ISSN 2788-5518

ІНФОКОМУНІКАЦІЙНІ ТА КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ

№ 2 (08), 2024

НАУКОВИЙ ЖУРНАЛ

Видавничий дім
«Гельветика»
2024

ISSN 2788-5518

INFOCOMMUNICATION AND COMPUTER TECHNOLOGIES

№ 2 (08), 2024

SCIENTIFIC JOURNAL

Publishing House
“Helvetica”
2024

Засновник, видавець та виробник журналу: «Відкритий міжнародний університет розвитку людини «Україна».

Реєстрація суб'єкта у сфері друкованих медіа: Рішення Національної ради України з питань телебачення і радіомовлення № 2317 від 11.07.2024 року

Журнал включений до переліку наукових фахових видань України категорія «Б», додаток 4 до наказу МОН України № 735 від 29.06.2021 р. та додаток 2 до наказу МОН України № 320 від 07.04.2022 р.

Журнал відповідає вимогам наказу МОН України № 32 від 15.01.2018 р.

Має міжнародні коди: ISSN 2788-5518 та DOI 10.36994/2788-5518.

Журнал індексується у міжнародних базах та каталогах Google Scholar, Index Copernscus, Vernadsky National Library of Ukraine.

Мова видання: українська, англійська

Рекомендовано до друку вченою радою відкритого міжнародного університету розвитку людини «Україна» (протокол № 8 від 26.12.2024 р.)

URL-адреса видання: visn-icct.uu.edu.ua

Тематика журналу:

- Теоретичні основи інфокомунікацій та комп’ютерних наук
- Інфокомунікаційні системи
- Безпроводові технології
- Технічна електродинаміка та антени
- Волоконно-оптичні системи
- Радіорелейні та супутникові системи
- Телевізійні системи
- Системи відеоконференцзв’язку та відеоспостереження
- Комп'ютерні системи та мережі
- Комп'ютерна і програмна інженерія
- Системи SCADA
- Захист інформації та кібербезпека
- Прикладна математика та моделювання
- Метрологія та інформаційно-вимірювальні системи
- Медична та екологічна електроніка
- Автоматизація управління технологічними процесами

В журналі публікуються статті для захисту дисертацій зі спеціальностей: 121 Інженерія програмного забезпечення. 122 Комп’ютерні науки. 123 Комп’ютерна інженерія. 125 Кібербезпека та захист інформації. 151 Автоматизація та комп’ютерні інтегровані технології. 152 Метрологія та інформаційно-вимірювальна техніка. 172 Телекомунікації та радіотехніка

Founder, publisher and producer of the journal: "Open International University of Human Development" Ukraine".

Registration of Print media entity: Decision of the National Council of Television and Radio Broadcasting of Ukraine No. 2317 as of 11.07.2024

The journal is included in the list of scientific professional publications of Ukraine – category "B", Annex 4 to the order of the Ministry of Education and Science of Ukraine № 735 from 29.06.2021 and Annex 2 to the order of the Ministry of Education and Science of Ukraine № 320 from 07.04.2022

The journal meets the requirements of the order of the Ministry of Education and Science of Ukraine № 32 from 15.01.2018.

It has international codes: ISSN 2788-5518 and DOI 10.36994 / 2788-5518.

The journal is indexed in international databases and directories of Google Scholar, Index Copernicus, Vernadsky National Library of Ukraine

Language of publication: Ukrainian, English

Recommended for publication by the Academic Council of the Open International University of Human Development "Ukraine" (Protocol № 8 dated 26.12.2024)

Publication URL: visn-icct.uu.edu.ua

Subject journal:

- Theoretical foundations of infocommunications and computer science
- Infocommunication systems
- Wireless technology
- Technical electrodynamics and antennas
- Fiber optic systems
- Radio relay and satellite systems
- Television systems
- Video conferencing and video surveillance systems
- Computer systems and networks
- Computer and software engineering
- SCADA systems
- Information security and cybersecurity
- Applied mathematics and modeling
- Metrology and information-measuring systems
- Medical and environmental electronics
- Automation of process control

The journal publishes articles for the defense of dissertations in the following specialties: 121 Software engineering. 122 Computer science. 123 Computer engineering. 125 Cybersecurity. 151 Automation and computer integrated technologies. 152 Metrology and information and measurement technology. 172 Telecommunications and radio engineering

ЗМІСТ

ІНФОКОМУНІКАЦІЙНІ ТЕХНОЛОГІЇ

КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ

CONTENTS

INFOCOMMUNICATION TECHNOLOGIES

COMPUTER TECHNOLOGIES

ІНФОКОМУНІКАЦІЙНІ ТЕХНОЛОГІЇ

УДК 004.624
DOI https://doi.org/10.36994/2788-5518-2024-02-08-01

СЦЕНАРІЙ ВИКОРИСТАННЯ АНАЛІТИКИ КЛІНІЧНОГО ІНТЕРНЕТУ РЕЧЕЙ (IOT)

Кирсанов О. О., *Харківський національний університет радіоелектроніки, Харків, Україна. https://orcid.org/ 0009-0001-5987-4728, oleksandr.kyrsanov@nure.ua*

Кривенко С. А., *к.т.н., Харківський національний університет радіоелектроніки, Харків, Україна. https://orcid.org/0000-0002-5244-1276, stanislav.kryvenko@nure.ua*

Анотація. *Стаття присвячена сценарію використання аналітики Інтернету речей (IoT) для збору та аналізу даних з електровелосипедів, обладнаних IoT-пристроями, з метою підвищення фізичної активності населення, що сприятиме профілактиці серцево-судинних захворювань. Запропонована архітектура системи базується на AWS і охоплює три ключові рівні: прийом та обробку даних, їхнє зберігання, а також аналіз і візуалізацію результатів.*

На першому етапі підключення IoT-пристроїв до AWS IoT Core забезпечує передачу даних з електровелосипедів через протокол MQTT. Дані регулярно надходять до Kinesis Data Firehose, де вони можуть бути збагачені функціями Lambda, які додають до них додаткову інформацію, зокрема геолокаційні дані про користувачів. Це забезпечує більш точний і деталізований аналіз зібраних даних.

Другий етап включає налаштування OpenSearch Service для зберігання та індексації отриманих даних, що дозволяє проводити детальний аналіз за допомогою інструментів OpenSearch Dashboards. У результаті створюються різноманітні візуалізації, такі як кругові діаграми, теплові карти, гістограми та інші графічні інструменти, які дозволяють не лише відстежувати використання електровелосипедів, а й проводити аналіз впливу програми на фізичну активність громадян, а також визначати ключові фактори, що впливають на ефективність цієї програми.

На завершальному етапі результати аналізу надаються зацікавленим сторонам через зручні інформаційні панелі, доступ до яких забезпечується за допомогою Amazon Cognito та AWS Identity and Access Management (IAM). Це гарантує безпечний і надійний доступ до даних, що є особливо важливим для прийняття обґрунтованих управлінських рішень. Запропонована архітектура демонструє значний потенціал IoT у клінічній аналітиці, сприяючи ефективному використанню зібраних даних для покращення здоров'я населення. Таким чином, впровадження цього підходу підвищує ефективність управління програмами фізичної активності, а також сприяє запобіганню захворюванням, що дозволяє суттєво покращити рівень громадського здоров'я та зробити цей підхід особливо актуальним у сучасних умовах швидкого розвитку технологій.

Ключові слова: *IoT, мобільний пристрій, смартфон, адаптивний кейс-менеджмент, структуровані та неструктуровані дані.*

A USE CASE FOR INTERNET OF THINGS (IOT) ANALYTICS

Oleksandr Kyrsanov, *Kharkiv National University of Radio Electronics, Kharkiv, Ukraine. https://orcid.org/0009-0001-5987-4728, oleksandr.kyrsanov@nure.ua*

Stanislav Kryvenko, *Ph.D., Ass. Prof., Kharkiv National University of Radio Electronics, Kharkiv, Ukraine. https://orcid.org/0000-0002-5244-1276, stanislav.kryvenko@nure.ua*

Abstract. *The article is dedicated to a scenario of using Internet of Things (IoT) analytics for collecting and analyzing data from electric bicycles equipped with IoT devices, aiming to enhance public health by increasing physical activity, which contributes to the prevention of cardiovascular diseases. The proposed system architecture is based on AWS and encompasses three key levels: data ingestion and processing, storage, and analysis with visualization of the results.*

The first stage involves connecting IoT devices to AWS IoT Core, enabling the regular transmission of data from electric bicycles via the MQTT protocol. This data is transmitted to Kinesis Data Firehose, where it can be further enriched using Lambda functions, which add supplementary information, such as geolocation data about users. This enrichment process ensures a more accurate and detailed analysis of the collected data.

The second stage involves configuring OpenSearch Service for storing and indexing the received data, allowing for in-depth analysis using OpenSearch Dashboards tools. Consequently, various visualizations are created, such as pie charts, heatmaps, histograms, and other graphical tools, which not only help track bicycle usage but also analyze the impact of the program on citizens' physical activity. Additionally, these visualizations help identify key factors influencing the program's effectiveness, providing valuable insights for optimization.

In the final stage, the analysis results are presented to stakeholders through user-friendly dashboards, with access secured via Amazon Cognito and AWS Identity and Access Management (IAM). This ensures safe and reliable data access, which is crucial for making informed management decisions. The proposed architecture highlights the significant potential of IoT in clinical analytics, contributing to the effective use of collected data to improve public health. Thus, implementing this approach enhances the efficiency of managing physical activity programs, as well as contributes to disease prevention, significantly improving the level of public health and making this approach especially relevant in the context of rapid technological advancement.

Key words: *IoT, mobile device, Adaptive Case Management, security, data protection, UI/UX, structured and unstructured data.*

Вступ

У цій роботі розглядається сценарій використання аналітики Інтернету речей (IoT). Дослідження пропонує архітектуру для підтримки рівня прийому та обробки, рівня зберігання та рівня аналізу та візуалізації.

Проблема бізнесу

Сценарій, представлений тут, відповідає одному варіанту планування конвеєра. Важливо працювати у зворотному напрямку від бізнес-проблеми, щоб визначити, який тип конвеєра потрібно створити.

Робота в зворотному напрямку від бізнес-проблеми

Стартап-компанія з електровелосипедів співпрацює з місцевою владою, щоб запровадити пілотну програму прокату електровелосипедів у кількох приміських районах.

Мета влади – профілактика серцево-судинних захворювань на основі підвищення фізичної активності громадян.

На цьому етапі їм потрібно зібрати й проаналізувати тенденції у даних про те, як використовуються велосипеди. Їм також потрібно визначити, які дані вони хочуть відстежувати, розширюючи програму.

Вони знають, що вони хочуть дивитися на те, де подорожували велосипеди, час роботи акумулятора, рівні використання двигуна (вимкнено, низький, середній, високий) та життєві показники велосипедистів.

Кожен із 50 велосипедів оснащений пристроєм Інтернету речей (IoT), який живиться від акумулятора велосипеда та використовує бездротову телефонну мережу для передачі даних. Пристрої використовують протокол повідомлень MQTT для надсилання та отримання даних JSON. Наразі пристрої налаштовані на надсилання повідомлення кожні 2 хвилини [6].

Потрібно створити конвеєр для підтвердження концепції, щоб збирати й аналізувати дані з пристроїв Інтернету речей, вбудованих в електровелосипеди.

Які ключові характеристики описані в бізнес-проблемі, яка керуватиме рішеннями, які ви приймаєте для створення конвеєра?

Виконання досліджень

Еталонна архітектура IoT

Рисунок 1 ілюструє пропоновану архітектуру AWS для програми для підтвердження концепції.

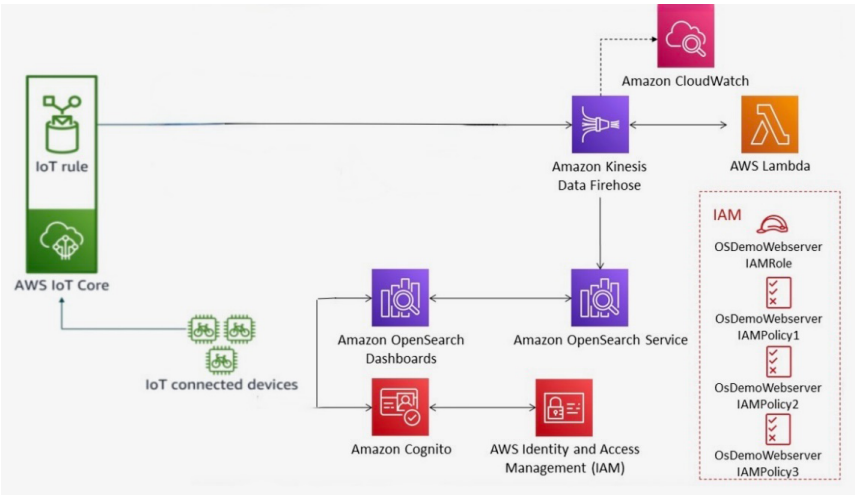


Рис. 1. Пропонована архітектура AWS

Пристрої IoT на велосипедах надсилатимуть повідомлення до AWS IoT Core. Правило IoT трансформує повідомлення та надсилає їх у потік доставки Amazon Kinesis Data Firehose. Потік доставки налаштовано як підписник на тему та може виконувати додаткові перетворення, щоб підготувати дані для аналізу. Kinesis Data Firehose забезпечує захоплення потокових даних з журналів вебсервера та їх передачу до OpenSearch Service для подальшого аналізу. Потік доставки налаштований на використання джерела даних «Direct PUT» та місця призначення «Amazon OpenSearch Service». Lambda обробляє та збагачує дані в режимі реального часу, додаючи додаткову інформацію, таку як географічне розташування відвідувачів на основі їхніх IP-адрес. Функція Lambda «os-demo-lambda-function» пов'язана з потоком доставки Kinesis Data Firehose і використовується для трансформації журналів доступу перед їх передачею до OpenSearch Service. OpenSearch Service індексує та зберігає дані, даючи можливість аналізувати їх за допомогою OpenSearch Dashboards. За допомогою Dashboards можна створювати різні візуалізації для відображення даних у зручному форматі [3].

Прийом та переробка

По-перше, розглядається рівень прийому та обробки.

Підключення пристроїв до AWS IoT Core

Схема підключення пристроїв до AWS IoT Core і початку публікації повідомлень на визначену тему наведена на рисунку 2.

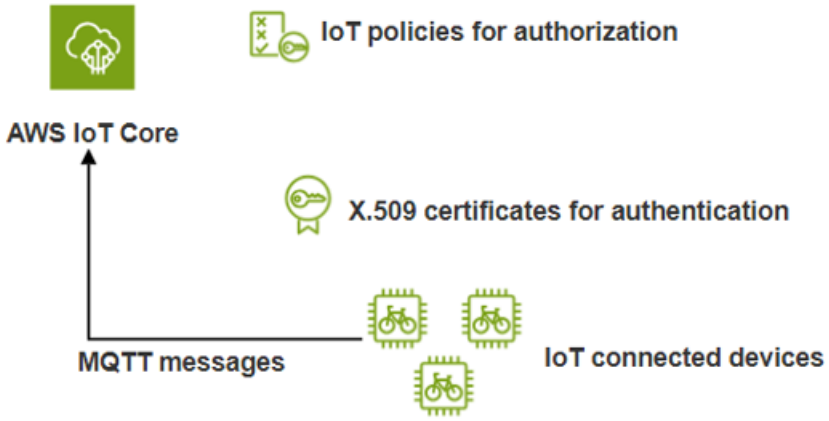


Рис. 2. Підключення пристроїв до AWS IoT Core

Для цього прикладу темою є ebike/stats. Підключалися фактичні пристрої для тестування та використовувались тестові дані із симулятором пристрою, щоб перевірити потік. В AWS IoT Core використовувались параметри в меню «Підключити», щоб підключити один або багато пристроїв. Це передбачає створення об'єктів-речей в AWS IoT Core для представлення пристроїв. Також налаштовані сертифікат безпеки, який контролює доступ до пристрою, і IoT-політики, необхідні для контролю доступу до ресурсів AWS IoT. За допомогою тестового клієнта MQTT у AWS IoT Core підписалися на тему ebike/stats, щоб переконатися, що повідомлення з пристрою публікуються в темі. Під час використання пристроїв IoT задіяні компоненти безпеки. Як правило, пристрої використовують сертифікати X.509 для захисту доступу між пристроями IoT. AWS IoT Core використовує ці сертифікати для захисту зв'язку між пристроями та AWS IoT Core, а також для автентифікації пристроїв IoT. Політики AWS IoT Core надають дозволи для авторизації дій у AWS IoT Core, наприклад підключення до брокера повідомлень. Основні політики AWS IoT є документами JSON і дотримуються тих самих умов, що й політики AWS Identity and Access Management (IAM). Як політики AWS IoT Core, так і політики IAM використовуються з AWS IoT Core для керування операціями, які може виконувати особа (також звана принципалом). Політики IoT використовуються для авторизації пристроїв із сертифікатами X.509, а політики IAM додаються до користувача, групи або ролі IAM [5].

Приклад: повідомлення MQTT

На рис. 3 показано приклад того, як повідомлення MQTT, надіслане до теми ebike/stats, може надати атрибути для такого пристрою IoT.

Налаштування правила IoT для маршрутизації та фільтрації повідомлень

Коли дані пристрою публікуються та пристрої успішно підключилися до AWS IoT Core, було створено правило (рис. 4) для вибору повідомлень і їх маршрутизації передплатникам.

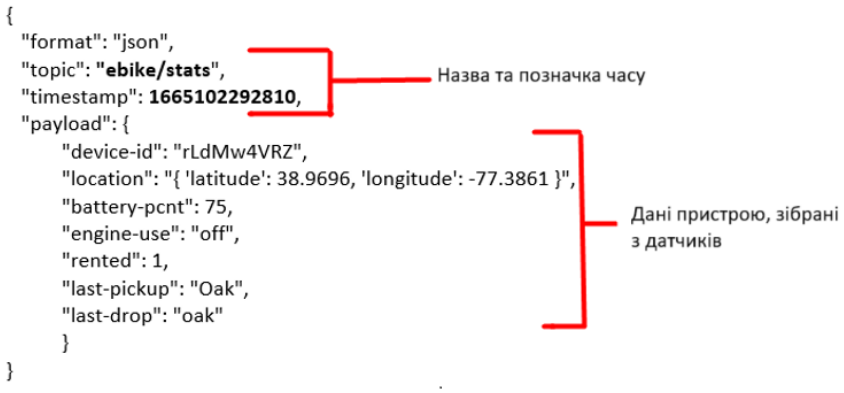


Рис. 3. Повідомлення MQTT

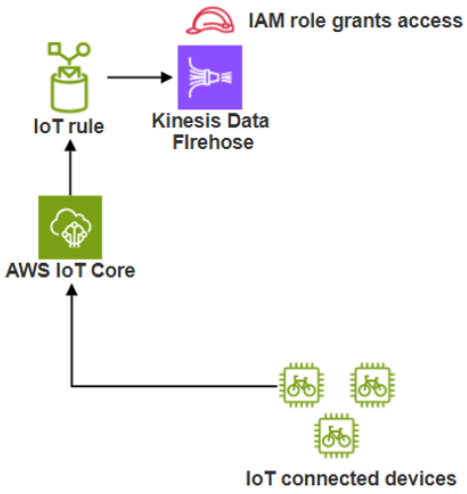


Рис. 4. Правило IoT

У цьому прикладі правило направляло повідомлення з теми ebike/stats до потоку доставки Kinesis Data Firehose. Роль IAM використовувалась для надання дозволів правилу IoT для запису даних у потік доставки.

Приклад: правило IoT

У цьому прикладі (рис. 5) наведені компоненти, які становлять правило IoT – інструкцію SELECT і дію, яку потрібно виконати. Інструкція SELECT визначає, які атрибути даних слід включити.

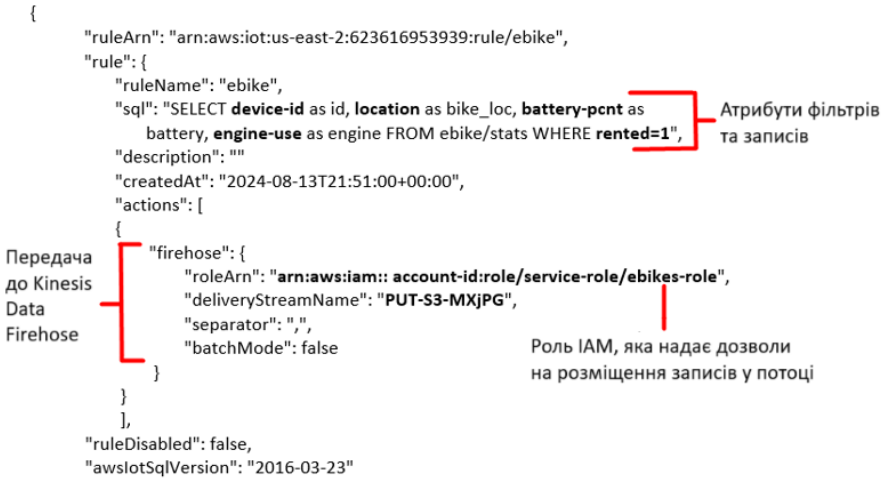


Рис. 5. Компоненти правила IoT

Речення FROM визначає відповідну тему MQTT, і речення WHERE використовується, щоб фільтрувати, які записи слід включити. Розділ дій вказує правилу, куди направляти результати запиту. У цьому прикладі оператор SELECT обмежує атрибути, які потрібно включати (ідентифікатор пристрою, розташування, акумулятор-pcnt і двигун-використання) і зіставляє їх з різними іменами полів. Правило також може доповнювати дані, додаючи додаткові атрибути [1]. Наприклад, було включено поле для ідентифікації цих записів як частину пілотної програми, оновлено оператор SELECT, щоб включити «пілотний етап як стадію» після «двигун-використання як механізм». Речення FROM вказує на тему запитувати – ebike/stat у цьому прикладі. Речення WHERE обмежує обсяг записів лише тими, де значення rented дорівнює 1.

Дії в цьому прикладі повідомляють правилу про доставку результатів запиту до потоку доставки Kinesis Data Firehose, який називається PUT-S3-MXjPG. У цьому прикладі користувач вирішив створити потік доставки як частину створення правила в консолі керування AWS. Це полегшило створення потоку та дозволів IAM, необхідних для доступу до потоку.

Потік доставки

Рішення використовує Kinesis Data Firehose для завантаження потокових журналів доступу до вебсервера, а потім використовує функції Lambda для збагачення даних [2]. Після використання Kinesis та Lambda для обробки даних використовувався OpenSearch Service для завантаження та індексації даних. Використовуючи OpenSearch Dashboards, створюються візуалізації даних, щоб отримати уявлення про стан користувачів електровелосипедів. Візуалізації поширюються серед зацікавлених сторін, використовуючи Amazon Cognito як інструмент аутентифікації та AWS Identity and Access Management (IAM) для авторизації доступу до інформаційної панелі.

Налаштування Kinesis Data Firehose

Наступним етапом було налаштування Kinesis Data Firehose. Для цього було використано консоль Kinesis та переглянуто конфігурацію потоку доставки os-demo-firehose-stream. У пошуковому рядку праворуч від «Services» вибрано «Kinesis». У лівій навігаційній панелі вибрано «Delivery streams» та натиснуто на потік доставки «os-demo-firehose-stream». Переглянуто розділ «Delivery stream details», де було вказано «Source» як «Direct PUT» і «Destination» як «Amazon OpenSearch Service». Якщо прокрутити сторінку вниз, можна було помітити увімкнений моніторинг, який генерував метрики потоку доставки. Також переглянуто конфігурацію трансформації даних. У нижній частині сторінки деталей потоку доставки було вибрано вкладку «Configuration» та розділ «Transform records». Там знаходиться функція Lambda «os-demo-lambda-function», яка використовується для трансформації журналів доступу перед їх передачею до OpenSearch Service.

Налаштування OpenSearch Service

Конфігурація кластера OpenSearch Service здійснювалася через консоль OpenSearch. На вкладці «Configuration» для потоку доставки у розділі «OpenSearch Service destination» під «Domain» було вибрано посилання «os-demo». Відкрилася консоль OpenSearch Service, де було переглянуто конфігурацію кластерата налаштування безпеки. URL-адреса OpenSearch Dashboards знадобилася для подальшого використання. Також на вкладці «Security configuration» переглянуто політики доступу, які використовували дві ролі: os_demo_firehose_delivery_role та osdemocognitoauthuserrole.

Створення індексу OpenSearch

Для створення індексу OpenSearch використано OpenSearch Dashboards. Для входу було використано такі облікові дані: ім'я користувача LabUser та пароль Passw0rd1!. Після входу система запропонувала змінити пароль. Було введено новий пароль Passw0rd1!2. Після входу видалено наявні журнали вебсервера. Для цього через значок меню запущено «Dev Tools». У консолі Dev Tools командою DELETE /apache_logs було видалено індекс, якщо він існував. Далі здійснено створення нового індексу за допомогою методу PUT. Приклад коду для створення індексу наведено нижче (рис. 6):

```
PUT apache_logs
{
  "settings": {
    "index": {
      "number_of_shards": 10,
      "number_of_replicas": 0
    }
  },
  "mappings": {
    "properties": {
      "agent": {
        "type": "text"
```

Рис. 6. Код для створення індексу

```
        },
        "browser": {
          "type": "keyword"
        },
        "bytes": {
          "type": "text"
        },
        "city": {
          "type": "keyword"
        },
        "country": {
          "type": "keyword"
        },
        "datetime": {
          "type": "date",
          "format": "dd/MMM/yyyy:HH:mm:ss Z"
        },
        "host": {
          "type": "text"
        },
        "location": {
          "type": "geo_point"
        },
        "referer": {
          "type": "text"
        },
        "os": {
          "type": "keyword"
        },
        "request": {
          "type": "text"
        },
        "response": {
          "type": "text"
        },
        "webpage": {
          "type": "keyword"
        },
        "refering_page": {
          "type": "keyword"
        }
      }
    }
}
```

Рис. 6. Код для створення індексу (продовження)

Відобразилася така відповідь (рис. 7):

```
{
  "acknowledged": true,
  "shards_acknowledged": true,
  "index": "apache_logs"
}
```

Рис. 7. Результат створення індексу

Ця команда створила новий індекс під назвою apache_logs. Коли журнали вебсервера оновлювалися через трафік вебсайту, потік доставки Kinesis Data Firehose наповнював кластер OpenSearch Service даними на основі відповідностей у команді. Команда встановила типи даних для полів

у файлах журналів сервера у базі даних OpenSearch Service. Після створення індексу було здійснено генерацію журналів доступу до вебсервера, а потім створено візуалізації на основі даних у цьому індексі [4].

Генерація журналів доступу до вебсервера

Після створення індексу було згенеровано журнали доступу до вебсервера. Це здійснювалося шляхом відвідування вебсайту (рис. 8), розміщеного на EC2-інстансі, з використанням різних веб-браузерів.

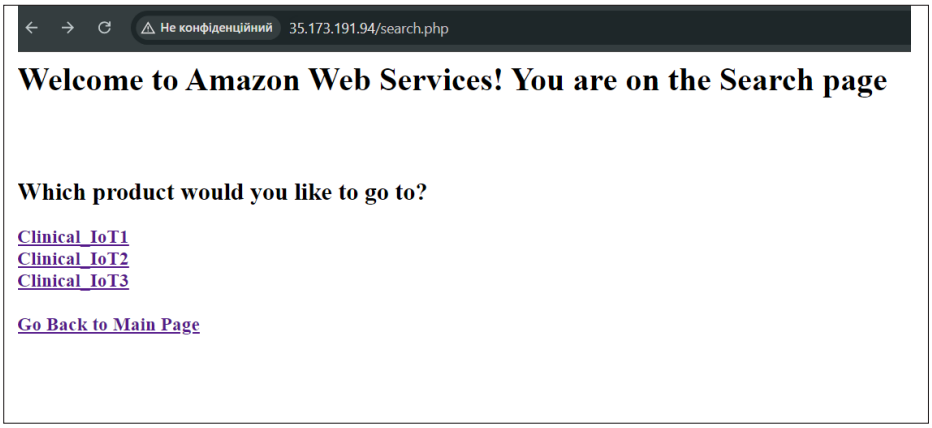


Рис. 8. Результат відвідування вебсайту

Для цього використовувалася нова вкладка або вікно браузера та здійснено перехід за URL-адресою http://<PUBLIC-IP>/main.php, замінивши <PUBLIC-IP> на публічну IP-адресу, яку було збережено раніше.

Шляхом симуляції користувацької активності в кількох веб-браузерах, згенеровано журнали вебсервера, відвідуючи різні сторінки сайту.

Спостереження за подіями журналу в CloudWatch

Для спостереження за процесом завантаження та обробки даних використовувалися журнали CloudWatch. Вони дозволяють побачити, як журнали доступу відвідувачів завантажувалися з вебсервера до Kinesis Data Firehose, збагачувалися функцією Lambda і передавалися до OpenSearch Service для подальшого аналізу [2]. Повернувшись до AWS Management Console, було вибрано «CloudWatch». У лівій навігаційній панелі було вибрано «Logs» > «Log groups» і групу журналів /aws/lambda/aes-demo-lambda-function. Переглянуто потік журналів (рис. 9), вибравши останній потік журналів.

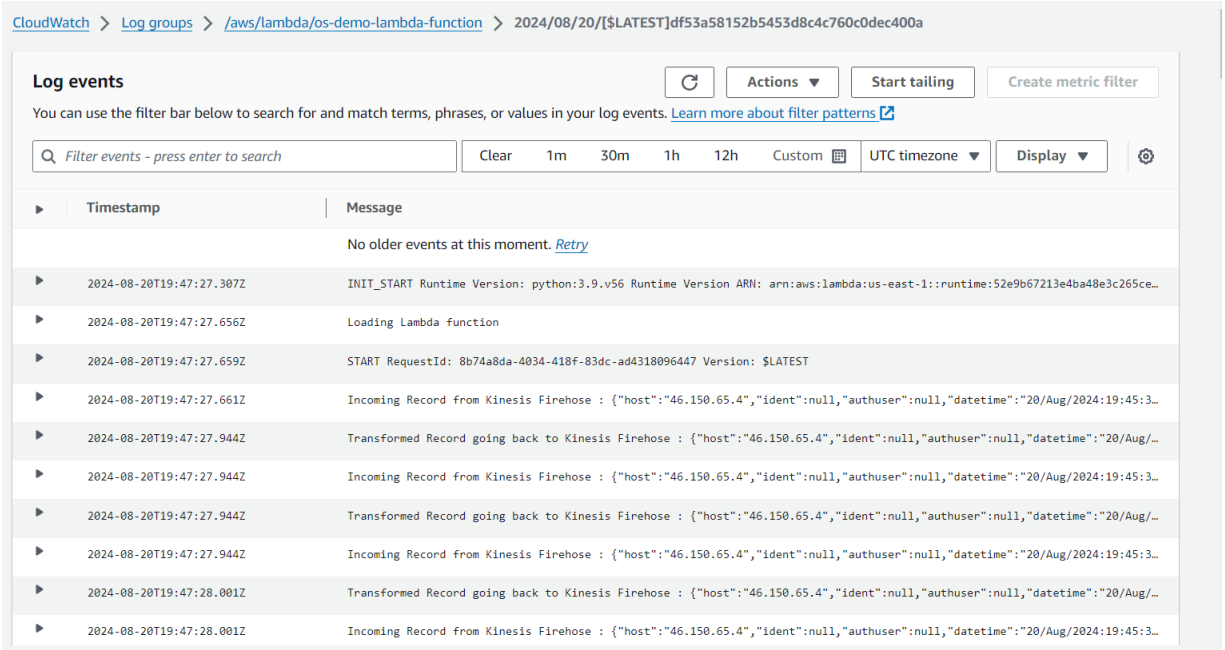


Рис. 9. Потік журналів

Деталі подій журналів було переглянуто, розгорнувши журнали з повідомленням «Incoming Record from Kinesis Firehose», щоб побачити початкові дані журналу доступу, та «Transformed Record going back to Kinesis Firehose», щоб побачити збагачені дані журналу. Деталі події подібні до таких (рис. 10):


```
Incoming Record from Kinesis Firehose :
{
    "host": "46.150.65.4",
    "ident": null,
    "authuser": null,
    "datetime": "20/Aug/2024:19:45:32 +0000",
    "request": "GET /Clinical_IoT1.php HTTP/1.1",
    "response": "200",
    "bytes": "297",
    "referer": "http://35.173.191.94/search.php",
    "agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like
Gecko) Chrome/127.0.0.0 Safari/537.36"
}
Transformed Record going back to Kinesis Firehose :
{
"host": "46.150.65.4",
    "ident": null,
    "authuser": null,
    "datetime": "20/Aug/2024:19:45:32 +0000",
    "request": "GET /Clinical_IoT1.php HTTP/1.1",
    "response": "200",
    "bytes": "297",
    "referer": "http://35.173.191.94/search.php",
    "agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like
Gecko) Chrome/127.0.0.0 Safari/537.36",
    "webpage": "Clinical_IoT1",
    "refering_page": "search",
    "city": "Zaporizhia",
    "country": "Ukraine",
    "browser": "Chrome",
    "os": "Windows",
    "location": {
        "lat": 47.85,
        "lon": 35.2833
  }
}
```


Рис. 10. Збагачені дані журналу

Аналіз та візуалізація

У цьому підрозділі описано, як можна використовувати OpenSearch Dashboards для роботи з даними IoT, які наповнювали кластер OpenSearch Service.

