

Національний університет
“Полтавська політехніка імені Юрія Кондратюка”

National University
“Yuri Kondratyuk Poltava Polytechnic”

Системи управління, навігації та зв'язку

Control, navigation and communication systems

Випуск 1 (79)

Issue 1 (79)

Щоквартальне видання

Засноване у 2007 році

У журналі відображені результати наукових досліджень з розробки та удосконалення систем управління, навігації та зв'язку у різних проблемних галузях.

Засновник і видавець:

Національний університет
“Полтавська політехніка імені Юрія Кондратюка”

Телефон:

+38 (050) 302-20-71

E-mail редколегії:

kuchuk_nina@ukr.net

Інформаційний сайт:

<http://journals.nupp.edu.ua/sunz>

Quarterly

Founded in 2007

Journal represent the research results on the development and improvement of control, navigation and communication systems in various areas

Founder and publisher:

National University
"Yuri Kondratyuk Poltava Polytechnic"

Phone:

+38 (050) 302-20-71

E-mail of the editorial board:

kuchuk_nina@ukr.net

Information site:

<http://journals.nupp.edu.ua/sunz>

За достовірність викладених фактів, цитат та інших відомостей відповідальність несе автор

Журнал індексується міжнародними наукометричними базами: Index Copernicus (ICV = **82.05**),
General Impact Factor, Google Scholar, Academic Resource Index, Scientific Indexed Service

Затверджений до друку Вченою Радою Національного університету

“Полтавська політехніка імені Юрія Кондратюка” (протокол від 12 березня 2025 року № 5).

Ідентифікатор медіа R30-04135 згідно з рішенням Національної ради України
з питань телебачення і радіомовлення від 25.04.2024 № 1416

Включений до “Переліку наукових фахових видань України, в яких можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора наук, кандидата наук та ступеня доктора філософії” до категорії Б – наказами МОН України від 17.03.2020 № 409 та від 09.02.2021 № 157

Полтава • 2025

Редакційна колегія

Головний редактор:

КОСЕНКО Віктор Васильович
(*д-р техн. наук, проф., Полтава, Україна*).

Заступники головного редактора:

НЕСТЕРЕНКО Катерина Сергіївна
(*д-р техн. наук, проф., Київ, Україна*);
ШЕФЕР Олександр Віталійович
(*д-р техн. наук, проф., Полтава, Україна*).

Члени редакційної колегії:

БЛАУНШТЕЙН Натан Олександрович
(*д-р техн. наук, проф., Ізраїль*);
БОГОМ'Я Володимир Іванович
(*д-р техн. наук, проф., Київ, Україна*);
ВАРБАНЕЦЬ Роман Анатолійович
(*д-р техн. наук, проф., Одеса, Україна*);
ВЕСОЛОВСЬКИЙ Кшиштоф
(*д-р техн. наук, проф., Польща*);
ГАВРИЛКО Євген Володимирович
(*д-р техн. наук, проф., Київ, Україна*);
ГАШИМОВ Ельшан Гіяс огли
(*д-р наук, проф., Баку, Азербайджан*);
ГЛИВА Валентин Анатолійович
(*д-р техн. наук, проф., Київ, Україна*);
ДАКІ Олена Анатоліївна
(*д-р техн. наук, доц., Ізмаїл, Україна*);
КОВАЛЕНКО Андрій Анатолійович
(*д-р техн. наук, проф., Харків, Україна*);
КОЛОМІЙЦЕВ Олексій Володимирович
(*д-р техн. наук, проф., Харків, Україна*);
КОРОБКО Богдан Олегович
(*д-р техн. наук, доц., Полтава, Україна*);
КРАСНОБАЄВ Віктор Анатолійович
(*д-р техн. наук, проф., Харків, Україна*);
КУЧУК Георгій Анатолійович
(*д-р техн. наук, проф., Харків, Україна*);
ЛЕВЧЕНКО Лариса Олексіївна
(*д-р техн. наук, доц., Київ, Україна*);
ЛУКОВА-ЧУЙКО Наталія Вікторівна
(*д-р техн. наук, проф., Київ, Україна*);
ЛУНТОВСЬКИЙ Андрій Олегович
(*д-р техн. наук, проф., Німеччина*);
МИРОНЦОВ Микита Леонідович
(*д-р ф.-м. наук, с.н.с., Київ, Україна*);
ПЄВНЄВ Володимир Яковлевич
(*д-р техн. наук, доц., Харків, Україна*);
ПИСАРЧУК Олексій Олександрович
(*д-р техн. наук, проф., Київ, Україна*);
ПОДКОПАЄВ Сергій Вікторович
(*д-р техн. наук, проф., Покровськ, Україна*);
СЕМЕНОВ Сергій Геннадійович
(*д-р техн. наук, проф., Краків, Польща*);
ТРИСТАН Андрій Вікторович
(*д-р техн. наук, проф., Черкаси, Україна*);
ФРОЛОВ Євгеній Андрійович
(*д-р техн. наук, проф., Полтава, Україна*);
ЧОРНИЙ Олексій Петрович
(*д-р техн. наук, проф., Кременчук, Україна*);

Відповідальний секретар:

КУЧУК Ніна Георгіївна
(*д-р техн. наук, проф., Харків, Україна*).

Технічні секретарі:

ЗАХАРЧЕНКО Руслан Володимирович
(*канд. техн. наук, доц., Полтава, Україна*);
ПЕТРОВСЬКА Інна Юріївна
(*д-р філософії, Харків, Україна*).

Editorial board

Editor-in-Chief:

Viktor KOSENKO
(*Dr. Sc. (Tech.), Prof., Poltava, Ukraine*).

Associates editor:

Katerina NESTERENKO
(*Dr. Sc. (Tech.), Prof., Kyiv, Ukraine*);
Oleksandr SHEFER
(*Dr. Sc. (Tech.), Prof., Poltava, Ukraine*).

Editorial board members:

Nathan BLAUNSTEIN
(*Dr. Sc. (Tech.), Prof., Israel*);
Volodymyr BOHOMYA
(*Dr. Sc. (Tech.), Prof., Kyiv, Ukraine*);
Roman VARBANETS
(*Dr. Sc. (Tech.), Prof., Odesa, Ukraine*);
Krzysztof WESOŁOWSKI
(*Dr. Sc. (Tech.), Prof., Poland*);
Yevhen HAVRILKO
(*Dr. Sc. (Tech.), Prof., Kyiv, Ukraine*);
Elshan Giyas oglu HASHIMOV
(*Dr. Sc., Prof., Baku, Azerbaijan*);
Valentyn GLYVA
(*Dr. Sc. (Tech.), Prof., Kyiv, Ukraine*);
Olena DAKI
(*Dr. Sc. (Tech.), Ass. Prof., Izmail, Ukraine*);
Andriy KOVALENKO
(*Dr. Sc. (Tech.), Prof., Kharkiv, Ukraine*);
Oleksii KOLOMIITSEV
(*Dr. Sc. (Tech.), Prof., Kharkiv, Ukraine*);
Bohdan KOROBKO
(*Dr. Sc. (Tech.), Ass. Prof., Poltava, Ukraine*);
Viktor KRASNOBAYEV
(*Dr. Sc. (Tech.), Prof., Kharkiv, Ukraine*);
Heorhii KUCHUK
(*Dr. Sc. (Tech.), Prof., Kharkiv, Ukraine*);
Larysa LEVCHENKO
(*Dr. Sc. (Tech.), Ass. Prof., Kyiv, Ukraine*);
Natalia LUKOVA-CHUIKO
(*Dr. Sc. (Tech.), Prof., Kyiv, Ukraine*);
Andryy LUNTOVSKYY
(*Dr. Sc. (Tech.), Prof., Kharkiv, Ukraine*);
Mykyta MYRONTSOV
(*Dr. Sc. (Ph.&M.), Senior Res., Kyiv, Ukraine*);
Volodymyr PEVNEV
(*Dr. Sc. (Tech.), Ass. Prof., Kharkiv, Ukraine*);
Oleksii PYSARCHUK
(*Dr. Sc. (Tech.), Prof., Kyiv, Ukraine*);
Serhii PODKOPAIEV
(*Dr. Sc. (Tech.), Prof., Pokrovsk, Ukraine*);
Serhii SEMENOV
(*Dr. Sc. (Tech.), Prof., Krakow, Poland*);
Andrii TRYSTAN
(*Dr. Sc. (Tech.), Prof., Cherkasy, Ukraine*);
Yevhen FROLOV
(*Dr. Sc. (Tech.), Prof., Poltava, Ukraine*);
Oleksii CHORNYI
(*Dr. Sc. (Tech.), Prof., Kremenchuk, Ukraine*).

Responsible secretary:

Nina KUCHUK
(*Dr. Sc. (Tech.), Prof., Kharkiv, Ukraine*).

Technical secretaries:

Ruslan ZAKHARCHENKO
(*PhD (Tech.), Ass. Prof., Poltava, Ukraine*);
Inna PETROVSKA
(*PhD (Tech.), Kharkiv, Ukraine*).

З М І С Т

АВТОМОБІЛЬНИЙ, РІЧКОВИЙ, МОРСЬКИЙ ТА АВІАЦІЙНИЙ ТРАНСПОРТ

<i>Костікова М. В., Карпішен Б. С., Неронов С. М., Плехова Г. А., Очеретенко С. В.</i> Рухомий пристрій для контролю відхилень положень рейкових шляхів	5
<i>Положий Д. С., Орехов О. О.</i> Прогнозування ризиків дорожньо-транспортного руху на основі нейронної мережі	10
<i>Чумак О. О.</i> Модель планування маршрутів руху однорідної групи безпілотних літальних апаратів ...	17

УПРАВЛІННЯ В СКЛАДНИХ СИСТЕМАХ

<i>Заповловський М. Й., Мезенцев М. В., Оліфір М. В.</i> Моделювання тягового електропривода дизель-поїзда з урахуванням використання гумо-кордової муфти в процесі буксування колісних пар	23
<i>Запорожець О. В., Соболев О. Р., Гриньов Д. В., Зиков І. С.</i> Інтеграція стандартів ISO в управління програмними проектами	31
<i>Kapiton A., Franchuk T., Tyshchenko D., Desiatko A., Sas N.</i> Modeling of management objects and processes	37
<i>Nevlyudov I., Yevsieiev, V., Gurin, D.</i> Model development of dynamic representation a model description parameters for the environment of a collaborative robot manipulator within the Industry 5.0 Framework	42
<i>Решетнікова П. Е., Заковоротний О. Ю.</i> Моделювання вертикальних коливань поїзда, що виникають під час руху залізничною колією зі стиковою нерівністю.....	49
<i>Fedorovich O., Malietiev L.</i> Matrix model of stratified representation of the high-tech products modernization project based on component and risk-oriented approaches	55
<i>Царьков Р. М., Єрмілов Р. О., Шефер О. В.</i> Використання високовольтних генераторів, керованих напругою, для підвищення точності систем фазового автопідстроювання частоти	62
<i>Shapovalova S., Baranichenko O.</i> Modeling of functional processes between components of a hierarchical control object	67
<i>Янко А. С., Крук О. О., Гончаренко С. О.</i> Теоретичне обґрунтування математичної моделі безвідмовності системи обробки інформації автоматизованих робототехнічних платформ у системі залишкових класів	72

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

<i>Huts V., Gorokhovatskyi O.</i> Efficient fault detection in industrial equipment using PCA and SMOTE enhanced neural networks	77
<i>Заковоротний О. Ю., Сапальський О. А.</i> Вирішення проблем продуктивності за допомогою асинхронних методів в JavaScript	83
<i>Зачепило М. О.</i> Багаторівневі екосистеми для адаптивних агентів: від моделювання до практичних застосувань	87
<i>Іващенко Г. С., Кононенко А. І., Тимошенко Д. О.</i> Метаевристичні методи вирішення задачі гільйотинного розкрою	91
<i>Ігнатюк Є. О., Попов А. В.</i> Методико-інструментальні засоби автоматизованого аналізу даних результатів UX-досліджень з використанням систем штучного інтелекту	96
<i>Клівець С. І., Кулешов О. В., Зубко Р. В.</i> Метод формування підсистеми обробки оперативних транзакцій Інтернету речей	107
<i>Леонов С. Ю., Тиртишний Д. А.</i> Розробка програмної платформи для тестування продуктивності клієнтської частини вебзастосунку	111
<i>Носко С. В., Бульба С. С., Коломійцев О. В., Лисиця Д. О., Молчанов Г. І.</i> Пропозиції щодо авторизації в сайдкар компоненті мікросервісної архітектури	116
<i>Protsenko A., Fedorchenko V.</i> Mobile application security analysis model based on artificial intelligence ...	124
<i>Skakalina O., Holub A.</i> Large language models: business applications and development prospects	129
<i>Шостак А. В.</i> Про особливості формування триплетів для навчання сіамської нейронної мережі	134
<i>Shupyliuk M., Martovytskyi V.</i> Analysis of personality detection and writer identification methods	138
<i>Barkovska O., Ni Ya., Havrashenko A., Peretiaka Ye., Romanenko A.</i> System for detecting critical human health conditions based on the analysis of physiological indicators	143
<i>Льбіна І. В., Костін К. Д.</i> Сучасні методи OLAP у партнерському маркетингу: оптимізація прибутку за допомогою багатовимірного аналізу даних	150
<i>Коваленко А. А., Куценко Т. Г., Шаповал А. С., Ситник О. В., Ні Я. С.</i> Огляд методів аналізу безпечного програмного забезпечення	156

ЦИВІЛЬНА БЕЗПЕКА

<i>Бірук Я. І., Климчук А. В.</i> Дослідження умов максимальної ефективності перфорованих електромагнітних екранів	162
--	-----

<i>Бурдейна Н. Б., Зозуля Л. А., Петруньок Т. Б.</i> Застосування композиційного залізовмісного матеріалу для екранування електромагнітних полів ультрависоких і вищих частот	165
<i>Глива В. А., Довгановський М. О.</i> Проектування одягу для захисту від термічних впливів та збереження тепла	169
<i>Резнік Д. В., Ченчева О. О., Лашко Є. Є., Петренко І. С., Гаврилець Д. С.</i> Дослідження стану повітряного середовища в приміщеннях закладів передвищої та професійної (професійно-технічної) освіти	172
<i>Третьяков О. В., Халмурадов Б. Д., Кічата Н. М., Ремська А. В.</i> Підхід до кількісної оцінки стійкості об'єктів критичної інфраструктури	178

