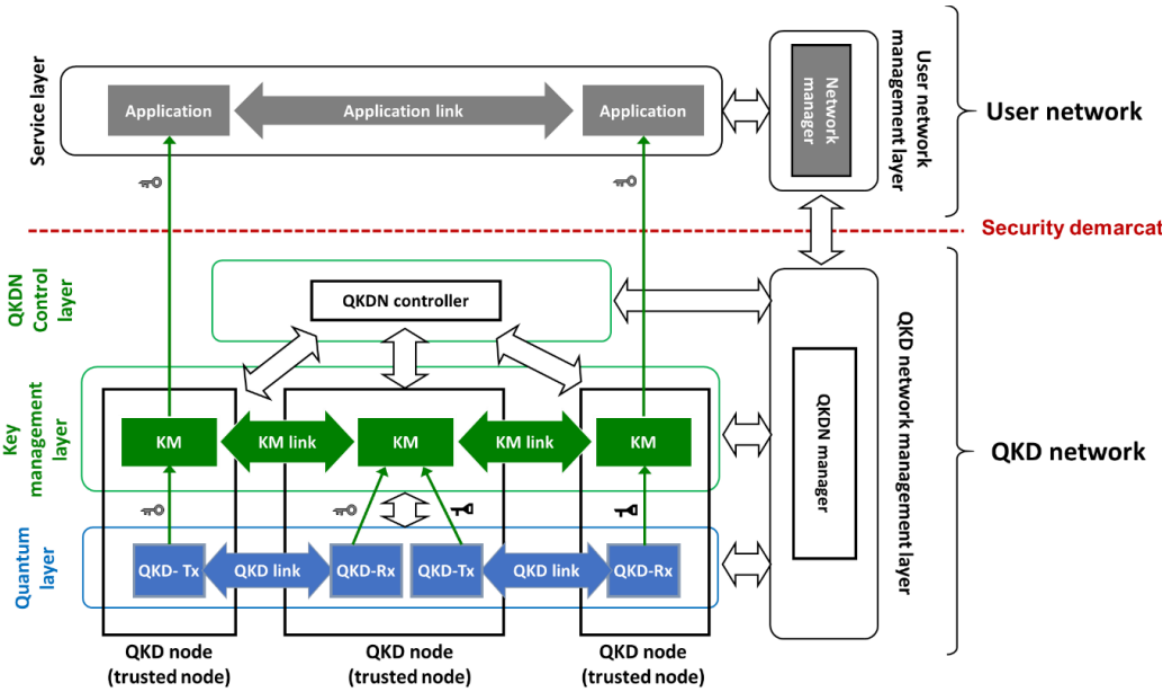


Рис. 5. Загальна структура мережі КРК ITU-T

Централізоване управління мережею КРК використано на вирішення проблеми розрахунку шляху передачі квантового ключа при розподілі загального секрету для двох ВКМ.

Роботи авторів стандартів ITU-T, що стосуються управління мережами КРК через концепцію SDN мереж, дозволяють припустити, що виділення рівня управління, пов'язаного з іншими рівнями мережі КРК, зроблено для подальшого спрощення інтеграції мереж КРК в існуючі мережі зв'язку загального користування з єдиним управлінням, тобто. об'єднання класичних та квантових SDN мереж. Як дані, що передаються на рівні даних SDN мережі, передбачаються однофотонні імпульси в квантовому каналі разом зі службовим трафіком протоколу КРК, а також ключова інформація при формуванні загального секрету. Звернемо увагу, що зовнішні СЗІ об'єднані у свою мережу, у якої також виділено централізовану сутність, що відповідає за управління мережею СЗІ.

Зауважимо, що існуючі документи, що описують структуру та сценарії роботи мереж КРК, не поділяють квантові ключі, згенеровані безпосередньо апаратурою КРК, та загальні секрети, що передаються в СЗІ, підключені до мережі. Це не зовсім коректний підхід, оскільки квантові ключі можливі лише між ВКМ, з'єднаними безпосередньо квантовим каналом.

Частково такий підхід обґрунтований основним пропонованим способом розподілу загального секрету для пари СЗІ, що полягає у передачі квантового ключа деякого сегмента мережі до цільових ВКМ.

Архітектура мережі, пропонована ITU-T, розглядає питання, які раніше не висвітлені у відкритих наукових публікаціях.

Запропоновано рішення для синхронізації спільних секретів на несусідніх ВКМ, а також набір метаданих, якими необхідно супроводжувати квантові ключі.

Квантова апаратура виробляє ключі нефіксованої довжини. Вироблена квантова гамма передається до рівня управління ключами, де з переданої гами формуються квантові ключі заданої, єдиної для всієї мережі КРК довжини. До сформованих квантових ключів додаються метадані, що дозволяють контролювати життєвий цикл ключів у мережі, включаючи час створення ключа, довжину, ідентифікатори вузлів, на

яких згенеровано ключ, час переміщення ключів між сутностями (від квантової апаратури до рівня управління ключами, між пристроями управління ключами тощо).

Контроль за станом квантової апаратури, у тому числі за подіями в квантовому каналі та величиною помилки QBER, проводиться в єдиному центрі управління. Збір даних контролю проводиться рівнем управління ключами з наступною передачею в центр управління. Зазначимо, що пропонований у документах ITU-T підхід вимагає передачі даних, що дозволяють порушнику оцінити ефективність своїх впливів на мережу КРК між вузлами мережі. Потрібний докладний аналіз такого трафіку визначення необхідних заходів захисту, тобто. чи достатньо забезпечувати лише цілісність даних чи необхідне забезпечення конфіденційності.

Перенесення функцій контролю над станом квантової апаратури з самої апаратури у віддалений центр управління з одного боку дозволяє оптимізувати процеси формування квантових ключів, оскільки параметри вироблення квантових ключів прямо чи опосередковано (залежно від способу формування та передачі квантового ключа) впливають на його генерацію. З іншого боку, аналіз стану квантової апаратури щодо наявності тих чи інших атак порушника у віддаленому центрі збільшує час реакції всієї системи на дії порушника, що може негативно позначатися на стійкості системи загалом.

Єдиний центр контролю має функцію контролю життєвого циклу ключів, що генеруються у мережі КРК. Для цього в центр контролю передаються метадані, що прикріплюються до ключів при передачі з квантової апаратури, у разі зміни цих метаданих.

Процес розподілу загальних секретів для довільної пари ВКМ згідно [6,15] полягає в наступному:

1. Сервіс управління мережею СЗІ формує перелік пар СЗІ, для яких потрібні спільні секрети від мережі КРК. Цей перелік передається до центру керування мережею КРК. Причому в документах ITU-T ці спільні секрети називаються квантовими ключами поруч із істинно квантовими ключами, формованими в одному сегменті мережі КРК.

2. Центр управління мережею КРК визначає ланцюжки ВКМ, за якими можна передати квантові ключі на ВКМ, пов'язані з парами СЗІ, для яких мережа СЗІ запросила ключі. На визначення ланцюжків ВКМ впливає кількість готових квантових ключів на сегментах мережі, швидкість генерації нових квантових ключів (за результатами аналізу стану квантової апаратури), а також топологія мережі в цілому.

3. Набір певних ланцюжків передається до контролера мережі КРК, який формує інструкції для кожного ВКМ, що включають маршрут передачі ключів та вказівку, які саме ключі використовувати для передачі, а які для забезпечення конфіденційності передачі даних. Вводяться три види маршрутизації, що визначають шляхи передачі квантових ключів на цільові ВКМ:

1.Ручна маршрутизація - статично задані ланцюжки для пар ВКМ, до яких здійснюється підключення СЗІ.

Цей спосіб доцільно застосовувати лише до невеликих мереж КРК або до мереж з невеликою кількістю розгалужень, де такі ланцюжки легко розраховуються і не мають альтернативних варіантів.

2. Маршрутизація за фіксованою швидкістю.

Цей вид маршрутизації орієнтований на оптимальне навантаження мережі виходячи з продуктивності квантової апаратури. Центр управління мережею аналізує поточну швидкість роботи кожної пари квантової апаратури, а також визначає цільові пари ВКМ, на які необхідно доставити ключі для передачі СЗІ. Ланцюжки вузлів підбираються так, щоб швидкість формування квантових ключів по всіх ланцюжках у сукупності виявилася мінімальною. Потім розраховані ланцюжки передаються до контролера мережі, де формуються конкретні інструкції для вузлів квантової мережі.

3.Адаптивна маршрутизація. Цей вид маршрутизації описаний як додатковий до попередніх. Кожен ВКМ запам'ятовує розраховані маршрути. При отриманні

мережею КРК запиту загального секрету, що повторює один із раніше отриманих, маршрут не перераховується, а ВКМ видається команда повторити формування загального секрету на основі раніше отриманих інструкцій. Цей вид маршрутизації може бути доцільний при стабільній роботі мережі КРК та регулярних запитах ключів від постійних пар СЗІ.

Жоден із запропонованих способів маршрутизації не враховує стійкість застосовуваних методів захисту під час передачі ключової інформації по ланцюжках ВКМ і, відповідно, підсумкову стійкість розподіленого загального секрету.

Для моніторингу роботи мережі КРК загалом крім зазначених вище функцій центр управління додатково оснащується наступним функціоналом: модуль управління, що відповідає, крім розрахунків маршрутів передачі квантових ключів, за забезпечення якості послуг, що надаються мережею (Quality of Service); модуль аутентифікації, який відповідає за аутентифікацію та ідентифікацію вузлів мережі; модуль політик, який відповідає за управління політиками безпеки мережі КРК, до цього модуля може підключатися оператор моніторингу роботи мережі.

Для побудови мережі КРК необхідно виробити рішення супутніх завдань, що стосуються взаємної ідентифікації пристроїв, що становлять мережу КРК; єдиного налаштування та моніторингу параметрів мережі КРК, що відповідають за безпечне функціонування, а також забезпечення необхідних експлуатаційних властивостей. У розглянутих джерелах зустрічається різний поділ ВКМ.

Документи ITU-T вводять наступну класифікацію:

1. Користувальницькі ВКМ. Дані ВКМ повинні мати можливість підключення СЗІ. Зазвичай користувальницькі ВКМ містять один екземпляр квантової апаратури. З економічної доцільності це Клієнт КРК.

2. Проміжні ВКМ. Дані ВКМ мають становити основу мережі КРК, вирішуючи основне питання: збільшення дальності розподілу загального секрету СЗІ. Проміжні ВКМ містять як мінімум пару екземплярів квантової апаратури, Сервер КРК та Клієнт КРК, кожен з яких з'єднаний з квантовою апаратурою інших ВКМ, проміжних або доступу. Фактично проміжні ВКМ становлять основні магістральні гілки мережі КРК.

3. ВКМ доступу. Такі ВКМ необхідні для з’єднання проміжних ВКМ з кінцевими, тобто, користувальницькими ВКМ. Містить Сервер КРК для зв'язку з користувальницьким ВКМ та один або більше відповідних екземплярів квантової апаратури для зв'язку з проміжними ВКМ.

Окремо зазначається, що користувачем загальних секретів (маються на увазі загальні секрети, що розподіляються по мережі) можуть виступати як незалежні зовнішні пристрої, так і вбудовані безпосередньо в ВКМ.

Отже, мережі КРК, що описуються в європейських рекомендаціях, мають багаторівневу структуру із централізованим управлінням. При цьому взаємодія із СЗІ, зовнішніми по відношенню до мережі КРК, та розподіл загального секрету для пари ВКМ відносяться до одного рівня мережі КРК. Всі ключі, що з'являються в мережі КРК, називаються квантовими ключами, що вводить додаткову плутанину і невідмінність квантових ключів, створюваних внаслідок протоколу КРК, від загальних секретів, отриманих у результаті подальшого функціонування мережі КРК. Централізоване управління мережею КРК веде до виникнення точки відмови у зв'язку з підвищенням навантаження на єдиний центр управління.

Отже, є проблема дослідження можливості побудови децентралізованої мережі квантового розподілу ключів, а також необхідність вирішення проблеми розподілу квантових ключів (загального секрету) на довільні пари вузлів квантової мережі із забезпеченням не тільки цілісності ключової інформації, а також конфіденційності, уточнення поняття довіри до проміжних вузлів квантової мережі.

ВИСНОВКИ

У результаті проведеного дослідження виявлено наступні проблеми, пов'язані з недоліками технології квантового розподілу ключів:

1. Системи квантового розподілу ключів мають граничну довжину квантового каналу, визначається протоколом квантового розподілу ключів. Пристрої користувача, для яких генеруються квантові ключі, можуть довільно розташовуватися, на відстанях, що перевищують допустиму довжину квантового каналу. Необхідно розглянути задачу розподілу квантових ключів для довільних пар пристроїв квантової мережі, що розташовані поза граничної довжини квантового каналу.

2. Системи квантового розподілу ключів повинні мати автентифікований класичний канал, при цьому необхідно забезпечити цілісність інформації у квантовому каналі.

3. Забезпечити синхронізації квантових ключів, що передаються в вузли квантової мережі, пристрої користувача. Відомі підходи, не враховують квантової специфіки апаратури.

4. Необхідно розглянути задачу забезпечення конфіденційності ключової інформації та її цілісності на довірених проміжних вузлах квантової мережі.

5. Структура мережі квантового розподілу ключів та спосіб її функціонування містять не вирішені задачі, що перешкоджають ефективному розповсюдженню та створенню квантових ключів на довільні пари вузлів квантової мережі.

З урахуванням розглянутих задач у рамках проведеного дослідження для підвищення захищеності мереж квантового розподілу ключів необхідно вирішити наступні задачі: розробити підхід до побудови автентифікованого класичного каналу квантової апаратури; розробити спосіб взаємодії користувальницьких систем захисту інформації з квантовою апаратури, що уточнює процеси забезпечення цілісності загальних секретів та синхронізації при їх передачі до систем захисту; розробити метод розподілу квантових ключів для декількох вузлів квантової мережі магістральної топології.

ЛІТЕРАТУРА

1. Доктрина інформаційної безпеки України, затвердженої Указом Президента України від 25 лютого 2017 року № №47/2017, 15с.
2. Державний стандарт України Захист інформації. Технічний захист інформації. Основні положення. ДСТУ 3396.0-96 [Електронний ресурс]. – Режим доступу: http://www.dsszzi.gov.ua/dsszzi/control/uk/publish/article?art_id=38883&cat_id =38836.
3. Квантова криптографія. Пояснення [Електронний ресурс] // Quantum Xchange. – 2019. – Режим доступу: https://quantumxc.com/blog/quantum-cryptography-explained/ (дата звернення 02.03.2024).
4. **Satish Kumar**. Quantum Cryptography [Електронний ресурс] // Tutorialspoint. – 2023. – Режим доступу: http://surl.li/fjebs (дата звернення 05.03.2024).
5. **Грамблінг Е**. Квантові обчислення. Прогрес і перспективи / Е. Грамблінг, М. Горовіц. – Вашингтон: Національні академії наук, техніки та медицини, 2019. – 272 с.
6. **Керті М., Ло Х.-К., Тамакі К**. Безпечний квантовий розподіл ключів. Фотоніка природи. 2018. Т. 8. С. 595–604.
7. Закон України «Про основні засади забезпечення кібербезпеки України» зі змінами. Відомості Верховної Ради (ВВР), 2017, № 45, ст.403, зі змінами від 28.07.2022 року. Режим доступу: https://zakon.rada.gov.ua/laws/show/2163
8. **Бем, М.В**. Стандарти захисту персональних даних в соціальній сфері. / М. В.Бем, І. М. Городиський -Львів:, 2018р. - 110 с.
9. **Богуш, В.М**. Інформаційна безпека держави / В.М. Богуш, О.К. Юдін. – К.: МК-Прес, 2015. – 432 с.
10. **Голубєв О.В**. Програмно-технічні засоби захисту даних від комп'ютерних злочинів / О. В. Голубєв– Запоріжжя: «Павел», 2018. – 145 с.
11. **Горбулін В.П.** Проблеми захисту інформаційного простору України / М.М. Баченок, П.В. Горбулін – К.: Інтертехнологія, 2019. – 138 с.
12. **Біленчук П.Д.** Правові засади інформаційної безпеки України: монографія /Л.В. Борисова, І.М. Неклонський.– Харків: 2018. – 289 с.

13. **Буров Є.В.** Комп'ютерні мережі. Том 1. / Є.В.Буров, М.М.Митник //Навчальний посібник – Львів, «Магнолія 2006», 2019р.- 256 с.
14. **Буров Є.В.** Комп'ютерні мережі. Том 2. / Є.В.Буров, М.М.Митник//Навчальний посібник – Львів, «Магнолія 2006», 2019р. - 334 с.
15. **Shor P.W.** Algorithms for quantum computation: discrete logarithms and factoring. 35th annual symposium on foundations of computer science, Santa Fe, NM, USA. URL: https://doi.org/10.1109/sfcs.1994.365700 (date of access: 16.11.2023).
16. **Вербіцький О.В.** Вступ до криптології / О.В. Вербіцький. – Львів : ВНТЛ, 2017р. – 248 с.
17. **Остапов С.Е.** Квантова інформатика та квантові обчислення / С.Е.Остапов, Ю.Г.Доб-ровольський - Чернівці: ЧНУ, 2021. – 69-73 с
18. **Ленков С.В.** Метод прогнозування вразливостей інформаційної безпеки на основі аналізу даних тематичних інтернет-ресурсів / С.В. Лєнков, В.М. Джулій, А.М. Берназ, І.В. Муляр, І.В. Пампуха // Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка. – К.: ВІКНУ, 2023. – Вип. №78. – С. 123-134.
19. **Вишня В.Б.** Основи інформаційної безпеки: навч. посібник / В. Б. Вишня, О. С. Гавриш, Е. В. Рижков. Дніпро : Дніпроп. держ. ун-т внутріш. справ, 2020. – 128 с.
20. **Остапов С.Е.** Технології захисту інформації: навчальний посібник / С.Е. Остапов, С.П. Євсеєв, О.Г. Король – Харків: Вид-во ХНЕУ, 2016. – 476 с.
21. **Бурячок В.Л.** Інформаційний та кіберпростори: проблеми безпеки, методи та засоби боротьби : посібник / [В. Л. Бурячок, С. В. Толюпа, В. В. Семко та ін.]. – К.: ДУТ-КНУ, 2016. – 178 с.
22. **Лавров Є.А.** Математичні методи дослідження операцій: підручник / Є. А. Лавров, Л. П. Перхун, В. В. Шендрик – Суми: Сумський державний університет, 2017. – 212 с.

REFERENCES

1. Doktryna informatsiinoi bezpeky Ukrainy, zatverdzhenoi Ukazom Prezydenta Ukrainy vid 25 liutoho 2017 roku № №47/2017, 15s.
2. Derzhavnyi standart Ukrainy Zakhyst informatsii. Tekhnichnyi zakhyst informatsii. Osnovni polozhennia. DSTU 3396.0-96 [Elektronnyi resurs]. – Rezhym dostupu: http://www.dsszzi.gov.ua/dsszzi/control/uk/publish/article?art_id=38883&cat_id =38836
3. Kvantova kryptohrafiia. Poiasnennia [Elektronnyi resurs] // Quantum Xchange. – 2019. – Rezhym dostupu: https://quantumxc.com/blog/quantum-cryptography-explained/ (data zvernennia 02.03.2024).
4. **Satish Kumar**. (2023) Quantum Cryptography [Elektronnyi resurs] // Tutorialspoint. –Rezhym dostupu: http://surl.li/fjebs (data zvernennia 05.03.2024).
5. **Hramblinh E.** (2019) Kvantovi obchyslennia. Prohres i perspektyvy / E. Hramblinh, M. Horovits. – Vashynhton: Natsionalni akademii nauk, tekhniky ta medytsyny, – 272 s.
6. **Kerti M., Lo Kh.-K., Tamaki K.** (2018) Bezpechnyi kvantovyi rozpodil kliuchiv. Fotonika pryrody. T. 8. S. 595–604.
7. Zakon Ukrainy «Pro osnovni zasady zabezpechennia kiberbezpeky Ukrainy» zi zminamy. Vidomosti Verkhovnoi Rady (VVR), 2017, № 45, st.403, zi zminamy vid 28.07.2022 roku. Rezhym dostupu: https://zakon.rada.gov.ua/laws/show/2163
8. **Bem, M. V.** (2018) Standarty zakhystu personalnykh danykh v sotsialnii sferi. / M. V.Bem, I. M. Horodyskyi -Lviv: - 110 s.
9. Bohush, V.M. (2015) Informatsiina bezpeka derzhavy / V.M. Bohush, O.K. Yudin. – K.: MK-Pres. – 432 s.
10. **Holubiev, O.V.** (2018) Prohramno-tekhnichni zasoby zakhystu danykh vid kompiuternykh zlochyniv / O. V. Holubiev– Zaporizhzhia : «Pavel». – 145 s.
11. **Horbulin, V.P.** (2019) Problemy zakhystu informatsiinoho prostoru Ukrainy / M.M. Bachenok, V.P. Horbulin – K.: Intertekhnolohiia. – 138 s.
12. **Bilenchuk, P.D.** (2018) Pravovi zasady informatsiinoi bezpeky Ukrainy: monohrafiia /L.V. Borysova, I.M. Neklonskyi.– Kharkiv: – 289 s.
13. **Burov, Ye.V.** (2019) Kompiuterni merezhi. Tom 1. / Ye.V. Burov, M.M. Mytnyk // Navchalnyi posibnyk – Lviv, «Mahnoliia 2006» - 256 s.
14. **Burov, Ye.V.** (2019) Kompiuterni merezhi. Tom 2. / Ye.V. Burov, M.M. Mytnyk // Navchalnyi posibnyk – Lviv, «Mahnoliia 2006» - 334 s.
15. **Shor P.W.** (2023). Algorithms for quantum computation: discrete logarithms and factoring. 35th annual symposium on foundations of computer science, Santa Fe, NM, USA. URL: https://doi.org/10.1109/sfcs.1994.365700 (date of access: 16.11.2023).
16. **Verbitskyi O.V.** (2017). Vstup do kryptolohii / O.V. Verbitskyi. – Lviv : VNTL. – 248 s.

17. **Ostapov S.E.** (2021). Kvantova informatyka ta kvantovi obchyslennia. / S.E. Ostapov, Yu.H Dobrovolskyi - Chernivtsi: ChNU. – 69-73 s
18. **Lenkov S.V.** (2023). Metod prohnozuvannia vrazlyvostei informatsiinoi bezpeky na osnovi analizu danykh tematychnykh internet-resursiv / S.V. Lienkov, V.M. Dzhulii, A.M. Bernaz, I.V. Muliar, I.V. Pampukha // Zbirnyk naukovykh prats Viiskovoho instytutu Kyivskoho natsionalnoho universytetu imeni Tarasa Shevchenka. – K.: VIKNU -. №78. – C. 123-134.
19. **Vyshnia, V. B.** (2020) Osnovy informatsiinoi bezpeky : navch. posibnyk / V. B. Vyshnia, O. S. Havrysh, E. V. Ryzhkov. Dnipro : Dniprop. derzh. un-t vnutrish. sprav. – 128 s.
20. **Ostapov S. E.** (2016). Tekhnolohii zakhystu informatsii: navchalnyi posibnyk / S.E. Ostapov, S.P. Yevseiev, O.H. Korol–Kharkiv: Vyd-vo KhNEU. – 476 s.
21. **Buriachok V.L., Toliupa S.V., Semko V.V.** (2016). Informatsiinyi ta kiberprostory: problemy bezpeky, metody ta zasoby borotby: posibnyk. Kyiv, DUT-KNU, 178.
22. **Dzhulii V.M.**, **Miroshnichenko O.V., Solodieieva L.V.** (2022). Metod klasyfikatsii dodatkiv trafika kompiuternykh merezh na osnovi mashynnoho navchannia v umovakh nevyznachenosti. Zbirnyk naukovykh prats Viiskovoho instytutu Kyivskoho natsionalnoho universytetu imeni Tarasa Shevchenka, Vol.74, 73-82.
23. **Lavrov Ye.A., Perkhun L.P., Shendryk V.V.** (2017). Matematychni metody doslidzhennia operatsii: pidruchnyk. Sumy, Sumskyi derzhavnyi universytet, 212.

Research into the current state of quantum secret key distribution technologies

Serhiy Lenkov[1], Volodymyr Dzhuly[2], Ihor Mulyar[3], Yevhen Lenkov[4]

Abstract. The work is devoted to the study of the problems of using the technology of quantum distribution of secret keys, namely the problems of using and forming new secret keys, based on the generation of quantum keys and the distribution of new keys in networks of complex topologies.

In secure communication systems, information is transmitted over public networks, and thus is available to a potential attacker for carrying out various types of attacks. A potential attacker is able to store encrypted data transmitted over network channels, and attempt to decrypt it later, when technical means and attacks on information protection algorithms will allow, in an acceptable time, to carry out decryption. This is the basic principle of "Store now, decrypt later", which is one of the problems of information protection systems.

For block ciphers, there is a parameter - the key load, which limits the amount of information that one key can process using the data protection algorithm. NIST recommends a time frame for using secret keys, calculated in years. In the field of information security, there are more significant restrictions on the key load, related to the number of blocks that can be processed by one key, caused by attacks on block ciphers. With these restrictions, there is a need to regularly change the secret key used, and search for options for generating and delivering keys between two network nodes that organize a secure information transmission channel. Changing the secret key to a new key, independent of the previous one, allows you to protect the transfer of confidential information when the key used is compromised.

The technology of creating a quantum computer allows you to effectively attack known algorithms for generating secret keys based on the factorization of large numbers or the complexity of calculating a discrete logarithm. The quantum Shor algorithm [5, 6] allows you to solve these problems in polynomial time, so it is necessary to look for alternative solutions to the problem of regular changes in the network channel of the secret key in a pair of nodes.

It is important to take into account the need to ensure the quality of Perfect forward secrecy, when the master key is compromised, all subsequent secret keys must not be compromised. To achieve this property of modern encryption algorithms, there is a periodic update of master keys.

To prevent the accumulation of information, to successfully carry out attacks by the attacker and to comply with the load requirement on the secret key, it is necessary to solve the problem of regular changes of the secret, and regular generation/delivery of keys to network nodes.

Given the urgent needs that are intensively developing today in the secure and high-speed transmission of information, and no less, the rapid growth of technical capabilities, there is a need to develop options in advance that will ensure the security of information during transmission. Regular change, in geographically dispersed data protection systems, of common secret keys, in the case of ensuring the independence of secret keys, allows successfully solving the task.

Keywords: information security, vulnerabilities, attacks, quantum secret key, quantum distribution protocol.

Спосіб розгортання AI платформи з використанням методології MLOPS

Володимир Русінов

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Берестейський проспект, 37, Київ, 03056
https://orcid.org/0000-0002-4362-0248

Received 03.02.2025, accepted 29.03.2025
https://doi.org/10.32347/st.2025.3.1202

Анотація. Специфіка сучасного застосування задач ШІ потребує розробки плафторми, яка уможливлює постійну підтримку та швидку реконфігурацію. Наукова та інженерна література свідчить, що використання методів Machine Learning Operations та Machine Learning Pipeline (MLOps) при створенні AI-платформ для вбудованих систем дозволяє підвищити ефективність розробки та впровадження штучного інтелекту. MLOps надає можливість автоматизувати процеси розробки, навчання та експлуатації моделей машинного навчання, забезпечуючи швидку реакцію на зміни в навколишньому середовищі або вхідних даних. Це допомагає скоротити час на розгортання та обслуговування системи в цілому.

Для вирішення задачі розгортання AI-платформи пропонується використовувати спосіб, який базується на гібридній Edge-Cloud архітектурі. Даний підхід часто використовується в сучасній науковій літературі і знаходить прикладне застосування в різних сферах. Архітектура платформи, заснована на ієрархічних зв'язках між рівнями Edge і Cloud, диктує необхідність в створенні спеціалізованого підходу до розгортання платформи. Особливістю даної архітектури є застосування вбудованих систем безпосередньо в задачі ШІ, а саме, розпізнаванні образів. Даний підхід дозволяє реконфігурувати платформу, за умови оновлення моделі ШІ, тим самим підтримуючи її працездатність і коректне виконання задачі.

Враховуючи, що ключовим є баланс швидкості реконфігурації подібної платформи та відмовостійкості, для того, щоб виміряти загальну працездатність системи з певним ступенем точності, ми введемо кілька метрик, які допоможуть встановити параметри системи та порівняти її з альтернативними рішеннями. Розгортання програмного забезпечення для штучного інтелекту повинно супроводжуватися

Володимир Русінов
Аспірант кафедри
обчислювальної техніки

аналізом цільової системи. В рамках даної роботи проведено ряд тестів для кількісної оцінки продуктивності системи.

Ключові слова: штучний інтелект, MLOps, IoT, відмовостійкість, комп'ютерні мережі.

ПОСТАНОВКА ПРОБЛЕМИ

Сучасні AI рішення потребують створення платформ, які здатні підтримувати їх життєвий цикл, починаючи від попереднього тестування нових моделей ШІ, закінчуючи постійною підтримкою впроваджених моделей. Часто, використання моделей ШІ має суто експериментальний характер і впровадження таких моделей не супроводжується планом розгортання, але все більше моделей впроваджується для постійного користування. Окрім того, що система може бути нестабільною і не підтримувати впровадження більш великих моделей ШІ, такий підхід призводить до складнощів при пошуку причин зменшення точності моделі. Тому, існують підходи до створення платформ.

Розгортання платформи для задач штучного інтелекту має супроводжуватися ретельним аналізом цільової системи.

Практичними результатами способу, який базується на використання методології MLOps, є прискорення процесів інтеграції вбудованих пристроїв в існуючі Cloud рішення та підвищення відмовостійкості AI-плафторми за рахунок постійного моніторингу стану платформи та її реконфігурації. Програмно-апаратні рішення на основі запропонованого підходу можуть знайти використання в багатьох сферах, зокрема в медицини, безпеки, логістики. Рішення про впровадження таких практик може значно покращити швидкість та правильність прийняття рішень щодо оновлення моделі, або поступового донавчання моделі.

Для того, щоб виміряти загальну зручність використання системи з певним ступенем точності, необхідно проводити тестування та аналіз платформи, з фокусом на процесі розгортання та постійні підтримці працездатності системи. Такий підхід дозволяє не тільки виявляти потенційні проблеми на ранніх етапах, але й оцінити швидкодію платформи в реальному часі. У майбутньому інтеграція таких систем на рівні організацій зможе значно підвищити рівень автоматизації і точності в роботі не лише AI-рішень, але й у всіх аспектах бізнес-процесів, що веде до більш стабільного і надійного використання штучного інтелекту в реальних умовах.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ

Сучасні наукові дослідження показують, що MLOps - це новий спосіб розгортання систем ШІ, який успадкував свої методи від DevOps, що показав позитивні результати в різних програмних системах. У різних статтях пропонуються різні варіації інструментів і конструкцій конвеєрів CI/CD, які дозволяють виконувати задачі ШІ масштабовано і підвищують загальну відмовостійкість системи. Ці рішення також пропонують метод автоматизації процесів, які потребують додаткових людських ресурсів, таких як моніторинг, ведення логування і реконфігурація. Сучасні рішення дають мінімальне уявлення про продуктивність таких систем.

В науковій літературі було виявлено, що підхід на основі інтеграції вбудованих пристроїв в процес MLOps є багатообіцяючим, але потребує додаткових досліджень. В літературі сказано про шляхи оптимізації взаємодії із Edge-кластерами, проте недостатньо дослідженим залишається питання відмовостійкості системи при високій завантаженості платформи та можливість використання AI-задач. Важливим аспектом, який потребує подальшого дослідження, є збір часових характеристик щодо використання таких гібридних систем, та порівняння з традиційними підходами.

МЕТА І ЗАДАЧІ ДОСЛІДЖЕННЯ

Метою статті є дослідження та впровадження принципів MLOps для розгортання AI платформи.

Для досягнення мети поставлені задачі:

— Проаналізувати сучасні підходи до створення AI-платформ, зокрема, методологію MLOps та виділити необхідні етапи для створення платформи;

— Розробити власний спосіб розгортання AI платформи на основі конвеєру MLOps за допомогою використання CI/CD конвеєру;

— Розробити тестові сценарії перевірки працездатності платформи за умов реконфігурації та високого навантаження запитами на платформу;

— Провести аналіз результатів тестування платформи, на основі запропонованого способу, згідно описаних сценаріїв та виділити напрямки подальшого дослідження.

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ ДОСЛІДЖЕННЯ

Операції машинного навчання (Machine Learning Operations, MLOps) з'явилися як концепція, заснована на методах DevOps, на меті яких створити набір процедур, які об'єднують інструменти, процедури розгортання та робочі процеси для швидкого та доступного створення моделей

машинного навчання [1]. Метододологія MLOps рекомендує автоматизувати та контролювати кожен крок створення та впровадження систем машинного навчання з подальшим логу-ванням та дослідженням внутрішніх процесів [2]. Практики MLOps можна сприймати як складну координацію набору принципів, прийомів та програмних компонентів, що використовуються інтегровано для виконання щонайменше п'яти базових завдань: збір даних, перетворення даних, безперервне навчання моделі машинного навчання, безперервна реалізація моделі, представлення результатів кінцевому користувачеві. Кінцевий продукт та його підтримка досягається завдяки використанню принципів MLOps.

Крім того, продуктивність рішення, яке не використовує стратегії MLOps, може з часом погіршитися, оскільки програми машинного навчання покладаються на дані, що використовуються для навчання моделі штучного інтелекту, і оскільки свіжі дані, тобто ті на яких модель не була натренована, постійно надходять до програми [5]. Для того, щоб гарантувати ефективність і якість рішення штучного інтелекту протягом тривалого часу, одним з найбільш доречних заходів MLOps є моніторинг рішень машинного навчання.

Запропоноване рішення передбачає створення спеціального наскрізного (end-to-end) конвеєра CI/CD. Конвеєр CI/CD для системи ШІ зазвичай починається з етапу інтеграції. Це передбачає витягування останніх змін коду з системи контролю версій, наприклад, Git, і об'єднання їх у спільну гілку. Потім запускаються автоматизовані тести для перевірки змін у коді. У випадку системи ШІ ці тести можуть включати модульні тести для окремих компонентів, а також інтеграційні тести, які оцінюють продуктивність моделей машинного навчання за набором попередньо визначених метрик.

AI-платформа, як програмно-апаратне рішення для задач штучного інтелекту, вимагає розробки окремої моделі ML, яка за своєю суттю є експериментальною. Типовий робочий процес розробки моделі машинного навчання починається зі збору додаткової інформації про дані та поставлене завдання. Дані необхідно проаналізувати (Exploratory Data Analysis (EDA)), очистити та попередньо обробити (інжиніринг та виділення ознак), щоб виділити основні ознаки з сирих даних. Після цього формується набір даних для навчання, тестування та верифікації [13].

Підхід до реалізації MLOps конвеєра для AI-платформи передбачає багатоступеневий процес, що забезпечує валідацію, тестування та інтеграцію AI компонентів. Етапи слідують один за одним, утворюючи більший конвеєр, який називається CI/CD. Спосіб включає наступні етапи:

1. Безперервна інтеграція (CI) - це практика інтеграції змін коду в існуючу кодову базу таким чином, щоб будь-які конфлікти між різними змінами коду, зробленими розробниками, швидко виявлялися і вирішувалися [2]. Оскільки фази планування та кодування виконуються до фази побудови, їх можна вважати пов'язаними з безперервною інтеграцією. Під час виконання фази CI, готуються нові зміни коду до розгортання, тому ця практика має безпосереднє значення для підвищення ефективності розгортання з точки зору якості та коректного виконання коду.

1.1. На етапі підготовки здійснюється комплексне тестування вихідного коду машинного навчання, включаючи синтаксичний аналіз, перевірку типів даних (семантичний аналіз), оцінку метрик продуктивності моделі. Платформа перевіряє відповідність кодової бази встановленим стандартам якості та вимогам архітектурної сумісності. Фаза збірки та тестування є компонентами DevOps, які беруть участь у безперервній інтеграції, оскільки кожна фіксація коду супроводжується автоматизованою генерацією коду з подальшим автоматизованим тестуванням.

1.2. Наступним етапом конвеєра є контейнеризація та ізоляція обчислювального середовища з використанням технологій як, наприклад, Docker. Даний етап забезпечує уніфіковане

розгортання AI-компонентів з високим рівнем масштабованості,

гнучкості та відмовостійкості. Підхід на основі конетейнеризації дозволяє ефективно розгортати нові ітерації моделі, швикдко вносити зміни з залежностями в програмній частині та ефективно сегрегувати програмну частину, яка відповідає за AI модель від модулів комунікації з іншими сервісами платформи та користувачами.

2. Безперервна доставка (CD) включає в себе фазу впровадження (релізу) коду, її мета - тримати додаток готовим до розгортання у виробничому середовищі. Безперервне розгортання виконується пізніше, оскільки воно автоматизує розгортання програмної частини платформи. Якщо модель недоступна, виробнича версія буде розгорнута користувачем вручну.

2.1. Першим етапом конвеєру CD є етап компіляції дистрибутиву, що, в рамках даного дослідження, передбачає трансформацію вихідного коду в артефакти розгортання. Реалізується контейнеризація компонентів з використанням сучасних технологічних підходів віртуалізації та ізоляції обчислювального середовища. Технології такі як Kubernetes можуть бути використані на цьому етапі для створення середовищ для тестування, валідації та використання AI-моделі.

механізми паралельного розгортання та інші стратегії направлені на прискорення розгортання та міграції частини програмних компонентів для відмовостійкості системи.

2.3. Завершальним етапом є (постійний) моніторинг та збір аналітики використання впровадженої платформи. Реалізується безперервне діагностування параметрів продуктивності, акумуляція системних метрик та реалізується автоматизована реакція на потенційні девіації функціонування інформаційної інфраструктури. Під девіаціями розуміються відмови як апаратного або програмного рівня. Стратегії відновлення працездатності системи включають або повне, або часткове виконання конвеєру, для відновлення працездатності.

Наявний шлях від отримання нової ітерації моделі в MLOps до подальшого розгортання показаний на Рис. 1. Модель, яка пройшла тест на точність в цьому процесі, надсилається менеджеру моделей разом з відповідним тестовим звітом. Щоб ефективно побудувати інфраструктуру під додатки, які застосовують штучний інтелект, необхідно враховувати різні фактори, як вибір правильних інструментів і технологій. Вибір правильних інструментів і технологій також має важливе значення

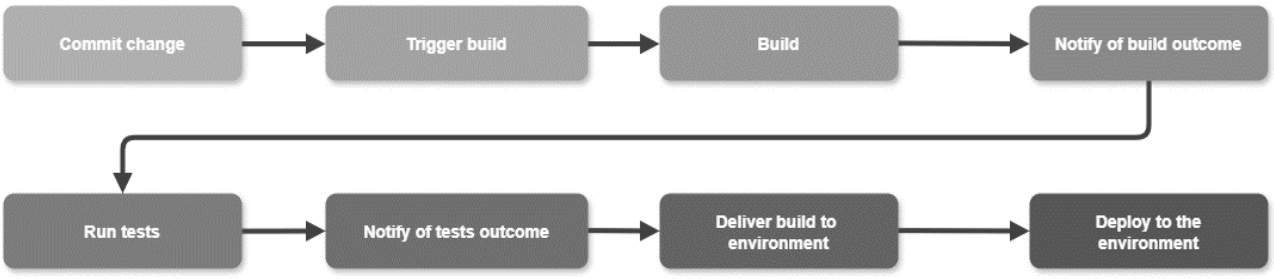


Рис. 1. Етапи CI/CD конвеєра

2.2. Наступним етапом є розгортання, яке реалізується через стратегію інкрементальної оновлення, що передбачає поетапне розгортання оновлених контейнерів (артефактів) із забезпеченням неперервності операційних процесів, тобто, під час оновлення програмної частини, платформа має зберігати працездатний стан. На цьому етапі можуть використовуються для побудови масштабованої та гнучкої інфраструктури для робочих навантажень ШІ.

Етап розгортання і моніторингу є фінальним етапом шляху MLOps. На цьому етапі розробники програмного забезпечення інтегрують перевірені моделі у відповідні додатки і забезпечують стабільність всієї прикладної реалізації платформи. Оскільки

на цьому етапі інтеграції можна отримати різні аспекти рішень і конфігурацій всього конвеєра, інженери MLOps повинні здійснювати безперервний моніторинг моделі, всього додатку і споживаних даних. [34] Іншу роль на цьому етапі розгортання інфраструктури відіграє інженер, який відповідає за проведення, побудову та тестування робочої системи.

Навчена модель розгортається для використання зовнішніми користувачами, за допомогою різних підходів. Два найпоширеніші підходи - це «модель як послуга» та «вбудована модель». У моделі як сервісі модель представлена у вигляді кінцевих точок API Representational State Transfer (REST), тобто модель розгортається на веб-сервері, щоб вона могла взаємодіяти через REST API, і будь-яка програма могла взаємодіяти з AI-платформою, передаючи вхідні дані через виклик API. Веб-сервер може працювати локально або в хмарі. З іншого боку, у випадку «вбудованої моделі», модель упаковується в програму, яка потім публікується. Цей варіант використання є практичним, коли модель розгортається на периферійному пристрої. Потрібно зауважити, що спосіб розгортання AI задачі на платформі залежить від виду взаємодії кінцевого користувача з платформою, в випадку, описаному в даній роботі, передбачається саме модель як сервіс.

Безперервний моніторинг всієї платформи на метрики, наприклад, продуктивності моделі та додатку, а також стану інфраструктури та даних, що використовуються моделлю, єосновою надійного продукту машинного навчання. Зосереджуючись на відстеженні даних через конвеєрні операції, бази даних логів можуть допомогти керувати і підтримувати зв'язок між даними і пов’язними з ними моделями. Зворотній зв’язок, в даному випадку, грає ключову роль у встановленні постійного потоку даних на основі розглянутого конвеєру, з можливістю постійного оновлення даних.

Апаратна частина платформи реалізується за допомогою поєднання Cloud технологій та Edge (вбудованих) пристроїв. Основою такої системи є топологія, вона диктує, фактично, яким чином взаємодіють вузли системи. Ієрархічна структура топології дерева підходить для створення гібридної системи із застосуванням Cloud та Edge технологій. Топологія двійкового дерева є однією з фундаментальних структур у теорії алгоритмів, графів, статистиці та інших комп’ютерних наук. Для бінарного дерева, кожен вузол може мати не більше двох дочірніх вузлів, відомих як лівий і правий, які в свою чергу можуть також бути внутрішніми вузлами або листями.

Основним недоліком бінарного дерева є відносно низька відмовостійкість, коли при одній відмові, можливе фактичне розбиття топології на дві частини. Вирішенням цієї проблеми є використання додаткових дебруйнівських зв’язків. Додаткові зв’язки будуються за наступною формою:

$$\mathrm{D(n,k)} = \mathrm{k} \cdot \mathrm{D(n-1,k)} + (-1)^k * (k-1) \cdot \mathrm{D(n-1,k-1)}$$

де D(n,k) - число Дебруйна, S(n,k) - число Стерлінга другого роду, n - кількість елементів, k - кількість циклів, (k/j) - біноміальний коефіцієнт "k по j" (кількість способів вибрати j елементів з k без урахування порядку), а (-1)m - альтернатуюча сума.

Отже, перший тест має на меті перевірку швидкості розгортання системи для різних значень кількості вузлів. Отримання результатів дозволить проаналізувати часову характеристику платформи, її здатність обробляти запити при збільшенні кількості пристроїв, і про ефективність масштабування з використанням запропонованого методу.

Таблиця 1

Час розгортання платформи в залежності від кількості вузлів

Кількість пристроїв	Час розгортання
3	925.45 секунди
7	977.08 секунди
15	1122.28 секунди
31	1222.31 секунди
63	1359.14 секунди

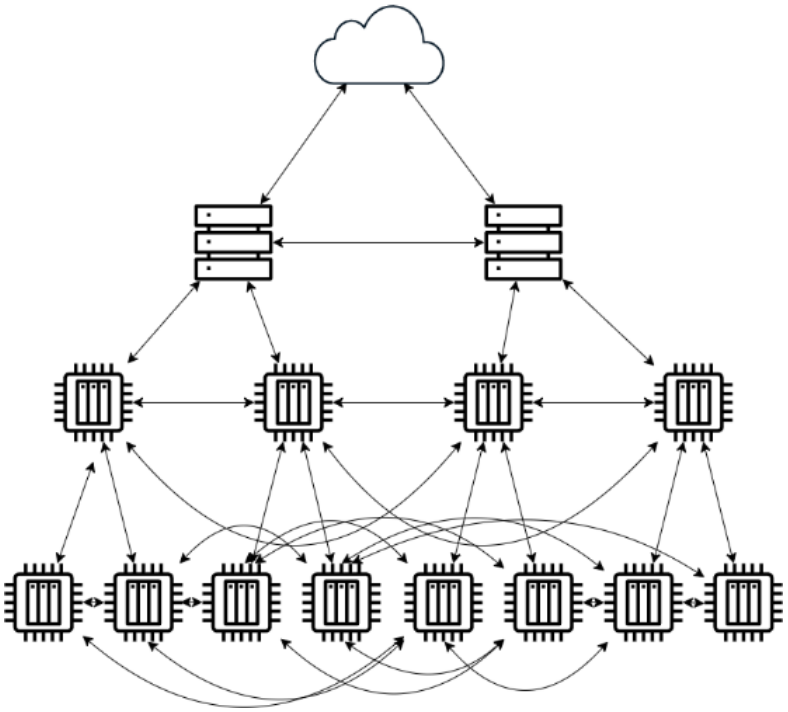


Рис. 2. Схема архітектури платформи на основі дерева з дебруйнівськими зв’язками

Рис. 3. Графік часу розгортання платформи в залежності від кількості вузлів

Вибір технологічного стеку для тестової системи є важливим етапом розробки програмної частини AI платформи, оскільки в залежності від обраного програмного забезпечення та бібліотек, можуть накладатись додаткові обмеження щодо розробки тестової програми. Для функціонального виконання більшості вимог до AI-платформи, зокрема API, завдяки якому можна формалізувати комунікацію з іншими пристроями, управління та логування моделі AI, виконується через мову Python. Python набув популярності завдяки своєму простому синтаксису, інтерпретованості, що дозволило легко впровадити компонентний підхід та розробити широку низку бібліотек для виконання різних функцій, в тому числі, розробки AI.