Створення шаблону індексу OpenSearch

Після генерування журналів доступу та спостереження за їх обробкою було створено шаблон індексу в OpenSearch Dashboards. Це дозволило ідентифікувати індекси OpenSearch, які використовувалися для створення візуалізацій. Для цього на сторінці OpenSearch Dashboards вибрано «Discover», потім «Create index pattern» та введено apache_logs як ім'я шаблону індексу. На другому кроці вибрано поле datetime для фільтрації даних та завершено створення шаблону (табл. 1).

Створення кругової діаграми

Кругові діаграми є важливим інструментом для візуалізації категорійних даних, оскільки вони дозволяють легко порівнювати частки різних сегментів. У контексті вебаналітики кругові діаграми допомагають візуалізувати дані про операційні системи та браузери, які використовували відвідувачі вебсайту [2]. Нижче описано покроковий процес створення кругової діаграми в OpenSearch Dashboards.

Спочатку здійснено вхід до OpenSearch Dashboards. У головному меню було вибрано розділ «Visualize» для створення нової візуалізації. Натиснуто кнопку «Create new visualization» і вибрано тип візуалізації «Pie».

Таблиця 1
Шаблон індексу в OpenSearch Dashboards

Ім'я	Тип	Формат	З пошуком	Агреговано	Виключено
id	string		✓	✓	
index_	string		✓	✓	
score	number			✓	
source	source				
type	string		✓	✓	
agent	string		✓		
browser	string		✓	✓	
bytes	string		✓		
city	string		✓	✓	
country	string		✓	✓	

Після вибору типу візуалізації відкрилося вікно для вибору джерела даних. Було вибрано індекс apache_logs*, що був створений раніше. Це забезпечило доступ до даних журналів вебсервера для подальшого аналізу.

У верхньому правому куті інтерфейсу було налаштовано часовий інтервал для аналізу даних. Вибрано опцію «Last 30 minutes» для обмеження даних до останніх 30 хвилин. Це дозволило отримати актуальні дані про відвідування вебсайту.

Кошики використовувалися для категоризації даних у круговій діаграмі. Для створення першого кошика було вибрано розділ «Buckets» і натиснуто «Add» > «Split slices». У полі «Aggregation» вибрано «Terms», а в полі «Field» – «webpage». Це налаштування дозволило розділити дані за різними вебсторінками.

Залишено значення за замовчуванням для «Order by», «Order» та «Size». Увімкнуто опцію «Group other values in separate bucket», щоб групувати інші значення в окремий кошик. Це забезпечило відображення всіх вебсторінок, навіть якщо деякі з них мали мало даних. Далі додано ще один кошик для категоризації даних за операційними системами. Знову вибрано «Add» > «Split slices» і «Terms» у полі «Aggregation». У полі «Field» натиснуто «os». Увімкнуто опцію «Group other values in separate bucket».

Останнім кошиком, який потрібно було додати, став кошик для категоризації даних за типом браузера. Повторено попередні кроки, «Add» > «Split slices» і «Terms» у полі «Aggregation». У полі «Field» натиснуто «browser». Увімкнуто опцію «Group other values in separate bucket».

Для більш детального аналізу застосовано фільтри до кругової діаграми. Наприклад, щоб відобразити лише дані про певну вебсторінку, вибрано «Add filter» у верхньому лівому куті. У полі «Field» вибрано «webpage», у полі «Operator» натиснуто «is», а у полі «Value» введено «kindle». Далі натиснуто «Save» для застосування фільтра. Для зміни стилю кругової діаграми на «donut» та налаштування інших параметрів вибрано «Options» на панелі праворуч. Опції «Donut» та «Show top level only» вимкнуто. Увімкнуто опцію «Show labels» для відображення міток на діаграмі. Натиснуто «Update» для застосування змін (рис. 12).

Ця візуалізація ілюструє відсоток переглядів сторінки продукту з різних браузерів та операційних систем. Після створення діаграми можна взаємодіяти з нею для отримання додаткових інсайтів. Курсор можно навести на кожен сегмент діаграми, щоб побачити деталі про дані, що включені до сегмента. Додано інші фільтри для перегляду даних за різними параметрами, такими як тип браузера або операційна система.

Ця кругова діаграма (рис. 13) подібна до попередньої, але включала фільтр для певного браузера.

Створення теплової карти

Теплова карта дає можливість дізнатися, чи більше клієнтів переходило на сторінки продуктів з пошукової сторінки або зі сторінки рекомендацій [2].

Для створення теплової карти використано OpenSearch Dashboards. У головному меню вибрано розділ «Visualize» для створення нової візуалізації. Натиснуто кнопку «Create new visualization» і тип візуалізації «Heat Map».

Після вибору типу візуалізації відкрилося вікно для вибору джерела даних. Було вибрано індекс apache_logs*, що був створений раніше. Це забезпечило доступ до даних журналів вебсервера для подальшого аналізу.

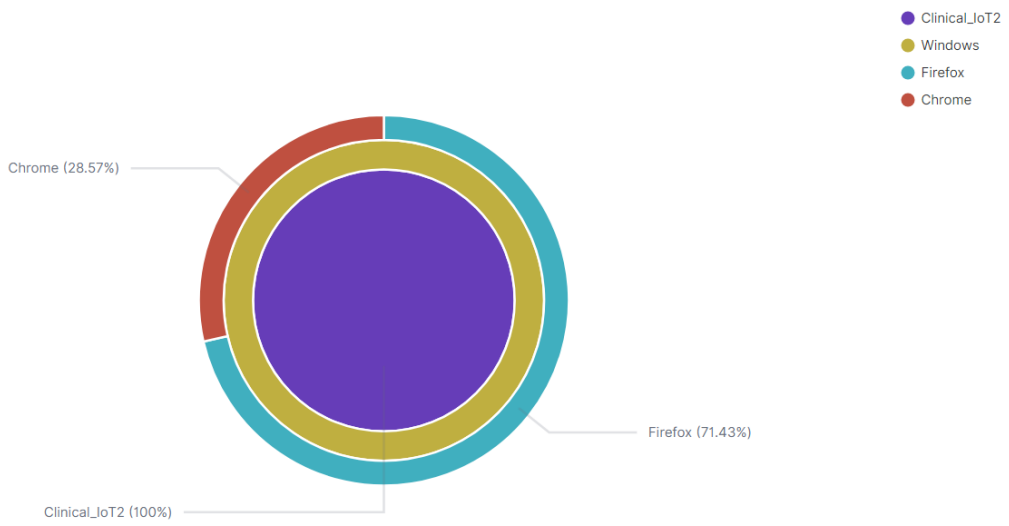


Рис. 12. Кругова діаграма

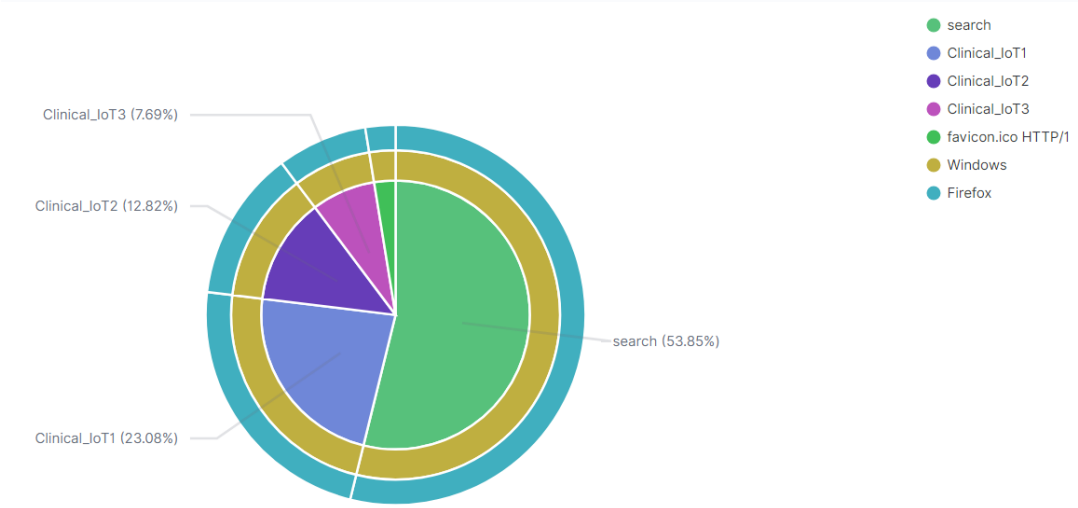


Рис. 13. Фільтр для браузера Firefox

Теплова карта використовує дві осі для відображення взаємозв'язків між різними категоріями даних. Почато з налаштування X-осьового кошика. У розділі «Buckets» натиснуто «Add» > «X-axis». У полі «Aggregation» вибрано «Terms», а в полі «Field» – «refering_page». Це дозволило розділити дані за сторінками, з яких користувачі переходили. Залишено значення за замовчуванням для «Order by», «Order» та «Size». Увімкнуто опцію «Group other values in separate bucket», щоб групувати інші значення в окремий кошик. Це забезпечило відображення всіх сторінок переходів, навіть якщо деякі з них мали мало даних.

Наступним кроком було налаштування Y-осьового кошика для категоризації даних за відвідуваними вебсторінками. Знову обрано «Add» > «Y-axis» і «Terms» у полі «Aggregation». У полі «Field» натиснуто «webpage». Залишено значення за замовчуванням для «Order by», «Order» та «Size» та увімкнуто опцію «Group other values in separate bucket». Після налаштування осей було налаштовано параметри відображення теплової карти. На вкладці «Options» і в розділі «Labels» увімкнуто опцію «Show labels» для відображення міток на тепловій карті. Це дозволило чіткіше бачити категорії даних, що аналізувалися.

Для отримання актуальних даних про відвідування вебсайту було налаштовано часовий інтервал. У верхньому правому куті інтерфейсу змінено тривалість на останню 1 годину і натиснуто «Refresh». Це дозволило тепловій карті відобразити найсвіжіші дані про активність користувачів.

Теплова карта відображає кількість переглядів вебсторінок, а також джерела переходів (з пошукової сторінки або зі сторінки рекомендацій). Чим вища інтенсивність кольору, тим більше переглядів зафіксовано для відповідної категорії. Теплова карта виглядала приблизно так (рис. 14), але вона

буде різною залежно від сторінок, які було відвідано під час генерації журналів доступу до вебсервера, та браузерів, які було використано.

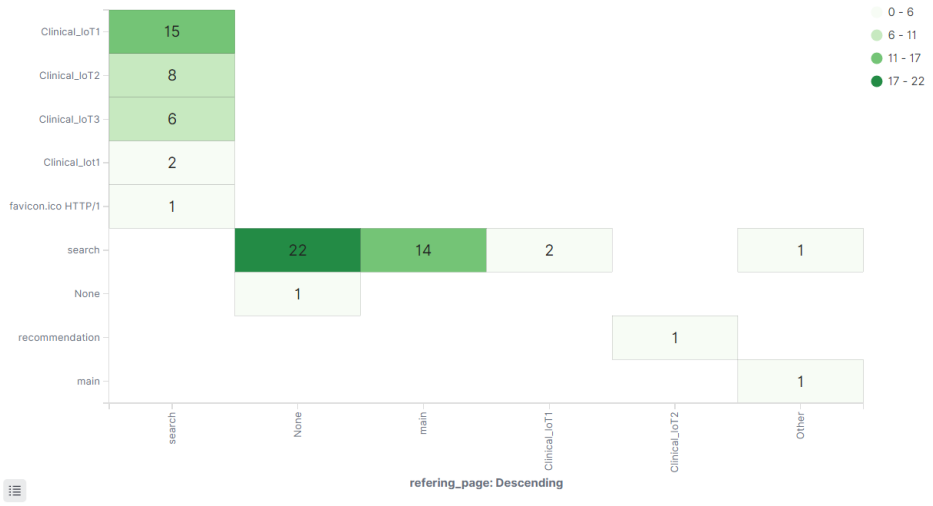


Рис. 14. Фільтр для браузера Firefox

Виходячи з цього зображення (рис. 14) візуалізації, відвідувачі частіше переходили на сторінки продуктів Clinical_IoT1 і Clinical_IoT2 з пошукової сторінки, ніж зі сторінки рекомендацій. Зроблено висновок, що пошукова сторінка є більш ефективною, ніж сторінка рекомендацій, для перенаправлення користувачів на сторінки продуктів.

Висновок

Створено конвеєр для підтвердження концепції, щоб збирати й аналізувати дані з пристроїв Інтернету речей, вбудованих в електровелосипеди. У складі конвеєру AWS IoT Core підписаний на тему, в якій публікуються дані з пристроїв IoT на електровелосипедах. Правило IoT направляє дані в потік доставки Kinesis Data Firehose. Рішення використовує Kinesis Data Firehose для завантаження потокових журналів доступу до вебсервера, а потім використовує функції Lambda для збагачення даних. Після використання Kinesis та Lambda для обробки даних використовувався OpenSearch Service для завантаження та індексації даних. Використовуючи OpenSearch Dashboards, створюються візуалізації даних, щоб отримати уявлення про стан користувачів електровелосипедів. Візуалізації поширюються серед зацікавлених сторін, використовуючи Amazon Cognito як інструмент аутентифікації та AWS Identity and Access Management (IAM) для авторизації доступу до інформаційної панелі.

Література

1. Evans D. The Internet of Things: How the Next Evolution of the Internet Is Changing Everything. *CISCO White Paper.* 2011.
2. Бондаренко Л. Д. Інтернет речей: Технології та застосування. *Системи управління, навігації та зв'язку*, 2020. 5(63), 120–123.
3. Perera C., Zaslavsky A., Christen P., Georgakopoulos D. Context Aware Computing for the Internet of Things: A Survey. *IEEE Communications Surveys & Tutorials*, 2014. 16(1), 414–454.
4. Rajaraman V. Big Data Analytics. *Resonance*, 2016. 21(7), 695–716.
5. Zhang Y., Qian C., Zhang W. Internet of Things: Security and Privacy in a Connected World. Springer. 2017.
6. Zaslavsky A., Perera C., Georgakopoulos D. Sensing as a Service and Big Data. *IEEE International Conference on Advances in Cloud Computing (ACC-2013)*, Bangalore, India, 2013. 21–29.

References

1. Evans, D. (2011). The Internet of Things: How the Next Evolution of the Internet Is Changing Everything. CISCO White Paper.
2. Bondarenko, L. D. (2020). Internet of things: Technologies and applications. *Control, navigation and communication systems*, 5(63), 120–123.
3. Perera, C., Zaslavsky, A., Christen, P., Georgakopoulos, D. (2014). Context Aware Computing for the Internet of Things: A Survey. *IEEE Communications Surveys & Tutorials*, 16(1), 414–454.
4. Rajaraman, V. (2016). Big Data Analytics. *Resonance*, 21(7), 695–716.
5. Zhang, Y., Qian, C., Zhang, W. (2017). Internet of Things: Security and Privacy in a Connected World. Springer.
6. Zaslavsky, A., Perera, C., Georgakopoulos, D. (2013). Sensing as a Service and Big Data. *IEEE International Conference on Advances in Cloud Computing (ACC-2013)*, Bangalore, India, 21–29.

УДК 657.004
DOI https://doi.org/10.36994/2788-5518-2024-02-08-02

РОЗРОБЛЕННЯ ІНФОРМАЦІЙНОЇ ОНЛАЙН-СИСТЕМИ ОБЛІКУ ВИТРАТ ТА ДОХОДІВ

Лаврів М. Р., *магістр, випускник, Івано-Франківський національний технічний університет нафти і газу, Івано-Франківськ, Україна. oksana7may@gmail.com*

Белей О. І., *к.т.н., доц., Івано-Франківський національний технічний університет нафти і газу, Івано-Франківськ, Україна. https://orcid.org/0000-0002-2386-4106, oksana7may@gmail.com*

Штаєр Л. О., *к.т.н., доц., Івано-Франківський національний технічний університет нафти і газу, Івано-Франківськ, Україна. https://orcid.org/0000-0003-1013-9869, oksana7may@gmail.com*

Анотація. *Метою цієї роботи є розроблення інформаційної онлайн-системи обліку витрат та доходів за допомогою сучасних технологій: React, Express.js, Mongo DB, Node.js. Для реалізації мети роботи було використано комплексний метод досліджень літературних джерел щодо понять «витрати» та «доходи»; проведений огляд наявних інформаційних онлайн-систем обліку витрат та доходів за допомогою сучасних технологій React, Express.js, Mongo DB, Node.js та сумісних з ними на ринку аналогічних продуктів, який дозволив виявити сильні сторони розроблювальної системи та визначити напрями подальшого розвитку; продемонстровано приклади реалізації системи обліку витрат та доходів; наведено приклади готових результатів; наведено настанову користувача до розробленої інформаційної онлайн-системи обліку витрат та доходів. Результати роботи: проведено аналіз наявних інформаційних онлайн-систем обліку витрат та доходів за допомогою технологій React, Express.js, Mongo DB, Node.js і супутніх до них, а також досліджено їх переваги та недоліки; продемонстровано і запропоновано для використання розроблену інформаційну онлайн-систему контролю витрат та доходів з можливістю додавання користувачів, виводу фінансової статистики по витратах та доходах; також є можливість додавати, видаляти і редагувати створені категорії витрат та доходів. Наукова новизна: використання новітніх технологій дозволило розробити інформаційну онлайн-систему обліку витрат та доходів, яка дозволить мати весь потрібний функціонал в одному місці з інтуїтивно простим інтерфейсом, безоплатним та необмеженим доступом до нього. Розроблена інформаційна онлайн-система обліку витрат та доходів допомагає користувачам оптимально управляти своїми фінансовими коштами з допомогою фінансового звіту, що дозволить більш раціонально управляти витратами та накопичувати кошти.*

Ключові слова: *витрати, інформаційна система, доходи, інтерфейс, Mongo DB, Node JS.*

DEVELOPMENT OF INFORMATION ONLINE ACCOUNTING SYSTEM EXPENSES AND INCOME

Maksym Lavriv, *master, graduate, Ivano-Frankivsk National Technical University of Oil and Gas, Ivano-Frankivsk, Ukraine. oksana7may@gmail.com*

Oksana Belei**,** *Ph.D., Ass. Prof., Ivano-Frankivsk National Technical University of Oil and Gas, Ivano-Frankivsk, Ukraine. https://orcid.org/0000-0002-2386-4106, oksana7may@gmail.com*

Lidiia Shtaier**,** *Ph.D., Ass. Prof., Ivano-Frankivsk National Technical University of Oil and Gas, Ivano-Frankivsk, Ukraine. https://orcid.org/0000-0003-1013-9869, oksana7may@gmail.com*

Abstract. *The goal of this work is to develop an online information system for accounting for expenses and income using modern technologies: React, Express.js, MongoDB, and Node.js. To achieve the goal of the work, a comprehensive method of research of literary sources regarding the concepts of expenses and income was used; a review of existing online information systems for accounting for expenses and income using modern technologies React, Express.js, MongoDB, and Node.js, and similar products on the market was conducted, which allowed identifying the strengths of the developed system and determining the directions for further development; examples of the implementation of the expense and income accounting system are demonstrated; examples of ready-made results are provided; user instructions for the developed online information system for accounting for expenses and income are provided. Results of the work: an analysis of existing online information systems for accounting for expenses and income using technologies React, Express.js, MongoDB, Node.js, and related technologies was conducted, and their advantages and disadvantages were studied; the developed online system for controlling expenses and income with the ability to add users, output financial statistics on expenses and income is demonstrated and proposed*

for use; there is also the possibility to add, delete, and edit created categories of expenses and income. Scientific novelty: the use of the latest technologies has made it possible to develop an online information system for accounting for expenses and income, which will allow having all the necessary functionality in one place with an intuitively simple interface, free and unlimited access to it. The developed online system for accounting for expenses and income helps users to optimally manage their finances with the help of a financial report, which will allow more rational management of expenses and accumulation of funds.

Key words: *expenses, information system, income, interface, Mongo DB, Node JS.*

Вступ. Облік та планування бюджету не є новинкою, що з'явилась із приходом цифрової ери. Відповідальне поводження із власними фінансами – це необхідність у капіталістичному світі, де гроші сильно впливають на людський добробут.

Щомісячно перед кожною дорослою людиною постає завдання – правильно розрахувати і контролювати власні доходи та витрати. Важливо зазначити, що маються на увазі дві визначені категорії осіб: працівники будь-якої сфери, а також малі підприємці, що не потребують масштабних систем із метриками, документообігом та іншим функціоналом. Вибрані категорії осіб будуть розглядатись упродовж усієї роботи, адже саме це є цільова аудиторія майбутньої системи.

Ось декілька причин чому потрібно вести планування та облік власних фінансів: контроль над витратами: може підвищити рівень вашої фінансової безпеки та допомогти уникнути зайвих витрат; досягнення фінансових цілей: заощадження на великі покупки, отримання освіти або підготовка до пенсії; раціональне використання ресурсів: може підвищити вашу фінансову стійкість та допомогти зберегти гроші на довгострокову перспективу; управління боргами: допомагає керувати боргами, які можуть стати проблемою для вашої фінансової безпеки [1].

Ніхто не забороняє використовувати власний підхід, головне – мати якусь систему, за якою можна ефективно розставляти пріоритети і правильно розподілити кошти на кожну категорію витрат і яка забезпечить відсутність заборгованостей. Одним із таких рішень є розроблення інформаційної онлайн-системи обліку витрат та доходів [1].

Аналіз останніх досліджень і публікацій. У сучасній цифровій економіці розвиток інформаційних онлайн-систем для обліку витрат і доходів має вирішальне значення. Різноманітні дослідження присвячені пошуку оптимальних рішень, що забезпечують зручність та ефективність фінансового обліку. Останні наукові роботи стосуються таких аспектів, як автоматизація процесів, інтуїтивність інтерфейсу та інтеграція із зовнішніми сервісами. У цьому розділі проаналізовано останні дослідження та публікації, присвячені розробці та впровадженню таких систем.

Авторами [2; 3] запропоновані методологічні підходи та розроблено схему структури бази даних системи обліку витрат та доходів, яка є основою для проєктування інформаційної системи моніторингу фінансових операцій, таких як витрати та доходи.

У науковій праці Syifa Siti Rahayu, Hery Dwi Yulianto є створення інформаційної системи бухгалтерського обліку громадянського управління водними ресурсами, яка базується на вебсайті. Використаним методом дослідження було описове опитування, під час якого автор систематично описує ситуацію чи явище в зрозумілій та детальній формі [4].

Розглянутий метод [5] управління персональною платіжною платформою стосується системи, яка дає можливість користувачам (фізичним особам, магазинам, професійним службам, торговим центрам, автозаправним станціям тощо) підключатися та отримувати доступ до широкого спектра фінансових і комунікаційних переваг, що робить її ідеальним центром для залучення клієнтів.

У методі та системі ефективного надання послуг з бухгалтерського обліку та корпоративного планування досліджуються технології, що застосовуються для автоматизації бухгалтерського обліку, включаючи обробку великих обсягів даних, створення звітів та використання алгоритмів машинного навчання для аналізу фінансової інформації [6].

Privat24 – це мобільний банкінг та фінансовий застосунок, що надає клієнтам ПриватБанку доступ до своїх фінансових ресурсів через смартфон або планшет. За допомогою Privat24 клієнти можуть переглядати баланс рахунків, оплачувати рахунки за комунальні послуги, мобільний зв'язок та інші послуги, переказувати гроші, купувати валюту та інші фінансові операції. Застосунок дозволяє користувачам контролювати свої витрати та вести облік своїх фінансів. Крім того, Privat24 має функціонал для збереження квитанцій та інших документів, які стосуються фінансів користувача. Застосунок також має низку безпечних функцій, таких як можливість змінити ПІН-код та встановити ліміти на операції, що дозволяє зберігати фінансові ресурси в безпеці [7].

Розглянуто також програми Google Sheets та Excel для створення та редагування електронних таблиць. Обидва застосунки мають схожий функціонал та дозволяють користувачам створювати таблиці, додавати дані, форматувати їх та виконувати різноманітні обчислення. Google Sheets – це безкоштовний онлайн-сервіс, який дозволяє користувачам створювати, редагувати та зберігати електронні таблиці в хмарі. За допомогою Google Sheets користувачі можуть спільно працювати над таблицями та відстежувати зміни, що були зроблені різними користувачами. Google Sheets також має вбудовані засоби для візуалізації даних та автоматичного оновлення інформації з інших джерел.

Excel – це програма, що входить до пакета Microsoft Office, та дає більш широкі можливості для обробки та аналізу даних. Excel має вбудовані функції для створення складних формул та фінансових розрахунків, а також можливості для автоматизації рутинних завдань, таких як обробка даних, форматування таблиць та інші. Excel також має розширений набір інструментів для візуалізації даних, що дозволяє користувачам створювати складні діаграми та графіки, використовуючи різні типи даних та джерела [8; 9].

Проаналізовано мобільний додаток Mono, який дозволяє користувачам отримувати доступ до своїх фінансових даних, що стосуються їхнього банківського рахунку в Monobank. За допомогою цього застосунку користувачі можуть переглядати історію транзакцій, переводити гроші, керувати бюджетом та планувати витрати. Mono має низку корисних функцій, таких як нагадування про платежі, створення автоматичних платежів та можливість налаштування категорій для кращого відстеження витрат. Застосунок також дозволяє користувачам контролювати свої витрати та аналізувати свої фінанси за допомогою графіків та статистики. Крім того, Mono має вбудовану підтримку клієнтів, що дозволяє швидко отримати відповіді на запитання та розв'язати проблеми з акаунтом [10].

У дослідженні [11] розглядаються фактори, які впливають на практику згладжування доходів, такі як розмір фірми, прибутковість, фінансовий левередж і норма чистого прибутку. У роботі розглянуто групування серед компаній, які здійснюють згладжування прибутку і які не виконують згладжування доходу за допомогою індексу Екеля до чистого прибутку для виробничих компаній.

Проте згадані застосунки володіють нерідко величезним функціоналом зі складним інтерфейсом або ж дають доступ до функціоналу лише за умови надання документів, як у випадку із Privat24, Mono, Google Sheets та Excel, або ж за умови щомісячної платної підписки на сервіс. Власне, і планування та облік там відносні, адже всі записи про доходи, витрати чи плани про такі існують тільки після безпосередньої фінансової операції, тому їх функціонал більше схожий на фінансову історію, часто без можливості редагування.

Отже, постає проблема, як можна одночасно відслідковувати власні доходи та витрати, не тільки фактичні, але й плановані, що означає: ми можемо створити запис про них перед тим, як власне реальні фінансові операції будуть здійснені, не надаючи документи банкам і не підписуючись на платні сервіси, що володіють у рази більшим функціоналом, ніж потрібно пересічному користувачу.

Для ефективного розроблення інформаційної системи обліку витрат та доходів за допомогою технологій React, Express.js, MongoDB та Node.js проведено дослідження з визначення чітких цілей та потреб користувачів. Детально проаналізовано наявні рішення на ринку, щоб виявити їхні сильні та слабкі сторони. Розроблено зручний та інтуїтивний інтерфейс користувача за допомогою React.

Виклад основного матеріалу й обговорення результатів. У розробленій інформаційній системі передбачено такий функціонал, виконаний згідно з ідеєю самої роботи: створення особистого профілю з інформацією про себе, а саме: повне ім'я, посада, навички, фотографія (рис. 1, 2); можливість вести облік витрат та доходів (рис. 3, 4), поділених за різними категоріями та створеними самим користувачем (рис. 5, 6); можливість перегляду фінансової статистики користувача по витратах та прибутках (рис. 7).

Для початку роботи у системі користувач повинен створити особистий кабінет користувача, перейшовши за посиланням на вебсайт та ввівши всі потрібні дані у форму реєстрації. Після заповнення полів форми користувач мусить натиснути на кнопку «SIGN UP» для відправки даних на сервер, після чого той поверне результат (рис. 1).

Рис. 1. Форма «Sign Up»

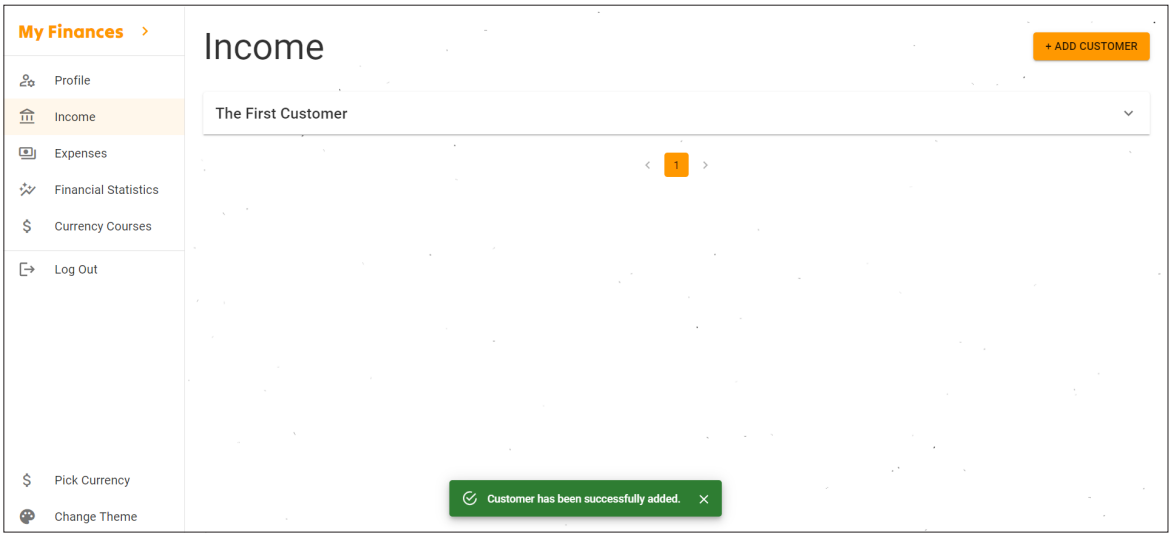


Рис. 2. Успішно доданий у таблицю клієнт

Після успішного створення користувачів інформаційної онлайн-системи обліку та доходів проєктуються вкладки «Income» та «Expenses», які відповідають за керування доходами та витратами (рис. 7).