ЗВ'ЯЗОК, ТЕЛЕКОМУНІКАЦІЇ ТА РАДІОТЕХНІКА 7

<i>Зоураб Ю. Р., Єрмілова Н. В.</i> Ідентифікація об'єктів поштового зв'язку для оптимізації опрацювання поштової кореспонденції	184
<i>Kuchuk N., Udyansky R., Usichenko V., Buslov P., Huk A.</i> Traffic providing holographic type of communications	188
<i>Леві Л. І., Зима О. Є.</i> Концептуальна модель ієрархічної організації інженерних мережних систем ...	193
<i>Почерняев В. М., Магомедова М. С., Сивкова Н. М.</i> Підвищення завадостійкості мобільних цифрових комбінованих радіотехнічних систем НВЧ діапазону	196
<i>Савченко М. В., Шиман М. В.</i> Метод аналізу завантаження вузлів кластеру MESH-мережі на основі математичної моделі мереж Джексона	201
<i>Serkov A., Breslavets V., Breslavets J., Yakovenko I.</i> Generation of surface oscillations of semiconductor structures by charged particle flows	205
<i>Штомпель М. А., Ковтун І. В., Гусєва І. І.</i> Оцінювання ефективності декодування за впорядкованими статистиками алгебраїчних згорткових кодів	209
<i>Yareschenko, V., Kosenko, V.</i> Method of generating energy efficient codes	213
АЛФАВІТНИЙ ПОКАЖЧИК	218

Організації авторів

Федеральний інститут Парани, Парана, Бразилія
 Вище професійне училище № 7, Кременчук, Україна
 Державний торговельно-економічний університет, Київ, Україна
 Державний університет «Київський авіаційний інститут», Київ, Україна
 Київський національний університет будівництва і архітектури, Київ, Україна
 Київський фаховий коледж зв'язку, Київ, Україна
 Кременчуцький національний університет імені Михайла Остроградського, Кременчук, Україна
 Національна академія Служби безпеки України, Київ, Україна
 Національна академія статистики, обліку та аудиту, Київ, Україна
 Національний аерокосмічний університет «ХАІ», Харків, Україна
 Національний технічний університет «Харківський політехнічний інститут», Харків, Україна
 Національний технічний університет України «КПІ імені Ігоря Сікорського», Київ, Україна
 Національний університет оборони України, Київ, Україна
 Національний університет «Полтавська політехніка імені Юрія Кондратюка», Полтава, Україна
 Харківський національний автомобільно-дорожній університет, Харків, Україна
 Харківський національний університет радіоелектроніки, Харків, Україна
 Харківський національний університет Повітряних Сил імені Івана Кожедуба, Харків, Україна
 Українська державна льотна академія, Кропивницький
 Український державний університет залізничного транспорту, Харків, Україна

Authors affiliation

Instituto Federal do Parana, Parana, Brazil
 Higher Vocational School No. 7, Kremenchuk
 State University of Trade and Economics, Kyiv, Ukraine
 State University "Kyiv Aviation Institute," Kyiv, Ukraine
 Kyiv National University of Construction and Architecture, Kyiv, Ukraine
 Kyiv Vocational College of Communication, Kyiv
 Kremenchuk Mykhailo Ostrohradskyi National University, Kremenchuk, Ukraine
 National Academy of the Security Service of Ukraine, Kyiv, Ukraine
 National Academy of Statistics, Accounting and Audit, Kyiv, Ukraine
 National Aerospace University "KHAI", Kharkiv, Ukraine
 National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine
 National Technical University "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
 National Defence University of Ukraine, Kyiv
 National University "Yuri Kondratyuk Poltava Polytechnic", Poltava, Ukraine
 Kharkiv National Automobile and Highway University, Kharkiv, Ukraine
 Kharkiv National University of Radio Electronics, Kharkiv, Ukraine
 Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine
 Ukrainian State Flight Academy, Kropyvnytskyi
 Ukrainian State University of Railway Transport, Kharkiv, Ukraine

Автомобільний, річковий, морський та авіаційний транспорт

УДК 625:681.786

doi: 10.26906/SUNZ.2025.1.5-9

М. В. Костікова, Б. С. Карпішен, С. М. Неронов, Г. А. Плехова, С. В. Очеретенко
Харківський національний автомобільно-дорожній університет, Харків, Україна

РУХОМИЙ ПРИСТРІЙ ДЛЯ КОНТРОЛЮ ВІДХИЛЕНЬ ПОЛОЖЕНЬ РЕЙКОВИХ ШЛЯХІВ

Анотація. Залізничний транспорт є важливою інфраструктурною основою економіки долаючи природні перешкоди. Експлуатація цих споруд вимагає постійного контролю технічного стану рейкових конструкцій. Ходові рейки разом зі шпалами та контактна рейка в метрополітені або контактний провід на залізниці та трамвайних коліях є єдиним комплексом, який створює необхідні умови для руху рейкового електротранспорту У статті проаналізовано пристрої для контролю відхилень положення рейкових колій, які містять світловипромінювач з регулятором положення та платформу, що рухається по балі. Недоліком цих пристроїв є недостатня функціональність на кривих ділянках. В роботі пропонується рухома система для вимірювання параметрів рейкових шляхів підвищеної точності на основі датчиків деформації та лазера. Пристрій належить до рейкового транспорту і може застосовуватися на наземному залізничному транспорті, метрополітені та на трамвайних коліях міста, що дозволить контролювати стан зміщення і прогини, що виникають в результаті впливу зовнішніх природно-кліматичних факторів.

Ключові слова: рейкові шляхи, лазер, коліматор, фотоприймач, рейковий візок.

Вступ

Ходові рейки разом зі шпалами та контактна рейка в метрополітені або контактний провід на залізниці та трамвайних коліях є єдиним комплексом, який створює необхідні умови для руху рейкового електротранспорту. Несанкціонованим змінам піддається взаємне положення по висоті ходових рейок, кривизна останніх, ширина рейкової колії і т. д [1]. Під час експлуатації можливі просідання лівої або правої рейки та (або) зміна ширини рейкової колії тощо [2, 3]. Тому варто подбати про автоматизований контроль стану рейкових шляхів, а саме контроль відхилень положення рейкових колій по горизонтальній і вертикальній осях, а також зазорів у стиках.

Аналіз публікацій. Відомо пристрої для контролю відхилень положення рейкових колій, що містять випромінювач світла (лазерне джерело) з регулятором положення, платформу що рухається вздовж променю по рейкам, забезпечену електродвигуном з приводом, на якій знаходяться фотоприймачі, дві камери для контролю зазорів у стиках направляючих рейок і взаємного зміщення торців рейок, що стикаються між собою, забезпечених калібрувальними елементами, і встановлених в торці платформи, блок контролю положення рейкового шляху, який забезпечений аналізатором зміщення та реєстратором, що містять контрольновимірювальні вузли, які служать для контролю відхилень положення рейкових колій по горизонтальній та вертикальній осях [4, 5].

Недоліком цих пристроїв є недостатня функціональна можливість, яка обумовлена виміром параметрів рейкової колії тільки на прямолінійних ділянках рейкового шляху і практична непрацездат-

ність пристрою на криволінійних ділянках та недостатня точність вимірювань параметрів рейкової колії, у зв'язку з тим, що не можливо чітко і точно визначити контури лазерної плями на прозорому екрані, що приведе до неточності контролю відхилень положення рейкових колій по горизонтальній і вертикальній осях [6–8].

Мета та постановка задачі. Метою цієї роботи є покращення точності існуючих систем особливо при їх використанні на криволінійних ділянках колій та поворотах. Також пристрій повинен бути компактним та рухомою системою. Тому була поставлена задача розробки пристрою для вимірювання параметрів рейкових шляхів підвищеної точності з підвищеними функціональними можливостями.

Виклад основного матеріалу

Побудова вимірювальної системи. Поставлена задача була вирішена розробкою пристрою для контролю рейкових шляхів (рис. 1, 2), який містить рухомий шляховимірювальний візок 1, забезпечений електродвигуном з приводом, лазер 2, оптичний коліматор 3, вузол 4 розгортки лазерного променя 6 з регулятором положення його в просторі 5 і перший 7 фотоприймач, встановлений на шляху пересування рухомого шляховимірювального візка 1, на якій встановлені другий 8 і третій 10 фотоприймачі, що служать для контролю відхилень положення рейкових колій по горизонтальній і вертикальній осях. Причому другий фотоприймач 8 знаходиться на гіроскопічній платформі 9 на лівому боці шляховимірювального візка 1 і контролює відхилення лівої рейки по горизонтальній і вертикальній осях, третій фотоприймач 10 знаходиться на гіроскопічній платформі 11 і контролює відхилення правої рейки по горизонтальній і вертикальній осях, крім того шляховимірю-

вальний візок 1 додатково забезпечений двома камерами 12 і 13 для контролю зазорів в стиках направляючих рейок і взаємного зміщення торців рейки, що стикуються між собою. Кожна з камер забезпечена

калібрувальними 14 і 15 елементами, і встановлена в торці шляховимірювального візка 1, які зв'язані з блоком 16 контролю відхилення положення рейкових колій по горизонталі та вертикалі.

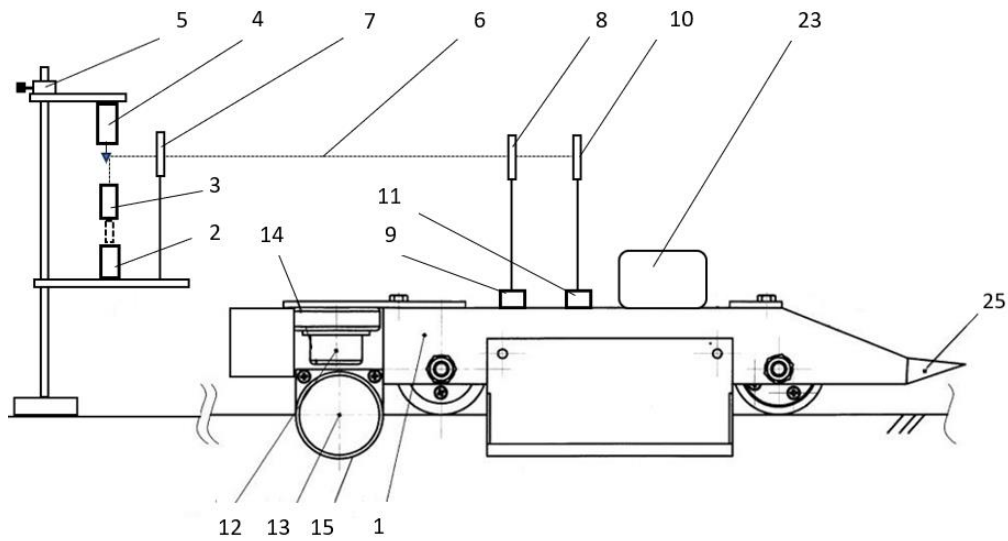


Рис. 1. Загальний вигляд пристрою

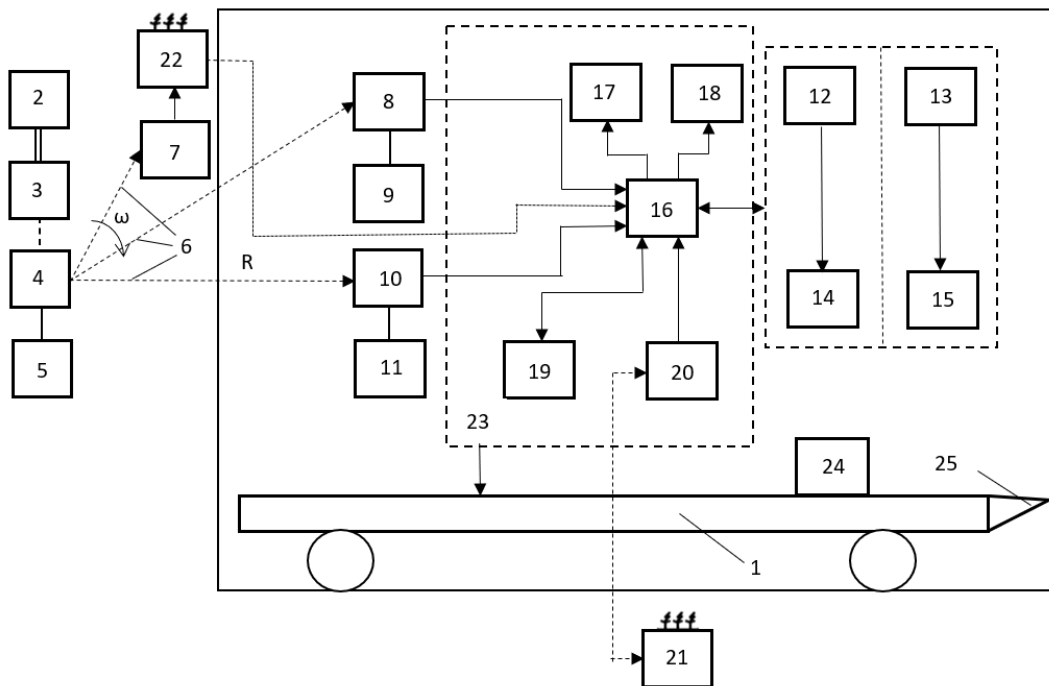


Рис. 2. Контрольно-вимірювальна структурна схема пристрою

Входи блока 16 контролю відхилення положення рейкових колій по горизонталі та вертикалі підключені до виходів першого 7, другого 8 та третього 10 фотоприймачів і до камер 12 та 13, до виходу датчика пройденого шляху 19 і виходу приймача команд 20 управління, пов'язаного з першим блоком 21 дистанційного керування, а виходи блока 16 контролю відхилення положення рейкових колій по горизонталі та вертикалі підключені, відповідно, до блока 17 реєстрації та аналізу інформації і блока 18 живлення.

Крім того перший фотоприймач 7 зв'язаний з другим блоком 22 дистанційного керування. Всі

блоки та вузли, що здійснюють прийом і обробку інформації, знаходяться у блоці 23 прийому і обробки інформації, розташованому на шляховимірювальному візку 1, а шляховимірювальний візок 1 забезпечений стопорним пристроєм 24 для його фіксування в потрібному місці і розсікачем-скидачем 25, виконаному на передньому торці шляховимірювального візка 1.