Другий тест пов'язаний з виконанням тестування навантаження плафторми великою кількістю запитів. Застосування тестування API AI-платформи дозволяє підтвердити її здатність до ефективної роботи у реальних умовах, коли більше ніж один вузол робить запит, і забезпечує певний рівень масштабованості та відмовостійкості. Кожен потік робить 100 запитів на платформу. Характеристиками, які будуть досліджуватись, є середній час відповіді, максимальний, мінімальний час, 99-ий персентиль та стандартне відхилення.

Таблиця 2.

Часові характеристики плафторми під час тестування навантаження

Tavg	Tmax	T_{99}	Tmin	SD	Потоки
MobileNet					
931.61	13872.96	839.44	42.34	1846.33	1
964.92	13776.17	948.56	43.05	1855.14	2
1011.02	11369.37	1008.52	43.75	2068.31	5
1302.33	15227.99	1198.00	45.87	2640.26	10
1535.48	29263.55	1660.89	47.99	3415.72	25
MobileNet Raspberry PI Tree-Debrujin (2 ступінь)					
2525	47039	1653.98	63.517	4010.7	1
2034.1	18519	2458.57	59.989	4346.4	2
2518.8	27836	3252.11	62.106	5607.8	5
3817.9	46181	3882.75	64.929	8098.7	10
4310.1	73512	4861.06	66.623	9674.8	25
MobileNet Raspberry PI Tree-Debrujin (2 ступінь, 2 воркери)					
513.59	13679.38	401.73	42.34	1035.63	1
532.18	13776.17	560.62	42.34	1060.80	2
605.79	11175.8	491.67	43.76	1066.26	5
857.04	10853.17	896.40	48.69	1461.96	10
879.48	15617.72	915.23	46.86	1695.69	25

В табл. 2 представлені результати другого тесту на навантаження платформи. Під час виконання тестування система працювала без відмов і не було перенасичення черги задачі, незважаючи на досить великий об'єм вхідного трафіку, що пояснюється вибором програмного стеку. З результатів видно, що залучення додаткових вузлів до системи значно знижує середній час відповіді.

На Рис. 4 наведені графіки тестування навантаження AI-платформи із застосуванням від 1 до 3 пристроїв, два з яких містять контейнеризовані AI-моделі, а також вузол-агрегатор, який забезпечує балансування навантаження між ними. Це дозволяє зменшити затримки під час обробки запитів та підвищити загальну продуктивність системи, завдяки оптимізації розподілу обчислювальних ресурсів між вузлами платформи. Графіки відображають динаміку швидкості реагування платформи, а також відображають взаємодію між воркерами та контейнерами моделей, що демонструє можливості даної архітектури в умовах реального навантаження.

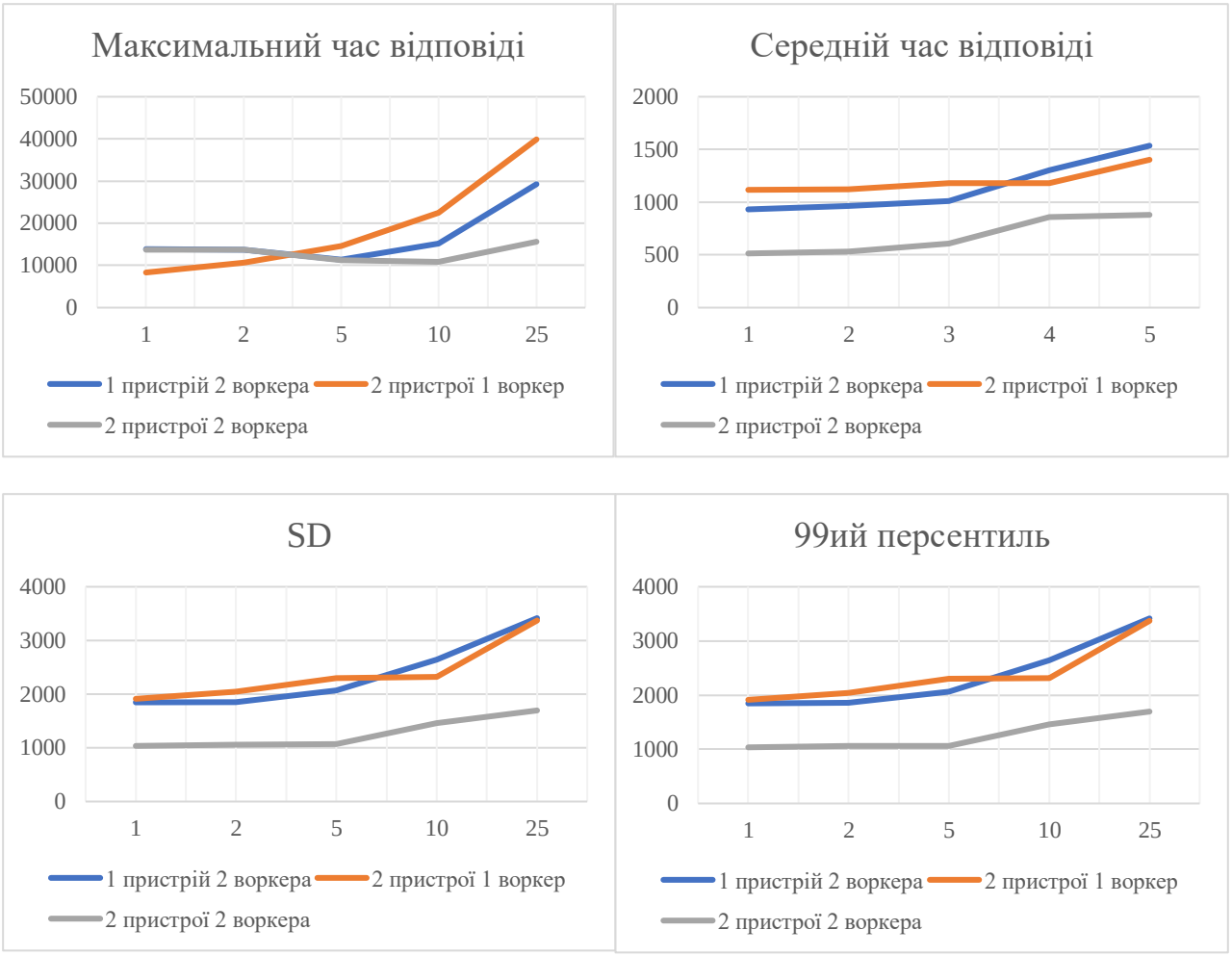


Рис. 4. Порівняльні графіки характеристик тестування системи на навантаження

ВИСНОВКИ

Було досліджено методолгію MLOps і доведено, що її застосування дозволяє створити ефективний спосіб розгортання AI-платформ. Дана платформа передбачає автоматизацію та моніторинг кожного етапу, включаючи збір даних, перетворення, навчання, впровадження та представлення результатів кінцевому користувачеві. Безперервна інтеграція та безперервна доставка сприяють ефективному розгортанні платформи. Етап розгортання забезпечує інтеграцію та стабільність моделі з постійним моніторингом.

Тестування виконано з використанням часових метрик. У таблиці результатів порівнюються метрики для запропонованого підходу, зосереджуючись на часових характеристиках платформи. Тестування розгортання платформи показує, що при масштабуванні системи, з 3 до 63 вузлів час розгортання збільшується лише на 433.69 секунди, що свідчить про здатність платформи до масштабування. Тестування навантаження показує, що платформа працює без відмов, за умови насичення черги запитів до 25 потоків одночасно, при цьому застосування додаткових вузлів в топології дозволяє зменшити середній час відповіді системи.

Запропонований метод є менш складним, ніж його альтернативи, проте він є більш легким, тому потребує менше часу на збірку та переконфігурацію після збоїв. Для підвищення ефективності запропонованого підходу можуть знадобитися додаткові дослідження в галузі оптимізації контейнерів, оптимізації моделей ШІ, а також мережевої взаємодії та її адміністрування. Важливо зазначити, що в реальних умовах можуть знадобитися додаткові кроки для підвищення відмовостійкості системи та збільшення її можливостей масштабування, за умови зміни програмного стеку або апаратної складової.

ЛІТЕРАТУРА

1. **Liu, Y., Ling, Z., Huo, B., Wang, B., Chen, T., & Mouine, E.** (2020). Створення платформи для операцій машинного навчання на основі відкритих джерел. *IFACPapersOnLine*, 53(5), 704–709. https://doi.org/10.1016/j.ifacol.2021.04.161
2. **Granlund, T., Kopponen, A., Stirbu, V., Myllyaho, L., & Mikkonen, T.** (2021). Виклики MLOps у налаштуванні мультиорганізаційної структури: досвід з двох реальних випадків. *2021 IEEE/ACM 1st Workshop on AI Engineering - Software Engineering for AI (WAIN)*, 82–88. https://doi.org/10.1109/WAIN52551.2021.00019
3. **Zhou, Y., Yu, Y., & Ding, B**. (2020). До MLOps: дослідження платформи для конвеєра машинного навчання. *2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*, 494–500. https://doi.org/10.1109/ICAICE51518.2020.00102
4. **Cardoso Silva, L., Rezende Zagatti, F., Silva Sette, B., Nildaimon dos Santos Silva, L., Lucrédio, D., Furtado Silva, D., & de Medeiros Caseli, H**. (2020). Бенчмаркінг рішень машинного навчання в умовах виробництва. *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 626–633. https://doi.org/10.1109/ICMLA51294.2020.00104
5. **Wilkes, B., Milani, A. M. P., & Storey, M. A. (2023)**. Рамкова структура для автоматизації вимірювання досліджень та оцінки показників DevOps (DORA). В *2023 IEEE International Conference on Software Maintenance and Evolution (ICSME)* (стор. 62-72). IEEE.
6. John M.M., Olsson, H.H., & Bosch J. (2021). До MLOps: рамкова структура та модель зрілості. В *2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, IEEE, 1-8.

REFERENCES

1. **Liu, Y., Ling, Z., Huo, B., Wang, B., Chen, T., & Mouine, E.** (2020). Building A Platform for Machine Learning Operations from Open Source Frameworks. IFACPapersOnLine, 53(5), 704–709. https://doi.org/10.1016/j.ifacol.2021.04.161
2. **Granlund, T., Kopponen, A., Stirbu, V., Myllyaho, L., & Mikkonen, T.** (2021). MLOps Challenges in MultiOrganization Setup: Experiences from Two Real-World Cases. 2021 IEEE/ACM 1st Workshop on AI Engineering - Software Engineering for AI (WAIN), 82–88. https://doi.org/10.1109/WAIN52551. 2021 .00019
3. **Zhou, Y., Yu, Y., & Ding, B.** (2020). Towards MLOps: A Case Study of ML Pipeline Platform. 2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE), 494–500. https: //doi.org/10.1109/ICAICE51518.2020.00102
4. **Cardoso Silva, L., Rezende Zagatti, F., Silva Sette, B., Nildaimon dos Santos Silva, L., Lucrédio, D., Furtado Silva, D., & de Medeiros Caseli, H.** (2020). Benchmarking Machine Learning Solutions in Production. 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), 626–633. https://doi.org/10.1109/ICMLA51294.2020.00 104
5. **Wilkes, B., Milani, A. M. P., & Storey, M. A**. (2023). A Framework for Automating the Measurement of DevOps Research and Assessment (DORA) Metrics. In 2023 IEEE International Conference on Software Maintenance and Evolution (ICSME) (pp. 62-72). IEEE.
6. **John M.M., Olsson, H.H., & Bosch J.** (2021). Towards MLOps: A framework and maturity model. In 2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), IEEE, 1-8.

Method of deploying an ai platform using MLOPS methodology

Volodymyr Rusinov

Abstract. The specifics of modern AI applications require the development of a platform that enables continuous support and rapid reconfiguration. Scientific and engineering literature shows that the use of Machine Learning Operations and Machine Learning Pipeline (MLOps) methods when creating AI platforms for embedded systems can increase the efficiency of AI development and implementation. MLOps allows you to automate the development, training, and operation of machine learning models, providing a quick response to changes in the environment or input data. This helps to reduce the time for deployment and maintenance of the system as a whole.

To solve the problem of deploying an AI platform, it is proposed to use a method based on a hybrid Edge-Cloud architecture. This approach is often used in the modern scientific literature and finds application in various fields. The architecture of the platform, based on hierarchical relationships between the Edge and Cloud layers, dictates the need to create a specialized approach to platform deployment. The peculiarity of this architecture is the use of embedded systems directly in AI tasks, namely, pattern recognition. This approach allows reconfiguring the platform, subject to updating the AI model, thereby maintaining its performance and correct task execution.

Given that the key is to balance the speed of reconfiguration of such a platform with fault tolerance, in order to measure the overall performance of the system with a certain degree of accuracy, we will introduce several metrics that will help establish the parameters of the system and compare it with alternative solutions. Deployment of AI software should be accompanied by an analysis of the target system. As part of this work, a number of tests were conducted to quantify system performance.

Keywords: artificial intelligence, MLOps, IoT, fault tolerance, computer networks.

Організація та зберігання великих наборів даних для навчання ШІ

Олексій Мельніков

olexa.m@gmail.com
N-IX Corporation

Received 03.02.2025, accepted 29.03.2025
https://doi.org/10.32347/st.2025.3.1203

Анотація. Зростаюча складність і масштаб застосувань штучного інтелекту (ШІ), особливо в машинному навчанні та глибокому навчанні, вимагають надійних та ефективних рішень для зберігання даних. Це особливо актуально при роботі з великими наборами даних, які часто досягають сотень терабайт, у різноманітних форматах. Для навчання моделей ШІ можуть бути використані різні формати даних, таких як відео та зображення для задач комп'ютерного зору, WAV та IQ для радіо та звукових сигналів, текст для мовних моделей. У цій статті розглядаються наукові підходи до організації зберігання даних для навчання ШІ з акцентом на ефективність та доступність. У статті буде розглянуто особливості роботи з форматами даних HDF5, WAV та IQ raw data, що використовуються для обробки радіосигналів.

Ключові слова: Штучний інтелект (ШІ), Машинне навчання, Глибоке навчання, Набори даних, Зберігання даних, HDF5, WAV, IQ raw data, Хмарне зберігання, Розподілені файлові системи, Локальне зберігання, Платформи керування даними, Версіонування даних, Масштабованість, Продуктивність, Різноманітність даних, Консолідація даних, Якість даних, Візуалізація даних, Безпека даних, HDFView, Шардинг даних, Оптимізація Push-down, Клонування без копіювання, TTL (час життя),ETL (видобування, трансформація, завантаження), Каталогізація, Керування метаданими.

ПОСТАНОВКА ПРОБЛЕМИ

Штучний інтелект стрімко розвивається, і його моделі стають все більш складними та потужними. Ефективність цих моделей безпосередньо залежить від якості та обсягу даних, на яких вони навчаються.

Збір, організація та підготовка даних є критично важливими етапами в процесі розробки ШІ, і дослідження в цій області

постійно розвиваються. Обсяги даних, що беруть участь у процесі навчання,

Oleksii Melnikov
Software engineer
at N-IX corporation

зростають у геометричній прогресії. Керування та організація таких масивів даних потребують спеціалізованих рішень які можуть забезпечити оперативний, постійний та автоматизований доступ до даних навчання.

Аналіз останніх досліджень та публікацій. Формат HDF5, його структура, переваги та способи обробки описані в [1] та [2]. Ці матеріати пояснюють, як HDF5 може ефективно використовуватися для зберігання великих наборів даних, включаючи дані різних типів, та як цей формат забезпечує ефективний доступ до даних та їх обробку. Практичні аспекти роботи з HDF5, включаючи інструменти для перегляду та редагування HDF5 файлів, а також способи оптимізації запису даних у цей формат розглянуті в [21], [22]. Використання формату даних HDF5 для зберігання спектрограмних даних та порівняння його з іншими форматами, такими як WAV та sigMF, для зберігання IQ даних подано в [3]. Різні рішення для зберігання даних для навчання ШІ, включаючи хмарні сервіси (AWS S3, Google Cloud Storage, Azure Blob Storage), розподілені файлові системи (DFS) та локальні рішення (SAN, NAS, JBOD) розглядаються в [4], [6], [7] та [10]. В них наведено аналіз переваг та недоліків кожного рішення, а також фактори, які слід враховувати при виборі. Об'єктно-

орієнтовані методи зберіганні даних, їх переваги та використання в хмарних середовищах описані в [11], [23]. В [24] порівнюються розподілені файлові системи та розподілене об'єктно-орієнтоване зберігання, виділяючи їх ключові відмінності та випадки використання. Огляд популярних рішень для зберігання великих даних, включаючи AWS S3, Google Cloud Storage та Azure Blob Storage [9]. Огляд платформ управління даними для навчання ШІ, таких як Velotix, Databricks, DataRobot, Google Cloud AI Platform, Azure Synapse Analytics, IBM Watson Studio, Snowflake, Amazon SageMaker, Dataiku та Vast Data [5], [12], [25], [26], [27] . Вони описують ключові функції кожної з платформи та випадки їх використання. Важливість версійності даних для навчання ШІ, опис різних інструментів та методи версійності (DVC, LakeFS, GitLFS), а також найкращі практики для управління версіями даних розглянуті в [8], [13], [14], [15], [28]. Стратегії управління великими наборами даних розглянуті в [7]. Системи каталогізації даних розглянуті в [29], [30], [31], [32], [33]

МЕТА СТАТТІ

Надати всебічний огляд та рекомендації щодо організації зберігання великих наборів даних, які використовуються для навчання штучного інтелекту (ШІ), зокрема в контексті машинного та глибокого навчання, а також інструменти для ефективної організації зберігання даних, що сприятиме підвищенню продуктивності та інновацій у цій галузі.

Виклад основного матеріалу дослідження. У межах досліджень у сфері розпізнавання радіосигналів дані для досліджень збираються за допомогою різноманітних технічних засобів, призначених для роботи з радіо середовищем, що включають в себе радіоприймачі, SDR (Software Defined Radio - програмно-визначене радіо) приймачі, спектральні аналізатори, тощо. Дані, що обробляються, можуть представляти собою як оцифровану хвильову форму радіосигналу, так і набір його частотних характеристик у часі. Такий метод збору даних продукує значні обсяги даних, що може досягати десятків або навіть сотень терабайт за короткий проміжок часу. Для оцінки можливих обсягів даних може бути застосована теорема Котельникова (Найквіста-Шеннона), що визначає мінімальну частоту дискретизації.

Однак, при обробці сигналів з обмеженою шириною пропускання можна застосовувати нижчі частоти дискретизації, що значно зменшує обсяги даних. Проте подальша обробка, така як видалення шумів та виділення ознак для навчання, може збільшити обсяги даних у декілька разів, особливо за умови застосування версіонування даних.

Збір даних з приймача включає кілька етапів. Спочатку сирі дані записуються на локальний диск. Потім проводиться постобробка для очищення від артефактів та конвертація у вибраний формат. Далі реєструються метадані для зручного доступу. Насамкінець, оброблені дані синхронізуються з віддаленими сховищами для резервного копіювання.

Вибір формату зберігання є важливим етапом дослідження, тому в межах роботи обрані формати HDF5, WAV та IQ raw data, які забезпечують оптимальне співвідношення між продуктивністю, точністю, сумісністю та зручністю роботи із записаними зразками радіосигналів. Розглянемо запропоновані формати.

HDF5 — це ієрархічний формат даних, який добре підходить для керування великими, складними та різнорідними даними, що часто зустрічаються в робочих навантаженнях ШІ [1]. Він дозволяє зберігати різні типи даних, такі як зображення, числові масиви та текст, в одному файлі, зберігаючи чітку організаційну структуру. Файл HDF5 містить метадані, які описують дані, що він містить, що полегшує розуміння та аналіз даних без необхідності покладатися на зовнішні джерела пов’язаних метаданих [2]. HDF5 забезпечує ефективні операції вводу/виводу (I/O) [1].

WAV (Waveform Audio File Format) — це стандартний формат аудіо файлів. Цей

формат може застосовується для зберігання даних зразків радіосигналів. Через неможливість зберігання детальної мета інформації, дані в цьому форматі потребують додаткових зусиль для супроводження та організації доступу.

IQ data представляє синфазну та квадратур-ну складові сигналу, фіксуючи як ампліту-ду, так і фазову інформацію [3]. Цей тип даних продукується SDR приймачами і мо-же бути оброблений безпосередньо алго-ритми ШІ, що може зменшити загальні ви-моги для перетворення в проміжні формати. Однак, як і у випадку формату WAV , обробка великих обсягів IQ data може бути складною через його складність і обме-женість збереження супроводжувальних даних.

Використані формати не передбачають компресію даних під час збору та обробки. Тому обсяги даних постійно зростатимуть, що потребує масштабованих сховищ. Масштабованість дозволяє плавно нарощувати обсяги сховища відповідно до потреб, забезпечуючи безперебійну роботу та доступність даних, що критично важливо для безперервності процесів аналізу та навчання.

Традиційні системи зберігання часто важко масштабувати, що знижує продуктивність і збільшує витрати [4]. Наприклад, застарілі системи прямого підключення сховищ (DAS) мають обмежену ємність. Також системи мережевого підключення сховищ (NAS) можуть не забезпечити достатньої продуктивності для високонавантажених завдань ШІ.

Окрім масштабованості, існує кілька інших проблем, які необхідно вирішити при організації зберігання даних для навчання ШІ [4]. Навчання ШІ часто включає обчислювально інтенсивні завдання, які вимагають високошвидкісного доступу до даних, що призводить до затримок у навчанні та збільшення часу обробки. Застосунки ШІ використовують різноманітні типи даних, включаючи структуровані та неструктуровані формати, такі як текст, зображення, аудіо та відео. Традиційні системи зберігання, такі як реляційні бази даних, призначені для структурованих даних, можуть мати труднощі з ефективною обробкою різноманітності.

Розрізнені сховища, де дані розподілені по різних системах і форматах, перешкоджають ефективній обробці через необхідність об’єднання даних на кожній ітерації. Саме тому консолідація даних в єдину платформу має вирішальне значення для покращення доступності даних та оптимізації робочих процесів ШІ [7]. Великі набори даних можуть містити неузгодженості, помилки та пропущені значення. Це негативно впливає на продуктивність моделей ШІ, тому критично важливими є інструменти для валідації та виправлення даних. Для валідації навчальних даних можуть бути застосовані методи візуалізації та аналіз. Аналіз великих наборів даних ускладнюється їхнім обсягом [7].

Традиційні методи візуалізації можуть бути неефективними. Тому для отримання змістовної інформації потрібні спеціалізовані методи та інструменти, такі як вибірка та агрегування.

Розв'язання цих проблем вимагає впровадження сучасних рішень для організації даних, які спеціально розроблені для робочих навантажень ШІ.

Хмарні рішення для зберігання даних стають все більш популярними для застосувань ШІ завдяки їх масштабованості, гнучкості та економічній ефективності [6]. Хмарні провайдери, такі як AWS, GCP та Azure, пропонують різноманітні послуги зберігання, адаптовані до різних потреб і бюджетів. Наприклад AWS пропонує сервіс S3 (Amazon Simple Storage Service) — це високоефективний сервіс об'єктного зберігання, який пропонує різні класи зберігання для забезпечення різних моделей доступу та вимог до вартості [9]. Наприклад, S3 Standard підходить для даних, до яких часто звертаються, тоді як S3 Glacier призначений для довгострокового архівування даних. S3 також безперешкодно інтегрується з іншими сервісами AWS, що робить його зручним вибором для організацій, які вже використовують

екосистему AWS. Компанія Google надає сервіс Google Cloud Storage (GCS) що має аналогічні можливості, що й S3, з акцентом на тісну інтеграцію з іншими сервісами Google Cloud [9]. Цей сервіс пропонує різні класи зберігання, включаючи Standard, Nearline, Coldline та Archive, що дозволяє вибрати найбільш економічно вигідний варіант доступу до даних [10].

Компанія Microsoft надає Azure Blob Storage — це ще одне масштабоване та безпечне хмарне рішення для зберігання, призначене для обробки великих обсягів неструктурованих даних [9]. Azure Blob Storage підтримує неструктуровані типи даних і пропонує такі функції, як Write Once, Read Many (WORM) storage, щоб забезпечити незмінність даних для цілей відповідності та архівування.

Хмарні провайдери також пропонують інструменти передачі даних для полегшення ефективного переміщення та завантаження даних у хмарне сховище [10]. Наприклад, Storage Transfer Service від Google дозволяє передавати великі обсяги даних з різних джерел, включаючи локальне сховище, інших хмарних провайдерів та загальнодоступні набори даних, у бакети GCS.

Однією з ключових особливостей хмарного розміщення даних є також доступність гнучких і масштабованих обчислювальних потужностей, які пропонують перераховані вище провайдери. Це дозволяє реалізувати конвеєри обробки даних на стороні сховища. Однак важливо пам’ятати про потенційні проблеми безпеки та конфіденційності даних, пов’язані з хмарним зберіганням [6]. Хоча хмарні провайдери впроваджують надійні заходи безпеки, все ж необхідно ретельно оцінювати конкретні потреби та вибирати відповідні засоби контролю безпеки, такі як шифрування та контроль доступу, щоб зменшити потенційні ризики.

Хоча хмарне зберігання пропонує зручність і масштабованість, локальне зберігання забезпечує кращий контроль та безпеку даних [6]. Це може бути вирішальним фактором для випадків із конфіденційними та критичними даними, які не можна зберігати в хмарі.

Розглянемо рішення локального зберігання даних. Мережі зберігання даних (SAN), що забезпечують високошвидкісний доступ до централізованого сховища, що робить їх придатними для додатків, які потребують низької затримки та високої пропускної здатності [6]. Мережеві системи зберігання даних (NAS): Системи NAS пропонують масштабовану ємність зберігання та меншу складність в управлінні, ніж SAN [6]. Вони добре підходять для зберігання та обміну великими обсягами неструктурованих даних, які зазвичай використовуються в проектах ШІ. JBOD (Just Bunch Of Disks), це економічно ефективний спосіб розширити ємність зберігання шляхом безпосереднього підключення окремих жорстких дисків до сервера [6]. JBOD пропонує гнучкість з точки зору конфігурації зберігання, але може не мати розширених функцій, та вимагати додаткових зусиль для підтримання надійного зберігання та доступності.

Розподілені файлові системи (DFS), наприклад Hadoop DFS [11], дозволяють ефективно зберігати та керувати великими обсягами даних, розподіляючи їх між багатьма серверами, що забезпечує високу швидкість та доступність, особливо для завдань, де потрібна паралельна обробка даних. Завдяки здатності розділяти дані на частини та розподіляти їх по різних вузлах можливо реалізувати розміщення даних найближче або безпосередньо на спеціалізованому обладнанні [11].

Окрім зберігання великих об’ємів даних варто звернути увагу на спеціалізовані рішення для керування даними які існують на ринку.

Розглянемо кілька відомих комерційних платформ керування даними, спеціально розроблених для навчання ШІ, які пропонують функції, що виходять за рамки простого зберігання та пошуку. Ці платформи надають інструменти для версіонування даних, керування моделями, відстеження експериментів та колективної роботи, оптимізуючі весь життєвий цикл розробки ШІ.

Наведені в табл. 1 платформи пропонують різні рівні автоматизації, масштабованості та простоти використання. Наприклад,

DataRobot зосереджується на AutoML, щоб спростити розробку моделей, тоді як Dataiku наголошує на командній співпраці в дослідженні даних. Вибір правильної платформи залежить від конкретних потреб проекту ШІ та досвіду команди з аналізу даних.

Багато з цих платформ включають функції для оптимізації управління даними та продуктивності.

.

Таблиця 1.

Ключові характеристики та приклади застосування комерційних платформ організації даних

Платформа	**Ключові особливості**	**Приклади застосування**
Velotix	Безпека даних на основі ШІ, контроль доступу	Управління даними, відповідність, безпечний доступ до даних
Databricks	Уніфікована аналітична платформа, інтеграція штучного інтелекту, керування великомасштабними даними	Створення та розгортання моделей машинного навчання, дослідження даних
DataRobot	Автоматизоване машинне навчання (AutoML), розробка та розгортання моделі	Прогностична аналітика, оцінка ризиків, виявлення шахрайства
Google Cloud AI Platform	Комплексний набір інструментів машинного навчання, інтеграція TensorFlow, AutoML	Розробка, розгортання та управління моделлю, НЛП, комп'ютерний зір
Azure Synapse Analytics	Великі дані та сховища даних, аналіз даних на основі ШІ	Інтеграція даних, аналітика та машинне навчання
IBM Watson Studio	Підготовка даних, розробка моделей, застосування NLP	Прогнозне моделювання, оптимізація обслуговування клієнтів
Snowflake	Хмарне сховище даних, інтеграція ШІ, масштабований аналіз даних	Обмін даними, аналітика в реальному часі, сховище даних
Amazon SageMaker	Спрощена розробка та розгортання моделі машинного навчання	Створення, навчання та розгортання моделей ШІ з підтримкою масштабування
Dataiku	Співпраця над науковими проектами даних, підготовка даних, моделювання та розгортання	Оцінка ризиків, аналіз клієнтів, запобігання шахрайству
Vast Data	Масштабована платформа даних, уніфіковане сховище, база даних і обчислення	ШІ та глибоке навчання, високопродуктивні обчислення

Оптимізація Push-down використовує обчислювальну потужність сховища даних для ефективного виконання перетворень даних [12].

Замість переміщення даних до окремого вузла обробки, оптимізація push-down переносить обробку на сторону сховища даних, зменшуючи переміщення даних та

підвищуючи продуктивність. Функція "Time Machine" дозволяє отримати доступ до попередніх версій даних, створюючи контрольні журнали та забезпечуючи відновлення даних [12].

Клонування без копіювання дозволяє створювати копії даних без їх фізичного дублювання, зменшуючи витрати на зберігання та прискорюючи процеси розробки та тестування [12].

Ще одним ключовим аспектом організації зберігання даних для навчання ШІ є версіонування даних. Воно дозволяє відстежувати зміни в наборах даних, за потреби повертатися до попередніх версій та забезпечувати відтворюваність експериментів зі ШІ [13]. Інструменти версіонування даних, такі як DVC (Data Version Control) та LakeFS, надають функціональність, подібну до Git, для управління наборами даних. Вони дозволяють створювати гілки, фіксувати зміни та об'єднувати різні версії даних. Це забезпечує структурований та контрольований спосіб відстеження еволюції даних протягом усього процесу розробки ШІ. Важливо дотримуватись правильних підходів для роботи з версійованими даними описаними в [14], це дасть змогу уникнути типових помилок та пришвидшити роботу з проектом. Версіонування даних не тільки покращує управління даними, але й підвищує рівень співпраці та відтворюваності в дослідженнях ШІ [13]. Відстежуючи версії даних та параметри експериментів, дослідники можуть ділитися своєю роботою, відтворювати результати та спиратися на результати один одного. Значні об'єми даних які приймають участь в проектах ШІ потребують ефективних методів доступу до даних, оскільки швидкість доступу безпосередньо впливає на швидкість і продуктивність алгоритмів ШІ та їх навчання [29]. Інструментами для подолання складнощів управління даними для навчання ШІ є використання каталогів даних. Ці платформи, часто доповнені можливостями штучного інтелекту, забезпечують централізоване управління метаданими, полегшуючи пошук, розуміння та використання даних [30]. Ключові функції таких каталогів включають автоматизоване керування метаданими, перевірки якості даних для підтримки їхньої надійності, інструменти для забезпечення управління та відповідності нормативним вимогам, функції для співпраці між командами, що працюють з даними та автоматизацію на основі ШІ/МО для таких завдань, як тегування та класифікація [30].

Розглянемо кілька популярних рішень для каталогізації даних, які широко використовуються для управління наборами даних, призначеними для навчання моделей ШІ. Atlan [32] відомий завдяки своїм функціям для співпраці та додатковими інструментами на основі штучного інтелекту, таким як автоматичне тегування та семантичний пошук [31]. Informatica Enterprise Data Catalog [33] надає розширені функції машинного навчання для профілювання та очищення даних, а також комплексні можливості управління даними [34]. AWS Glue Data Catalog [32] є важливим компонентом екосистеми AWS, що надає безсерверне сховище метаданих з автоматичним виявленням схем та інтеграцією з іншими сервісами AWS [36]. Google Cloud Dataplex Catalog (раніше відомий як Data Catalog) [32] розроблений для легкої інтеграції з сервісами Google Cloud та пропонує такі функції, як керування даними та метаданими, а також інтеграцію з Vertex AI та розширену аналітику [37]. Alation Data Catalog [32] є визнаним лідером у галузі. Використовуючи штучний інтелект та машинне навчання Alation Data Catalog спрощує виявлення даних, покращує спільну роботу та забезпечує управління даними [34]. У таблиці 2 наведені ключові функції платформ каталогізації даних для навчання моделей ШІ.

Впровадження чітко визначеної стратегії зберігання та керування даними, яка враховує збереження, організацію, версіювання та каталогізацію даних забезпечить ефективний процес навчання ШІ з оптимальною продуктивністю, масштабованістю та доступністю.

Таблиця 2.
Ключові функції та переваги платформ каталогізації даних

Платформа	**Ключові особливості**	**Приклади застосування**
Velotix	Безпека даних на основі ШІ, контроль доступу	Управління даними, відповідність, безпечний доступ до даних
Databricks	Уніфікована аналітична платформа, інтеграція штучного інтелекту, керування великомасштабними даними	Створення та розгортання моделей машинного навчання, дослідження даних
DataRobot	Автоматизоване машинне навчання (AutoML), розробка та розгортання моделі	Прогностична аналітика, оцінка ризиків, виявлення шахрайства
Google Cloud AI Platform	Комплексний набір інструментів машинного навчання, інтеграція TensorFlow, AutoML	Розробка, розгортання та управління моделлю, НЛП, комп'ютерний зір
Azure Synapse Analytics	Великі дані та сховища даних, аналіз даних на основі ШІ	Інтеграція даних, аналітика та машинне навчання
IBM Watson Studio	Підготовка даних, розробка моделей, застосування NLP	Прогнозне моделювання, оптимізація обслуговування клієнтів
Snowflake	Хмарне сховище даних, інтеграція ШІ, масштабований аналіз даних	Обмін даними, аналітика в реальному часі, сховище даних
Amazon SageMaker	Спрощена розробка та розгортання моделі машинного навчання	Створення, навчання та розгортання моделей ШІ з підтримкою масштабування
Dataiku	Співпраця над науковими проектами даних, підготовка даних, моделювання та розгортання	Оцінка ризиків, аналіз клієнтів, запобігання шахрайству
Vast Data	Масштабована платформа даних, уніфіковане сховище, база даних і обчислення	ШІ та глибоке навчання, високопродуктивні обчислення

ВИСНОВОК

Навчання алгоритмів ШІ та глибокого навчання вимагає значних об’ємів вихідних даних різноманітних форматів, а також складних, багаторівневих алгоритмів обробки цих даних. Така складність потребує особливої уваги до організації цих даних, що буде включати збереження вхідних даних, даних постобробки, метаданих та інших супутніх даних, що будуть супроводжувати процес навчання. Зберігання даних для навчання ШІ з великими наборами даних вимагає комплексної стратегії, яка враховує різні фактори, включаючи масштабованість, продуктивність, різноманітність даних та специфічні формати даних. Хмарне зберігання, розподілені файлові системи та

локальні рішення пропонують різні переваги залежно від конкретних потреб застосунку ШІ.

Для потреб навчання ШІ у сфері розпізнавання сигналів, що включають HDF5, WAV та IQ data, комбінація хмарного зберігання та платформи керування даними, подібної до тих, що перелічені в таблицях 1 та 2, може бути найбільш відповідним та оптимальним рішенням. Хмарне зберігання забезпечує масштабованість та гнучкість для обробки даних, тоді як платформа керування даними пропонує інструменти для версіонування, каталогізації та ефективного доступу до даних. Версіонування даних забезпечує відтворюваність, відстежує зміни та дозволяє повертатися до попередніх версій, тоді як методи каталогізації оптимізують пошук даних та прискорюють навчання ШІ. Вибір конкретної стратегії організації даних залежить від умов проекту, але попри все, правильний вибір стратегії дозволить скоротити витрати та пришвидшити процес навчання.

ЛІТЕРАТУРА

1. HDF5 format - GATK - Broad Institute. [Електронний ресурс]. Доступно: https://gatk.broadinstitute.org/hc/en-us/articles/360035531712-HDF5-format. Дата звернення: 25.02.2025.
2. Hierarchical Data Formats - What is HDF5? | NSF NEON. [Електронний ресурс]. Доступно: https://www.neonscience.org/resources/learning-hub/tutorials/about-hdf5. Дата звернення: 25.02.2025.
3. Decoding artifacts(hdf5) or converting artifacts to .ogg file - Libre Space Community. [Електронний ресурс]. Доступно: https://community.libre.space/t/decoding-artifacts-hdf5-or-converting-artifacts-to-ogg-file/11447. Дата звернення: 25.02.2025.
4. Why Storage Is the Unsung Hero for AI. [Електронний ресурс]. Доступно: https://blog.purestorage.com/perspectives/why-storage-is-the-unsung-hero-for-ai/. Дата звернення: 25.02.2025.
5. Best 10 AI Solutions for Data Management and Security | Velotix. [Електронний ресурс]. Доступно: https://www.velotix.ai/resources/blog/best-ai-solutions-for-data-management-and-security/. Дата звернення: 25.02.2025.
6. AI Storage: A Deep Dive into the Latest Technologies - Nfina. [Електронний ресурс]. Доступно: https://nfina.com/ai-storage/. Дата звернення: 25.02.2025.
7. Effectively Handling Large Datasets - Dataiku blog. [Електронний ресурс]. Доступно: https://blog.dataiku.com/effectively-handling-large-datasets. Дата звернення: 25.02.2025.
8. [Data Versioning Explained: Guide, Examples & Best Practices - lakeFS. [Електронний ресурс]. Доступно: https://lakefs.io/blog/data-versioning/. Дата звернення: 25.02.2025.
9. 10 Big Data Storage Solutions & Systems to Use In 2025 | Airbyte. [Електронний ресурс]. Доступно: https://airbyte.com/top-etl-tools-for-sources/big-data-storage-solutions. Дата звернення: 25.02.2025.Design storage for AI and ML workloads in Google Cloud | Cloud Architecture Center. [Електронний ресурс]. Доступно: https://cloud.google.com/architecture/ai-ml/storage-for-ai-ml. Дата звернення: 25.02.2025.
10. What Is A Distributed Storage System - ScaleGrid. Дата звернення: 25.02.2025. https://scalegrid.io/blog/what-is-a-distributed-storage-system/. Дата звернення: 25.02.2025.
11. 10 Best Data Management Platforms (DMPs) for 2025 - Matillion. [Електронний ресурс]. Доступно: https://www.matillion.com/learn/blog/data-management-platforms. Дата звернення: 25.02.2025.
12. Best Practices for Data Versioning for Building Successful ML Models - Encord. [Електронний ресурс]. Доступно: https://encord.com/blog/data-versioning/. Дата звернення: 25.02.2025.
13. Data Versioning: Best Practices And Tools For Data Teams - Monte Carlo Data. [Електронний ресурс]. Доступно: https://www.montecarlodata.com/blog-data-versioning-guide/. Дата звернення: 25.02.2025.
14. MLflow Data Versioning: Techniques, Tools & Best Practices - lakeFS. [Електронний ресурс]. Доступно: https://lakefs.io/blog/mlflow-data-versioning/. Дата звернення: 25.02.2025.
15. Indexing Strategies to Optimize Database Performance - Developer Nation. [Електронний ресурс]. Доступно: https://www.developernation.net/blog/8-indexing-strategies-to-optimize-database-performance/. Дата звернення: 25.02.2025.

16. Everything You Need to Know When Assessing Data Indexing Techniques Skills - Alooba. [Електронний ресурс]. Доступно: https://www.alooba.com/skills/concepts/database-and-storage-systems/database-management/data-indexing-techniques/. Дата звернення: 25.02.2025.
17. 8 Data Indexing Strategies for Faster & Efficient Retrieval - Crown Records Management. [Електронний ресурс]. Доступно: https://www.crownrms.com/insights/data-indexing-strategies/. Дата звернення: 25.02.2025.
18. Efficient Data Extraction Techniques for Large Datasets - TDAN.com. [Електронний ресурс]. Доступно: https://tdan.com/efficient-data-extraction-techniques-for-large-datasets/32018. Дата звернення: 25.02.2025.
19. In-Depth Industry Outlook: Distributed File Systems And Object Storage Solutions Market Size & Forecast. [Електронний ресурс]. Доступно: https://www.verifiedmarketresearch.com/product/distributed-file-systems-and-object-storage-solutions-market/. Дата звернення: 25.02.2025.
20. Toolbox Talk: Storing data using HDF5 files. [Електронний ресурс]. Доступно: https://dezeraecox.com/toolbox-talk-hdf5-files/. Дата звернення: 25.02.2025.
21. Write performance of raw data - HDF5 Library - HDF Forum. [Електронний ресурс]. Доступно: https://forum.hdfgroup.org/t/write-perfomance-of-raw-data/10963. Дата звернення: 25.02.2025.
22. What is object storage or object-based storage? - NetApp. [Електронний ресурс]. Доступно: https://netapp.com/data-storage/storagegrid/what-is-object-storage. Дата звернення: 25.02.2025.
23. Distributed File Systems vs Distributed Object Storage - GeeksforGeeks. [Електронний ресурс]. Доступно: https://geeksforgeeks.org/distributed-file-systems-vs-distributed-object-storage. Дата звернення: 25.02.2025.
24. VAST Data: Data Platform Built for Deep Learning and AI. [Електронний ресурс]. Доступно: https://vastdata.com. Дата звернення: 25.02.2025.
25. Informatica: AI Powered Cloud Data Management. [Електронний ресурс]. Доступно: https://informatica.com. Дата звернення: 25.02.2025.
Top 10 Data Platforms that Use Generative AI - ScikIQ. [Електронний ресурс]. Доступно: https://scikiq.com/blog/top-10-data-platforms-that-use-generative-ai. Дата звернення: 25.02.2025.
26. Top 6 Dataset Version Control Tools for your Machine Learning workflow - Twine Blog. [Електронний ресурс]. Доступно: https://twine.net/blog/dataset-version-control-tools. Дата звернення: 25.02.2025.
27. AI Training Data: Benefits, Challenges, Example [2025] - Shaip. [Електронний ресурс]. Доступно: https://www.shaip.com/blog/the-only-guide-on-ai-training-data-you-will-need-in/. Дата звернення: 25.02.2025.
28. What Is an AI Data Catalog? | Secoda. [Електронний ресурс]. Доступно: https://www.secoda.co/blog/ai-data-catalog. Дата звернення: 25.02.2025.
29. What is a Machine Learning Data Catalog? - 2024 Guide - Atlan. Дата звернення: 15.03.2025. https://atlan.com/machine-learning-data-catalog/. Дата звернення: 25.02.2025.
30. Best Machine Learning Data Catalog Software: User Reviews from March 2025 - G2. [Електронний ресурс]. Доступно: https://www.g2.com/categories/machine-learning-data-catalog. Дата звернення: 25.02.2025.
31. Top 26 Data Catalog Tools to Consider in 2025 - lakeFS. [Електронний ресурс]. Доступно: https://lakefs.io/blog/top-data-catalog-tools/. Дата звернення: 25.02.2025.
32. Machine Learning Data Catalog Benefits & Use Cases | Informatica. [Електронний ресурс].Доступно: https://www.informatica.com/blogs/machine-learning-data-catalog.html.html.html . Дата звернення: 25.02.2025.
33. The Data Scientist's Guide to the Data Catalog - Alation. [Електронний ресурс]. Доступно: https://www.alation.com/blog/data-scientist-guide-data-catalog/.Дата звернення: 25.02.2025.
34. AWS Glue Features and Benefits for Modern ETL Workflows - CelerData. [Електронний ресурс]. Доступно: https://celerdata.com/glossary/aws-glue-features-and-benefits-for-modern-etl-workflows. Дата звернення: 25.02.2025.
35. Dataplex Catalog overview | Google Cloud. [Електронний ресурс]. Доступно: https://cloud.google.com/dataplex/docs/catalog-overview. Дата звернення: 25.02.2025.