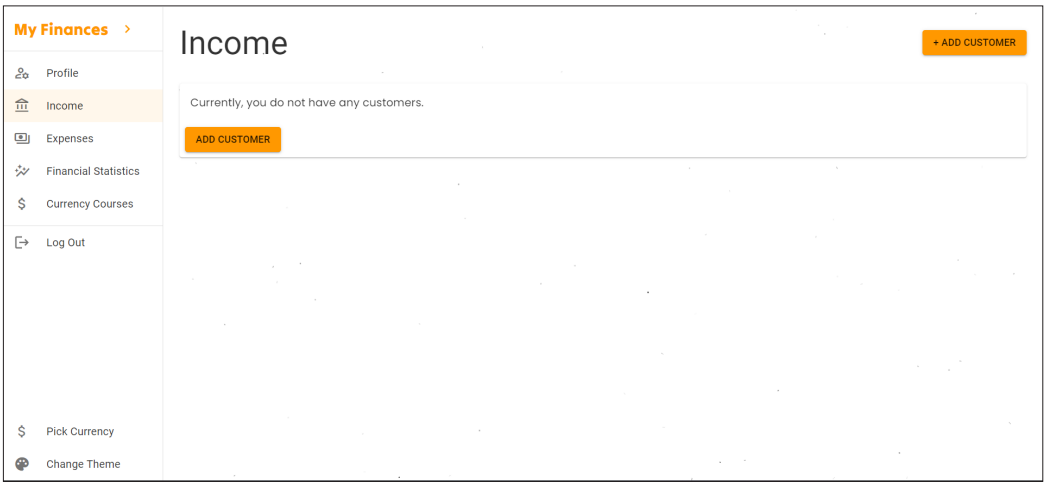


Рис. 3. Сторінка керування доходами

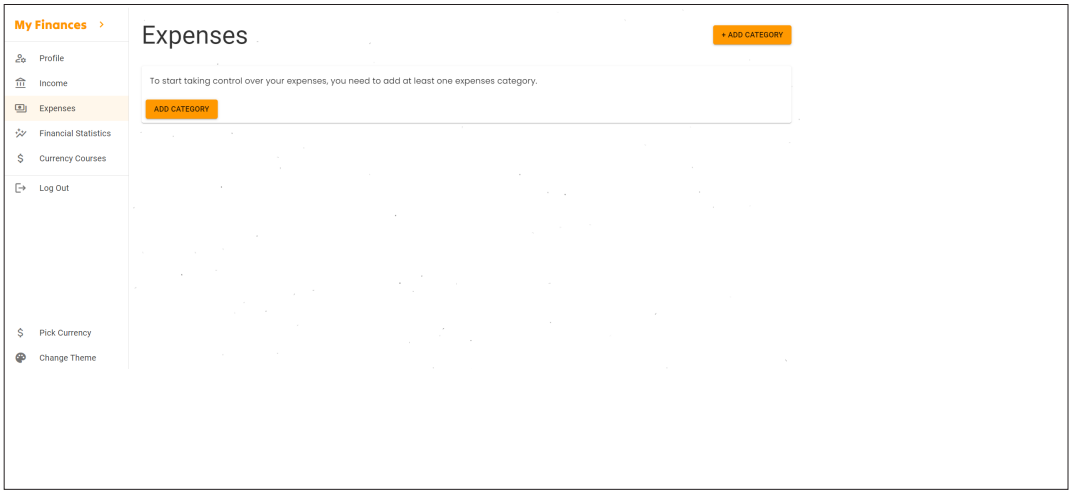


Рис. 4. Сторінка керування вкладкою «Expense»

Розроблена система дозволяє створити щонайменше одну категорію витрат для того, щоб розпочати роботу із функціональністю модуля. Користувач на вибір може скористатись для додавання категорії однією із кнопок «+ ADD CATEGORY», які працюють аналогічно до додавання клієнта, єдине, що форма всередині діалогового вікна зроблена для додавання категорії (рис. 5). Форма дає можливість вказати ім'я, іконку та опис категорії.

Категорії витрат та доходів є групами, за якими класифікуються всі витрати чи доходи користувачів. Така класифікація дозволяє краще контролювати фінансові потоки, аналізувати витрати та приймати обґрунтовані рішення. Наприклад, контроль бюджету; податкове планування; матеріальні витрати; постійні витрати; змішані витрати, заробітна плата, кошти від надання орендних послуг тощо (рис. 6).

Успішно створена форма для категорії витрат «Test Food Category» показана на рис. 5, а на рис. 6 показаний конкретний приклад витрат на білий хліб «Bread white».

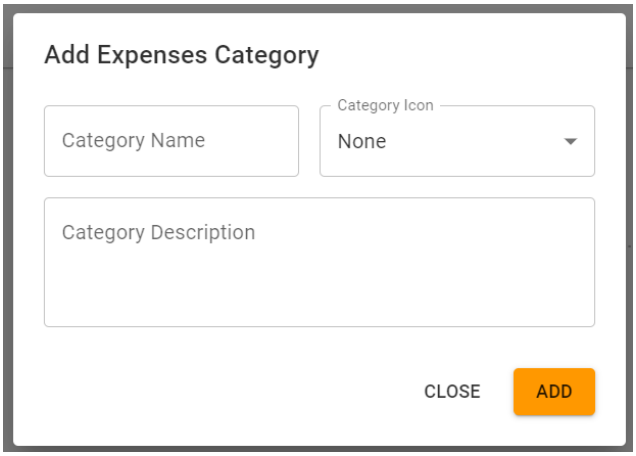


Рис. 5. Форма створення категорії витрат

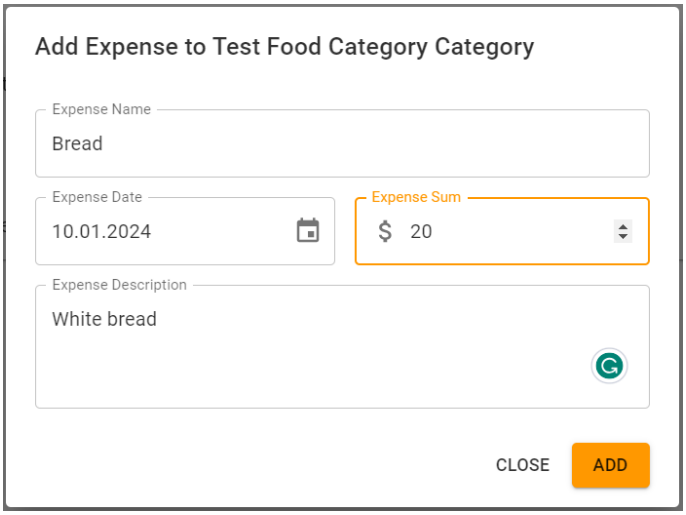


Рис. 6. Діалогове вікно додавання витрати

Financial Statistics – сторінка, яка тільки відображає статистичну інформацію, сформовану на основі введених користувачем виплат та витрат, рис. 7.

Переглянути статистику (рис. 7) можна окремо для прибутків і витрат. При цьому користувачам надається така інформація: Total Expenses/Income – загальна сума витрат та доходів, а також категорія, у якій їх найбільше; Expenses/Income This Month – сума витрат та доходів поточного місяця; Most Spending – найбільша за сумою витрата чи дохід за весь час; Monthly Expenses/Income – граф для зображення сум усіх витрат та доходів помісячно; Expenses/Income Breakdown – граф для розбиття категорій витрат у відсотковому відношенні, що дає можливість переглянути, скільки відсотків становить сума витрат чи доходів окремої категорії із загальної суми витрат чи доходів користувача.

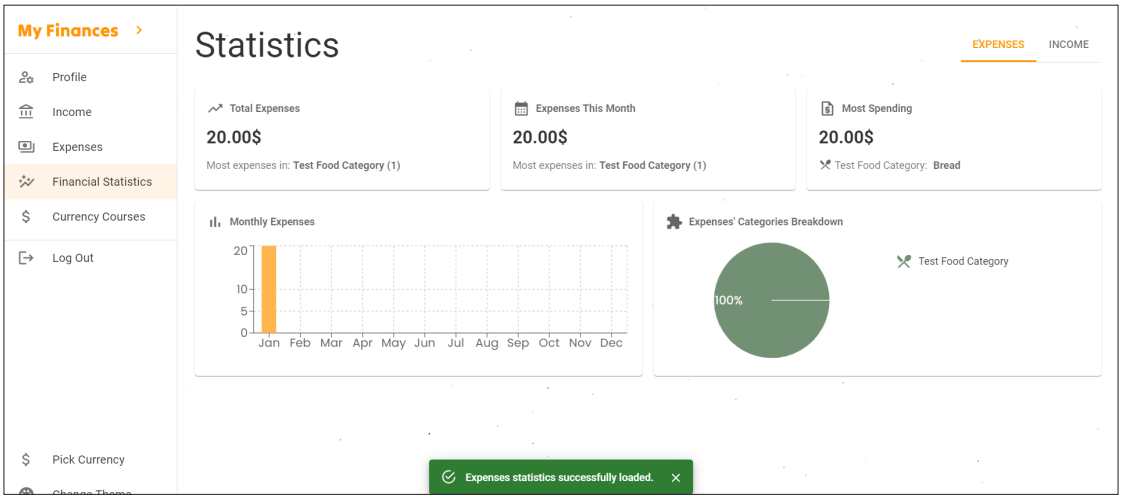


Рис. 7. Статистика витрат користувача

Висновки. На ринку досить платних систем, які мають можливість задовільнити потребу пересічного громадянина, але вони це роблять частково або ж за певну ціну. Натомість розроблювальна інформаційна система оригінально постачатиметься безоплатно. Запропонована інформаційна онлайн-система обліку витрат та доходів дозволяє одночасно відслідковувати власні доходи та витрати не тільки фактичні, але й плановані, що означає: ми можемо створити запис про них перед тим, як власне реальні фінансові операції будуть здійснені, не надаючи документи банкам і не підписуючись на платні сервіси, що володіють у рази більшим функціоналом, ніж потрібно пересічному користувачу. А практичне використання новітніх технологій дозволило спроєктувати реальну інформаційну онлайн-систему для обліку витрат і доходів, яка також характеризується інтуїтивно зрозумілим інтерфейсом з усіма необхідними функціями в одному місці та необмеженим доступом.

Література

1. Як спланувати сімейний бюджет на рік: поради економіста. *Suspilne.media.* 2021, 20 лютого. URL: https://suspilne.media/106678-ak-splanuvati-simejnij-budzet-na-rik-poradi-ekonomista/ (дата звернення: 14.10.2024).

2. Лаврів М. Р., Белей О. І., & Штаєр Л. О. (2023). Теоретичні передумови розроблення online-системи обліку витрат та доходів. *Приладобування: стан і перспективи*, с. 356–357. ПБФ КПІ ім. Ігоря Сікорського. URL: https://ela.kpi.ua/items/064f16c2-73f4-47f7-96aa-d3462db55170.

3. Лаврів М. Р., Белей О. І., Штаєр Л. О. (2024). Розроблення схеми структури бази даних системи обліку витрат та доходів. Міжнародний науковий журнал *«Інтернаука»,* 2024. 9. URL: https://doi.org/10.25313/2520-2057-2024-9-10309.

4. Rahayu S. S., & Yulianto H. D. Development of Website-Based Accounting Information Systems of Comprehensive Income Reports for Citizen Water Management. *@is The Best Accounting Information Systems and Information Technology Business Enterprise*. Vol. 8. Issue 1, 2023. 61–75. URL: https://doi.org/10.34010/aisthebest.v8i1.10890.

5. Martin Thomas Mcleod, & Christopher M. Hughes. Method of Managing a Personal Payment Platform (Patent United States of America № US 20200394638A1). 2020. URL: https://patentimages.storage.googleapis.com/86/b4/db/458bb5cb54cfd5/US20200394638A1.pdf

6. Walter Drangmeister, & R. Blake Ridgeway. METHODS AND SYSTEMS FOR EFFICIENT DELIVERY OF ACCOUNTING AND CORPORATE PLANNING SERVICES (Patent United States of America № US 11,393,045 B2). 2020. URL: https://patentimages.storage.googleapis.com/73/5b/da/73402ed64434c7/US11393045.pdf.

7. ПриватБанк. URL: https://privatbank.ua/ (дата звернення: 14.10.2024).

8. Google Sheets. WikipediA. URL: https://en.wikipedia.org/wiki/Google_Sheets (дата звернення: 14.10.2024).

9. Microsoft Excel. WikipediA. URL: https://en.wikipedia.org/wiki/Microsoft_Excel (дата звернення: 14.10.2024).

10. Monobank – банк у телефоні. Monobank. URL: https://www.monobank.ua/?lang=uk (дата звернення: 14.10.2024).

11. Alqirem R., Afifa M. A., Saleh I., & Haniah F. Ownership Structure, Earnings Manipulation, and Organizational Performance: The Case of Jordanian Insurance Organizations. *The Journal of Asian Finance, Economics and Business*, 2020. 7(12), 293–308. URL: https://doi.org/10.13106/jafeb.2020.vol7.no12.293.

References

1. Yak splanuvaty simeinyi biudzhet na rik: porady ekonomista. (2021, 20 February) [How to plan a family budget for the year: advice from an economist]. *Suspilne.media* (Last accessed: 14.10.2024) [in Ukrainian].

2. Lavriv, M., Belei, O., & Shtaier, L. (2023). Teoretychni peredumovy rozroblennia online-systemy obliku vytrat ta dokhodiv [Theoretical prerequisites for the development of an online system of accounting for expenses and income]. *Pryladobuvannia: stan i perspektyvy*: zbirnyk materialiv XXII Mizhnarodnoi naukovo-tekhnichnoi konferentsii, s. 356–357 [in Ukrainian].

3. Lavriv, M., Belei, O., & Shtaier, L. (2024). Rozroblennia skhemy struktury bazy danykh systemy obliku vytrat ta dokhodiv [Development of the scheme of the structure of the database of the cost and income accounting system]. Mizhnarodnyi naukovyi zhurnal *"Internauka"*, 9 [in Ukrainian].

4. Rahayu, S. S., & Yulianto, H. D. (2023). Development of Website-Based Accounting Information Systems of Comprehensive Income Reports for Citizen Water Management. *@is The Best Accounting Information Systems and Information Technology Business Enterprise*, 8(1), 61–75. Retrieved from: https://doi.org/10.34010/aisthebest.v8i1.10890/ [in English].

5. Martin Thomas Mcleod, & Christopher M. Hughes. (2020). Method of Managing a Personal Payment Platform (Patent United States of America № US 20200394638A1). Retrieved from: https://patentimages.storage.googleapis.com/86/b4/db/458bb5cb54cfd5/US20200394638A1.pdf [in English].

6. Walter Drangmeister, & R. Blake Ridgeway. (2020). METHODS AND SYSTEMS FOR EFFICIENT DELIVERY OF ACCOUNTING AND CORPORATE PLANNING SERVICES (Patent United States of America № US 11,393,045 B2). Retrieved from: https://patentimages.storage.googleapis.com/73/5b/da/73402ed64434c7/US11393045.pdf [in English].

7. PryvatBank [Privatbank]. Retrieved from: https://privatbank.ua/ (Last accessed: 14.10.2024) [in Ukrainian].

8. Google Sheets. WikipediA. Retrieved from: https://en.wikipedia.org/wiki/Google_Sheets (Last accessed: 14.10.2024) [in English].

9. Microsoft Excel. WikipediA. Retrieved from: https://en.wikipedia.org/wiki/Microsoft_Excel (Last accessed: 14.10.2024) [in English].

10. Monobank – bank u telefoni [Monobank – a bank in the phone]. Monobank. Retrieved from: https://www.monobank.ua/?lang=uk (Last accessed: 14.10.2024) [in Ukrainian].

11. Alqirem, R., Afifa, M. A., Saleh, I., & Haniah, F. (2020). Ownership Structure, Earnings Manipulation, and Organizational Performance: The Case of Jordanian Insurance Organizations. *The Journal of Asian Finance, Economics and Business*, 7(12), 293–308. https://doi.org/10.13106/jafeb.2020.vol7.no12.293 [in English].

КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ

УДК 519.85
DOI https://doi.org/10.36994/2788-5518-2024-02-08-03

ЗАДАЧІ СТОХАСТИЧНОЇ КОМБІНАТОРНОЇ ОПТИМІЗАЦІЇ: ОГЛЯД ОСТАННІХ РЕЗУЛЬТАТІВ

Барболіна Т. М., *д.ф.-м.н., доц., Полтавський національний педагогічний університет імені В. Г. Короленка, Полтава, Україна. https://orcid.org/0000-0002-4596-7907, tm-b@gsuite.pnpu.edu.ua*

Баранник Т. А., *к.ф.-м.н., доц., Полтавський національний педагогічний університет імені В. Г. Короленка, Полтава, Україна. https://orcid.org/0009-0003-0509-9935, barannykt@gmail.com*

Кононович Т. О., *к.ф.-м.н., доц., Полтавський національний педагогічний університет імені В. Г. Короленка, Полтава, Україна. https://orcid.org/0000-0002-8755-9020, ptkm@ukr.net*

Подошвелев Ю. Г., *к.ф.-м.н., доц., Полтавський національний педагогічний університет імені В. Г. Короленка, Полтава, Україна. https://orcid.org/0000-0002-3394-2809, optimist1618@outlook.com*

Анотація. *Стаття присвячена огляду результатів останніх років у галузі стохастичної комбінаторної оптимізації. Формулювання задач в умовах невизначеності, у тому числі ймовірнісної, вимагає уточнення поняття екстремуму. У роботах вітчизняних науковців запропоновано різні підходи до постановок задач комбінаторної оптимізації зі стохастичною невизначеністю. Один із підходів пропонується для досить широкого класу оптимізаційних задач і розуміє невизначеність як неоднозначність коефіцієнтів функціонала оптимізації. Пропонуються компромісні критерії, обґрунтовано алгоритми побудови компромісних розв'язків як для загальної постановки оптимізаційної задачі, так і для окремих класів задач (транспортна задача, задача дробово-лінійного програмування, задача календарного планування). Інший підхід до формалізації задач з імовірнісною невизначеністю ґрунтується на введенні відношення порядку на множині випадкових величин і допускає, що випадковими величинами можуть бути як коефіцієнти цільової функції, так і компоненти розв'язку. Досліджується розв'язування задач у такій постановці на загальній множині розміщень. Для задач з лінійною цільовою функцією без додаткових обмежень встановлено ряд властивостей екстремалі, які дозволяють зменшувати вимірність задачі, що розв'язується. Для задач з лінійною цільовою функцією і додатковими обмеженнями запропоновано алгоритм у рамках методології гілок і меж. Дослідження також проводяться і в напрямі врахування стохастичної невизначеності під час формулювання окремих класів оптимізаційних задач, зокрема задачі комівояжера або плану формування поїздів. Автори робіт розглядають задачі, у яких окремі параметри є нормально розподіленими випадковими величинами, та пропонують обчислювальні схеми для розв'язування сформульованих задач.*

Ключові слова: *комбінаторна оптимізація, стохастична оптимізація, задачі оптимізації на розміщеннях, задача комівояжера, компромісний критерій.*

STOCHASTIC COMBINATORIAL PROBLEMS: A REVIEW OF LATEST RESULTS

Tetiana Barbolina, *Dr. habil., Ass. Prof., Poltava V. G. Korolenko National Pedadogical University, Poltava, Ukraine. https://orcid.org/0000-0002-4596-7907, tm-b@gsuite.pnpu.edu.ua*

Tetiana Barannyk, *Ph.D., Ass. Prof., Poltava V. G. Korolenko National Pedadogical University, Poltava, Ukraine. https://orcid.org/0009-0003-0509-9935, barannykt@gmail.com*

Tetiana Kononovych, *Ph.D., Ass. Prof., Poltava V. G. Korolenko National Pedadogical University, Poltava, Ukraine. https://orcid.org/0000-0002-8755-9020, ptkm@ukr.net*

Yurii Podoshvelev, *Ph.D., Ass. Prof., Poltava V. G. Korolenko National Pedadogical University, Poltava, Ukraine. https://orcid.org/0000-0002-3394-2809, optimist1618@outlook.com*

Abstract. *The article is dedicated to reviewing recent results in the field of stochastic combinatorial optimization. The formulation of problems under uncertainty, including probabilistic uncertainty, requires a refinement of the concept of extremum. Domestic researchers have proposed various approaches to the formulation of combinatorial optimization problems under stochastic uncertainty. One of the approaches is proposed for a quite general class of optimization problems and interprets uncertainty as ambiguity in the*

coefficients of the optimization functional. Compromise criteria are proposed, and algorithms for constructing compromise solutions are substantiated, for the general formulation of the optimization problem as well as for specific classes of problems (such as the transportation problem, fractional-linear programming problem, and scheduling problem).

Another approach to the formalization of problems with probabilistic uncertainty is based on the introduction of an order relation on the set of random variables and assumes that the random variables can represent the coefficients of the objective function as well as the components of the solution. They investigate solving of problems on the general set of arrangements in such formulation. For problems with a linear objective function without additional constraints, a number of properties of extremal have been established. These properties allow to reduce the problem's dimension. For problems with a linear objective function and additional constraints, they propose the algorithm within the framework of the branch-and-bound methodology.

Research is also being conducted in the direction of accounting for stochastic uncertainty in the formulation of specific classes of optimization problems, such as the traveling salesman problem or train formation scheduling. The authors consider problems where certain parameters are normally distributed random variables and propose computational schemes for solving the formulated problems.

Key words: *combinatorial optimization, stochastic optimization, optimization problems on arrangements, travelling salesman problem, compromise criteria.*

Вступ

Оптимізаційні задачі різної природи привертають увагу багатьох дослідників в усьому світі. Важливим класом таких задач є задачі оптимізації з обмеженнями комбінаторного характеру, які мають численні застосування у різних галузях людської діяльності. Сучасні потреби моделювання вимагають також врахування у постановках задач невизначеностей вхідних даних. Досить широка бібліографія присвячена задачам з інтервальною, нечіткою невизначеністю, багатокритеріальних задач тощо. Актуальним напрямом досліджень також є розв'язування задач стохастичної комбінаторної оптимізації (див., наприклад, [1–4]).

Ця стаття присвячена огляду отриманих в останнє десятиліття результатів вітчизняних досліджень задач комбінаторної оптимізації з імовірнісною невизначеністю.

Виконання досліджень

Досить загальний клас задач комбінаторної оптимізації в умовах невизначеності введено О. А. Павловим в [5]. Автором розглядаються задачі в такій постановці:

$$\min_{\sigma\in\Omega}\sum_{i=1}^{s}\omega_i k_i(\sigma),$$

де Ω – множина допустимих розв'язків, ω_i – числа, $k_i(\sigma)$ – i-та числова характеристика допустимого розв'язку σ. Вважається, що сформульована задача має ефективний алгоритм розв'язування, який не може бути застосований у випадку зміни структури множини Ω.

Під невизначеністю у цій постановці розуміється невизначеність значень коефіцієнтів ω_i: існують набори значень ваг $\{\omega_i^l, i = 1, \dots, s\}, l = 1, \dots, L$, кожен з яких може бути набором коефіцієнтів задачі на етапі реалізації її розв'язку. У випадку імовірнісної невизначеності задаються додатні ймовірності для кожного можливого набору значень ваг.

$$\min_{\sigma\in\Omega}\sum_{i=1}^{s}\left(\sum_{l=1}^{L}a_l\omega_i^l\right)k_i(\sigma). \qquad (1)$$

Ідеї, запропоновані в [5], знайшли розвиток у наступних роботах О. А. Павлова та його учнів. Так, робота [6] присвячена питанням знаходження експертних вагових коефіцієнтів, використання яких передбачено окремими критеріями компромісних рішень. На основі попередніх робіт автора та його учнів сформульований алгоритм, який містить ефективну процедуру визначення експертних вагових коефіцієнтів на основі емпіричних матриць попарних порівнянь.

У роботах [7–10] та інших застосовано запропонований у [5] підхід до формалізації та розв'язування в умовах невизначеності окремих класів оптимізаційних задач. Зокрема, у [7] представлено розв'язування транспортної задачі як задачі комбінаторної оптимізації в умовах невизначеності вигляду (1), наведено приклади індивідуальних задач як ілюстрація ефективності вибраного підходу. Залежно від вибраного критерію пошук компромісного розв'язку сформульованої в роботі транспортної задачі може вимагати розв'язування однієї детермінованої транспортної задачі або застосування ітераційної процедури. Ще один підхід до розв'язування задачі (1), що демонструється на прикладі розв'язування транспортної задачі в умовах невизначеності, запропоновано у [8].

У роботі [9] показано можливість застосування викладеної в [5] конструктивної теорії знаходження компромісного розв'язку до розв'язування в умовах невизначеності задач дробово-лінійного програмування: наведено дві постановки задачі дробово-лінійного програмування в умовах невизначеності (у першій невизначеність стосується значень коефіцієнтів чисельника, у другій – значень коефіцієнтів знаменника), запропоновано кілька компромісних критеріїв, обґрунтовано побудову компромісного розв'язку для першої постановки задачі, наведено результати експериментів з вивчення залежності вихідних даних задачі від окремих параметрів.

У статті [10] в умовах невизначеності розглядається двоетапна задача календарного планування (неоднозначність стосується векторів вагових коефіцієнтів критеріїв для кожного пристрою на другому етапі задачі).

Дослідження оптимізаційних задач з комбінаторними обмеженнями, у тому числі в умовах невизначеності, здійснюється також у рамках такого наукового напряму, як евклідова комбінаторна оптимізація. У роботах засновника цього напряму Ю. Г. Стояна та численних представників цієї наукової школи вивчаються властивості евклідових комбінаторних множин, побудова математичних моделей практичних задач як евклідових задач комбінаторної оптимізації, обґрунтовуються методи розв'язування таких задач.

Для задач з імовірнісною невизначеністю у роботах [11; 12] запропоновано підхід до постановок задач, який ґрунтується на введенні відношення порядку на множині випадкових величин.

Нехай $(\Omega, \mathcal{F}, P)$ – деякий імовірнісний простір. Випадкові величини, визначені на $(\Omega, \mathcal{F}, P)$, позначатимемо літерами напівжирного накреслення (a, b).

Упорядкування дискретних випадкових величин, серед можливих значень яких є найменше, у [11] пропонується здійснювати на основі послідовного порівняння математичних сподівань, дисперсій, можливих значень та відповідних імовірностей випадкових величин. У другому підході до упорядкування (див., наприклад, [12]) розглядається характеристичний вектор випадкової величини, компонентами якого є визначені її числові характеристики. Здійснюється розбиття заданої множини незалежних випадкових величин за еквівалентністю на основі рівності характеристичних векторів, на отриманій фактор-множині вводиться лінійний порядок відповідно до лексикографічного порядку характеристичних векторів представників цих класів. Послідовне порівняння числових характеристик випадкових величин дозволяє краще враховувати особливості реальних ситуацій, ніж заміна випадкових величин однією з їх числових характеристик.

Ґрунтуючись на розглянутих підходах до впорядкування випадкових величин, запропоновано постановки оптимізаційних задач, які використовують розуміння максимуму (мінімуму) у відповідності до зазначених відношень порядку (якщо порядок вводиться відповідно до першого підходу, то говоритимемо про *S*-задачі, якщо відповідно до другого – *H*-задачі). Представлено як задачі, у яких випадковими величинами є коефіцієнти цільової функції, так і такі, де випадковими величинами є компоненти розв'язку.

В інших роботах цих авторів розглянуто розв'язування задач стохастичної комбінаторної оптимізації на загальній множині розміщень $E_\eta^k(G)$, під якою розуміють множину всіх упорядкованих k-вибірок із мультимножини $G = \{g_1, g_2, \dots, g_\eta\}$ вигляду $(g_{i_1}, g_{i_2}, \dots, g_{i_k})$, де $g_{i_j} \in G,\ i_j \neq i_t\ \ \forall i_j, i_t \in J_n\ \ \forall j, t \in J_k$ (тут і далі J_n позначає множину *n* перших натуральних чисел).

Розв'язування лінійної безумовної *H*-задачі стохастичної комбінаторної оптимізації на розміщеннях, тобто задачі вигляду

$$L(\boldsymbol{x}^*) = \operatorname*{extr}_{\boldsymbol{x} \in E_\eta^k(G)} \sum_{j=1}^{k} c_j \boldsymbol{x}_j, \qquad \boldsymbol{x}^* = \arg \operatorname*{extr}_{\boldsymbol{x} \in E_\eta^k(G)} \sum_{j=1}^{k} c_j \boldsymbol{x}_j,$$

розглядається в роботі [13].

Встановлено, що коли характеристичний вектор випадкової величини задовольняє умову

$$h_i(u\boldsymbol{a} + v\boldsymbol{b}) = u_i^{\lambda_i} h(\boldsymbol{a}) + v_i^{\lambda_i} h(\boldsymbol{b})\ \ \forall i \in J_s,$$

причому $H_{r-1}(\boldsymbol{g}_i) = H_{r-1}(\boldsymbol{g}_j)\ \forall i, j \in J_n, \boldsymbol{g}_i, \boldsymbol{g}_j \in G$ і з $c_j \neq c_j$ випливає $c_i^{\lambda_r} \neq c_j^{\lambda_r}\ \forall r \in J_s$, то *r*-та компонента однієї з мінімалей у розв'язку стохастичної задачі може бути одержана з розв'язку спеціальної детермінованої лінійної безумовної задачі комбінаторної оптимізації на розміщеннях. Крім того, показано, що мінімаль вихідної задачі може бути сформована із мінімалей у розв'язках *H*-задач стохастичної комбінаторної оптимізації на розміщеннях менших вимірностей.

У випадку, коли всі коефіцієнти цільової функції мають однаковий знак, мінімаль у розв'язку *H*-задачі може бути сформована без розв'язування детермінованих задач. Останні твердження покладені в основу редукційного методу розв'язування лінійної безумовної задачі стохастичної комбінаторної оптимізації на розміщеннях. У роботах [14–17] одержано аналогічні результати для *S*-задач.

Статті [12; 18] описують застосування алгоритму в рамках методології гілок і меж для розв'язування лінійних умовних задач стохастичної комбінаторної оптимізації на розміщеннях. Розглядаються

задачі, у яких або коефіцієнт цільової функції, або елементи мультимножини є незалежними випадковими величинами з додатним значенням першої компоненти характеристичного вектора.

Разом із розглянутими напрямами відзначимо також роботи, присвячені обґрунтуванню методів розв'язування окремих класів задач комбінаторної оптимізації зі стохастичною невизначеністю.

Л. В. Сухомлин у [19] розглядає застосування декомпозиційного генетичного алгоритму до розв'язування задачі комівояжера, у якій «відстані» між пунктами є випадковими величинами із заданими щільностями розподілу. Автором вводиться поняття «віддаленість одного пункту від іншого», яка тим більша, цим більша ймовірність того, що випадкова «відстань» між пунктами виявиться більшою певного порогового значення.

Для випадку, коли випадкові величини є нормально розподіленими, описано ітераційну процедуру розв'язування задачі. На першому етапі відбувається кластеризація пунктів обходу з урахуванням імовірнісної невизначеності у відстанях між пунктами. У результаті цього етапу одержується кілька груп пунктів, у кожній з яких розраховується положення центру ваги. Для кожної пари суміжних кластерів знаходиться найменша відстань поєднання. На наступному етапі розв'язується стохастична задача знаходження найкоротшого маршруту між заданими пунктами у кожному кластері. Розв'язання цієї задачі передбачає використання стохастичного аналогу генетичного алгоритму, для якого обґрунтовано критерій відбору у разі формування нової популяції. Розв'язок вихідної задачі одержується з'єднанням індивідуальних маршрутів кластерів за допомогою «перемичок», знайдених на попередніх етапах.

У роботах [20; 21] розглядається складання плану формування поїздів як задачі комбінаторної оптимізації. Розробка якісного плану формування поїздів розглядається як найбільш ефективний шлях мінімізації добових вагоно-годин накопичення составів. План формування поїздів може бути поліпшений, якщо вважати параметри накопичення випадковими величинами, які набувають значень у певному інтервалі, у межах якого можна на них впливати.

Урахування в такий спосіб імовірнісної невизначеності вимагає уточнення цільової функції для відображення можливого ризику неотримання значення для виконання плану формування поїздів, який виникає, якщо параметр накопичення набуває значення, які є відхиленнями у бік зменшення від математичного сподівання. Як кількісна міра ризику приймається добуток імовірності настання небажаної події на величину втрат у разі її настання.

Як небажана подія розглядається перевищення прийнятої у разі побудови плану формування поїздів величини параметра накопичення. Для випадку, коли параметри накопичення є нормально розподіленими випадковими величинами, імовірність цієї події обчислюється за формулою

$$P = 1 - \frac{1}{2}\left(1 + erf\left(\frac{c - \bar{\bar{c}}}{\sigma\sqrt{2}}\right)\right),$$

де erf(·) – функція похибок Лапласа, c – параметр накопичення, $\bar{\bar{c}}$ – математичне сподівання величини параметра накопичення, а σ^2 – його дисперсія.