Принцип роботи пристрою. Пристрій для контролю рейкових шляхів працює наступним чином: джерело лазерного випромінювання 2, оптичний коліматор 3, вузол 4 розгортки лазерного 6 променя з

регулятором положення його в просторі 5 і перший фотоприймач 7 встановлюють на елементах рейкових шляхів на опорі, через заздалегідь задані інтервали руху шляховимірювального візка 1.

Лазер 2, оптичний коліматор 3 і вузол 4 розгортки лазерного променя 6, які розташовуються на опорі на такій же висоті, як перший фотоприймач так і другий та третій фотоприймачі 8 і 10, що знаходяться на відповідних гіроскопічних платформах, розташованих по боках шляховимірювального візка 1.

Чутлива поверхня всіх фотоприймачів повернута у напрямку вузла розгортки лазерного променя. У пристрої застосовується лазер 2 з безперервним випромінюванням променя певної довжини хвилі. Послідовно з лазером розташовується оптичний коліматор 3, який складається з об'єктива, у фокальній площині якого розміщується вихід лазерного променя 6. Оптичний коліматор 3 забезпечує паралельність лазерного променя і тому, за рахунок, практично, нульової розбіжності, вся енергія лазерного променя буде зосереджена на чутливій поверхні кожного багатоеlementного фотоприймача, виконаного у вигляді лінійки елементів певної довжини h (рис. 3), причому кожний фотоприймач має однакову кількість елементів [9].

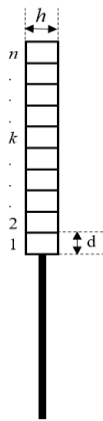


Рис. 3. Фотоприймач

Кожний елемент фотоприймача виконаний із певного матеріалу [10], що без особливих перешкод пропускає лазерний промінь 6. Всі фотоприймачі, що складаються із n елементів, містять фотодіоди, перед якими розташовуються чутлива поверхня і оптичний фільтр, що пропускає промінь тільки певної довжини хвилі лазера.

Всі елементи фотоприймача розміщуються у герметичному корпусі [10]. Вузол розгортки 4, наприклад, представляє собою двигун, на валу якого розташовується дзеркало або призма із дзеркальною гранню, що знаходиться під кутом 45° до лазерного променя, або круглий оптичний клин, за рахунок чого промінь лазера розгортається у горизонтальній площині. Регулятор положення в просторі 5 регулює вузол 4 розгортки лазерного променя по вертикалі таким чином, щоб лазерний промінь послідовно пробігав по чутливій поверхні елемента з відповідним однаковим номером k елемента кожного фотоприймача (рис. 4). Електричний сигнал з фотодіода кожного фотоприймача поступає на підсилювач електричного сигналу фотоприймача, який, після підсилення,

поступає на входи блока 16 контролю відхилення положення рейкових колій по горизонталі та вертикалі (рис. 2), причому електричний сигнал першого фотоприймача 7 передається через перший блок 21 дистанційного керування.

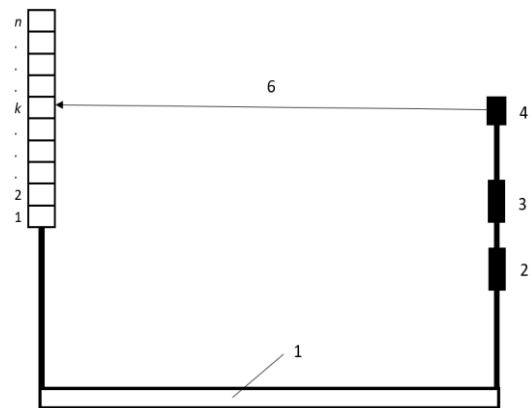


Рис. 4. Рейкова колія без відхилення по вертикалі

Якщо положення однієї із рейкових колій буде зміщено по вертикалі, то другий 8 або третій 10 фотоприймачі, в залежності яка із колій буде зміщена, змінять своє положення відносно лазерного променя, що розгортається, і лазерний промінь 6 буде обігати інший за номером (наприклад, j або p) елемент відповідного фотоприймача, а на першому фотоприймачі 7 номер елемента не поміняється. Підсилені електричні сигнали з відповідних елементів фотоприймачів надходять блока 16 контролю відхилення положення рейкових по горизонталі та вертикалі до блока 17 реєстрації та аналізу інформації. Різниця між номером k елемента першого 7 фотоприймача і номером елемента відповідного фотоприймача 8 або 10, що знаходиться над зміщеною по вертикалі колією помножена на певну довжину елемента d , дасть величину зміщення по вертикалі відповідної колії (рис. 5).

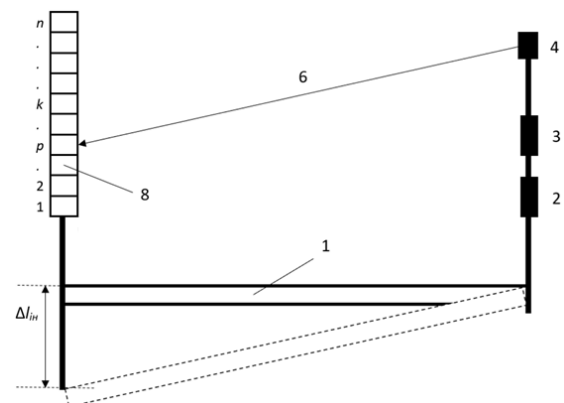


Рис. 5. Відхилення рейкової колії по вертикалі вверх

Таким чином величина зміщення по вертикалі відповідної колії у низ на одному періоді обертання лазерного променя буде обчислюватися, як

$$\Delta l_{\text{вн}} = (p - k) \cdot d, \quad (1)$$

а величина зміщення по вертикалі у верх (рис. 6), як

$$\Delta l_{\text{вв}} = (j - k) \cdot d, \quad (2)$$

де Δl_{iH} та Δl_{iB} – величина зміщення по вертикалі відповідної колії фотоприймачем; k – номер елемента фотоприймача без зміщення по вертикалі; p та j – номер елемента фотоприймача після зміщення по вертикалі; d – довжина елемента фотоприймача.

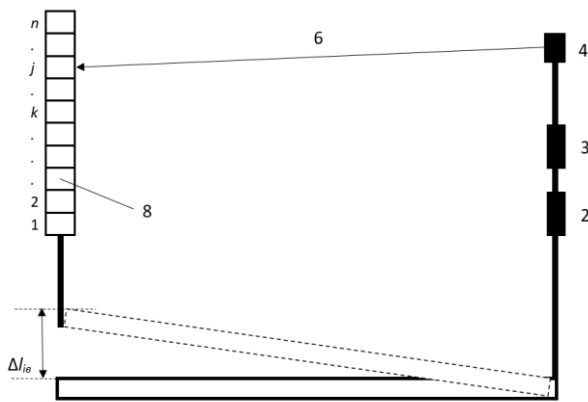


Рис. 6. Відхилення рейкової колії по вертикалі вниз

Знак в арифметичних виразах після обчислення вказує, відповідно, на характер зміщення по вертикалі, тобто на прогин чи опуклість рейкової колії.

Лазерний промінь 6, що розгортається, з вузла розгортки 4 послідовно пробігає по чутливій поверхні кожного 7, 8, 10 фотоприймача певної довжини h (рис. 3). Електричний сигнал з фотодіода або фотоелектричний перетворювач (ФЕП) поступає на підсилювач електричного сигналу фотоприймача, який, після підсилення, поступає на блок 16 контролю відхилення положення рейкових колій (рис. 3). Тривалість електричного імпульсу з виходу фотоприймача буде визначатися швидкістю проходження лазерного променя по чутливій поверхні фотоприймача, яка, у свою чергу, буде визначатися кутовою швидкістю ω двигуна вузла розгортки і відстанню R вузла розгортки до фотоприймача [4]. Таким чином, тривалість імпульсів τ_1 з фотоприймача 7 та τ_2 і τ_3 – з фотоприймачів 8 і 10 буде різною і в міру наближення фотоприймача до вузла розгортки буде збільшуватись (рис. 7, а).

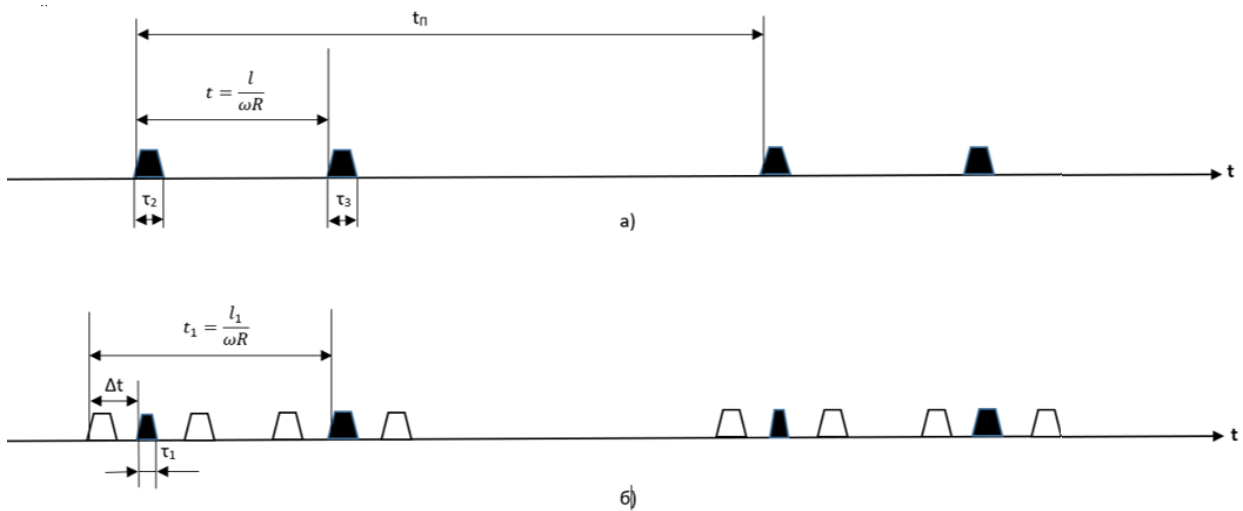


Рис. 7. Часові діаграми, що характеризують відхилення положення рейкової колії по горизонталі

Крім того чутлива поверхня фотоприймача 7 менша аніж фотоприймачів 8 і 10.

$$\tau_1 = \frac{h_1}{\omega R_1}, \quad (3)$$

де $\omega = 2 \cdot \pi \cdot f$ – кутова швидкість обертів двигуна; h_1 – чутлива поверхня фотоприймача 7; R_1 – відстань між блоком розгортки лазерного променя та першим фотоприймачем. Тоді для фотоприймача 8 тривалість імпульсу буде

$$\tau_2 = \frac{h_2}{\omega R_2}, \quad (4)$$

а для фотоприймача 10 тривалість імпульсу буде (рис. 7):

$$\tau_3 = \frac{h_3}{\omega R_3}, \quad (5)$$

Слід зазначити, що відстань (пауза) між електричними імпульсами теж буде збільшуватися в міру

віддалення фотоприймачів до вузла розгортки.

Відстань між фотоприймачами 8 і 10 становить стандартну ширину l колії, тому час, за який лазерний промінь переміститься від фотоприймача 8 до фотоприймача 10 становитиме (рис. 7)

$$t = \frac{l}{\omega R}. \quad (6)$$

Якщо стандартна ширина l колії буде порушена, тобто буде переміщення шляховимірювального візка 1 і, відповідно, другого 8 від першого 7 фотоприймачів по горизонталі, то час, за який лазерний промінь переміститься від переміщеного другого фотоприймача 8 до третього фотоприймача 10, якщо б він залишався на місті, було б (рис. 7, б):

$$t_1 = \frac{l_1}{\omega R}, \quad (7)$$

а час, за який лазерний промінь пройде між переміщеним другим фотоприймачем 8 і першим фотоприймачем 7 становитиме

$$\Delta t = t - t_1, \quad (8) \quad \text{звідки}$$

$$\text{або} \quad \Delta l = l - l_1, \quad (9)$$

$$\Delta l = \frac{l \cdot t_1 - l t}{t} = \frac{l \cdot (t_1 - t)}{t}, \quad (13)$$

де Δl – величина переміщення другого фотоприймача 8 відносно першого фотоприймача 7.

Знак в арифметичному виразі після обчислення вказує, відповідно, на характер зміщення по горизонталі – вправо, або вліво переміщена рейкова нитка колії. Тоді

$$\frac{t}{t_1} = \frac{l}{l_1}, \quad (10)$$

$$\text{звідки} \quad l_1 = \frac{l \cdot t_1}{t}. \quad (11)$$

З урахуванням виразу (9) вираз (11) можна записати у вигляді

$$l + \Delta l = \frac{l \cdot t_1}{t}, \quad (12)$$

але $\Delta t = t_1 - t$ (рис. 7, б).

Тоді вираз (13) можна переписати у вигляді

$$\Delta l = \frac{l \cdot \Delta t}{t}. \quad (14)$$

Висновки

Розроблений пристрій належить до рейкового транспорту, а саме до засобів контролю стану рейкового шляху в умовах експлуатації, і може застосовуватися на наземному залізничному транспорті, метрополітені та на трамвайних коліях міста. Він призначений для виявлення і вимірювання зовнішніх дефектів або пошкоджень на рейках, що може допомогти у забезпеченні безпеки та ефективності руху транспорту.