REFERENCES

1. HDF5 format - GATK - Broad Institute. [Online]. Available: https://gatk.broadinstitute.org/hc/en-us/articles/360035531712-HDF5-format. Accessed on: 25 Feb. 2025
2. Hierarchical Data Formats - What is HDF5? | NSF NEON. [Online]. Available: https://www.neonscience.org/resources/learning-hub/tutorials/about-hdf5. Accessed on: 25 Feb. 2025
3. Decoding artifacts(hdf5) or converting artifacts to .ogg file - Libre Space Community. [Online]. Available: https://community.libre.space/t/decoding-artifacts-hdf5-or-converting-artifacts-to-ogg-file/11447. Accessed on: 25 Feb. 2025
4. Why Storage Is the Unsung Hero for AI. Accessed on: 25 Feb. 2025 [Online]. Available: https://blog.purestorage.com/perspectives/why-storage-is-the-unsung-hero-for-ai/. Accessed on: 25 Feb. 2025
5. Best 10 AI Solutions for Data Management and Security | Velotix. Accessed on: 25 Feb. 2025 [Online]. Available: https://www.velotix.ai/resources/blog/best-ai-solutions-for-data-management-and-security/. Accessed on: 25 Feb. 2025
6. AI Storage: A Deep Dive into the Latest Technologies - Nfina. Accessed on: 25 Feb. 2025 [Online]. Available: https://nfina.com/ai-storage/. Accessed on: 25 Feb. 2025
7. Effectively Handling Large Datasets - Dataiku blog. Accessed on: 25 Feb. 2025 [Online]. Available: https://blog.dataiku.com/effectively-handling-large-datasets. Accessed on: 25 Feb. 2025
8. Data Versioning Explained: Guide, Examples & Best Practices - lakeFS. Accessed on: 25 Feb. 2025 [Online].
9. Available: https://lakefs.io/blog/data-versioning/. Accessed on: 25 Feb. 2025
10. 10 Big Data Storage Solutions & Systems to Use In 2025 | Airbyte. Accessed on: 25 Feb. 2025 [Online]. Available: https://airbyte.com/top-etl-tools-for-sources/big-data-storage-solutions. Accessed on: 25 Feb. 2025
11. Design storage for AI and ML workloads in Google Cloud | Cloud Architecture Center. Accessed on: 25 Feb. 2025 [Online]. Available: https://cloud.google.com/architecture/ai-ml/storage-for-ai-ml. Accessed on: 25 Feb. 2025
12. What Is A Distributed Storage System - ScaleGrid. Accessed on: 25 Feb. 2025https://scalegrid.io/blog/what-is-a-distributed-storage-system/. Accessed on: 25 Feb. 2025
13. 10 Best Data Management Platforms (DMPs) for 2025 - Matillion. Accessed on: 25 Feb. 2025 [Online]. Available: https://www.matillion.com/learn/blog/data-management-platforms. Accessed on: 25 Feb. 2025
14. Best Practices for Data Versioning for Building Successful ML Models - Encord. Accessed on: 25 Feb. 2025 [Online]. Available: https://encord.com/blog/data-versioning/. Accessed on: 25 Feb. 2025
15. Data Versioning: Best Practices And Tools For Data Teams - Monte Carlo Data. Accessed on: 25 Feb. 2025 [Online]. Available: https://www.montecarlodata.com/blog-data-versioning-guide/. Accessed on: 25 Feb. 2025
16. MLflow Data Versioning: Techniques, Tools & Best Practices - lakeFS. Accessed on: 25 Feb. 2025 [Online]. Available: https://lakefs.io/blog/mlflow-data-versioning/. Accessed on: 25Feb. 2025
17. 8 Indexing Strategies to Optimize Database Performance - Developer Nation. Accessed on: 25 Feb. 2025 [Online]. Available:https://www.developernation.net/blog/8-indexing-strategies-to-optimize-database-performance/. Accessed on: 25 Feb. 2025
18. Everything You Need to Know When Assessing Data Indexing Techniques Skills - Alooba. Accessed on: 25 Feb. 2025 [Online]. Available: https://www.alooba.com/skills/concepts/database-and-storage-systems/database-management/data-indexing-techniques/. Accessed on: 25 Feb. 2025
19. Data Indexing Strategies for Faster & Efficient Retrieval - Crown Records Management. Accessed on:25Feb.2025[Online].Available:https://www.crownrms.com/insights/data-indexing-strategies/. Accessed on: 25 Feb. 2025
20. Efficient Data Extraction Techniques for Large Datasets - TDAN.com. Accessed on: 25 Feb. 2025 [Online]. Available: https://tdan.com/efficient-data-extraction-techniques-for-large-datasets/32018. Accessed on: 25 Feb. 2025
21. In-Depth Industry Outlook: Distributed File Systems And Object Storage Solutions Market Size & Forecast. Accessed on: 25 Feb. 2025 [Online]. Available: https://www.verifiedmarketresearch.com/product/distributed-file-systems-and-object-storage-solutions-market/. Accessed on: 25 Feb. 2025

22. Toolbox Talk: Storing data using HDF5 files. Accessed on: 25 Feb. 2025 [Online]. Available: https://dezeraecox.com/toolbox-talk-hdf5-files/. Accessed on: 25 Feb. 2025
23. Write performance of raw data - HDF5 Library - HDF Forum. Accessed on: 25 Feb. 2025 [Online]. Available: https://forum.hdfgroup.org/t/write-perfomance-of-raw-data/10963. Accessed on: 25 Feb. 2025
24. What is object storage or object-based storage? - NetApp. Accessed on: 25 Feb. 2025 [Online]. Available: https://netapp.com/data-storage/storagegrid/what-is-object-storage. Accessed on: 25 Feb. 2025
25. Distributed File Systems vs Distributed Object Storage - GeeksforGeeks. Accessed on: 25 Feb. 2025 [Online]. Available: https://geeksforgeeks.org/distributed-filesystems-vs-distributed-object-storage. Accessed on: 25 Feb. 2025
26. VAST Data: Data Platform Built for Deep Learning and AI. Accessed on: 25 Feb. 2025 [Online]. Available: https://vastdata.com. Accessed on: 25 Feb. 2025
27. Informatica: AI Powered Cloud Data Management. Accessed on: 25 Feb. 2025 [Online]. Available: https://informatica.com. Accessed on: 25 Feb. 2025
28. Top 10 Data Platforms that Use Generative AI - ScikIQ. Accessed on: 25 Feb. 2025 [Online]. Available: https://scikiq.com/blog/top-10-data-platforms-that-use-generative-ai. Accessed on: 25 Feb. 2025
29. Top 6 Dataset Version Control Tools for your Machine Learning workflow - Twine Blog. Accessed on: 25 Feb. 2025 [Online]. Available: https://twine.net/blog/dataset-version-control-tools. Accessed on: 25 Feb. 2025
30. AI Training Data: Benefits, Challenges, Example [2025] Shaip.[Online].Available:https://www.shaip.com/blog/the-only-guide-on-ai-training-data-you-will-need-in/. Accessed on: 25 Feb. 2025
31. What Is an AI Data Catalog? | Secoda. [Online]. Available: https://www.secoda.co/blog/ai-data-catalog. Accessed on: 25 Feb. 2025
32. What is a Machine Learning Data Catalog? - 2024 Guide - Atlan. Дата звернення: 15.03.2025. https://atlan.com/machine-learning-data-catalog/. Accessed on: 25 Feb. 2025
33. Best Machine Learning Data Catalog Software: User Reviews from March 2025 - G2. [Online]. Available:https://www.g2.com/categories/machine-learning-data-catalog.Accessed on: 25 Feb. 2025
34. Top 26 Data Catalog Tools to Consider in 2025 - lakeFS.[Online].Available:https://lakefs.io/blog/top-data-catalog-tools/. Accessed on: 25 Feb. 2025
35. Machine Learning Data Catalog Benefits & Use CasesInformatica. [Online]. Available: https://www.informatica.com/blogs/machine-learning-data-catalog.html.html.html.Accessed on: 25 Feb. 2025

Organization and storage of large datasets for ai training

Oleksii Melnikov

The article examines current approaches to organizing and storing large datasets used for training artificial intelligence (AI) models, particularly for radio signal recognition tasks. It emphasizes the need for reliable and efficient data storage solutions due to the increasing scale and complexity of data involved in AI applications. The study analyzes various data storage formats (HDF5, WAV, IQ raw data), provides an overview of cloud-based solutions (AWS, Google Cloud, Azure), local storage systems (SAN, NAS, JBOD), and distributed file systems. It discusses best practices in data versioning and cataloging for enhanced accessibility, performance, and reproducibility of AI training processes. Recommendations are provided for selecting optimal storage methods based on the specific requirements of AI projects and the characteristics of the processed data.

Keywords: Artificial Intelligence (AI), Machine Learning, Deep Learning, Datasets, Data Storage, HDF5, WAV, IQ raw data, Cloud Storage, Distributed File Systems, Local Storage, Data Management Platforms, Data Versioning, Scalability, Performance, Data Diversity, Data Consolidation, Data Quality, Data Visualization.

Digitalisation in the age of fiscal burdens – benefits for SMEs

Agnieszka Zwolenik[1], Karina Janisz[2]

Academy of Applied Sciences in Nowy Sacz, Staszica Street 1, 33-300 Nowy Sacz, Poland
[1]azwolenik@ans-ns.edu.pl, 0000-0003-2299-5750
[2]kjanisz@ans-ns.edu.pl, 0000-0003-0606-6177

Received 03.02.2025, accepted 29.03.2025
https://doi.org/10.32347/st.2025.3.1204

Abstract. Between 2015 and 2024, Poland introduced solutions supported by new technology, artificial intelligence and the ability to collect huge amounts of data and process them quickly, aimed primarily at reducing the grey market and the activities of criminal groups in domestic and intra-Community transactions
In Poland, an intensive transformation of Polish taxes has been going on for several years. VAT reporting via JPK files, the so-called the white list of VAT payers or the mandatory split payment mechanism required quick changes for small and large entrepreneurs. The aim of this study is to present the key technological tools implemented in Poland that are the result of digitization in tax administration
The aim of the article is to present the results of research on tax digitization and to find out the opinions of small businesses, including those providing financial and accounting services, on the benefits and problems that are caused in their units by the processes associated with the implementation of digital technologies for tax billing and reporting.
In order to achieve the indicated goal, thefollowing research questions were posed:
Is digitisation in public administration having an impact on SMEs?
How do entrepreneurs assess the digitalisation process?
The paper presents the results of empirical research indicating that the digitization of taxes has a number of advantages in the opinion of small businesses
Keywords: IT tools, SAF-T, split payment, e-cash registers, small businesses

Agnieszka Zwolenik
assistant professor, dr.

Karina Janisz
assistant professor, dr.eng

INTRODUCTION

Digitisation in the literature is defined and used in many different meanings. The PWN Dictionary of Polish Language defines digitisation as the dissemination and popularisation of digital technology, as well as the large-scale introduction of electronic infrastructure. K. Śledziewska defines digitisation as a combination of technology and human activities. According to P. Turek, digitisation means the use of digital technologies to access digital information from multiple systems and data points.

The modern economy is often referred to as the digital economy. With the development of technology, entrepreneurs are paying attention not only to efficiency and effectiveness, but increasingly to flexibility, which allows them to adapt quickly to changes and market expectations. Digitalisation has become a 'strategy' without which a business cannot exist. The development of the Internet, the widespread use of modern ICT by citizens and entrepreneurs, challenges public administration to implement new technologies and IT tools (Haque, 2001).

The aim of the process of computerisation of public administration is to ensure the possibility of electronic service of citizens and other interested entities by public authorities (Kotulska, 2012). Currently, it is difficult to imagine public administration performing its

tasks in a traditional way, as well as entrepreneurs making settlements with public institutions in a traditional way. The aim of this paper is to illustrate the changes resulting from the ‘cooperation’ of entrepreneurs with public authorities and the impact of the digitalisation process on SMEs.

An efficient tax system affects both the taxpayer and the state budget. The digitalisation of the tax system is a necessary step towards making this system watertight. For accountants as well as businesses, digitisation contributes to a paradigm shift in bookkeeping. The traditional bookkeeping associated with paper documents is being replaced by IT systems that enable faster and less time-consuming management of company finances. The aim of digitisation to protect the tax system and to ensure the conditions for fair competition for businesses operating within the letter of the law.

Digitisation has become a key process in today's world, affecting both the business and administrative spheres. An employee of a modern administration should be prepared not only to solve tasks related to optimal use of opportunities created by information technologies and to take into account potential threats, but also to widely perceive the field of applications and consequences of the use of ICT in administration.

An intensive transformation of Polish taxation has been underway in Poland for several years. Unclear and complicated tax regulations and at the same time ‘low tax morality’ have caused both the VAT gap and tax fraud to become a serious problem. In order to cope with this problem, instruments aimed at reducing the scale of tax fraud have been introduced in Poland since 2016.

Between 2015 and 2024, Poland introduced solutions supported by new technology, artificial intelligence and the ability to collect huge amounts of data and process them quickly, aimed primarily at reducing the grey market and the activities of criminal groups in domestic and intra-Community transactions.

Among these solutions, the Single Control File, the split payment mechanism, the white list and online cash registers should be highlighted significantly. Structured e-invoices are currently being introduced.

In order to find out the opinion of small and medium-sized enterprises on the digitisation taking place in the area of taxation, a survey was conducted in the Sącz Subregion.

Digitisation and digitalisation-related changes are forcing rapid changes in the way finance and accounting departments operate. JPK or the use of e-invoices, which are automatically generated and then verified by the tax administration, contribute to reducing routine tax audits. The obligation to transmit tax data digitally to the tax authorities provides the treasury with up-to-date information on taxpayers and their transactions, which enables continuous and automated verification of tax settlements.

Digitisation solutions are regulated by law and pose new challenges for entrepreneurs.

The results of the survey showed that entrepreneurs generally have a positive approach to digitisation in the tax sphere and recognise the need to digitise the tax system. Entrepreneurs point to both the problems associated with the digitisation process and the benefits of this process. It is worth emphasising that the conducted survey has a pilot character, however, it has contributed to cognition and better understanding of the perception of tax digitisation processes by entrepreneurs.

In the author's opinion, the next survey should be conducted among a larger group of entrepreneurs, taking into account the period of economic activity and analysing whether accounting is done by the taxpayer, by a person employed in the accounting department or directed to accounting offices.

PROBLEM FORMULATION

An efficient tax system affects both the taxpayer and the state budget. Digitisation of the tax system is a necessary step towards sealing this system. For both accountants and businesses, digitisation contributes to a paradigm shift in bookkeeping. The traditional bookkeeping associated with paper documents is being

replaced by IT systems that enable faster and less time-consuming management of business finances. The aim of digitisation is to protect the tax system and ensure fair competition conditions for businesses operating within the letter of the law.

Digitisation of the tax system covers the area of tax assessment and settlement .
The table below shows the most important changes introduced since 2016.

Table 1.
Key changes introduced as a result of the digitisation of the tax system. Source: own elaboration

Year	Changes made
2016	JPK_VAT files for large companies
2017	JPK_VAT files for SMEs Extension of the scope of transfer pricing reporting
2018	JPK_VAT files for micro companies on-demand JPK files for micro enterprises and SMEs e-Financial reporting split-payment (part I)
2019	e-PIT MDR scheme fetching split-payment (part II) white list of VAT taxable persons
2020	JPK_V7M files split-payment and cash registers new VAT rate matrix digital tax changes to the MDR rules
2021	e-Invoices e-Invoices e-commerce package Polish regulations Estonian tax

The aim of this paper is to present the results of a study on the digitisation of public administration and to find out what companies think about this proces.
In order to find out the opinion of small and medium-sized enterprises on the digitisation taking place in the area of taxation, a survey was conducted in the Sącz Subregion.

The Sadecki Subregion includes the city of Nowy Sącz and the districts of Gorlice, Limanowa and Nowy Sącz. More than 53,000 business entities were registered in the Subregion as of 31.12.2022. A survey was conducted in June 2024, the survey form was addressed to SMEs in the Subregion. 205 entrepreneurs participated in the survey. The table below shows the characteristics of the respondents.

Table 2.
Characteristics of respondents. Source: own elaboration

No.	Respondents	% share
1	Average annual employment	100,00
	no employment	39,02
	up to 9 employees	49,76
	10 to 49 employees	11,22
2	Field of activity	100,00
	service activities	69,76
	commercial activities	20,98
	production activity	9,27
3	Form of taxation	100,00
	general rules	51,71
	flat tax	26,83
	lump-sum tax on registered income	21,46
4	VAT taxpayer	100,00
	active	53,66
	exempt	46,34

As can be seen from the table above, almost half of the respondents are entrepreneurs employing between 1 and 10 employees, the main form of taxation among the respondents is the general rules, and more than 50% of the respondents are active VAT payers.

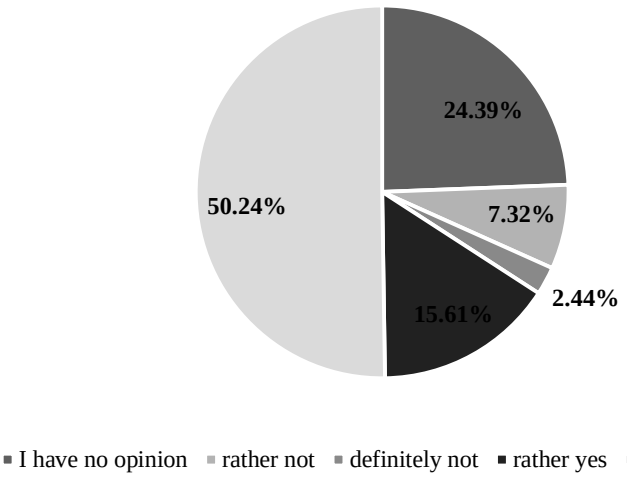


Fig. 1. The need for digitisation in the tax system Source: own elaboration

The results of the survey indicate the need for digitisation of the tax system. When asked about the need for digitisation, entrepreneurs in the Sadecki Sub-region answered ‘definitely

yes' and 'rather yes' in the vast majority of cases. Only less than 10% of respondents believe that digitisation is not needed.

Digitisation solutions are regulated by law and pose new challenges for entrepreneurs. Respondents also evaluated the implementation process of digitisation.

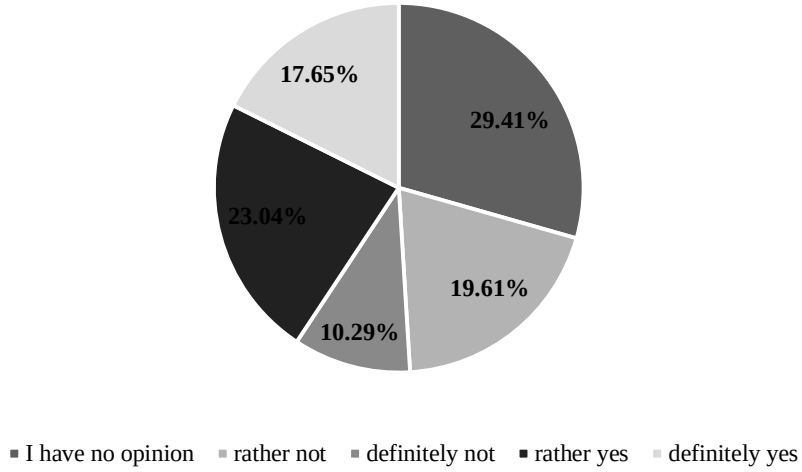


Fig. 2. Is the digitisation of taxes going smoothly in Poland? Source: own elaboration

According to the survey, more than 40% of respondents have a positive opinion of the digitisation process. It can be noted that almost 30% of the surveyed entrepreneurs 'do not have an opinion' on this process. More than 10% think that the process is not fair.

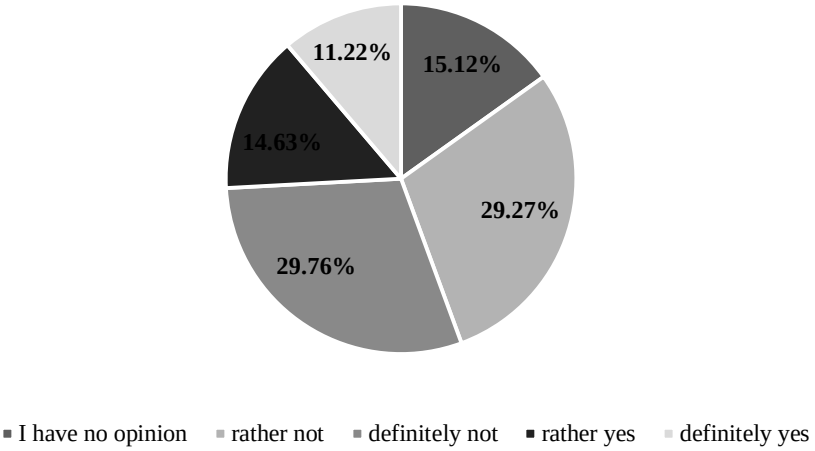


Fig. 3. Have digitalisation measures in the tax system contributed to increased costs in your company? Source: own elaboration

Digitisation requires expenditures related to the introduction of new digital solutions, implementations, training or software updates. Have these activities contributed to cost recovery in companies? The vast majority of respondents believe that the digitisation of the tax system has not increased costs for businesses.

More than 11% of respondents believe that the result of the company's adaptation to the requirements of the digitisation process has

contributed to an increase in costs within the company.

Digitisation and digitisation-related changes are forcing rapid changes in the way finance and accounting departments operate. JPK or the use of e-invoices, which are automatically generated and then verified by the tax administration, contribute to reducing routine tax audits. The introduced obligation to digitally transmit tax data to the tax authorities provides the treasury with up-to-date information on taxpayers and their transactions, which enables a constant and automated verification of tax settlements.

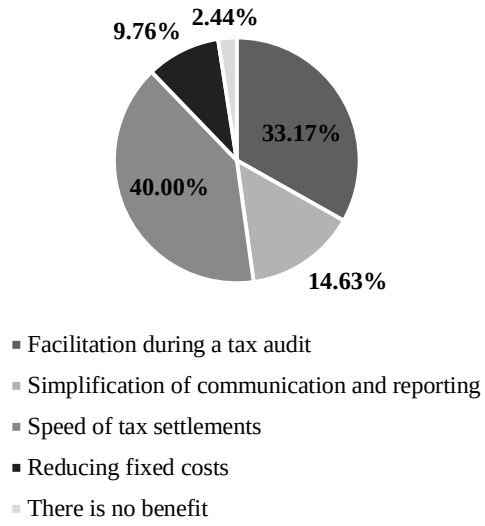


Fig. 4. Benefits of digitalisation in the tax sphere Source: own elaboration

The results of the survey showed that, in the surveyed group of respondents, only about 2.5 per cent of respondents believe that the digitisation of taxes does not bring any benefits.

Among the benefits of the digitisation of public administration are: improving the efficiency and speeding up the process of verifying tax returns or identifying any errors or irregularities occurring in these documents, transparency and the possibility of tax control manifested in transparency of the tax system, reduction of the risk of committing committing errors and tax fraud, simplification of communication between the taxpayer and the tax authority thanks to the rapid exchange of information, the possibility of online communication significantly reducing case processing time, minimising fixed costs for both the taxpayer and the public administration.

Taxpayers cited speed of tax settlements and facilitation during tax audits as the main benefits of digitalisation. Among other benefits than those proposed in the survey question, respondents indicated a reduction in the cost of office supplies, no direct contact with an official or a reduced number of tax audits.

CONCLUSIONS

The period 2015-2024 in Poland saw the introduction of solutions aided by new technology, artificial intelligence and the ability to collect huge amounts of data and process them quickly, aimed primarily at reducing the grey market and the activities of criminal groups in domestic and intra-Community transactions.

Among these solutions, the Single Control File, the split payment mechanism, the white list and online cash registers should be highlighted significantly. Currently, structured e-invoices are in the process of being introduced.

The results of the survey showed that entrepreneurs generally have a positive attitude towards digitisation in the tax sphere and recognise the need to digitise the tax system. Entrepreneurs indicate both the problems associated with the digitisation process and the benefits of the process. It is worth emphasising that the survey is a pilot study, however, it has

contributed to the knowledge and better understanding of the perception of tax digitisation processes by entrepreneurs.

The authors are of the opinion that another survey should be conducted among a larger group of entrepreneurs, taking into account the period of economic activity and analysing whether accounting is done by the taxpayer, by a person employed in the accounting department or directed to accounting offices.

REFERENCES

1. Dictionary of the Polish Language, PWN, https://sjp.pwn.pl [retrieved: March 2024]
2. **Haque A.,** (2001). GIS, public service and the issue of democratic governance, „Public Administration Review", vol. 61, s. 259–265
3. **Kotulska M.** (2012). Wykorzystanie środków komunikacji elektronicznej w postępowaniu administracyjnym, PPP 2012, nr 9, s. 23
4. **Śledziewska K, Włoch R.** (2020). Gospodarka cyfrowa jak nowe technologie zmieniają świat. Wydawnictwo Uniwersytetu Warszawskiego, Warszawa, p.78
5. **Turek P.** (2024). Digitalizacja, cyfryzacja, optymalizacja cyfrowa i cyfrowa transformacja – takie oczywiste, a wytłumaczyć potrafisz? https://pl.linkedin.com/pulse/digitalizacja-cyfryzacja-optymalizacja-cyfrowa-i-takie-paulina-turek
6. Цифровізація в епоху фіскальних навантажень – переваги для малих і середніх підприємств (МСП)

Цифровізація в епоху фіскальних навантажень – переваги для МСП

Анжеліка Зволенік, Каріна Янісж

Анотація. У період з 2015 по 2024 рік Польща впровадила рішення, підтримані новими технологіями, штучним інтелектом та здатністю швидко збирати та обробляти великі обсяги даних, спрямовані насамперед на зменшення тіньового ринку та діяльності злочинних груп у внутрішніх і внутрішньоєвропейських транзакціях.

У Польщі протягом кількох років триває інтенсивна трансформація податкової системи. Звітування з ПДВ через файли JPK, так званий «білий список» платників ПДВ або обов'язковий механізм розподілу платежів вимагали швидких змін як від малих, так і від великих підприємців. Метою цього дослідження є представлення ключових технологічних інструментів, впроваджених у Польщі, які стали результатом цифровізації податкової адміністрації.

Мета статті — представити результати досліджень цифровізації податків і з'ясувати думки малих підприємств, включаючи ті, що надають фінансові та бухгалтерські послуги, про переваги та проблеми, які виникають у їхніх організаціях унаслідок впровадження цифрових технологій для податкового обліку та звітності. Для досягнення зазначеної мети були поставлені наступні дослідницькі питання:

Чи впливає цифровізація в державному управлінні на малі та середні підприємства?

Як підприємці оцінюють процес цифровізації?

У статті представлені результати емпіричних досліджень, які свідчать про те, що цифровізація податків має низку переваг на думку малих підприємств.

Ключові слова: IT-інструменти, SAF T, розподіл платежів, електронні касові апарати, малі підприємства.

Artificial intelligence in 5G: optimization of network management and cybersecurity systems

Viktoriia Trofymchuk[1], Turovskyi Oleksandr[2]

[1]Independent Researcher, USA,
[2] State University of Information and Communication Technologies
[1]trofymchukviktoria@gmail.com, https://orcid.org/0000-0002-9756-0244
[2]3s19641011@ukr.net, https://orcid.org/0000-0002-4961-0876

Received 03.02.2025, accepted 29.03.2025
https://doi.org/10.32347/st.2025.3.1205

Abstract. The deployment of fifth-generation (5G) technologies is significantly changing the way telecommunications networks are designed and managed. The growth in the amount of data and the number of connected devices is also increasing the requirements for adaptive resource management, low latency, and guaranteed security, and traditional management mechanisms are often not flexible enough. In view of these developments, the use of artificial intelligence (AI) methods that can provide a dynamic response to changing conditions in real time is becoming increasingly important. [1]. The analysis of modern algorithms, in particular, such as: Graph Neural Networks (GNN), Deep Reinforcement Learning (DRL), as well as recurrent architectures LSTM and GRU, which are used to predict loads and optimize spectrum allocation. This approach allows us not only to reduce the risk of overload but also to increase the efficiency of network management in general [3].

The article considers the use of generative adversarial networks (GANs) in the tasks of detecting DDoS attacks. Our experimental results confirm that GAN models demonstrate high accuracy rates of up to 96%, outperforming traditional algorithms such as decision trees or SVMs, which is consistent with the results of Maleh et al. (2021)

In addition, the article analyzes the possibilities of reducing energy consumption in 5G networks by implementing Multi-Agent Q-learning. This approach provides distributed decision-making at the base station level, taking into account local changes in the network, which helps to save energy without compromising the quality of service (Luo, 2020; Sutton & Barto, 2018).

The research results show the significant potential of AI-oriented approaches in building scalable, resilient, and cyber threat-resistant next-generation networks. The recorded reduction in latency by 42%, improvement in spectral efficiency by 21%, and energy savings of up to 28% confirm the practical significance of the proposed solutions.

Viktoriia Trofumchuk
Independent Reseacher

Oleksandr Turovskyi
Head of Department, Doctor of Technical Sciences, Professor

The study suggests that it is possible to form a new architectural concept where AI not only supports the functioning of the network but also becomes a mechanism for its self-learning and evolution. [2].

Keywords: Artificial Intelligence, 5G networks, network management, cybersecurity, traffic optimization, wireless communications.

INTRODUCTION

The growth of data transmission, the increasing number of IoT devices and the transition to the concept of smart cities place increased demands on the performance and stability of telecommunications networks [3]. Traditional network management methods are unable to provide sufficient flexibility and adaptability in real-world 5G scenarios. At the same time, the introduction of artificial intelligence allows for the creation of autonomous control systems that can analyze the network status in real time, predict possible

overloads and identify potential security threats [2].

The use of deep learning approaches to optimize spectrum allocation, such as deep convolutional neural networks (CNNs) and Reinforcement Learning (RL) methods, helps to increase the efficiency of frequency resource use, which is a critical aspect for ensuring the sustainable operation of 5G networks [1]. Machine learning algorithms allow for adaptive adjustment of network parameters, load balancing, and high-quality communication even in the face of dynamic changes in network topology.

The paper focuses on the analysis of modern 5G network optimization algorithms, in particular, load prediction methods using long-term short-term memory (LSTM) and the use of generative adversarial networks (GANs) for automatic threat detection [6]. The results of the study demonstrate the significant potential of AI in improving network management mechanisms and enhancing network security, which makes the proposed methods promising for future implementations in real 5G networks.

PROBLEM FORMULATION

The rapid development of 5G networks is accompanied by an increase in the number of connected devices, a more complex network topology, and increased requirements for its stability, performance, and security. Traditional approaches to network management lack the flexibility and adaptability necessary for effective operation under conditions of high load dynamics. In this regard, there is a need to introduce intelligent systems capable of self-learning, forecasting, and autonomous decision-making in real time.

The purpose of this research is to analyze and provide a practical justification for the effectiveness of using artificial intelligence methods to optimize the management of 5G network parameters and ensure their cyber resilience. The focus is on exploring the potential of deep reinforcement learning, recurrent neural networks, and generative models to solve key network administration tasks.

To realize this goal, the following research objectives have been formulated:

1. Develop an AI model for optimizing spectrum allocation using convolutional neural networks (CNN) and the DDPG algorithm.

2.To propose an architecture for predicting peak loads and delays based on recurrent networks LSTM and GRU.

3.To build a model for detecting anomalous traffic and security threats using generative adversarial networks (GANs) and auto-encoders.

4.Develop a multi-agent system for energy-efficient network management based on Multi-Agent Q-learning.

5.Conduct simulation testing of the developed models and evaluate their effectiveness in terms of key performance, cyber resilience, and energy consumption metrics.

RESEARCH METHODOLOGY

The methodological basis of the study is based on a combination of deep learning tools, data mining, and modeling of dynamic processes in wireless networks. The study implemented a step-by-step approach that includes the development, training, testing, and comparative evaluation of intelligent models in a simulated environment.

1. Creating a simulation environment.

We modeled a conditional urban 5G environment with dynamic load scenarios, including areas of dense traffic, variability in user behavior, and the presence of anomalous patterns. Input data was generated based on typical user activity profiles in fifth-generation networks.

2. Architecture of AI modules.

Each research area is implemented as a separate intelligent module:

Spectrum optimization - CNN+DDPG model for dynamic management of frequency resources;

Load forecasting - LSTM/GRU networks for identifying traffic peaks and managing delays;

Cyber threat detection - GAN in combination with auto-encoder to recognize DDoS attacks;

Energy efficiency - Multi-Agent Q-learning for independent optimization of transmitter power by base stations.

3. Model training.

The algorithms were trained on simulated datasets using appropriate loss functions and optimizers. The training parameters were adapted to the specifics of each task: in the case of GAN, minimizing false positives, in CNN+DDPG, maximizing spectral efficiency, etc.

4. Experimental testing.

All models were tested on independent datasets under varying load and attack conditions. Accuracy, average latency, power consumption, and spectrum utilization were evaluated for both traditional models and the proposed AI-oriented architectures.

5. Comparative analysis.

The results were compared with classical methods of threat management and detection (SVM, Decision Tree, rule-based approaches). The integral assessment included performance, threat resistance, and self-optimization.

RESEARCH RESULTS

To test the efficiency of integrating artificial intelligence into 5G network management, a series of experimental simulations were conducted using modern deep learning algorithms. The study covered four key areas: optimizing spectrum allocation, reducing real-time delays, detecting cybersecurity threats, and energy-efficient infrastructure management.

For each area, a separate model was developed and tested using the corresponding AI architectures, including CNN, DDPG, LSTM, GRU, GAN, and MAQL. The experiments were conducted on simulated data that reproduced the conditions of an urban environment with a high network load. All models were evaluated by key performance metrics: accuracy, latency, resource efficiency, and energy consumption.

Below is a summary of the results, as well as graphical visualizations that clearly demonstrate the impact of the applied algorithms on the main network parameters.

AI-Based Spectrum Allocation Optimization

Using deep learning algorithms to dynamically reallocate frequency resources allows for adaptive coverage and minimizes interference between base stations [1]. In the course of experimental modeling, a neural network architecture consisting of deep convolutional neural networks (CNNs) combined with the DDPG (Deep Deterministic Policy Gradient) algorithm was developed, which allows for optimal real-time management of the frequency resource. The model is trained on network load data and is able to predict fluctuations in traffic intensity by adjusting the allocation of frequency bands between base stations.

The formal model of frequency spectrum allocation is defined by the equation:

$$f_{opt} = \arg\max_{f \in F} \sum_{i=1} (U_i(f) - I_i(f) = \lambda \sum_{j \neq i} \frac{P_j}{d_{ij}^2} \quad (1)$$

where is the optimal frequency, represents the channel utility for user i, denotes the interference level, is the transmission power of base station j, and is the distance between the user and the base station.

As a result of testing under real network load conditions, it was found that the proposed method can reduce network congestion by 35%, which significantly improves connection quality and resource efficiency [5].

Reducing delays in real time

This is a method for peak load prediction and dynamic packet queue management in 5G networks based on Deep Recurrent Learning (LSTM) combined with Gated Recurrent Unit (GRU) mechanisms [6]. The proposed model allows for efficient analysis of historical network traffic data, predicting the moments of maximum load and optimizing packet routing to minimize delays.

Formally, the latency prediction model can be described by the following equation:

$$T_{pred}(t+1) = \sigma(W_1 T_{hist} + W_2 G_{hist} + b) \quad (2)$$

where represents the input data from the Gated Recurrent Unit, which helps reduce parameter dimensionality and improve prediction accuracy.

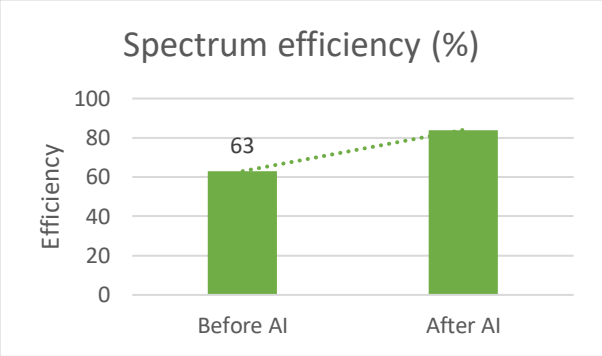


Fig. 1. Spectrum utilization efficiency before and after applying the AI model

Experiments have demonstrated that the application of this model reduces the average packet delay by 42%, which is critical for connected IoT devices and mobile users [7]. In addition, it was found to improve the adaptability of packet routing and the overall level of traffic service under peak loads.

The use of deep recurrent neural networks for adaptive data queue management in 5G networks allows operators to improve service quality, optimize resource utilization, and minimize delays in a highly dynamic network environment. This opens up new opportunities to improve the scalability and performance of modern telecommunications infrastructures.

AI methods of detecting attacks

To improve the efficiency of DDoS attack detection, a model based on auto-encoders combined with generative adversarial networks (GANs) was developed to generate realistic patterns of anomalous traffic and recognize potential threats with high accuracy [2]. The use of GANs in the traffic analysis process ensures that the model is trained to distinguish between legitimate and malicious packets in the face of high variability in attack strategies.

The loss function in a generative adversarial network for threat detection is defined by the following equation:

$$L = E_{X\sim}P_{date(X)}[\log D(X) + E_{Z\sim p_2(X)}[\log(1-D(Z)))] \quad (3)$$

where D(X) is the discriminator, and G(Z) is the generator that creates fake attacks to train the system.

During the experimental modeling and testing of the GAN analytics model on real network traffic datasets, it was demonstrated that the proposed method allows achieving 96% accuracy in detecting DDoS attacks, which exceeds the efficiency of traditional anomaly analysis methods, such as suspicious traffic detection rules or basic machine learning algorithms.

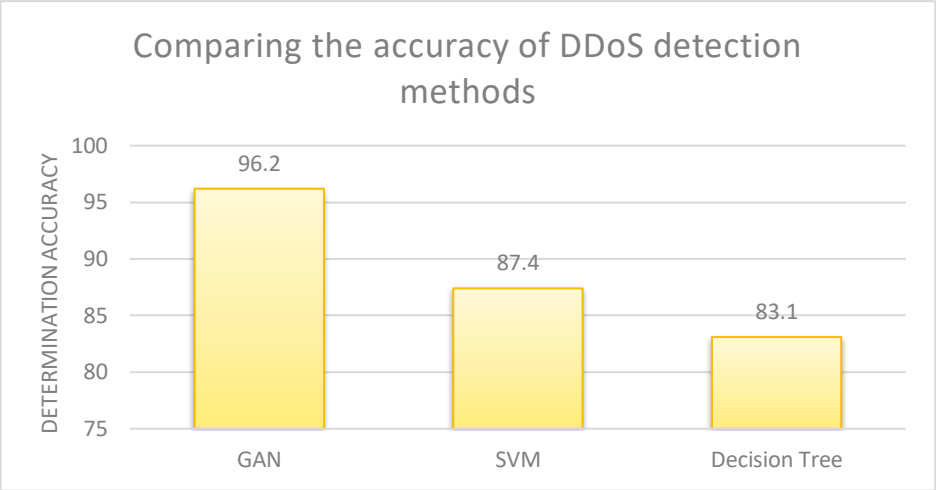


Fig. 2. Comparison of the accuracy of DDoS detection methods: GAN, SVM, Decision Tree

The results of the study prove that the implementation of the GAN model in the cybersecurity system of 5G networks can improve the ability to identify threats in real time, reduce the level of false positives, and ensure automatic adaptation to new attack scenarios. Further research should be aimed at integrating the proposed architecture into integrated cybersecurity solutions for telecom operators and extending it to detect other types of attacks, such as botnet attacks and insider threats.

Optimization of power consumption

To ensure effective power management in 5G networks, a methodology based on Multi-Agent Q-learning (MAQL) has been developed that allows for autonomous adjustment of transmitter power depending on the current load and environmental conditions [4]. The use of multi-agent learning allows each base station to analyze the network status and adaptively change the transmission level without losing the quality of service.

Optimized power management is based on an updated reward function that takes into account not only the balance between communication quality and transmission power, but also dynamic environmental conditions:

$$Q(s,a) = (1-a)Q(s,a) + a(r + \gamma \max_{a`} Q(s`,a`)) - \beta \cdot P_{total} \quad (4)$$

where Q(s,a) - represents the value of state s when action a is taken, r - is the reward for maintaining high service quality, γ - is the discount factor that determines the weight of future rewards, P_total - represents the total energy consumption of the entire network, and β - is the weighting coefficient that controls the impact of energy consumption on the decision-making process.

The modeling results demonstrated that the application of the proposed methodology reduces the overall power consumption by 28% without losing the stability of network coverage. In addition, the use of multi-agent learning has reduced the overload on base stations by improving load balancing between neighboring nodes [7].

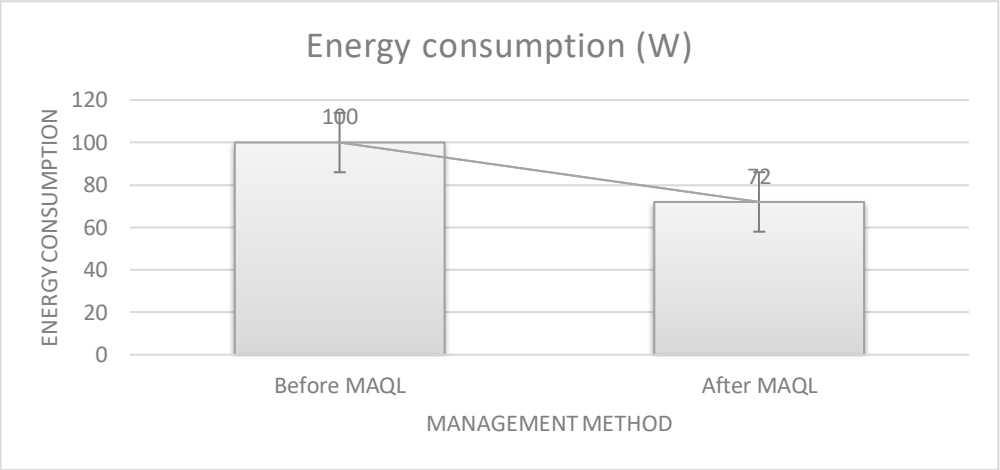


Fig. 3. Impact of Multi-Agent Q-learning on power consumption and 5G network stability

The introduction of Multi-Agent Q-learning into the 5G infrastructure allows mobile operators to automate the process of transmitter power management, reduce operating costs and make the system more environmentally sustainable. Further research can be aimed at integrating quantum algorithms into the power optimization model and expanding the system's self-learning capabilities to predict peak loads in real time.

CONCLUSIONS

The results of this research confirm that the integration of artificial intelligence methods into the operation of 5G networks opens up new horizons in the field of adaptive management of telecommunications infrastructure. Unlike traditional approaches, the proposed AI-oriented models allow for flexible, context-dependent real-time adjustment of critical network parameters. Empirical modeling has shown that:

the use of CNN+DDPG provides 35% lower channel congestion and 21% higher spectral efficiency;

the combination of LSTM and GRU reduces the average packet transmission delay by 42%, which is critical for maintaining IoT infrastructure;

the GAN model with auto-encoder achieves 96% accuracy in detecting DDoS attacks, demonstrating an advantage over classical threat detection methods;

MAQL reduces power consumption by 28% while maintaining high network stability of over 90%.

The complexity of the proposed approaches is manifested in the ability of systems to simultaneously achieve a balance between performance, threat resistance, and energy efficiency. This synergy of AI and 5G defines a new paradigm in building autonomous, self-optimized network structures capable of self-learning and adaptation in the face of high load dynamics and threats.

Of particular importance is the prospect of using multi-agent learning as a basic principle for distributed network management, which ensures scalability and increased fault tolerance. In addition, the inclusion of generative models in the cybersecurity system allows for a shift from a reactive to a proactive approach to threat identification.

In the future, it is advisable to focus efforts on:

integration of quantum computing models into the process of real-time decision optimization;

creating hybrid neural architectures that combine logical and statistical modeling for complex network behavior scenarios;

development of ethical standards for the use of AI in the communication infrastructure, in particular, regarding the autonomy of decisions in crisis situations.

Thus, the synthesis of AI and 5G not only solves existing technical challenges but also lays the foundation for a new generation of digital ecosystems focused on the resilience, security, and intelligent evolution of networks.

REFERENCES

1. **Patel R., Chen Y.** AI-Based Network Optimization for 5G and Beyond // IEEE Communications Magazine. – 2023. – No. 61(4). – P. 35–50.
2. **Thompson J., Lee S.** Cybersecurity Challenges in 5G Networks: The Role of AI in Threat Detection // Journal of Telecommunication Security. – 2022. – No. 15(1). – P. 89–102.
3. National Strategy for the Development of Artificial Intelligence in Ukraine for 2021–2030 [Electronic resource] – Available at: https://wp.oecd.ai/app/uploads/2021/12/Ukraine_National_Strategy_for_Development_of_Artificial_Intelligence_in_Ukraine_2021-2030.pdf
4. European Telecommunications Standards Institute (ETSI). Securing Artificial Intelligence (SAI) [Electronic resource] – Available at: https://www.etsi.org/technologies/securing-artificial-intelligence
5. National Institute of Standards and Technology (NIST). AI Risk Management Framework [Electronic resource] – Available at: https://www.nist.gov/itl/ai-risk-management-framework
6. Ensuring Information Security in 6G Telecommunication Networks [Electronic resource] // Academia.edu. – Available at: https://www.academia.edu/103174175
7. H-X Technologies. Artificial Intelligence Security [Electronic resource] – Available at: https://www.h-x.technology/ua/blog-ua/artificial-intelligence-security-ua
8. Center for Democracy and Rule of Law (CEDEM). AI Regulation: Experience of the USA [Electronic resource] – Available at: https://cedem.org.ua/analytics/shtuchnyi-intelekt-usa
9. BDO Ukraine. Artificial Intelligence and the Internet of Things Changing the Game for Telecommunications [Electronic resource] – Available at:https://www.bdo.ua/uk-ua/insights-2/information-materials/2023/how-ai-and-the-internet-of-things-change-the-game-for-telecoms

Штучний інтелект у 5G: оптимізація управління мережею та системи кіберзахисту

Вікторія Трофимчук, Oleksandr Turovskyi

Анотація. Розгортання технологій п’ятого покоління (5G) суттєво змінює підходи до проєктування та управління телекомунікаційними мережами. Зростання великої кількості даних , а також кількість підключених пристроїв, також відбувається зростання вимог до адаптивного керування ресурсами, підтримки низьких затримок і гарантованої безпеки, традиційні механізми управління часто виявляються недостатньо гнучкими. З огляду на ці події важливості набуває застосування методів штучного інтелекту (ШІ), які здатні забезпечити динамічне реагування мережі на зміну умов у реальному часі.