Величина втрат визначається як різниця між математичним сподіванням параметра накопичення і його значенням, помножена на вартість вагоно-годин і норму кількості вагонів у складі поїзда. Крім зазначених втрат, цільова функція відображає витрати з окремими операціями, що здійснюються у процесі формування поїздів. Розв'язування сформульованої задачі здійснюється за допомогою генетичного алгоритму.

Висновок

Урахування невизначеності вхідних даних, у тому числі ймовірнісної, у постановках оптимізаційних задач вимагає уточнення розуміння того, який розв'язок є допустимим, яким чином визначається оптимальний розв'язок. Останніми роками було запропоновано різні підходи до формалізації задач стохастичної комбінаторної оптимізації та обґрунтовано алгоритми розв'язування окремих класів задач у таких постановках. Актуальною є розробка ефективних методів розв'язування задач стохастичної комбінаторної оптимізації як у загальновідомих, так і нових постановках.

Література

1. Angel A. Juan, Javier Faulin, Scott E. Grasman, Markus Rabe, Gonçalo Figueira. A review of simheuristics: Extending metaheuristics to deal with stochastic combinatorial optimization problems. *Operations Research Perspectives*. 2015. Vol. 2. P. 62–72. DOI: https://doi.org/10.1016/j.orp.2015.03.001.

2. Buchheim C., Pruente J. K-adaptability in stochastic combinatorial optimization under objective uncertainty. *European Journal of Operational Research*. 2019. Vol. 277. Issue 3. P. 953–963. DOI: https://doi.org/10.1016/j.ejor.2019.03.045.

3. Archetti C., Feillet D., Mor A., Speranza M. G. Dynamic traveling salesman problem with stochastic release dates. *European Journal of Operational Research*. 2020. Vol. 280, Issue 3. P. 832–844. DOI: https://doi.org/10.1016/j.ejor.2019.07.062.

4. Oyola J., Arntzen H., Woodruff D. L. The stochastic vehicle routing problem: a literature review. Part I: models. *EURO J Transp Logist*. 2018. Iss.7. P. 193–221. DOI: https://doi.org/10.1007/s13676-016-0100-5.

5. Pavlov A. A. Optimization for one class of combinatorial problems under uncertainty. *Адаптивні системи автоматичного управління*. 2019. Т. 1. № 34. С. 81–89. DOI: https://doi.org/10.20535/1560- 8956.1.2019.178233.

6. Pavlov A. Combinatorial optimization under uncertainty and formal models of expert estimation. *Вісник Національного технічного університету «ХПІ». Серія : Системний аналіз, управління та інформаційні технології*. 2019. № 1. С. 3–7. DOI: https://doi.org/10.20998/2079-0023.2019.01.01.

7. Pavlov A. A., Zhdanova E. G. The transportation problem under uncertainty. *J. Autom. Inform. Sci*. 2020. T. 52, № 4. P. 1–13. DOI: https://doi.org/10.1615/JAutomatInfScien.v52.i4.10.

8. Pavlov A. A., Zhdanova E. G. Finding a compromise solution to the transportation problem under uncertainty. *Адаптивні системи автоматичного управління*. 2020. Т. 1. № 36. С. 60–72. DOI: https://doi.org/10.20535/1560-8956.36.2020.209764.

9. Павлов О., Вознюк О., Жданова О. Задача дробово-лінійного програмування в умовах невизначеності. *Вісник Національного технічного університету «ХПІ». Серія : Системний аналіз, управління та інформаційні технології*. 2021. № 1 (5). С. 20–28. DOI: https://doi.org/10.20998/2079-0023.2021.01.04.

10. Павлов О., Халус О., Місюра О., Мельников О., Медведєв М. ПДС-алгоритми для двоетапної задачі календарного планування в детермінованій постановці та в умовах невизначеності. *Адаптивні системи автоматичного управління*. 2023. № 1(42). С. 184–196. DOI: https://doi.org/10.20535/1560-8956.42.2023.279170.

11. Iemets O. O., Barbolina T. N. Combinatorial Optimization Model of Packing Rectangles with Stochastic Parameters. *Cybernetics and Systems Analysis*. 2015. Vol. 51. Iss. 4. P. 583–593. DOI: https://doi.org/10.1007/s10559-015-9749-2.

12. Ємець О. О., Барболіна Т. М. Метод гілок і меж розв'язування задач оптимізації лінійної цільової функції на розміщеннях з імовірнісною невизначеністю. *Вісник Запорізького національного університету. Фізико-математичні науки*. 2018. № 2. С. 43–54.

13. Iemets O. O., Barbolina T. M. Solving Linear Unconstrained Problems of Combinatorial Optimization on Arrangements Under Stochastic Uncertainty. *Cybernetics and Systems Analysis*. 2016. Vol. 52. Iss. 3. P. 457–466. DOI: https://doi.org/10.1007/s10559-016-9846-x.

14. Ємець О. О., Барболіна Т. М. Побудова і дослідження математичної моделі задачі директора зі стохастичними параметрами. *Вісник Черкаського університету. Серія : Прикладна математика. Інформатика*. 2014. № 18 (311). С. 3–11.

15. Ємець О. О., Барболіна Т. М. Лінійні оптимізаційні задачі на розміщеннях з імовірнісною невизначеністю: властивості і розв'язання. *Системні дослідження та інформаційні технології*. 2016. № 1. С. 107–119. DOI: 10.20535/SRIT.2308-8893.2016.1.11

16. Iemets O. O., Barbolina T. M. Properties of the linear unconditional problem of combinatorial optimization on arrangements under probabilistic uncertainty. *Cybernetics and Systems Analysis*. 2016. Vol. 52. Iss. 2. P. 285–295. DOI: https://doi.org/10.1007/s10559-016-9825-2.

17. Ємець О. О., Барболіна Т. М. Властивості лінійних безумовних задач оптимізації на розміщеннях з імовірнісною невизначеністю. *Доповіді НАН України*. 2016. № 2. С. 31–37. DOI: http://dx.doi.org/10.15407/dopovidi2016.02.031.

18. Ємець О. О., Барболіна Т. М. Стохастична оптимізація на розміщеннях: властивості лінійних безумовних задач. *Вісник Запорізького національного університету. Фізико-математичні науки*. 2017. № 1. С. 147–158.

19. Сухомлин Л. В. Стохастична задача маршрутизації високої розмірності в умовах неточно заданих вихідних даних. *Вісник Кременчуцького національного університету імені Михайла Остроградського.* 2015. № 5. С. 149–154. URL: https://visnikkrnu.kdu.edu.ua/statti/2015_5_149-5-2015.pdf (дата звернення: 26.10.2024).

20. Прохоров В. М. Розробка методу розрахунку плану формування поїздів на основі стохастичної комбінаторної оптимізації. *ScienceRise*. 2016. Т.12, № 2(29). С. 53–56. DOI: 10.15587/2313-8416.2016.84109.

21. Прохоров В. М., Рябушка Ю. А. Розрахунок плану формування поїздів на основі стохастичної комбінаторної оптимізації. *Збірник наукових праць Українського державного університету залізничного транспорту*. 2016. Вип. 165. С. 215–224.

References

1. Juan, A. A., Faulin, J., Grasman, S. E., Rabe, M., & Figueira, G. (2015). A review of simheuristics: Extending metaheuristics to deal with stochastic combinatorial optimization problems. *Operations Research Perspectives*, 2, 62–72. https://doi.org/10.1016/j.orp.2015.03.001.

2. Buchheim, C. & Pruente, J. (2019). K-adaptability in stochastic combinatorial optimization under objective uncertainty. *European Journal of Operational Research,* 277(3), 953–963. https://doi.org/10.1016/j.ejor.2019.03.045.

3. Archetti, C., Feillet, D., Mor, A. & Speranza, M. G. (2020). Dynamic traveling salesman problem with stochastic release dates. *European Journal of Operational Research,* 280(3), 832–844. https://doi.org/10.1016/j.ejor.2019.07.062.

4. Oyola, J., Arntzen, H. & Woodruff, D. L. (2018). The stochastic vehicle routing problem: a literature review. Part I: models. *EURO J Transp Logist*, 7, 193–221. https://doi.org/10.1007/s13676-016-0100-5.

5. Pavlov, A. A. (2019). Optimization for one class of combinatorial problems under uncertainty. *Adaptive systems of automatic control*, 1(34). 81–89. https://doi.org/10.20535/1560-8956.1.2019.178233.

6. Pavlov, A. (2019). Combinatorial optimization under uncertainty and formal models of expert estimation. *Bulletin of National Technical University "KhPI". Series: System Analysis, Control and Information Technologies*, 1, 3–7. https://doi.org/10.20998/2079-0023.2019.01.01.

7. Pavlov, A. A. & Zhdanova, E. G. (2020). The transportation problem under uncertainty. *J. Autom. Inform. Sci.* 52(4), 1–13. https://doi.org/10.1615/JAutomatInfScien.v52.i4.10.

8. Pavlov, A. A. & Zhdanova, E. G. (2020). Finding a compromise solution to the transportation problem under uncertainty. *Adaptive systems of automatic control*, 1(36), 60–72. https://doi.org/10.20535/1560-8956.36.2020.209764.

9. Pavlov, O., Vozniuk, O. & Zhdanova, O. (2021). Zadacha drobovo-liniinogo programuvannia v umovah nevyznachenosti [The linear-fractional programming problem under uncertainty conditions]. *Bulletin of National Technical University "KhPI". Series: System Analysis, Control and Information Technologies*, 1 (5), 20–28. https://doi.org/10.20998/2079-0023.2021.01.04 [in Ukrainian].

10. Pavlov, A., Khalus E., Misura E., Melnikov O. & Medvediev M. (2023). PDS-alhorytmy dlia dvoetapnoi zadachi kalendarnoho planuvannia v determinovanii postanovtsi ta v umovakh nevyznachenosti [PSC-algorithms for a two-stage scheduling

problem in the deterministic formulation and under uncertainty conditions]. *Adaptive systems of automatic control*, 1(42), 184–196. https://doi.org/10.20535/1560-8956.42.2023.279170 [in Ukrainian].

11. Iemets, O. O. & Barbolina, T. N. (2015). Combinatorial Optimization Model of Packing Rectangles with Stochastic Parameters. *Cybernetics and Systems Analysis*, 51(4), 583–593. https://doi.org/10.1007/s10559-015-9749-2.

12. Iemets, O. O. & Barbolina, T. M. (2018). Metod hilok i mezh rozviazuvannia zadach optymizatsii liniinoi tsilovoi funktsii na rozmishchenniakh z imovirnisnoiu nevyznachenistiu [Branch and bound method for solving problems of optimization of linear objective function on arrangements under probabilistic uncertainty]. *Visnyk of Zaporizhzhya National University. Physical and Mathematical Sciences*, 2, 43–54 [in Ukrainian].

13. Iemets, O. O. & Barbolina, T. M. (2016). Solving Linear Unconstrained Problems of Combinatorial Optimization on Arrangements Under Stochastic Uncertainty. *Cybernetics and Systems Analysis,* 52(3), 457–466. https://doi.org/10.1007/s10559-016-9846-x.

14. Iemets, O. O. & Barbolina, T. M. (2014). Pobudova i doslidzhennia matematychnoi modeli zadachi dyrektora zi stokhastychnymy parametramy. *Cherkasy University Bulletin: Applied Mathematics. Informatics*, 18, 3–11.

15. Iemets, O. O. & Barbolina, T. M. (2016). Liniini optymizatsiini zadachi na rozmishchenniakh z imovirnisnoiu nevyznachenistiu: vlastyvosti i rozviazannia [Linear optimization problems on arrangements under probabilistic uncertainty: properties and solution]. *System research and information technologies,* 1, 107–119. 10.20535/SRIT.2308-8893.2016.1.11 [in Ukrainian].

16. Iemets, O. O. & Barbolina, T. M. (2016). Properties of the linear unconditional problem of combinatorial optimization on arrangements under probabilistic uncertainty. *Cybernetics and Systems Analysis*, 52(2), 285–295. https://doi.org/10.1007/s10559-016-9825-2.

17. Iemets, O. O. & Barbolina, T. M. (2016). Vlastyvosti liniinykh bezumovnykh zadach optymizatsii na rozmishchenniakh z imovirnisnoiu nevyznachenistiu [Properties of linear unconditional optimization problems on arrangements under probabilistic uncertainty]. *Reports of the National Academy of Sciences of Ukraine*, 2, 31–37. http://dx.doi.org/10.15407/dopovidi2016.02.031 [in Ukrainian].

18. Iemets, O. O. & Barbolina, T. M. (2017). Vlastyvosti liniinykh bezumovnykh zadach optymizatsii na rozmishchenniakh z imovirnisnoiu nevyznachenistiu [Stochastic optimization on arrangements: properties of linear unconstrained problems]. *Visnyk of Zaporizhzhya National University. Physical and Mathematical Sciences,* 1, 147–158 [in Ukrainian].

19. Sukhomlyn, L. (2015). Stokhastychna zadacha marshrutyzatsii vysokoi rozmirnosti v umovakh netochno zadanykh vykhidnykh danykh [Stochastic routing problem of high dimension in conditions inaccurately given basic data]. *Transactions of Kremenchuk Mykhailo Ostrohradskyi National University*, (5), 149–154.

20. Prokhorov, V. (2016). Rozrobka metodu rozrakhunku planu formuvannia poizdiv na osnovi stokhastychnoi kombinatornoi optymizatsii [Development of calculation method for railway train formation plan on the basis of stochastic combinatorial optimization]. *ScienceRise*, 12(2), 53–56. https://doi.org 10.15587/2313-8416.2016.84109 [in Ukrainian].

21. Prokhorov, V. M. & Riabushka, Yu. A. (2016). Rozrakhunok planu formuvannia poizdiv na osnovi stokhastychnoi kombinatornoi optymizatsii [Computation of trains formation plan on the base of stochastic combinatorial optimization]. *Collected scientific works of Ukrainian State University of Railway Transport,* 165, 215–224 [in Ukrainian].

УДК 004.056:004.056.5:004.056.2
DOI https://doi.org/10.36994/2788-5518-2024-02-08-04

МЕТОД ЗАХИСТУ ІНФОРМАЦІЇ НА МОБІЛЬНИХ ПРИСТРОЯХ НА ОСНОВІ АДАПТИВНОЇ ПОВЕДІНКОВОЇ МОДЕЛІ

Бровченко Є. М., *Відкритий міжнародний університет розвитку людини «Україна», Інститут комп'ютерних технологій, Київ, Україна. https://orcid.org/0000-0002-1416-0385, yebrovchenko@hotmail.com*

Самарай В. П., *к.т.н., с.н.с., доцент, Відкритий міжнародний університет розвитку людини «Україна», Інститут комп'ютерних технологій, Київ, Україна. https://orcid.org/0000-0003-4419-1366, samaraj@ukr.net*

Анотація. *Стаття присвячена проблемі захисту неструктурованої інформації на мобільних пристроях на основі адаптивної поведінкової моделі. Досліджено сучасні виклики інформаційної безпеки, що виникають у зв'язку з активним використанням мобільних пристроїв у цифровому середовищі. Запропоновано новий підхід до захисту даних, що включає машинне навчання, поведінкову аналітику та адаптивні механізми ідентифікації загроз. Використання штучного інтелекту та поведінкових моделей безпеки сприяє вдосконаленню механізмів ідентифікації загроз та адаптивного реагування на можливі ризики. Розглянуто основні методи, такі як багатофакторна автентифікація, криптографічні алгоритми, блокчейн та квантово-стійке шифрування, які забезпечують комплексний рівень безпеки. Проведені експериментальні дослідження продемонстрували ефективність адаптивного підходу у виявленні та попередженні потенційних загроз для інформаційної безпеки. Методи захисту інформації засновувалися на засадах інтеграційного машинного навчання, математичних моделей прогнозування загроз, алгоритмів шифрування та багатофакторної автентифікації. Аспекти захисту: моніторинг, адаптація, прогнозування загроз, прийняття рішень, шифрування та біометрична автентифікація. Кожен етап включає підхід із використанням формул і відповідних алгоритмів. Отримані результати підтверджують, що впровадження інноваційних моделей захисту дозволяє підвищити рівень кібербезпеки мобільних пристроїв, зменшуючи ризики несанкціонованого доступу до конфіденційної інформації. Використання адаптивних моделей дозволить ефективно виявляти загрози та реагувати на аномальну поведінку користувачів у реальному часі. Блокчейн та квантово-стійке шифрування сприятимуть підвищенню безпеки даних, а концепція «приватності за замовчуванням» гарантуватиме інтеграцію кібербезпеки на всіх етапах розробки. Комплексний підхід, що поєднує технологічні рішення, підвищення обізнаності користувачів та нормативне регулювання, стане ключем до ефективного захисту мобільної інформації.*

Ключові слова: *мобільні пристрої, захист інформації, неструктурована інформація, кібербезпека, поведінкова аналітика, машинне навчання, багатофакторна автентифікація.*

ADAPTIVE BEHAVIORAL MODEL-BASED METHOD FOR INFORMATION PROTECTION ON MOBILE DEVICES

Evgen Brovchenko, *Open International University of Human Development "Ukraine", Kyiv, Ukraine. https://orcid.org/0000-0002-1416-0385, yebrovchenko@hotmail.com*

Valery Samarai, *Ph.D. in Technical Sciences, Senior Research Fellow, Associate Professor, Open International University of Human Development "Ukraine", Kyiv, Ukraine. https://orcid.org/0000-0003-4419-1366, samaraj@ukr.net*

Abstract. *This paper addresses the issue of protecting unstructured information on mobile devices based on an adaptive behavioral model. The increasing role of mobile devices in digital ecosystems necessitates the development of advanced security solutions that ensure data integrity, confidentiality, and resilience to cyber threats. The research focuses on modern cybersecurity challenges arising from the widespread use of mobile devices, including data breaches, unauthorized access, phishing attacks, and malware exploitation. The study explores an innovative approach that integrates behavioral analytics, machine learning, and adaptive threat identification mechanisms to enhance security measures. Key methods such as multi-factor authentication (MFA), cryptographic algorithms, blockchain, and quantum-resistant encryption are examined to establish a comprehensive security framework. Experimental research was conducted using various mobile security scenarios, including different levels of threat exposure, data transmission security, and real-time behavioral monitoring. The results demonstrated that adaptive behavioral models significantly improve the detection and mitigation of cyber threats, dynamically adjusting security policies*

based on user interactions and environmental conditions. By leveraging artificial intelligence and behavioral biometrics, this approach enhances authentication accuracy, reduces the risk of unauthorized access, and strengthens protection against evolving cyberattacks. The study findings highlight the importance of combining advanced encryption techniques with behavioral-based security mechanisms to create a proactive cybersecurity strategy. Implementing adaptive security models enables mobile devices to function as secure elements within digital infrastructures, ensuring both user convenience and high-level protection. The proposed methodology contributes to the ongoing development of intelligent mobile security systems, offering effective solutions for safeguarding sensitive information in an increasingly connected world.

Key words: *mobile devices, information security, unstructured data protection, cybersecurity, behavioral analytics, machine learning, multi-factor authentication.*

Вступ

Інтенсивний розвиток сучасних інформаційних технологій вимагає впровадження інноваційних рішень із захисту даних. При цьому безпека інформації на мобільних пристроях стає критично важливою з урахуванням їх ключової ролі і в приватному житті, і в професійній діяльності людей.

Використання штучного інтелекту та поведінкових моделей безпеки сприяє вдосконаленню механізмів ідентифікації загроз та адаптивного реагування на можливі ризики. Мобільні пристрої – вже не просто засоби комунікації, а складові частини інформаційних екосистем, що зберігають значні обсяги конфіденційних даних. У майбутньому інтеграція між користувачем та пристроєм стане ще більш динамічною, що підвищить вимоги до ефективності захисних механізмів. З огляду на постійний розвиток кіберзагроз та непередбачуваність атак перед фахівцями стоїть завдання розробки гнучких, адаптивних методів захисту, що можуть автоматично аналізувати поведінкові характеристики користувачів, визначати аномальні дії та застосовувати відповідні контрзаходи. Запровадження адаптивної поведінкової моделі безпеки дозволяє створити проактивний підхід до захисту даних, що забезпечує не лише виявлення загроз у реальному часі, а й їхнє прогнозування на основі аналізу поведінки користувача та змін у його середовищі. Таким чином, мобільні пристрої мають бути розглянуті не лише як частина цифрової екосистеми, а й як динамічні елементи, що інтегрують передові засоби інформаційної безпеки.

Безпека неструктурованих даних, що зберігаються на мобільних пристроях, є ключовим аспектом кіберзахисту [3; 8]. Такі дані включають текстові файли, електронну пошту, мультимедійний контент (фотографії, відео, аудіозаписи) та інші інформаційні об'єкти, що можуть бути вразливими до несанкціонованого доступу. Основні загрози для інформаційної безпеки мобільних пристроїв включають ризик втрати або крадіжки пристрою [7], експлуатацію вразливостей операційної системи [9] та додатків, недостатній рівень шифрування, слабкі паролі та нехтування принципами кібергігієни [4; 6; 10]. Запровадження адаптивних поведінкових моделей безпеки дозволяє ефективніше протидіяти цим загрозам, аналізуючи звички користувача та виявляючи аномальну активність [2]. Такий підхід дає змогу в режимі реального часу адаптувати механізми захисту залежно від контексту використання пристрою, рівня ризику та зовнішніх факторів. Водночас сучасні архітектурні рішення безпеки не завжди враховують необхідність гнучкої інтеграції з адаптивними методами захисту [7], що може створювати додаткові виклики з погляду продуктивності та економічної ефективності впровадження.

Аналіз сфери кібербезпеки та результати досліджень у галузі поведінкової аналітики, адаптивних систем автентифікації [1; 2], методів шифрування даних на мобільних пристроях [3; 7], а також сучасних підходів до виявлення аномалій [6; 8] підкреслюють важливість захисту інформації та зростаючий попит на надійні рішення для захисту інформаційних систем і мобільних пристроїв. Захист інформації є комплексним завданням: від етапу розробки до контролю під час експлуатації, включаючи інформаційну обізнаність і відповідальність кінцевого користувача [9]. Кібербезпека переходить на новий рівень, де лише комплексний підхід здатен забезпечити ефективні рішення [10; 11].

Аналіз останніх досліджень і публікацій

Аналіз особливостей використання мобільних пристроїв як у особистих, так і корпоративних цілях дає змогу виокремити ключові тенденції та методи, що застосовуються для адаптивного захисту інформації [2]. Основою такої системи є біометричний доступ [1]. Постійний розвиток у сфері біометрії сприяє підвищенню точності та надійності цих технологій. Додатково важливим компонентом інформаційного захисту є шифрування, яке вдосконалюється завдяки новітнім алгоритмам та методам безпечного зберігання ключів [6]. Серед актуальних рішень для мобільних пристроїв варто відзначити використання віртуальних приватних мереж (VPN), які, однак, не завжди ефективно застосовуються користувачами через складність налаштування або недостатню обізнаність [8]. Окрему увагу слід приділити захисту від шкідливого програмного забезпечення та вірусних загроз. Фахівці з кібербезпеки безперервно аналізують нові загрози, такі як фітингові атаки, експлойти нульового дня, мобільні троянські програми тощо [4], та розробляють ефективні методи їх виявлення й нейтралізації [5; 7; 10]. Таким чином, адаптивний захист інформації на мобільних пристроях базується на комбінації біометричної автентифікації, сучасних методів шифрування, VPN-технологій та антивірусного захисту, що разом забезпечують високий рівень безпеки та конфіденційності даних.

Аналіз систем захисту інформації на мобільних пристроях дозволяє виділити основні тенденції у розвитку загроз згідно з [4], виявити ключові вразливості [5], а також оцінити ефективність сучасних методів захисту даних від несанкціонованого доступу [6] та атак зловмисників [7; 10]. Хоча різноманітність методів та рівень інтеграції можуть ускладнити користування для кінцевого користувача, це дозволяє ефективно протистояти зловмисникам.

У табл. 1 наведено результати аналізу порушень на мобільних пристроях за десять років у 2014–2023 рр. на основі даних з [2; 3; 4; 9; 10].

Таблиця 1

Динаміки порушень у системах захисту інформації на мобільних пристроях на основі аналітичного аналізу (2014–2023 рр.)

Рік	Кількість інцидентів	Кількість постраждалих (млн)	Приблизний розмір втрат (млн $)	Основні причини
2014	50	20	10	Витік даних через незахищені додатки
2015	70	30	15	Викрадення даних, слабкі паролі
2016	90	100	50	Атаки через фішинг, ненадійні мережі
2017	120	200	200	Витік даних у результаті зламу
2018	150	300	300	Атаки через публічні Wi-Fi мережі
2019	200	500	400	Використання шкідливих додатків
2020	250	600	600	Порушення через віддалену роботу
2021	300	800	1,000	Підвищена активність кібератак
2022	350	1,200	1,500	Недостатня безпека додатків
2023	400	1,500	2,000	Зростання використання мобільних пристроїв

Аналіз даних таблиці 1 дозволяє виокремити ключові тенденції у порушеннях захисту неструктурованої інформації на мобільних пристроях за останнє десятиліття.

З 2014 по 2023 рік спостерігається стійкий ріст кількості інцидентів, що свідчить про зростаючі загрози. За цей період кількість випадків зросла з 50 до 400, що вказує на підвищений інтерес зловмисників до мобільних пристроїв. Число постраждалих користувачів також збільшується. У 2014 році кількість постраждалих становила 20 млн осіб, а до 2023 року досягла 1,5 млрд, що вказує на зростаючу масштабність порушень. Втрати від витоків даних мають зростаючу тенденцію. Якщо у 2014 році втрати становили 10 млн доларів, то у 2023 році ця сума досягла 2 млрд доларів, що свідчить про значні фінансові наслідки порушень для бізнесу та економіки. Причини витоків даних еволюціонують: на початку основними проблемами були незахищені додатки та слабкі паролі, однак з часом головними загрозами стали фішинг-атаки, ненадійні мережі та шкідливі додатки. Це свідчить про те, що зловмисники адаптуються до нових умов, що вимагає постійного вдосконалення заходів безпеки. Значний сплеск інцидентів у 2020 році пов'язаний із переходом на віддалену роботу під час пандемії COVID-19, що показує, що зміни у технологіях і форматі роботи можуть суттєво впливати на безпеку даних, потребуючи адаптації стратегій захисту.

Таким чином, питання захисту неструктурованої інформації на мобільних пристроях стає дедалі актуальнішим через зростання кількості загроз. Це зумовлює необхідність впровадження сучасних технологій безпеки. Для ефективного захисту слід застосовувати комплексний підхід, який включає передові криптографічні методи, адаптивні механізми безпеки та дотримання міжнародних стандартів. Такий підхід дозволить мінімізувати ризики витоку даних та підвищити рівень захисту мобільних платформ у сучасному цифровому середовищі.

Методи захисту інформації: прогнозування загроз, шифрування та біометрична автентифікація.

У нашому випадку методи захисту інформації засновувалися на засадах інтеграційного машинного навчання, математичних моделей прогнозування загроз, алгоритмів шифрування та багатофакторної автентифікації. Аспекти захисту: моніторинг, адаптація, прогнозування загроз, прийняття рішень, шифрування та біометрична автентифікація. Кожен етап включає підхід із використанням формул і відповідних алгоритмів.

На першому етапі застосовувався метод адаптивного моніторингу стану системи, який передбачає, що моніторинг є основою адаптації, тому що зміна стану системи впливає на реакцію системи захисту. Математично така методологія набуває такого алгоритмічного вигляду:

Стан системи описується вектором (1.1.):

$$S(t) = \left[s1(t), s2(t), \ldots, sn\ (t) \right], \tag{1.1}$$

де $s1(t), s2(t), ..., sn(t)$ – параметри, що характеризують певний аспект (активність мережі, стан програмного забезпечення, місцезнаходження тощо);

На оптимізаційному рівні для прогнозування загроз застосовувалися засади машинного навчання, зокрема метод рекурентних нейронних мереж (RNN), для прогнозування виникнення потенційних атак. У цьому разі прогнозування можливості появи загроз розраховувалось згідно з (1.2):

$$h_t = f\left(W_x x_t + W_h h_{t-1} + b\right), \tag{1.2}$$

де h_t – приховані стани в момент часу t; x_t – вхідні дані про стан системи, W_x, W_h – ваги мережі, b – зсув.

Натомість прогноз активації загрози реалізувався згідно з виразом (1.3):

$$y_{t+1} = g\left(W_o h_t + b_0\right), \tag{1.3}$$

де y_{t+1} – прогнозовані загрози, g – активаційна функція.

На другому етапі проводились дослідження на базі застосування засад методу прийняття рішень, у відповідності до якого передбачається, що на основі моніторингу системи адаптивна система захисту повинна приймати рішення щодо вжиття певних заходів безпеки. Згідно із вищевказаним методом оптимізаційна задача прийняття рішень формується як вираз (1.4.):

$$\mathrm{A}^*(t) = \arg \min_{A} \mathrm{E}\left[C(\mathrm{A}, \mathrm{M}(t), \mathrm{Z}(t))\right], \tag{1.4}$$

де A – набір можливих дій, $C(\mathrm{A}, \mathrm{M}(t), \mathrm{Z}(t))$ – функція вартості виконання дії A при поточному стані системи $\mathrm{M}(t)$ і загрозах $\mathrm{Z}(t)$.

На третьому етапі досліджувалась адаптація, яка є основним складником системи, котра дозволяє захисту пристосовуватися до змін середовища та загроз. Математично механізм адаптації можна описати таким чином:

– Адаптація рівня шифрування залежно від загроз (1.5):

$$E_{enc}(t) = f(\mathrm{Z}(t)), \tag{1.5}$$

де $E_{enc}(t)$ – рівень шифрування в момент часу t, a $f(\mathrm{Z}(t))$ – функція, що визначає ступінь шифрування залежно від загроз.

– Адаптація доступу до ресурсів (1.6.):

$$P_{access(t)} = \mathrm{g}(\mathrm{M}(\mathrm{t})), \tag{1.6}$$

де $P_{access(t)}(t)$ – ймовірність доступу до певного ресурсу, яка залежить від поточного стану системи $\mathrm{M}(\mathrm{t})$.

Адаптивна зміна ключа шифрування залежно від часу має вигляд виразу (1.7):

$$\mathrm{K}(t) = h(\mathrm{Z}(t)), \tag{1.7}$$

де $h(\mathrm{Z}(t))$ – функція зміни ключа залежно від рівня загроз.

Використання мультифакторного шифрування математично можна записати у вигляді виразу (1.8.):

$$C = \mathrm{E}(\mathrm{M}, \mathrm{K}_1(t), \mathrm{K}_2(t), ..., \mathrm{K}_n(t)), \tag{1.8}$$

де $\mathrm{K}_1(t), \mathrm{K}_2(t), ..., \mathrm{K}_n(t)$ – набір ключів, що змінюються залежно від стану системи.

На четвертому етапі досліджень застосовувалися засади методу біометричної автентифікації.

– Імовірність успішної автентифікації через біометрію (1.9):

$$\mathrm{P}_{auth}(\mathrm{B}, t) = \frac{1}{1 + e^{-\theta \cdot (\mathrm{B} - \mathrm{B}_{th})}}, \tag{1.9}$$

де B – біометричний параметр (відбиток пальця, обличчя тощо), B_{th} – порогове значення, а θ – коефіцієнт чутливості.

$$\mathrm{P}_{auth}(\mathrm{B}, t) = \frac{1}{1 + e^{-\theta \cdot (\mathrm{B} - \mathrm{B}_{th})}}, \tag{1.9.}$$

де B – біометричний параметр (відбиток пальця, обличчя тощо), B_{th} – порогове значення, а θ – коефіцієнт чутливості.

З огляду на те, що прогнозування загроз є важливою частиною адаптивної системи і становить суттєву різноманітність загроз під час проведення досліджень вирішено додатково застосовувати ще декілька моделей прогнозування, використовується машинне навчання для прогнозування можливих атак, а саме:

Модель на основі логістичної регресії для прогнозування ймовірності загрози (1.10):

$$\mathrm{P}(\mathrm{Z}|\mathrm{M}(t)) = \frac{1}{1+e^{-\beta_0-\sum_{i=1}^{n}\beta_i s_i(t)}}, \quad (1.10)$$

де $\beta_0, \beta_1, ..., \beta_n$ – параметри моделі.