СПИСОК ЛІТЕРАТУРИ

1. D. M. Kurhan, "Modeling of development vertical deformation of railway track," *Nauka ta progres transportu*, No. 1(61), pp. 100–108, Feb. 2016, doi: <https://doi.org/10.15802/stp2016/61003>
2. Курган Д. М. До вирішення задач розрахунку колії на міцність із урахуванням нерівнопружності підрейкової основи / Д. М. Курган // *Наука та прогрес транспорту*. – 2015. – № 1 (55). – С. 90–99. URL: <https://stp.ust.edu.ua/article/view/38250/34809>
3. Курган М. Б. Дослідження нерівностей колії в межах залізничних переїздів / М. Б. Курган, Д. М. Курган, О. Ф. Лужицький // *Наука та прогрес транспорту*. – 2015. – № 5 (59). – С. 84–96. URL: <https://stp.ust.edu.ua/article/view/55341/52850>
4. Levterov A. I., Storchak M. O. Measuring system to monitor deformation of large size structure members // *Вісник Харківського національного автомобільно-дорожнього університету*. № 83. 2018. С. 66–73, doi: <https://doi.org/10.30977/BUL.2219-5548.2018.83.0.66>
5. Z. Xiong, Q. Li, Q. Mao, and Q. Zou, "A 3D Laser Profiling System for Rail Surface Defect Detection," *Sensors*, vol. 17, No. 8, p. 1791, Aug. 2017, doi: <https://doi.org/10.3390/s17081791>
6. L. Chen, Y. Li, Z. Ma, H. Liu, and W. Mao, "Vision-Based Position Deviation Measurement of Rail Corrugation Chord Measuring Points Under Bi-Linear Laser Assistance," *IEEE access*, vol. 9, pp. 36207–36217, Jan. 2021, doi: <https://doi.org/10.1109/access.2021.3062663>
7. Z. Ma, Y. Dong, H. Liu, X. Shao and C. Wang, "Forecast of non-equal interval track irregularity based on improved grey model and PSO-SVM", *IEEE Access*, vol. 6, pp. 34812-34818, 2018. URL: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8371594>
8. Dingfei Jin, Hua Ma, "A Fast Global Optimal Strategy for Iteration Closest Point Using 2D-BnB and Its Application to Rail Profile Registration", *Journal of Sensors*, vol. 2023, 11 p., 2023. doi: <https://doi.org/10.1155/2023/6353247>
9. Пат. КМ 135379 Україна: МПК G01B11/16 (2006.01). Левтеров А. І. Пристрій для вимірювання деформацій елементів конструкцій великих розмірів. Заявка у 2019 01026 31.01.2019. Опубл. 25.06.2019, бюл. «Промислова власність», № 12. Книга 1. С. 4.73–4.74. URL: <https://ukrpatent.org/uk/articles/bulletin-ip>
10. В. М. Балабан, Є. П. Тимофеев, А. С. Литвиненко, О. М. Ляшенко, «Забезпечення дозиметричного контролю лазерного випромінювання», *Український метрологічний журнал*, № 2, С. 27–32, 2017. URL: <http://umj.metrology.kharkov.ua/issue/view/6574>

Received (Надійшла) 26.11.2024

Accepted for publication (Прийнята до друку) 12.02.2025

A movable device for monitoring deviations in the position of the rail tracks

M. Kostikova, B. Karpishen, S. Neronov, G. Pliekhova, S. Ocheretenko

Abstract. Problem. The running rails together with sleepers and the contact rail in the subway or the contact wire on the railway and tram tracks are a single complex that creates the necessary conditions for the movement of rail electric transport. The operation of these facilities requires constant monitoring of the technical condition of rail structures. The relative position of the height of the running rails, the curvature of the latter, and the width of the rail gauge are subject to unauthorised changes. **Objective.** The paper analyses devices for controlling deviations in the position of rail tracks, which contain a light emitter with a position controller and a platform moving along the beam. The disadvantage of these devices is insufficient functionality on curved sections. **Operation.** The task was set to develop a device for measuring the parameters of railways with increased accuracy and enhanced functionality. **Results.** A movable device for monitoring deviations in the position of rail tracks using an optical laser system and a group of photodetectors was developed and described. The developed device belongs to rail transport, namely, to the means of monitoring the state of the rail track in operation, and can be used on land railway transport, subway and tramways of the city.

Keywords: rail tracks, deployable optical system, laser, collimator, photodetector, rail truck.

Д. С. Положий, О. О. Орехов

Національний аерокосмічний університет «ХАІ», Харків, Україна

ПРОГНОЗУВАННЯ РИЗИКІВ ДОРОЖНЬО-ТРАНСПОРТНОГО РУХУ НА ОСНОВІ НЕЙРОННОЇ МЕРЕЖІ

Анотація. На підставі структурного аналізу статистики ДТП в Україні виділені найбільш небезпечні їх причини: перевищення безпечної швидкості, порушення правил маневрування, порушення правил проїзду перехресть та пішохідних переходів. Розглянуто запропоновані моделі прогнозування безпеки дорожньо-транспортного руху і ризику водіння. Досліджено та систематизовано фактори безпеки дорожньо-транспортного руху. Запропонована модель прогнозування ризику дорожньо-транспортного руху із використанням інтелектуальної транспортної системи (ITS) та нейронної мережі. Модель побудована на технології обробки нейронною мережею зважених статистичних та динамічних масивів вхідних даних, що характеризують внутрішнє і зовнішнє середовище транспортного засобу з метою отримання оцінки ризику дорожньо-транспортного руху. Запропоновано алгоритм прогнозування ризику нейронною мережею. Навчання багатопарової нейронної мережі здійснюється за алгоритмом зворотного поширення помилки. Тестування роботи системи прогнозування продемонструвало точність передбачень 85-90%.

Ключові слова: безпека дорожньо-транспортного руху, фактори безпеки, моделювання, транспортний засіб, нейронна мережа, інтелектуальна транспортна система.

Вступ та опис проблеми

Стійкою тенденцією сучасності є збільшення інтенсивності дорожнього руху, а внаслідок цього і зростання кількості ДТП та жертв і травмованих на дорогах людей. Аналіз статистики патрульної поліції України [1] свідчить про те, що в 2023 році на дорогах України в середньому щодоби гинуло понад 8 осіб і було травмовано понад 80 осіб. У довоєнні роки (2020-2021) ці показники були ще вище. Тому проблема забезпечення безпеки дорожнього руху залишається актуальною і потребує втілення сучасних технологічних рішень. Концепція безпеки дорожнього руху на даний час ґрунтується на розпізнаванні об'єктів перешкоди транспортному руху, а саме транспортних засобів, пішоходів, тварин, інших рухомих і

нерухомих об'єктів. В той же час проведений аналіз статистики причин дорожньо-транспортних пригод по Україні за 2020-2023 рр. вказує на те, що структура причин ДТП щороку майже повторюється, спостерігається також їх часова залежність продовж доби, а найбільш небезпечними причинами виникнення аварій є наведені у табл. 1.

Найбільша кількість ДТП (39,1%) за даними табл. 1 трапляється через перевищення транспортними засобами безпечної швидкості – саме такі аварії призвели до найбільшої кількості загиблих (51,6% від загальної кількості загиблих). На другому місці серед причин ДТП – порушення правил маневрування, що привели до зіткнень. Значна кількість аварій траплялась також через порушення при проїздах перехресть та пішохідних переходів.

Таблиця 1 – Структура ДТП з загиблими та травмованими за причинами за період 2020-2023 рр.

Причини ДТП	Частка у загальній кількості, %		
	ДТП	Осіб, які загинули	Травмованих осіб
Перевищення безпечної швидкості	39,1	51,6	38,7
Порушення правил маневрування	21,1	14,1	21,1
Порушення правил проїзду перехресть	8,5	2,5	9,5
Порушення правил проїзду пішохідних переходів	7,3	3,6	6,0
Недотримання дистанції	5,2	2,5	5,3
Інше	18,8	25,7	19,4

Джерело: складено і розраховано автором за даними [1].

Але будь яке ДТП, як правило, є наслідком збігу декількох причин. Тому перед сучасними системами запобігання аваріям стоїть завдання отримання і обробки великих обсягів різномірної і мінливої інформації, яку надають учасники дорожнього руху, дорожня інфраструктура, метеопрогнози, і виокремлювати найважливішу для безпеки інформацію та виконувати її класифікацію для прогнозування дорожньо-транспортних пригод на різних ділянках дороги. Вирішення таких завдань можливо за допо-

могою сучасних технологій високошвидкісної передачі даних та використання нейронних мереж для їх обробки у хмарі.

Аналіз останніх досліджень і публікацій. Спектр наукової думки що до проблем безпеки дорожнього руху та прогнозування ризиків водіння на даний час є достатньо широким і різноманітним.

Сучасну концепцію інтелектуальної системи автомобільної безпеки на основі хмарної архітектури проаналізували Д. С. Положий та О.О. Орехов

[2]. Їми доведено необхідність розподіленої та динамічно масштабованої архітектури мережевої програми, щоб задовольнити зростаючі вимоги до розширюваної системи додатків IoT (Інтернету речей), до категорії яких можна віднести і прилади сучасних транспортних засобів для збору інформації або виконання команд управління під дистанційним керуванням.

А.А. Кашкановим і О.В. Пальчевським [3] проведено аналіз методів прогнозування трафіку та дорожньої безпеки ITS, за результатами якого вони прийшли до висновку, що найбільш перспективними моделями для швидкого та ефективного вирішення неординарних ситуацій дорожнього руху є моделі із застосуванням штучного інтелекту або комбінації із ними, засновані на алгоритмах глибокого навчання.

Зарубіжними авторами запропоновані різні методи прогнозування ризику водіння та зіткнень, які поділяють на два класи:

- засновані на математичних моделях;
- засновані на навчанні на основі попереднього досвіду [4].

Методи, що засновані на математичних моделях, спираються на апріорні знання про транспортний засіб та інфраструктуру дорожнього руху, представлені у формалізованому вигляді їх динаміки або імовірності подій. Найбільш розповсюдженими є моделі інтелектуального драйвера, моделі на основі локально зваженої регресії та моделі на основі звичайних нейронних мереж.

Найпоширенішою моделлю динаміки транспортного засобу є модель часу до зіткнення при умові, що транспортний засіб продовжує поточний шлях з поточною швидкістю [5]. Інші моделі включають інтегрований час до зіткнення, час експозиції до зіткнення, час після вторгнення і швидкість уповільнення для уникнення зіткнення. Однак більшість моделей безпеки дорожнього руху базуються лише на аналізі швидкості автомобіля [6, 7].

Nyothiri Aung із співавторами [8] запропонували систему прогнозування аварій для автомобільних спеціальних мереж (VANET) у міських умовах, в якій ризик розглядається як прихована змінна, отримана за допомогою спостережень за іншими факторами безпеки, такими як швидкість руху, погодні умови, місце дії, інтенсивність руху транспортних засобів поблизу та втома водія. Для моделювання кореляції між цими спостереженнями та латентною змінною ризику використана прихована модель Маркова. За свідченням авторів, запропонована система має кращі характеристики з точки зору чутливості та точності порівняно з моделями, заснованими на формалізації динаміки одного фактора. Недоліком цієї моделі є те, що вона не враховує небезпеки від транспортних засобів, які рухаються поруч і часто стають причинами аварій.

В публікаціях останніх років великої популярності для прогнозування аварій на дорозі набули також підходи, які включають різноманітні алгоритми машинного навчання. Найпопулярнішим є використання технології глибокого навчання, яка надає можливість отримувати від системи рішення у

неординарних обставинах. Наприклад, Стрикленд та співавтори [9] реалізували модель глибокого навчання з байєсівською згортковою довготривалою пам'яттю (LSTM).

А. Iqbal із співавторами [10] запропонували систему виявлення небезпеки можливих зіткнень з транспортними засобами, що рухаються поруч, шляхом надання попереджувальної інформації. Для об'єднання даних кількох датчиків в системі була використана динамічна байєсівська мережа (DBN). Характеристики руху зустрічних транспортних засобів система збирала за допомогою бортової камери та інерціального вимірювального блоку (IMU), яким фіксувались швидкість, прискорення та напрямки руху транспортного засобу. Зібрані дані подавались в класифікатор на основі градієнтних ознак.

Чен та співавтори [11] використали нейронну мережу для обробки інфраструктурної інформації, отриманої за допомогою зв'язку V2I (транспортний засіб – інфраструктура) та інформації, яку отримано за допомогою зв'язку V2V (транспортний засіб – транспортний засіб) для запобігання зіткненню.

Halim Z. із співавторами [12] представили систему на основі глибокого навчання та штучного інтелекту для виявлення ризиків водіння для легкових транспортних засобів, яка генерує ранні попередження перед очікуваною аварією. Система збирає та вивчає різні шаблони стилю водіння водія на симуляторі, а потім застосовує глибоке навчання для віднесення поведінки водія до одного з попередньо визначених класів.

S. Orgea із співавторами [13] наводять огляд та класифікацію методів глибокого навчання для прогнозування ДТП. Інтелектуальні системи на основі моделі машинного навчання для уникнення ДТП автори пропонують будувати з використанням алгоритмів штучної нейронної мережі (ANN), логістичної регресії (LR), дерева рішень (DT), K-найближчого сусіда (KNN), опорної векторної машини (SVM), байєсівської системи (NB), стохастичного градієнтного спуску (SGD) та ін. [14]

Оскільки значна кількість інформації про навколишнє середовище транспортного засобу отримується через відеодатчики, запропоновано багато систем запобігання зіткнень, які побудовані на обробці відео спостережень за допомогою нейронних мереж.

Так В. Maaloul розглядає алгоритми виявлення ДТП на основі відео інформації [15]. Robles-Serrano [16] пропонує автоматичне виявлення ДТП за допомогою відео спостереження з використанням методів глибокого навчання.

R. Coll-Josifov із співавторами [17] пропонують систему, в якій нейронні мережі YOLO навчаються на наборі даних з метою виявлення та класифікації предметів із картинок вулиці, яку видно зверху з безпілотної. Навчені моделі використовуються для виявлення та класифікації об'єктів у міському сценарії дорожнього руху в режимі реального часу. Нарешті, пропонується алгоритм відстеження об'єктів, прогнозування їх майбутніх траєкторій та потенційних зіткнень.

Отже, методи засновані на навчанні використовують параметри, які мінімізують функцію небезпеки, що оцінюється на наборі навчальних даних. Значною мірою вони залежать від якості та обсягу таких даних, що дає можливість перенавчання при зміні умов. Недоліком таких систем є те, що у разі недостатньо повного масиву навчальних даних, такі системи зазвичай сходяться до локальних оптимумів, що може привести до прийняття помилкового рішення. Або такі системи навчають оперувати тільки декількома факторами безпеки, які є найважливішими на думку розробників, і не оцінюють всього спектру важливих факторів, що впливають на безпеку дорожнього руху. Внаслідок цього система може не спрацювати у неординарній ситуації.

Постановка завдання. Метою статті є дослідження факторів небезпеки дорожньо-транспортного руху та пропозиція рішень по системі прогнозування ризиків на дорожній дільниці з використанням інфраструктури ITS на основі технології, заснованій на нейронній мережі.

Виклад основного матеріалу

Систематизація факторів небезпеки. Системи, побудовані на технологіях штучного інтелекту, швидко поширюються в тих секторах, де потрібно виявляти як закономірності у великих обсягах даних, так і відхилення від таких закономірностей, і на основі результатів обробки даних прогнозувати ситуацію та приймати рішення.

Саме до такого сектору з великими обсягами різноманітних динамічних даних відноситься система безпеки дорожнього руху, яка відстежує інформацію з великої кількості транспортних засобів, компонентів інфраструктури дороги, враховує інформацію про погодні та часові умови, рухомі та нерухомі перешкоди дорожньому руху (пішоходи, травини тощо).