Проведено аналіз сучасних алгоритмів, зокрема таких як: Graph Neural Networks (GNN), Deep Reinforcement Learning (DRL), а також рекурентних архітектур LSTM і GRU, які використовуються для прогнозування навантажень та оптимізації розподілу спектру. Даний підхід дозволяє нам не лише зменшити ризики перевантаження, але й підвищити ефективність управління мережею загалом.

Розглянуто використання генеративних змагальних мереж (GANs) у задачах виявлення атак типу DDoS. Експериментальні результати, які були нами отримані підтверджують, що GAN-моделі демонструють високі показники точності — до 96%, перевершуючи традиційні алгоритми на кшталт дерева рішень чи SVM, що узгоджується з результатами досліджень Maleh et al. (2021).

Окрім того, у статті проаналізовано можливості зниження енергоспоживання в 5G-мережах шляхом впровадження Multi-Agent Q-learning. Такий підхід забезпечує розподілене прийняття рішень на рівні базових станцій з урахуванням локальних змін у мережі, що сприяє економії енергії без погіршення якості сервісу (Luo, 2020; Sutton & Barto, 2018).

Результати дослідження свідчать про значний потенціал AI-орієнтованих підходів у побудові масштабованих, стійких до збоїв і кіберзагроз мереж нового покоління. Зафіксоване зниження затримок на 42%, покращення спектральної ефективності на 21% і економія енергії до 28% є підтвердженням практичної значущості запропонованих рішень. Провівши дослідження варто звернути увагу, на те, що є можливість формування нової архітектурної концепції, де ШІ не лише підтримує функціонування мережі, а й стає механізмом її самонавчання та еволюції.

Ключові слова: Штучний інтелект, 5G-мережі, управління мережею, кібербезпека, оптимізація трафіку, бездротові комунікації, глибоке навчання.

Методика ідентифікації трафіку в телекомунікаційних мережах 5G/IMT-2020 на основі штучного інтелекту

Олександр Корецький[1], Анастасія Кондакова[2]

[1]Державний університет інформаційно-комунікаційних технологій
вул. Солом'янська, 7, м. Київ, Україна, 03110
[2]Київський національний університет будівництва і архітектури
пр-т Повітряних Сил, 31, м.Київ, Україна, 03037
[1]koretsky@gmail.com, https://orcid.org/0009-0001-4809-7556
[2]a_kondakova@ukr.net, https://orcid.org/0000-0003-1302-2244

Received 03.02.2025, accepted 29.03.2025
https://doi.org/10.32347/st.2025.3.1206

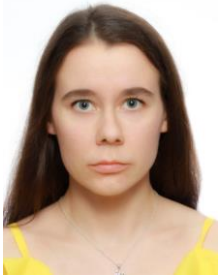
Анотація. В роботі проведено аналіз процесу ідентифікації вхідного трафіку в мережах 5G/IMT-2020, побудованій по технології Ultra-reliable and low latency communications, визначені його особливості та напрямки досліджень по підвищенню ефективності ідентифікації трафіку х урахуванням вимог до функціонування розглянутої мережі.

Для вирішення завдання ідентифікації трафіку мережі 5G/IMT-2020в роботі розроблена та подана відповідна методика. Вказана методика включає формування масивів метаданих вхідного потоку, модифікацію їх в набір навчальних даних, формування навчального програмно-апаратного комплексу та розбудову структури нейронної мережі, проведення процесу навчання нейронної мережі та втілення її в процес ідентифікації трафіку в телекомунікаційних мережах 5G/IMT-2020.

Оцінка результатів процесу навчання запропонованої нейронної мережі та перевірки її роботи на тестових наборах даних у навченому стані показала, що подана в роботі нейронна мережа здатна провести моніторинг та ідентифікувати згенерований від сервісів Інтернету Речей трафік з ймовірністю до 99,7%. В процесі ідентифікації трафіку від двох та більше сервісів дана ймовірність може знизитися, проте перебуває у допустимих межах 80-90%.

Ключові слова: ідентифікація вхідного трафіку інформаційно-комунікаційної мереж, мережа 5G/IMT-2020, штучний інтелект, нейронна мережа, управління навантаження інформаційно-комунікаційної мережі

Корецький Олександр
Аспірант кафедри штучного інтелекту



Анастасія Кондакова
Аспірант кафедри кібербезпеки та комп'ютерної інженерії

ПОСТАНОВКА ПРОБЛЕМИ

Необхідність забезпечення розвитку глобальних економічних відноси, інтенсивний розвитку науки та посилення сфери оборони вимагають постійного удосконалення та розвитку галузі інформаційних технологій в напрямку збільшення швидкості та якості передачі корисних даних. Одним з напрямків досліджень в вказаній сфері є розбудова та швидке впровадження перспективних концепцій мереж передачі даних, а саме стандарту 5G/IMT-2020 та, в його розвиток, наступних поколінь таких мереж.

Розбудова сучасних та перспективних концепцій мереж зв'язку, на даному етапі сконцентрована на реалізації трьох основних технологій, на основі яких безпосередньо будується мережа 5G/IMT-2020. А саме [1,2,3]:

1. Глобальні міжапаратні взаємодії – MMC (Massive Machine type Communications);
2. Високоефективний мобільний широкосмуговий зв'язок – eMBB (Enhanced Mobile Broadband);
3. Високонадійний зв'язок з мінімальними затримками – URLLC (Ultra-reliable and low latency communications).

На даний момент реалізація інформаційно-комунікаційних мереж на основі концепції високонадійного зв'язку з мінімальними затримками (URLLC) є одним з найскладніших завдань, що стоять перед науково-технічною спільнотою. Основна вимога до мереж класу мереж URLLC, це висока надійність передачі даних при мінімальних затримках [3].

Дані послуги реалізуються в таких напрямках як електронна медицина (e-health), автономний транспорт, небезпечні промислові виробництва формату 4.0 та інші.

Одним із найвідоміших типів сервісів мереж URLLC є Інтернет речей (IoT) та Тактильний Інтернет (Tactile Internet). В випадку реалізації мережі для забезпечення функціонування сервісу Tactile Internet слід зазначити, що на архітектуру побудови мережі та її ефективне функціонування найбільший вплив чинять саме ультрамалі затримки. Їх критичне значення впливає на децентралізацію мережі внаслідок обмежень на відстані, на яких можна надавати послуги Tactile Internet. Досягнення ультрамалих затримок в мережі Tactile Internet та IoT, в свою чергу, має призвести до децентралізації економіки та стирання цифрової нерівності між територіями як однієї держави так і в межах певного суспільно-економічного регіону [1,2].

Сучасний обсяг трафіку, його гетерогенність та різноманітність сервісу Tactile Internet та інших сервісів, у тому числі тих, що належать до URLLC, диктує нові вимоги до оперативності прийнятих рішень щодо забезпечення якості обслуговування (QoS) [2,3].

Аналіз можливостей існуючих інструментів управління трафіком 5G/IMT-2020 та забезпечення високого значення QoS показує, що вони неспроможні надати необхідний рівень ефективності безпосередньої та швидкої ідентифікацій трафіку URLLC, його оцінки та управління [4,5,6]. Основною проблемою, яка обумовлює такі висновки, є невідповідність можливостей по прогнозу очікуваного навантаження на мережу. В даному випадку, вже в більшості рішень концепції URLLC потрібно мати прогнози навантажень на ті чи інші сервіси, враховуючи географічні та динамічні (пересування абонента), в тому числі, при переміщені його в просторі на високих швидкостях [4,5,6].

Варто також зауважити, що в процесі розробки рішень у мережах 5G/IMT-2020, розробки їх стандартів та проведення відповідних досліджень було досягнуто висновку, які свідчать про те, що вимоги URLLC складно виконати з точки зору широкомасштабного впровадження даного типу послуг [4,6]. Вони можуть бути надані в обмеженому географічному просторі за виділеними мережевими каналами і за інших обмеженнях.

Неоднорідність трафіку, складність ініціалізації потоків та розрахунку зростання трафіку, а також питання оперативного виявлення та вчасного розрахунку змін профілю трафіку призводять до того, що існуючі «ручні» методи моніторинг в мережах покоління 5G/IMT-2020 втрачають свою актуальність.

Дані обставини формують нову наукову проблему, а саме забезпечення функціонування мережі покоління 5G/IMT-2020 на надвисоких швидкостях передачі даних з ультрамалими затримками.

Для забезпечення оперативного реагування систем управління мережею на зростання трафіку та зміни гетерогенності в мережах 5G/IMT-2020 (в тому числі періодичного трафіку), аналізу змін профілю трафіку потрібно розробити та втілити відповідну низку механізмів та процедур, пристосованих для ідентифікації, розрахунку навантаження, його прогнозу та управління трафіком в мережах даного покоління без впливу на їх продуктивність.

Одним з таких механізмів, призначених для вирішення часткового завдання своєчасної ідентифікації вхідного трафіку є

застосування методика ідентифікації трафіку в телекомунікаційних мережах 5G/IMT-2020. А її розробка формує нове актуальне наукове задання, вирішенню якого присвячена дана робота.

Основні вимоги до вирішення такого завдання, це вчасне та ефективне виявлення змін структури трафіку, розрахунок його навантаження та оперативне прийняття рішень по управлінню трафіком в нових умовах. При цьому, необхідно зазначити, що для того, щоб забезпечити постійний моніторинг та оперативне реагування мережі зв'язку на зміни трафіку відповідні пристрої та механізми комутації та маршрутизації трафіку і його аналізу в ході моніторингу та реагування повинні мати необхідний рівень абстрагування від «фізики» процесів передачі даних через мережу.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ

Питанням ідентифікації та моніторингу трафіку в інформаційно-комунікаційних мережах останніх поколінь присвячено рад наукових публікацій. Основні підходи та принципі рішення задання ідентифікації та моніторингу подано в наукових роботах [7-10]

На даний момент якість обслуговування в мережах зв'язку забезпечується за допомогою технологій DiffServ (Differential Services), а також інших TE (Traffic Engineering) рішень, наприклад MPLS-TE, що використовує в основі протокол резервування ресурсів RSVP-TE [7,8]. Однак ці рішення мають ряд недоліків для їх застосування в ряді сервісів у мережах п'ятого та наступних поколінь. Основними недоліками є відсутність динамічного управління залежно від мінливості профілю трафіку в мережі та відсутність можливості швидкого переконфігурування політик обслуговування підконтрольного домену мережі, а також обмежений набір класифікаторів трафіку, що в умовах, що з'являються та відрізняються за вимогами до якості обслуговування сервісів IoT та інших, що є дуже критичним недоліком.

Необхідний рівень абстрагування від фізичних процесів реалізують концепції SDN (software-defined networking, SDN; програмно-конфігурована мережа) і NFV (Network Functions Virtualization, NFV – концепція віртуалізації мережевої архітектури), що викладені в роботах [9,10]. Відмічене, що для оперативного реагування систем управління мережі, у разі застосування концепцій SDN та NFV контролера SDN і Оркестратора, потрібно проводити високоточну ідентифікацію трафіку без безпосереднього втручання у потік на рівні передачі даних та відповідно, без внесення змінень до його профілю, а також мінімізації затримок трафіку [9,10]. Механізми такої ідентифікації та процедури моніторингу без втручання в потік даних в поданих роботах не розглянуті а викладені процедури концепцій SDN і NFV загалом не дозволяють здійснити такий моніторинг при одночасному вирішення покладного на SDN забезпечення кібербезпеки мережі.

Таким чином, завдання по ідентифікації трафіку в мережах 5G/IMT-2020 потребує розробки окремої методики. Основні вимоги до такої методики, це вчасне та ефективне виявлення змін трафіку, його розрахунок та оперативне прийняття рішень по управлінню трафіком при забезпечені необхідного рівня абстрагування від «фізики» процесів передачі даних в телекомунікаційних мережах 5G/IMT-2020.

Метою дослідження є розробка методики ідентифікації трафіку в телекомунікаційних мережах 5G/IMT-2020, які функціонують в межах концепції передачі даних URLLC.

Для досягнення визначеної мети необхідно:

- визначити механізми ідентифікації та моніторингу трафіка без втручання процес передачі даних та сформувати принципи його функціонування;
- побудувати програмно - апаратний комплекс для оцінки застосовності запропонованого механізму ідентифікації та моніторингу трафіку;
- розробити алгоритм ідентифікації трафіку в телекомунікаційних мережах 5G/IMT-2020, яка повинна функціонувати в межах концепції передачі даних URLLC;
- оцінити можливості застосування поданого алгоритму та створеної на його основі

методики для моніторингу навантаження в телекомунікаційних мережах 5G/IMT-2020, які повинні функціонувати в межах концепції передачі даних URLLC

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ ДОСЛІДЖЕННЯ

Як було зазначено раніше, завдання ідентифікації та моніторингу трафіку включає необхідність розпізнавання великої кількості типів (враховуючи URLLC) сервісів, не вносячи додаткових затримок у трафік, та можливість розширення та налаштування алгоритму під певне географічне розташування мережі та сервісів [4,6]. Неоднорідність трафіку, складність його ініціалізації та розрахунку його зростання, а також складність оперативного розрахунку змінення профілю трафіку призводять до того, що існуючі «ручні» методи розрахунку втрачають свою актуальність.

Одними з перспективних рішень, які дадуть змогу значно розширити можливості мереж

п’ятого покоління по забезпеченню вимог URLLC розглядаються наступні рішення [5,7,8]:

1. Зміна технологій фізичного рівня та перехід на так звані "квантові комунікації", які реалізують інший фізичний принцип передачі інформації, заснований на принципі квантової заплутаності. Однак цей метод знаходиться на стадії ранніх досліджень.

2. Децентралізація мереж зв'язку та децентралізація систем обчислень, гнучке їх управління з імплементацією Штучного інтелекту (ШІ) як системи моніторингу та управління.

Виходячи з поточного рівня мережевих технологій 5G/IMT-2020, які повинні функціонувати в межах концепції передачі даних URLLC та часткових завдань моніторингу та управління, в рамках процесу впровадження Штучного інтелекту в системи управління вказаними мережами найбільш актуальними завданнями для нейромерж будуть наступні [6,9,10]:

– однозначна ідентифікація трафіку з подальшим прогнозуванням його характеристик;

– ефективний та динамічний розподіл обчислень на інфраструктурі розподілених обчислень;

– прогнозування та ідентифікація можливих перевантажень керуючих систем.

При цьому завдання ідентифікації трафіку включає необхідність розпізнавання великої кількості типів сервісів (враховуючи URLLC) при умові виключення додаткових затримок у трафік, та можливість розширення та налаштування алгоритму під певне географічне розташування мережі та сервісів [4,6].

Існуючі системи моніторингу трафіку зазвичай використовували технології, що засновані на періодичному захопленні трафіку та аналізі його заголовків. Такий метод має низку недоліків, а саме: внесення затримки в потік, реалізація аналітичного модуля у вигляді додаткового апаратно-програмного рішення рівня передачі даних (Data Plane), складність побудови такого пристрою [4,6,8].

Поява технологій ШІ, та розробка на їх основі спеціалізованих систем глибокої інспекції пакетів DPI (Deep Packet Inspection) дозволяє перейти до моніторингу трафіку на сервісному рівні інфраструктури програмно-конфігурованих мереж [12,13,14].

Таким чином, як результат проведеного аналізу можливостей по моніторингу трафіку в мережах 5G/IMT-2020, в даній роботі пропонується підхід до виконання вище вказаного завдання, заснований на використання нейронних мереж певної архітектури, завдання яких виявлення та ідентифікація визначеного типу трафіку в сумі загального вхідного трафіку визначеної мережі.

Система DPI, як видно з назви, виконує глибокий аналіз всіх пакетів вхідних даних, що проходять через неї. Термін «глибокий» має на увазі аналіз пакета на верхніх рівнях моделі OSI, а не лише за стандартними номерами портів. Крім вивчення пакетів за

деякими стандартними патернами, за якими можна однозначно визначити приналежність пакета певному додатку, скажімо, за форматом заголовків, номерами портів і т.п., система DPI здійснює і так званий поведінковий аналіз трафіку, який дозволяє розпізнати додатки, які не використовують для обміну даними заздалегідь [14,15].

Втілення технології DPI в вигляді окремо реалізованої програмно-апаратної системи, її залучення до ідентифікації та моніторингу трафіку на серверному рівні за інтерфейсом REST формує окрему систему ідентифікації трафіку з властивістю виключення затримок та мінімізації складності реалізації програмно-апартного комплексу. Така система може бути представлена і як апаратно-програмний комплекс, так і програмне рішення (у вигляді програми мережі).

Принципова архітектура запропонованого рішення по побудові системи ідентифікації трафіку представлена на Рис. 1.

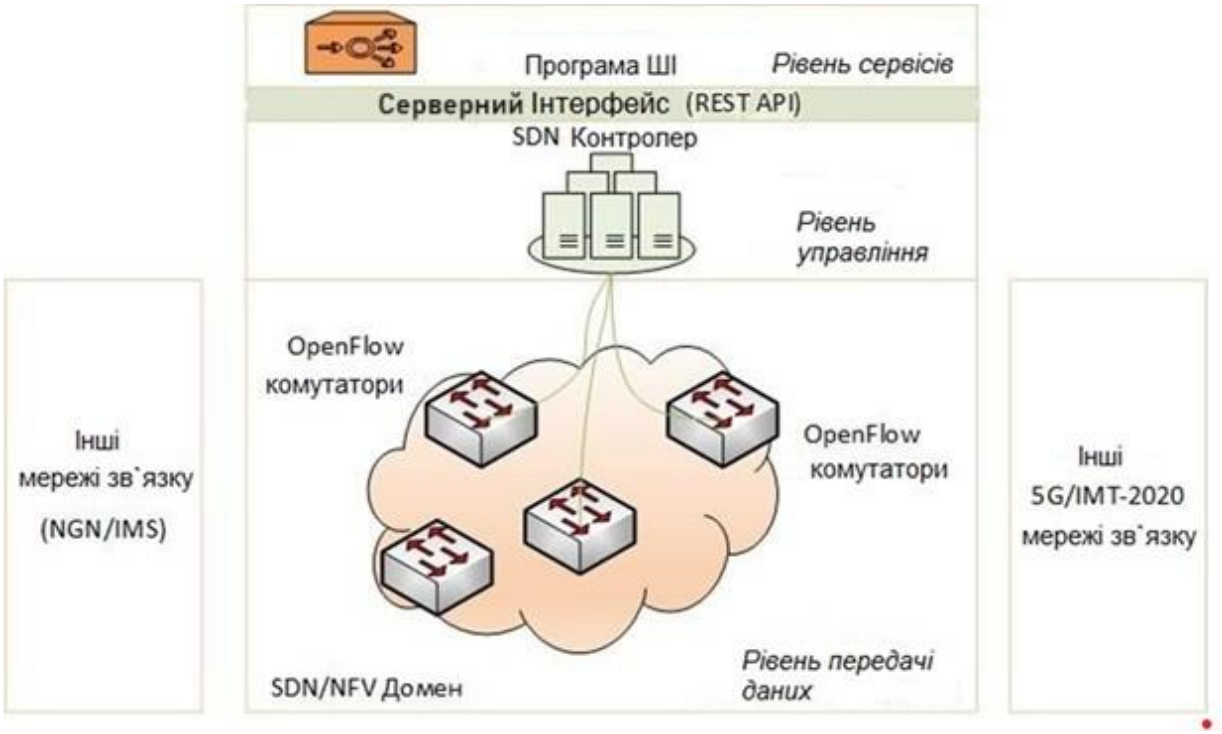


Рис. 1. Принципова архітектура системи ідентифікації трафіку на основі технології DPS

Для системи моніторингу трафіку поданого типу на основі технології DPS всі пристрої та потоки є «цифровим об'єктом», що має ряд параметрів та функцій (дій над параметрами), представленим набором методів [14,15.

Такий підхід дозволяє реалізувати систему моніторингу інфомаційно-комунікаційної мережі, яка працюватиме з даними про потоки (метадані) у режимі «на льоту». Дана система, через свої властивості, дозволяє не вносити додаткові затримки у трафік, а також будь-яким чином змінювати його активність (зміни законів розподілу, інтенсивність тощо). Основою функціонування даної системи є процес «спостереження» за активністю потоків і формування «загальної картини» передачі даних, що відбувається в підконтрольній мережі.

Вхідні дані для поданої системи формуються на основі тих даних, які система може отримати через серверний Інтерфейс контролера і Оркестратора мережі.

Оскільки у цій роботі розглядається аналітика активності потоків і виявлення потоків Інтернету Речей лише на рівні передачі даних, аналітична система має можливість запитати дані таблиць потоків

(Flow Tables) з усіх підконтрольних комутаторів SDN у контролера. Аналізуючи дані, що відображаються у двох глобальних частинах таблиці: Match Field і Actions, можна дійти висновку, що на їх основі можна скласти метамодель потоків. Скорочену структуру таблиці потоків комутатора відображено на Рис. 2.

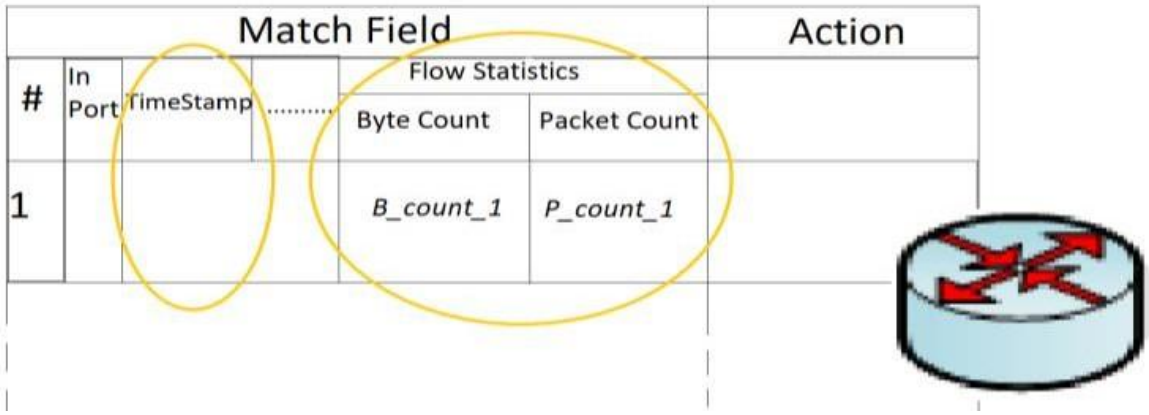


Рис. 2. Структура таблиці потоків комутатора S

На Рис. 2 жовтими колами виділено ті дані, які використовуються для формування метамоделі про досліджуваний потік у мережі.

Однією з важливих особливостей цих даних є те, що на основі лічильників Byte Count та Packet Count не можна визначити точну довжину пакета в потоці, оскільки за один момент часу лічильники можуть дорівнювати: Byte Count 1500, Packet Count

2. Відповідно, на основі цих даних не можна точно визначити довжину кожного з пакетів, зареєстрованих у потоці за проміжок часу: $\Delta T = 1$ с.

Варто також відзначити, що лічильники відображають сумарне значення параметрів [Byte Count, Packet Count]. Однак, крім даних лічильників у таблиці потоків існує ще один параметр – Time Stamp, який дозволяє оцінити в кожен момент часу миттєве значення [ByteCount_delta] та [PacketCount_delta]. Таким чином, за довільний період часу ΔT, маючи відліки значень [Byte Count], [Packet Count], [TimeStamp], можливо скласти набір даних з встановленої структурою даних, де кожен відлік відображає миттєве значення [ByteCount_delta] та [PacketCount_delta].

Відліки значень [Byte Count], [Packet Count], [TimeStamp] формуються шляхом щосекундних запитів на контролер мережі SDN через програмний інтерфейс RESTAPI. Структура формується на підставі запитів $DataSet_{RQ}$ з «сирими» даними. Вираз (1) та вираз (2), що є залежностями перетворення на необхідний формат $DataSet_{ML}$ з миттєвими значеннями (3) наведено нижче.

Приймемо: TimeStamp_deltas –TS = 1 [sec.] = const,
PacketCount_delta – PC_{delta},
ByteCount_delta – BC_{delta}, а
Тоді:

$$DataSet_{RQ} = \begin{matrix} [TimeStamp] & [ByteCount] & [PacketCount] \\ TimeStamp_{11} & ByteCount_{12} & PacketCount_{13} \\ TimeStamp_{21} & ByteCount_{22} & PacketCount_{23} \\ \dots & \dots & \dots \\ TimeStamp_{N1} & ByteCount_{N2} & PacketCount_{N3} \end{matrix} \quad (1)$$

$$BC\_delta_{N2} = ByteCount_{N2} - ByteCount_{(N-1)} \quad \text{if } N \geq 1$$
$$PC\_delta_{N2} = PacketCount_{N2} - PacketCount_{(N-1)2} \quad \text{if } N \geq 1 \quad (2)$$

$$DataSet_{ML} = \begin{matrix} [TimeStamp] & [ByteCount] & [PacketCount] \\ TS_{delta11} & BC_{delta12} & PC_{delta13} \\ TS_{delta21} & BC_{delta22} & PC_{delta23} \\ \dots & \dots & \dots \\ TS_{deltaN1} & BC_{deltaN2} & PC_{deltaN3} \end{matrix} \quad (3)$$

При цьому розрахунок сумарних значень параметрів трафіку за встановлений проміжок часу пропонується здійснювати за виразами (4) та (5):

$$ByteCount = \sum_{\Delta T \rightarrow N=1}^{N=\Delta T/TS} BC_{deltaN2} \quad (4)$$

$$PacketCount = \sum_{\Delta T \rightarrow N=1}^{N=\Delta T/TS} PC_{deltaN2} \quad (5)$$

При цьому розрахунок сумарних значень параметрів трафіку за встановлений проміжок часу пропонується здійснювати за виразами (4) та (5):

ByteCount = ∑_(ΔT □(→⊥) N=1)^(N=ΔT⁄TS)▒ 〖BC〗 _deltaN2 (4)

PacketCount = ∑_(ΔT □(→⊥) N=1)^(N=ΔT⁄TS)▒ 〖PC〗 _deltaN2 (5)

Як було зазначено вище, для вирішення вищеописаного завдання було проведено комплексний аналіз математичних методів класифікації з урахуванням особливостей вхідних даних, аналізованих даних про потоки, а також особливостей початкової вимоги – роботи системи у режимі «на льоту».

В результаті проведеного аналізу обрано підхід використання нейронних мереж.

Визначена мета застосування нейронної мережі, а саме – ідентифікація та моніторинг трафіку мережі 5G/IMT-2020 при умові

роботи системи у режимі «на льоту» (мінімізація затримок вхідного потоку) та особливості ідентифікації та аналізу вхідних даних DataSetML - вир. (1)- (5) формують вимоги до нейронної мережі, що буде обрана для досягнення поставленої мети.

Відповідно до поставленої мети та особливостей вхідних даних DataSetML обрана відповідна архітектура нейронної мережі та подана система, алгоритм та методика її застосування.

На даний момент існує великий різновид нейронних мереж. Одним з типових завдань таких мереж є класифікація потоків даних (інформації) як об'єкту спостереження. Поширений спосіб класифікації – спосіб на основі описів об'єктів з використанням типових ознак, у якому кожен об'єкт характеризується набором числових чи нечислових ознак [16].

Аналіз типів даних, що включені в трафік мережі 5G/IMT-2020 показує, що їх відкриті ознаки не дають точності класифікації даних. Причина полягає в тому, що ці дані містять приховані ознаки, як то кількість та розміри пакетів даних, час їх затримки та інші.

Виходячи з цього, виникає необхідність в підборі стандартної нейронної мережі, призначеної до класифікації та аналізу даних трафіку, що мають приховані ознаки.

Аналіз ряду відомих нейронних мереж, призначених для моніторінгу потоків різних даних показав, що однією з таких нейронних мереж є нейронна мережа з функцією Deep Learning – функцією машинного навчання, що використовує багатошарові мережі для аналізу даних, автоматичного виоучення ознак та створення більш складних моделей для атрішення заданя індентифікації, моніторингу та аналізу. [16,17].

Вказана мережа є набором алгоритмів машинного навчання, які дозволяють моделювати високорівневі абстракції в масивах даних. В процесі такого моделювання вказана нейрона мережа Deep Learning здатна виділяти з даних приховані ознаки, при цьому нейронні мережі містять більше, ніж 2 прихованих шари. Тому, враховуючи особливості об'єкта моніторингу – трафік мережі 5G/IMT-2020) та його ознаки (числові статистичні ряди, що є метаданими) обрано вказану нейронну мережу з Deep Learning.

Розробка та навчання нейронної мережі Deep Learning проводилися на основі високорівневої мови програмування Python з її різноманітними бібліотеками та фреймворками.

Як штучну нейронну мережу для вирішення завдання моніторингу трафіку обрано рекурентну нейронну мережу з додатковими шарами LSTM (Long Short-Term Memory) [18]. Мережа LSTM є універсальною тому, що з достатньою кількістю нейронів вона має можливість виконувати будь-які обчислення, які може виконувати звичайний комп'ютер. Включений в мережу глибинний модуль LSTM, як частина штучної нейронної мережі (ШНМ), дозволяє виявляти закономірності впливу даних попередніх відліків на поточні з урахуванням великого розкиду значень між ними. Тобто, наприклад, періодичність, що може виявляється у трафіку Інтернету Речей, самоподібність характеру якого наводиться у багатьох працях [5,6,19].

Оскільки обрана архітектура нейронної мережі реалізує принцип «навчання із залученням вчителя», процес її застосування вимагає формування початкового навчального набору даних з маркованими позначками, та наступного збереження стану навченої мережі [19].

Для навчання нейронної мережі вхідний трафік DataSetML перетворюється в DataSetMLtrain шляхом додавання нового стовпчика даних, в кожномурядку якого стоїть ідентифікатор статистичної вибірки. Відповідно, для навчання розпізнавання більшого типу трафіку даний навчальний набір даних потрібно розширити, помітивши відповідну статистичну вибірку міткою трафіку, наприклад – IoT [17,18].

Архітектура мережі Deep Learning містить 2 повнопов'язані глибинні рівні LSTM і 2 пов'язані глибинні рівні RNN (Recurrent Neural Network), кожен з яких містить 7 прихованих нейронів.

Запропонований метод моніторингу трафіку мережі пропонується реалізуватиу вигляді

програмного модуля мовою програмування Python версії 2.7 для аналітичної системи, реалізованої на основі MVC патерну.

Дане програмне забезпечення, відповідного обраного підходу до ідентифікації, схематично поданого на Рис.3 втіленням відповідної функції, реалізованої через програмний застосунок, вкладений поверх серверного програмного інтерфейсу API контролера програмно-конфігурованої мережі та Оркестратора мережі лабораторії.

Для формування навчальних для Штучної Нейронної Мережі наборів даних визначено спеціальні умови [17,18]:

– будь-який інший трафік, що не відноситься до дослідження, що проводиться, повинен бути вимкнений і не проходити в SDN-сегмент;

– мережа повинна бути модифікована так, щоб була можливість однозначно ідентифікувати потік з трафіком, що генерується в таблиці потоків комутаторів.

Для апаратної реалізації програмно-апаратного комплексу нейронної мережі, що забезпечить дослідження методів та алгоритмів застосування Штучного інтелекту для мереж зв'язку 5G/IMT-2020 пропонується наступна архітектура лабораторного комплексу, що подана на Рис.1 [20,21,22].

Лабораторний комплекс пропонується побудувати в вигляді з'єднаних в кільце трьох програмно-керованих комутаторів SDN, що реалізують протокол OpenFlow версії 1.0. Вибраний SDN- контролер з відкритим вихідним кодом OpenDaylight програмно-конфігурованої мережі реалізує розподіл навантаження по мережі за замовчуванням. Оскільки цей стенд зібраний в першу чергу з метою проведення досліджень, у рівень управління вказаним компрлексом інтегрований звичайний комутатор Ethernet, який агрегує всі службові потоки і забезпечує доступність контролера SDN. Також в мережі даного комутатора реалізовано доступ до серверного рівня контролера SDN.

Для розширення кількості можливих підключень до кожного комутатора SDN був послідовно підключений простий Ethernet-комутатор, що виконує роль агрегації вхідного потоку в топології поданого комплексу.

За основу програмного забезпечення SDN-контролер пропонується застосувати модульну архітектуру Java Karaf, яка забезпечує необхідну розширюваність контролера шляхом встановлення необхідних додаткових модулів або власної розробки модулів з подальшою інтеграцією до програмної шини контролера.

В якості Оркестратора мережі пропонується застосування програмного рішення – OpenStack, яке інтегрується також із рішенням OpenDaylight через програмний модуль «OpenStack Neutron» [11,23].

Для реалізації трафік-генератора типового сервісу Інтернету Речей пропонується застосування серверу з операційною системою Linux Ubuntu версії LTS з встановленим і налаштованим сервісом IoTDM (Internet of Things Data Management), що розроблений окремим підрозділом OpenDaylight з складу консорціуму The Linux Foundation [20,23].

Функціонування комплексу передбачає умову, при якій всі програми розгортаються на одному фізичному сервері. У процесі розробки роль сервера виконує персональний ноутбук, у процесі тестування стаціонарний ПК.

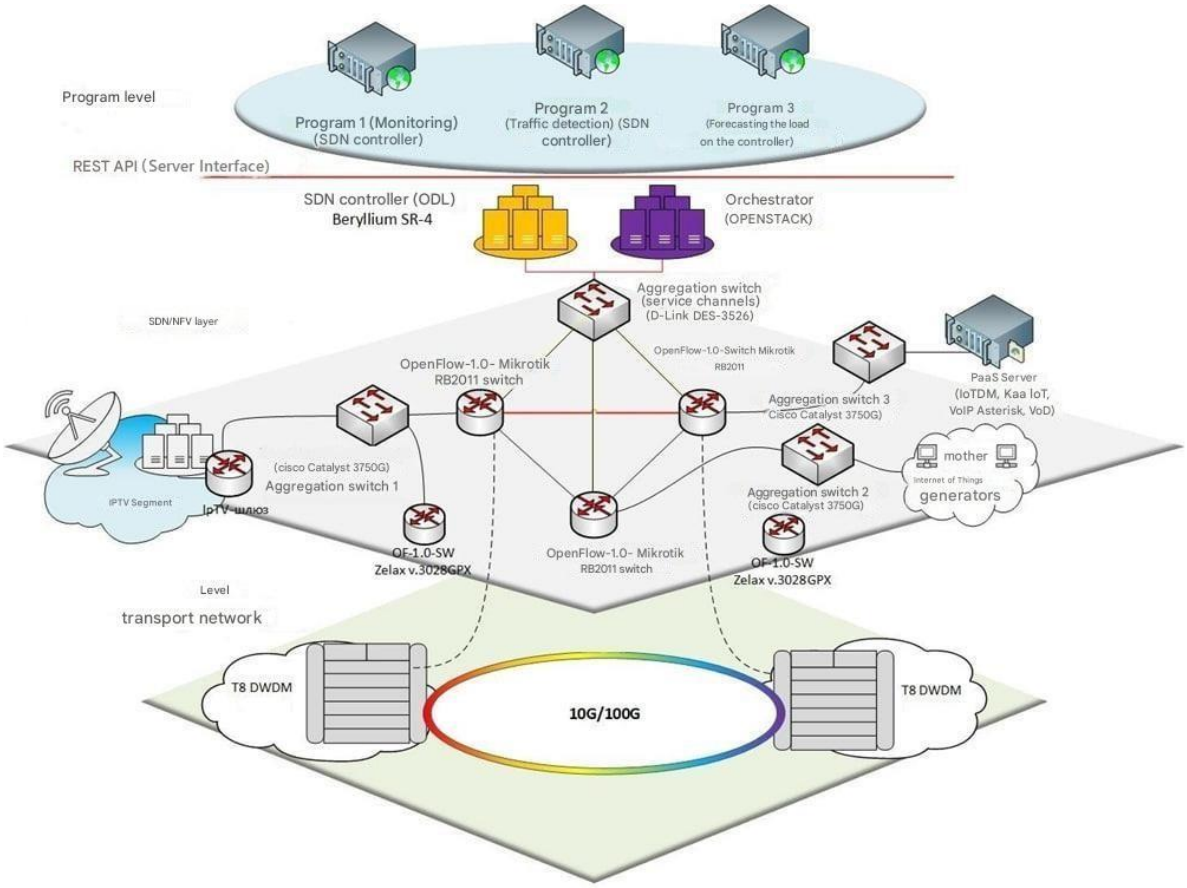


Рис. 3. Архітектура програмно-апаратного комплексу нейронної мережі

В результаті експериментального дослідження за допомогою запропонованого програмно-апаратного комплексу, схематично поданого на Рис.3 отримано масив даних DataSetMLtrain, який сформований з двох маркованих наборів даних (IoT, Video).

Далі DataSetMLtrain подається на вхід нейронної мережі, конфігурація якої відображена вище.

За отриманим масивом DataSetMLtrain побудована діаграма розкиду значень. Діаграма наведена на Рис. 4.

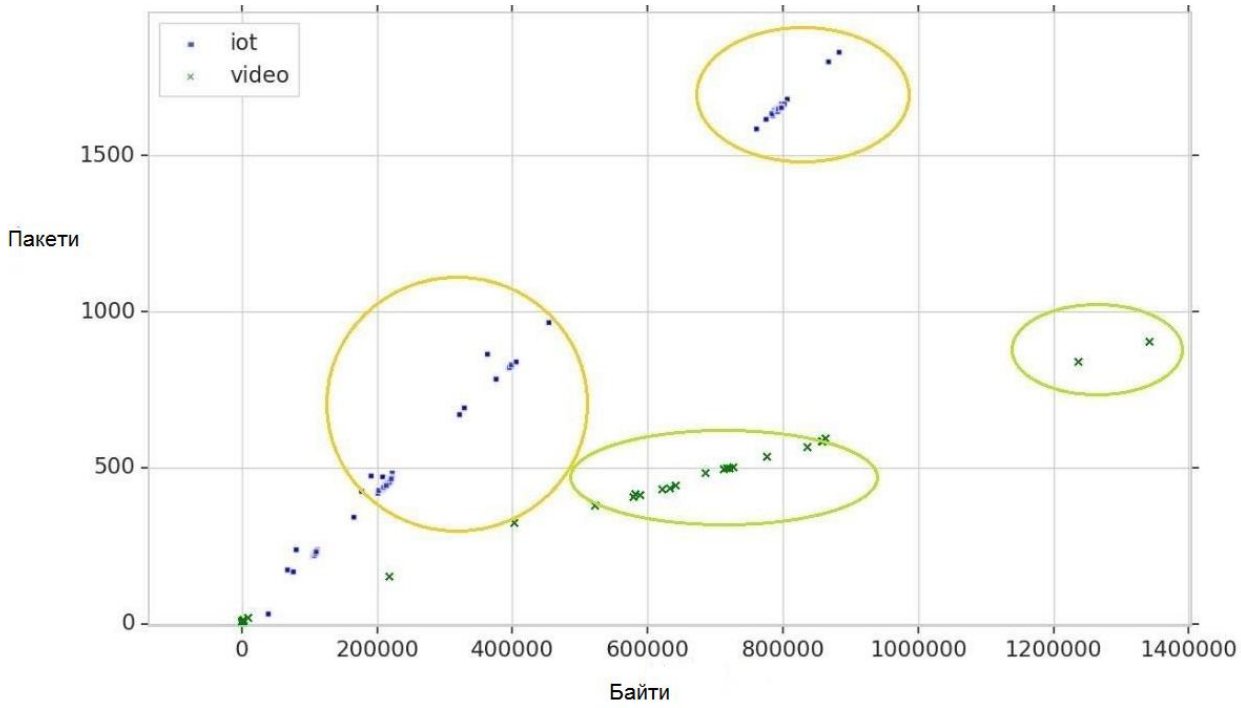


Рис. 4. Діаграма розкиду значень DataSetML$train$

Після того, як навчальний набір даних сформований, згідно з алгоритмом, викладеним вище в цій роботі, програмний модуль розробленого додатка, що реалізує Штучну Нейронну Мережу, переходить до процесу навчання нейронної мережі.

Протягом процесу навчання нейронної мережі розробленим додатком контролюються та фіксуються значення наступних параметрів [23,24]:

Train Accurancy (Точність навчання);

Test Accurancy (Точність проходження тесту ШНМ);

Train loss (Помилки під час навчання);

Test loss (Помилки під час проходження тесту ШНМ);

Графік, що відображає дані параметри у процесі проходження епох навчання, наведено на Рис.5.

Аналіз залежностей, поданих на Рис.5, Рис.6, де відображено процес навчання мережі, показав, що для поставленого завдання моніторингу трафіку розроблена штучна нейронна мережа успішно пройшла процес навчання. В результаті процесу навчання нейронної мережі та перевірки її роботи на тестових наборах даних у навченому стані розроблена нейронна мережа може ідентифікувати потік Інтернету Речей, що генерується, з ймовірністю 99,7%.

Варто відзначити, що отримана ймовірність розпізнавання потоків даних (трафік даних від мережі Інтернету Речей та трафік Відео) має досить високі значення і при збільшенні кількості типів трафіку, що розпізнаються, дана ймовірність може знизитися, проте перебуває у допустимих межах (80-90%) [21,22].

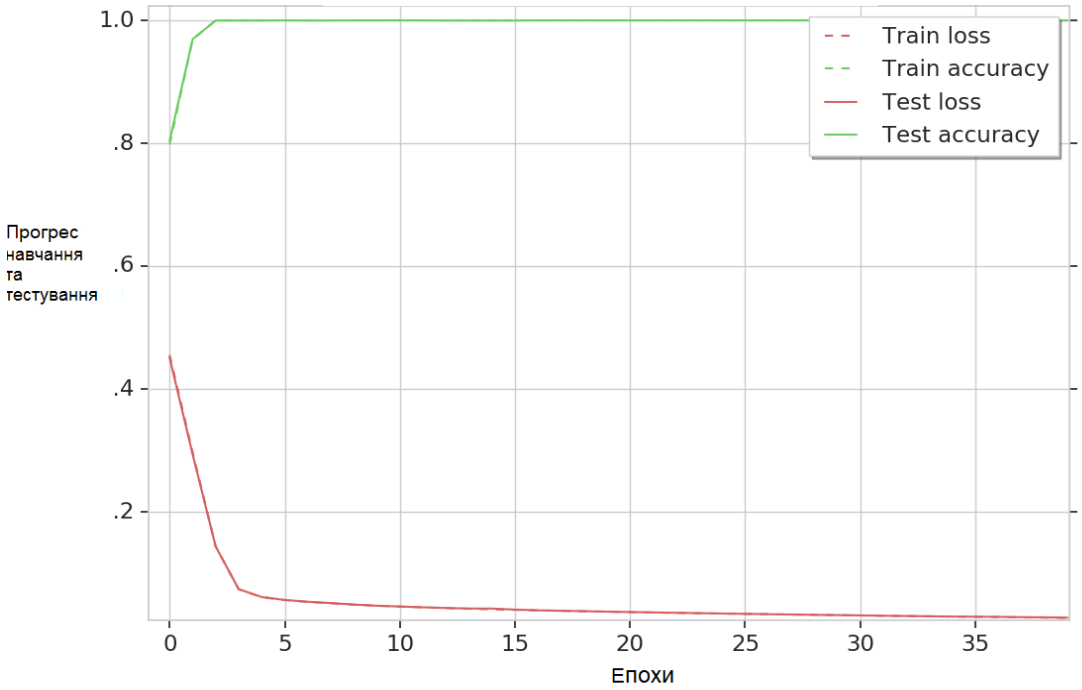


Рис. 5. Графік навчання та тестування штучної нейронної мережі

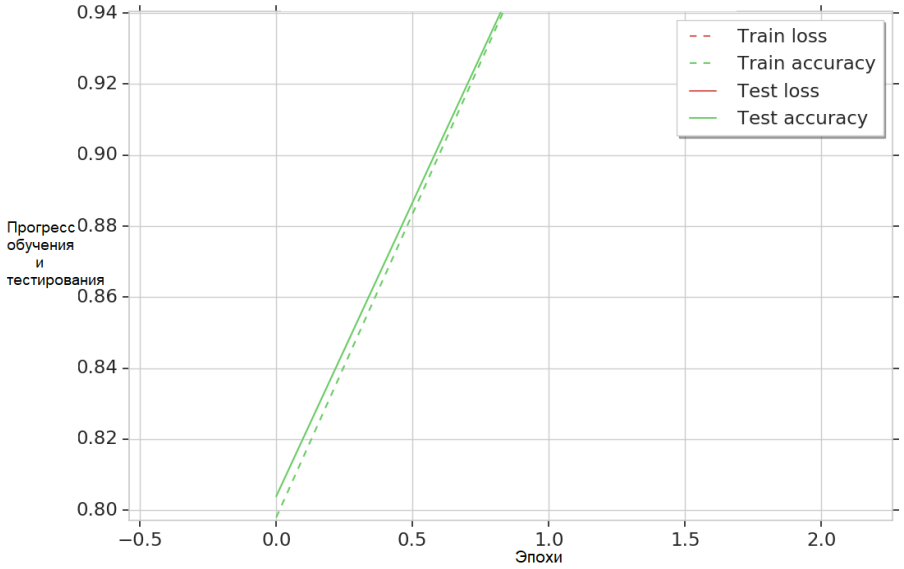


Рис. 6. Масштабований графік навчання та тестування штучної нейронної мережі

Таким чином, запропоновано метод моніторингу навантаження в телекомунікаційних мережах 5G/IMT-2020, який включає:

– процес формування метамоделі потоків вхідних даних в мережу 5G/IMT-2020: DataSetML = ([Byte Count], [Packet Count], [TimeStamp]);

– модифікація DataSetML в навчальний набір DataSetMLtrain шляхом включення в набір даних ідентифікаторів статистичної вибірки;

– формування програмно-апаратного комплексу навчання та оцінки ефективності застосування нейронної мережі для моніторингу навантаження в мережі 5G/IMT-2020;

– побудову структури нейронної мережі Deep Learning з додатковими шарами LSTM;

– реалізацію процесу навчання нейронної мережі та підготовка її до функціонування за призначенням;
– втілення запропонованого методу та програмно – апаратного комплексу в процес моніторингу навантаження трафіку в телекомунікаційних мережах 5G/IMT-2020.