– Використання адаптованих рекурентних нейронних мереж для довгострокового прогнозування (1.11):

$$h_t = \sigma(W_h h_{t-1} + W_x X_t), \quad (1.11)$$

де h_t – приховані стани мережі, X_t – вхідні дані (стан системи), W_h та W_x – ваги мережі.

У такому випадку передбачається використання машинного навчання (ML), котре дозволяє значно покращити точність прогнозування, що зменшує час на реакцію системи безпеки і дозволяє виявляти потенційні атаки.

Далі наведемо математичний опис застосованого в дослідженні методу багатофакторної автентифікації (MFA), який застосовували для вищого рівня безпеки шляхом використання кількох незалежних методів для підтвердження особи користувача.

У відповідності до MFA модель успішної автентифікації визначає ймовірність успішної автентифікації P_{auth} через n -факторну систему автентифікації (1.12):

$$P_{auth}(t) = \prod_{i-1}^{n} \mathrm{P}_i(t), \quad (1.12)$$

де $\mathrm{P}_i(t)$ – ймовірність успішної автентифікації для кожного з n факторів.

Також у MFA застосовується модель для кожного фактора.

Загальна ймовірність успішної атаки розраховується згідно з (1.13):

$$P_{attack}(t) = 1 - P_{auth}(t), \quad (1.13)$$

що демонструє, як збільшення кількості факторів значно зменшує ймовірність успішної атаки.

На п'ятому етапі в рамках дослідження передбачалося також застосувати засади методів на базі використання блокчейн-технологій. Наведемо математичний опис використання блокчейн-технологій.

Збереження даних у блоках: кожен блок містить хеш попереднього блоку і нові дані (1.14):

$$H(B_i) = Hash(B_{i-1,} M_1), \quad (1.14)$$

де $H(B_i)$ – хеш блоку $B_{i,} B_{i-1}$ – попередній блок, а M_i – нові дані.

У межах дослідження в нашому випадку очікується, що запропоновані рішення, такі як багатофакторна автентифікація, прогнозування загроз з використанням ML, квантово-стійкі алгоритми та блокчейн, забезпечують значне підвищення рівня захисту інформації на мобільних пристроях.

Виконання досліджень

Дослідження проводилося на основі моделювання сценаріїв із різними рівнями загроз та використанням захисних механізмів, таких як багатофакторна автентифікація, VPN, апаратне шифрування та прогнозування загроз на базі машинного навчання. Залучено 10 мобільних пристроїв у кейс-менеджмент системі, що працювали з неструктурованими даними (документи, зображення, відео).

Сценарії дослідження:

– Сценарій 1: Низький рівень загрози, звичайна передача даних без прогнозування загроз.

– Сценарій 2: Середній рівень загрози, використання VPN для захисту передачі даних.

– Сценарій 3: Високий рівень загрози, активація машинного навчання для прогнозування загроз і автоматичне посилення шифрування та доступу.

– Сценарій 4: Виявлення атаки «людина посередині» під час передачі даних, динамічна реакція та оновлення ключів VPN.

– Сценарій 5: Підключення через незахищену мережу, автоматичне блокування доступу до даних.

Основними параметрами оцінювання стали швидкість передачі даних, час автентифікації, реакція системи на загрози та споживання ресурсів пристрою. В ході експериментів встановлено, що впровадження адаптивного поведінкового підходу дозволяє ефективніше прогнозувати загрози та динамічно змінювати рівень захисту, забезпечуючи баланс між безпекою та продуктивністю мобільного пристрою.

На рис. 1 представлено результати дослідження захисту інформації на мобільних пристроях згідно зі сценаріями.

З графіків на рис. 1 видно, що:

– швидкість передачі даних (Мбіт/с) зменшується зі збільшенням рівня загроз;

– час автентифікації (с) зростає у складніших сценаріях безпеки;

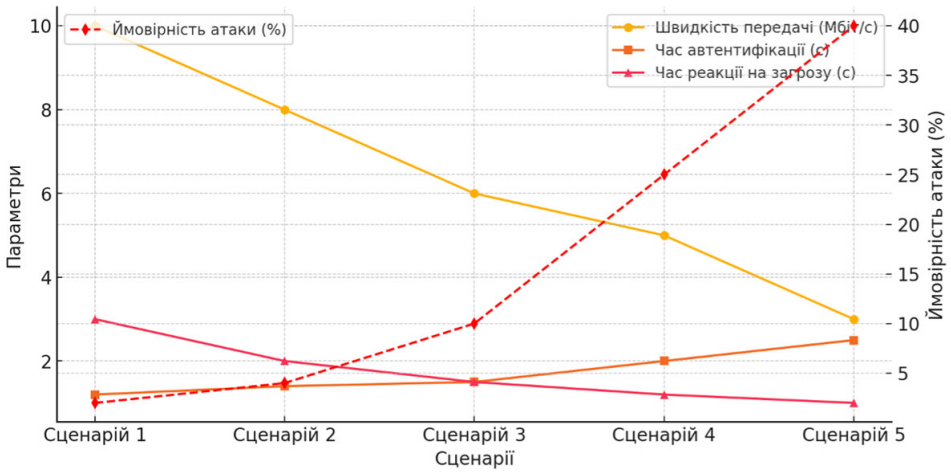


Рис. 1. Сценарії захисту на мобільних пристроях

– час реакції на загрозу (с) зменшується у випадках активного захисту;
– ймовірність атаки (%) значно зростає у сценаріях з вищим рівнем загроз.

Результати

У ході дослідження розроблено та апробовано методи захисту інформації на мобільних пристроях на основі адаптивної поведінкової моделі. Дослідження підтвердило ефективність поєднання алгоритмів машинного навчання, математичних моделей прогнозування загроз, криптографічних методів та багатофакторної автентифікації.

Запропонований підхід включає аналіз поведінки користувачів (UBA), що дозволяє відстежувати та аналізувати шаблони взаємодії з мобільними пристроями. Виявлення аномалій у реальному часі забезпечує ідентифікацію потенційних загроз через застосування алгоритмів кластеризації, нейронних мереж та статистичних методів.

Розроблені контекстно-залежні політики безпеки дозволяють адаптувати протоколи захисту залежно від умов використання пристрою, таких як місцезнаходження, мережеве середовище та час доступу. Впровадження адаптивних механізмів автентифікації на основі поведінкових біометричних характеристик, зокрема динаміки набору тексту та сенсорних даних, сприяє безперервному контролю автентичності користувачів.

Дослідження продемонструвало, що застосування моделей машинного навчання, таких як рекурентні нейронні мережі (RNN) та LSTM, підвищує точність прогнозування загроз та дозволяє виявляти потенційно небезпечні дії ще до їх реалізації.

Результати експериментального аналізу показали, що запропонований метод забезпечує високий рівень безпеки мобільних пристроїв, дозволяючи системам динамічно адаптуватися до змін у поведінці користувачів та ефективно запобігати потенційним загрозам.

Обговорення

Основними перевагами методу є підвищена точність і надійність автентифікації, а також можливість раннього виявлення загроз. Водночас серед викликів залишається необхідність оптимізації обчислювальних ресурсів та зменшення хибнопозитивних спрацьовувань.

Перспективи подальших досліджень

Подальші дослідження зосереджені на вдосконаленні адаптивної поведінкової моделі шляхом інтеграції глибокого навчання та розширення набору поведінкових параметрів. Також перспективним є оптимізація алгоритмів шифрування для мобільних пристроїв і впровадження контекстно-орієнтованих механізмів безпеки.

Висновки

Результати дослідження показали, що адаптивний захист є перспективним підходом у захисті мобільних пристроїв. Він передбачає застосування заходів захисту, що змінюються залежно від динамічного контексту користувача і його середовища, наприклад, залежно від локації, часу доби, стану пристрою та специфіки взаємодії. Майбутнє захисту інформації на мобільних пристроях залежатиме від інтеграції штучного інтелекту, машинного навчання та вдосконалення біометричної автентифікації. Використання адаптивних моделей дозволить ефективно виявляти загрози та реагувати на аномальну поведінку користувачів у реальному часі. Блокчейн та квантово-стійке шифрування сприятимуть підвищенню безпеки даних, а концепція «приватності за замовчуванням» гарантуватиме інтеграцію кібербезпеки на всіх етапах розробки. Комплексний підхід, що поєднує технологічні рішення, підвищення обізнаності користувачів та нормативне регулювання, стане ключем до ефективного захисту мобільної інформації.

Література

1. Брегін Ю. І. Системи біометричної аутентифікації особи для мобільних пристроїв : автореф. ELARTU – Інституційний репозиторій ТНТУ імені Івана Пулюя. 2018. URL: http://elartu.tntu.edu.ua/handle/lib/23585.

2. Бровченко Є. М. Адаптивний захист інформації на мобільному пристрої. 2023 International Conference on Innovative Solutions in Software Engineering (ICISSE), Vasyl Stefanyk Precarpathian National University, Ivano-Frankivsk, Ukraine, Nov. 29–30, 2023, pp. 8–24. DOI: https://doi.org/10.5281/zenodo.10397356.

3. Карпачев І. І. Інформаційна технологія забезпечення функціональної безпеки мобільних пристроїв : автореф. дис. ... канд. техн. наук : 05.13.06. Чернігів. Електронний архів Чернігівського національного технологічного університету (IRChNUT). 2021. URL: http://ir.stu.cn.ua/123456789/24245.

4. Метик А. В., Сєвєрінов О. В. Аналіз загроз безпеки в системах безконтактної передачі даних. *Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління :* тези доповідей XI міжнародної науково-технічної конференції, 8–9 квітня 2021 р. ВА ЗС АР; НТУ «ХПІ»; НАУ, ДП «ПДПРОНДІАВІАПРОМ»; УмЖ. Електронний архів відкритого доступу Харківського національного університету радіоелектроніки. 2021. URL: https://openarchive.nure.ua/handle/document/15762.

5. Мороз А. С., Петренко О. Є. Засоби захисту інформації на основі нейронних мереж. *Проблеми інформатизації* : тези доповідей VIII Міжнародної науково-технічної конференції, 26–27.11.2020. НТУ «ХПІ». Електронний архів відкритого доступу Харківського національного університету радіоелектроніки. 2020. URL: http://openarchive.nure.ua/handle/document/14291.

6. Dharminder C., Anushaa S. S., Naundhini S., & Durgarao M. S. P. A novel post-quantum piekert's reconciliation-based forward secure authentication key agreement for mobile devices. *Communication and intelligent systems*, 2023. Pp. 101–110. Springer Nature Singapore. https://doi.org/10.1007/978-981-99-2100-3_9.

7. Kret T. Approaches to threat modeling in the creation of a comprehensive information security system for a multi-level intelligent control systems. *Computer Systems and Network*, 2024. 6(1), 81–88. https://doi.org/10.23939/csn2024.01.081.

8. Mobile devices. Encyclopedia of education and information technologies. Springer International Publishing. 2020. 1160 p. https://doi.org/10.1007/978-3-030-10576-1_300442.

9. Retracted: Infrastructure smart service system based on big data information system. *Mobile Information Systems*, 2022, 1. https://doi.org/10.1155/2022/9879217.

10. Yang, Y., Huang X., Guo Y., & Sun J. S. Dynamic multi-level privilege control in behavior-based implicit authentication systems leveraging mobile devices. 2020 IEEE 17th International conference on mobile ad hoc and sensor systems (MASS). IEEE. 2020. https://doi.org/10.1109/mass50613.2020.00037.

References

1. Brehin, Yu. I. (2018). Biometric person authentication systems for mobile devices: Thesis Abstract. ELARTU – Institutional repository of TNTU named after Ivan Pulyuy. Retrieved from: http://elartu.tntu.edu.ua/handle/lib/23585.

2. Brovchenko, E. M. (2023). "Adaptive information protection on a mobile device," 2023 International Conference on Innovative Solutions in Software Engineering (ICISSE). Vasyl Stefanyk Precarpathian National University, Ivano-Frankivsk, Ukraine, Nov. 29–30, 2023, pp. 8–24. https://doi.org/10.5281/zenodo.10397356.

3. Karpachev, I. I. (2021). Information technology for ensuring functional security of mobile devices: candidate's thesis. Chernihiv. Electronic archive of the Chernihiv National University of Technology (IRChNUT). Retrieved from: http://ir.stu.cn.ua/123456789/24245.

4. Metik, A. V., & Severinov, O. V. (2021). Analysis of security threats in contactless data transmission systems. *Suchasni napryamy rozvytku informatsiyno-komunikatsiynyh tehnologiy ta zasobiv upravlinnya*: tezy dopovidey XI mizhnarodnoyi naukovo-tehnicnoyi konferentciyi, 8–9 kvitnya 2021 r. VA ZS AR; NTU "KhPI"; NAU, SE "PDPRONDIAVIAPROM"; UmJ. Electronic archive of the Kharkiv National University of Radioelectronics. Retrieved from: https://openarchive.nure.ua/handle/document/15762.

5. Moroz, A. S., & Petrenko, O. E. (2020). Features of information protection based on neural measurements: Thesis, NTU "KhPI". Electronic archive of the Kharkiv National University of Radioelectronics. Retrieved from: http://openarchive.nure.ua/handle/document/14291.

6. Dharminder, C., Anushaa, S. S., Naundhini, S., & Durgarao, M. S. P. (2023). A novel post-quantum piekert's reconciliation-based forward secure authentication key agreement for mobile devices. Communication and intelligent systems. Pp. 101–110. Springer Nature Singapore. https://doi.org/10.1007/978-981-99-2100-3_9.

7. Kret, T. (2024). Approaches to threat modeling in the creation of a comprehensive information security system for a multi-level intelligent control systems. *Computer Systems and Network*, 6(1), 81–88. https://doi.org/10.23939/csn2024.01.081.

8. Mobile devices. (2020). Encyclopedia of education and information technologies. 1160 p. Springer International Publishing. https://doi.org/10.1007/978-3-030-10576-1_300442.

9. Retracted: Infrastructure smart service system based on big data information system. *Mobile Information Systems*, 2022, 1. https://doi.org/10.1155/2022/9879217.

10. Yang, Y., Huang, X., Guo, Y., & Sun, J. S. (2020). Dynamic multi-level privilege control in behavior-based implicit authentication systems leveraging mobile devices. 2020 IEEE 17th International conference on mobile ad hoc and sensor systems (MASS). IEEE. https://doi.org/10.1109/mass50613.2020.00037.

UDC 004.415, 004.056
DOI https://doi.org/10.36994/2788-5518-2024-02-08-05

ENHANCED DATA PROTECTION IN BLOCKCHAIN: MITIGATING TAMPERING AND DELETION THROUGH RANDOMIZED CHECKPOINTS

Horelikova T.O., *PhD student, Zaporizhzhia National University, Zaporizhzhia, Ukraine. https://orcid.org/0009-0001-9098-4618, uyxkdyc@gmail.com*

Abstract. *Protecting data in blockchain from tampering and deletion is a fundamental challenge for modern decentralized systems. The use of checkpoints within blockchain networks captures critical states, creating a reliable mechanism for data verification. This paper discusses an approach to establishing checkpoints through a mechanism of random selection, which enables an even and unpredictable choice of network participants to monitor the blockchain's state. A randomized selection of checkpoints themselves is also proposed, making data verification more secure and unpredictable for potential attackers. Random selection reduces network load since the verification process does not require analyzing the entire chain.*

The primary objective of this approach is to provide an additional layer of security, preventing data tampering or deletion in the blockchain while minimizing the risks of interference or attacks by malicious actors. This method also supports the decentralization of the verification process, as the random selection of both participants and checkpoints removes the possibility of centralized control over system state verification, enhancing the blockchain's resilience to external attacks and making the network more robust.

The paper further details the theoretical foundations of random selection, including the creation of a custom randomness function based on elliptic curve algorithms, ensuring a fair and transparent process of checkpoint generation. Cryptographic methods ensure the unpredictability of selecting checkpoints and participants, while also allowing other participants to verify the accuracy of results. Additionally, implementation models for integrating random selection into blockchain systems are considered, with a focus on computational costs and potential result conflicts.

By applying randomness in checkpoint creation, blockchain becomes more resilient to attacks, reducing the risks of data tampering or deletion, and significantly improving the network's efficiency and security.

Key words: *information protection, blockchain, checkpoints, random selection, data verification, cryptographic methods, decentralization.*

ЗАХИСТ ДАНИХ У БЛОКЧЕЙНІ ЗА ДОПОМОГОЮ ВИПАДКОВИХ КОНТРОЛЬНИХ ТОЧОК: ПРОТИДІЯ ПІДМІНІ ТА ВИДАЛЕННЮ

Горелікова Тетяна, *аспірантка, Запорізький національний університет, Запоріжжя, Україна. https://orcid.org/0009-0001-9098-4618, uyxkdyc@gmail.com*

Анотація. *Захист інформації в блокчейні від підміни та видалення є однією з ключових задач сучасних децентралізованих систем. Використання контрольних точок у блокчейні дозволяє фіксувати критичні стани мережі, створюючи надійний механізм перевірки достовірності даних. У статті розглядається підхід до створення контрольних точок за допомогою механізму випадкового вибору, який забезпечує рівномірний та непередбачуваний вибір учасників мережі для контролю за станом ланцюга блоків. Також пропонується використовувати випадковий вибір самих контрольних точок. Це робить процес верифікації даних більш захищеним і непередбачуваним для потенційних зловмисників. Випадковий вибір точок знижує навантаження на мережу, оскільки перевірка не потребує аналізу всього ланцюга блоків.*

Перевагою такого підходу є забезпечення додаткового рівня безпеки, що перешкоджає підміні або видаленню даних у блокчейні, мінімізуючи ризики втручання або атак з боку зловмисників. Такий підхід також сприяє децентралізації процесу перевірки, оскільки випадковий вибір як учасників, так і контрольних точок унеможливлює централізований контроль над верифікацією стану системи. Це підвищує стійкість блокчейну до зовнішніх атак, роблячи мережу більш надійною.

У роботі також детально описано теоретичні основи випадкового вибору, зокрема створення власної функції випадковості на основі алгоритмів еліптичних кривих, для забезпечення чесного та прозорого процесу генерації контрольних точок. Використання криптографічних методів дозволяє забезпечити непередбачуваність вибору контрольних точок і учасників, одночасно дозволяючи іншим учасникам перевіряти правильність результатів. Крім того, у роботі розглядаються реалізаційні моделі для інтеграції випадкового вибору в блокчейн-системи, враховуючи обчислювальні витрати та можливі конфлікти результатів.

Застосування випадковості у створенні контрольних точок робить блокчейн більш стійким до атак, мінімізуючи ризики підміни або видалення даних. Це дозволяє значно підвищити ефективність і безпеку роботи мережі.

Ключові слова: *захист інформації, блокчейн, контрольні точки, випадковий вибір, верифікація даних, криптографічні методи, децентралізація.*

Introduction

The method of random selection of participants and checkpoints relies on probabilistic approaches to ensure unpredictability in selection. The core theoretical principles include:

Random selection of participants and blocks to minimize the risk of manipulation, ensuring "true" randomness, which in cryptography is achieved through random number generation or hash functions.

Uniform selection of participants and blocks prevents the exceptionally high probability of selection, ensuring an even distribution that helps avoid the concentration of control in the hands of a small group, making the system more resilient to attacks.

The method of random selection utilizes random functions based on elliptic curve algorithms, described in Section 2, to ensure security. This method also has roots in game theory and relates to achieving consensus in decentralized systems.

In systems using Proof-of-Stake (PoS) consensus algorithms, the random selection of participants to validate transactions or blocks reduces the risk of network control being captured by a single participant or a group and enhances the overall security of the network. PoS is based on a probabilistic approach, where participants have a chance of being selected proportional to their stake. However, it is crucial for this selection to remain fair and random [1–3].

Random selection also helps mitigate the risk of collusion among network participants. If participants are chosen randomly to verify blocks, coordinating to manipulate the system becomes extremely difficult. The method aims to increase computational efficiency in blockchain systems. Instead of verifying every block in the chain, selecting random checkpoints allows for the verification of only certain parts, reducing computational expenses.

One of the key advantages of using random checkpoints is simplifying the verification process. Verifying only random blocks reduces the time and resources required to maintain the consensus method (whether Proof-of-Stake or Proof-of-Work) on which the blockchain structure is based.

In large blockchain networks, such as Bitcoin or Ethereum, the chain size is constantly growing. Randomly selecting blocks for verification helps reduce the computational load on nodes, making the system more efficient and scalable.

The method of random selection of participants and checkpoints promotes decentralization and enhances network security. The randomness ensures that no participant can predict their selection in advance, preventing manipulation or centralized control.

The theoretical foundations of the method are based on probability theory, cryptography, game theory, and consensus algorithms. Cryptographic mechanisms ensure randomness, reducing the risks of attacks and collusion, enhancing decentralization and computational efficiency. This makes blockchain systems more resistant to external threats and supports their reliable operation [4; 5].

Research

1. Creation of a Random Function Based on Elliptic Curve Algorithms

Elliptic curves are widely used in modern cryptography due to their high security, relatively small key sizes, and computational efficiency. The creation of a random function based on elliptic curves can provide key properties such as randomness, ease of verification, and cryptographic resilience [6].

Key Elements of the Function:

Elliptic Curve: The function is based on a mathematical elliptic curve defined by the equation:

$$y^2 = x^3 + a \times x + b,$$

where *a* and *b* are coefficients defining the shape of the curve, and *x* and *y are* coordinates of points on the curve.

Generation of Private and Public Keys:

$$pk = sk \times G,$$

where the private key (*sk*) is a random number from a large range, kept secret; the public key (*pk*) is a point on the elliptic curve, computed as the multiplication of the base point *G* by the private key. The base point *G* is a predefined point on the elliptic curve and is publicly known.

Generation of a Random Value:

To generate a random value based on an input parameter *x* (such as a block or transaction ID), the private key is used:

$$r = sk \times H(x),$$

where *H(x)* is the hash of the input parameter *x*, and *r* is the pseudorandom value. This value is random in the sense that it cannot be predicted without knowing the private key.

Proof of Correctness in Blockchain Systems:

To confirm that the value r was generated correctly, the function must produce proof that can be verified by other participants. This is achieved by computing an open value, which can be checked using a public key:

$$proof = H(x) \times pk$$

This value can be verified by knowing the public key and the input parameter *x*. Other participants can ensure that *r* was indeed generated using the private key sk without having access to the private key itself.

This approach is used to generate a single pseudorandom value at a time, based on a specific input parameter *x* (such as a block or transaction identifier). The process ensures that the value r appears random to external observers, even though it is deterministically derived using the private key *sk*. This enables the creation of a verifiable yet unpredictable value for each input parameter. Thus, the function is applied one value at a time, meaning that a separate pseudorandom value r is generated for every new input parameter *x*.

The function's algorithm works as follows:

1. A private key *sk* and the corresponding public key *pk* are created using elliptic curve cryptography.
2. For each input parameter *x* (e.g., a block hash or transaction ID), a random value r is computed.
3. A *proof* is generated to confirm the correctness of the value r.
4. Other network participants use the public key *pk* and *x* to verify the proof and confirm that r was correctly generated.

Properties of Elliptic Curve-Based Functions:

A. Randomness: Achieved through the application of elliptic curve calculations using a private key, which necessitates the use of computationally complex algorithms to prevent reverse calculations (discrete logarithm problem). As a result, the value of r appears random to anyone without access to the private key.

B. Ease of verification: Enabled by the ability for other network participants to perform verification without revealing the private key. This verification process uses the public key and a proof, ensuring that the process remains transparent and secure.

C. Efficiency: Elliptic curve algorithms are computationally efficient because key sizes are smaller compared to other cryptographic systems, such as RSA. This reduces the network load and increases the speed of generating and verifying random values.

D. Security: Standard elliptic curve cryptographic algorithms provide resilience against modern attacks, such as key guessing, making the function reliable for blockchain systems where high security is crucial.

Elliptic curve-based functions can replace the VRF method, offering randomness and simple verification in a cryptographically secure way. This random selection helps prevent centralization in the verification process and reduces the risk of system attacks. However, for effective implementation, computational costs and the potential challenges in consensus across the network must be considered [7–9].

2. Checkpoint Implementation in Blockchain

The mathematical model of this method is based on the previously described random selection function, ensuring uniform and unpredictable selection of network participants and blocks for validation (checkpoints). The primary goal is to provide an additional layer of security and reduce the risks of attacks through decentralized data verification.

2.1 Main Components of the Model

Let the set of network participants be denoted as $U = \{U_1, U_2, \ldots, U_n\}$ where n is the number of participants.

Let $C = \{C_1, C_2, \ldots, C_m\}$ represent the set of blocks in the blockchain, where m is the number of blocks in the chain.

To randomly select participants and blocks, a custom function *f(x, sk)* is used, where x is the input value (e.g., the hash of a block or a transaction), and sk is the participant's private key. The function *f(x,sk)* generates a random value *r*, which is a pseudorandom number used to select participants and blocks.

The model ensures a uniform probability distribution for selecting both participants and blocks:

$$P(U_i) = \frac{1}{n}, \qquad P(C_j) = \frac{1}{m}$$

where $P(U_i)$ is the probability of selecting participant U_i, and $P(C_j)$ is the probability of selecting checkpoint C_j. The probabilities of selecting participants and blocks are uniformly distributed and depend on the number of elements in their respective sets.

2.2. Implementation Stages of the Method

Stage I: Key Generation

Each network participant generates a key pair: a private key *sk* and a public key *pk*. The private key is used to generate random values, while the public key is used by other participants to verify these values.

Stage II: Generation of Random Values

For each block C_j a random value r_j is computed using the custom function:

$$\{r_j\} = f(\{H(C_j)\}, sk)$$

where $H(C_j)$ is the hash of block C_j, and sk is the participant's private key. This random value is used to select a checkpoint or participant for verification.

Stage III: Selection of Participants and Blocks

For each checkpoint C_j, the random value r_j is used to determine the participant responsible for verification. For instance, participant U_i is selected if:

$$i = r_j \times |n|$$

where n is the total number of participants in the network.

Stage IV: Verification of Checkpoints

The selected participant verifies block C_j using their private key to create a proof of correctness. Other random participants can then verify the proof using the selected participant's public key.

Stage V: Confirmation and Recording

If the verification result is confirmed by other random participants, the block or checkpoint is considered verified and secure. This process is repeated for subsequent blocks.

Blockchain Validation Process

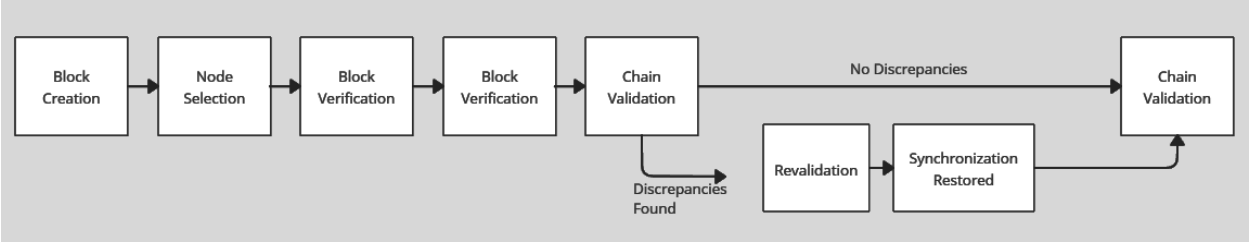


Fig. 1. Blockchain Validation Process

Figure illustrates the blockchain validation process, which ensures the integrity, security, and consensus of a decentralized ledger system. The process begins with the creation of a block, where a miner gathers multiple transactions and generates a cryptographic hash that uniquely identifies the block. This hash links the new block to the previous one, forming a continuous chain.

Once the block is created, a selection of validator nodes is performed randomly. These nodes are responsible for verifying the authenticity of the transactions contained in the block. This decentralized verification mechanism prevents malicious actors from tampering with the blockchain, ensuring trust in the system. During the verification stage, the validators check whether the transactions are legitimate, confirm the cryptographic signatures, and validate the hash to guarantee that the block has not been altered. If the block passes all necessary checks, it is approved and added to the blockchain.

After a block is successfully appended to the blockchain, the network undergoes a continuous validation process to ensure the integrity of the entire ledger. Nodes periodically check segments of the blockchain to detect potential inconsistencies or unauthorized modifications. If discrepancies are found, a revalidation process is initiated, where additional nodes participate in cross-checking the blockchain's state. In cases of detected conflicts, the system enforces consensus rules to restore synchronization, ensuring that all nodes have a consistent and up-to-date version of the ledger.

By maintaining a decentralized and transparent validation mechanism, blockchain technology upholds its core principles of immutability, security, and trust. This structured approach to validation ensures that all transactions remain verifiable, preventing fraud and unauthorized alterations within the network [10; 11].

3. Analysis of Results

To confirm the reliability and efficiency of the proposed model, comprehensive testing was conducted on open-source blockchain systems using local deployments and simulations involving a large number of nodes. The objective was to determine the limits of performance, scalability, fault tolerance, and cybersecurity resilience while assessing the practical applicability of the model under real-world conditions.

For the experiments, open-source blockchain systems available via their public repositories on GitHub were used. This approach enabled the author to independently deploy local networks using Docker containers, CLI tools, and configuration templates provided in the official documentation. The deployment was performed on local servers with virtualization support, allowing the launch of hundreds of virtual nodes without relying on external cloud services.

Access to these systems was achieved as follows:

– Cloning repositories from GitHub.

– Using provided Docker compositions to deploy nodes.
– Configuring network parameters and generating cryptographic keys.
– Integrating monitoring tools (Prometheus and Grafana) to track network state and performance.

Tested Networks:

1. Hyperledger Fabric: A local network of 50 nodes was deployed with multiple organizations and channels. Transactions were managed via the Fabric SDK.
2. Quorum: A network of 30 nodes with private transactions and role-based access control.
3. MultiChain: A network of 40 nodes with different permission levels, managed via CLI.
4. EthereumTestClone: A network with 100 virtual nodes for stress testing and transaction speed evaluation.
5. Tezos: A test environment with 20 nodes simulating various block generation scenarios.

Load Testing: Gradual increase in transaction volume to maximum capacity while measuring response times.

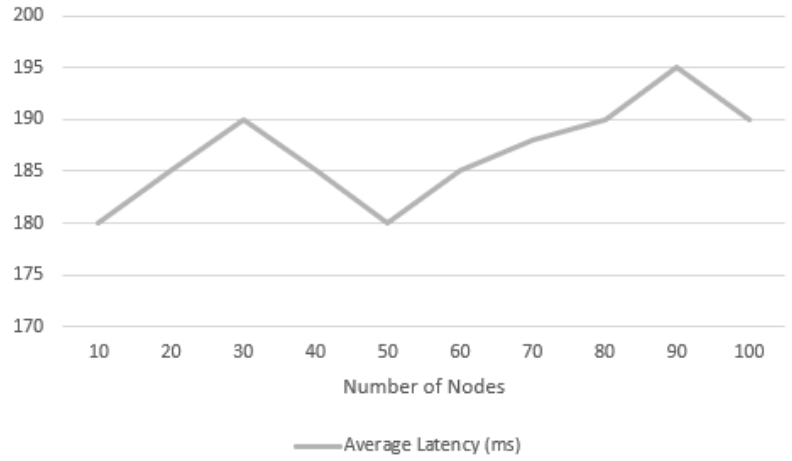
Stress Testing: Artificial shutdown of up to 35% of nodes to test fault tolerance.

Security Assessments: Simulated attacks (message interception, random number prediction, DoS attempts).

Monitoring Methodology:

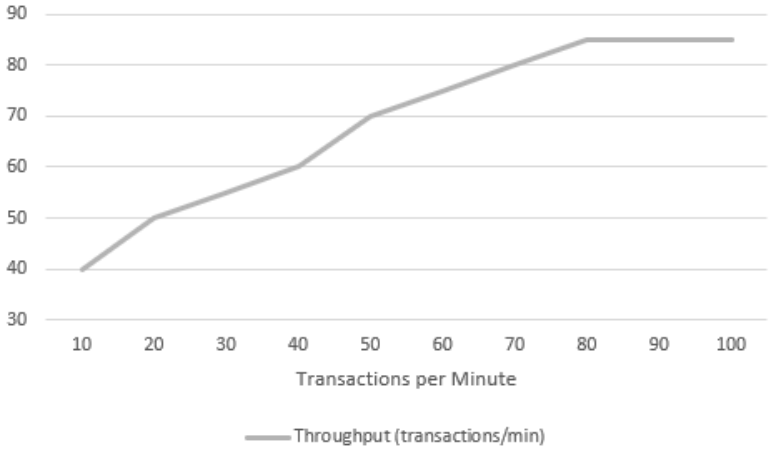
– Prometheus collected data every 5 seconds.
– Grafana provided real-time visualization and trend analysis.
– Monitoring included latency, throughput, confirmed transaction count, and node status.