Підхід до забезпечення безпеки такої системи полягає у визначенні загального ризику від сукупності можливих перешкод від всіх елементів системи і впровадження заходів що до мінімізації таких ризиків з урахуванням часу їх можливого настання.

Сучасні технології високошвидкісної передачі інформації 5G та хмарні технології зберігання та обробки великих даних дозволяють проводити обробку і передачу інформації із великої кількості датчиків одночасно з незначними затримками, тому система забезпечення безпеки на дорозі технологічно в спроможності враховувати інформацію про значну кількість факторів, як високо динамічних, так і тих, що змінюються повільно або періодично.

З системної точки зору основними підсистемами джерел небезпеки в дорожньому русі є: водій, транспортний засіб, трафік на дільниці дороги, дорожнє полотно та навколишнє середовище (рис. 1).

До атрибутів навколишнього середовища відносять: погодні умови, просторово-часові умови, можливість появи тварин на дорозі. Несприятливі погодні умови – дощ, туман, ожеледь і вітер – впливають на видимість і, як наслідок, на безпеку дорожнього руху.

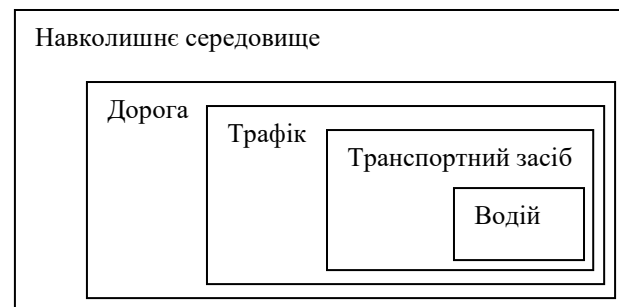


Рис. 1. Основні підсистеми джерел небезпеки
Джерело: здобуток автора

Час доби також впливає як на видимість, освітленість, так і на втому водія та інтенсивність трафіку на дорозі. Тварини також можуть становити небезпеку на ділянках проїжджій частині, якщо маршрути їх міграції пролягають в прилеглих районах. Тож мають бути вжиті відповідні заходи, такі як огорожі чи попереджувальні знаки, а в ITS повинна зберігатися відповідна інформація.

До атрибутів проїжджій частини відносяться як тип та стан дорожнього покриття на дорожній дільниці, дорожні смуги, так і геометрія дороги, розв'язки, перехрестя, пішохідні переходи. Крім цього до даної підсистеми відноситься інформація про перешкоди, аварійні дільниці, огорожені дільниці, слизьке дорожнє покриття в зимових умовах, що за своєю природою є динамічною. Атрибути проїжджій частини можуть збиратися як спеціалізованими компаніями технічного обслуговування доріг, так і учасниками дорожньо-транспортного руху з датчиків вібрації автомобіля та з камери, що надає дані та зображення тротуарів, смуг та перешкод.

Атрибути трафіку на певній ділянці дороги включають: обсяг трафіку, швидкість транспортних засобів і події, такі як випадкові аварії, аварійні маневри тощо. Збір таких даних є першочерговим питанням для моделей прогнозування ризиків небезпеки, і може здійснюватися за допомогою транспортних засобів, смартфонів та інших пристроїв, щоб досягти повного покриття мережі.

Обсяг трафіку є наріжним камнем аналізу безпеки, але його часто бракує в моделях прогнозування ризику. Якщо вихідні дані про обсяг трафіку походять із розміру парку транспортних засобів, то їм бракує репрезентативності. Достовірні оцінки обсягу трафіку на всій мережі доріг можуть бути зібрані за допомогою ITS з наземних приладів оцінки трафіку з передачею даних на сервер.

До цієї категорії відноситься також інформація про транспортні засоби навколо, небезпеку їх маневрів та манери водіння.

Швидкість руху – один із найважливіших факторів безпеки на дорозі, як показує статистика ДТП (табл. 1). Дані про швидкість мають бути пріоритетним пунктом у списку даних, пов'язаних з безпекою. Звичайне надання даних про швидкість усіма виробниками автомобілів на сервер ITS могло б пом'якшити помилки, виявлені в телематичних даних навігації придорожніми приладами. Технологія GPS за певних обставин також може посилити збір

даних про швидкість руху, але ці дані недостатньо точні в міських умовах і недоступні в тунелях.

Атрибути транспортного засобу можна поділити на такі, що характеризують його технічний стан та динамічні характеристики (швидкість, маневри, відстань до перешкоди, наявність перешкоди). Сучасні автомобілі оснащені системою ADAS, яка за допомогою сенсорів, що під'єднанні до Інтернету речей, допомагає уникати зіткнень та ДТП, попереджає водіїв про потенційні проблеми, здійснює запобіжні заходи, а в деяких випадках перебирає на себе контроль над автомобілем. Інформація з системи ADAS може передаватися до серверу ITS для оцінки ризику водіння на ділянці дороги.

До атрибутів водія, які впливають на безпеку дорожньо-транспортного руху, відносяться: стан здоров'я водія, втома, манера водіння, присутність алкоголю або наркотиків в організмі. Такі показники може збирати та передавати до захищеного сервера бортова система контролю стану водія [18] або оснащена відповідними сенсорами сучасна автомобільна система ADAS.

Таким чином, із розвитком дорожньої інфраструктури та розвитком автомобільних систем ADAS кількість атрибутів дорожньо-транспортного руху, які можуть бути передані до серверу ITS, зростає, і це може збільшити прогностичну силу моделей прогнозування ризику з одночасним їх ускладненням.

Визначення важливості окремих вхідних змінних для прогностичної сили моделі є складною проблемою. Рішення залежатиме передусім від достатніх історичних даних. Доцільно обрати ті атрибути, які мали великий вплив на виникнення ДТП в минулому.

Тоді можна передбачити ймовірність ризиків подібних подій у найближчому майбутньому.

Для моделювання відбираємо найбільш значущі атрибути, наведені у табл. 2.

Таблиця 2 – Перелік вхідних змінних

Підсистема	Вхідні змінні, X_i
Водій	Стан водія
	Манера водіння
Транспортний засіб	Швидкість
	Технічний стан
Трафік	Інтенсивність руху
	Наближення перешкод
	Небезпечне авто поблизу
Дорога	Тип дорожнього покриття
	Стан покриття
	Слізка дорога
	Складність геометрії
Зовнішнє середовище	Видимість
	Погодні умови
	День тижня, час доби

Джерело: здобуток автора

Фактичні значення атрибутів, отримані за допомогою сенсорів або із баз даних, для цілей прогнозування ДТП потрібно нормалізувати відносно мінімуму або відносно середнього.

Треба зазначити, що кожен з наведених у табл. 2 атрибутів, як правило, є комплексним і складається із групи ознак. Наприклад, стан водія складається із показників його фізичного стану (тиск, пульс, температура), наявності алкоголю або наркотиків в організмі, сонливості, втоми та ін. Отже, оцінку стану водія виконує окрема підсистема по п'ятибальній шкалі.

Також виконується комплексна оцінка і інші вхідні змінні.

Формалізація задачі. Стан безпеки дорожнього руху на ділянці дороги постійно змінюється і для його оцінки в системі запропоновано використовувати класифікатор за рівнем ризику, наведений у табл. 3.

Таблиця 3 – Класифікатор ризиків безпеки

Рівень ризику	Значення ризику	Категорія ризику
Катастрофічний	0,511÷0,89	3
Критичний	0,141÷0,51	2
Граничний	0,08÷0,14	1
Незначний	0,01÷0,08	0

Джерело: складено автором на основі даних з [19]

Оскільки ризик має імовірнісну природу, то може бути оцінений як математичне очікування збитку від усіх факторів небезпеки за формулою:

$$R(t) = \sum_{i=1}^n W_i P_i(t), \quad (1)$$

де $R(t)$ – ризик небезпеки системи;

$P_i(t)$ – імовірність збитку від i -го фактору;

W_i – вагова функція i -го фактору, за допомогою якої наслідки різних факторів небезпеки приводяться до єдиної шкали;

n – кількість задіяних в системі факторів ризику.

Функціонування нейрона визначається співвідношенням:

$$s = \sum_{i=1}^n w_i x_i, \quad (2)$$

$$y = f(s, b),$$

де x_i – вхідні сигнали;

w_i – вагові коефіцієнти;

s – зважена сума вхідних сигналів;

b – поріг рівень нейрона;

f – функція активації.

Порівнюючи формули (1) і (2), можна прийти до висновку, що для завдання визначення ризику системи безпеки може бути використана нейронна мережа, яка складається з декількох шарів: вхідного, внутрішніх (прихованих) та вихідного.

Вхідний шар реалізує зв'язок з вхідними даними, вихідний – з вихідними. На кожному шарі міститься декілька нейронів, з'єднаних між собою вагами.

На вхід якої подається набір вхідних змінних, що є оцінками вагомих факторів безпеки на ділянці дороги:

$$X = (X_1, X_2, \dots, X_n). \quad (3)$$

У якості вхідних змінних використано обсяг різнорідних даних: транспортно-технічних, метеорологічних та ін., що наведені у табл. 2. Мета полягає в тому, щоб передбачити єдину змінну, якою є ймовірність ризику аварії, використовуючи інструмент багатшарової нейронної мережі.

На виході нейронної мережі маємо одну вихідну змінну, яка може приймати одне із чотирьох значень, наведених у табл. 3, які є категоріями ймовірності ризику ДТП на даній ділянці дороги для даного транспортного засобу:

$$R = (0, 1, 2, 3). \quad (4)$$

Ці категоріальні змінні не можна безпосередньо вставити в матрицю плану, тому що нейронна мережа передбачає фіктивні змінні:

$$Y = (Y_1, Y_2, Y_3, Y_4) = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}. \quad (5)$$

Для набору значень вхідних змінних за допомогою нейронної мережі можна, наприклад, отримати наступний вихідний вектор, що представляє ймовірності настання конкретних ризикових станів:

$$Y = (0,15; 0,58; 0,35; 0,74)$$

Після перекодування найбільшого значення у векторі Y приймає значення 1, а решти – 0, в результаті отримуємо вектор ймовірності категорії ризику:

$$YD = (0, 0, 0, 1).$$

Отриманий результат показує, що для даного набору вхідних змінних ризик ДТП є «катастрофічним», отже потрібні негайні дії для його усунення.

Оскільки проект має дослідницьку мету, то використовувалася багатшарова нейронна мережа з функцією активації.

Проектування нейромережі. За теоремою Колмогорова функція n змінних може бути подана як суперпозиція одновимірних функцій у кількості $2n+1$. Отже, кількість прихованих елементів для даної системи можна обрати $14 \times 2 + 1 = 29$. Але не можна просто вибрати теоретичний максимум кількості вагомих коефіцієнтів, бо в цьому разі мережа навчиться мати справу тільки з тими даними, що надавалися в процесі тренування, і в результаті узагальнююча здатність мережі буде слабкою. Тому розмір прихованих шарів вибираємо емпірично з урахуванням особливостей розв'язуваної задачі, оцінюючи точність прогнозу моделі з різною структурою. Загальна структура двошарової нейронної мережі наведена на рис. 1.

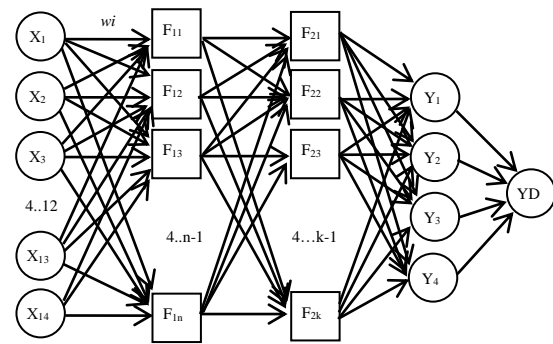


Рис. 1. Структура нейронної мережі

Вхідний шар вважається нульовим. Кожен нейрон наступних шарів видає на вихід сигнал y , перетворюючи зважену суму виходів нейронів попереднього шару.

Завдання навчання багатшарових штучних нейронних мереж вирішує алгоритм зворотного поширення помилки. Обираємо для нього гладку функцію активації:

$$y = \frac{1}{1 + e^{-s}}. \quad (6)$$

Ця функція зручна тим, що її похідна виражається через саму функцію:

$$y' = y(1 - y). \quad (7)$$

Кількість тренувальних прикладів має приблизно дорівнювати кількості ваг мережі, помноженій на зворотну величину помилки, тобто

$$w < n/e, \quad (8)$$

де w – кількість тренувальних прикладів;
 e – припустима помилка,

$$w < 14 / 0,03 = 467.$$

Отже, приймаємо кількість тренувальних прикладів приблизно 500.

В процесі навчання одноразова перехресна перевірка передбачає розбиття вибірки на взаємно доповнювані підвибірки з метою проведення аналізу на одній частині (навчальному наборі) та перевірки результатів на іншій частині (тестовому наборі). Для зниження дисперсії здійснюється багаторазова перехресна перевірка, а результати перевірок усереднюються.

Проектування системи. Система запрограмована на Python з використанням FastAPI як HTTP API серверу, та Tensorflow для побудови нейронної мережі.

Алгоритм прогнозування ризику дорожньо-транспортного руху включає наступні кроки:

- 1) Отримати файл з історичними даними про аварії;
- 2) Перетасувати дані (для зменшення ризику переналаштування нейронної мережі);
- 3) Завантажити вхідні змінні X_i (зчитування змінних на основі прогнозування ризику ДТП, змінні можна додавати та видаляти, модель адаптована);
- 4) Перекодувати вихідні змінні X_i в числа, а потім в дуальні матриці;

5) Розділити дані на навчальний і тестовий набір (тестовий набір даних становить 20% від загального набору даних);

6) Поділити набір навчальних даних на дані навчання та перевірки;

7) Задати первісні ваги нейронної мережі в інтервалі $(-1, 1)$;

8) Встановити максимальну кількість періодів навчання на 500 та виконати ініціалізацію нейронної мережі з коефіцієнтом навчання 0,01.

9) Якщо кількість епох навчання більше або дорівнює максимальній кількості епох, навчання припиняється, інакше для кожної епохи навчання виконати:

- вибір і подачу одного вхідного вектора нейрона з навчального набору;

- отримання реакції нейрона методом зворотного поширення,

- корекцію ваг нейронів на основі помилки;

- оцінку помилки класифікації в усьому навчальному наборі – якщо помилка менша за прийнятну межу, навчання завершується;

- оцінку успішності мережевої діяльності на тестовому наборі.