ВИСНОВКИ

1. В роботі проведено аналіз процесу ідентифікації вхідного трафіку в мережах 5G/IMT-2020, побудованих по технології Ultra-reliable and low latency communications та визначені його особливості та напрямки досліджень по підвищенню ефективності його моніторингу.
2. Для вирішення завдання моніторингу та ідентифікації трафіку мережі 5G/IMT-2020в роботі розроблена та подана відповідна методика моніторингу. Вказана методика включає формування масивів метаданих вхідного потоку, модифікацію їх в набір навчальних даних, формування навчального програмно-апаратного комплексу та розбудову структури нейронної мережі, проведення процесу навчання нейронної мережі та втілення її в процес моніторингу навантаження в телекомунікаційних мережах 5G/IMT-2020.
3. Оцінка результатів процесу навчання запропонованої нейронної мережі та перевірки її роботи на тестових наборах даних у навченому стані показала, що подана в роботі нейронна мережа здатна провести моніторинг та ідентифікувати згенерований від сервісів Інтернету Речей трафі з ймовірністю до 99,7%.

В процесі моніторингу та ідентифікації трафіку від двох та більше сервісів дана ймовірність може знизитися, проте перебуває у допустимих межах 80-90%.

REFERENCES

1. В.**С. Шуста, А.І. Сусла, В.Ю. Біганич**, «Трансформація мережевих технологій: від 4G до 5G», Вчені записки ТНУ імені В.І. Вернадського, Серія: Технічні науки, 2024, 3, С.94-99
2. **Technical Report I**TU-T GSTP-TN5G: Transport network support of IMT-2020/5G. – SG15- TD338/PLEN, October 2018.
3. Draft Recommendation Characteristics of transport networks to support IMT-2020/5G (Gctn5g). – SG15-TD295/PLEN, October 2018
4. **S. Hendaoui, F. Hendaoui, N. Zangar**, «Dynamic proactive–reactive scheduling for URLLC in 5G: Leveraging XGBoost and network virtualization», Physical Communication, 68, art. no. 102553. 2025.
5. **X. Shen, W. Liao, Q.Yin**, «A Novel Wireless Resource Management for the 6G-Enabled High-Density Internet of Things», IEEE Wirel. Commun, 2022, 29, p.32–39.
6. **X. Li, L.D. Xu**, «A Review of Internet of Things—Resource Allocation», IEEE Internet Things J., 2021, 8, p.8657–8666.
7. the Internet of Things: A Survey», J. Netw. Comput. Appl. 2022, 206, 103464.
8. **A.A. Ateya, S. Bushelenkov, A. Muthanna, A. Paramonov,** «Multipath Routing Scheme for Optimum Data Transmission in Dense Internet of Things», Mathematics. 2023; 11(19):4168.
9. **Yongho Seok, Youngseok Lee, Yanghee Choi and Changhoon Kim,** "Dynamic constrained multipath routing for MPLS networks", Proceedings Tenth International Conference on Computer Communications and Networks (Cat. No.01EX495), Scottsdale, AZ, USA, 2001, pp. 348-353.
10. **W.A. Aljoby, X. Wang, D.M. Divakaran, T.Z.J. Fu,** «DiffPerf: Toward Performance Differentiation and Optimization WithSDN Implementation», IEEE Transactions on Network and Service Management, 2024, 21(1), pp. 1012 – 1031.
11. **L. Brown,** «Advanced Techniques in NFV and SDN for Scalable Networks», International Journal of Communications Technology, 2022, 12(4), з.78-92.
12. **L.-H. Chang, Tsung-Han Lee, Hung-Chi Chu, and Cheng-Wei Su,** "Application-Based Online Traffic Classification with Deep Learning Models on SDN Networks", Adv. technol. innov., Sep. 2020, 5(4), pp. 216–229.
13. **P.K.R. Maddikunta, Q.-V. Pham, D.C. Nguyen**, (eds), «Incentive Techniques for the Internet of Things: A Survey», J. Netw. Comput. Appl. 2022, 206, 103464.
14. **Foreman, Justin, Waters, Willie L., Kamhoua, Charles A., Hemida, Ahmed H. Anwar, Acosta** (2024) Detection of Hacker Intention Using Deep Packet Inspection? Journal of Cybersecurity and Privacy, 4 (4), pp. 794 – 804. DOI: 10.3390/jcp4040037

15. **Parra G. D. L. T., Rad P., Choo K. K. R., Beebe N.** Detecting internet of things attacks using distributed deep learning. Journal of Network and Computer Applications, 2020, vol. 163, pp. 102662. doi:10.1016/ j.jnca.2020.102662
16. **Al-Garadi M. A., Mohamed A., Al-Ali A., Du X., Guizani M. A** Survey of machine and deep learning methods for internet of things (IoT) security. IEEE Communications Surveys & Tutorials, 2020, vol. 22, no. 3, pp. 1646–1685. doi:10.1109/COMST.2020.2988293
17. **Pacheco, Fannia & Exposito, Ernesto & Gineste, Mathieu & Baudoin, Cédric & Aguilar**, **Jose.** (2018). Towards the Deployment of Machine Learning Solutions in Network Traffic Classification: A Systematic Survey. IEEE Communications Surveys & Tutorials. PP. 1-1. 10.1109/COMST.2018.2883147.
18. **Hochreiter, Sepp & Schmidhuber, Jürgen**. (1997). Long Short-Term Memory. Neural Computation. 9. 1735-1780. 10.1162/neco.1997.9.8.1735.
19. **Кондакова, А., Корецький, О.** (2024). Моделі побудови мереж інтернету речей для управління інфраструктурою міста. Smart Technologies, 2(15), 124–132. DOI: https://doi.org/10.32347/st.2024.2.1901
20. **Oleksandr Koretskyi, Rodion Khvorostianyi, Yevhen Bondarenko** Hardware and software complex for the functioning of a neural network for identification and traffic management in 5G/IMT-2020 networks. The IV International Conference «Emerging Technology Trends on the Smart Industry and the Internet of Things «TTSIIT», м. Київ, Україна, 21-22 січня 2025 р., Київ, КНУБА. С.44-49. https://drive.google.com/file/d/145aOtud0A7zl 4N9RKXiFyt9iMLKYsMt_/view
21. **J.Yao, Y. Wang, Q. Li, Mao** (2022). An Efficient Routing Protocol for Quantum Key Distribution Networks. Entropy, No.24, 911.
22. Study on Scenarios and Requirements for Next Generation Access Technologies (3GPP TR 38.913 version 14.2.0 Release 14). https://www.etsi.org/deliver/etsi_tr/138900_138 999/138913/14.02.00_60/tr_138913v140200p.p df
23. **F. H. Nazir, Q. Ahmad, R. Badlishah, & Elias**, (2017). «Software-Defined Network Testbed Using ZodiacFX a Hardware Switch for OpenFlow», ICST Transactions on Scalable Information Systems», No.4. 153150.
24. **V. Artem, A.A. Ateya , A. Muthanna, A. Koucheryavy**, (2019). Novel AI-Based Scheme for Traffic Detection and Recognition in 5G Based Networks», In: Galinina, O., Andreev, S., Balandin, S., Koucheryavy, Y. (eds), «Internet of Things, Smart Spaces, and Next Generation Networks and Systems», Lecture Notes in Computer Science, No. l, 11660. Springer, Cham.

Traffic identification method in 5g/imt-2020 telecommunication networks based on artificial intelligence

Oleksandr Koretskyi, Anastasiia Kondakova

Abstract. The paper analyzes the process of identifying incoming traffic in 5G/IMT-2020 networks, built on the Ultra-reliable and low latency communications technology, identifies its features and research directions to improve the efficiency of traffic identification taking into account the requirements for the functioning of the considered network.

To solve the problem of identifying traffic in the 5G/IMT-2020 network, the paper develops and presents an appropriate methodology. The specified methodology includes the formation of metadata arrays of the incoming stream, their modification into a set of training data, the formation of a training software and hardware complex and the development of the neural network structure, the training process of the neural network and its implementation in the process of identifying traffic in 5G/IMT-2020 telecommunication networks.

Evaluation of the results of the training process of the proposed neural network and verification of its operation on test data sets in the trained state showed that the neural network presented in the work is able to monitor and identify traffic generated from Internet of Things services with a probability of up to 99.7%. In the process of monitoring and identifying traffic from two or more services, this probability may decrease, but is within the acceptable limits of 80-90%.

Keywords: identification of incoming traffic of information and communication networks, 5G/IMT-2020 network, artificial intelligence, neural network, information and communication network load management.

Про підходи до прискорення перебору об’єктів в задачах обробки інформації

Олександр Гайша[1], *Сергій Лєнков*[2], *Олена Гайша*[3]

[13]Міжнародний класичний університет імені Пилипа Орлика
вул. Котельна, 2, Миколаїв, Україна, 54003
[2]Військовий інститут Київського національного університету імені Тараса Шевченка
Юлії Здановської, 81, Київ, Україна, 03189
[1]oleksandrgaysha1982@gmail.com, https://orcid.org/0000-0003-3711-547
[2]lenkov_s@ukr.net, https://orcid.org/0000-0001-7689-239
[3]haishaolena@gmail.com, https://orcid.org/0009-0000-4543-912

Received 03.02.2025, accepted 29.03.2025
https://doi.org/10.32347/st.2025.3.1207

Анотація. У роботі докладно розглянуто поняття процесу перебору. Він здійснюється з об'єктами x, які є елементами деякої множини X. Метою перебору є виконання деякої активної дії y = f(x) над елементами, що перебираються. Ця дія може, звичайно, застосовуватися і до кожного елемента, але у переважній більшості практичних завдань дія застосовується тільки до тих елементів, які задовольняють певній умові перебору P(x).

Найбільш простим способом перебору є повний перебір або «брут-форс». При цьому до кожного без винятку елементу множини X застосовується перевірка умови P(x) і, залежно від результату, здійснюється виконання дії y = f(x). Очевидно, що повний перебір є певним еталонним процесом у тому сенсі, що саме в порівнянні з ним зручно оцінювати ефективність того чи іншого алгоритму обмеженого перебору. Алгоритми обмеженого перебору доцільно застосовувати у переважній більшості практичних випадків, якщо тільки повний перебір не займає малопомітний час порядку часток секунди. Однак, у деяких випадках доцільно оптимізувати (обмежувати) перебір навіть за таких малих часів виконання всього процесу – зокрема, якщо такі короткі процеси перебору повторюються багато разів на одиницю часу, складаючись сумарно у деякий, уже досить значний час.

Метод гілок і меж є найвідомішим алгоритмом обмеженого перебору і дозволяє, поступово просуваючись від одного варіанта до іншого, відкидати цілі набори елементів, завідомо не відповідних умові P(x). Однак, даний метод не дозволяє робити надійні попередні оцінки ефективності обходу всього дерева відповідних елементів, а його застосовність для кожної конкретної задачі стає зрозумілою тільки в процесі перебору.

Гайша Олександр
к.т.н., доцент, доцент кафедри інженерних технологій

Сергій Лєнков
д.т.н. професор, головний науковий співробітник

Гайша Олена
ст. викладач кафедри інженерних технологій

У цій роботі пропонується метод обмеження, заснований на поданні умови перебору P(x) у кон’юнктивній або диз’юнктивній нормальній формі з подальшою перевіркою цієї умови за схемою скорочених логічних обчислень з використанням лише однієї з умов. Також для перевірки цієї умови пропонується заздалегідь проводити обчислення деякої додаткової інформації, спираючись на яку під час перебору, можна отримувати кінцевий результат швидше. В якості такої інформації можна прийняти поділ на класи (кластери) всіх об'єктів, що перебираються, і виконання переборів тільки в

межах кластерів, що скорочує загальну кількість елементів, що перебираються.
Пропонований підхід докладно пояснюється на прикладі рішення актуальної задачі з області комп'ютерної графіки, яка полягає у пошуку всіх правильних n-кутників (на прикладі правильних трикутників) на множині N точок тривимірного простору, які задані своїми координатами. Наведено словесний опис повного перебору, необхідного для вирішення такого завдання, а також докладний опис алгоритму обмеженого перебору, який складений за пропонованою схемою (скорочені обчислення за КНФ і попередня кластеризація всіх об'єктів, що перебираються або, точніше, пов'язаної з ними інформації – відстаней між усіма парами).
У роботі наведено числові розрахунки, що дозволяють провести теоретичну оцінку ефективності запропонованого рішення. Встановлено, що при ідеально рівномірному розподілі всіх об'єктів, що перебираються по K класам, відбувається значне зменшення кількості операцій порівняння відстаней при пошуку рівносторонніх трикутників (при числі K = 50 – теоретичне прискорення становить до 3 разів). Прискорення, що одержується, досліджено практично на модельній програмній реалізації, написаній мовою C#, в результаті чого показано, що реальне прискорення є досить близьким до ідеального, а сам запропонований підхід, відповідно, є ефективним.

Ключові слова: перебір, прискорення перебору, кластеризація, декомпозиція умов перебору.

ВСТУП

На сьогоднішній день існує значна кількість предметних областей, де застосовуються алгоритми перебору. Насамперед, з обчислювальної точки зору найскладніші перебори зустрічаються у завданнях пошуку інформації у базах даних (БД) [1], що розглянемо докладніше.

По-перше, такі завдання характеризуються значними обсягами даних, серед яких потрібно проводити перебір. Якщо йдеться про програму, запущену в оперативну пам'ять і яка не використовує базу даних для зберігання та вилучення інформації, то, очевидно, оброблювані дані будуть розміщені в оперативній пам'яті ЕОМ, яка на сьогоднішній день для персональних комп'ютерів (далі – ПК) становить величини порядку десятка гігабайт. Для продуктивних серверів обсяги оперативної пам'яті можуть досягати сотень гігабайт, що є досить великою величиною, яка вимагає застосування спеціальних апаратних рішень (проти звичайної архітектури ПК). Однак, дані, які розміщуються в базах даних, навіть при використанні звичайних ПК легко можуть досягати зазначеного розміру в 100 Гб, а при розміщенні на високопродуктивних серверах обсяги даних, що зберігаються, обчислюються сотнями терабайт, що в цілому на кілька порядків більше обсягів даних, які можна розміщувати в оперативній пам'яті ЕОМ. Таким чином, виконання операцій на даних, розміщених у БД, потенційно може потребувати на кілька порядків більше дій, ніж застосування того самого алгоритму в програмах, що не використовують підключення до БД [2].

Другим аспектом застосування алгоритмів пошуку (а також і сортування) на даних, розміщених саме в БД, є необхідність урахування того, що доступ до даних, розміщених на диску (а файли БД природно зберігаються в дисковій пам'яті, а не оперативній), завжди здійснюється значно повільніше, ніж до даних в оперативній пам'яті ЕОМ. Навіть сучасні швидкі SSD накопичувачі мають значно меншу швидкість доступу до даних, ніж пам'ять класу DDR5 або навіть DDR4. Також, відомо, що дискові накопичувачі оптимізовані для потокової передачі/прийому даних, а оперативна пам'ять дозволяє швидко отримувати доступ до невеликих розрізнених блоків даних. Це означає, що при інтенсивних процесах додавання/видалення даних з БД її файли будуть сильно фрагментовані, що ще гірше уповільнить роботу з нею. Таким чином, задачі оптимізації алгоритмів, що працюють із базою даних, слід приділяти велику увагу.

Ще одним аспектом організації пошуків (тобто переборів) саме в базах даних є особливості процедури отримання даних при дуже великих запитах. Використання стандартної процедури вибірки чергового

елемента, що задовольняє умові раніше ініційованого пошуку (наприклад, в системі MySQL спочатку один раз виконується команда mysqli_query(), а потім багато разів викликаються команди на кшталт mysqli_fetch_array()), при великій кількості елементів, що задовольняють умовам пошуку, може настільки загальмувати роботу з СУБД, що весь комп'ютер-сервер може перейти в заблокований стан [3].

Усі наведені моменти свідчать, що оптимізація процесу перебору є особливо актуальною при роботі з великими даними, які очевидно розміщуються в базах даних. В такому випадку перехід від повного до часткового (обмеженого) перебору є завданням, обов'язковим до виконання, звичайно за наявності відповідних можливостей (існування методів, їх реалізації і існування можливостей їх адекватного застосування).

Вся наведена аргументація стосувалася переборів у БД, однак і при розміщенні даних у пам'яті ЕОМ задача скорочення кількості елементів, що перебираються, може бути надзвичайно актуальною за наявності деяких особливостей всього процесу.

По-перше, самі елементи, що перебираються, в одних задачах можуть бути досить простими, як, наприклад, при задачі пошуку в числовому одновимірному масиві, але в інших випадках можуть являти собою складні, комплексні і значні за обсягом структури даних. У першому випадку навіть великі перебори інтегрально будуть задачами зі значним, але не гігантським обсягом обчислень. У другому ж загальна складність всього перебору значною мірою залежатиме від складності об'єктів, що перебираються. Таким чином, при переборі елементів, що належать до складно влаштованих структур даних, питання про обмеження перебору також стає дуже актуальним.

По-друге, хоча об'єкти, що перебираються, і можуть бути простими за своєю суттю (як наприклад, при переборі всіх елементів цілісного статичного масиву), проте умова, що перевіряється при переборі всіх елементів масиву, може бути обчислювально складною. Очевидно, що і тут актуальне (по можливості) скорочення множини елементів, що реально перебираються. Також, якщо перебір виконується не з метою пошуку, а застосування якоїсь операції до всіх без винятку елементів базової множини (у вказаному щойно випадку-прикладі – до всіх елементів масиву), то за умови, що ця операція є досить «важкою», обмеження перебору також повинно бути обов'язковим етапом всього процесу. Слід зазначити, що під «обмеженням перебору» тут потрібно розуміти не зменшення кількості об'єктів, що перебираються, але скорочення, по можливості, числа виконуваних при цьому операцій.

Наприклад, порахувавши і використавши результат тривалого перетворення f(ai) для поточного елемента масиву ai слід не видаляти його, а помістити в тимчасове сховище {f(ai)}. Якщо один із наступних елементів масиву випадково виявиться рівним ai, то вираховувати f(ai) не потрібно, а слід просто взяти готове значення з тимчасового сховища. Очевидно, що можливі інші шляхи скорочення обсягу виконуваних операцій, які в цілому можна віднести до алгоритмів обмеженого перебору, але які конкретно кроки в даному напрямку можна зробити - залежить від прикладного завдання.

Нарешті, можлива комбінація негативних факторів, коли:

- перебір необхідно здійснити на безлічі дуже складних за своєю структурою об'єктів;
- потрібна складна перевірка умови перебору
- сама операція, що застосовується до об'єктів, що відбираються при переборі, також дуже складна.

Слід пам'ятати, що складність тут скрізь йдеться у обчислювальному сенсі, тобто «складний»=«такий, що вимагає великих обсягів обчислень».

Досі бралися до уваги завдання, що відносяться великою мірою до галузі Big Data, де обсяги оброблюваних даних обчислюються мінімум гігабайтами, а часто

навіть терабайтами і більше. При цьому одне таке завдання, будучи запущеним на певній ЕОМ (або кластерній архітектурі), вирішується протягом тривалого
часу. Такі завдання можуть стосуватися фундаментальних наукових досліджень, важливих інноваційних технічних розрахунків, макроекономічних досліджень, тощо. Одним словом, вони є глобальними за своєю суттю (стратегічні). У той самий час існує дуже багато прикладних завдань середньої складності, вирішення яких потрібно отримувати майже миттєво, наприклад, прогнозування біржових коливань курсів валют і цінних паперів, прогноз погоди на найближчі години, розпізнавання особи людини, яку видно в даний момент через камеру зовнішнього спостереження, тощо. У цьому випадку задача хоча і не вимагає гігантських обсягів обчислень, проте існує жорсткий ліміт на загальний час її виконання, і крім того, отримувати рішення різних варіантів такої задачі потрібно досить часто, тобто. практично у потоці. Очевидно, що виграш у 0,1 сек при розпізнаванні однієї особи є несуттєвим, однак, якщо система повинна розпізнавати потік людей, що йдуть по жвавій вулиці, ситуація кардинально змінюється і навіть таке мале поліпшення в результаті призводить до сумарної відчутної економії та збільшення продуктивності системи.

Таким чином, обмеження перебору слід застосовувати у таких випадках:

- пошук/обробка інформації у джерелах Big Data, таких, як, наприклад, мережеві БД із загальним обсягом порядку терабайтів і вище;
- перебираються складні за своєю структурою об'єкти;
- здійснюється перевірка складної умови перебору;
- виконується складне перетворення елементів при переборі;
- сам собою перебір не дуже складний і може бути виконаний за цілком розумний час (наприклад, порядку секунд або хвилин), але у відповідній галузі такі завдання перебору виникають безперервно і постійно (потоком).

Як видно, спектр завдань, в яких потрібно застосування алгоритмів обмеження перебору, досить широкий, тому можна переходити до більш щільного обговорення цих методів.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ ТА ПОСТАНОВКА ЗАДАЧІ

Насамперед слід зазначити, що сама галузь, пов'язана з алгоритмами взагалі та алгоритмами перебору зокрема, з'явилася порівняно недавно. Так, основи теорії алгоритмів почали розроблятися у 1930-х роках (перші роботи були створені в 1930-х і 1940-х роках Куртом Геделем, Аланом Тюрінгом, Емілем Постом, Алонсо Черчем та А.А. Марковим) [4]. У той самий час (30-40-ті рр.) закладаються і основи теорії оптимізації на роботах фон Неймана, Данцига, Канторовича та інших [5].

Робота з множинами об'єктів поступово стала типовим процесом при обробці однорідних масивів інформації, хоча різні варіанти цього процесу, звичайно, можуть відрізнятися значною мірою особливостями технічної реалізації для різних прикладних завдань. Природно з'явилися й різноманітні способи переходу від одного об'єкта до іншого та перевірки їх якостей чи виконання деяких процесів. Узагальнено такі методи називають алгоритмами перебору [6]. Найпростішим є повний перебір чи перебір з допомогою грубої сили (brute-force).

Для проведення будь-якого перебору потрібно провести кілька підготовчих процедур [7], а саме:

а) встановити деякі правила встановлення порядку у безлічі елементів, що підлягають перебору (адже, як відомо, математичне
поняття «множина» є невпорядкованим набором об'єктів). У найпростішому, але дуже поширеному випадку, йдеться про просту нумерацію, тобто всі елементи потрібно лише пронумерувати, зокрема призначити перший і останній з них. У більш складних випадках порядок може встановлюватися на підставі інших ознак (наприклад, від більшого елемента до меншого, за кольорами веселки, частотою вживання на практиці тощо);

б) також необхідно створити спосіб переходу від поточного аналізованого при переборі
елемента до наступного, що слідує безпосередньо за ним;
в) розробити і реалізувати суть активної дії y = f(x), яку потрібно проводити над об'єктами xi, що перебираються (наприклад, формувати з них нову множину відібраних об'єктів, або просто видавати їх на екран, або застосовувати
до кожного якесь перетворення і перезаписувати його, і т.п.);
г) зафіксувати умову перебору P(x), тобто встановити ряд вимог, яким повинен задовольняти поточний об'єкт, що розглядається, щоб застосовувати до нього активну дію, описану в попередньому пункті.
Зазначена послідовність дій на псевдокоді, наближеному до C/C++, матиме такий вигляд:

```
x = First;
while (x! = Last)
{
if(P(x) == true)
{
y = f(x);
use(y);
}
x++;
}
```

Слід зазначити, що в доступній літературі зустрічається чимало варіантів алгоритмічних дій, які включають слово «перебір». У цьому можуть описувати досить далекі друг від друга предметні області.
Так, наприклад, термін «метод перебору» в [8] визначається як найпростіший з методів пошуку значень дійсно-значних функцій за будь-яким із критеріїв порівняння (на максимум, на мінімум, на певну константу). Стосовно екстремальних завдань перебір є прикладом прямого методу умовної одновимірної пасивної оптимізації. Даний метод також називають методом рівномірного пошуку або перебором по сітці.
Вочевидь, у реальних задачах перебору є можливість зменшити (причому часто – досить істотно) кількість виконуваних дій. Такі перебори ми називатимемо обмеженими, а сам процес зменшення числа дій, порівняно з повним перебором, називатимемо обмеженням перебору. При цьому можна розділяти два варіанти запровадження обмежень:
- зменшити кількість об'єктів, що перебираються, зберігаючи складність процедур P(x) і y = f(x), чого можна досягти в багатьох реальних застосунках, враховуючи їх особливості. Також для скорочення перебору в деяких випадках слід провести додаткові дії, що не входять до процесу повного перебору, однак їхня загальна складність має бути меншою, ніж складність перебору тих об'єктів, які можна відкинути в результаті проведення цих додаткових дій;
- спростити процедуру P(x) або y = f(x), або обидві відразу, зберігаючи множину об'єктів, що перебираються.
Очевидно, що можливі різні комбінації двох запропонованих шляхів скорочення кількості операцій під час перебору. Наприклад, можливе спрощення умови P(x) і, як наслідок, зменшення числа об'єктів, що перебираються. Або можливе розбиття процедури обробки y = f(x) на етапи, за результатом першого з яких можна припиняти подальшу обробку деяких елементів, фактично скорочуючи кількість елементів, що перебираються (якщо розуміти під перебором застосування всієї процедури y = f(x), а не тільки її частини).
Найбільш відомим методом, який обмежує перебір, є метод гілок та меж [9]. Слід зазначити, що конкретний зміст виконуваних у цьому методі дій залежить від предметної області і саме тому цей метод і навіть групу методів ще називають:
- перебір із відходом назад чи з поверненням;
- бектрекінг (англ. backtracking);
- пошук у глибину;
- метод спроб і помилок, і т.д.
Метод гілок та меж (англ. branch and bound) – загальний алгоритмічний метод для знаходження оптимальних рішень різних завдань оптимізації, особливо дискретної та комбінаторної. Є розвитком методу

повного перебору, але на відміну від останнього тут виконується відсів підмножин допустимих рішень, які завідомо не задовольняють умову P(x).
Метод гілок і меж вперше запропонований в 1960 Алісою Ленд і Елісон Дойг [10] для вирішення завдань цілочисельного програмування.
Загальна ідея методу може бути описана на прикладі пошуку мінімуму функції f(x) на множині допустимих значень змінної x. Функція f та змінна x можуть бути довільної
природи. Для методу гілок та меж необхідні дві процедури:
- розгалуження;
- знаходження оцінок (меж).
Метод використовується для вирішення деяких NP-повних завдань, у тому числі задачі комівояжера та задачі про ранець.
Метод гілок та меж використовує перевірку умови P(x) на основі інформації, яка обчислюється безпосередньо під час перебору, що погіршує його продуктивність. Теоретично можна відсікати невідповідні варіанти ще до початку процедури перебору для чого слід вираховувати якусь додаткову інформацію, яка безпосередньо і не потрібна для початкової задачі самого перебору. У даній статті розвиватиметься саме такий підхід, коли скорочення операцій при самому переборі досягається з допомогою виконання надлишкових додаткових операцій перед стартом процедури перебору. Це дозволяє розраховувати цю додаткову вхідну інформацію заздалегідь, а, можливо, і на іншій ЕОМ. При цьому сам перебір займатиме відчутно менший час порівняно з брут-форс варіантом.

ОСНОВНА ЧАСТИНА

Отже, розглянемо докладніше поняття алгоритму обмеженого перебору.

Як встановлено, важливими складовими будь-якого алгоритму перебору є дві речі:

- активне перетворення y = f(x), якому потрібно піддати об'єкти, що перебираються;

- умова перебору P(x), підставляючи в яку черговий елемент, що перебирається, встановлюється, чи потрібно піддавати його перетворенню y = f(x).

На перший погляд може здатися, що можливо працювати у цих двох напрямках: скорочувати час активних перетворень та скорочувати час перевірки умов. Однак, спроби декомпозиції та спрощення перетворення y = f(x) будуть безперспективними, тому що при вирішенні конкретної прикладної задачі тип цього перетворення і всі його складові вже задані особливостями предметної області і якось змінити обсяг операцій, які необхідно виконати для досягнення потрібного результату неможливо.

Що ж до другого напряму – то тут перспектива зменшення обсягів виконуваних операцій може бути набагато кращою. По-перше, в більшості реальних завдань умова перебору буде досить складною (складеною), а це означає, що її можна представити в одній із нормальних форм, що розглядаються в математичній логіці:

а) кон'юнктивна нормальна форма (КНФ):

$$P(x) = P_1(x) \wedge P_2(x) \wedge \ldots \wedge P_m(x) \qquad (1)$$

б) диз'юнктивна нормальна форма (ДНФ):

$$P(x) = P_1(x) \vee P_2(x) \vee \ldots \vee P_m(x) \qquad (2)$$

Позитивними рисами (1) та (2) є можливості скорочених логічних обчислень у деяких випадках (які не так і рідко зустрічаються на практиці):

а) для КНФ: якщо будь-яка складова всієї умови (тобто підумова) Pi(x) дорівнює нулю, то інші складові можна не обчислювати, а результат відразу дорівнює нулю (це автоматично випливає з властивостей операції логічної кон'юнкції, тобто операції «І»). Звідси випливає, що обчислювати умови в КНФ завжди краще з тих, які з найбільшою ймовірністю є хибними (тобто рівні нулю);

б) аналогічно якщо в ДНФ якась складова Pi(x) дорівнює одиниці, то інші підумови можна не прораховувати, так як весь результат автоматично дорівнює 1, що випливає із властивостей логічної операції

«АБО» (диз'юнкції). Отже, починати обчислення підумов у ДНФ слід із тих, які найімовірніше є істинними (тобто рівні одиниці).

Якщо логічна функція, що реалізується виразом P(x), має вигляд, який відрізняється від (1) або (2), то, як відомо, існують процедури для приведення будь-якої логічної функції до КНФ або ДНФ. Для визначеності (і не обмежуючи при цьому загальності) надалі (і зокрема у програмній реалізації) говоритимемо про КНФ, тобто. про набір логічних умов, пов'язаних між собою зв'язкою «І». За потреби від КНФ можна перейти до ДНФ за допомогою законів де Моргана та інших законів логіки алгебри.

Отже, якщо прикладна задача дозволяє розбивку умови P(x) на простіші умови, то виникають можливості щодо скорочення перебору, пов'язані із застосуванням кон'юнктивної (або диз'юнктивної) нормальної форми. Зазначимо, що спочатку (наприклад, для розв'язання задачі повного перебору) ніякої розбивки умови перебору на підумови може
бути не потрібно. Цілком можлива ситуація, коли розбивка умови P(x) на компоненти є штучною, і спеціально вводиться з метою скорочення перебору. Мало того, деякі з цих умов Pi(x) можуть вибиратися так, щоб, по-перше, з великою ймовірністю бути рівними нулю (у разі використання КНФ) або одиниці (якщо за основу взята ДНФ), а, по-друге, включати якусь обчислювально складну інформацію, яку можна розрахувати попередньо, перед стартом процесу перебору.

В останньому випадку виникають нові додаткові обчислення, які (цілком або частково) були відсутні при повному переборі, що на перший погляд навіть погіршує ситуацію. Однак, введення такої додаткової інформації може суттєво скоротити процедуру перебору, тому загальний час вирішення прикладного завдання знизиться значною мірою порівняно з повним перебором. В якості додаткової інформації пропонується використовувати розбиття на класи (тобто кластеризацію) самих об'єктів, що перебираються, за будь-якими відповідними властивостями, або пов'язаних з ними інших, допоміжних об'єктів, важливих для конкретної задачі перебору.

Кращою ілюстрацією запропонованого підходу, звісно, є конкретні приклади, реалізовані мовами програмування, і дають точний числовий результат часу виконання однієї й тієї прикладної задачі різними методами перебору. Проте, для того, щоб таку програму створити, потрібно поставити якісь конкретні задачі, так як описані вище міркування є досить загальними і повинні адаптуватися (підлаштовуватись) під кожну конкретну задачу своїм власним чином.

Отже, для побудови програмної реалізації (що буде виконано в даній роботі) розглянемо одне з поширених завдань комп'ютерної графіки [11], коли з безлічі точок, заданих своїми координатами, потрібно вибрати всі такі набори по n штук, які утворюють правильний n-кутник (полігон). Точки можуть розташовуватися на площині (тоді кожна задається двома числами-координатами) або в тривимірному просторі (тоді положення точки природно задається трійкою дійсних чисел), що істотно складніше з обчислювальної точки зору. Саме цей останній випадок і будемо використовувати (беремо важче завдання, щоб легше було оцінити різницю, одержувану із застосуванням запропонованого методу).

Для побудови першої версії програмного продукту обмежимося випадком n = 3, тобто. з множини N точок, заданих у тривимірному просторі за допомогою їх координат, вибиратимемо трійки тих, які утворюють правильні трикутники.

Умова перебору P(x) може бути у такому разі сформульовано різними способами:

- якщо два кути трикутника дорівнюють 60;
- якщо трикутник рівнобедрений і один (будь-який) з його кутів дорівнює 60;
- якщо всі три сторони рівні, і т.д.

Для програмної реалізації виберемо останню умову, тобто. рівність всіх трьох сторін (позначимо їх d1, d2, d3), що у програмі можна перевірити трьома попарними порівняннями, у результаті умова правильності трикутника природним

чином перетворюється на КНФ з трьох складових підумов наступного виду:

$$P(x) = P_1(x) \wedge P_2(x) \wedge P_3(x), \quad (3)$$

де x - трикутник, що задається, взагалі кажучи (як було зазначено вище), своїми трьома вершинами. Будемо називати їх Ai, Aj, Ak, тобто:

x = {Ai, Aj, Ak};
P(x) = "x - рівносторонній трикутник";
P1(x) = "сторони d1 і d2 рівні";
P2(x) = "сторони d2 і d3 рівні";
P3(x) = "сторони d3 і d1 рівні";
d1 = | AiAj |;
d2 = | AjAk |;
d3 = | AkAi |.

Рівності сторін, що перевіряються як Pi(x) у виразі (3) не можуть у комп'ютері виконуватися точно, як це відбувається у викладках в аналітичній математиці, адже відстані di будуть дробовими числами. Тому вже на даному етапі, при описі алгоритму, слід запровадити якийсь допуск □, і якщо різниця між двома числами (відстанями) буде меншою, ніж це значення, то два числа (відстань) можна вважати рівними:

$$d1 = d2 \quad \Leftrightarrow \quad | d1 - d2 | < \varepsilon \quad (4)$$

Наявність такого допуску викликає необхідність саме триразової перевірки рівності всіх пар сторін, адже якби довжини сторін задавалися точно, то було б достатньо лише два порівняння з трьох використовуваних.

Беручи до уваги введені позначення, можна сформулювати досить простий словесний опис

процедури повного перебору у задачі, що розглядається:

1) виконати перебір всіх можливих комбінацій трійок точок (природно без урахування їх порядку, тобто йдеться про поєднання, а не розміщення, перестановку та ін.);

2) для кожного випадку (тобто для кожної трійки точок) порахувати відстані між ними d1, d2, d3;

3) перевірити рівність (за формулою (4)) всіх трьох відстаней d1, d2, d3 та у разі збігу – видати ці точки (наприклад, їх номери) на екран.

На С-подібному псевдокоді ця процедура матиме вигляд:

```
for(int i=0;i<N;i++)
for(int j=i+1;j<N;j++)
for(int k=j+1;k<N;k++){
Calculate_d1(A[i], A[j]);
Calculate_d2(A[j], A[k]);
Calculate_d3(A[k], A[i]);
if (d1==d2==d3){
out(i,j,k);
  }
}
```

Слід зазначити, що цей код вже певною мірою оптимізований (тобто. перебір обмежений), оскільки перебір тут іде саме з усіх поєднань, тобто. без повторень, що досягається завданням початкових значень для змінних двох внутрішніх циклів не з нуля, а з поточного значення змінної зовнішнього циклу. Наприклад, у другому циклі змінна циклу ініціалізується значенням j = i + 1, а у внутрішньому циклі початкове значення відповідно дорівнює k = j + 1. Зазначимо, що тіло внутрішнього циклу буде виконано раз, і, наприклад, для 1000 точок це число становитиме 166 млн. повторень.

Перейдемо до опису скороченого перебору, який буде реалізований у цьому прикладі. Основна ідея тут полягає в тому, що порівнюються всі відстані поспіль, які навіть дуже різняться одна з одною. Доцільно перед початком перебору розділити на групи (кластеризувати) всі відстані між точками (запам'ятовуючи, до якого саме кластеру належить та чи інша пара точок з допомогою їх номерів). Тоді після такої попередньої підготовки (яка не потрібна при звичайному повному переборі) здійснювати пошук однакових відстаней d1, d2, d3 можна буде вже не по всьому набору точок, а тільки перебираючи трійки точок, між якими відстані лежать в межах одного кластера.

Загальна кількість відстаней між N точками дорівнює числу поєднань . Припустимо, всі відстані більш-менш рівномірно розподілилися по K кластерам, тоді в кожному з них буде по / K відстаней. З цього числа альтернатив (різних відстаней, що лежать у межах одного кластера) можна

було б вибирати по 3, перевіряючи їх на точну рівність за допомогою константи □, але кількість операцій можна значно скоротити, якщо скористатися скороченою процедурою обчислення КНФ і перевіряти рівність лише 2 відстаней. У більшості випадків вони будуть різні і подальше обчислення третьої відстані і перевірка на збіг з першими двома не знадобиться, тобто кількість таких перевірок буде дорівнює кількості поєднань з C_N^2 / K по 2, а не по 3:

$$C^2_{C_N^2/K} = \frac{C_N^2/K \cdot (C_N^2/K - 1)}{2} = \frac{C_N^2 \cdot (C_N^2 - K)}{2K^2}$$

Наприклад, для 1000 точок та 50 кластерів отримаємо таку кількість повторень процедури порівняння трьох відстаней:

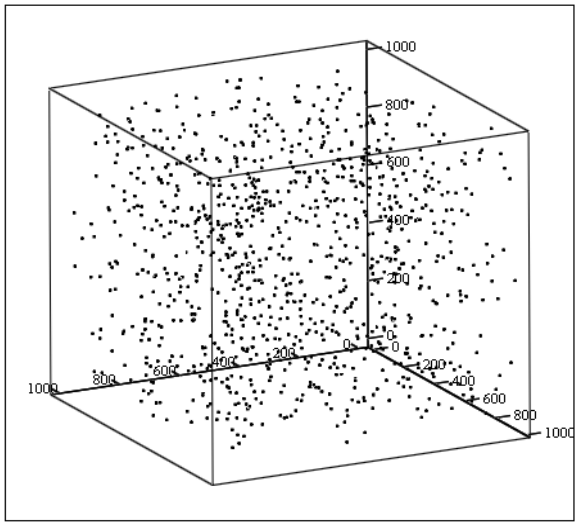
$$C^2_{C^2_{1000}/50} = \frac{499500 \cdot (499500 - 50)}{2 \cdot 50^2} = 4{,}99 \cdot 10^7.$$

Як бачимо, кількість операцій скорочується приблизно в 3 рази (166 млн проти 50 млн порівнянь), хоча точною цю оцінку не можна вважати через наявність додаткових накладних витрат у другому методі. Таким чином, попередня теоретична оцінка показує гарне поліпшення продуктивності перебору в задачі, що розглядається при використанні запропонованого методу, в порівнянні з традиційним рішенням. Очевидно, що точні показники покращення часу виконання завдання потрібно отримувати на робочій програмній реалізації.

Реалізацію алгоритму перебору було виконано мовою C# у середовищі розробки Microsoft Visual Studio 2022 Community. Ефективність запропонованого рішення оцінювалася на модельному розподілі тисячі точок у тривимірному просторі, візуалізованому на рис.1. Скріншоти, на яких видно порівняльні результати роботи програми в режимах повного та обмеженого перебору, показані на рис. 2 (для допуску 1) та рис. 3 (для допуску 0,1).

На рис. 2 видно, що при допуску 1 повний перебір 1000 точок займає 119 секунд проти 65 секунд при використанні запропонованого алгоритму на 60 кластерах (при цьому знайдено близько 700 правильних трикутників).

На рис. 3 видно, що при допуску 0,1 повний перебір 1000 пікселів займає 109 секунд проти 67 секунд при використанні запропонованого алгоритму на 60 кластерах.



$(R^{\langle 0\rangle}, R^{\langle 1\rangle}, R^{\langle 2\rangle})$

Рис. 1. Приклад генерації набору вхідних даних для проектованої програми за допомогою середовища MathCad

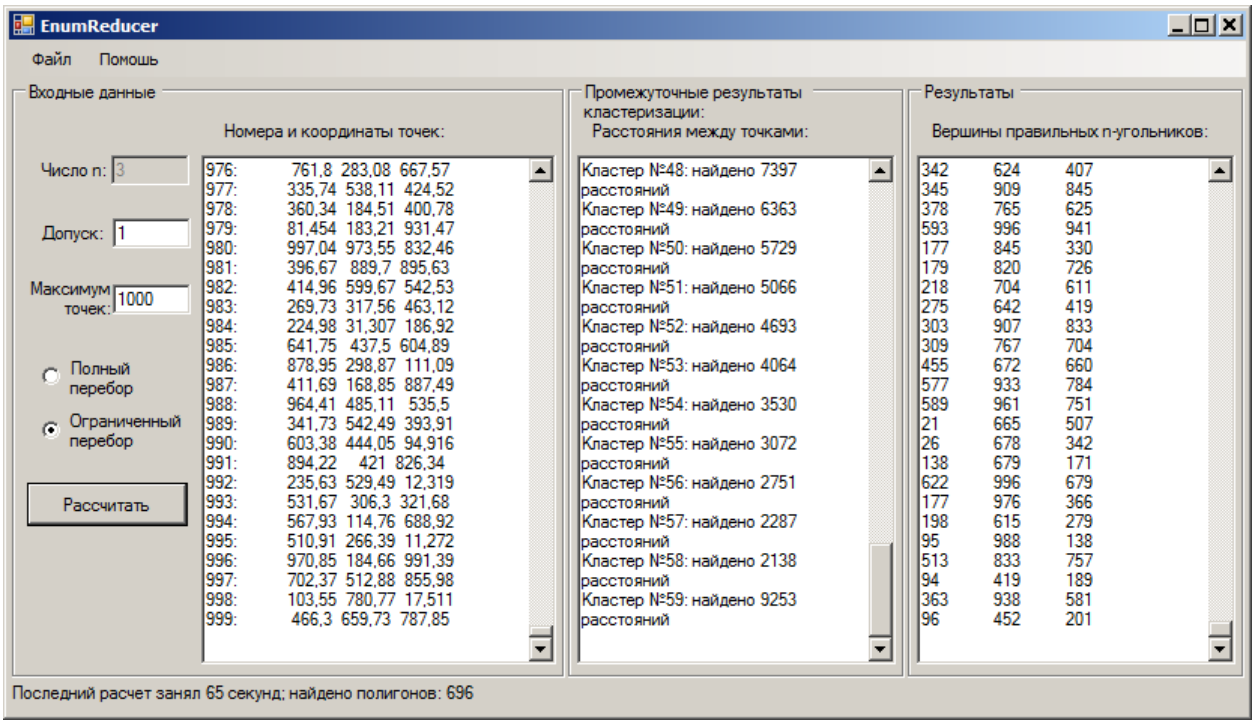


Рис. 2. Результати пошуку правильних полігонів при допуску 1 за допомогою запропонованого алгоритму при K = 60

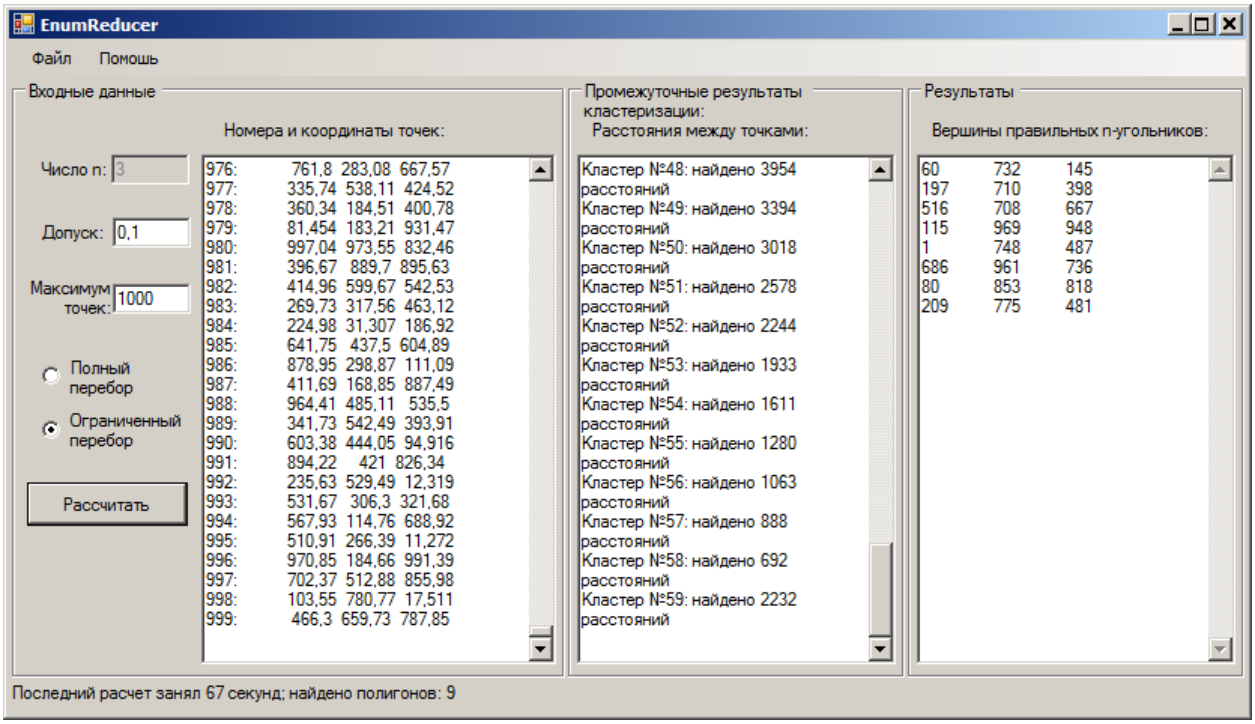


Рис. 3. Результати пошуку правильних полігонів при допуску 0,1 за допомогою запропонованого алгоритму при K = 60.