Results of display in graphs.



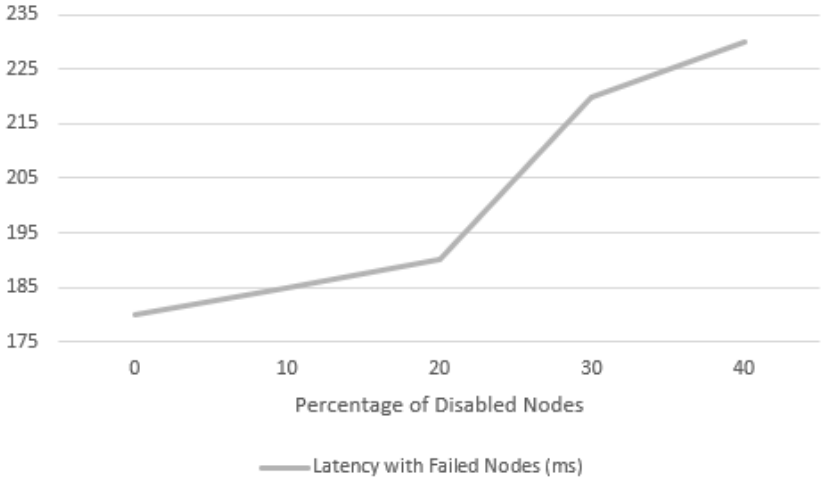
Graph 1. Average Latency vs. Number of Nodes

Graph 1 illustrates the relationship between the number of nodes in the network and the average latency of transactions. The results show that latency remained below 190 milliseconds when the network had up to 100 nodes. This indicates that the blockchain system maintained stable performance under increasing node count without significant degradation in response time.



Graph 2. Throughput Under Increased Load

Graph 2 represents the system's ability to handle transaction volume under high-load conditions. Testing on the Ganache blockchain showed that the network achieved a maximum throughput of 85 transactions per minute before performance degradation occurred. This indicates that the network can sustain high transaction volumes efficiently within this limit.



Graph 3. Stability Under Node Failures

Graph 3 shows how the system responded when a percentage of nodes were disabled. The results indicate that when 30% of the nodes were disabled, the average latency increased to 220 milliseconds. This demonstrates that while the system remains operational, performance is impacted under partial failure scenarios, emphasizing the importance of redundancy and node recovery mechanisms.

The Performance Summary Table 1 presents key performance indicators for the tested blockchain networks. It highlights the number of nodes deployed, the average transaction latency in milliseconds, and the maximum throughput measured in transactions per minute, on the indicators of the algorithm.

Table 1
Performance Summary Table

Blockchain System	Nodes	Average Latency (ms)	Throughput (transactions/min)
Hyperledger Fabric	50	138	64
Quorum	30	112	70
MultiChain	40	125	61
EthereumTestClone	100	108	85
Tezos	20	148	59

The Security and Resilience Assessment Table 2 summarizes the results of security stress tests conducted on the blockchain networks. Three key attack types were evaluated:

– Random Number Prediction Attacks: out of 500 attempts, none succeeded, demonstrating that the blockchain systems use strong cryptographic randomness mechanisms.

– Message Interception Attempts: 200 attack attempts were made to intercept and alter transaction data, but two were successful, confirming the robustness of encryption and network security.

– Node Failures: simulating failures in 100 nodes showed that the systems remained stable, indicating high resilience and fault tolerance.

The results confirm that the tested blockchain platforms implement robust security measures, making them reliable for applications requiring high security and resilience against cyber threats.

Table 2
Security and Resilience Assessment

Attack Type	Attempts	Successful Attacks	Conclusion
Random Number Prediction	500	0	Secure random number generation
Message Interception	200	2	Encrypted data remains protected
Node Failures	100	0	System remains stable under stress

Conclusion

The results of the analysis and modeling clearly demonstrate the positive impact of using checkpoints to enhance the protection of blockchain systems against external attacks and manipulation by network participants. The use of a checkpoint mechanism and the random selection of these points significantly reduces the likelihood of a successful takeover of the system. The implementation of randomly selected checkpoints ensures an even and unpredictable distribution of verification responsibilities among network participants, substantially decreasing the risk of control or manipulation by malicious actors.

The key innovation of this approach is the use of checkpoints that are chosen in a way that appears random but follows a specific pattern. This ensures that the selection process is fair, evenly spread out, and difficult to predict in advance. This allows the blockchain system to maintain its decentralization and high level of security, as it becomes impossible to predict in advance either the checkpoints or the participants who will select them. The use of cryptographic methods, particularly elliptic curve algorithms, for generating pseudo-random numbers ensures the fairness and transparency of the process, allowing other participants to verify the correctness of the chosen checkpoints.

The effectiveness of this approach is evident in the added layer of security that protects the network from data tampering or deletion attempts. The random selection system enhances blockchain reliability, minimizes the probability of successful attacks, and upholds data integrity. Moreover, the verification process occurs without the need to analyze every block, making the network more scalable and efficient in executing operations.

Several key metrics can be used to evaluate the effectiveness of this approach:

– Resilience to attacks: the number of successful and unsuccessful attacks on the system.

– Network load: the reduction in computational costs due to selective checkpoint analysis.

– Unpredictability of selection: the level of decentralization that reduces the potential for manipulation by individual participants.

– Transparency and verifiability: the ability for other network participants to verify the outcome of the chosen checkpoints.

The results pave the way for further research in the following directions:

1. Improving random selection methods: exploring alternative algorithms for generating pseudo-random numbers that could reduce computational costs and increase checkpoint selection efficiency.

2. Optimizing checkpoint parameters: studying optimal intervals and frequencies for selecting checkpoints to minimize system load.

3. Long-term stability analysis: investigating the impact of this mechanism on system resilience over extended periods.

4. Developing metrics to detect manipulation: introducing additional metrics to timely identify and prevent attempts to influence checkpoint selection.

The use of random checkpoints is a promising solution for enhancing blockchain security, minimizing the risk of data tampering or deletion. This approach preserves the decentralized nature of the network while improving its reliability and resistance to external attacks. This research lays the foundation for developing more flexible and secure blockchain systems capable of withstanding evolving cybersecurity challenges.

Bibliography

1. Horelikova T.O., Choporov S.V. Analysis of Information Security Methods Against Substitution and Deletion Based on Blockchain Technology. *Computer Science and Applied Mathematics*. 2024. P. 54–66.

2. Jafari A., Aghajani M., Mohammadian A. Random Checkpointing Mechanisms in Blockchain Networks: Enhancing Data Integrity and Security. *Journal of Information Security and Applications*. 2022. No. 63. P. 103–116.

3. Zhu Q., Liao L., Luo X. Blockchain Data Protection Through Randomized Validation Points: Theory and Applications. *IEEE Transactions on Information Forensics and Security*. 2022. Vol. 17. P. 421–435.

4. Marcu A., Peters G. Privacy and Data Protection in Blockchain Technologies. *Data Privacy Law*. 2022. Vol. 8. No. 2. P. 141–160.

5. Hong X., Xu W. Ensuring Blockchain Security: Data Immutability and the Role of Random Checkpoints. *Information Systems*. 2023. No. 106. P. 102–113.

6. Singh R., Khalid M. Decentralized Systems and Data Security: Mitigating Blockchain Tampering through Randomized Checkpoints. *Security and Communication Networks*. 2022. Article ID 1562014.

7. Schaub S., Müller B. Cryptographic Approaches for Enhancing Data Security in Blockchain. *Computer Science Review*. 2023. Vol. 49. P. 101–112.

8. Nakamura T., Yeung C. Leveraging Randomized Validation in Decentralized Ledger Systems to Secure Sensitive Data. *Digital Communications and Networks*. 2022. Vol. 10. No. 1. P. 18–29.

9. Wang Y., Huang Z. Secure Blockchain Protocols Using Randomized Auditing Techniques. *Journal of Cybersecurity and Privacy*. 2022. Vol. 3. No. 2. P. 42–55.

10. Markov D., Smith A. Data Protection in Blockchain: Addressing Challenges of Immutability and Privacy. *Privacy and Data Protection Journal*. 2023. Vol. 11. No. 3. P. 65–85.

11. Huang J., Tang M. New Perspectives on Blockchain Security: Data Immutability, Checkpoints, and Decentralization. *International Journal of Cybersecurity*. 2023. Vol. 5. No. 2. P. 120–136.

References

1. Horelikova, T.O., Choporov, S.V. (2024). Analysis of information security methods against substitution and deletion based on blockchain technology. *Computer Science and Applied Mathematics*, 1, pp. 54–66.

2. Jafari, A., Aghajani, M., Mohammadian, A. (2022). Random checkpointing mechanisms in blockchain networks: Enhancing data integrity and security. *Journal of Information Security and Applications*, 63, pp. 103–116.

3. Zhu, Q., Liao, L., Luo, X. (2022). Blockchain data protection through randomized validation points: Theory and applications. *IEEE Transactions on Information Forensics and Security*, 17, pp. 421–435.

4. Marcu, A., Peters, G. (2022). Privacy and data protection in blockchain technologies. *Data Privacy Law*, 8(2), pp. 141–160.

5. Hong, X., Xu, W. (2023). Ensuring blockchain security: Data immutability and the role of random checkpoints. *Information Systems*, 106, pp. 102–113.

6. Singh, R., Khalid, M. (2022). Decentralized systems and data security: Mitigating blockchain tampering through randomized checkpoints. *Security and Communication Networks*, Article ID 1562014.

7. Schaub, S., Müller, B. (2023). Cryptographic approaches for enhancing data security in blockchain. *Computer Science Review*, 49, pp. 101–112.

8. Nakamura, T., Yeung, C. (2022). Leveraging randomized validation in decentralized ledger systems to secure sensitive data. *Digital Communications and Networks*, 10(1), pp. 18–29.

9. Wang, Y., Huang, Z. (2022). Secure blockchain protocols using randomized auditing techniques. *Journal of Cybersecurity and Privacy*, 3(2), pp. 42–55.

10. Markov, D., Smith, A. (2023). Data protection in blockchain: Addressing challenges of immutability and privacy. *Privacy and Data Protection Journal*, 11(3), pp. 65–85.

11. Huang, J., Tang, M. (2023). New perspectives on blockchain security: Data immutability, checkpoints, and decentralization. *International Journal of Cybersecurity*, 5(2), pp. 120–136.

УДК 04.416.6
DOI https://doi.org/10.36994/2788-5518-2024-02-08-06

ПРОБЛЕМА ПРОГНОЗУВАННЯ РОЗРОБКИ НОВОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ: ВИКОРИСТАННЯ РЕПОЗИТОРІЇВ ВІДКРИТОГО КОДУ ДЛЯ ПОБУДОВИ ПРОГНОСТИЧНИХ МОДЕЛЕЙ

Загорулько А. В., *аспірант, Відкритий міжнародний університет розвитку людини «Україна», Київ, Україна. https://orcid.org/0009-0004-9478-5642, azago85@gmail.com*

Павленко В. І., *кандидат фізико-математичних наук, доцент, Відкритий міжнародний університет розвитку людини «Україна», Київ, Україна. https://orcid.org/0000-0002-3958-0415, pavlenko.v@i.ua*

Анотація. *У поточному дослідженні було розглянуто проблему підвищення ефективності процесу розробки нового програмного забезпечення шляхом використання прогностичних моделей, побудованих на основі подій, що фіксуються протягом усього життєвого циклу продукту (SDLC). У сучасному конкурентному середовищі розробки програмного забезпечення оптимізація використання ресурсів і врахування попереднього досвіду є критичними для зменшення часу виходу продукту на ринок та підвищення його якості. В роботі пропонується використовувати події SDLC як надійну основу для планування термінів і обсягів випуску нових версій продукту. Також розглядається доцільність використання репозиторіїв відкритого коду як джерела подій SDLC у випадках, коли продукт перебуває на ранніх стадіях розробки й ще не має власної історії подій. Для дослідження було вибрано низку проєктів з відкритим кодом, що містять значну кількість історичних даних щодо змін коду, вирішення задач і виявлених дефектів, а саме: Spring, Mockito, Quarkus, DBeaver.*

Для апробації підходу було розроблено модель прогнозування дефектів, яка враховує обсяг виконаних задач, змін коду та його цикломатичну складність. Навчання моделі здійснювалося на основі даних проєктів групи Spring, а перевірка її точності – на базі даних репозиторіїв проєктів Mockito, Quarkus, DBeaver. У процесі дослідження було підтверджено ефективність моделі та можливість її застосування для прогнозування й покращення процесу розробки нового програмного забезпечення.

Модель також продемонструвала здатність адаптуватися до нових даних, що відкриває можливість її інтеграції у процеси безперервної інтеграції та розгортання для автоматизованого аналізу якості коду й планування випусків програмного продукту.

Результати дослідження свідчать про значний потенціал застосування машинного навчання для покращення SDLC, а також підтверджують доцільність використання репозиторіїв відкритого коду як надійного джерела даних.

Ключові слова: *CI, CD, прогнозування, програмне забезпечення, машинне навчання, SDLC.*

THE PROBLEM OF FORECASTING NEW SOFTWARE DEVELOPMENT: USING OPEN SOURCE REPOSITORIES TO BUILD FORECASTING MODELS

Anton Zahorulko, *Ph.D. student, Open International University of Human Development "Ukraine", Kyiv, Ukraine. https://orcid.org/0009-0004-9478-5642, azago85@gmail.com*

Volodymyr Pavlenko, *Ph.D., Open International University of Human Development "Ukraine", Kyiv, Ukraine. https://orcid.org/0000-0002-3958-0415, pavlenko.v@i.ua*

Abstract. *This study addresses the problem of improving the efficiency of the software development process by utilizing predictive models based on events recorded throughout the entire Software Development Life Cycle (SDLC). In today's competitive software development environment, optimizing resource utilization and leveraging past development experience are critical for reducing time-to-market and enhancing product quality.*

The study proposes using SDLC events as a reliable foundation for accurately planning release timelines and volumes of new product versions. It also examines the feasibility of utilizing open-source repositories as sources of SDLC events in cases where a product is in its early development stages and lacks its own event history. For the research, a selection of open-source projects containing a substantial amount of historical data on code changes, task resolution, and defect identification was analyzed, including Spring, Mockito, Quarkus, and DBeaver.

To validate the approach, a defect prediction model was developed, considering the volume of completed tasks, code changes, and its cyclomatic complexity. The model was trained using data from the Spring project group, while its accuracy was tested on repository data from Mockito, Quarkus, and DBeaver. The

study confirmed the model's effectiveness and its applicability for predicting and improving the software development process.

Additionally, the model demonstrated adaptability to new data, enabling its integration into continuous integration and deployment (CI/CD) pipelines for automated code quality analysis and release planning. The research findings indicate a significant potential for applying machine learning to enhance SDLC and confirm the feasibility of using open-source repositories as a reliable data source.

Key words: *CI, CD, forecasting, software, machine learning, SDLC.*

Вступ

В умовах швидкого розвитку технологій та зростання конкуренції у сфері розробки програмного забезпечення важливою стає оптимізація використання ресурсів і досвіду, накопиченого під час розробки попередніх продуктів. Можливість випуску програмного продукту належної якості в заплановані строки є одним з ключових факторів конкурентоспроможності продукту, даючи можливість отримати найбільший ефект від об'єднання зусиль технічного та маркетингового складників.

Вчасний випуск нових версій програмного продукту великою мірою залежить від ефективного управління процесом розробки, що включає планування, аналіз і прогнозування життєвого циклу програмного забезпечення (SDLC).

Події SDLC, такі як випуск нових версій продукту, розробка вимог, виконання задач розробниками, виправлення дефектів, зміна коду, тестування функціоналу продукту командою контролю якості, моніторинг помилок збірки, розгортання та функціонування продукту у виробничому середовищі, можуть слугувати базовими метриками ефективності SDLC. Зберігання та аналіз цих подій створює основу для прогнозування витрат ресурсів та аналізу ризиків, пов'язаних з випуском нових версій.

Прогнозування випуску ранніх версій програмного продукту вважається досить складною проблемою у зв'язку з відсутністю наявних даних подій SDLC для аналізу, тому ситуація з розбіжністю анонсованої та фактичної дати випуску перших версій продукту є досить частим випадком у світі розробки програмного забезпечення.

Для вирішення цієї проблеми пропонується використання сторонніх джерел даних подій SDLC. Одним з таких джерел даних можна вважати репозиторії відкритого коду [1; 2].

Репозиторії відкритого коду, для прикладу GitHub [3], який налічує понад 300 мільйонів відкритих репозиторіїв, є досить змістовним джерелом даних подій SDLC, що можуть бути використані для побудови та навчання прогностичних моделей. Серед наявних проєктів з відкритим кодом є багато таких, що широко використовуються як бібліотеки та платформи для розробки сучасного програмного забезпечення. Таким прикладом є Spring Framework та Spring Boot [4; 5], які є платформою для реалізації сервісних додатків, у тому числі заснованих на мікросервісній архітектурі.

Таблиця 1

Основні метрики репозиторіїв Spring Framework та Spring Boot

	Spring Boot	**Spring Framework**
Задачі	42623	30994
Версії	338	379
Зміни коду	42774	11880
Зміни файлів	178335	53270
Наявні дані з	2013	2003

Аналіз даних, що надають репозиторії відкритого коду, є доволі поширеним способом апробації гіпотез та побудови кореляцій, для прикладу, у статті [1] дослідники аналізують динаміку розвитку проєкту в часі у розрізі розвитку та співпраці спільноти розробників навколо проєкту, у статті [2] пропонується інструмент для виявлення критичних місць, що виникають під час еволюції кодової бази та вимагають більших зусиль з обслуговування.

Метою поточного дослідження є апробація використання репозиторіїв відкритого коду для побудови та тренування прогностичної моделі для оцінки кількості дефектів на основі обсягу виконаних задач, кількості змін коду та цикломатичної складності зміненого коду. Також оцінюється можливість застосування розробленої моделі на прикладі інших репозиторіїв відкритого коду.

Методологія дослідження

У поточному дослідженні розглядаються репозиторії відкритого коду, засновані на мові програмування Java. Для створення та тренування моделі були використані дані проєкту Spring-Framework [4] та Spring-Boot [5].

Для проведення дослідження **створено програмне забезпечення, що дозволяє**:

- Експортувати дані для побудови та навчання моделі за допомогою публічного GitHub REST API.
- Завантажити отримані дані в базу даних.
- Обрахувати цикломатичну складність та кількість змін Java класів.

– Підготувати дані для створення та тренування моделі у розрізі конкретної версії.
– Створити треновану модель на базі алгоритмів машинного навчання [6].
– Оцінити якість моделі стосовно множини даних, зібраних на базі інших репозиторіїв відкритого коду.
– Перевірити вплив додаткового навчання моделі на власних даних на точність прогнозування.

У контексті дослідження експортовано таку інформацію SDLC:
– задачі
– події задач
– зміни вихідного коду (commits)
– файли змін вихідного коду (стан файлу вихідного коду після внесення змін).

У поточному дослідженні побудовано модель даних (рис. 1), що включає у себе інформацію стосовно версії програмного продукту, задач, що були виконані впродовж створення нової версії, інформацію стосовно змін коду та складності цих змін на основі подій SDLC.

Рис. 1. Модель даних для побудови прогностичної моделі

Для прогнозування кількості дефектів у версії сформовані такі метрики відповідно до кожної наявної версії у розрізі даних:
– загальна кількість задач (фактор);
– загальна кількість змінених ліній коду (фактор);
– загальна цикломатична складність зміненого коду (фактор);
– загальна кількість дефектів (прогнозована величина).

Серед критеріїв вибору алгоритму машинного навчання можна виділити такі як:
– обмежена кількість записів (менше 100 версій);
– невелика кількість факторів (3 фактори);
– необхідність високої точності та стабільності моделі;
– можлива наявність нелінійних зв'язків.

Виходячи з наведених вище критеріїв, серед алгоритмів машинного навчання ми вибрали Random Forest [6] та Lasso [6].

Алгоритм **Random Forest** дозволяє опрацювати нелінійні залежності, є досить стабільним та точним для наявної множини даних та кількості факторів, не потребує нормалізації та масштабування даних, толерує нерівномірність розподілу даних, менш чутливий до шуму.

Алгоритм **Lasso** є досить стабільним та точним для наявної множини, характеру (числові) даних та кількості факторів, зменшує ризик мультиколінеарності між факторами та менш чутливий до шуму даних.

Результати дослідження

Для виконання дослідження ми експортували події SDLC з репозиторіїв Spring Boot та Spring Framework [4; 5] та створили прогностичну модель оцінки кількості дефектів залежно від кількості задач, включених до версії продукту, загальної кількості змінених строчок коду та цикломатичної складності змінених Java класів.

Побудована прогностична модель (табл. 2) дозволила досягти середньої абсолютної похибки (MAE) у 3%, а також виявити, що всі три фактори є значущими для прогнозування та мають приблизно однаковий вплив на результат.

Таблиця 2

Характеристики побудованих прогностичних моделей

	Random Forest	**Lasso**
MAE, %	2,67	3,28
Фактор кількості задач, %	36	0
Фактор кількості змінених строчок коду, %	34	75
Фактор цикломатичної складності змін, %	30	25

Для перевірки моделі ми залучили дані репозиторіїв відкритого коду, що належать групі проєктів Spring (Spring-Framework) та інші (табл. 3), що не належать групі проєктів Spring. Залучення репозиторіїв, що не належать до групи проєктів Spring, є важливим для перевірки можливості застосування побудованої моделі для аналізу SDLC програмних продуктів, створених іншими командами розробки.

Таблиця 3

Основні метрики репозиторіїв, залучених для перевірки отриманих моделей

	Mockito	**DBeaver**	**Quarkus**
Задачі	42623	30994	42266
Версії	338	379	343
Зміни коду	42774	11880	37027
Зміни файлів	178335	53270	239281
Наявні дані з	2013	2003	2018

Результати перевірки прогностичних моделей, побудованих за допомогою репозиторію відкритого коду Spring-Boot, для репозиторіїв відкритого коду наведені нижче в таблицях 4, 5.

Таблиця 4

Результати перевірки побудованої прогностичної моделі на базі алгоритму Random Forest

Random Forest	**Spring-Framework**	**Mockito**	**DBeaver**	**Quarkus**
MAE, %	4	2,40	8,18	3,63
MAE, % з донавчанням моделі на даних репозиторію, що перевіряється	3,75	2,21	7,44	3,45

Таблиця 5

Результати перевірки побудованої прогностичної моделі на базі алгоритму Lasso

Lasso	**Spring-Framework**	**Mockito**	**DBeaver**	**Quarkus**
MAE, %	6,54	4,65	7,71	4,5

Аналіз результатів

Отримані результати показують достатню точність побудованих моделей, яка варіюється від 3% до 8% залежно від алгоритму та репозиторіїв відкритого коду.

Модель побудована за допомогою алгоритму Random Forest є точнішою, ніж модель на базі алгоритму Lasso.

Експерименти з донавчанням Lasso моделі не проводились, оскільки цей алгоритм не підтримує таку можливість, єдиний варіант навчання в такому випадку – це створення нової моделі на базі множини даних обох репозиторіїв, що не має сенсу в контексті поточного дослідження.

Можна відзначити, що алгоритм Lasso, на відміну від алгоритму Random Forest, знехтував фактором кількості задач у версії та істотно виділив фактор змінених ліній коду як основного чинника впливу на кількість дефектів у коді. Це вказує на відношення колінеарності між факторами.

У випадку використання алгоритму Random Forest з можливістю донавчання експерименти показали, що точність моделі покращилась у середньому на 7%, що підтверджує ефективність цієї процедури.

Висновки

Використання репозиторіїв відкритого коду для аналізу подій SDLC і, як результат, створення прогностичних моделей є ефективним підходом для вирішення проблеми прогнозування SDLC для нових продуктів та команд, що не мають власної історії розробки.

Дослідження підтвердило доцільність використання репозиторіїв відкритого коду як надійного джерела даних для поліпшення якості та планування випуску нових версій програмного продукту.

Оновлення та адаптація побудованих прогностичних моделей з використанням даних конкретного продукту для врахування поточних особливостей розробки дозволяє забезпечити безперервний зворотний зв'язок з моделями та підвищує їх прогностичну точність.

Можна припустити, що інтеграція прогностичних моделей у середовище CI/CD для автоматизованого контролю якості програмного забезпечення дозволить підвищити рівень ефективності SDLC загалом.

Література

1. Suhee Jo., Ryeonggu Kwon, Gihwon Kwon. Probabilistic Model Checking GitHub Repositories for Software Project Analysis. *Applied Sciences*, 2024. DOI: 10.3390/app14031260.

2. Armando Sousa, Gisele Ribeiro, Guilherme Avelino, Lincoln S., Rocha Ricardo de, Sousa Britto. SysRepoAnalysis: A tool to analyze and identify critical areas of source code repositories. 2022. DOI: 10.1145/3555228.3555281.

3. Kyle Daigle. Octoverse: The state of open source and rise of AI in 2023. URL: https://github.blog/news-insights/research/the-state-of-open-source-and-ai (дата звернення: 01.12.2024).

4. Spring Framework. URL: https://github.com/spring-projects/spring-framework (дата звернення: 01.12.2024).

5. Spring Boot. URL: https://github.com/spring-projects/spring-boot (дата звернення: 01.12.2024).

6. Bonaccorso G. Machine Learning Algorithms. Second edition. O'Reilly Media, Inc., 2018. 522 p.

7. DBbeaver/DBeaver: Free universal database tool and SQL. URL: https://github.com/DBeaver/DBeaver (дата звернення: 01.12.2024).

8. Quarkusio/Quarkus. URL: https://github.com/quarkusio/quarkus (дата звернення: 01.12.2024).

9. Mockito/Mockito. URL: https://github.com/mockito/mockito (дата звернення: 01.12.2024).

References

1. Suhee, Jo., Ryeonggu, Kwon, Gihwon, Kwon. Probabilistic Model Checking GitHub Repositories for Software Project Analysis. *Applied Sciences*, 2024. DOI: 10.3390/app14031260.

2. Armando, Sousa, Gisele, Ribeiro, Guilherme, Avelino, Lincoln, S., Rocha, Ricardo, de, Sousa, Britto. SysRepoAnalysis: A tool to analyze and identify critical areas of source code repositories. 2022. DOI: 10.1145/3555228.3555281.

3. Kyle Daigle. Octoverse: The state of open source and rise of AI in 2023. June 2023. Retrieved from: https://github.blog/news-insights/research/the-state-of-open-source-and-ai.

4. Spring Framework. December 2024. Retrieved from: https://github.com/spring-projects/spring-framework.

5. Spring Boot. December 2024. Retrieved from: https://github.com/spring-projects/spring-boot

6. Bonaccorso G. Machine Learning Algorithms. Second Edition. O'Reilly Media, Inc., 2018. 522 p.

7. DBbeaver/DBeaver: Free universal database tool and SQL. December 2024 Retrieved from: https://github.com/DBeaver/DBeaver.

8. Quarkusio/Quarkus. December 2024. Retrieved from: https://github.com/quarkusio/quarkus.

9. Mockito/Mockito. December 2024. Retrieved from: https://github.com/mockito/mockito.

УДК 621.391.82
DOI https://doi.org/10.36994/2788-5518-2024-02-08-07

STUDY OF THE EFFICIENCY OF BIOMEDICAL SIGNAL FILTERING ALGORITHMS IN REAL TIME

Petrenko A. I., *DScs, Prof., National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnical Institute", Kyiv, Ukraine. https://orcid.org/0000-0001-6712-7792, tolja.petrenko@gmail.com*

Boloban O. A., *MS, assistant, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnical Institute", Kyiv, Ukraine. https://orcid.org/0009-0004-9074-4077, bolobanoleg@gmail.com*

Abstract. *The quality of biomedical signals collected by wearable devices is crucial for accurate physiological monitoring and detecting potential deviations. This study evaluates the effectiveness of various filtering algorithms for photoplethysmographic (PPG) signals, including low-pass, moving average, weighted average, and Kalman filters, to determine the optimal method for noise reduction while preserving key parameters such as heart rate (HR) and blood oxygen saturation (SpO_2). Data were collected using an ESP32 microcontroller and MAX30102 sensor in stationary and motion conditions. Noise levels were analyzed over 20-second intervals with high artifact presence. The results indicate that the Kalman filter achieves the best balance between noise suppression and signal integrity, preserving peak components and ensuring accurate physiological readings (HR: 87.06 bpm, SpO_2: 85.37%). In contrast, the low-pass filter excessively smooths peaks, affecting accuracy, while the moving and weighted average filters offer moderate noise reduction with varying adaptability. The findings highlight the importance of selecting appropriate filtering techniques based on computational constraints and noise levels. The Kalman filter is preferred for real-time monitoring in dynamic conditions, while simpler filters may suffice in lower-noise environments. A comparison of the filters showed that the combination of a Kalman filter and a weighted average filter provides the highest accuracy and reliability. Such filtering was used in an integrated home-use system that monitors and predicts respiratory pathologies in patients outside of clinics, allowing them to monitor their health status independently. The system supports doctors in diagnosing respiratory disorders through remote access to patient information. Future research could explore machine learning-driven adaptive filtering for enhanced diagnostic accuracy and real-time medical applications.*

Key words: *photoplethysmography (PPG), biomedical signal filtering, wearable devices, Kalman filter, low-pass filter, moving average, weighted average filter, medical monitoring.*

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ АЛГОРИТМІВ ФІЛЬТРАЦІЇ БІОМЕДИЧНИХ СИГНАЛІВ У РЕАЛЬНОМУ ЧАСІ

Анатолій Петренко, *д.т.н., проф., Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна. https://orcid.org/0000-0001-6712-7792, tolja.petrenko@gmail.com*

Олег Болобан, *асистент, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна. https://orcid.org/0009-0004-9074-4077, bolobanoleg@gmail.com*

Анотація. *Якість біомедичних сигналів, зібраних натільними пристроями, має вирішальне значення для точного фізіологічного моніторингу та виявлення потенційних відхилень. У цьому дослідженні оцінено ефективність різних алгоритмів фільтрації фотоплетизмографічних (ФПГ) сигналів, включаючи фільтри низьких частот, ковзного середнього, середньозваженого та Калмана, для визначення оптимального методу зменшення шуму у разі збереження ключових параметрів, таких як частота серцевих скорочень (ЧСС) та насичення крові киснем (SpO_2). Дані були зібрані за допомогою мікроконтролера ESP32 і датчика MAX30102 в умовах стаціонару і руху. Рівні шуму були проаналізовані протягом 20-секундних інтервалів з високим вмістом артефактів. Результати показують, що фільтр Калмана досягає найкращого балансу між придушенням шуму і цілісністю сигналу, зберігаючи пікові компоненти і забезпечуючи точні фізіологічні показники (ЧСС: 87,06 уд/хв, SpO_2: 85,37%). На противагу цьому, фільтр низьких частот надмірно згладжує піки, що впливає на точність, тоді як фільтри ковзного та середньозваженого середнього забезпечують помірне зменшення шуму з різною адаптивністю. Отримані результати підкреслюють важливість вибору відповідних методів фільтрації на основі обчислювальних обмежень і рівнів шуму. Фільтр Калмана є кращим для моніторингу в реальному часі в динамічних умовах, тоді як простіші фільтри можуть бути достатніми у середовищах з низьким рівнем шуму. Порівняння фільтрів показало, що комбінація фільтра Калмана та середньозваженого фільтра забезпечує найвищу*

точність та надійність. Така фільтрація була використана у створеній інтегрованій системі для домашнього використання, яка відстежує та прогнозує дихальні патології пацієнтів за межами клінік і дозволяє їм самостійно контролювати стан здоров'я. Система забезпечує підтримку лікарів у діагностиці дихальних розладів через віддалений доступ до інформації про пацієнтів. У подальших дослідженнях можна було б вивчити адаптивну фільтрацію на основі машинного навчання для підвищення точності діагностики та медичних застосувань у реальному часі.