10) Виконати прогнозування вектора значень вихідної змінної (ризик) для конкретних значень вхідних змінних.

Архітектура системи прогнозування ризику складається з декількох компонентів, діаграма взаємодії яких наведена на рис. 2.

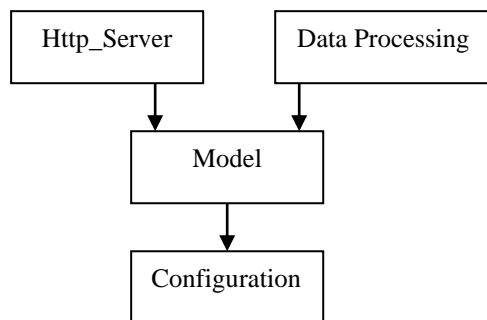


Рис. 2. Діаграма взаємодії компонентів

Загальна архітектура системи включає:

1) Компонент обробки даних Data Processing:

- dataset.csv та test_dataset.csv - набори даних для тренування та тестування моделі;

- make_dataset.com - скрипт для підготовки набору даних.

2) Компонент моделі Model:

- model.com - код для побудови та тренування нейронної мережі;

- model директорія, що містить файли, пов'язані з натренованою моделлю, включаючи параметри, збережені змінні та метадані моделі.

3) Компонент Http Server:

- http_api директорія, що містить код для реалізації HTTP API серверу;

- schemas.com - схеми даних для валідації вхідних та вихідних даних API;

- server.com - код, який обробляє HTTP запити, отримує вхідні дані та надає прогноз рівня небезпеки на дорозі.

4) Конфігураційний компонент Configuration:

- config.com - налаштування проекту, такі як параметри підключення до бази даних та налаштування HTTP сервера;

- gunicorn.conf.com - файл конфігурації для Gunicorn, що використовується для розгортання та обробки HTTP API.

5) Інші компоненти:

- check.com - скрипт для перевірки наявності необхідних залежностей та конфігурації перед запуском;

- requirements.txt - перелік залежностей, необхідних для розгортання та роботи проекту.

Результати тестування. Було оброблено кілька джерел ДТП, метеорологічних та інших пов'язаних даних. Попередньо дані були відсортовані, оброблені належним чином, а потім класифіковані. Всього було використано 14 показників, наведених у табл. 2.

В результаті аналізу отримано оптимальні параметри нейронної мережі, яка демонструє найвищу точність прогнозування: багатощарова нейронна мережа, що містить два прихованих шари, у першому прихованому шарі 80 нейронів, у другому прихованому шарі 40 нейронів. Емпіричне тестування виявило, що точність прогнозування зростала із зростанням кількості нейронів у кожному прихованому шарі. Кількість нейронів 80/40 демонструє точність 85-90%. З більшою кількістю нейронів точність прогнозу не зростає, але значно ускладнюються обчислення алгоритму.

Точність результату нейронної мережі для класифікації та прогнозування ризику дорожньо-транспортного руху склала 85-90%. Її порівнювали з точністю прогнозування за допомогою класифікації з використанням дерева рішень.

Використання дерев рішень показує набагато точність 70-80%. Точність у цьому випадку залежить від кількості дерев рішень і глибини кожного дерева рішень. Дерево випадкових рішень складається з кількох дерев рішень, і його кінцева точність є середнім значенням точності кожного дерева. Можна стверджувати, що із збільшенням кількості дерев рішень і зростанням глибини дерева рішень точність кінцевого прогнозу зростає.

Висновки

У роботі розглянуто можливості прогнозування та класифікації відносного ризику дорожньо-транспортних рухів з реальними вхідними даними, що збираються в ITS, за допомогою нейронних мереж.

Систематизовано фактори безпеки дорожньо-транспортного руху та зроблено вибір значущих факторів для вхідних даних нейронної мережі системи прогнозування ризику.

Розроблено алгоритм прогнозування ризику дорожньо-транспортного руху нейронною мережею та архітектуру системи.

Система запрограмована на Python з використанням FastAPI, як HTTP API серверу, та Tensorflow для побудови нейронної мережі.

Точність передбачення нейронної мережі порівнювали з точністю прогнозів дерева рішень, і нейронні мережі показали найвищу точність.

Отримані результати можуть мати практичне застосування для попередження водіїв про високий рівень небезпеки. Перспективами подальших досліджень є розширення кількості вхідних змінних нейронної мережі, проведення експериментів з різноманітною вхідною інформацією.

СПИСОК ЛІТЕРАТУРИ

1. Офіційний сайт патрульної поліції України. Статистика. URL: <https://patrolpolice.gov.ua/statystyka/>
2. Положий Д. С., Орехов О. О. Інтелектуальні системи автомобільної безпеки на основі хмарних архітектур // Системи управління, навігації та зв'язку, 2023. № 4. С. 91-95.
3. Кашканов А.А., Пальчевський О.В. Інтелектуалізація управління дорожнім рухом як засіб підвищення ефективності транспортної мережі міста в неординарних ситуаціях // DOI: <https://doi.org/10.31649/2413-4503-2022-16-2-42-50>
4. Anukruthi Karre, Akramul Azim. Collision Prediction of Smart Cities Using Co-simulation // Procedia Computer Science 224 (2023) 205-214. URL: www.sciencedirect.com
5. Hou J., List G. F., Guo X. 2014. New Algorithms for Computing the Time-to-Collision in Freeway Traffic Simulation Models URL: <https://doi.org/10.1155/2014/761047>
6. Sharma, B.; Katiyar, V.K.; Kumar, K. Kranti Kumar Traffic Accident Prediction Model Using Support Vector Machines with Gaussian Kernel. In Proceedings of Fifth International Conference on Soft Computing for Problem Solving; Pant, M., Deep, K., Bansal, J.C., Nagar, A., Das, K.N., Eds.; Springer: Singapore, 2016; pp. 1–10.
7. Wu, Q.; Hui, L.C.K.; Yeung, C.Y.; Chim, T.W. Early car collision prediction in VANET. In International Conference on Connected Vehicles and Expo (ICCVE); IEEE: Shenzhen, China, 2015; pp. 19–23.
8. Aung N., Zhang, W., Dhelim, S., Yibo Ai. Accident Prediction System Based on Hidden Markov Model for Vehicular Ad-Hoc Network in Urban Environments //MDPI, 2018, 9, 311. doi:10.3390/info9120311
9. Strickland, M., Fainekos, G., & Ben Amor, H. (2018). Deep predictive models for collision risk assessment in autonomous driving. In *2018 IEEE International Conference on Robotics and Automation, ICRA 2018* (pp. 4685-4692). Institute of Electrical and Electronics Engineers Inc.. URL: <https://doi.org/10.1109/ICRA.2018.8461160>
10. Iqbal, A., Busso, C., Gans, N.R. (2015) Adjacent Vehicle Collision Warning System using Image Sensor and Inertial Measurement Unit. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction—ICMI '15, Seattle, WA, USA, 9–13 November 2015; ACM Press: New York, NY, USA, 2015; pp. 291–298.
11. C. Chen, L. Liu, T. Qiu, Z. Ren, J. Hu, and F. Ti, "Driver's intention identification and risk evaluation at intersections in the Internet of vehicles," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1575–1587, Jun. 2018.
12. Halim Z., Sulaiman M., Waqas M. (2022). Deep neural network-based identification of driving risk utilizing driver dependent vehicle driving features: a scheme for critical infrastructure protection // Journal of Ambient Intelligence and Humanized Computing, 2022, 14(2):1-19. DOI:10.1007/s12652-022-03734-y
13. Oprea,S., Martinez-Gonzalez,P., Garcia-Garcia,A., Castro-Vargas,J.A., Orts-Escolano,S., Garcia-Rodriguez,J., Argyros,A. (2020). A review on deep learning techniques for video prediction / IEEE Transactions on Pattern Analysis and Machine Intelligence.
14. Kushwaha, M.; Abirami, M.S. (2023). Intelligent Model for Avoiding Road Accidents Using Artificial Neural Network, *International Journal of Computers Communications & Control*, 18(5), 5317, 2023. <https://doi.org/10.15837/ijccc.2023.5.5317>
15. Maaloul, B. (2018). Video-based algorithms for accident detections. Ph.D. thesis. Universit. Ede Valenciennes set du Hainaut-Cambresis; Universit. edeMons.
16. Robles-Serrano, S., Sanchez-Torres,G., Branch-Bedoya,J., 2021. Automatic detection of traffic accidents from video using deep learning techniques. Computers 10, 148.
17. Coll-Josifov, R., Masip-Alvarez, A., Lav`ernia-Ferer, D., 2022. Deep learning classification applied to traffic accident prediction, in: XLIII Jornadasde Autom.atica, Universidadeda Coru`na .Servizode Publicaci. ons. pp. 964–971.
18. Сергиенко Н.Е., Маренич А.Н. Бортова система контролю стану водія // Автомобільний транспорт, 2010. Вип. 27. С. 148-152.
19. Purkrábková, Z., Růžička, J., Bělinová, Z., Korec, V. Traffic accident risk classification using neural networks // Neural Network World, 2021, Issue 5, p.343

Received (Надійшла) 12.11.2024

Accepted for publication (Прийнята до друку) 19.02.2025

Prediction of traffic risks based on a neural network

Denys Polozhyi, Oleksandr Oriekhov

Abstract. On the basis of the structural analysis of traffic accident statistics in Ukraine, the most dangerous causes were identified: exceeding the safe speed, violation of maneuvering rules, violation of the rules for crossing intersections and pedestrian crossings. The proposed models for predicting the danger of road traffic and the risk of driving are considered. The factors of road traffic danger have been studied and systematized. A traffic risk prediction model is proposed using an intelligent transport system (ITS) and a neural network. The model is built on the technology of neural network processing of weighted statistical and dynamic arrays of input data characterizing the internal and external environment of the vehicle in order to obtain a traffic risk assessment. A neural network risk prediction algorithm is proposed. The learning of a multilayer neural network is carried out using the backpropagation algorithm. Testing of the forecasting system demonstrated the accuracy of predictions of 85-90%.

Keywords: traffic safety, risk factors, modeling, vehicle, neural network, intelligent transport system.

МОДЕЛЬ ПЛАНУВАННЯ МАРШРУТІВ РУХУ ОДНОРІДНОЇ ГРУПИ БЕЗПІЛОТНИХ ЛІТАЛЬНИХ АПАРАТІВ

Анотація. В статті визначено, що на теперішній час існує значна кількість моделей, методів, які дозволяють ефективно планувати маршрут руху БпЛА для виконання завдань, однак наступним етапом розвитку роботизованих технологій є колективне застосування різномірних систем для виконання спільних завдань. Одночасне прибуття у необхідне місце або місця – один з елементів виконання місії, що потребує спільного планування та координації між БпЛА у групі. Мета місії полягає в тому, що група БпЛА повинна досягти місця призначення одночасно, що досягається шляхом створення траєкторій руху однакової довжини, допускаючи, що всі БпЛА летять з однаковою швидкістю. В статті отримала подальший розвиток модель планування маршрутів руху однорідної групи безпілотних літальних апаратів, яка, на відміну від існуючих, враховує обмеження та вимоги дотримання безпеки руху БпЛА та забезпечує одночасне прибуття групи до кінцевої точки виконання місії.

Ключові слова: безпілотний літальний апарат; генетичні алгоритми; інформаційні технології; оптимізація траєкторії руху; планування маршрутів руху групи БпЛА; управління повітряним рухом; штучний інтелект.

Вступ

Постановка проблеми. Застосуванню безпілотних літальних апаратів (далі – БпЛА) у всіх сферах людської діяльності присвячено значну кількість публікацій та наукових досліджень. На теперішній час існує значна кількість моделей, методів, які дозволяють ефективно планувати маршрут руху БпЛА для виконання завдань, однак наступним етапом розвитку роботизованих технологій є колективне застосування різномірних систем для виконання спільних завдань. Одночасне прибуття у необхідне місце або місця – один з елементів виконання місії, що потребує спільного планування та координації між БпЛА у групі. Мета місії полягає в тому, що група БпЛА повинна досягти місця призначення одночасно, що досягається шляхом створення шляхів однакової довжини, допускаючи, що всі БпЛА летять з однаковою швидкістю. Для створення шляхів однакової довжини використовується аналогія з розтягуванням ланцюжка пружин, яка з'єднує точки маршруту.

Аналіз останніх досліджень і публікацій. Завдання побудови оптимальних (раціональних) маршрутів руху груп БпЛА для виконання спільних місії є предметом дослідження значної кількості вчених у галузі системного аналізу, математичного моделювання, оптимізації, штучного інтелекту.

Дослідження, що відображені у статті, ґрунтуються на роботах [1-4], в яких з позиції побудови генетичних алгоритмів вирішується завдання планування маршрутів, приводяться оригінальні методи, що залежать від різних зовнішніх факторів.

Слід також відзначити роботи [5-8], в яких розвиваються методи пошуку та моніторингу, як основного призначення БпЛА, що дозволяє зробити висновки про результативність планування маршрутів.

В роботах [9-12] розроблені оптимізаційні постановки та математичні моделі планування маршрутів польотів БпЛА, а також однорідної групи БпЛА при виконанні різноманітних завдань.

Результати аналізу схожих робіт свідчать про те, що дослідники менше звертають увагу на шви-

дкість обчислень при сучасному рівні розвитку апаратних засобів. В той же час можливість оперативно перепланувати маршрут руху в залежності від умов обстановки та впливу сукупності внутрішніх та зовнішніх факторів є важливою спроможністю групового застосування БпЛА.

Метою статті є удосконалення моделі планування руху однорідної групи БпЛА.

Виклад основного матеріалу

Розглянемо N БпЛА, які починають рух з точки старту у момент часу t_s та повинні досягти цільового пункту призначення у момент часу t_f , де $t_s < t_f$. Точки старту та точки цілі місії можуть бути з'єднані через набір маршрутних точок, як це наведено на рис. 1.