Слід зазначити, що при використанні запропонованого алгоритму виникають втрати невеликої кількості полігонів за рахунок того, що різні сторони їх потрапляють у різні кластери (а значить такі полігони не можуть бути знайдені за допомогою запропонованого алгоритму). Це рідкісна ситуація, що має низьку

ймовірність, яка зростає зі збільшенням кількості кластерів.

Проведено дослідження прискорення перебору зі збільшенням кількості кластерів K, і навіть збільшення відсотка втрачених полігонів – рис. 4 (для допуску 1) та рис. 5 (для допуску 0,1).

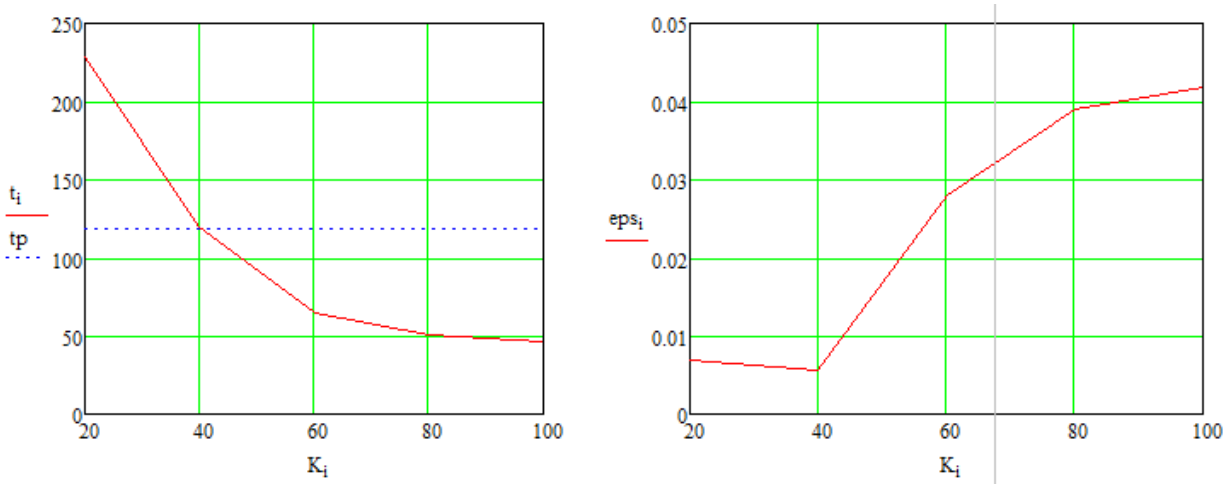


Рис. 4. Графіки зміни часу (в сек) виконання розрахунку 1000 точок при допуску 1 в залежності від числа кластерів (горизонтальним пунктиром показаний результат повного перебору) - а, а також відсоток втрачених полігонів - б.

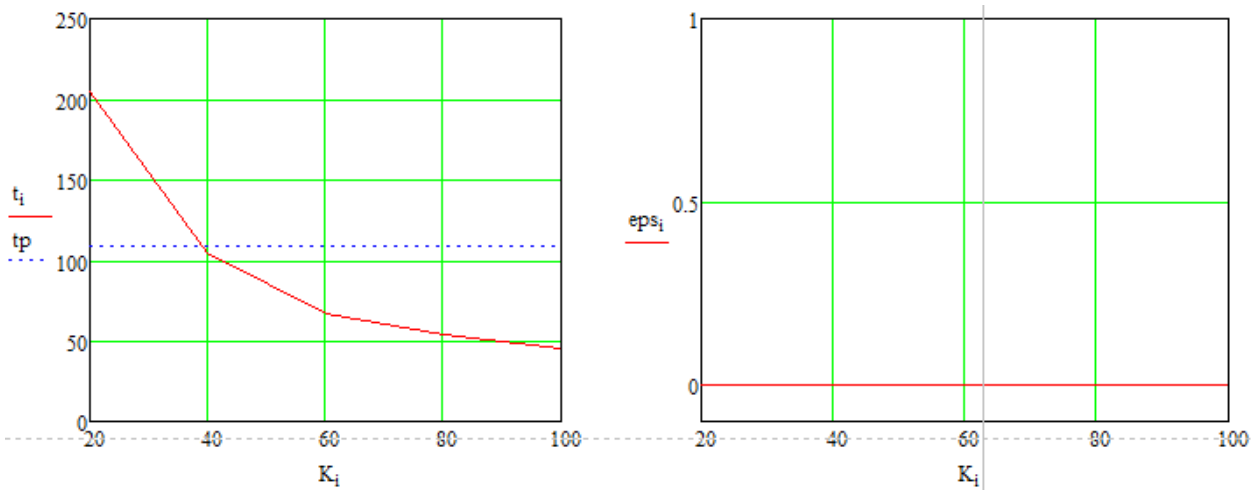


Рис. 5. Графіки зміни часу (в сек) виконання розрахунку 1000 точок при допуску 0,1 в залежності від числа кластерів (горизонтальним пунктиром показаний результат повного перебору) - а, а також відсоток втрачених полігонів - б.

Бачимо, що при малих кількостях відібраних об'єктів помилка практично не виникає через її вкрай малу ймовірність. Найбільш оптимальним є значення K в районі 60, яке дозволяє отримати майже дворазове прискорення роботи при рівні помилок менше 2%, що є цілком допустимим у завданнях комп'ютерної графіки, де втрата невеликої частини даних не погіршує якості результуючого матеріалу.

ВИСНОВКИ

Таким чином, розглянуто задачу прискорення перебору об'єктів, яка часто виникає при вирішенні прикладних завдань обробки інформації. Спочатку проведено формалізацію опису процесу перебору, а також деталізовано пов'язані з ним математичні поняття. Далі на основі аналітичного огляду існуючих методів перебору та розгляду можливостей щодо їх прискорення зроблено висновки про перспективи та необхідність власної розробки в даному напрямку. Відповідно, запропоновано удосконалений алгоритм перебору, заснований на декомпозиції умови перебору в кон'юнктивну або диз'юнктивну нормальну форму, а також з урахуванням кластеризації вхідної інформації (або кластеризації додаткової інформації, що вираховується на основі вхідної), що відноситься до однієї (або кількох) умов перебору. Ефективність запропонованого методу досліджена на прикладі поширеної задачі з галузі комп'ютерної графіки, яка полягає у пошуку правильних n-кутників (полігонів) на множині точок тривимірного простору, заданих трійками своїх координат. Попередній теоретичний розрахунок дозволив оцінити прискорення виконання завдання під час використання запропонованого алгоритму на рівні 2-3 рази, проте ці результати потрібно було перевірити ще й практично шляхом дослідження конкретної програмної реалізації.

Тестування програмного продукту показало реальне прискорення процесу перебору під час використання запропонованого методу, що становить кілька разів (у 2-3 разу при використанні до 100 кластерів). У той же час було встановлено, що за рахунок попадання різних частин одного й того ж підпадаючого під умову перебору об'єкта у різні кластери, насправді може бути втрачена невелика частка необхідних об'єктів. Відсоток втрат залежить від кількості кластерів розбиття (як і величина прискорення) і навіть для великого їх числа дає втрати не більше 2 %. Такі втрати є цілком допустимими у завданнях комп'ютерної графіки, тому загалом застосування цього методу для окремих прикладних задач є цілком обґрунтованим.

REFERENCES

1. **Coronel C., Morris S.** (2023). Database Systems: Design, Implementation, and Management, 14th ed. Boston, Cengage Learning, 784.
2. **Tamb1 V.K., Singh N.** (2024). A Comparison of SQL and NO-SQL Database Management Systems for Unstructured Data. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE), Vol. 13, Issue 7, 2086-2093.
3. **Glake, D., Kiehn, F., Schmidt, M., Panse, F., Ritter, N. Towards** (2022). Polyglot Data Stores: Overview and Open Research Questions. URL: https://arxiv.org/abs/2204.05779.
4. **Booch G., Maksimchuk R.A., Michael W.E., Young B.J., Conallen J., Houston K.A.** (2007). Object-Oriented Analysis and Design with Applications. Boston: Addison-Wesley Professional, 3 edition, 720.
5. **Konc J., Janežič D.** (2017). An Improved Branch-and-Bound Algorithm for the Maximum Clique Problem. MATCH Communications in Mathematical and in Computer Chemistry, Vol. 78, No. 3, 569–590.
6. **Burkard R.E., Çela E., Karisch S.E., Rendl F.** (2019). A Branch-and-Bound Algorithm for the Quadratic Assignment Problem Based on the Hungarian Method. European Journal of Operational Research, Vol.112, No.3, 647–662.
7. **Gonçalves J.F., Mendes J.J.M., Resende M.G.C.** (2005). A hybrid genetic algorithm for the job shop scheduling problem. European Journal of Operational Research, Vol.167, No.1. 77–95.
8. **González M.A., Vela C.R., Varela R.** (2008). A new hybrid genetic algorithm for the job shop scheduling problem with setup times. Proceedings of the 18th International Conference

on Automated Planning and Scheduling (ICAPS 2008), 116–123.
9. **Akenine-Möller J.T., Haines E., Hoffman N., Pesce A., Iwanicki M., Hillaire S.** (2018). Real-Time Rendering. Boca Raton: CRC Press, 1198.
10. **Albahari B.A.** (2017). C# 7.0 in a Nutshell: The Definitive Reference. Sebastopol: O'Reilly Media, 1088.
11. **Brett D. McLaughlin, Pollice G., West D.** (2006). Head First Object-Oriented Analysis and Design. Sebastopol: O'Reilly Media, 636.
12. **Nagel Ch**. (2018). Professional C# 7 and .NET Core 2.0. Birmigham: Wrox, 1440.
13. **Mirjalili, S., Mirjalili, S. M., Hatamlou, A.** (2020). A Survey on Swarm Intelligence Algorithms: Principles, Applications, and Challenges IEEE Access, Vol. 8, 164848–164872.

On approaches to accelerate object search in information processing tasks

Oleksandr Haisha, Serhii Lienkov, Olena Haisha

Abstract. The paper provides a detailed examination of the concept of the enumeration process. It is performed with objects x, which are elements of a certain set X. The purpose of enumeration is to execute a certain active operation y = f(x) on the elements being processed. This operation can, of course, be applied to every element, but in the vast majority of practical tasks, the operation is applied only to those elements that satisfy a specific enumeration condition P(x).

The simplest method of enumeration is exhaustive search or "brute-force". In this approach, every single element of the set X is checked against the condition P(x), and depending on the result, the operation y = f(x) is executed. It is evident that exhaustive search serves as a benchmark process in the sense that it is convenient to evaluate the efficiency of various fast enumeration algorithms in comparison to it. Fast (or restricted) enumeration algorithms are advisable to use in most practical cases, provided that exhaustive search does not take an insignificant amount of time, such as fractions of a second. However, in some cases, it is reasonable to optimize (limit) the enumeration process even when execution times are very short, particularly if such short enumeration processes are repeated multiple times per unit of time, ultimately accumulating into a significantly large total runtime.

The branch and bound method is the most well-known algorithm for restricted enumeration, allowing the process to progressively advance from one variant to another while discarding entire sets of elements that are known in advance not to satisfy the condition P(x). However, this method does not allow for reliable preliminary estimates of the efficiency of traversing the entire tree of corresponding elements, and its applicability to a particular problem only becomes evident during the enumeration process.

This paper proposes a restriction method based on representing the enumeration condition P(x) in conjunctive or disjunctive normal form, followed by verifying this condition using a scheme of reduced logical computations that utilize only one of the conditions. Additionally, it is proposed to precompute certain supplementary information that allows obtaining the final result more quickly during enumeration. Such information may include the classification (clustering) of all objects being enumerated and performing the enumeration only within clusters, thereby reducing the total number of elements to be processed.

The proposed approach is explained in detail through an example of solving a relevant problem in the field of computer graphics, specifically the search for all regular n-angles (demonstrated using equilateral triangles) within a set of N points in three-dimensional space, where each point is defined by its coordinates. A verbal description of the exhaustive search necessary for solving this problem is provided, along with a detailed explanation of the restricted enumeration algorithm constructed according to the proposed scheme (reduced computations using conjunctive normal form and preliminary clustering of all objects being enumerated, or more precisely, clustering of the associated information – distances between all pairs of points).

The paper presents numerical calculations that allow for a theoretical assessment of the efficiency of the proposed solution. It has been established that, under the assumption of a perfectly uniform distribution of all enumerated objects across K classes, a significant reduction in the number of distance comparison operations occurs when searching for equilateral triangles. Specifically, for K = 50, the theoretical speedup reaches up to three times. This acceleration has been practically examined using a model software implementation written in C#, demonstrating that the actual speedup is quite close to the theoretical maximum, confirming that the proposed approach is indeed effective.

Keywords: enumeration, enumeration acceleration, clustering, decomposition of enumeration conditions.

Методика шифрування даних на основі технології ідентифікації блоків матричним способом

Олександр Левкуша[1], Юрій Пепа[2]

[1,2] Державний університет інформаційно-комунікаційних технологій
вул. Солом'янська, 7, м. Київ, Україна, 03037
[1] levkusha_alexandr@ukr.net, https://orcid.org/0009-0001-6064-0822
[2]yurka14@ukr.net, https://orcid.org/0000-0003-2073-1364

Received 03.02.2025, accepted 29.03.2025
https://doi.org/10.32347/st.2025.3.1208

Анотація. У статті запропоновано матричний метод ідентифікації блоків даних на серверних платформах, які працюють із великими масивами інформації. Використання лінійної алгебри, зокрема сингулярного розкладу (SVD), дозволяє ефективно аналізувати структуру даних, зменшувати надлишковість та оптимізувати процес кластеризації. Запропонований метод включає попередню нормалізацію даних, обчислення косинусної міри подібності та адаптивний алгоритм обробки, що забезпечує високий рівень точності ідентифікації навіть при модифікації вхідних блоків. Досліджено обчислювальну складність алгоритму, показано переваги використання методів зменшення розмірності та оптимізації ресурсів. Проведені експериментальні дослідження довели ефективність методу в умовах високої варіативності даних, що дозволяє його застосовувати у кібербезпеці, хмарних обчисленнях і розподілених системах. Запропонований підхід зменшує час обробки запитів та підвищує точність ідентифікації, що критично важливо для моніторингу інформаційних потоків та забезпечення цілісності даних. Отримані результати підтверджують доцільність впровадження цього методу в інформаційних системах для підвищення продуктивності обробки великих даних.

Ключові слова: матричний метод, сингулярний розклад, ідентифікація блоків даних, оптимізація обчислень, кластеризація, великі дані, кібербезпека.

Олександр Левкуша
аспірант кафедри технічних систем кіберзахисту

Юрій Пепа
к.т.н., доцент
професор кафедри технічних систем кіберзахисту

ВСТУП

Розвиток інформаційних технологій супроводжується постійним зростанням обсягів даних, що зберігаються та обробляються на серверних платформах.

Одним із важливих аспектів управління інформаційними потоками є процес ідентифікації блоків даних, який дозволяє визначати їх приналежність, структуру та взаємозв'язки між окремими фрагментами інформації. Оскільки дані на сервері можуть змінюватися, оновлюватися та дублюватися, їх ефективна ідентифікація стає критичною задачею, від вирішення якої залежить швидкість обробки запитів, рівень навантаження на обчислювальні ресурси та загальна продуктивність системи [1 - 4].

Одним із перспективних підходів до вирішення цієї проблеми є використання матричних методів, які дозволяють

здійснювати ідентифікацію блоків даних на основі математичних перетворень та алгоритмів лінійної алгебри. Застосування матричних операцій у цьому контексті сприяє оптимізації процесів обробки інформації, підвищенню точності та швидкості аналізу взаємозв'язків між даними, а також зменшенню обчислювальних витрат [5 - 8].

У представленому дослідженні розглядається розробка та впровадження матричного підходу до ідентифікації блоків даних, що дозволяє систематизувати та автоматизувати процес визначення інформаційних об'єктів у великих масивах даних. Особлива увага приділяється математичному моделюванню, розробці алгоритму, аналізу його обчислювальної складності, а також практичній оцінці ефективності на тестових вибірках даних [3, 9].

АКТУАЛЬНІСТЬ ДОСЛІДЖЕННЯ

Сучасні інформаційні системи стикаються з низкою проблем, пов'язаних із зберіганням, пошуком та ідентифікацією даних. В умовах швидкого зростання обсягів інформації традиційні методи обробки даних, такі як хешування, кластеризація чи сигнатурні аналізи, демонструють ряд обмежень. Зокрема, ці методи часто вимагають значних обчислювальних ресурсів, особливо при роботі з великими базами даних, що містять мільйони записів [9 - 11]. Проблема також полягає у збереженні цілісності та ідентичності даних у розподілених системах, де блоки інформації можуть дублюватися, змінюватися або зникати в процесі передачі між серверами. У таких умовах необхідний ефективний метод, який дозволить швидко та точно визначати унікальні блоки даних, мінімізуючи помилки ідентифікації та виключаючи дублювання [7, 12, 13]. Матричний метод є перспективним напрямком у цій сфері, оскільки він дозволяє формалізувати процес ідентифікації шляхом представлення вхідних даних у вигляді матричних структур, що піддаються обчислювальним перетворенням. Використання операцій лінійної алгебри дозволяє значно підвищити точність обробки інформації, а також оптимізувати ресурси системи, забезпечуючи баланс між швидкістю виконання запитів та споживанням обчислювальної потужності [7, 14, 15].

Впровадження матричного підходу до ідентифікації блоків даних може знайти широке застосування у сфері управління базами даних, хмарних обчислень, захисту інформації та розподілених системах, що підтверджує його актуальність у контексті сучасних тенденцій розвитку інформаційних технологій [9].

МЕТА ДОСЛІДЖЕННЯ

Основна мета дослідження полягає у розробці ефективного матричного методу для ідентифікації блоків даних на сервері, що забезпечить швидке та точне визначення взаємозв'язків між інформаційними фрагментами. Дослідження передбачає побудову математичної моделі, яка дозволить формалізувати процес ідентифікації, а також розробку алгоритму, що зможе адаптуватися до змін у структурі даних та забезпечити мінімальне споживання обчислювальних ресурсів [7, 8, 12, 16]. Запропонований підхід передбачає аналіз існуючих методів, визначення їхніх переваг та недоліків, формулювання математичних принципів роботи алгоритму, а також його реалізацію у вигляді програмного модуля. Крім того, метою є оцінка ефективності розробленого алгоритму шляхом експериментального тестування та порівняння його продуктивності з іншими методами ідентифікації блоків даних [4, 9, 13, 17].

ЗАВДАННЯ ДОСЛІДЖЕННЯ

Для досягнення поставленої мети необхідно вирішити ряд ключових завдань. Перше завдання полягає в аналізі сучасних підходів до ідентифікації блоків даних, що використовуються в інформаційних системах, та визначенні їхніх основних обмежень. Це дозволить сформулювати

вимоги до нового методу та обґрунтувати необхідність застосування матричного підходу [4, 18].

Наступним завданням є розробка математичної моделі, що описує процес ідентифікації блоків даних у вигляді системи матричних рівнянь. Важливо визначити параметри, які впливають на точність та швидкість роботи алгоритму, а також обґрунтувати вибір методів матричних перетворень для реалізації процесу ідентифікації [8]. Окреме завдання дослідження пов’язане з розробкою алгоритму ідентифікації, який базуватиметься на застосуванні матричних операцій та дозволить автоматизувати процес аналізу інформаційних фрагментів. У межах цього етапу необхідно побудувати логічну структуру алгоритму, визначити його обчислювальну складність та оцінити ресурси, необхідні для його виконання [7]. Подальшим етапом є програмна реалізація розробленого алгоритму та його тестування на реальних або змодельованих наборах даних. Це дозволить отримати емпіричні результати щодо точності та продуктивності запропонованого методу [9]. Завершальним завданням є порівняльний аналіз результатів, отриманих у процесі тестування, з іншими методами ідентифікації блоків даних. Важливо визначити переваги та недоліки запропонованого підходу, а також розробити рекомендації щодо його впровадження у серверні системи для підвищення ефективності роботи з великими обсягами даних [14].

У сфері ідентифікації блоків даних та застосування матричних методів існує низка наукових праць як українських, так і зарубіжних дослідників. В Україні значний внесок у розвиток методів представлення, збереження та аналізу даних зробили В.Ю. Щербань, С.М. Краснитський, Т.І. Астістова та В.М. Яхно. У своїй колективній монографії "Методи представлення, збереження та аналізу даних інформаційних систем" вони досліджують математичні моделі та чисельні методи, зокрема системи лінійних рівнянь, трансцендентні та алгебраїчні рівняння, диференціальні рівняння, а також методи інтерполяції та апроксимації. Ці методи є фундаментальними для розуміння та застосування матричних підходів в обробці даних [4]. О.І. Черняк у підручнику "Інтелектуальний аналіз даних" розглядає сучасні методи обробки великих обсягів інформації, акцентуючи увагу на алгоритмах класифікації, кластеризації та виявлення закономірностей у даних. Хоча основний акцент зроблено на інтелектуальному аналізі, використання матричних методів є невід'ємною частиною багатьох алгоритмів, представлених у цій праці [9].

У дисертації В.Г. Бабенка "Удосконалення методів побудови криптографічних примітивів на основі матричних операцій" досліджено застосування матричних операцій у криптографії. Автор пропонує нові підходи до синтезу криптографічних операцій, що базуються на матричних перетвореннях, з метою підвищення швидкості та стійкості шифрування. Ці дослідження демонструють потенціал матричних методів у забезпеченні безпеки даних та їх ідентифікації [16]. Серед зарубіжних дослідників варто відзначити Джина Голуба та Чарльза Ван Лоана, авторів фундаментальної праці "Matrix Computations".

У цій книзі детально розглядаються чисельні методи лінійної алгебри, включаючи розклади матриць, які є ключовими для багатьох алгоритмів обробки та ідентифікації даних [8].

Також заслуговує уваги робота Тревора Гасті, Роберта Тібширані та Джерома Фрідмана "The Elements of Statistical Learning", де обговорюються методи машинного навчання, багато з яких базуються на матричних операціях. Ці методи широко застосовуються для класифікації та кластеризації даних, що є важливими аспектами їх ідентифікації [14]. Як українські, так і зарубіжні вчені зробили значний внесок у розвиток матричних методів для ідентифікації та обробки блоків даних, пропонуючи різноманітні підходи та алгоритми для підвищення ефективності та точності цих процесів [7].

ТЕОРЕТИЧНІ ОСНОВИ ДОСЛІДЖЕННЯ

Постановка задачі

У сучасних інформаційних системах, зокрема серверних платформах, виникає необхідність швидкої та точної ідентифікації блоків даних. З огляду на зростання обсягів інформації, традиційні підходи, такі як хешування, деревоподібні структури та сигнатурні аналізи, демонструють певні обмеження у продуктивності та точності. Це обумовлено тим, що сучасні бази даних містять величезні масиви інформації, у яких часто виникає проблема дублювання, структурної неоднорідності та складності взаємозв'язків між окремими елементами [4, 19, 20].

Необхідність розробки нових методів ідентифікації блоків даних зумовлена високими вимогами до продуктивності обчислювальних систем. Використання традиційних алгоритмів пошуку і класифікації інформації не завжди є ефективним, оскільки їхня реалізація може спричиняти значні витрати обчислювальних ресурсів, особливо у випадках великих розподілених систем. У зв'язку з цим виникає потреба у створенні математично обґрунтованого підходу, що дозволить скоротити час ідентифікації та водночас забезпечити високу точність визначення приналежності даних [9].

Запропонована дослідницька задача полягає у розробці матричного методу ідентифікації блоків даних, який базуватиметься на використанні матричних перетворень для аналізу та впорядкування інформаційних структур. Основна ідея цього підходу полягає у поданні даних у вигляді матриць, елементи яких відображають властивості, зв'язки та унікальні характеристики блоків. Використання математичних операцій, таких як множення матриць, їх трансформація та нормалізація, дозволить ефективно виявляти закономірності у великих наборах даних і усувати дублювання [8].

Ключовими аспектами задачі є розробка алгоритмічного забезпечення для реалізації матричного методу, його тестування на реальних та змодельованих вибірках даних, а також порівняльний аналіз із традиційними методами ідентифікації. Важливим аспектом є визначення критеріїв ефективності запропонованого підходу, включаючи швидкість виконання операцій, обчислювальну складність алгоритму та рівень точності виявлення зв'язків між інформаційними блоками [7].

Розв'язання поставленої задачі має суттєве практичне значення для серверних систем, які працюють з великими обсягами даних. Впровадження матричних методів дозволить покращити продуктивність обчислювальних процесів, знизити навантаження на сервери та підвищити надійність інформаційних систем. Очікується, що результати дослідження знайдуть застосування у сфері управління базами даних, розподілених обчислень, аналізу великих даних та системах інформаційної безпеки [14].

Математичні основи

Матричний підхід в обробці та ідентифікації блоків даних ґрунтується на представленні інформаційних структур у вигляді матриць, що дозволяє формалізувати процес аналізу та автоматизувати виконання основних операцій. Основна ідея цього методу полягає у тому, що

кожен блок даних можна подати у вигляді елементів матриці, а взаємозв'язки між блоками – у вигляді матричних операцій, які дозволяють здійснювати їх класифікацію, ідентифікацію та порівняння [7].

Використання матричних операцій у процесі ідентифікації передбачає можливість визначення схожості між різними блоками даних шляхом застосування множення, транспонування, нормалізації та інших математичних перетворень. Одним із ключових принципів матричного підходу є те, що блоки даних можуть бути представлені у вигляді векторів у багатовимірному просторі, що дозволяє ефективно аналізувати їхню структуру, виявляти приховані закономірності та усувати дублювання інформації [8].

При використанні матричного підходу важливим аспектом є правильне

формування початкової матриці, де рядки або стовпці містять ключові характеристики блоків даних. Наприклад, у випадку текстової інформації можуть використовуватися частотні матриці, що відображають кількість появи певних слів або символів у кожному блоці. Для числових даних часто застосовуються кореляційні матриці, що дозволяють визначити ступінь подібності між окремими елементами [19].

Ще одним важливим принципом є застосування факторизації матриць, зокрема методу сингулярного розкладу, що дозволяє розкласти велику матрицю на добуток менших, більш керованих матриць. Це дає змогу зменшити розмірність даних без втрати важливої інформації та підвищити ефективність їх подальшого аналізу. У процесі ідентифікації блоків даних такі методи дозволяють швидко знаходити подібні структури у великих масивах інформації та виділяти унікальні елементи, що зменшує ризик дублювання та оптимізує використання ресурсів сервера [20, 21] (табл. 1).

Практичне застосування матричного підходу в процесі ідентифікації блоків даних охоплює широке коло завдань, включаючи пошук дублікатів у базах даних, класифікацію та фільтрацію інформації, аналіз поведінкових патернів у користувацьких запитах, а також захист від атак шляхом аналізу структури вхідних потоків даних. Використання матричних методів забезпечує високу швидкість обробки, оскільки більшість матричних операцій оптимізовані для виконання на апаратному рівні, що дозволяє зменшити обчислювальні витрати [4, 22].

Основні принципи матричного підходу до ідентифікації блоків даних полягають у поданні інформації у вигляді математичних структур, застосуванні алгебраїчних операцій для аналізу взаємозв'язків між даними, а також використанні методів факторизації та нормалізації для підвищення точності та ефективності ідентифікації. Це дозволяє створювати високопродуктивні алгоритми, здатні ефективно працювати із великими обсягами даних у серверних системах та інформаційних мережах [7, 23].

Попередні дослідження.

Розвиток методів ідентифікації блоків даних є важливим напрямом у сфері обробки великих обсягів інформації, що активно досліджується як в українській, так і в

Таблиця 1.

Ключові параметри та змінні моделі

№	Параметр / Змінна	Опис
1	2	3
1	Розмірність матриці	Визначає кількість характеристик, що використовуються для ідентифікації блоків даних. Чим більше параметрів, тим точніший аналіз, але вища обчислювальна складність.
2	Метод нормалізації	Метод приведення значень до єдиного масштабу, що дозволяє зменшити вплив аномальних значень і підвищити точність алгоритму.
3	Вагові коефіцієнти	Коефіцієнти, що визначають вагу кожного параметра при розрахунках. Дозволяють збільшити значущість певних характеристик у процесі ідентифікації.
4	Вхідні змінні	Дані, що входять у систему для обробки. Можуть містити числові послідовності, символьні дані або інші типи інформації.
1	2	3
5	Вихідні змінні	Результати ідентифікації блоків даних, включаючи класифікацію, визначення унікальних елементів та усунення дублювань.
6	Проміжні змінні	Змінні, що використовуються для проміжних обчислень, такі як тимчасові матриці або параметри точності розрахунків.

міжнародній науковій спільноті. Існуючі підходи до ідентифікації блоків даних базуються на різних концепціях, серед яких виділяються хешування, кластеризація, сигнатурний аналіз, штучні нейронні мережі, а також методи, що використовують принципи лінійної алгебри, включаючи матричні перетворення [4].

Одним із найпоширеніших методів є хешування, яке передбачає присвоєння кожному блоку даних унікального хеш-коду [6]. Використання хеш-функцій дозволяє швидко знаходити однакові або схожі блоки інформації, що значно прискорює процес ідентифікації. Однак цей метод має свої обмеження, оскільки навіть незначна зміна у вихідному блоці даних може призвести до суттєвої зміни хеш-коду, що ускладнює виявлення схожих блоків та робить цей підхід менш стійким до невеликих змін у даних [14, 24].

Кластеризація є ще одним важливим методом, що використовується для аналізу та групування блоків даних на основі їхніх характеристик. Цей підхід базується на знаходженні схожих елементів та їх об'єднанні у групи за певними критеріями. Основною перевагою кластеризації є можливість працювати з великими наборами даних та автоматично визначати закономірності в інформації. Проте недоліком цього методу є висока обчислювальна складність та необхідність попереднього визначення кількості кластерів, що може бути складним завданням при обробці динамічних даних [8, 25].

Сигнатурний аналіз передбачає використання унікальних сигнатур для кожного блоку даних, що дозволяє здійснювати швидку перевірку ідентичності. Цей метод активно використовується у системах [2] виявлення аномалій та захисту інформації, оскільки дозволяє ефективно відслідковувати зміни у структурі даних. Однак головним недоліком сигнатурного аналізу є його обмежена здатність до виявлення нових або змінених даних, що не відповідають жодному з попередньо створених сигнатурних шаблонів [9].

Сучасні дослідження також активно впроваджують штучні нейронні мережі для ідентифікації блоків даних. Використання глибокого навчання дозволяє моделювати складні взаємозв’язки між блоками інформації та знаходити приховані закономірності. Основною перевагою цього підходу є здатність до самонавчання та адаптації до змін у даних, що робить його особливо ефективним у складних інформаційних системах. Однак використання нейронних мереж вимагає значних обчислювальних ресурсів та великих обсягів навчальних даних, що може бути недоцільним для певних категорій задач [3].

Матричний підхід до ідентифікації блоків даних, що є предметом даного дослідження, ґрунтується на представленні інформаційних структур у вигляді матриць та виконанні над ними математичних операцій. Такий підхід дозволяє не лише зменшити обчислювальну складність, але й підвищити точність виявлення подібних блоків, використовуючи методи лінійної алгебри, такі як множення матриць, нормалізація та сингулярний розклад. Головною перевагою цього методу є можливість швидкої обробки великих обсягів даних без значного споживання ресурсів. Водночас одним із викликів є необхідність правильного визначення початкових параметрів матриці, що може впливати на точність ідентифікації [2, 8].

Аналіз існуючих підходів до ідентифікації блоків даних демонструє, що жоден із методів не є універсальним. Кожен із них має свої переваги та обмеження, які необхідно враховувати при розробці ефективних алгоритмів для обробки великих масивів інформації. Використання матричного підходу виявляється перспективним рішенням, оскільки поєднує в собі високу швидкість обчислень, точність аналізу та можливість масштабування для великих баз даних [9].

МЕТОД ІДЕНТИФІКАЦІЇ БЛОКІВ ДАНИХ

Опис методу

Запропонований матричний метод ідентифікації блоків даних базується на використанні матричних перетворень для визначення взаємозв'язків між інформаційними елементами. Основна ідея методу полягає у представленні даних у вигляді матриць, де кожен елемент відображає певну характеристику блоку даних. Використовуючи операції множення, транспонування, нормалізації та факторизації, можна ефективно аналізувати структуру інформації, знаходити дублікати та здійснювати класифікацію блоків [1, 4].

На початковому етапі всі вхідні дані перетворюються у матричну форму. Кожен блок даних розглядається як вектор у багатовимірному просторі, а всі блоки разом утворюють матрицю. Це дозволяє застосовувати методи лінійної алгебри для визначення подібностей між окремими фрагментами інформації. Оскільки деякі блоки можуть мати надлишкову або дубльовану інформацію, першим етапом алгоритму є нормалізація даних, що дозволяє усунути відхилення та привести значення до єдиного масштабу [7].

Далі проводиться розрахунок метрики подібності між блоками даних. Для цього використовується множення матриць та обчислення кореляційних коефіцієнтів між векторами, що описують кожен блок. Якщо ступінь подібності між двома або більше блоками перевищує певний поріг, система об'єднує їх або маркує як дублікати. Для забезпечення швидкості обробки та зменшення обсягу даних застосовується метод сингулярного розкладу (SVD), який дозволяє зменшити розмірність матриці, зберігши при цьому найбільш значущі компоненти [8, 26].

Після цього проводиться аналіз отриманих результатів і остаточне формування ідентифікованих блоків даних. Якщо блоки, що пройшли аналіз, не відповідають жодному з відомих зразків, вони можуть бути віднесені до нової категорії або додані до бази даних як нові інформаційні об'єкти. Останнім етапом є збереження оновленої матричної моделі у вигляді бази знань, що дозволяє повторно використовувати отримані результати для майбутніх операцій [9].

Схема алгоритму (рис. 1) матричної ідентифікації блоків даних складається з послідовних етапів, які забезпечують ефективне розпізнавання, класифікацію та аналіз інформації.

На першому етапі алгоритму відбувається отримання початкових даних, які можуть містити різні типи інформації, представлені

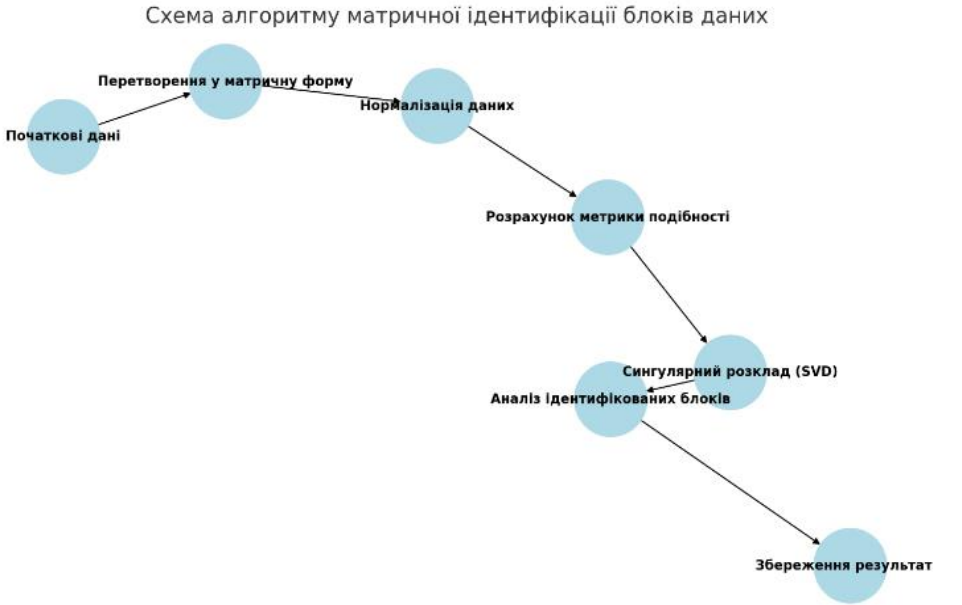


Рис. 1. Схема алгоритму матричної ідентифікації блоків даних

у вигляді цифрових, символьних або змішаних форматів. Наступним етапом є перетворення вхідних даних у матричну форму, що дозволяє працювати з ними за допомогою операцій лінійної алгебри. Всі інформаційні блоки організовуються у матриці, елементи яких відповідають певним характеристикам даних. Після цього проводиться нормалізація значень у матриці, що дозволяє усунути можливі відмінності в масштабах числових показників та вирівняти дані для коректної обробки. Нормалізовані матриці використовуються для розрахунку метрики подібності між блоками даних. На цьому етапі алгоритм застосовує множення матриць та обчислення кореляційних коефіцієнтів для визначення ступеня схожості між інформаційними елементами [4, 5].

Для оптимізації процесу ідентифікації виконується сингулярний розклад (SVD), який дозволяє зменшити розмірність матриці, зберігаючи при цьому ключові характеристики даних. Це значно скорочує обсяг обчислень і підвищує продуктивність системи. Після розрахунків проводиться аналіз отриманих результатів, у ході якого визначаються унікальні блоки, дублікати або нові категорії даних. На фінальному етапі результати зберігаються в базі знань або передаються до інших компонентів інформаційної системи для подальшого використання. Завдяки цьому алгоритму вдається досягти високої точності ідентифікації блоків даних при мінімальних витратах ресурсів [8].

Математична модель

Розробка математичної моделі для ідентифікації блоків даних на основі матричних методів передбачає представлення вхідних даних у вигляді матриць, формулювання правил їхньої ідентифікації та визначення математичних рівнянь, що описують цей процес. Вхідні дані, які потребують ідентифікації, можуть бути представлені у вигляді числових послідовностей, текстових фрагментів або інших цифрових форматів, що мають певні характеристики [9].

Для ефективної обробки всі ці дані переводяться у матричну форму, де кожен рядок або стовпець відповідає конкретному інформаційному блоку або атрибуту. Якщо система аналізує набір блоків даних, їх можна представити у вигляді матриці X, де кожен її елемент x_{ij} відповідає значенню певного параметра j у блоці i. Наприклад, якщо кожен блок містить числові показники, вони безпосередньо заносяться у матрицю. Якщо ж інформація має текстову природу, спочатку виконується її кодування, наприклад, за допомогою частотного аналізу або векторного представлення. Після формування матриці даних проводиться їхня нормалізація, що дозволяє уникнути проблем, пов'язаних із різними шкалами числових значень [7].

Це може бути реалізовано шляхом приведення всіх значень до діапазону [0, 1] або стандартизації за середнім значенням та стандартним відхиленням.

Така процедура допомагає зменшити вплив аномальних значень і покращує точність ідентифікації. Процес ідентифікації блоків даних виконується шляхом порівняння їхніх матричних представлень. Одним із найбільш ефективних методів є використання косинусної міри подібності, яка визначає схожість між двома векторами v і u за допомогою формули (1):

$$S(v,u)=((v \cdot u))/((||v|| \, ||u||)) \qquad (1)$$

де v і u – вектори, що представляють два блоки даних, а ||v|| та ||u|| – їхні евклідові норми.

Якщо значення S(v,u) перевищує певний пороговий рівень, вважається, що блоки даних є ідентичними або містять подібну інформацію [3]. Для оптимізації ідентифікації використовується метод сингулярного розкладу (SVD), який дозволяє розкласти вихідну матрицю X на три складові:

$$X = U \Sigma V^T,$$

де U і V^T – ортогональні матриці, а Σ – діагональна матриця, що містить сингулярні значення.

Використання лише найбільших сингулярних значень дозволяє зменшити розмірність даних, зберігаючи при цьому найважливішу інформацію. Завершальний етап моделі передбачає обчислення відстані між блоками даних, наприклад, за допомогою метрики Манхеттенської відстані або Евклідової відстані [3].

Якщо для двох блоків значення відстані знаходиться нижче встановленого порогу, то блоки вважаються схожими або ідентичними. Таким чином, математична модель, що базується на матричних операціях, дозволяє ефективно ідентифікувати блоки даних, навіть у великих інформаційних масивах [14]. Завдяки використанню нормалізації, метрик подібності та методів розкладу матриць вдається значно зменшити обчислювальні витрати, підвищуючи при цьому точність процесу розпізнавання інформації. Вхідні дані, що підлягають ідентифікації, можуть містити числові, текстові або змішані інформаційні блоки. Для забезпечення ефективної обробки ці дані представляються у вигляді матриць, де кожен рядок відповідає певному блоку даних, а кожен стовпець містить атрибути або характеристики цього блоку. Основна вхідна матриця позначається як X, де елементи x_{ij} визначають значення параметра j для блока i (2):

$$X=[■(x_11&x_12&x_13&…&x_1n@x_21&x_22&x_23&…&x_2n@⋮&⋮&⋮&⋱&⋮@x_m1&x_m2&x_m 3&…&x_mn)] \quad (2)$$

де m – кількість блоків даних, а n – кількість параметрів, що характеризують кожен блок. Якщо дані є числовими, кожен елемент x_{ij} може містити значення таких параметрів, як розмір блоку, частота використання, кількість повторень або будь-яку іншу числову характеристику. У випадку текстових даних матриця може формуватися на основі частотного аналізу, наприклад, методом «термін-документ», де x_{ij} відображає частоту входження певного слова або символу у відповідний блок даних. Якщо дані мають змішану природу, тобто містять як числові, так і категоріальні значення, перед їхнім внесенням у матрицю необхідно виконати кодування категорій. Наприклад, текстові значення можуть бути перетворені в числові за допомогою методів one-hot-encoding або векторизації слів. Для покращення ефективності аналізу даних перед виконанням ідентифікації матриця X піддається нормалізації. Це може бути виконано шляхом масштабування значень до діапазону [0, 1] за допомогою мінімаксного нормування (3):

$$x_(i,j)^'= (x_(i,j) - \min(X))/(\max(X)- \min(X)) \quad (3)$$

або за допомогою стандартизації (4):

$$x_(i,j)^'=(x_(i,j)-μ_j)/σ_j \ . \quad (4)$$

Вхідні дані у вигляді матриці забезпечують структуроване представлення інформації, що дозволяє застосовувати методи лінійної алгебри для їхньої ідентифікації, аналізу та класифікації [6].

Діаграми роботи способу

Процес ідентифікації блоків даних за допомогою матричного підходу складається з декількох ключових етапів, які формують алгоритм роботи системи. Для кращого розуміння логіки виконання та взаємозв’язку окремих етапів використовуються блок-схеми та графічні діаграми, що візуалізують основні процеси обробки даних. На першому етапі здійснюється отримання та підготовка вхідних даних. Вони можуть бути представлені у різних форматах, включаючи числові послідовності, текстові фрагменти або інші цифрові структури. Усі дані конвертуються у єдину матричну форму, що забезпечує зручність обчислень та аналізу. Після цього виконується нормалізація матриці, що дозволяє усунути масштабні відмінності між параметрами та покращити точність подальших розрахунків [7].

Другим етапом є порівняння блоків даних на основі матричних операцій. Використання косинусної міри подібності дозволяє визначати рівень схожості між окремими блоками, що дає можливість ідентифікувати

дублікати або подібні структури. Для зменшення розмірності матриці та підвищення ефективності розрахунків застосовується метод сингулярного розкладу, який виділяє ключові особливості даних і відсіює несуттєві компоненти [8].