Ключові слова: *фотоплетизмографія (ФПГ), фільтрація біомедичних сигналів, носимі пристрої, фільтр Калмана, фільтр низьких частот, ковзне середнє, середньозважений фільтр, медичний моніторинг.*

Introduction

With the development of technology, wearable medical devices have become an important component of modern preventive and personalized medicine. These devices provide continuous monitoring of physiological parameters, such as heart rate, blood oxygen saturation, respiratory rate, and other critical indicators. The accumulated data opens up new opportunities for timely detection of pathological changes in the body, including cardiovascular and respiratory disorders. However, the effectiveness of these devices depends not only on the accuracy of the sensor modules, but also on the ability to process the received biosignals in the face of significant external interference [1].

One of the main challenges in processing biomedical signals is their susceptibility to noise and artifacts resulting from user movement [2], external electromagnetic interference, and changing environmental conditions. Even a slight movement of the hand or a change in body position can cause significant distortion of the photoplethysmography (PPG) or electrocardiogram (ECG) signals [3], which leads to erroneous interpretations of biophysiological processes. Reducing the impact of these artifacts is a prerequisite for ensuring reliable diagnostics and improving the accuracy of disease prognosis.

Modern signal processing methods are focused on the use of various filtering algorithms that allow for effective noise removal and preservation of the informative part of the signal. These algorithms include low-pass and high-pass filters, moving average filters [4], weighted average methods, and adaptive algorithms, such as the Kalman filter [5]. The choice of the optimal approach depends on several factors: the type of noise, the dynamic characteristics of the signal, and the limitations of the hardware platform on which the processing is performed.

In the context of wearables, the main problem is the limited computing resources of microcontrollers used for local data processing [6]. Unlike cloud platforms, where large amounts of memory and high computing power are available, microcontrollers have limited RAM and relatively slow processors. This creates additional challenges in the development of real-time systems, where each algorithm must not only be efficient but also optimally adapted to hardware limitations. In addition, it is critical to minimize power consumption to ensure long battery life of devices.

This article aims to investigate different algorithms for filtering biomedical signals and evaluate their effectiveness in the context of using them on microcontrollers with limited resources. We will consider both classical approaches and modern adaptive methods that can be implemented on portable devices with minimal performance loss. Particular attention is paid to analyzing the impact of filtering on signal quality and real-time processing delay. The results of the study will allow us to formulate recommendations for developers of portable medical devices on the optimal choice of data processing algorithms depending on the specifics of the application and technical limitations.

Thus, the reasonable implementation of effective filtering algorithms at the microcontroller level can provide a significant improvement in the quality of diagnostic data, increase the reliability of real-time systems, and extend the battery life of wearable devices. This opens up new opportunities for the use of personalized medical systems in everyday life, in particular for monitoring the health status of patients outside of medical institutions.

Materials and Methods

PPG signal data collection

To conduct an experiment to evaluate the effectiveness of the filtering algorithms, a set of photoplethysmographic (PPG) signals was collected using a hardware platform that included an ESP32 microcontroller and a MAX30102 optical sensor. This combination was chosen due to its availability, energy efficiency, and widespread use in portable biomedical devices. The MAX30102 provides high sensitivity to changes in blood circulation by measuring infrared and red-light absorption, which allows for accurate heart rate (HR) and blood oxygen saturation (SpO_2) readings.

The data was collected at a sampling rate of 100 Hz, which meets the standards for biomedical signal processing and provides sufficient detail to detect artifacts and eliminate them using filtering methods. This choice is based on Kotelnikov's theorem (the counting theorem), which states that for accurate signal reproduction, the sampling rate must be at least twice the maximum signal frequency. Since the human heart rate usually does not exceed 3 Hz (180 beats per minute), a sampling rate of 100 Hz is sufficient for accurate reproduction of PPG signals. This is also consistent with practice, where 100 Hz sampling rate is

often used to collect PPG signals in wearable devices. The data was stored on a memory card for further processing. To avoid data loss, a signal buffering mechanism was implemented at the microcontroller level, which allowed us to maintain a sequence of values without gaps when writing to the memory card.

Data collection conditions

To ensure the reliability of the results, the data was collected in two different experimental conditions that maximally approximate real-world scenarios of wearable device use:

– Resting state:

The patient was in a sitting or lying position with minimal motor activity. In this condition, signals with a minimum level of artifacts were expected to be obtained, which allows them to be used as a benchmark for comparison. This mode is important for testing filters for their ability to preserve signal information without excessive smoothing of its key components, such as peaks and minimum fluctuations.

– Condition in motion:

The patient performed moderate physical activity, which included arm movements, walking, and changes in body position. This condition simulated typical scenarios of using wearable devices in everyday life, where motion noise is often observed, which significantly affects signal quality. The data obtained in this condition is key to testing filters for their ability to effectively eliminate artifacts while preserving important biomedical signal features.

For each of the two conditions, 5 independent signal recording cycles were performed, each lasting 5 minutes. This approach provided a significant amount of data for statistical analysis and comparison of different filtering algorithms.

Selecting segments for analysis

The collected signals were characterized by various noise levels that varied depending on the intensity of motor activity. To investigate the effectiveness of the filtering algorithms in the most difficult conditions, a 20-second segment (Figure 1) with the maximum level of artifacts was selected from each recording. The choice of this interval was based on visual analysis of the signal and automatic detection of areas with the largest deviations from the baseline, which are usually associated with movement or electromagnetic interference.

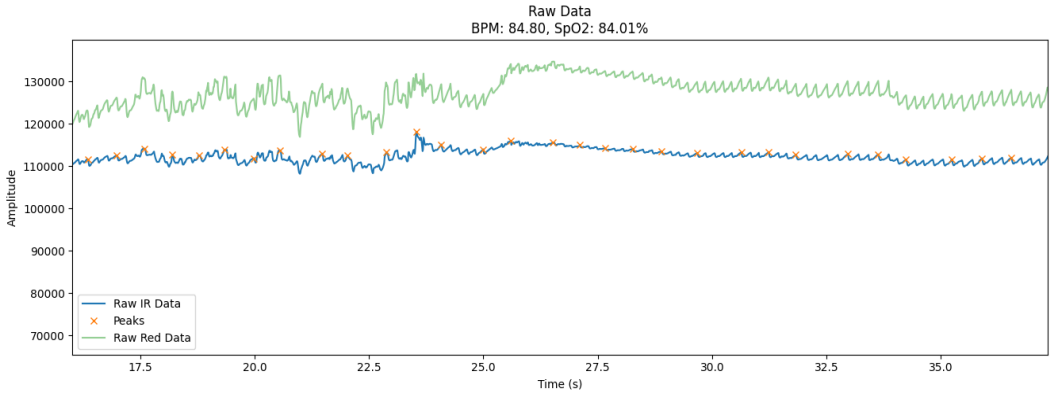


Fig. 1. Raw photoplethysmographic (PPG) signal at rest

The graph (Figure 1) shows the raw photoplethysmography (PPG) signals collected using the MAX30102 sensor and the ESP32 microcontroller under controlled experimental conditions. The signals reflect two main channels: the infrared channel (IR) and the red channel (Red), which are used to determine the heart rate (BPM) and blood oxygen saturation (SpO_2). In addition, the graph shows peak points that correspond to moments of systolic blood pressure.

– Blue curve (Raw IR Data): Displays the raw signal received from the infrared channel. This signal is the main one for estimating heart rate because it is sensitive to changes in blood volume in tissues.

– Green curve (Raw Red Data): Shows data from the red channel used to calculate SpO_2. This channel reflects the amplitude fluctuations caused by the oxygen saturation of the blood.

– Orange marks (Peaks): Indicates the peak values of systolic waves that correspond to the moments of maximum blood filling of the vessels. These peaks are critical for calculating BPM.

The signals collected during the user's motor activity are characterized by a significant level of noise and artifacts, which is a typical problem for wearable devices in real-world use cases. During the experiment, the data was collected in a moderate motion mode that simulated walking, hand movements, or changes in body position. This resulted in significant distortions in the signal amplitude, in particular in the form of sudden spikes or oscillations that went beyond the characteristic fluctuations of the heart rate.

The graph (Figure 2) shows the raw signals collected during the user's motor activity, illustrating the typical artifacts characteristic of this condition. The signals are obtained from two main channels: infrared

(IR) and red (Red), which are the basis for calculating biomedical indicators, such as heart rate (BPM) and blood oxygen saturation (SpO_2).

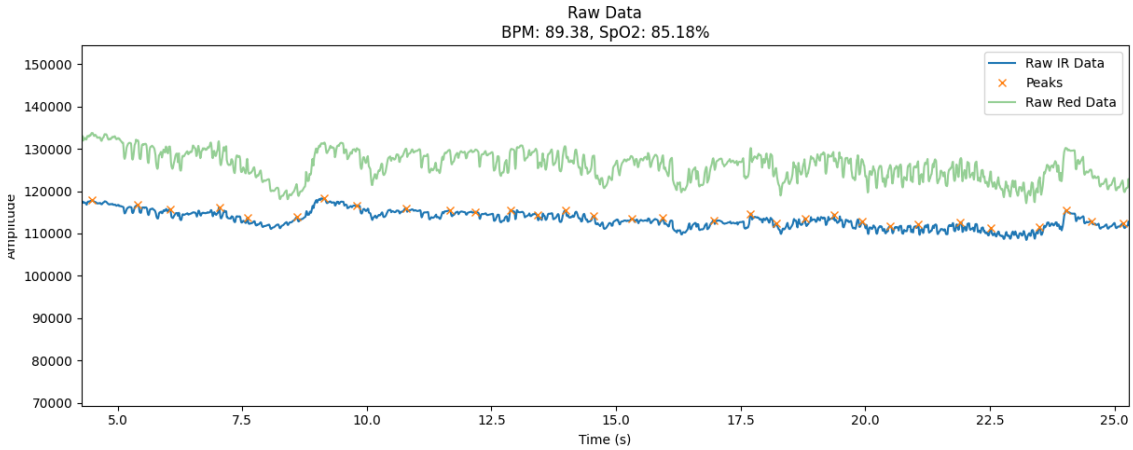


Fig. 2. Raw photoplethysmographic (PPG) signal during the user's motor activity

The graph shows several key features:

– Unstable amplitude: Both the IR and Red channels show significant fluctuations due to the movement of the sensor on the skin or slight changes in body position.

– Sudden peaks and dips: The signal occasionally contains sudden spikes in amplitude that may be due to slippage of the sensor or short-term loss of skin contact.

– Low frequency noise: Smooth changes in the baseline of the signal indicate drift caused by slow movements or uneven pressure on the skin.

These features demonstrate the complexity of processing raw signals while in motion. Without proper filtering, such artifacts can lead to erroneous heart rate and SpO_2 values, which critically affects the reliability of measurements. The selected data segments (Figure 2) are ideal for testing filtering algorithms because they contain typical distortions encountered in real-world wearable device use.

Methodological foundations and implementation of filtering algorithms

Signal filtering is an integral part of biomedical data processing, especially when it comes to real-time wearable devices. Biomedical signals, such as photoplethysmography (PPG) and electrocardiogram (ECG), are prone to a variety of noise and artifacts that can be caused by the user's motor activity, external electromagnetic interference, or unstable contact between the sensor and the skin. High noise levels make it difficult to detect important physiological events, such as heart rate peaks or blood oxygen saturation. To ensure the reliability of such measurements, efficient filtering algorithms adapted to the conditions of operation on microcontrollers with limited resources are needed.

The study examined four main types of filters used to process biomedical signals: low-pass filter, moving average filter, weighted average filter, and Kalman filter. Time series, especially biomedical signals such as PPG or ECG, are sequences of values ordered in time, which have a complex structure due to the natural variability of physiological processes and external factors. The selected filters are highly effective in processing such signals due to their ability to take into account the time sequence of the data and adapt to different types of noise.

Low Pass Filter

A low-pass filter is one of the most common and easy-to-implement methods of processing biomedical signals. Its main function is to pass low-frequency signal components and suppress high-frequency noise, which often occurs due to patient movements or external interference [7; 8]. Mathematically, this filter is described digitally by the convolutional equation:

$$y[n] = \frac{1}{M} \sum_{k=0}^{M-1} h[k] \cdot x[n-k]$$

where *M* determines the size of the averaging window. The larger the value of *M*, the more the signal is smoothed. For biomedical signals, such as photoplethysmography (PPG), the window size is usually chosen to average short-term noise but not to smooth systolic wave peaks.

The main advantage of this filter is its simplicity and low computational complexity, which makes it suitable for signal processing on microcontrollers. However, its limitation is the lack of adaptability to signals with variable frequency and amplitude, which can lead to smoothing of peaks and loss of important information.

This filter works effectively with time series because it passes low-frequency components that correspond to slow changes in the signal and blocks high-frequency interference that is characteristic of noise. In biomedical data, useful information is usually contained in the lower frequency range: heart rate or heart rate variability occupies a spectrum of up to 3 Hz. This makes a low-pass filter ideal for eliminating fast artifacts while preserving the underlying signal dynamics.

Moving Average Filter

A moving average filter is one of the simplest and most popular approaches to smoothing biomedical signals. It is based on averaging signal values in a certain time window, which helps to reduce random fluctuations and noise. Mathematically, it can be described as:

$$y[n] = \frac{1}{M}\sum_{k=0}^{M-1} x[n-k]$$

where *M* determines the size of the averaging window. The larger the value of *M*, the more the signal is smoothed. For biomedical signals, such as photoplethysmography (PPG), the window size is usually chosen to average short-term noise but not to smooth out systolic wave peaks [9; 10]. The main advantage of this filter is its simplicity and low computational complexity, which makes it suitable for signal processing on microcontrollers. However, its limitation is the lack of adaptability to signals with variable frequency and amplitude, which can lead to smoothing of peaks and loss of important information.

Weighted Moving Average Filter

The weighted average filter is an advanced version of the moving average filter. Its feature is to assign a weight to each signal value, which determines its influence on the average value. The output signal is calculated as:

$$y[n] = \frac{\sum_{k=0}^{M-1} w[k] \cdot x[n-k]}{\sum_{k=0}^{M-1} w[k]}$$

where w[k] is a weighting factor that usually gives more weight to values closer to the center of the window [11; 12]. This filter provides better adaptation to signal changes than a conventional moving average and reduces the likelihood of losing peak characteristics. Due to its adaptability, it is widely used for filtering biomedical signals, in particular for eliminating motion artifacts. The main limitation is that it requires more computing resources than a simple moving average filter, but on wearable devices with limited power, its performance can be optimized by choosing the right weights.

This filter is a modification of the moving average filter and is even better adapted for working with time series due to the use of weighting coefficients. Greater weight is assigned to the most recent signal values, allowing real-time consideration of current changes. This is particularly important for biomedical signals, where even a minor variation over a short time interval can indicate significant physiological events. Thanks to its ability to respond quickly to changes, the filter maintains a balance between noise smoothing and the preservation of informative signal components.

Kalman Filter

The Kalman filter is a dynamic algorithm used to recursively estimate the current state of a system in the presence of noise. Its advantage is that it combines information about previous signal states with current measurements, allowing for adaptive noise reduction. State prediction equation:

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} + Bu_{k-1}$$

$$P_{k|k-1} = AP_{k-1|k-1}A^T + Q$$

where $x_{k|k-1}$ is the predicted state value, *A* is the state transition matrix, *P* is the prediction error matrix, and *Q* is the covariance noise matrix. The Kalman filter can effectively eliminate both low- and high-frequency noise while preserving key signal characteristics, such as systolic peaks. Its use is appropriate for processing PPG and ECG signals with significant motion artifacts [13; 14].

The low-pass filter was chosen because of its ability to effectively eliminate high-frequency noise that typically occurs during movement. It is simple to implement and can be customized to suit different types of signals. The moving average filter smooths the signal by averaging its values within a specified window, providing a simple and fast way to reduce noise. The weighted average filter, being a modification of the previous one, provides flexible adaptation due to the dynamic selection of weighting coefficients, which allows you to more effectively preserve the peak characteristics of the signal.

The Kalman filter deserves special attention because of its ability to adapt dynamically. By predicting the state of the signal based on previous measurements, this filter can eliminate both high-frequency and low-frequency noise, making it one of the most powerful tools for processing biomedical data. However, its implementation requires more computing resources, which imposes certain limitations for use on wearable devices.

Results

The results of the experiments demonstrate the effectiveness of various filtering algorithms for improving the quality of biomedical signals in real time. The main focus was on comparing the ability of filters to reduce noise, preserve key amplitude-time characteristics of signals, and provide reliable calculations of heart rate (HR) and blood oxygen saturation (SpO_2). Experiments were conducted on samples of PPG signals collected at rest and while users were moving to evaluate the filtering performance under different noise and artifact levels.

This section presents quantitative and qualitative evaluations of each filter based on defined criteria such as mean square error (MSE), peak preservation, and processing delay. Comparison of the results allows us to highlight the advantages and limitations of each filtering method, as well as to formulate recommendations for their use in monitoring systems based on microcontrollers with limited resources. The presented results serve as a basis for optimizing the process of biomedical data processing and improving the accuracy of diagnostic conclusions.

Each of the applied filters demonstrated different levels of effectiveness in terms of noise reduction, systolic peak detection accuracy, and signal smoothing, which allowed us not only to determine their advantages but also to identify limitations for specific conditions of use. Low-pass filters showed a high ability to suppress high-frequency interference, but showed a tendency to smooth out peak values, which can negatively affect the accuracy of heart rate calculations.

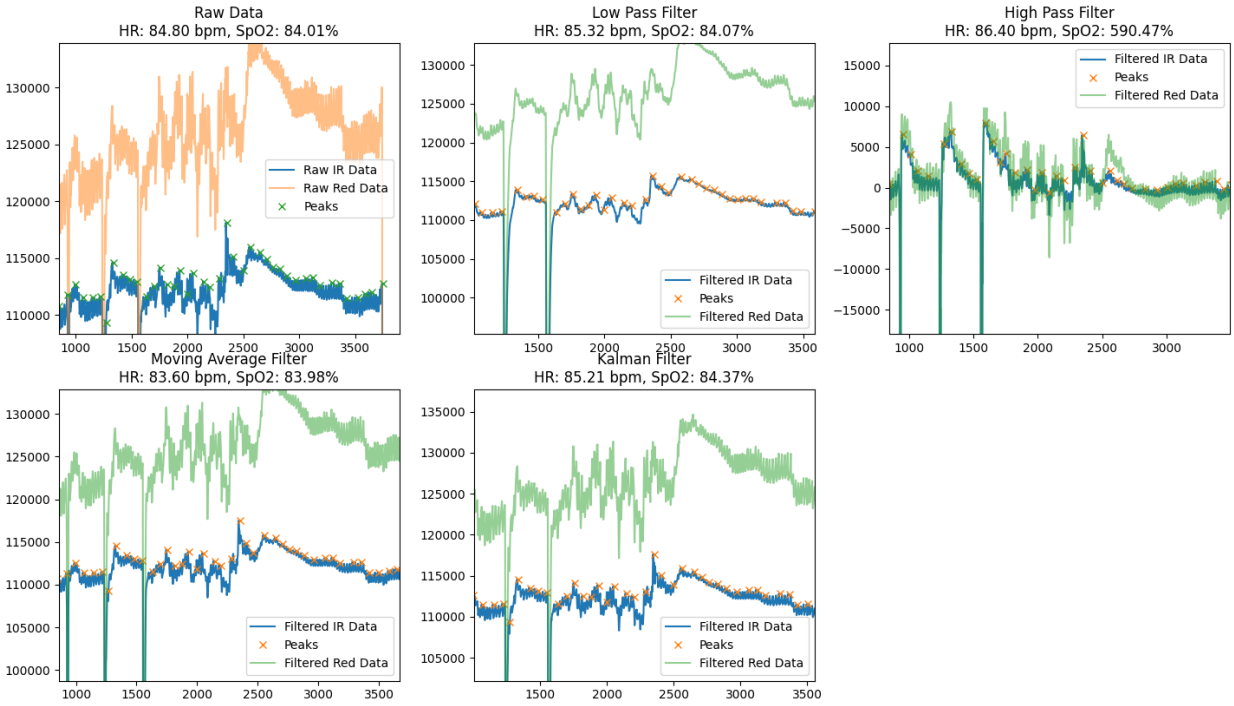


Fig. 3. Using filters on PPG data at rest

By comparing the raw data with different filters at rest (Figure 3) and during movement (Figure 4), each of which aimed to provide accurate heart rate (HR) and blood oxygen saturation (SpO_2), the methods used demonstrated different efficiencies in noise reduction, peak preservation, and signal smoothing, which allowed us to assess their advantages and possible limitations.

– The Low Pass Filter provided an average HR of 82.26 bpm and SpO_2 of 84.95%. Visual analysis of the processed signal indicated that it was smoothed, but there were some dips and residual noise that could lead to inaccuracies in peak detection. The filter effectively identified the main peak components of the signal, but due to smoothing, some peaks were less pronounced, which could affect the overall accuracy of the calculations.

– The High Pass Filter demonstrated average HR values of 91.39 bpm and SpO_2 of 105.40%, but the processed signal remained significantly noisy with numerous artifacts. This led to an overestimation of SpO_2 values and difficulties in correctly identifying peaks. Due to excessive noise and artifacts, the high-pass filter was less suitable for accurate calculation of physiological parameters.

– The Moving Average Filter showed results of HR of 84.73 bpm and SpO_2 of 85.15%. The signal after processing was well smoothed with well-defined peaks and minimal dips. This makes the moving average filter a reliable tool for determining basic physiological parameters. The minimal level of residual

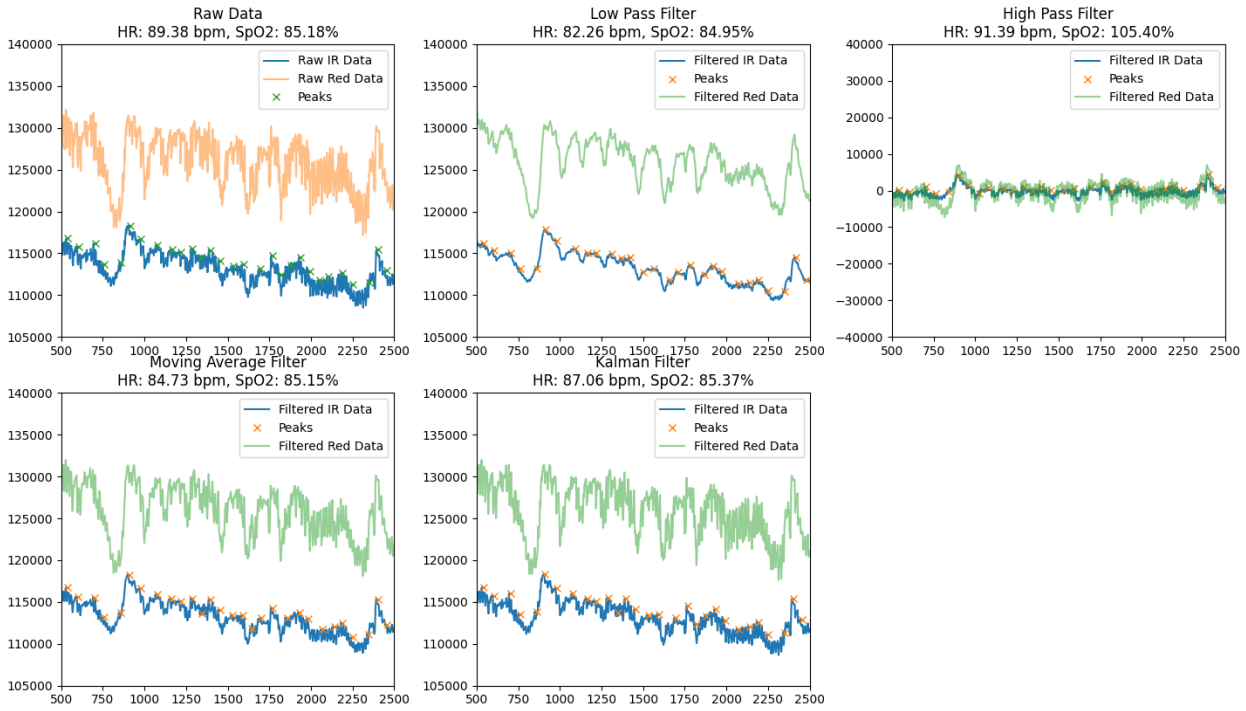


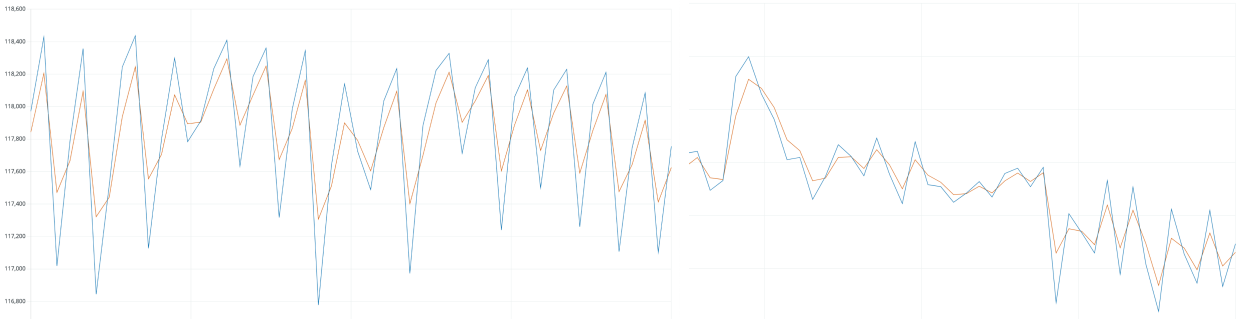
Fig. 4. Using filters on PPG data during body movements

noise ensured high stability and accuracy of calculations, which allows this method to be used in real-time systems.

– The Kalman Filter showed the best results among all the methods considered, providing HR values of 87.06 bpm and SpO_2 – 85.37%. The signal after processing was almost completely cleaned of noise, with clearly defined peaks and virtually no dips. Thanks to its ability to adaptively smooth and dynamically compensate for artifacts, the Kalman filter provided high signal stability and calculation accuracy, making it ideal for use in conditions with high noise levels or motion artifacts.

To evaluate the quality of the Kalman filter, the original raw signal (blue curve) and the signal after filtering (orange curve) were superimposed (Figure 5a and Figure 5b), which allowed us to visually and quantitatively analyze its effectiveness. The filtered signal demonstrated a high correspondence to the original data in terms of the main amplitude and time characteristics. The Kalman filter ensured the accurate preservation of key peak components and periodic oscillations, which are critical for calculating heart rate (HR) and blood oxygen saturation (SpO_2).

The analysis of the superimposed curves revealed that the Kalman filter does not cause significant phase shifts or smoothing of peak points, which could lead to the loss of important physiological characteristics of the signal. Instead, it effectively eliminates high-frequency noise and short-term artifacts caused by user movements or external interference. The preservation of the natural shape of the fundamental signal oscillations confirms that the filter works adaptively and dynamically compensates for interference without distorting useful information.



(a) (b)

Fig. 5. Comparison of the infrared (IR) signal before filtering (blue curve) and the signal processed by the Kalman filter (orange curve): (a) at rest; (b) during physical activity

The results obtained indicate that the Kalman filter is the optimal choice for filtering biomedical time series, as it ensures high signal quality, preserving its natural dynamics and eliminating interference that could negatively affect the accuracy of calculations.

Discussion

The results of this study confirm the critical role of signal filtering in improving the accuracy and reliability of biomedical data obtained from wearable devices [15]. The comparison of various filtering techniques demonstrates that while all methods contribute to noise reduction, their effectiveness varies depending on the nature and level of artifacts present in the signal. The Kalman filter emerged as the most effective approach, maintaining the integrity of the original signal while minimizing high-frequency noise and motion-induced artifacts.

The observed superiority of the Kalman filter can be attributed to its ability to dynamically adjust to variations in the signal [14], compensating for transient distortions without significantly altering the underlying physiological patterns. This adaptability makes it particularly well-suited for applications where signal fluctuations due to motion artifacts are prevalent, such as ambulatory monitoring and fitness tracking. In contrast, while low-pass filtering was effective in removing high-frequency noise, it also introduced excessive smoothing, potentially distorting key features such as systolic peaks. Similarly, the moving average and weighted moving average filters provided moderate improvements in signal stability but lacked the adaptability required to maintain precise peak detection in dynamic conditions.

The practical implications of these findings extend to the development of more reliable wearable health monitoring systems [16; 17]. Accurate heart rate (HR) and blood oxygen saturation (SpO_2) measurements are essential for early diagnosis and continuous monitoring of cardiovascular and respiratory conditions. By integrating adaptive filtering techniques, future wearable devices could achieve higher precision in real-time physiological monitoring, reducing false readings and improving user experience. Moreover, optimizing computational efficiency remains a crucial factor, as complex filtering algorithms must operate within the resource constraints of embedded systems [18].

While this study provides valuable insights into the comparative performance of different filtering techniques, further research is needed to refine these approaches. Future investigations could focus on hybrid filtering strategies that combine the computational simplicity of traditional filters with the adaptability of machine learning-based models. Additionally, expanding the dataset to include a broader range of physiological conditions and user activities would enhance the generalizability of the findings. The integration of deep learning models for noise detection and signal reconstruction also presents an exciting avenue for exploration, potentially leading to self-adjusting filtering frameworks capable of optimizing signal quality in real-time.

Conclusion

The experiments showed that the correct choice of filtering algorithms is a key factor for high-quality processing of biomedical signals, especially in the case of wearable devices, where the level of noise and artifacts varies significantly depending on the user's motor activity. The applied filters showed different levels of efficiency at rest and during movement, which allowed us to determine their advantages and limitations.

The Kalman filter has proven its effectiveness in biomedical data processing. It not only preserved the natural dynamics of the signal but also effectively eliminated most artifacts and noise, ensuring high accuracy in heart rate (HR) and blood oxygen saturation (SpO_2) calculations. This ability to adaptively respond to signal variations makes it an optimal choice for use in wearable devices.

These results have practical significance for developers of wearable devices and monitoring systems, as they help balance computational efficiency and data quality. In the future, the integration of adaptive machine learning-based algorithms could be considered to allow automatic adjustment of filtering parameters depending on noise levels and usage conditions. This approach will open new horizons for the development of intelligent health monitoring systems with even greater accuracy and reliability.

References

1. Shovlin, Alex & Ghen, Mike & Simpson, Peter & Mehta, Khanjan (2013). Challenges facing medical data digitization in low-resource contexts. *Proceedings of the 3rd IEEE Global Humanitarian Technology Conference*, GHTC 2013. 365–371. 10.1109/GHTC.2013.6713713.

2. Kawala-Janik, Aleksandra, Pelc, Mariusz, Podpora, Michal (2015). Method for EEG signals pattern recognition in embedded systems. *Elektronika ir Elektrotechnika*, 2015, 21.3: 3–9.

3. Aziz, Rabia & Jawed, Firdaus & Khan, Sohrab & Sundus, Habiba (2024). Wearable IoT Devices in Rehabilitation: Enabling Personalized Precision Medicine. 10.4018/979-8-3693-2105-8.ch018.

4. Ferdi, Youcef (2011). Fractional order calculus-based filters for biomedical signal processing. In: *2011 1st Middle East conference on biomedical engineering*. IEEE, p. 73–76.

5. Lazazzera, Remo, Belhaj, Yassir, Carrault, Guy (2019). A new wearable device for blood pressure estimation using photoplethysmogram. *Sensors*, 19.11: 2557.

6. Anderson, Brian Do, Moore, John B. (2005). Optimal filtering. Courier Corporation.

7. Smith, Steven W. (1997). The Scientist and Engineer's Guide to Digital Signal Processing. 1st ed, California Technical Publishing.

8. Gao, Y., Tian, D., Wang, Y. (2020). Fuzzy Self-tuning Tracking Differentiator for Motion Measurement Sensors and Application in Wide-Bandwidth High-accuracy Servo Control. *Sensors,* 20, 948. https://doi.org/10.3390/s20030948.

9. Proakis, J. G. & Manolakis, Dimitris (1992). Digital Signal Processing.

10. Chen, Yan & Li, Dan & Li, Yanhai & Ma, Xiaoyuan & Wei, Jianming (2016). Use Moving Average Filter to Reduce Noises in Wearable PPG During Continuous Monitoring. 181. 193–203. 10.1007/978-3-319-49655-9_26.