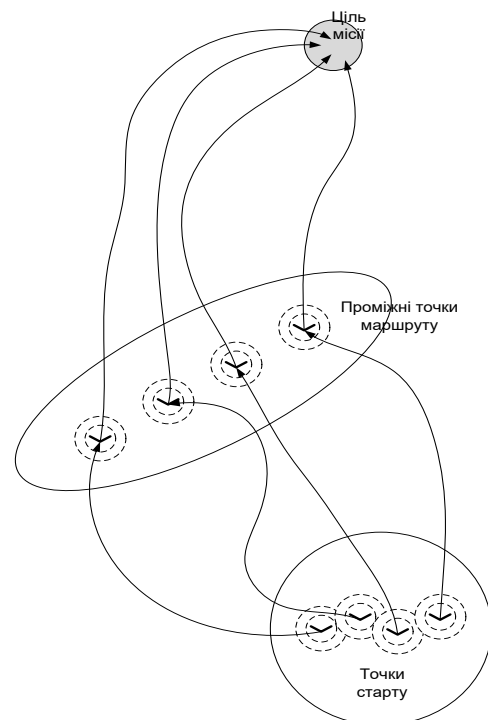


Рис. 1. Сфери безпеки та дальності руху БпЛА (джерело: розроблено автором)

Дана модель є спрощеною, оскільки навколишнє середовище може мати перешкоди, які вимагають модифікації маршруту руху для уникнення зіткнення. Позиції кожного БПЛА у точках старту відомі, а позиції в точці цілі визначені заздалегідь для максимізації ефективності виконання місії кожним БПЛА. Вважаємо, що група БПЛА є однорідною, кожен елемент групи має однакові кінематичні та динамічні характеристики. Допустимо, що кожен БПЛА знаходиться в центрі двох концентричних сфер, які лежать в центрі мас БПЛА. Внутрішня сфера є сферою безпеки з радіусом R_s , тоді як зовнішня сфера зв'язку радіусом R_r , яка представляє радіус дії передавача таким чином, що $R_r > R_s \geq 1/k_{max}$. Отже, шляхи можуть бути промодельовані у вигляді труб з радіусами R_s навколо маршрутів руху, як це наведено на рис. 2.

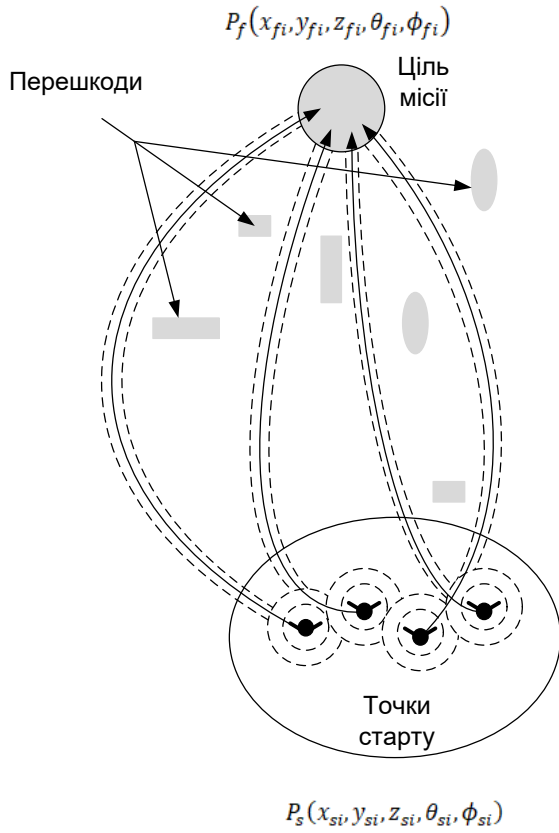


Рис. 2. Сценарій планування маршрутів декількох БПЛА (джерело: розроблено автором)

Загальну постановку завдання планування маршрутів руху БПЛА можна записати так: сформувавши найкоротші шляхи $\{r_1(q), r_2(q), \dots, r_N(q)\}$ для множини БПЛА $\{UAV_1, UAV_2, \dots, UAV_N\}$ таких, що для i -го та j -го БПЛА виконується:

$$h_{f,i} = h_{f,j} \quad (1)$$

при обмеженнях:

$$C = \begin{cases} |v_i| = |v_j|; \\ C_k: k \leq k_{max}; \\ C_\tau: \tau \leq \tau_{max}; \\ C_{safe}: r_i(q) \cap r_j(q) = \emptyset, i \neq j; \\ R_r > R_s \geq 1/k_{max}. \end{cases} \quad (2)$$

Розглянемо планування маршруту для одного БПЛА від точки старту до цільової точки (точки виконання місії). Планувальник маршруту створює шлях $r(q)$, що з'єднує початкову позицію: $P_s(x_s, y_s, z_s, \Theta_s)$ та цільову точку $P_f(x_f, y_f, z_f, \Theta_f)$:

$$P_s(x_s, y_s, z_s, \Theta_s) \xrightarrow{r(q)} P_f(x_f, y_f, z_f, \Theta_f). \quad (3)$$

Розширення рівняння (3) для врахування маршрутизації N БПЛА виконується так, що кожен БПЛА проходить через n_p точок, що представляється як:

$$\begin{aligned} &P_{s,i,j-1}(x_{s,i,j-1}, y_{s,i,j-1}, z_{s,i,j-1}, \theta_{s,i,j-1}, \psi_{s,i,j-1}) \\ &\xrightarrow{r_{i,j-1}(q), C_{safe}, C_{len}} P_{f,i,j}(x_{f,i,j}, y_{f,i,j}, z_{f,i,j}, \theta_{f,i,j}, \psi_{f,i,j}); \end{aligned} \quad (4)$$

$$|k_i(q)| < k_{max}, |\tau_i(q)| < \tau_{max}, i = 1, \dots, N, \\ j = 2, \dots, n_p,$$

де τ – кручення; k – кривизна; k_{max} – обмеження на максимальну кривизну; τ_{max} – обмеження на максимальне кручення; C_{safe} , та C_{len} – обмеження на безпеку руху (огинання перешкод) та довжину шляху. Обмеження на довжину шляху $h(q)$ для групи БПЛА має вигляд:

$$C_{len} = \min \int_{s_1}^{s_2} |h(q)| dq, \quad (5)$$

де

$$h(q) = \int_{q_1}^{q_2} \sqrt{\dot{x}(q)^2 + \dot{y}(q)^2 + \dot{z}(q)^2} dq, \quad (6)$$

де $q \in [q_1, q_2]$; $\dot{x}(q) = dx/dq$; $\dot{y}(q) = dy/dq$; $\dot{z}(q) = dz/dq$ є годографами.

Для польотів на постійній висоті формула (4) зводиться до двовимірної форми:

$$\begin{aligned} &P_{s,i,j-1}(x_{s,i,j-1}, y_{s,i,j-1}, z_{s,i,j-1}, \theta_{s,i,j-1}) \\ &\xrightarrow{r_{i,j-1}(q), C_{safe}, C_{len}} P_{f,i,j}(x_{f,i,j}, y_{f,i,j}, z_{f,i,j}, \theta_{f,i,j}); \end{aligned} \quad (7)$$

$$|k_i(q)| < k_{max}, i = 1, \dots, N, j = 2, \dots, n_p.$$

Після того, як початкові маршрути визначені, необхідно переконатися:

- у безпеці для польоту;
- досягнення БПЛА мети місії (одночасне прибуття до місця призначення).

При цьому виникає питання часу планування маршрутів руху. Це залежить від наявної інформації про навколишнє середовище та місцезнаходження БПЛА. Один із способів – інтегрувати процес генерації маршруту руху БПЛА з плануванням. Даний підхід схожий на планування маршрутів за допомогою методів оптимізації [1, 2] Однак, він потребує подальшого доопрацювання для створення раціональних маршрутів, а також має значну обчислювальну складність.

Інша можливість полягає в тому, щоб розділити два процеси: генерацію маршрутів та планування руху БПЛА [4]. Тепер виникають дві можливості:

коли планувати – до або після генерації маршруту. Математичну модель руху БпЛА у загальному випадку можна описати кривою. Тому планування маршруту можна розглядати як планування геометричної еволюції кривої.

Маршрут руху у рівнянні (4) повинен задовольняти обмеженням. Значення k_{max} та τ_{max} визначають кінематичні обмеження БпЛА. Також це єдині два параметри, які повністю визначають криву в просторі. У двовимірному просторі для визначення кривої потрібна лише кривизна. Крива, що задовольняє обмеженням на кривизну, є здійсненою траєкторією. Таким чином, можливий маршрут визначається як придатний для руху (відповідає кінематичним та динамічним обмеженням), так і безпечний для польоту (без зіткнень). Тобто, маршрут руху – це або

одна поліноміальна крива або складена крива. Такий маршрут корисний для прогнозування положення БпЛА протягом всієї місії до того, як БпЛА вийдуть на траєкторію. Крім того, це допомагає планувальнику маршруту врахувати кінематичні обмеження на ранній стадії планування маршруту.

Рішення завдання одночасного прибуття групи БпЛА повинно відповідати обмеженням (4). Основними обмеженнями є максимальні обмеження на кривизну та уникнення зіткнень. Модель складається з трьох етапів.

На першому етапі створюються допустимі маршрути, на другому етапі отримані маршрути модифікуються для отримання маршрутів з обмеженням на безпеку. На третьому етапі створюються маршрути однакової довжини. Схема розробленого планувальника маршрутів наведена на рис. 3.

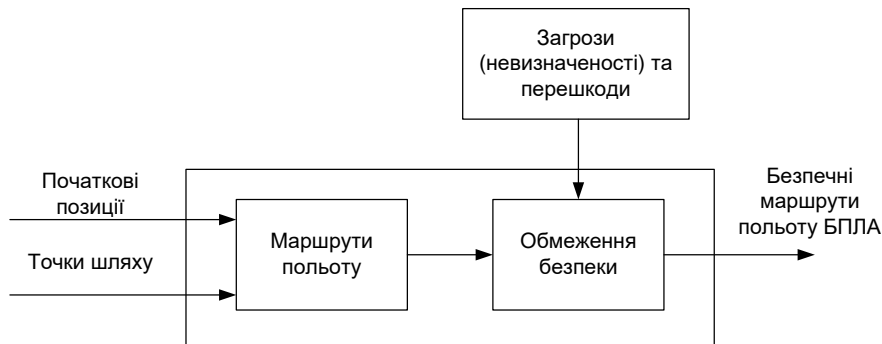


Рис. 3. Узагальнена схема планувальника маршруту (джерело: розроблено автором)

Залежно від обмежень швидкості БпЛА, одночасне прибуття до пункту призначення може бути досягнуто двома способами:

- 1) створення шляхів рівної довжини для БпЛА з постійною швидкістю;
- 2) шляхом створення шляхів різної довжини для БпЛА зі змінною швидкістю.

Рівняння (4) можна розв'язати за допомогою методів оптимізації, але це вимагатиме значних обчислювальних затрат, і отриманий в результаті маршрут може виявитися не оптимальним.

На першому етапі створюються змінні маршрути шляхом з'єднання стартової та фінішної позиції кожного БпЛА. Ці маршрути є початковими траєкторіями. Конструкцію змінних маршрутів можна записати у вигляді:

$$P_{si}(x_{si}, y_{si}, z_{si}, \theta_{si}, \psi_{si}) \xrightarrow{r_i(q)} P_{fi}(x_{fi}, y_{fi}, z_{fi}, \theta_{fi}, \psi_{fi}), \quad (8)$$

$$|k_i(q)| < k_{max}, |\tau_i(q)| < \tau_{max}, \quad i = 1, \dots, N,$$

$$j = 2, \dots, n_p$$

При цьому можна використовувати три типи маршрутів:

- 1) маршрути Дубінса;
- 2) маршрути піфагорійського годографа (ПГ)
- 3) клоатоїдні маршрути (далі – КМ).

Маршрути Дубінса та КМ – це складові шляхи, які є кусково-неперервними, тоді як ПГ – це єдиний

маршрут, який отримується шляхом інтерполяції. Контур Дубінса забезпечує контур з максимальною кривизною [6].

Для безперервного маршруту, яким БпЛА з фіксованим крилом слідує легше, можна використовувати два типи траєкторій: маршрути ПГ та КМ.

На другому етапі допустимі маршрути модифікуються під обмеження безпеки. Виконання цих обмежень забезпечує уникнення зіткнень і достатній простір для маневрування, уникаючи будь-яких перешкод. Отже, побудований маршрут є одночасно і маневреним, і безпечним. Побудований маршрут задовольняє головному обмеженню на максимальну кривизну. Однак, цей маршрут не гарантує безпеку, доки не будуть накладені обмеження безпеки на допустимі маршрути. Для уникнення зіткнення БпЛА з іншими перешкодами визначено два обмеження безпеки:

- 1) мінімальна відстань між БпЛА;
- 2) відсутність зіткнення маршрутів.

Ці обмеження наведені на рис. 4.

Вважається, що кожен БпЛА має безпечний радіус R_s , розташований в центрі маси. Цей радіус безпеки менший за радіус дії датчика, але більший за мінімальний радіус кривизни.

Визначення мінімальної відстані. Розглянемо спочатку випадок, коли кожний з маршрутів польоту не перетинається. Відстань між будь-якими двома маршрутами польоту повинна бути більшою за порогове значення, яке задається мінімальною роздільною відстанню, d_{sep} . Тут за порогове значення

приймається подвоєний безпечний радіус R_{safe} . Це можна інтерпретувати як перевірку на неперетин маршрутів польоту, а також зміщення кривих руху

на відповідність запасу міцності. Це наведено на рис. 4, де маршрути $r_1(q)$ та $r_2(q)$ розділені відстанню d_{sep} .