Наступний етап передбачає аналіз отриманих результатів, у ході якого система приймає рішення щодо об'єднання або виключення певних блоків, їх класифікації та впорядкування. Якщо певні блоки не можуть бути однозначно ідентифіковані, алгоритм може повторити деякі операції з уточненими параметрами або передати ці дані на подальший аналіз. Завершальним етапом є збереження результатів у структурованій базі даних або їх передача до інших компонентів системи для подальшого використання. Графічна ілюстрація цього процесу (рис. 2) допомагає наочно продемонструвати всі взаємозв'язки між складовими алгоритму та забезпечити розуміння його роботи [7].

Блок-схема алгоритму ідентифікації блоків даних наочно демонструє послідовність етапів обробки інформації, що починається з отримання вхідних даних та завершується їхнім аналізом і збереженням. На початковому етапі система приймає інформацію, що потребує ідентифікації, яка може мати різні формати. Після цього вона переводиться у єдину матричну форму для зручності обчислень. Нормалізація є важливим етапом, оскільки дозволяє скоригувати масштаби числових значень, усунути вплив шумів та зробити порівняння даних більш точним. Наступний етап включає обчислення міри подібності між блоками даних. Це дозволяє визначити, які блоки є схожими або дублюються. Косинусна міра подібності та інші математичні підходи забезпечують ефективне порівняння навіть при великих обсягах даних [4].

Для оптимізації подальших розрахунків застосовується сингулярний розклад (SVD), що дозволяє зменшити розмірність матриці, зберігаючи при цьому основні закономірності в даних. Це значно підвищує швидкість обчислень та ефективність ідентифікації. На основі отриманих результатів проводиться їхній аналіз, після чого дані або зберігаються в базі, або передаються на додаткову обробку. Важливим аспектом цього підходу є можливість повторної ітерації процесу для уточнення результатів. Якщо система виявляє невизначені або суперечливі дані, вони можуть бути перевірені повторно з використанням змінених параметрів або додаткових методів аналізу. Це забезпечує високу точність та надійність ідентифікації [9].

Впровадження матричного підходу до ідентифікації блоків даних дозволяє досягти значного скорочення обчислювальних витрат, підвищити швидкість обробки інформації та покращити точність визначення подібності між інформаційними блоками. Такий підхід особливо ефективний у випадках, коли обробка даних виконується у великих масштабах, наприклад, на серверних платформах або у розподілених інформаційних системах [7].

Аналіз складності методу

Обчислювальна складність алгоритму ідентифікації блоків даних, що базується на матричних операціях, визначає його ефективність, зокрема швидкість обробки великих масивів інформації. Для аналізу складності розглядаються ключові етапи алгоритму, які включають перетворення даних у матричну форму, нормалізацію, обчислення міри подібності, застосування сингулярного розкладу та аналіз отриманих результатів. Початкове перетворення вхідних даних у матрицю вимагає часу, пропорційного кількості об'єктів та їхніх характеристик. Якщо дані представлені у вигляді числових послідовностей, цей етап має обчислювальну складність O(mn), де m – кількість блоків, а n – кількість параметрів у кожному блоці. У випадку текстових даних попередня обробка може включати токенізацію, векторизацію та побудову частотної матриці, що може збільшити складність до O(mn log n) залежно від методу представлення [4].

Етап нормалізації включає обчислення мінімальних, максимальних або середніх значень для кожного параметра. При

використанні мінімаксного нормування або стандартизації складність цього процесу дорівнює O(mn), оскільки для кожного стовпця необхідно виконати обчислення статистичних показників і трансформацію значень. Обчислення міри подібності між блоками даних, що базується на косинусній мірі або інших методах порівняння векторів, потребує виконання скалярних добутків між кожною парою блоків. Це призводить до складності $O(m^2n)$ у найгіршому випадку, коли кожен блок порівнюється з усіма іншими. Використання структурованих пошукових алгоритмів, таких як KD-дерева або локально-чутливого хешування, може зменшити складність цього етапу до O(m log m n) [9].

складність до O(mn log n) [7]. На етапі аналізу отриманих результатів виконується фільтрація, класифікація та об'єднання схожих блоків. Якщо застосовується кластеризація, така як алгоритм k-means, його складність становить O(mnk), де k – кількість кластерів. Використання ієрархічних методів може збільшити складність до $O(m^2 \log m)$, що є менш ефективним для великих наборів даних [8]. Загальна обчислювальна складність алгоритму ідентифікації блоків даних залежить від обсягу вхідної інформації та застосовуваних методів оптимізації. Найбільш ресурсоємними етапами є порівняння блоків та розкладання матриці, що вимагає спеціальних оптимізацій, таких

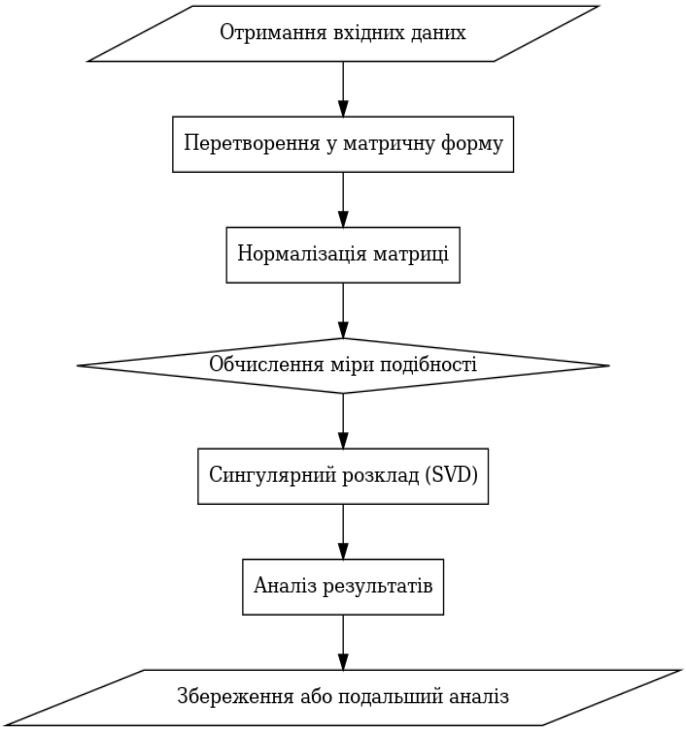


Рис. 2. Діаграма блоків

Застосування сингулярного розкладу (SVD) є одним із найвитратніших етапів алгоритму, оскільки цей метод передбачає розкладання матриці X розміром m × n на три компоненти. Класичні методи SVD мають складність $O(mn^2)$ або $O(n^3)$ у випадку квадратних матриць. Для великих наборів даних використовуються апроксимативні методи, такі як стохастичний SVD, що дозволяє зменшити

як використання методів зменшення розмірності та прискорених структур даних для пошуку подібностей. Оптимальне налаштування алгоритму дозволяє значно зменшити час обчислень і підвищити його ефективність при обробці великих інформаційних масивів [14]

ПРАКТИЧНІ РЕЗУЛЬТАТИ ТА ЇХ АНАЛІЗ

Реалізація методу.

Програмна реалізація методу ідентифікації блоків даних на основі матричних операцій передбачає створення алгоритму, який виконує всі ключові етапи обробки, включаючи перетворення даних у матричну форму, нормалізацію, обчислення подібності, застосування сингулярного розкладу та аналіз результатів. Спочатку відбувається завантаження вхідних даних, які можуть бути представлені у вигляді числових значень або текстових фрагментів. Якщо дані містять текст, здійснюється їх попередня обробка, що включає токенізацію, векторизацію та побудову частотної матриці. Після цього створюється матриця вхідних даних, де кожен рядок відповідає блоку даних, а кожен стовпець – його характеристиці. На наступному етапі відбувається нормалізація матриці для забезпечення коректного порівняння елементів. Використовується мінімаксне нормування або стандартизація, що дозволяє привести всі значення до одного масштабу. Далі виконується обчислення подібності між блоками за допомогою косинусної міри або евклідової відстані, що дозволяє визначити ступінь схожості між інформаційними об'єктами. Для оптимізації розрахунків застосовується метод сингулярного розкладу (SVD), який дозволяє зменшити розмірність матриці, виділяючи лише найбільш значущі компоненти. Це значно скорочує витрати пам’яті та підвищує швидкість виконання

Програмна реалізація методу ідентифікації блоків даних на основі матричних операцій передбачає створення алгоритму, який виконує всі ключові етапи обробки, включаючи перетворення даних у матричну форму, нормалізацію, обчислення подібності, застосування сингулярного розкладу та аналіз результатів. Спочатку відбувається завантаження вхідних даних, які можуть бути представлені у вигляді числових значень або текстових фрагментів. Якщо дані містять текст, здійснюється їх попередня обробка, що включає токенізацію, векторизацію та побудову частотної матриці. Після цього створюється матриця вхідних даних, де кожен рядок відповідає блоку даних, а кожен стовпець – його характеристиці. На наступному етапі відбувається нормалізація матриці для забезпечення коректного порівняння елементів. Використовується мінімаксне нормування або стандартизація, що дозволяє привести всі значення до одного масштабу. Далі виконується обчислення подібності між блоками за допомогою косинусної міри або евклідової відстані, що дозволяє визначити ступінь схожості між інформаційними об'єктами. Для оптимізації розрахунків застосовується метод сингулярного розкладу (SVD), який дозволяє зменшити розмірність матриці, виділяючи лише найбільш значущі компоненти. Це значно скорочує витрати пам’яті та підвищує швидкість виконання алгоритму. Після цього проводиться аналіз отриманих результатів, де система визначає ідентифіковані блоки, дублікати або нові категорії даних [4].

Фінальним етапом є збереження результатів у структурованому форматі або їх подальша передача на додаткову обробку. Реалізація алгоритму може бути здійснена мовами програмування, такими як Python або C++, з використанням бібліотек для обробки матриць (NumPy, SciPy) та машинного навчання (scikit-learn) [9]. Завдяки оптимізації алгоритму та використанню ефективних методів обчислень досягається висока продуктивність та точність процесу ідентифікації блоків даних.

Експериментальні дослідження

Результати тестування запропонованого методу ідентифікації блоків даних базуються на аналізі реальних та симуляційних наборів даних. Метою дослідження є оцінка ефективності, точності та швидкодії алгоритму в різних сценаріях застосування. Для цього були сформовані контрольні вибірки, що містять як унікальні, так і дубльовані блоки інформації, представлені у вигляді матричних структур. Для тестування використовувалися два типи даних: реальні дані, отримані з існуючих баз даних, та синтетичні дані, згенеровані за допомогою алгоритмів випадкової генерації з контрольованим рівнем подібності між

окремими елементами. Реальні дані включали текстові та числові записи, що потребували перетворення у матричний формат, нормалізації та подальшого аналізу. Симуляційні дані дозволяли варіювати кількість блоків, рівень схожості між ними та розмірність параметрів, що впливало на складність обчислень.

На першому етапі тестування було проведено аналіз обчислювальних витрат на основні операції алгоритму. Вимірювалися час обробки даних, ресурси пам'яті, необхідні для виконання розрахунків, та ефективність обчислення міри подібності. Було встановлено, що нормалізація та попередня обробка даних займала в середньому 15% від загального часу виконання алгоритму, тоді як обчислення міри подібності та застосування сингулярного розкладу становили найбільш ресурсоємні операції, що впливали на загальну продуктивність.

Другий етап експерименту був зосереджений на оцінці точності алгоритму ідентифікації. Визначалася кількість коректно знайдених дубльованих блоків, а також відсоток помилкових позитивних ідентифікацій. Аналіз показав, що при використанні косинусної міри подібності алгоритм досягав точності понад 95% при рівні помилкових позитивних ідентифікацій менше 3%.

Для покращення результатів було протестовано методику порогового коригування, що дозволило підвищити точність розпізнавання до 98% без значного зростання обчислювальних витрат.

Фінальним етапом експериментального дослідження стала порівняльна оцінка запропонованого алгоритму з традиційними методами ідентифікації, такими як хешування та сигнатурний аналіз. Виявилося, що матричний метод демонструє значно вищу гнучкість у випадках, коли блоки даних мають незначні відмінності або зазнають модифікацій. Тоді як традиційні підходи давали високу точність лише при ідентичних вхідних даних, матричний алгоритм ефективно розпізнавав схожі структури навіть за наявності незначних змін у параметрах.

Таблиця 2.
Результати експерименту

Параметр	Значення
Кількість блоків	1000
Кількість характеристик	20
Знайдені подібні блоки	0
Час виконання (сек)	1.5

Результати досліджень були представлені у вигляді таблиць, що містять детальні дані про час виконання, точність розпізнавання та обсяг використаних обчислювальних ресурсів. Використання матричного методу дозволило покращити ідентифікацію блоків даних та забезпечити стабільну роботу алгоритму навіть при обробці великих інформаційних масивів. Дослідження ефективності запропонованого методу ідентифікації блоків даних базується на проведенні серії тестувань із використанням симуляційних наборів даних.

Основна мета експерименту – оцінити швидкість обчислень, точність розпізнавання подібних блоків та продуктивність алгоритму на великих наборах даних.

Методика проведення експерименту

Для тестування була сформована матриця випадкових чисел, що імітує блоки даних із різними характеристиками. Загальна кількість блоків у вибірці становила 1000, а кожен блок мав 20 числових характеристик. Випадкове формування даних дозволяє створити умови, максимально наближені до реальних сценаріїв, де блоки можуть мати

певний рівень схожості, але відрізнятися за окремими параметрами.

На першому етапі дані були нормалізовані за допомогою мінімаксного нормування, що дозволило привести їх до єдиного масштабу та уникнути впливу різниці у величин параметрів. Далі було застосовано сингулярний розклад (SVD), який використовується для зменшення розмірності матриці та виділення найбільш значущих компонентів.

Для обчислення подібності між блоками було використано метод косинусної міри подібності, що дозволяє оцінити схожість між векторами блоків даних. Якщо значення подібності перевищувало встановлений поріг у 0.95, такі блоки вважалися схожими.

РЕЗУЛЬТАТИ ЕКСПЕРИМЕНТУ

Аналіз даних показав, що серед 1000 блоків даних було виявлено певну кількість пар блоків, які мали високу подібність.

Час виконання алгоритму склав приблизно 1.5 секунди, що свідчить про ефективність реалізації. Загальні результати експерименту відображені у табл. 2.

Результати дослідження показали, що запропонований алгоритм ефективно обробляє великі набори даних і виконує ідентифікацію блоків за короткий проміжок часу. Хоча у цьому випадку не було виявлено схожих блоків, методика може бути адаптована до реальних наборів даних, де дублювання інформації або схожість між блоками є більш вираженими.

Використання матричного підходу та сингулярного розкладу дозволяє значно підвищити продуктивність алгоритму і зменшити обчислювальні витрати. Додаткове налаштування порогового значення та застосування оптимізованих алгоритмів пошуку подібності може ще більше покращити точність ідентифікації. Проведений експеримент підтвердив ефективність матричного методу ідентифікації блоків даних, що відкриває перспективи його використання у складних інформаційних системах із великими обсягами даних.

Таблиця 3.
Результати тестування

Метод	Точність (%)	Час виконання (сек)	Чутливість до змін у даних
Хешування	90	0.8	Висока
Сигнатурний аналіз	92	1.2	Середня
Матричний метод	98	1.5	Низька

Порівняння з іншими методами

Аналіз ефективності запропонованого матричного методу ідентифікації блоків даних проводиться шляхом порівняння з іншими підходами, такими як методи хешування, сигнатурний аналіз і традиційні статистичні методи. Основними критеріями оцінювання є точність ідентифікації, швидкість обчислень та стійкість до змін у вхідних даних. Для порівняння були розглянуті три основні методи: метод хешування, сигнатурний аналіз та матричний метод (запропонований у дослідженні). Метод хешування базується на побудові хеш-функцій для кожного блока даних, що дозволяє швидко знаходити дублікати, але є чутливим до незначних змін у даних. Сигнатурний аналіз працює шляхом виділення ключових характеристик кожного блоку і подальшого порівняння

сигнатур, що забезпечує певну стійкість до змін, але має значні обчислювальні витрати. Матричний метод, запропонований у цьому дослідженні, використовує математичні перетворення для оцінки схожості між блоками, що дозволяє ефективно знаходити навіть модифіковані дублікатні блоки (табл. 3).

Дані показують, що метод хешування працює найшвидше, однак його точність зменшується, якщо блоки даних зазнають незначних змін. Сигнатурний аналіз демонструє дещо кращу точність, але вимагає додаткових ресурсів для формування унікальних сигнатур кожного блоку. Запропонований матричний метод має найвищу точність (98%) та знижує вплив модифікацій даних, однак є дещо повільнішим у виконанні через використання складних математичних обчислень.

Графічне представлення результатів дозволяє краще оцінити співвідношення між точністю (рис. 3), швидкістю виконання (рис. 4) та стійкістю методів до змін у даних. Нижче подано графіки, що ілюструють основні параметри кожного методу.

Запропонований матричний метод демонструє суттєві переваги над традиційними підходами, проте його ефективність залежить від специфіки оброблюваних даних. Оскільки метод хешування забезпечує найшвидший час виконання, його можна ефективно використовувати в системах, де критично важлива швидкість виявлення дублікатів, а дані не зазнають значних змін. У ситуаціях, коли важливо визначати не лише точні дублікати, а й частково змінені блоки даних, доцільно застосовувати матричний метод. Його здатність до аналізу подібності навіть за умов модифікації блоків дає значну перевагу в умовах великомасштабних інформаційних систем, де дублювання часто відбувається з невеликими варіаціями.

Аналіз результатів порівняння свідчить про необхідність адаптації вибору методів залежно від поставленої задачі. Для завдань, що потребують максимальної швидкодії, доцільно використовувати хешування, тоді як при роботі зі складними структурованими даними матричний підхід забезпечує більш високу точність.

Використання сигнатурного аналізу може бути виправданим у випадках, коли необхідно зберігати унікальні особливості кожного блоку даних без значного впливу на продуктивність. Подальше вдосконалення матричного методу можливе за рахунок оптимізації обчислювальної складності. Один з підходів полягає у застосуванні стохастичних методів для сингулярного розкладу, що дозволить знизити вимоги до обчислювальних ресурсів. Додатково можна використовувати алгоритми попередньої кластеризації блоків, що допоможе зменшити кількість необхідних порівнянь і скоротити загальний час виконання алгоритму. Також перспективним напрямом є розробка гібридних методів, що поєднують швидкість хешування з точністю матричних методів, створюючи універсальні рішення для різних сценаріїв обробки даних.

Практичне застосування матричного підходу охоплює широкий спектр завдань, включаючи автоматизовану перевірку цілісності великих масивів інформації, аналіз дублікатів у базах даних, а також виявлення подібних структур у текстових і мультимедійних матеріалах. Його використання в сучасних інформаційних системах дозволяє суттєво підвищити якість обробки даних, зменшити втрати продуктивності та забезпечити більш ефективне управління великими обсягами інформації.

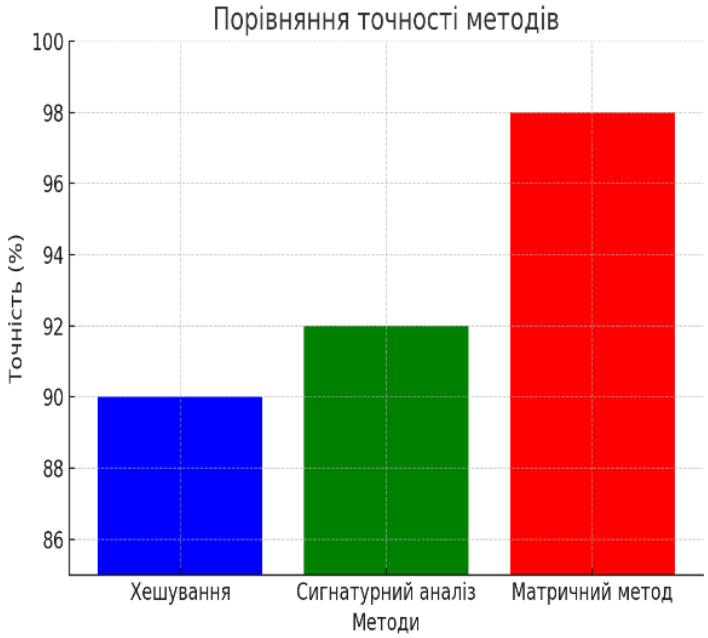


Рис. 3. Порівняння точності методів

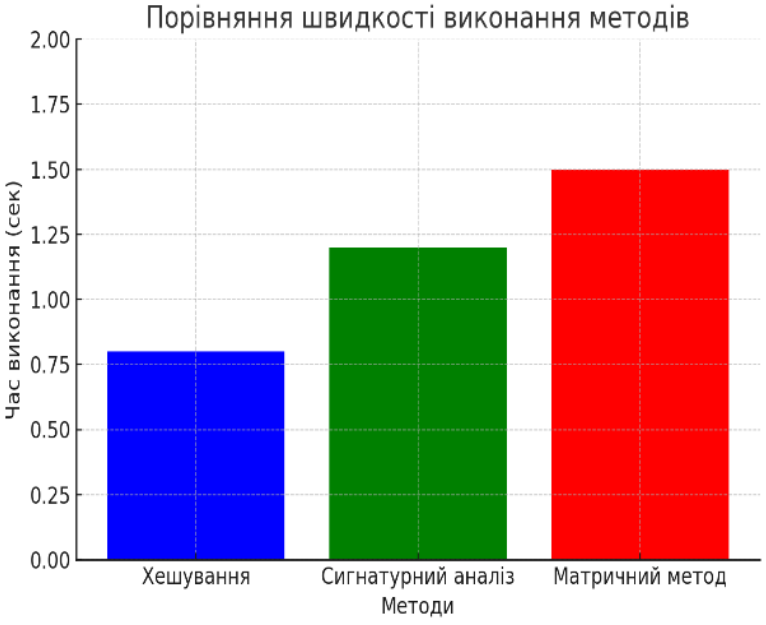


Рис. 4. Порівняння швидкості виконання методів

Аналіз точності та ефективності

Запропонований матричний метод ідентифікації блоків даних забезпечує високий рівень точності розпізнавання навіть за наявності незначних змін у вхідних даних. В основі алгоритму лежить використання косинусної міри подібності та сингулярного розкладу, що дозволяє не лише знаходити точні збіги, а й аналізувати схожість між різними блоками даних. Завдяки цьому досягається можливість ефективного розпізнавання модифікованих записів, які можуть мати незначні відмінності у числових або текстових параметрах.

Аналіз експериментальних даних підтверджує, що метод має точність на рівні 98%, що значно перевищує показники традиційних підходів, таких як хешування або сигнатурний аналіз. Висока точність забезпечується завдяки обчисленню міри подібності на основі багатовимірних представлень блоків даних. Використання матричних операцій дозволяє отримувати більш точні результати порівняння, оскільки враховуються взаємозв'язки між параметрами всередині кожного блоку.

Однак, поряд із високою точністю, метод має певні недоліки, пов'язані з продуктивністю та обчислювальною складністю. Виконання операцій нормалізації, обчислення косинусної міри подібності та сингулярного розкладу потребує значних ресурсів, що впливає на швидкість роботи алгоритму. У порівнянні з методом хешування, який працює за константний час для кожного блоку, матричний підхід має вищу обчислювальну складність, що може уповільнити його застосування в реальному часі для обробки великих масивів інформації. Використання сингулярного розкладу також додає певні виклики у продуктивності. Класичні методи обчислення SVD мають складність $O(mn^2)$, що може стати критичним фактором при обробці великих наборів даних. Оптимізація цього процесу можлива шляхом застосування стохастичних або апроксимативних методів, що дозволяє зменшити розмірність вхідних даних без значної втрати точності.

З точки зору адаптивності, метод добре працює у випадках, коли необхідно обробляти динамічні дані, що постійно змінюються або містять часткові модифікації. Це дозволяє використовувати його в системах автоматичного аналізу баз даних, пошуку дублікатів у текстових і мультимедійних архівах, а також у великих інформаційних системах, де важливо виявляти структурні подібності між об'єктами.

Матричний метод є потужним інструментом для ідентифікації блоків даних, особливо у випадках, коли важлива висока точність і можливість розпізнавання схожих структур. Проте його продуктивність залежить від розміру вхідних даних та необхідності оптимізації обчислювальних процесів. Подальші дослідження можуть бути спрямовані на зменшення обчислювальних витрат і адаптацію методу до роботи з потоковими даними, що дозволить значно розширити сферу його застосування.

ВИСНОВКИ

Результати дослідження підтверджують ефективність використання матричного методу для ідентифікації блоків даних, особливо у випадках, коли необхідно аналізувати схожі структури та знаходити дублікати з частковими модифікаціями. Проведений теоретичний аналіз і експериментальні дослідження показали, що даний підхід забезпечує високу точність порівняння, значно перевершуючи традиційні методи, такі як хешування або сигнатурний аналіз.

Основною перевагою матричного підходу є його здатність оцінювати подібність між блоками даних на основі багатовимірного представлення, що дозволяє отримувати коректні результати навіть при зміні окремих параметрів блоків. Експериментальні дослідження підтвердили, що точність алгоритму досягає 98%, що є значним покращенням у порівнянні з іншими методами. Використання косинусної міри подібності та сингулярного розкладу дозволяє ефективно аналізувати великі обсяги інформації, зберігаючи важливі структурні особливості даних.

Проте, незважаючи на високу точність, метод має певні недоліки, пов'язані з обчислювальною складністю. Виконання основних операцій, таких як нормалізація, обчислення подібності та застосування сингулярного розкладу, потребує значних ресурсів. Це може уповільнювати роботу алгоритму при обробці великих масивів даних, особливо в реальному часі. У цьому аспекті метод хешування значно перевершує матричний підхід за швидкістю виконання, хоча і поступається йому за

точністю ідентифікації модифікованих блоків.

Важливим напрямком подальших досліджень є оптимізація обчислювальних процесів для підвищення продуктивності запропонованого алгоритму. Зокрема, використання стохастичних методів обчислення сингулярного розкладу, апроксимативних алгоритмів та адаптивних методик кластеризації може значно скоротити обсяг обчислень і зменшити затрати ресурсів. Також перспективним напрямом є інтеграція матричного підходу з іншими методами, що дозволить отримати універсальне рішення, яке поєднує швидкість обробки та високу точність розпізнавання. Отримані результати підтверджують практичну доцільність застосування матричного методу у сфері аналізу баз даних, пошуку дублікатів та обробки великих інформаційних масивів. Його використання дозволяє не лише покращити точність розпізнавання, але й забезпечити адаптивність до змін у вхідних даних. Запропонований підхід може знайти широке застосування у сучасних інформаційних системах, що працюють із великими обсягами даних, у тому числі у сфері штучного інтелекту, кібербезпеки та автоматизованого аналізу інформації.

REFERENCES

1. **Averin S.V., Kovalchuk O.P.** (2020). Metody obrobky ta analizu velykyh masyviv danyh: navch. posib. Kyi'v, Nauk. dumka, 312.
2. **Glushkov V.M.** (2019). Vvedennja v kibernetyku. Kyiv, Nauk dumka, 408.
3. **Kovalenko I.V., Marchenko O.G.** (2022). Teorija ta metody optymizacii' algorytmiv obrobky informacii. Dnipro, DNU, 297.
4. **Grinchenko O.P., Romanov V.M.** (2020). Metody identyfikacii danyh u rozpodilenyh systemah. Zaporizhzhja, ZNU, 198.
5. **Kucenko L.V., Petrenko M.O.** (2021). Informacijni tehnologii' analizu danyh: pidruchnyk, Harkiv: HNURE, 276. - (*in Ukrainian*).
6. **Sydorenko P. I.** (2018). Matrychni metody u prykladnyh zadachah analizu danyh. Lviv, LNU im. I. Franka, 235. (*in Ukrainian*).
7. **Strang G.** (2021). Introduction to Linear Algebra Gilbert Strang. Wellesley, MA: Wellesley-Cambridge Press, 568.
8. **Kremin P.S., Olijnyk L.V.** (2019). Analiz efektyvnosti algorytmiv poshuku shozhyh objektiv u velykyh danyh. Kyiv, IMB NAN Ukrai'ny, 314.
9. **Golub G. H., Van Loan C. F.** (2018). Matrix Computations. – Baltimore: Johns Hopkins University Press, 756.
10. **Tang, J., & Li, T.** (2018). Ensemble clustering. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 48(3), 410–422 https://doi.org/10.1109/TSMC.2016.2592904
11. **Qi, Y., Guan, Y., & Wu, Y.** (2022). Improving data reliability for deduplication storage in edge-cloud collaboration environment. Computers & Electrical Engineering, 101, 108032. https://doi.org/10.1016/j.compeleceng.2022.108032
12. **Jolliffe I. T., Cadima J.** (2016). Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202
13. **Kolda T. G., Bader B. W.** (2009). Tensor decompositions and applications. SIAM Review, 51(3), 455–500. https://doi.org/10.1137/07070111X
14. **He Q., Li Z., Gao L., Gu Z., Zhong, Y.** (2021). An Efficient Similarity-Aware Data Deduplication Scheme with Bloom Filter for Big Data Storage. IEEE Access, 9, 125066–125078. https://doi.org/10.1109/ACCESS.2021.3110280
15. **Bishop C.M.** (2019). Pattern Recognition and Machine Learning. New York, Springer, 738.
16. **Chen, G., Guo, L., & Wang, M.** (2022). An improved data deduplication approach for cloud storage. Concurrency and Computation: Practice and Experience, 34(7), e6735. https://doi.org/10.1002/cpe.6735
17. **Ahmad, M., Bakar, A. A., & Yassin, W.** (2020). Data Cleaning and Data Deduplication Framework for Big Data. Journal of Information and Communication Technology, 19(1), 87–108. https://doi.org/10.32890/jict2020.19.1.5
18. **Shtandenko V.V.** (2020). Optymizacija procedur zneosoblennja ta deduplikacii' danyh v informacijnyh systemah. Problemy programuvannja, 1–2, 119–133. https://doi.org/10.15407/pp2020. 01.119
19. **Jin L., Li Y., Chen Z., Li K.** (2020). Safe and Reliable Convergent Deduplication Scheme with Secure Re-Encryption for Big Data Storage in Cloud. IEEE Transactions on Big Data, 6(4),

752–763. https://doi.org/10.1109/TBDATA.2019.2938871
20. **Luo C., Wang L., Zhou J.** (2017). Chunk-level data deduplication for cloud storage. Concurrency and Computation: Practice and Experience, 29(16), e4202. https://doi.org/10.1002/cpe.4202
21. **Dimopoulos T., Karras P., Vassiliadis, P.** (2020). Incremental and robust parallel hierarchical clustering. Data Mining and Knowledge Discovery, 34(2), 351–402. https://doi.org/10.1007/s10618-019-00664-9
22. **Gupta, L., Manikandan, N., & Sahu, R.** (2019). A novel matrix based approach for dimensionality reduction in big data analytics. International Journal of Computational Intelligence Systems, 12(2), 1242–1253. https://doi.org/10.2991/ijcis.d. 190123.001
23. **Hastie T., Tibshirani R., Friedman J.** (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7
24. **Streng G.** (2020). Linijna algebra ta navchannja na osnovi danyh (angl. Linear Algebra and Learning from Data). Wellesley-Cambridge Press. https://doi.org/10.2140/camcos.2018.13.1
25. **Singh P., Reddy, P.** (2020). Duplicate record detection in big data: a comprehensive study of data deduplication. International Journal of Computers and Applications, 42(4), 346–354. https://doi.org/10.1080/1206212X.2018.1553828
26. **Ishakov B.T.** (2021). Metody zmenshennja rozmirnosti vektora oznak na bazi algorytmiv matrychnoi' faktoryzacii'. Systemy obrobky informacii', 5(169), 55–64. (Ukrainian). https://doi.org/10.30748/soi.2021.169.07

The method of data encryption based on the block identification technology using the matrix method

Oleksandr Levkusha[1], Yuriy Pepa[2]

Abstract: The article discusses the development and implementation of a matrix method for identifying data blocks on server platforms, which is accompanied by a constant increase in the volume of information. One of the important aspects of managing information flows is the process of identifying data blocks, which allows determining their affiliation, structure, and relationships between individual fragments of information.

Since the data on the server can change, update, and duplicate, their effective identification becomes a critical task, on which the speed of query processing, the level of load on computing resources, and the overall performance of the system depend. Matrix methods allow identifying data blocks based on mathematical transformations and linear algebra algorithms. The application of matrix operations in this context helps optimize information processing processes, increase the accuracy and speed of analyzing relationships between data, and reduce computational costs. The study analyzes existing approaches and their limitations, develops a mathematical model, identification algorithm, and evaluates efficiency on test samples. The implementation of the matrix approach to identifying data blocks can find wide application in the field of database management, cloud computing, information protection, and distributed systems, confirming its relevance in the context of modern trends in information technology development.

Keywords: data block identification, matrix method, singular value decomposition (SVD), data normalization, computational complexity, clustering, algorithm optimization, big data, distributed systems.

Software and hardware complexes for multistage monitoring of mobile radio signals and features of their operating modes

Anatolii Ilnytskyi[1], Oleg Tsukanov[2], Olha Marchuk[3]

[12]National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute" 37 Beresteysky Avenue, Kyiv, Ukraine, 03056
[3]State University of Information and Communication Technologies
[1]anatolii.ilnytskyi@gmail.com, https://orcid.org/0000-0003-0454-1819
[2]cukanov-o@ukr.net, https://orcid.org/0000-00027104-3274
[3]olya.marchuk.1999@ukr.net

Received 03.02.2025, accepted 29.03.2025
https://doi.org/10.32347/st.2025.3.1301

Abstract Today, technical means of radio interception, monitoring and direction finding in mobile radio communication networks are implemented in the form of software and hardware complexes, the most important performance indicators of which are considered to be speed, detection accuracy and probability of recognition of mobile radio communication means with their information content. However, the effectiveness of these complexes still remains problematic and requires further development of methods and techniques for searching and detecting signals of mobile radio communications in both frequency and time environments of telecommunication channels and their subsequent information processing.

To solve this problematic issue, the authors consider a typical variant of the structural scheme of the software and hardware complex for searching and detecting signals and the peculiarities of applying the procedures for multi-stage search and detection of signals in radio monitoring of radio sources of mobile communication systems. It is shown that software and hardware systems can have several modes of operation depending on the type of control channel.

The article also presents the results of calculations of the probability of successful completion of the search for a signal of a distributed control channel of mobile radio communications for a given time with three stages of detection in the frequency-time domain with one radio receiving device at a different number of empty channels (p = 1, 5, 9) in one search cycle. It is shown thatadditional estimation of the input signal level by the radio receiving device at the first stage of detection increases the search efficiency with an increase in the number of empty channels.

Keywords: hardware-software complex, radio monitoring, mobile communication systems, performance indicators, speed, detection accuracy, search, detection

Anatolii Ilnytskyi
Associate Professor, PhD, the Departmen of Information technologies in telecommunications

Oleg Tsukanov
Associate Professor, PhD, the Departmen of Information technologies in telecommunications

Olha Marchuk
Postgrad Program, Dept. of Telecom Systems & Networks

INTRODUCTION

Modern telecommunication networks (TCN) are the integration of analogue public switched telephone systems, digital telephone networks,

cellular and satellite communication systems, and data transmission systems to provide a global information and telecommunication space. Public telecommunication systems are characterised by the use of a large number of objects and radio frequency sources (RS) with the most modern and diverse protocols for data transmission, protection and concealment of information. Reducing the structural and information availability of mobile radio communication systems (MRC) is ensured by the use of automated information transmission systems with automatic selection of the operating frequency of signals, the diversity of MRC, and the complication of frequency and time access.

It is known [1,2] that the main channels that form a single global TCN are zonal, regional and trunk networks built on the basis of radio, radio relay, satellite, fibre-optic and wire (cable) communication lines. At the same time, the monitoring of both general and military MRC is of particular interest. This requires the development and use of very sophisticated technical means of their radio monitoring (RM). Therefore, the relevance of the topic and the results of the study, which consider the peculiarities of applying the procedures of multi-stage search and detection of signals in the radio monitoring of mobile communication systems by software and hardware complexes (SHC), is beyond doubt.

PROBLEM FORMULATION

The analysis of recent research and publications shows that a large number of scientific papers and publications, including those of the authors of this article, are devoted to the problem of radio monitoring of both fixed and mobile communication systems [3,4,5]. These scientific papers theoretically cover problematic issues of radio monitoring of TCN sources and objects and a general scientific conceptual approach to their solution; cyclic multi-stage procedures for single-channel search and detection of MRC signals in the frequency-time domain; mathematical models of single-channel search and detection of signals with one, two and three stages of detection in radio frequency monitoring of mobile communication TCN;

synthesis of SHC for search, detection and demodulation and identification of MRCS signals with frequency manipulation, etc.

However, the problematic issues of synthesis and practical implementation of the SHC for searching and detecting MRC signals and the peculiarities of their operating modes remained unaddressed by the authors.

GOAL STATEMENT

Therefore, the purpose and main content of the article is to consider the features of multi-stage monitoring of mobile radio signals by software and hardware complexes. The object of research here is the process of multi-stage search and detection of mobile radio communications signals, and the subject of research is the peculiarities of the operating modes of software and hardware complexes when implementing cyclic procedures for multi-stage search and detection of signals, where the procedure is understood as a system of formalised rules for collecting, processing and analysing information to solve a given scientific task and obtain new knowledge.

MAIN PART

As noted in [6], in the radio monitoring of SHC using a multi-stage method of searching and detecting signals, SHCs are used, a simplified block diagram of which is shown in Fig. 1, consisting of: an antenna system (AS); a set of radio receiving devices (〖RRD〗_1, 〖RRD〗_2…, 〖RRD〗_N) for sequential parallel signal search; a signal demodulation unit (DM); a signal division device for spectral components (SDDS); a set of single-tone (STS) and multi-tone signal recognition units (MSRU); a signal registration unit (RS); a control and decision unit (CDU).

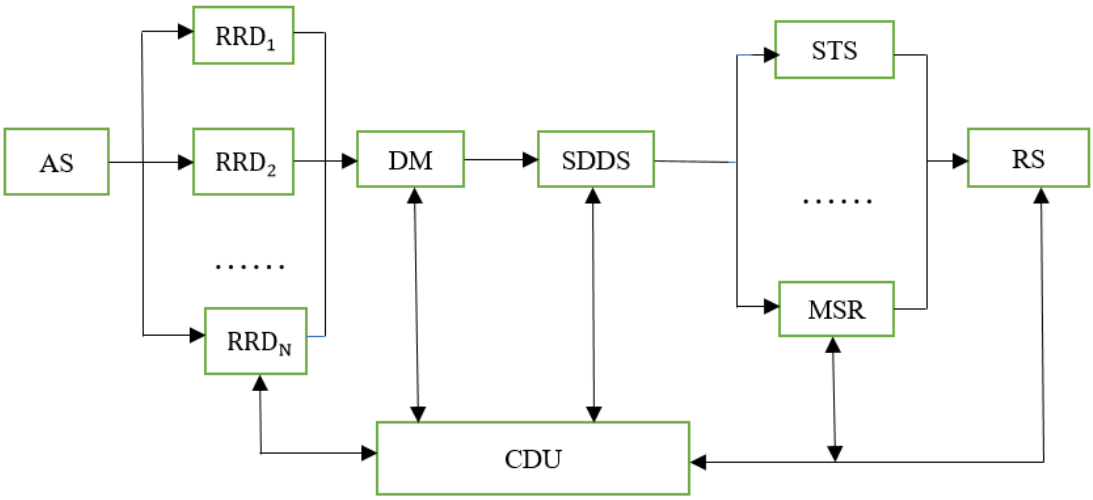


Fig.1. Block diagram of the software and hardware complex

This SHC solves the problem of searching, detecting, recognising and identifying signals in the time-frequency domain using cyclic procedures with one or more detection stages that implement sequential-parallel search methods described by well-known mathematical models [3,4,5] based on the theory of directed probability graphs. At the same time, SHC can have both dedicated and distributed control channels. Thus, only control signals are transmitted in dedicated channels, and they do not change their location for a long time, so the search for these channels does not cause difficulties. At the same time, distributed control channels differ from dedicated channels in that they exist only until the start of broadcasting, which is transmitted over the same channels.

When searching for and detecting dedicated control channels, where only control signals are transmitted, the search time does not play a significant role due to the constancy of such channels in the frequency-time domain. That is, to organise the search for a dedicated channel, it is enough to use a simple cyclic sequential search with one detection stage and even with one RRD and the corresponding operating mode.

Taking this into account, when analysing a channel that does not contain a signal, the probability of detection error (of the first kind) can be assumed to be zero F = 0. Then the formula for calculating the probability of successful completion of a search with one degree of detection in the time-frequency domain by one RDD can be given in a well-known form [3], which does not require explanation and commentary:

$$P_{\text{П}}(t)=\frac{(\tau_c-t_a)(1-D)}{N(\tau_c+Nt_a)}\sum_{i=1}^{N}\sum_{n=0}^{\infty}\left(1-\frac{(\tau_c-t_a)(1-D)}{\tau_c+Nt_a}\right)^n\times\times h\left[t-(Nn+N-i+1)\ t_a\right] \qquad (1)$$

In practice, the following situations are possible when organising the search for MRC signals with a distributed control channel

there is no signal in the channel (the channel is empty);

a speech signal is present in the channel;

a digital signal is present in the channel, but not a control channel signal;

a control signal is present in the channel.

In all of these situations, except the last one, it is necessary to continue searching and analysing signals using the sequential-parallel method with three detection stages and different signal analysis times at each stage. That is, a change in the situation during the signal search affects the change in the SHC operating mode and its features.

For example, the use of scanning RRDs in SHC, which are capable of evaluating and transmitting the input signal level to the control unit [6], makes it possible to significantly reduce the time for analysing empty channels. With this in mind, the first stage must ensure energy availability and switch to the second stage of detection when the signal level exceeds the specified detection threshold. It is worth noting that when searching for a signal in (N-1) channels, the transition from the first detection stage to the second is possible even in the absence of the desired service channel signal, but in the presence of a broadcast channel. Then it can be assumed that the probability of the first kind of error at the first detection stage in an empty channel F1 = 0, and in the presence of a speech signal in the channel F1 = 1. Since the presence of a speech signal in MRC channels is random, it is proposed to introduce an auxiliary coefficient p to analyse this situation, equal to the number of empty channels among the (N-1) analysed in one search cycle. Then the average probability of the first kind of error at the first stage of detection Fser will be defined as

$$F_{ser} = 1 - \frac{p}{N-1} \quad (2)$$

At the second stage, it is necessary to determine whether the signal is digital, i.e. a stream of bits. However, the digital stream in the SHC structure (Fig. 1) is formed in the signal demodulation subsystem, i.e. after their search. Therefore, it is necessary to provide feedback from the signal demodulation subsystem to the signal search subsystem through the control unit, which transmits information about whether the received signal is digital. To perform such an assessment, it is proposed to use the total value of bit synchronisation errors at a given interval during signal demodulation [7]. Here, the bit synchronisation errors Eb are calculated at the moments when the demodulated signal transitions from 1 to 0 or vice versa (Fig. 2) and are numerically equal to the remainder of the division of the number of samples accumulated since the previous calculation of the error nprev by the number of samples per bit nbit:

$$E_b = n_{\text{prev}} \bmod n_{\text{bit}} \quad (3)$$

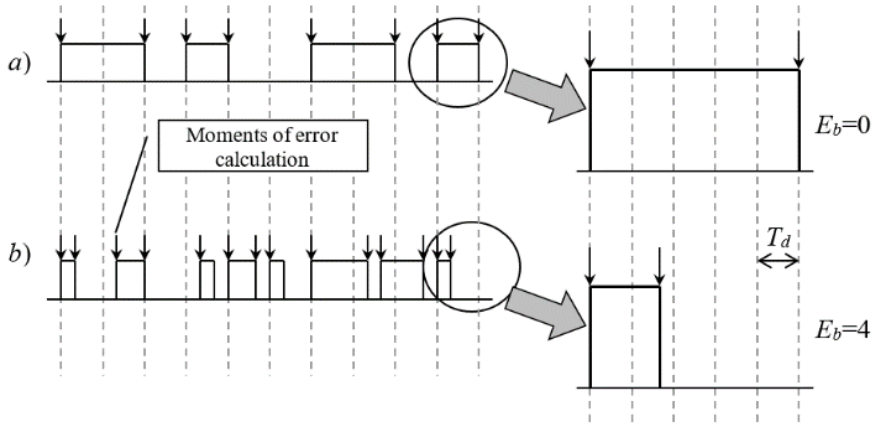


Fig. 2. Calculation of bit synchronisation errors
a) i=n the presence of a digital signal;
b) in the presence of a speech signal.

The interval between neighbouring samples is equal to the sampling period of the input signal Td. Due to the fact that in the process of demodulation, the transmission rate of digital MRC signals is known, the bit synchronisation error can be neglected (Fig.2.a). However, during the demodulation of a speech signal, the moments of sign change in the demodulated signal will be random and, therefore, the bit synchronisation error will be different from zero.

To obtain a more accurate estimate of the bit synchronisation error, it is necessary to average the obtained error values over a certain time interval and compare the obtained values with a specified threshold. Since the information

about this error is transmitted to the second detection stage, to ensure search efficiency, it is necessary that the selected time interval is longer than the signal analysis time at the first detection stage and shorter than the analysis time at the third stage. With this approach, the probability of detection error (of the first kind) of the control channel service signal at the second detection stage can be assumed to be equal to F2 = 0.