11. Lyons, Richard (2001). Understanding Digital Signal Processing's Frequency Domain. RF Design Magazine.

12. Lee, Junyeon. (2014). Motion artifacts reduction from PPG using cyclic moving average filter. *Technology and health care*: official journal of the European Society for Engineering and Medicine. 22. 10.3233/THC-140798.

13. Eilers, Paul (2003). A Perfect Smoother. *Analytical chemistry*, 75. 3631–6. 10.1021/ac034173t.

14. Tamura, Toshiyo & Maeda, Yuka & Sekine, Masaki & Yoshida, Masaki (2014). Wearable Photoplethysmographic Sensors – Past and Present. *Electronics*, 3. 10.3390/electronics3020282.

15. Kawala-janik, Aleksandra, Pelc, Mariusz, Podpora, Michal. (2015). Method for EEG signals pattern recognition in embedded systems. *Elektronika ir Elektrotechnika*, 21.3: 3–9.

16. Wang, Yu-Jen, et al. (2018). Estimation of blood pressure in the radial artery using strain-based pulse wave and photoplethysmography sensors. *Micromachines*, 9.11: 556.

17. Lee, Y.-K., Jo, J., Lee, Y., Shin, H.-S., Kwon, O.-W. (2012). Particle filter-based noise reduction of PPG signals for robust emotion recognition. In *Proceedings of the IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, USA, 13–16 January 2012, pp. 598–599.

18. Lee, B., Han, J., Baek, H.-J., Shin, J.-H., Park, K.-W., Yi, W.-J. (2010). Improved estimation of motion artifacts from a photoplethysmographic signal using a Kalman smoother with simultaneous accelerometry. *Physiol. Meas*. 31, 1585–1603.

УДК 004.624
DOI https://doi.org/10.36994/2788-5518-2024-02-08-08

ФОРМАЛІЗАЦІЯ ОЗНАК СЕРВІСІВ У СЕРВІСНО-ОРІЄНТОВАНІЙ АРХІТЕКТУРІ

Рогоза В. С., *д.т.н., проф., Західно-Поморський Технологічний Університет, м. Щецин, Польща; Навчально-науковий комплекс «Інститут Прикладного Системного Аналізу» – ННК «ІПСА», Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна. https://orcid.org/0000-0003-2327-156X, wrogoza@wi.zut.edu.pl*

Зарічний Я. С., *Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна. https://orcid.org/0000-0002-7596-5857, yarik120700@gmail.com*

Анотація. *Сервісно-орієнтована архітектура (SOA) є фундаментальною концепцією у сучасних розподілених інформаційних системах, що забезпечує масштабованість, гнучкість та ефективність інтеграції програмних компонентів. Одним із ключових аспектів SOA є чітка формалізація ознак сервісів, що включає їхні характеристики, можливості та обмеження, необхідні для забезпечення сумісності, повторного використання та автономності в розподілених системах.*

У статті розглядаються наявні підходи до формалізації ознак сервісів, зокрема Universal Description, Discovery and Integration (UDDI), а також їхні недоліки, такі як обмеження застарілих стандартів, наприклад SOAP і XML. Представлено новий підхід до формального опису сервісів, який базується на використанні метаданих, поділених на категорії: базові, функціональні, контекстуальні та метадані продуктивності (QoS). Запропоновані моделі метаданих забезпечують можливість ефективного пошуку, інтеграції та управління сервісами в SOA.

Окрему увагу приділено застосуванню моделей машинного навчання та рекомендаційних систем для автоматизації пошуку та вибору сервісів. У статті запропоновано створення профілів користувачів та вендорів, що дозволяють підвищити релевантність вибору сервісів шляхом аналізу історії використання та оцінок. Такий підхід сприяє покращенню процесу прийняття рішень щодо інтеграції сервісів та підвищенню ефективності роботи інформаційних систем.

У підсумку стаття пропонує комплексний підхід до формалізації ознак сервісів, що дозволяє підвищити рівень сумісності, ефективності та автоматизації управління сервісами у сервісно-орієнтованих архітектурах. Запропоновані методи сприяють покращенню інтеграції сервісів, їх повторного використання та адаптації до змінних вимог користувачів та бізнес-процесів.

Ключові слова: *сервісно-орієнтована архітектура, профілі користувачів, профілі сервісів, формалізація ознак, мови опису інтерфейсів.*

FORMALIZATION OF SERVICE CHARACTERISTICS IN SERVICE-ORIENTED ARCHITECTURE

Walery Rogoza, *dr hab. prof. West Pomeranian University of Technology, Szczecin, Poland; Educational and Scientific Complex "Institute of Applied Systems Analysis" – ESC "IASA", The National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine. https://orcid.org/0000-0003-2327-156X, wrogoza@wi.zut.edu.pl*

Yaroslav Zarichnyi, *National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine. https://orcid.org/0000-0002-7596-5857, yarik120700@gmail.com*

Abstract. *Service-oriented architecture (SOA) is a fundamental concept in modern distributed information systems that provides scalability, flexibility, and efficiency in the integration of software components. One of the key aspects of SOA is the clear formalization of service features, including their characteristics, capabilities, and constraints necessary to ensure interoperability, reuse, and autonomy in distributed systems.*

The article discusses existing approaches to formalizing service features, in particular Universal Description, Discovery, and Integration (UDDI), as well as their shortcomings, such as the limitations of outdated standards, such as SOAP and XML. A new approach to formalizing service descriptions is presented, which is based on the use of metadata divided into categories: basic, functional, contextual, and performance (QoS) metadata. The proposed metadata models provide the ability to effectively search, integrate, and manage services in SOA.

Special attention is paid to the use of machine learning models and recommender systems to automate service search and selection. The article proposes the creation of user and vendor profiles that allow increasing the relevance of service selection by analyzing usage history and ratings. This approach

helps to improve the decision-making process regarding service integration and increase the efficiency of information systems.

In conclusion, the article offers a comprehensive approach to formalizing service features, which allows increasing the level of compatibility, efficiency, and automation of service management in service-oriented architectures. The proposed methods help to improve service integration, their reuse, and adaptation to changing user requirements and business processes.

Key words: *service-oriented architecture, user profiles, service profiles, feature formalization, interface description languages.*

Вступ

Сервісно-орієнтована архітектура (SOA) є ключовою концепцією сучасних розподілених інформаційних систем, що забезпечує гнучкість, масштабованість та ефективність інтеграції програмних компонентів. У межах SOA кожен сервіс розглядається як самостійна функціональна одиниця, яка взаємодіє з іншими сервісами через стандартизовані інтерфейси. Проте для ефективного пошуку, вибору, композиції та управління сервісами необхідна їхня чітка формалізація, що включає визначення характеристик, можливостей та обмежень кожного сервісу. Формалізація ознак сервісу відіграє ключову роль у підвищенні сумісності сервісів. Через відсутність чітких підходів до опису характеристик та навичок сервісів веде до складнощів щодо їх повторного перевикористання, композиції тощо.

Вебсервіси забезпечують надання функціональних можливостей, які постачальники послуг пропонують різним споживачам у децентралізованому та розподіленому середовищі. З часом необхідність змін у вебслужбах спричиняє появу їхніх нових версій. Часто для задоволення потреб різних користувачів потрібне масштабне налаштування служб, що змушує постачальників розгортати декілька варіантів вебслужб, адаптованих для кожної групи користувачів [1].

Клієнтам не потрібно знати розташування серверів безпосередньо; їм досить мати доступ до служби пошуку, який може бути забезпечений за допомогою механізмів початкового завантаження, таких як початкові посилання або широкомовні повідомлення. Уся інша необхідна інформація про місцезнаходження отримується через службу пошуку.

Мета статті

У цій статті досліджуються та обґрунтовуються підходи до формалізації ознак сервісу у сервісно-орієнтованій архітектурі (SOA) з метою підвищення ефективності, масштабованості та гнучкості інформаційних систем. Приділено увагу наявним рішенням, таким як UDDI, оглянуто їх недоліки. У статті розглядаються ключові характеристики сервісів, їхня класифікація та принципи формального опису, що забезпечують сумісність, повторне використання та автономність сервісів у розподілених системах. Окрему увагу приділено практичним аспектам впровадження формалізованих ознак у сучасних SOA-платформах, включаючи використання метаданих та моделей машинного навчання для вдосконалення управління сервісами. Крім того, аналізуються перспективи використання семантичного пошуку для автоматизації процесів взаємодії між сервісами та користувачами, що може суттєво підвищити ефективність управління сервісно-орієнтованими системами.

Огляд схожих рішень

На сьогодні були деякі підходи, в яких були спроби сформувати загальні ознаки та характеристики сервісів. Серед таких підходів можна відзначити UDDI (Universal Description Discovery and Integration). Така система була спроєктована як реєстр вебсервісів, що дає можливості щодо опису та пошуку й інтеграції вебсервісів користувачами [2].

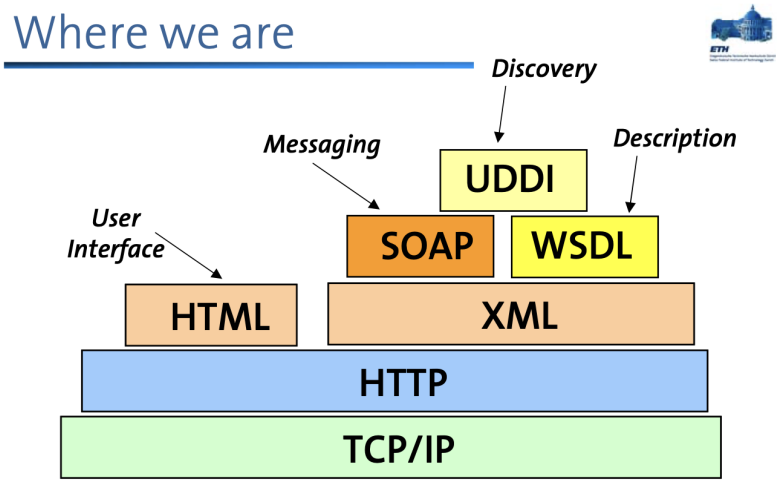


Рис. 1. Загальна структура UDDI [3]

UDDI базується на таких протоколах, як SOAP, XML, WSDL. SOAP використовується як основа для обміну повідомленнями між клієнтами та реєстром. Такий протокол забезпечує стандартний формат XML повідомлень для пошуку та реєстрації й отримання інформації про вебсервіси. Дані, які описують вебсервіс, представляються у форматі XML, також може використовуватися мова опису WSDL як формат опису інтерфейсів вебсервісів [3].

Така технологія доволі застаріла й не набула високого попиту серед користувачів. Серед недоліків можна відзначити підтримку застарілих протоколів комунікації в мережі – SOAP та XML. А також неможливість використання більш сучасних протоколів обміну та передачі даних.

Формалізація ознак сервісу

У Service-Oriented Architecture (SOA) будь-який сервіс може бути описаний за допомогою різних типів метаданих та типів даних. Ознаки сервісів можна поділити на деякі категорії: базові метадані, функціональні метадані, контекстуальні дані, дані про продуктивність і якість обслуговування тощо.

Метою базових метаданих є надання важливої інформації високого рівня про послугу, яка полегшує її ідентифікацію, виявлення та категоризацію. Він утворює фундаментальний рівень опису сервісу у сервісно-орієнтованій архітектурі (SOA).

Таблиця 1

Профіль сервісу

Елемент метаданих	**Ціль**	**Приклад**
Name	Унікально ідентифікує службу, дозволяючи посилатися на неї, отримати доступ і відрізняти її від інших служб	Платіжний шлюз
Description	Пояснення того, що робить служба	Обробляє платежі кредитними картками
Service ID/Endpoint	Знаходження та доступність до сервісу	https://api.example.com/v1/pay
Category/Tags	Дозволяє групувати службу з подібними послугами, полегшуючи організацію та навігацію каталогом послуг	Платежі, фінанси
Version	Відстежування та керування оновленнями або сумісністю	v1.2.3

Функціональні метадані описують робочі аспекти служби, такі як її входи, виходи, методи та протоколи. Його основна мета – надання докладної інформації про те, як взаємодіяти зі службою, що дозволяє розробникам і системам ефективно інтегрувати, споживати та використовувати службу в робочих процесах або програмах.

Таблиця 2

Профіль сервісу, який описує функціональні метадані

Елемент метаданих	Ціль	Приклад
Input Specifications	Визначає дані, необхідні для виклику служби, включаючи типи, формати та структури.	JSON object: {“user_id”: “string”}
Output Specifications	Описує дані, які повертає служба, включаючи типи, формати та структури.	JSON object: {“status”: “success”}
Methods/Operations	Перелічує конкретні функції або дії, які може виконувати служба.	createUser, getUserDetails
Service Protocol	Вказує протокол зв'язку, який використовується для доступу до служби.	HTTP, gRPC, SOAP
Supported Formats	Визначає формати даних, які підтримуються для введення/виведення.	JSON, XML
Authentication Requirements	Описує будь-які механізми автентифікації, необхідні для доступу до служби.	API Key, OAuth, JWT

Метадані продуктивності та якості обслуговування (QoS) надають показники та характеристики, які описують, наскільки добре сервіс працює за певних умов. Його основна мета полягає в тому, щоб користувачі та системи могли оцінити надійність, ефективність і масштабованість служби, допомагаючи встановити очікування та підтримувати прийняття обґрунтованих рішень під час інтеграції або вибору послуг.

Також, будуючи таку систему, корисно знати профілі вендорів-постачальних цих сервісів і користувачів. Профіль користувача має містити історію використання вебсервісів.

Таблиця 3
Профіль метаданих сервісу

QoS Element	Purpose	Example
Latency	Вимірює час, потрібний службі для відповіді на запит.	200 ms
Throughput	Вказує кількість запитів, які сервіс може обробити за секунду.	500 requests/second
Uptime/Availability	Показує відсоток часу, протягом якого послуга працює та доступна.	99.99% uptime
Error Rate	Надає відсоток невдалих або помилкових запитів за певний період.	0.02% error rate
Response Time Distribution	Надає детальні показники часу відгуку, включаючи середні значення та процентилі (наприклад, P90, P99).	P99: 300 ms

Таблиця 4
Профіль вендора сервісу

Used Services	Purpose	Example
Service Identifier	Ідентифікатор сервісу.	UUID-string
Category	Категорія, до якої належить сервіс.	«Фінанси»
Last used	Дата у вигляді UNIX timestamp, коли останній раз користувач звертався до сервісу або шукав його.	1738615470
Rating	Оцінка сервісу, яку поставив користувач.	4

Профілі користувачів можуть стати основою для побудови рекомендаційної моделі, яка допомагає автоматизувати пошук, вибір і використання сервісів у сервісно-орієнтованій архітектурі (SOA). Профілі користувачів також допомагають створювати рекомендаційні системи, які дозволяють автоматично підбирати сервіси на основі попереднього досвіду користувача та його подібності до інших споживачів. Це значно скорочує час на пошук потрібного сервісу, підвищує ефективність роботи та сприяє глибшій інтеграції сервісно-орієнтованих рішень у бізнес-процеси. Оцінку сервісу можна буде збирати таким чином, що після використання сервісу або його знаходженням питати у користувача, чи був сервіс корисний. Таким чином, можна буде закрити повний цикл рекомендації та пошукової видачі сервісу [4; 5].

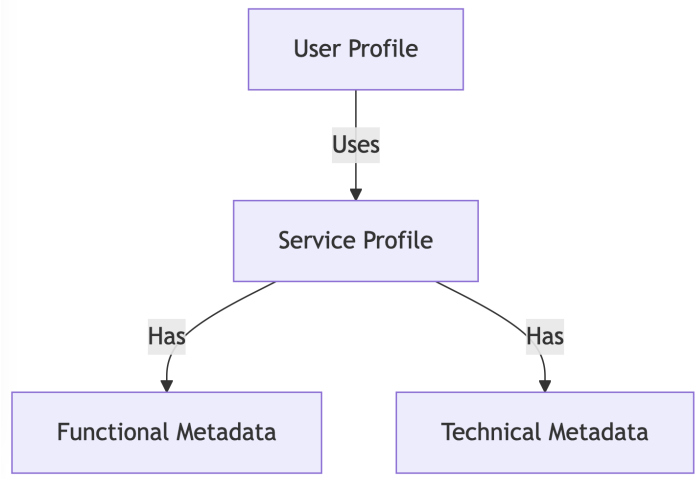


Рис. 2. Загальний вид реєстру

Також пропонується використання цієї системи як системи для виявлення нових та пошуку потрібних сервісів. Комунікація між сервісом і клієнтом може здійснюватися за допомогою будь-яких протоколів та форматів передачі й представлення даних. На відміну UDDI використовує SOAP (Simple Object Access Protocol) та WSDL (Web Services Description Language), що є досить важкими та складними для налаштування і підтримки. У випадку прямої комунікації користувач і сервіс можуть взаємодіяти через будь-які доступні технології, такі як REST, GraphQL, WebSockets, gRPC, MQTT тощо.

Ця технологічна нейтральність надає більшу гнучкість у виборі інструментів для розробки та розгортання сервісів. Розробники можуть самостійно вибирати найкращі способи інтеграції без обмежень, пов'язаних із застарілими стандартами чи необхідністю підтримки централізованих реєстрів сервісів.

Таким чином, пряма взаємодія між користувачем і сервісом не тільки підвищує продуктивність та безпеку, а й дозволяє вибирати будь-які сучасні технології для передачі даних, що робить систему більш гнучкою, адаптивною та ефективною.

Уніфікація контракту взаємодії між компонентами в розподілених системах

Розподілена система складається з безлічі компонентів, розташованих на різних машинах, які взаємодіють, обмінюючись даними та узгоджуючи свої дії, щоб для кінцевого користувача система виглядала як єдине ціле.

Для відображення інформації, що використовується у запущених програмах, застосовуються структури даних. Інформація представляється у вигляді послідовності байтів у повідомленнях, що передаються між компонентами розподіленої системи. Таким чином, перед передаванням даних потрібно перетворити структуру даних на послідовність байтів. Після отримання повідомлення ці дані також повинні бути здатні перетворюватися назад у початкову структуру даних [6].

Комп'ютери обробляють різні типи даних, які можуть відрізнятися залежно від того, де саме ці дані передаються. Примітивні типи даних можуть мати різні формати зберігання та представлення. Наприклад, цілі числа можуть впорядковуватися у двох способах: **big-endian**, де старший байт (Most Significant Byte) знаходиться першим, та **little-endian**, де старший байт (Most Significant Byte) знаходиться останнім, а молодший байт (Least Significant Byte) – першим. Архітектури комп'ютерів також відрізняються у способі представлення чисел із плаваючою комою [7].

Ще одна проблема пов'язана із кодуванням символів. У більшості UNIX-систем використовується кодування ASCII, яке використовує один байт на символ. Натомість стандарт Unicode застосовує два байти на символ, дозволяючи підтримувати тексти багатьма мовами.

Щоб забезпечити успішну передачу даних між комп'ютерами, необхідно конвертувати їх у стандартний формат. Якщо обидва комп'ютери мають однакову архітектуру, конвертація може бути пропущена. У всіх інших випадках дані перетворюються на стандартний зовнішній формат перед передачею та відновлюються у локальний формат після отримання. Для цього дані передаються у форматі відправника разом із описом цього формату, щоб одержувач міг виконати необхідну конвертацію.

Варто зауважити, що самі байти залишаються незмінними під час передачі. Усі типи даних, які передаються як параметри чи результати, повинні бути здатні до перетворення у прийнятний формат для підтримки механізмів Remote Procedure Call (RPC) або Remote Method Invocation (RMI). Таким чином, стандарт зовнішнього представлення даних визначає узгоджений спосіб представлення структур даних і примітивних значень [7].

Питання використання IDL для опису профілей сервісу

Мова опису інтерфейсу або мова визначення інтерфейсу (IDL) – це загальний термін для мови, яка дозволяє програмі чи об'єкту, написаним однією мовою, спілкуватися з іншою програмою, написаною невідомою мовою. IDL зазвичай використовуються для опису типів даних та інтерфейсів у незалежний від мови спосіб, наприклад, між написаними на C++ і написаними на Java. IDL зазвичай використовуються в програмному забезпеченні для віддаленого виклику процедур. У цих випадках машини на обох кінцях зв'язку можуть використовувати різні операційні системи та комп'ютерні мови. IDL пропонують міст між двома різними системами [8].

Різні мови опису інтерфейсів (IDL), такі як OAS, WSDL, Protocol Buffers (Protobuf), CORBA та GraphQL Schema Definition Language (SDL), відіграють важливу роль у розробці програмного забезпечення. Вони забезпечують стандартизовані засоби для опису та визначення інтерфейсів, слугуючи мостом між різними мовами програмування та системами. Використовуючи IDL як проміжне представлення, ці мови стають ключовими інструментами для автоматизованого створення API. Вони охоплюють структурні та поведінкові аспекти програмних компонентів, виступаючи основою для створення API у стандартизованому та нейтральному до мови програмування форматі. Такий підхід спрощує процес розробки, забезпечує узгодженість, зменшує кількість помилок і сприяє створенню API, що відповідають галузевим стандартам і найкращим практикам. У результаті ці мови опису інтерфейсів підвищують ефективність і надійність розробки програмного забезпечення, надаючи єдину основу для визначення та створення інтерфейсів [9].

Використання мови визначення інтерфейсу (IDL) для опису профілів послуг у сервісно-орієнтованій архітектурі (SOA) або подібних середовищах забезпечує структурований і стандартизований спосіб визначення послуг, забезпечуючи плавний зв'язок, інтеграцію та автоматизацію.

Використання спільних об'єктів у розподілених системах створює унікальну низку викликів:

Створення та видалення об'єктів: У розподілених системах об'єкти можуть створюватися та видалятися динамічно. Для керування їхнім життєвим циклом можна використовувати фабрики об'єктів, що забезпечують ефективний розподіл і вивільнення ресурсів.

Збереження стану об'єкта: Розподілені об'єкти часто потребують зберігати свій стан упродовж часу. Для цього можна використовувати об'єктні бази даних або механізми серіалізації, що забезпечують довготривале збереження стану.

Пошук об'єктів: Клієнтам необхідні механізми для пошуку об'єктів у розподіленому середовищі. Це може включати служби імен або реєстраційні служби, які відстежують місцеперебування об'єктів.

Керування версіями об'єктів: Забезпечення сумісності між клієнтами та різними версіями об'єктів є важливим завданням. Стратегії керування версіями дозволяють клієнтам взаємодіяти з відповідною версією об'єкта без збоїв.

Розглянемо варіант використання Protobuf [10] як IDL для опису сервісів.

Protobuf або Protocol Buffers – це бінарний формат серіалізації, розроблений Google для ефективного обміну даними між сервісами.

Розглянемо структуру ***.proto*** файлів.

Насамперед файл .proto – це схема даних для Protobuf. Це місце, де описується структура даних, яку планується серіалізувати або десеріалізувати.

Все починається з указання версії синтаксису. Наприклад, *syntax = "proto3"*. Це говорить компілятору, які правила використовувати під час аналізу вмісту. Потім іде оголошення пакета *package mypackage*. Це допомагає уникнути конфліктів, а також організувати код.

Серце *.proto* файлу – це оголошення повідомлення. Кожне повідомлення – це структура даних, яку можна серіалізувати.

Повідомлення визначаються за допомогою синтаксису повідомлення *MyMessage {}*. Всередині фігурних скобок ви описуєте поля даних. Поля можуть бути стандартними типами даних, такими як *int32*, *float*, *double*, *string* або *bool*. Також можуть бути типи користувачів або інші повідомлення.

У Protobuf є три типи правил для полів: *singular* для одиничних значень, *repeated* для масивів значень і карта для асоціативних масивів. Кожному полю присвоюється унікальний номер. Ці номери використовуються в бінарному представленні та дуже важливі для сумісності даних.

Починаючи номера полів з 1, будьте обережні, не перестрибуючи через десятки і сотні, це може бути корисно для майбутніх версій вашого протоколу. Якщо поле більше не потрібне, позначте його як *reserved*.

Протокол Protobuf є бінарним форматом серіалізації, який має кілька ключових переваг, що роблять його ідеальним вибором для сучасних розподілених систем. Protobuf є значно компактнішим і швидшим порівняно з текстовими форматами, такими як JSON або XML. Завдяки використанню бінарного формату дані можуть бути передані з мінімальними накладними витратами, що робить Protobuf ідеальним для високопродуктивних систем, де важлива швидкість передачі даних. Це особливо важливо в умовах великих обсягів даних, де зменшення обсягу передаваних повідомлень може суттєво покращити загальну ефективність системи.

Також Protobuf забезпечує строгий контроль над типами даних завдяки типізованим структурам, що дозволяє уникнути багатьох помилок, характерних для текстових форматів. Це дозволяє забезпечити високу надійність під час передачі даних між різними компонентами, що особливо важливо в умовах розподілених систем, де взаємодіють різні технології та мови програмування.

Використання Protobuf як IDL для опису профілів сервісів дозволяє отримати значну перевагу в плані ефективності, надійності та масштабованості, що робить його оптимальним вибором для розподілених систем та сервісно-орієнтованих архітектур (SOA). Крім того, його використання сприяє автоматизації створення API, підвищуючи узгодженість і зменшуючи кількість помилок, що часто виникають під час використання менш строгих форматів.

Висновки

У статті було розглянуто підходи до формалізації ознак сервісів у сервісно-орієнтованій архітектурі (SOA) й наявні напрацювання в цій галузі досліджень. Було проаналізовано структуру метаданих, що описують сервіси, а також визначено їхню роль у забезпеченні сумісності, повторного використання.

Окрему увагу приділено профілям користувачів та постачальників сервісів, оскільки вони відіграють ключову роль у вдосконаленні інтеграційних процесів. Зокрема, профілі користувачів можуть слугувати основою для побудови рекомендаційних моделей, що автоматизують вибір сервісів, підвищуючи ефективність та гнучкість роботи системи.

На відміну від UDDI технології, запропонований варіант не має обмежень щодо протоколів комунікації між споживачем та вебсервісом. Відсутність такого обмеження дає можливість використовувати різні протоколи та формати передачі даних, а це призводить до більшої гнучкості, якої не було раніше.

У цьому контексті важливим є питання використання мови опису інтерфейсу (IDL), яка дозволяє створювати стандартизовані інтерфейси для різних компонентів системи, незалежно від мов програмування, що використовуються. Використання IDL забезпечує чітке визначення типів даних та методів, що робить взаємодію між компонентами більш передбачуваною та ефективною. Одним з найефективніших інструментів у цьому напрямі є використання Protocol Buffers (Protobuf) як IDL для опису профілів сервісів. Протокол Protobuf є бінарним форматом серіалізації, який, на відміну від JSON, є компактнішим, швидшим і типізованим, що дозволяє зменшити накладні витрати на передавання даних у розподілених системах. Використання Protobuf як IDL сприяє більш ефективному опису структур даних та забезпечує сумісність під час передачі даних між різними системами. Крім того, Protobuf забезпечує високу гнучкість для динамічного оновлення контрактів без порушення сумісності, що особливо важливо для розвитку сервісно-орієнтованих архітектур.

Важливим напрямом розвитку цієї галузі є вдосконалення методів семантичного пошуку, зокрема через використання більш складних моделей для кращого розуміння запитів користувачів та точнішого підбору сервісів. Застосування технологій, таких як семантичний веб та контекстний аналіз, може значно покращити точність і швидкість пошуку відповідних сервісів, що базуються на більш глибокому розумінні запитів та потреб користувачів. Таким чином, запропонована модель опису сервісів може слугувати підґрунтям для інтелектуалізації пошуку й рекомендацій сервісів у системі, що включає механізми семантичного аналізу, адаптивні рекомендації та вдосконалене управління сервісами на основі формалізованих ознак.

Література

1. Marwaha P., Banati H., Bedi P. UDDI Extensions for Temporally Customized Web Services. *International Conference on Computational Intelligence: Modeling Techniques and Applications,* 2013. P. 184–190.

2. UDDI (Universal Description, Discovery and Integration), 2023. URL: https://www.techtarget.com/whatis/definition/UDDI-Universal-Description-Discovery-and-Integration (дата звернення: 01.02.2025).

3. Pautasso C. Distributed Systems UDDI and beyond, Information and Communication System. URL: https://vs.inf.ethz.ch/edu/WS0405/VS/VS-050131.pdf (дата звернення: 01.02.2025).

4. Relationship between UDDI and WSDL, 2021. URL: https://www.ibm.com/docs/en/rsas/7.5.0?topic=uddi-relationship-between-wsdl (дата звернення: 01.02.2025).

5. Лобур М. В., Шварц М. Є., Стех Ю. В., Моделі і методи прогнозування рекомендацій для колаборативних рекомендаційних систем. *Інформаційні системи та мережі*, 2018. № 901. С. 68–75.

6. Мелешко Є. Проблеми сучасних рекомендаційних систем та методи їх рішення. *Системи управління, навігації та зв'язку*, 2018, № 4(50). С. 120–125.

7. Marinov M., Valova I. Component Interaction in Distributed Knowledge-Based Systems. *TEM Journal*. Volume 8. Issue 3, 2019. P. 721–727.

8. Marshaling in Distributed System 2024. URL: https://www.geeksforgeeks.org/marshalling-in-distributed-system (дата звернення: 03.02.2025).

9. Interface Description Language. URL: https://en.wikipedia.org/wiki/Interface_description_language (дата звернення: 03.02.2025).

10. What is IDL in Computing? URL: https://60sec.site/terms/what-is-idl-in-computing-interface-definition-language (дата звернення: 03.02.2025).

11. Protobuf 3.0 specification. URL: https://protobuf.dev/reference/protobuf/proto3-spec/ (дата звернення: 03.02.2025).

References

1. Marwaha P., Banati H., Bedi P. (2013). UDDI Extensions for Temporally Customized Web Services. *International Conference on Computational Intelligence: Modeling Techniques and Applications*. P. 184–190.

2. UDDI (Universal Description, Discovery and Integration), (2023). [Online]. Retrieved from: https://www.techtarget.com/whatis/definition/UDDI-Universal-Description-Discovery-and-Integration.

3. Pautasso, C. Distributed Systems UDDI and beyond, Information and Communication System. [Online]. Retrieved from: https://vs.inf.ethz.ch/edu/WS0405/VS/VS-050131.pdf.

4. Relationship between UDDI and WSDL. (2021). [Online]. Retrieved from: https://www.ibm.com/docs/en/rsas/7.5.0?topic=uddi-relationship-between-wsdl.

5. Lobur, M. V., Schwartz, M. E., Stekh, Y. V. (2018). Models and methods of forecasting recommendations for collaborative recommender systems. *Information Systems and Networks*. No. 901. P. 68–75 [in Ukrainian].

6. Meleshko, E. (2018). Problems of modern recommender systems and methods for their solution. *Control, navigation and communication systems*, No. 4(50). P. 120–125 [in Ukrainian].

7. Marinov, M., Valova, I. (2019). Component Interaction in Distributed Knowledge-Based Systems. *TEM Journal*. Volume 8. Issue 3. P. 721–727.

8. Marshaling in Distributed System. (2024). [Online]. Retrieved from: https://www.geeksforgeeks.org/marshalling-in-distributed-system.

9. Interface Description Language. [Online]. Retrieved from: https://en.wikipedia.org/wiki/Interface_description_language.

10. What is IDL in Computing? [Online]. Retrieved from: https://60sec.site/terms/what-is-idl-in-computing-interface-definition-language.

11. Protobuf 3.0 specification. [Online]. Retrieved from: https://protobuf.dev/reference/protobuf/proto3-spec/.

НОТАТКИ

НАУКОВЕ ВИДАННЯ

ІНФОКОМУНІКАЦІЙНІ ТА КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ

INFOCOMMUNICATION AND COMPUTER TECHNOLOGIES

№ 2 (08) 2024

Підписано до друку: 27.12.2024.
Формат 60x84/8. Arial.
Папір офсет. Цифровий друк. Ум. друк. арк. 8,14. Замов. № 0425/319. Наклад 300 прим.

Видавництво і друкарня – Видавничий дім «Гельветика»
65101, Україна, м. Одеса, вул. Інглезі, 6/1
Телефон +38 (095) 934 48 28, +38 (097) 723 06 08
E-mail: mailbox@helvetica.ua
Свідоцтво суб'єкта видавничої справи
ДК № 7623 від 22.06.2022 р.