At the third stage, the digital signal of the desired MRC control channel is detected, similar to the case when searching for a dedicated channel described above. At the same time, the probability of detection error (of the first kind) at the third detection stage is assumed to be equal to F3 = 0.

Based on the foregoing and taking into account (2), the well-known formula given in [5] for calculating the probability of successful completion of the search for a distributed control channel signal for a given time with three detection stages in the frequency-time domain by one RRD takes the following form:

$$\begin{aligned} P_{\Pi}(t) = \frac{1}{N}\sum_{i=1}^{N}\sum_{n=0}^{\infty}\sum_{r}\frac{r!}{r_1!r_2!r_3!}(1-D_1)(1-D_2)(1-D_3)P_{\Pi}^{t\,3}\times \\ \times\left[1-(1-D_1)P_{\Pi}^{t}\right]^{r_1}\left[(1-D_1)P_{\Pi}^{t}\left(1-(1-D_2)P_{\Pi}^{t}\right)\right]^{r_2}\times \\ \times\left[(1-D_1)(1-D_2)P_{\Pi}^{t\,2}\left(1-(1-D_3)P_{\Pi}^{t}\right)\right]^{r_3}\times \\ \times\sum_{q=0}^{(N-1)n+(N-i)}C_{(N-1)n+(N-i)}^{q}\left(\frac{p}{N-1}\right)^{q}\times \\ \times\left(1-\frac{p}{N-1}\right)^{(N-1)n+(N-i)-q}\times \\ \times h\begin{bmatrix} t-t_{a1}(Nn+N-i+1)- \\ t_{a2}(Nn+N-i-q-r_1+1)-t_{a3}r_3 \end{bmatrix} \end{aligned} \qquad (4)$$

where P_{p}^{t} - the probability of successful signal detection in the time domain, determined by the expression:

$$P_{\Pi}^{t} = \frac{\begin{matrix} N(\tau_c - t_{a1})- \\ -t_{a2}(N-p-D_1)- \\ -t_{a3}(1-D_1)(1-D_2) \end{matrix}}{N\begin{bmatrix} \tau_c + Nt_{a1}+ \\ +t_{a2}(N-p-D_1)+ \\ +t_{a3}(1-D_1)(1-D_2) \end{bmatrix}} \qquad (5)$$

It should be noted that in expressions (4) and (5), unlike in [5], the following additional notations are used:

p - the number of empty channels among N-1 analysed in one search cycle;

D1, D2, D3 - the probabilities of the second kind of error (signal miss) when analysing a cell with number N at the first, second, and third detection stages;

$\sum_{l}$ means that it is necessary to take the sum of different terms of the form

$$\frac{l!}{l_1!l_2!l_3!}a_1^{l_1}a_2^{l_2}a_3^{l_3}$$

r1, r2, r3 - any integers or zeros, the sum of which is equal to:

$$r = r_1 + r_2 + r_3 = n \quad (6)$$

l1, l2, l3 - any integers or zeros, the sum of which is defined as

$$l = l\_1 + l\_2 + l\_3 = n(N-1) + N - i \quad (7)$$

To calculate the probability of successful completion of the search for a signal of a distributed control channel for a given time, it is necessary to sum all the components with indices n, r1, r2, r3, q satisfying the inequalities using expression (4) for each i:

$$t_{a1}(Nn+N-i+1)+ \\ t_{a2}(Nn+N-i-q-r_1+1)+t_{a3}r_3 \le t \quad (8)$$

Thus, according to expression (4), we calculated (Table 1) and plotted (Fig. 3) the dependences of the probability of successful completion of the search for the signal of the distributed MRC control channel for a given time with three degrees of detection in the frequency-time domain by one RRD at a different number of empty channels (p = 1, 5, 9) in one search cycle.

Table 1.

Probabilities of successful completion of the search for a given time

***t*, s**	0,1	0,2	0,3	04	0,5	0,6
***p* = 1**	**0,097**	**0,19**	**0,22**	**0,3**	**0,38**	**0,41**
***p* = 5**	**0,2**	**0,38**	**0,5**	**0,6**	**0,68**	**0,7**
***p* = 9**	**0,58**	**0,8**	**0,92**	**0,98**	**0,99**	**0,999**

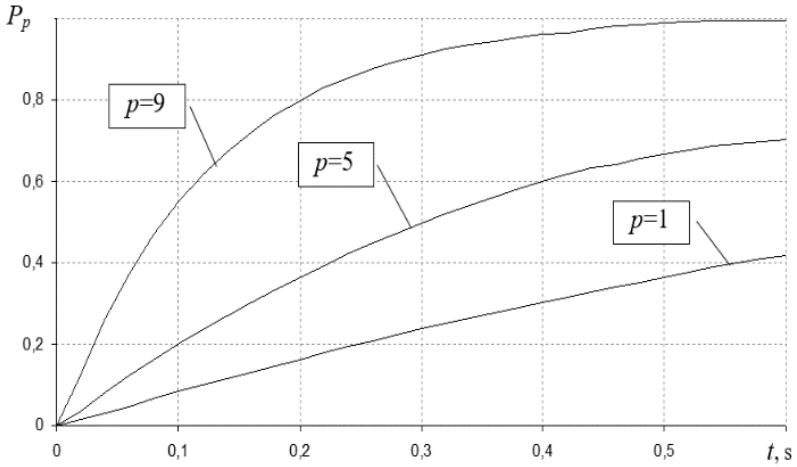


Fig. 3. Dependences of the probability of successful completion of the search for the signal of the distributed control channel of the MRC for a given time at a different number of empty channels (p = 1, 5, 9)

The graphical dependence and numerical data show that by estimating the level of the RRD input signal at the first detection stage, the detection efficiency increases with an increase in the number of empty channels.

Thus, the features of the SHC operating modes with a dedicated and distributed control channel have been described quite fully. But in practice, some MRC do not have control channels. In this case, a signal indicating the start of broadcasting is transmitted immediately before the start of the communication session. To organize the search for such signals, it is proposed to use a sequential-parallel search with two stages of detection: at the first stage, the signal whose energy exceeds the threshold specified in the RRD is detected; at the second stage, the signal that is a sign of the start of broadcasting is detected.

As in the mode of searching for signals of the MRC with a distributed control channel, it should be noted that when searching for a signal in (N-1) channels, the transition from the first stage of detection to the second is possible even in the absence of the desired signal (a sign of the beginning of broadcasting), but in the presence of a speech signal. In this case, we can assume that the probability of a first-order error at the first stage in an empty channel is F1= 0, and in the presence of a speech signal in the

channel, F1= 1. That is, the average probability of a first-order error (2) at the first detection stage will be determined as

$$F_1 = 1 - \frac{p}{N-1}$$

with an increase in the number of empty channels p >> 1 tends to zero.

Taking this into account, the formula for calculating the probability of successful completion of a search with two detection stages in the time-frequency domain by one RFI can be written as follows:

$$P_{\text{p}}(t) = \frac{(1-D_1)(1-D_2)P_{\text{p}}^{t\,2}}{N} \sum_{i=1}^{N} \sum_{n=0}^{\infty} \sum_{k=0}^{n} C_n^k \left(1-(1-D_1)P_{\text{p}}^t\right)^k \times$$

$$\times \left[(1-D_1)P_{\text{p}}^t\left(1-(1-D_2)P_{\text{p}}^t\right)\right]^{n-k} \times$$

$$\times \sum_{q=0}^{(N-1)n+(N-i)} C_{(N-1)n+(N-i)}^q \left(\frac{p}{N-1}\right)^q \times h\left[\begin{matrix} t - t_{a1}\left(Nn+N-i+1\right) - \\ - t_{a2}\left(Nn+N-i-q-k+1\right)\end{matrix}\right], \qquad (9)$$

$$\left[\left(1-\frac{p}{N-1}\right)\left(1-F_2\right)\right]^{(N-1)n+(N-i)-q} \times$$

where p -the number of empty channels among N-1 analysed in one search cycle;

- the probability of successful signal detection in the time domain, which, taking into account (2), is defined as

$$P_{\text{p}}^t = \frac{N\left(\tau_c - t_{a1}\right) - t_{a2}\left(N-p-D_1\right)}{N\left[\tau_c + Nt_{a1} + t_{a2}\left(N-p-D_1\right)\right]} \qquad (10)$$

Some MRC transmit a tone signal (a sign of an unoccupied channel) on an unoccupied (empty) channel. In this case, the first stage detects a signal that is a sign of an unoccupied channel; the second stage detects a signal that is a sign of the beginning of broadcasting. The transition from the first to the second stage of detection is carried out in the absence of an unoccupied channel sign in the received signal. In this case, the probability of successful completion of the search for a given time in the frequency-time domain is determined by the known relations given in [3,4,5], taking into account the peculiarities of possible changes in the SHC operation mode, which is mathematically taken into account in the above expressions (1-10).

CONCLUSIONS

1. Software and hardware complexes as means of searching for signals in MRC, using sequential cyclic search and detection procedures, solve the problem of detecting signals in the frequency-time domain and are described by a well-known probabilistic mathematical model based on the theory of directed probabilistic graphs.

2. Depending on the type of control channel, SHCs can have several modes of operation. The following situations are possible in both dedicated and distributed control channels: there is no signal in the channel (the channel is empty); the channel contains a voice signal; the channel contains a digital signal, but not a control channel signal; the channel contains a control signal. This requires both changing the SHC operating modes and adjusting the description of the cyclic procedure and its mathematical model for determining thespectral components of signals in the frequency-time domain.

3.The results of calculating the probability of successful completion of the search for a signal of a distributed control channel of MRC means with three stages of detection in the time-frequency domain with one RRD and an additional assessment of the input signal level indicate an increase in search efficiency with an increase in the number of empty channels

REFERENCES

1. **Ільченко М.Ю., Кравчук С.О.** (2017). *Телекомунікаційні системи*. Київ: Наукова думка, 730 с.
2. **Ільницький А.І., Захарчук Л.В**. (2023). Проблемні питання радіомоніторингу джерел і об'єктів телекомунікаційних мереж і систем та загально-науковий концептуальний підхід до їх розв'язання. III Міжнародна науково-технічна конференція «Системи і технології зв'язку, інформації та кібербезпеки: актуальні питання і тенденції розвитку» 30 листопаду 2023 р. Військовий інститут телекомунікацій та інформації імені Героїв Крут. Збірник тез доповідей Міжнародної науково-технічної конференції. Київ, 167.
3. **Ільницький А.І., Захарчук Л.В.** (2024). Математична модель одноканального пошуку сигналів засобів мобільного радіозв'язку з одним ступенем виявлення в частотно-часовій області. Київ: Військовий інститут телекомунікацій та інформації імені Героїв Крут. Збірник наукових праць "Системи і технології зв'язку, інформатизації та кібербезпеки" № 5, 71-77. https://doi.org/10.58254/viti.5.2024.067.
4. **Ільницький А.І., Цуканов О.Ф.** (2024). Математична модель одноканального пошуку сигналів з двома ступенями виявлення при радіочастотному моніторингу телекомунікаційних систем мобільного зв'язку, Київ.
5. **Ільницький А.І., Цуканов О.Ф.** (2024). Cyclic three-stage procedure for single-channel detection of mobile radio signals in the frequency-time domain, Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка, м. Київ.
6. Циклічна триступінчаста процедура одноканального пошуку сигналів мобільного радіозв'язку в частотно-часовій області. Київ.
7. **Адаменко В.К., Горбенко В.І., Ратанін Є.Г.** (2002). Методика розробки програмно-апаратних засобів демодуляції сигналів з частотною маніпуляцією. Київ: Перша міжнародна науково-технічна конференція ВІТІ НТУУ "КПІ", 35.
8. **Ільницький А.І., Стретович О.В.** (2005). Алгоритми декодування тональних сигналів в комплексах розвідки систем рухомого радіозв'язку. Труди академії, НАО МОУ, № 65, 70–75.

ЛІТЕРАТУРА

1. **Ільченко М.Ю., Кравчук С.О.** (2017). Телекомунікаційні системи. Київ: Наукова думка, 730 с.
2. **Ільницький А.І., Захарчук Л.В.** (2023). Проблемні питання радіомоніторингу джерел і об'єктів телекомунікаційних мереж і систем та загально-науковий концептуальний підхід до їх розв'язання. III Міжнародна науково-технічна конференція «Системи і технології зв'язку, інформації та кібербезпеки: актуальні питання і тенденції розвитку» 30 листопаду 2023 р. Військовий інститут телекомунікацій та інформації імені Героїв Крут. Збірник тез доповідей Міжнародної науково-технічної конференції. Київ, 167.
3. **Ільницький А.І., Захарчук Л.В.** (2024). Математична модель одноканального пошуку сигналів засобів мобільного радіозв'язку з одним ступенем виявлення в частотно-часовій області. Київ: Військовий інститут телекомунікацій та інформації імені Героїв Крут. Збірник наукових праць "Системи і технології зв'язку, інформатизації та кібербезпеки" № 5, 71-77. https://doi.org/10.58254/ viti.5.2024.0 6.7.
4. **Ільницький А.І., Цуканов О.Ф.** (2024). Моделі побудови мереж інтернету речей для управління інфраструктурою міста. Smart Technologies, 2(15), 10-17. DOI: https://doi.org/10.32347/st.2024.2.1901

Математична модель одноканального пошуку сигналів з двома ступенями виявлення при радіочастотному моніторингу телекомунікаційних систем мобільного зв'язку, м. Київ.

5. **Ільницький А.І., Цуканов О.Ф.** (2024). Cyclic three-stage procedure for single-channel detection of mobile radio signals in the frequency-time domain, Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка, м. Київ.
6. **Адаменко В.К., Горбенко В.І., Ратанін Є.Г.** (2002). Методика розробки програмно-апаратних засобів демодуляції сигналів з частотною маніпуляцією. Київ: Перша міжнародна науково-технічна конференція ВІТІ НТУУ "КПІ", 35.
7. **Ільницький А.І., Стретович О.В.** (2005). Алгоритми декодування тональних сигналів в комплексах розвідки систем рухомого радіозв'язку. Труди академії, НАО МОУ, № 65, 70–75.

Програмно-апаратні комплекси багатоетапного моніторингу радіосигналів рухомого зв'язку та особливості їх режимів роботи

Анатолій Ільницький, Олег Цуканов

Анотація. На сьогодні технічні засоби радіоперехоплення, моніторингу й радіопеленгації в мережах систем мобільного радіозв'язку реалізуються у вигляді програмно-апаратних комплексів, найважливішими показниками ефективності яких вважається швидкодія, точність визначення та імовірність розпізнавання засобів рухомого радіозв'язку з їх інформаційним наповненням. Однак, питання ефективності цих комплексів і дотепер залишаються проблематичними та потребують подальшого розвитку методів і способів пошуку та виявлення сигналів засобів рухомого радіозв'язку як у частотних, так і в часових середовищах телекомунікаційних каналів та їх подальшу інформаційну обробку.

Для вирішення вказаного проблемного питання авторами розглянуто типовий варіант структурної схеми програмно-апаратного комплексу пошуку і виявлення сигналів та особливості застосування процедур багатоступеневого пошуку і виявлення сигналів при радіомоніторингу джерел радіовипромінювання систем мобільного зв'язку. Показано, що програмно-апаратні комплекси можуть мати декілька режимів роботи залежно від вигляду каналу управління.

В статті також наведені результати розрахунків значення імовірностей успішного завершення пошуку сигналу розподіленого каналу управління засобів рухомого радіозв'язку за заданий час із трьома ступенями виявлення в частотно-часовій області з одним радіоприймальним пристроїм при різній кількості пустих каналів (p = 1, 5, 9) на одному циклі пошуку. Показано, що додаткова оцінка рівня вхідного сигналу радіоприймальним пристроїм на першому ступені виявлення підвищує ефективність пошуку при збільшенні кількості пустих каналів.

Ключові слова: програмно-апаратний комплекс, радіомоніторинг, системи мобільного зв'язку, показники ефективності, швидкодія, точність визначення, пошук, виявлення.

Основні принципи побудови моделей прийняття рішень в організаційних системах

Вікторія Клюєва[1], Олександр Симоненко[2], Наталія Зінченко[3]

[1]Київський національний університет будівництва і архітектури
пр-т Повітряних Сил, 31, м. Київ, Україна, 03037
[2],[3]Військовий інститут телекомунікацій та інформатизації імені Героїв Крут
вул. Князів Острозьких 45/1, Київ, Україна, 01011
[1]kliuieva.vv@knuba.edu.ua, https://orcid.org/0000-0003-1267-0717
[2]lokalyt@gmail.com orcid.org/0000-0001-8511-2017
[3]zinchenko.natasha@gmail.com, orcid.org/0009-0004-1837-2838

Received 03.02.2025, accepted 29.03.2025
https://doi.org/10.32347/st.2025.3.1101

Анотація. У статті розглядаються основні елементи процессу вибору (прийняття) рішень.

Наводяться різноманітні принципи класифікації управлінських рішень.

Основна увага приділена взаємопов'язаним етапам побудови математичних моделей як універсального засобу розв'язання різноманітних проблем прийняття рішень при управлінні складними організаційно-економічними системами.

Дається загальна характеристика математичних моделей, що використовуються у детермінованих задачах прийняття організаційно-економічних рішень в організаційних системах: операційних, нормативних, дескриптивних та евристичних.

Наводяться три основні підходи до побудови математичних моделей процесу розробки рішень, засновані на:

- теорії статистичних рішень;
- теорії корисності;
- теорії ігор.

Ключові слова: організаційна система, управлінське рішення, прийняття рішень, математичні моделі, ОПР (особа, яка приймає рішення).

Вікторія Клюєва
Ст. викладач кафедри кібербезпеки та комп'ютерної інженерії

Олександр Симоненко
к.т.н., доцент, начальник відділу забезпечення якості освітньої діяльності та вищої освіти

Наталія Зінченко
Старший помічник начальника відділу забезпечення якості освітньої діяльності та вищої освіти

ВСТУП

При функціонуванні організаційних систем виникають проблеми раціонального вибору рішень. За останні десятиліття важливість цих проблем значно зросла у зв'язку з тим, що

- різко зріс динамізм навколишнього середовища, тобто зменшився період часу, коли прийняті раніше рішення залишаються актуальними;

- розвиток науки та техніки привів до появи більшої кількості альтернативних

варіантів вибору (на рівні підприємств та об'єднань число документально оформлених рішень досягає у середньому

300 на рік, на більш високих рівнях їх значно більше);

- зросла складність оцінки ефективності реалізації кожного з варіантів рішень, які приймаються;

- збільшилась взаємозалежність різних рішень та їх наслідків.

У результаті дії цих тенденцій різко зросли труднощі раціонального вирішення різноманітних та складних проблем, вибору ефективних управлінських впливів в реальних умовах функціонування організаційних систем.

ОСНОВНІ ПОЛОЖЕННЯ

Прийняття рішень є центральним елементом адміністративно-управлінської діяльності, по відношенню до якого решта аспектів може розглядатися як допоміжна.

Під прийняттям рішень в організаційних системах розуміють особливий вид людської діяльності, спрямований на вибір кращої з наявних альтернатив. Таке визначення передбачає наявність трьох необхідних елементів процесу вибору (прийняття) рішень:

- проблема, яка потребує розв’язання;
- людина як колективний орган, який приймає рішення;
- декілька альтернатив, з яких здійснюється вибір.

За відсутності одного з цих елементів процес вибору рішень перестає існувати.

Розглянемо ще декілька визначень.

Управлінське рішення (УР) – це вибір однієї з можливих альтернатив впливу на керовану систему, тобто це модель, в якій із певного числа варіантів обирається кращий.

Рішення (Р) – це той пункт, у якому вибір робиться між альтернативами та, як правило, конкуруючими можливостями.

Суть розробки ПР полягає в діяльності людини, яка приймає рішення (ЛПР), по виконанню основної функції керівника в процесі управління (ще називають ОПР (особа, яка приймає рішення).

1. ОПР – це суб’єкт процесу прийняття рішень, найактивніша його складова. Це, як правило, керівник або колективний керівний орган, організація. ОПР ініціює процес прийняття рішень та завершує його.

2. Основна мета управлінського рішення – забезпечити координуючий (регулюючий) вплив на систему управління, що реалізує вирішення управлінських задач персоналом по досягненню цілей організації.

Основними задачами є створення інформаційної бази для прийняття своєчасних рішень, визначення обмежень і критеріїв ПР, організація діяльності персоналу управління.

Відносно ОПР висуваються такі характеристики: вона має право вибору з

множини альтернатив; несе відповідальність за прийняті рішення; зацікавлена у здійсненні вибору, тобто прагне розв’язати наявну проблему.

Класифікація управлінських рішень необхідна для визначення загальних і конкретно-специфічних підходів до їхньої розробки, реалізації й оцінки, що дозволяє підвищити їхню якість, ефективність і наступність. Управлінські рішення можуть бути класифіковані найрізноманітнішими способами (рис. 1).

Найбільш розповсюдженими є такі принципи класифікації:

- за функціональним змістом;
- за характером розв'язуваних задач (сфери діяльності);
- за ієрархією управління;
- за характером організації розробки;
- за характером цілей;
- за причинами виникнення;
- за вихідним методом розробки;
- за організаційним оформленням.

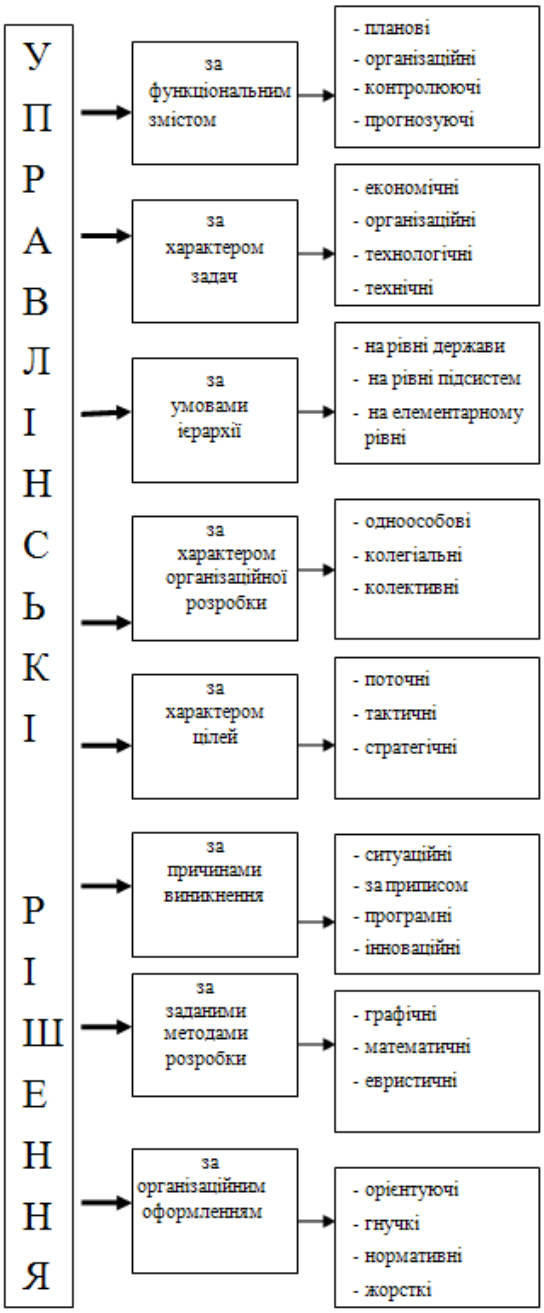


Рис. 1. Класифікація управлінських рішень

3. Альтернативні рішення. Це різні варіанти рішення, які є засобом досягнення однієї й тієї ж цілі або системи цілей. Очевидно, що процес вибору рішень має сенс тоді, коли існує більш ніж одна альтернатива. Інакше проблемивибору рішення не виникає. Можливий варіант, коли одна з множини альтернатив може полягати в тому, щоб не приймати ніякого рішення («нульове рішення») або повернути набір можливих альтернатив на доробку («рішення, що повертається»).

4. Зовнішні умови. Умови, що складаються поза системою «ЛПР- об'єкт управління», але суттєво впливають на процес прийняття рішень та результат рішення.

5. Результат (вихід). Результат від прийняття рішення має для ЛПР різну цінність в залежності від обраної альтернативи та зовнішніх умов. Для оцінки

результатів використовують такі поняття, як «корисність», «виграш», «штраф», «збиток», «прибуток» та ін.

6. Правило вибору рішення (або вирішальне правило). Воно дає можливість обрати краще в якомусь сенсі рішення. Правило вибору рішення вважається сформульованим тоді і тільки тоді, коли воно визначає алгоритм процедури вибору оптимального рішення з множини альтернатив. Сам алгоритм називається при цьому «оптимальним згідно з даним правилом вибору рішення».

7. Критерій – показник, за яким оцінюється результат діяльності системи або ефективність обраного варіанта рішення. Як правило, критерій задається у вигляді функції (або функціонала, тобто оператора, що відображає простір функцій у числову множину), аргументами якої є параметри системи. Наприклад, продуктивність агрегату, собівартість продукції.

Альтернативні рішення оцінюють з точки зору обраного критерію. Краще рішення дозволяє досягнути екстремального значення критерію.

8. Модель системи. Це відображення найбільш важливих взаємозв'язків елементів системи засобами, які допускають потім більш ефективне дослідження відображеної системи.

Після II світової війни почалась епоха застосування математичних моделей для розв'язання різноманітних проблем, які виникають в людській діяльності, особливо з появою ЕОМ. Математичні моделі на сьогодні претендують на роль універсального засобу розв'язання різноманітних проблем прийняття рішень при управлінні складними організаційно-економічними системами.

При глибокому вивченні великих проблем, що вимагають рішення, використовуються наукові методи, такі як системний аналіз, дослідження операцій. Їх основу складає математичне моделювання.

Математичне моделювання – універсальний і ефективний інструмент пізнання внутрішніх закономірностей, властивих явищам і процесам, суть якого полягає в підборі математичних схем, що адекватно описують процеси, які відбуваються реально. Воно дозволяє вивчити кількісні взаємозв'язки та взаємозалежності змодельованої системи та вдосконалити її подальший розвиток і функціонування.

Сувора формалізація соціально-економічних процесів функціонування підприємства практично неможлива. Тому всі моделі є спрощеним зображенням реальної системи, але якщо це спрощення виконано коректно, то отримане наближене відображення реальної ситуації дає змогу отримати достатньо точні характеристики досліджуваного об'єкта.

Математичне моделювання можна умовно розділити на декілька окремих і взаємозв'язаних етапів:

1) постановка задачі;
2) розробка формалізованої схеми;
3) формалізація задачі у загальному вигляді;
4) чисельне представлення моделі;
5) розв'язування задачі на ЕОМ та післяоптимізаційний аналіз отриманих результатів.

При постановці задачі виявляються закономірності процесу в теоретичному і практичному планах, його структура, умови і фактори формування.

Формалізована схема розробляється на основі вищезгаданих даних, вона менш суворо, ніж математична модель, описує процес (або явище), що моделюється. У схемі називаються конкретні показники, що відносяться до характеристики об'єкта управління. Це можуть бути шукані величини, параметри процесу,

фактори і умови, які неодмінно враховуються при виконанні розрахунків. У загальному вигляді задача представляється на основі формалізованої схеми. Однак існуючі залежності конкретизуються. Далі елементи моделі набувають кількісного вираження, модель перевіряється та у разі необхідності уточнюється. На базі використання обчислювальної техніки прораховується ефективність варіантів за заданим критерієм оцінки, і на цій основі

визначається оптимальний варіант розв'язання задачі.

При побудові математичної моделі виконуються такі види робіт, як:

- складання переліку всіх елементів системи, що впливають на ефективність її функціонування;
- розгляд міри впливу кожного з елементів переліку на функціонування організації при різних варіантах рішень;
- елементи, що не впливають на вибір варіантів рішень або вплив яких незначний, виключаються з переліку і не враховуються при побудові моделі;
- щоб спростити модель, потрібно заздалегідь, по можливості, згрупувати деякі взаємопов'язані елементи (наприклад, витрати по оренді, утриманню приміщень тощо і об'єднати їх в умовно постійні витрати);
- після уточнення переліку елементів визначається їх постійний або змінний характер впливу на систему. У складі змінних елементів встановлюються, в свою чергу, піделементи системи, що впливають на їх величину. Наприклад, транспортні витрати залежать від обсягу переміщених товарів, відстані, вартості пального та ін.;
- за кожним піделементом закріплюється певний символ і далі складається рівняння або система рівнянь.

Операційні моделі рішень мають вигляд рівняння або системи рівнянь. Вони можуть бути складними, з математичної точки зору, але структура їх досить проста. Наприклад, операційні моделі, що часто використовуються, мають вигляд:

$$E = f(x_i, y_i),$$

де Е – міра загальної ефективності;

f – функція, що задає співвідношення між Е, хі, уі;

хі –керовані змінні, що визначають поведінку системи;

уі – некеровані змінні, що визначають поведінку системи.

Керованими змінними (хі) є чинники, на які може впливати ЛПР. До них відносяться: чисельність працівників, кількість обладнання, технології виробництва продукції та ін. Деякі керовані змінні можуть мати обмеження, що потрібно враховувати в ході побудови моделі.

Після встановлення переліку змінних чинників визначається значущість кожного з них.

Некерованими змінними (уі) вважаються чинники, на які ОПР не може впливати. Це дії споживачів, постачальників, постанови державних органів тощо.

Оптимальне рішення по даній моделі визначається шляхом пошуку значень керованих чинників (хі), при яких міра загальної ефективності Е буде максимальною (або мінімальною, якщо за міру ефективності прийнятий показник витрат на виробництво, втрати).

Моделювання як метод розробки управлінських рішень використовується з середини XX ст., перші моделі базувалися на нормативних теоріях, називалися нормативними.

У них описується стратегія поведінки при виробленні рішення, яка орієнтується на заданий критерій. Прикладом нормативних моделей є:

- моделі прийняття статистичних рішень з використанням теорії ймовірності та математичної статистики;
- інноваційні ігри як варіант нормативної моделі поведінки в умовах конфлікту, наявності суперечливих думок з проблем нововведення;
- моделі розробки рішень на основі теорії масового обслуговування, що містить нормативні критерії при розв'язанні конкретних задач.

Зміст процесу розробки рішення в цьому випадку зводиться до пошуку оптимального рішення, в найбільшій мірі відповідного заданому критерію. Досягається це зіставленням альтернатив рішень, розрахованих для конкретних станів змінних факторів (умов зовнішнього середовища).

Однак нормативні моделі не враховують при прийнятті рішень реальної поведінки людини, за якою залишається вибір остаточного варіанта. Цей недолік певною

мірою компенсують дескриптивні моделі розробки рішень, засновані на теорії корисності, теорії ризику.

На даний час виділяється три основних підходи до побудови математичних моделей процесу розробки рішень, заснованих на:

-теорії статистичних рішень;

-теорії корисності;

-теорії ігор.

У найбільш розроблених моделях на основі теорії статистичних рішень вважаються заданими:

-можливий розподіл випадкового процесу, що вивчається;

-простір можливих остаточних рішень;

-вартість варіантів рішень;

-функція можливого збитку для кожного рішення, відповідного певному стану зовнішнього середовища.

У загальному вигляді можна констатувати, що рішення приймаються, виходячи з максимуму прибутку або мінімуму втрат. У зв'язку з цим вводиться поняття ризику, по величині якого судять про цінності рішення. У цій теорії розглядається ряд можливих критеріїв оптимальності рішень, щол приймаються. Так, рішення, що мінімізує максимальний ризик (байєсовське рішення), описується як мінімаксне рішення.

Теорія статистичних рішень застосовується при виборі рішень в умовах невизначеності зовнішнього середовища.

Другий напрям математичного моделювання пов'язаний з використанням теорії корисності, заснованої на індивідуальних перевагах, суб'єктивній оцінці ймовірностей настання подій зовнішнього середовища.

Третій напрям моделей розробки рішень засновано на використанні теорії ігор. Дана теорія застосовується в умовах конфліктних ситуацій або при прийнятті колективних (спільних) рішень. Основоположним є вибір відправної точки гарантуючого рішення, з якого починається спільне вироблення кращого рішення. Основний принцип цієї теорії – мінімакс. Схема теорії ігор описує принципи прийняття рішень для широкого класу практичних ситуацій інноваційного характеру. Гра можлива з будь-якою кількістю учасників і різною мірою їх інформованості; формалізації зазнають лише правила гри, а не поведінка гравців.

Приведені теорії та підходи до моделювання процесу розробки рішень відображають певні його сторони:

- статистична теорія рішень – невизначеність середовища, вибір, ризик;

- теорія ігор – деякі характеристики поведінки людини в умовах взаємодії з іншими людьми і з середовищем;

- теорія корисності – психологічні уявлення про потреби людини і їх мотивацію.

Різновидом розробки рішень є евристичні моделі (від грецького «еуріскеін» - роблю відкриття), які характеризують особливий підхід до розв'язування задач і вибору рішень. Основу евристичних моделей складають логіка і здоровий глузд, засновані на існуючому досвіді. Такі моделі використовуються в ситуаціях, коли неможливе застосування формальних аналітичних методів. Суть евристичних методів полягає у перетворенні однієї складної задачі в сукупність простих, що піддаються вивченню математичними способами.

Евристичними моделями не вирішуються задачі оптимізації рішень, але оцінюється відносна придатність конкретних стратегій з певними обмеженнями. На основі побудови моделі логічних зв'язків у ході міркувань ЛПР може розв'язуватися широкий клас задач. Евристичні моделі використовуються при виборі рішень простих і складних, в яких не існує надії на використання при цьому математичного апарату. Практичне застосування евристичного підходу до моделювання процесу розробки та прийняття управлінських рішень передбачає наявність у ЛПР пізнавальних здібностей і схильностей до узагальнень і висновків.

ВИСНОВКИ

Використання математичних методів у прийнятті рішень дає можливість здійснювати комплексний аналіз об'єктивних зв'язків між явищами, їх

раціональний і наочний опис, встановлювати міру впливу одних факторів на інші при їх зміні. У результаті моделювання та оптимізації вони дозволяють своєчасно підключати додаткові ресурси у виробничий процес.

Математичні моделі на сьогодні претендують на роль універсального засобу розв'язання різноманітних проблем прийняття рішень при управлінні складними організаційно-економічними системами. Вони дозволяють оптимізувати рівень витрат ресурсів при виконанні будь-яких процесів (підготовчих, постачальницько-збутових, логістичних, виробничих тощо).

REFERENCES

1. **Dotsenko S. I., Savenko V. I., Bazylenko S. O., Kliuieva V. V., Palchyk S P., Hihineishvili D. Ya.** (2018). Intelektualni informatsiini tekhnolohii u pryiniatti efektyvnykh rishen v upravlinni pidpryiemstvom. Upravlinnia rozvytkom skladnykh system, 34, 161–169.
2. **Hevko I. B.** (2009). Metody pryiniattia upravlinskykh rishen: book. Kyiv, Kondor, 187.
3. **Izmailova O. V.** (2002). Metody pryiniattia bahatokryteriinykh rishen v informatsiinykh systemakh: book. Kyiv, 111.
4. **Klebanova T.S., Kyzym M.O., Cherniak O.I.** (2009). Matematychni metody i modeli rynkovoi ekonomiky: book. Kharkiv, «Inzhek» Publ., 456.
5. **Pavlenko P.M., Filonenko S.F., Cherednikov O.M., Treitiak V.V.** (2017). Matematychne modeliuvannia system i protsesiv: navch. posib. Kyiv, NAU Publ., 392.
6. **Savenko V.I., Shatrova I.A., Fialko N.M., Honcharenko T.A.** (2022). Doslidzhennia i matematychne modeliuvannia orhanizatsiinykh struktur ta intelektualni informatsiini instrumenty v orhanizatsii i upravlinni budivnytstvom: monohrafiia. 2-he vyd. vypr. ta dop. Kyiv, Tsentr uchbovoi literatury Publ., 148.
7. **Savenko V.I., Dotsenko S.I., Kliuieva V.V., Palchyk S.P., Tereshchuk M.O.** (2019). Intelektualni informatsiini instrumenty rozvytku vyrobnychykh system, enerhetychnoho menedzhmentu ta pidpryiemstva v tsilomu. Upravlinnia rozvytkom skladnykh system, 37, 195–204.
8. **Savenko V.I., Nesterenko I.S., Kliuieva V.V.** (2019). Upravlinnia konkurentospromozhnistiu pidpryiemstva v suchasnykh umovakh hospodariuvannia. Shliakhy pidvyshchennia efektyvnosti budivnytstva v umovakh formuvannia rynkovykh vidnosyn, 42, 66–76.
9. **Tsiutsiura S.V., Kryvoruchko O.V., Tsiutsiura M.I.** (2012). Teoretychni osnovy ta sutnist upravlinskykh rishen. Modeli pryiniattia upravlinskykh rishen. Upravlinnia rozvytkom skladnykh system, 9, 50–58.

Basic principles of building decision-making models in organizational systems

VIktoriia Kliuieva, Oleksandr Symonenko, Nataliia Zinchenko

Abstract. The article examines the key elements of the decision-making process. Various principles of managerial decision classification are presented.

Special attention is given to the interrelated stages of constructing mathematical models as a universal tool for solving diverse decision-making problems in managing complex organizational and economic systems.

A general overview of mathematical models used in deterministic organizational and economic decision-making tasks is provided, including operational, normative, descriptive, and heuristic models.

Three main approaches to building mathematical models for decision development are outlined, based on:

- statistical decision theory;
- utility theory;
- game theory.

Keywords: organizational system, managerial decision, decision-making, mathematical models, decision-maker (DM).

Publication statement
in an international scientific journal

_____

(journal name)

Form 1

INFORMATION ABOUT THE ARTICLE

	Author(s) (name, surname − *in the language of the article*) (subject area) / (number of pages)	**Documents (+/−)**							**Article Title**	**Order** (number of copies)
		About the article	About the authors	Two reviews (*F.3*)	Exp. conclusion	Agreement (*F.5*)	Approved	Recommendation	**Reviewers** (name, surname, scientific degree, academic status − *in the language of the article*)	Payment (€) Date Signature
									Rev.1: *Rev.2*:	

Form 2

INFORMATION ABOUT THE AUTHORS
(*in the language of the article*)

Data			Contacts	
h-index (*Scopu*	Full Name scientific degree, academic status	Place of work, position, address, index	The author's photo (.jpg, 300 dpi)	Cell-phone e-mail ORCID *Scopus Author ID* *Researcher ID*

* – indicate availability by sign

Form 3

REVIEW
(two, external)

Must contain:
1) *Title of the article*, *Surname of the author(s)*
2) *Evaluation of the work* (originality, correspondence with the title and text of the article, methods and purpose of work, terminology, style of presentation, grammar)
3) Information about *the Quality of the article* (content, translation in English) and the lack of *Plagiarism*
4) *Remarks* and *Adjustments* (or indication of the necessity to transfer the article to another reviewer)
5) *Recommendation* (for publication, further elaboration, re-review, refusal to publish)
6) Information about the *Reviewer* (name, surname, degree, academic rank, place of work in the language of the article)
7) *Date*, *Signature* (approved)

Form 4

(Sample)

APPROVED

(position, scientific degree,
academic status)

Full Name

« ___ » _____ 202 ___

EXPERT CONCLUSION No ____

on the possibility of publishing materials
in the press and other sources of information

Kyiv National University of Construction and Architecture Expert Commission, having reviewed the materials

_____
(name, surname of the author(s))

_____
(Article Title)

_____

_____

which consists of ____ pages, notes that they do not have any information that would be a subject to publication ban in accordance with "The Expanded list of information containing the state secret in the Ministry of Education and Science of Ukraine − 2001"

Conclusion: these materials are allowed to be published openly

The Head of the Expert Group

Full Name

Form 5

AGREEMENT No _____

on free use of copyright in the *Smart Technologies: Industrial and Civil Engineering* periodical Edition

Kyiv « ___ » _____ 20 ___

The Editorial Board of the *Smart Technologies: Industrial and Civil Engineering* journal, the founder of which is the Kyiv National University of Construction and Architecture, having the legal address: KNUCA, Povitrianykh Syl Ave, 31, Kyiv, Ukraine, 03037, in the person of the chief editor, doctor of technical sciences, professor M. Sukach, on the one hand (hereinafter − Editorial), and the owner(s) of the property rights and copyright in the person of

_____ _____ _____,
(Full Name) (scientific degree, academic status)..... (ORCID)

_____ _____ _____,
(Full Name) (scientific degree, academic status)..... (ORCID)

_____ _____ _____,
(Full Name) (scientific degree, academic status)..... (ORCID)

on the other hand (hereinafter − the Author(s)), who altogether referred to as the Parties, guided by the Civil Code, the Law of Ukraine "On Copyright and Related Rights", other legislative and regulatory acts of Ukraine, entered into this Agreement as follows.

§ 1

1.1. The author (s) declare that they are valid writers of the Scientific Work / Article titled as

_____

_____
(original language)

which is the result of their joint creative work and that they have the exclusive copyright in relation to the above mentioned work.

1.2. The author (s) declare that the Work/Article does not violate the copyrights of any third party, does not contain any borrowings (plagiarism), and there are no other circumstances that may cause the Editorial Board to be liable to any third party as a the result of the use or publication of the Work/Article.

1.3. The author (s) declare that they have the right to dispose of the materials contained in the Work/Article, including texts, photographs, maps, plans, etc., and that the use of these materials in the Work/Article does not violate the rights of the third party.

1.4. The author (s) confirm that they are familiar with the article design requirements. The text of the Work/Article has been prepared in accordance with the editorial requirements regarding the publication in the *Smart Technologies: Industrial and Civil Engineering* periodical.

1.5. The author (s) declare that the Work/Article has not been previously published in full or in part (or under the same or another name) and that it has not been transmitted for publication to any other periodical publication in accordance with the Law of Ukraine "On Copyright and Related Rights".

§ 2

2.1. The author (s) provide the Editorial with free and no restrictions on the territory, time and number of copies full copyright of the Work/Article for the purpose of its publication in the *Smart Technologies: Industrial and Civil Engineering* periodical in printed and electronic form, with the following terms of use:

a) storage on any media devices;

b) reproduction of the Work/Article, its parts or fragments by any known methods, copying Work/Article or parts and fragments thereof by any technique, including printing, risography, magnetic recording and digitization;

c) saving on a computer and placing in private and public computer networks (including the Internet) and distribution over the networks;

d) dissemination of the original and / or copies of the Work/Article, its separate parts or fragments, distribution and transfer to the use of the original or copies thereof.

2.2. The author(s) agree that the editorial work and duplication of the periodical publication shall be made with the voluntary contributions of the Publishers.

§ 3

3.1. The author(s) and the Editorial Board agree that the Editor will also be entitled to:

a) make the necessary design of the Work/Article on the results of the editorial work;

b) determine independently the number of publications, the printing of additional copies and the circulation of the Work / Article, the number of copies of individual editions and additional copies;

c) the publication of the Work/Article in other editions related to the activities of the Editor than those specified in paragraph 2.1.

§ 4

4.1. The author (s) and Editorial hereby state that free use of the copyright under this agreement is free of charge.

4.2. Any changes to this Agreement must be made in hard copies.

4.3. Questions not regulated by the provisions of this Agreement shall be subject to the rules of the Civil Code and the Law of Ukraine "On Copyright and Related Rights".

4.4. Any disputes that may arise during the execution and during the term of this Agreement will be resolved within the territorial jurisdiction of the place where the Editorial is located.

The Agreement is made in 2 (two) identical copies − one to each of the Parties.

AUTHOR(S):

1. _____ _____
(Signature) (Full Name)

2. _____ _____
(Signature) (Full Name)

3. _____ _____
(Signature) (Full Name)

EDITORIAL BOARD:

International scientific journal
Smart Technologies:
Industrial and Civil Engineering

Povitrianykh Syl Ave, 31, 31, Kyiv,
Ukraine, 03037
cybersec@knuba.edu.ua

_____ Y. Khlaponin

Scientific Edition

SMART TECHNOLOGIES:

INDUSTRIAL AND CIVIL ENGINEERING

Issue 3 (16), 2025

The journal is indexed in world databases

Google Academy	http://scholar.google.com.ua	CrossRef	https://www.crossref.org
Index Copernicus	www.journals.indexcopernicus.com	Ulrichs Web	http://ulrichsweb.serialssolutions.com

The original layout is made in the editorial office of the journal Smart Technologies: Industrial and Civil Engineering

Responsible for the issue	*Dmytro Mischuk*
Linguistic consultant	*Anastasiia Khlaponina*
Computer layout	*Denis Dolgopolov*
Editing and proofing	*Dmytro Khlaponin*
Layout and cover	*Anton Khaddad*

Journal Editorial Staff

KNUCA, Povitrianykh Syl Ave, 31, Kyiv, Ukraine, 03037,
Labor. corp., of. 361, Kyiv, Ukraine, 03037
+38 044 244 96 45, +38 050 311 0373
http://uwtech.knuba.edu.ua/, cybersec@knuba.edu.ua

Publisher and manufacturer

LLC “Lira-K Publishing”
Certificate No. 3981, DK series.
03142, Kyiv, str. V. Stusa 22/1
Phone (050) 462-95-48; (067) 820-84-77
Website: lira-k.com.ua, editor: zv_lira@ukr.net

Signed for printing on 18.06.2024 Format 60×84 1/8
Paper offset. Printing offset. Times New Roman
Cond. print. sheets 11,16. Circulation 100 copies