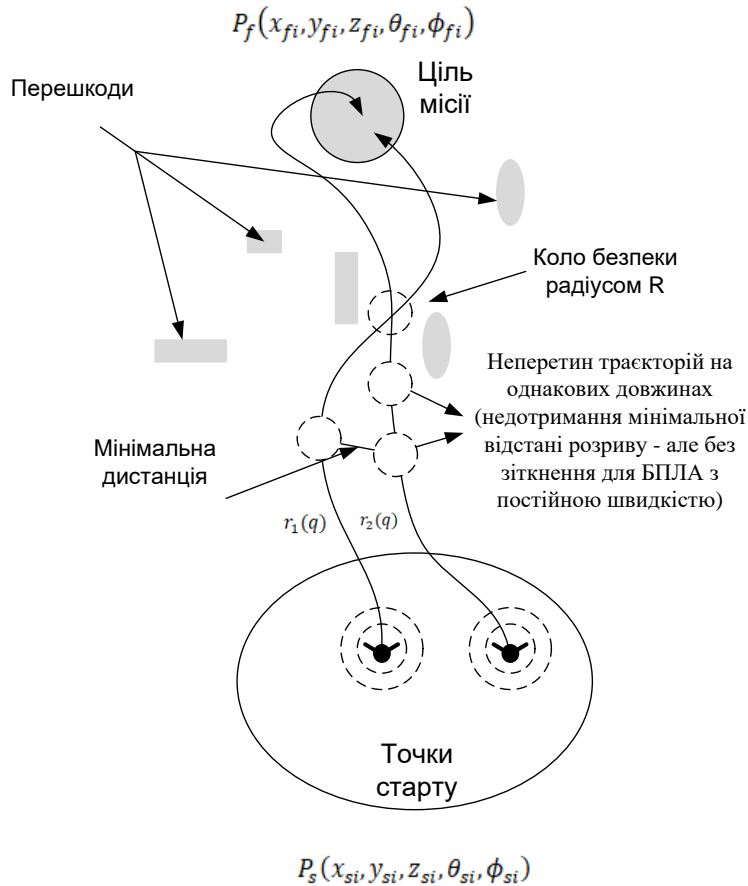


Рис. 4. Обмеження безпеки при побудові маршрутів (джерело: розроблено автором)

Реалізація мінімальної відстані відбувається наступним чином. На рис. 4 наведені кола безпеки, центри яких знаходяться у точках вздовж кожного маршруту. Якщо ці кола безпеки будь-яких двох маршрутів перетинаються, може статися зіткнення. Представляючи відстань між двома маршрутами як $r_k(q)$ та $r_m(q)$, через $d_{sep\ k,m}$, то обмеження мінімальної відстані розходження можна записати:

$$d_{sep\ k,m} = R_{s,k} + R_{s,m}, \quad (9)$$

де $R_{s,k}$ та $R_{s,m}$ – радіуси безпеки k -го та m -го маршрутів відповідно.

Отже, мінімальна відстань $d_{sep\ k,m}$ між будь-якими двома маршрутами БПЛА повинна дорівнювати або бути більшою за суму відповідних радіусів їх кіл безпеки. Відстань між двома БПЛА вимірюється шляхом обчислення евклідової відстані між двома точками на двох різних маршрутах. Відстань між k -м маршрутом та m -м маршрутом при певній довжині або часі визначається за формулою:

$$d_{sep\ k,m} = \sqrt{(z_m - z_k)^2 + (y_m - y_k)^2 + (x_m - x_k)^2}, \quad (10)$$

де (x_k, y_k, z_k) – точка на k -му маршруті, (x_m, y_m, z_m) – точка на m -му маршруті.

Для однорідних БПЛА з однаковими безпечними відстанями рівняння (9) зводиться до:

$$d_{sep\ k,m} > 2R_{safe} \quad (11)$$

Відсутність зіткнення маршрутів. У багатьох випадках побудовані маршрути можуть мати перетини. Це порушує обмеження на безпеку. У таких випадках тест на мінімальну відстань дасть хибний результат. Якщо маршрути перетинаються, це не обов'язково означає, що відбудеться зіткнення. БПЛА зіткнуться лише у тому випадку, якщо вони досягнуть точки перетину одночасно. Кола безпеки перетнуться, якщо БПЛА знаходяться в межах $d_{sep\ k,m}$ при наближенні до точки перетину.

Відповідно до рис. 4, шляхи $r_k(q)$ та $r_m(q)$, уникають зіткнення, якщо різниця між довжинами маршрутів до точки перетину d_{int} задовольняє наступній умові:

$$d_{int\ k,m} = |s_{int,k} - s_{int,m}| > d_{sep} > (R_{s,k} + R_{s,m}), \quad (12)$$

де $s_{int,k}$ та $s_{int,m}$ – довжини шляхів $r_k(q)$ та $r_m(q)$, відповідно, від їх початкових точок до точки перетину.

Якщо маршрути задовольняють обмеженням безпеки, визначеним у рівняннях (11) та (12), то

немає необхідності їх перепланувати. У іншому випадку, маршрути переплануються шляхом зміни кривизни або проходження через проміжні точки та/або позиції.

Якщо маршрут, який можна прокласти, задовольняє умові мінімального розділення у нерівності (11), то маршрути безпечні для проходження й немає необхідності перепланувати. Якщо ця умова не виконується, то перевіряється умова (12) на зіткнення маршрутів на перетині. Якщо обидві умови не виконуються, то потрібне повторне планування. Для групи з N БПЛА, яка використовує n БПЛА за один раз, кількість тестів безпеки T_{safe} , які необхідно провести, становить:

$$T_{safe} = 2 \frac{N!}{n!(N-n)!} \quad (13)$$

Збільшення або переміщення кола (сфери) вздовж маршруту польоту, а потім побудова еволюційної кривої створює або дві криві зсуву для двохвимірної кривої, або трубу (канал) для трьохвимірної кривої. У двохвимірному випадку мінімальна відстань розділення між маршрутами може бути представлена набором еволюційних маршрутів зміщення.

Маршрути зміщення $r_d(q)$ на відстані зміщення R_{safe} задаються формулою:

$$r_d(q) = r(q) \pm R_{safe}n(q), \quad (14)$$

де $n(q)$ – вектор нормалі до $r(q)$, а відстань зміщення дорівнює подвоєному радіусу безпеки.

Аналогічно, для тривимірної маршруту поверхні каналу або труби $r_s(q)$ з безпечним радіусом задається формулою:

$$r_s(q) = r(q) + R_{safe}[n(q), b(q)] \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix}, \quad (15)$$

де $n(q)$ – нормальний вектор, а $b(q)$ – бінормальні вектори, які можна аналітично обчислити через параметризацію кривої.

Однотимчасне прибуття до кінцевої точки (точки місії) вимагає вирівнювання всіх довжин маршрутів для групи БПЛА, які летять з однаковою швидкістю. Хоча на другому етапі планувальник створює можливі маршрути (ефективні та безпечні). Вони, як правило, не мають однакової довжини.

Еталонна довжина маршруту необхідна для того, щоб забезпечити мету для модифікації довжини маршруту. Маршрути можна подовжувати, змінюючи обмеження на кривизну. Довжини всіх маршрутів збільшуються до довжини еталонної траєкторії шляхом відповідних перетворень з обмеженнями на кривизну.

Якщо еталонний маршрут, що відповідає маршруту максимальної довжини, задано як r_{ref} , а його довжина як s_{ref} , то

$$r_{ref} = r(s_{ref}), i = 1, \dots, N, \quad (16)$$

Значення кривизни k модифікується так, щоб задовольнити умову:

$$h_{ref} - \{h_i\} = 0, i = 1, \dots, N, \quad (17)$$

де h_i – довжина i -го маршруту, а N – кількість БПЛА.

Еталонний маршрут не враховується в процесі модифікації.

Блок-схема алгоритму одночасного прибуття безпілотних літальних апаратів до пункту призначення наведена на рис. 5.

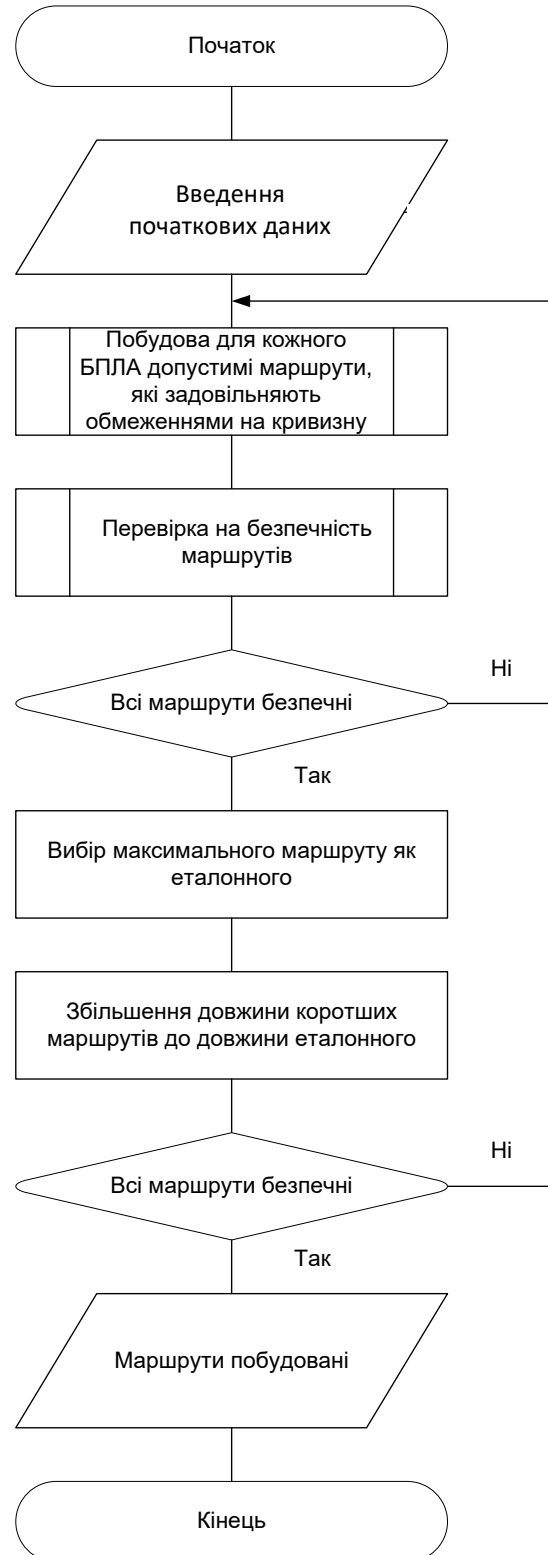


Рис. 5. Блок-схема алгоритму одночасного прибуття безпілотних літальних апаратів до пункту призначення (джерело: розроблено автором)

Швидкість обчислень можна прискорити, використовуючи глобальний планувальник для створення маршрутних точок та/або позицій або шляхом створення проміжних маршрутних точок.

Висновки

Одночасне прибуття БПЛА до точки місії досягається шляхом створення шляхів однакової довжини. Для створення шляхів однакової довжини використовується аналогія з розтягуванням ланцюжка пружин, що з'єднує точки маршруту.

Отримала подальший розвиток модель планування маршрутів руху однорідної групи безпілотних літальних апаратів, що на відміну від існуючих враховує обмеження та вимоги дотримання безпеки руху безпілотних літальних апаратів та забезпечує одночасне прибуття групи до кінцевої точки виконання місії.

Методи побудови маршрутів руху ґрунтуються на кривих Дубінса, годографічних кривих, клотоїдних траєкторіях та узгоджують дії групи щодо виконання місії.

СПИСОК ЛІТЕРАТУРИ

1. Chodnicki, M.; Siemiatkowska, B.; Stecz, W.; Stępień, S. Energy Efficient UAV Flight Control Method in an Environment with Obstacles and Gusts of Wind. *Energies* 2022, 15.
2. Huang, G.; Hu, M.; Yang, X.; Lin, P. Multi-UAV Cooperative Trajectory Planning Based on FDS-ADEA in Complex Environments. *Drones* 2023, 7, 55. <https://doi.org/10.3390/drones7010055>
3. Liu, Y.; Zhang, F.; Huang, P.; Zhang, X. Analysis, Planning and Control for Cooperative Transportation of Tethered Multi-Rotor UAVs. *Aerosp. Sci. Technol.* 2021, 113, 106673.
4. Xu, L.; Cao, X.; Du, W.; Li, Y. Cooperative Path Planning Optimization for Multiple UAVs with Communication Constraints. *Knowl. Based Syst.* 2022, 129, 110164.
5. NataliiaYeromina, Tarshyn, V., Petrov, S., Samoilenko, V.T., Tabakova, I., Dmitriiev, O., Surkova, K., Danylko, O., Kushnierova, N., Soroka, M., Salo, N., & Chumak, O. (2021). Method of reference image selection to provide high-speed aircraft navigation under conditions of rapid change of flight trajectory. *International Journal of Advanced Technology and Engineering Exploration*, Vol 8(85) ISSN (Print): 2394-5443 ISSN (Online): 2394-7454 <http://dx.doi.org/10.19101/IJATEE.2021.874814>
6. Tymochko O., Tymochko O., Trystan A., Matiushchenko O. and Berezhnyi A. Method of Automated Flight Route Planning for Unmanned Aerial Vehicles to Search for Stationary Objects. 2020 IEEE 11th International Conference on Dependable Systems, Services and Technologies (DESSERT), Kyiv, Ukraine, 2020. P. 242-246.
7. Бережний А. О., Крижанівський І. М. Комплекс задач системи підтримки прийняття рішення на планування маршрутів польотів безпілотних літальних апаратів. Системи управління, навігації та зв'язку. Збірник наукових праць. Полтава, 2020. Вип. 1(59). С. 3–6. DOI:10.26906/SUNZ.2020.1.003.
8. Бережний А.О., Крижанівський І.М., Барабаш О.В. Метод автоматизованого планування маршрутів безпілотних літальних апаратів з урахуванням виявлення стаціонарних об'єктів. Телекомунікаційні та інформаційні технології. Київ, 2019. Вип. 4(65). С. 90–98. DOI:10.31673/2412-4338.2019.049098.
9. Aabid, A.; Parveez, B.; Parveen, N.; Khan, S.A.; Zayan, J.; Shabbir, O. Reviews on design and development of unmanned aerial vehicle (drone) for different applications. *J. Mech. Eng. Res. Dev.* 2022, 45, 53–69.
10. Amici, C.; Ceresoli, F.; Pasetti, M.; Saponi, M.; Tiboni, M.; Zaroni, S. Review of propulsion system design strategies for unmanned aerial vehicles. *Appl. Sci.* 2021
11. Rojas-Perez, L.O.; Martínez-Carranza, J. On-board processing for autonomous drone racing: An overview. *Integration* 2021, 80, 46–59.
12. Tymochko O., Trystan A., Berezhnyi A., Sotnikov O., Matiushchenko O., Kryzhanivskiy I. Vehicles while Monitoring Behavior of Dynamic Objects in a Forest-Steppe Area. *International Journal of Emerging Trends in Engineering Research.* 2020. Vol.8, No 7. P. 3208 – 3215.

Received (Надійшла) 14.12.2024

Accepted for publication (Прийнята до друку) 05.03.2025

Model for planning routes for a homogeneous group of unmanned aerial vehicles

Oleksandr Chumak

Abstract. The article determines that currently there are a significant number of models and methods that allow for effective planning of UAV movement routes to perform tasks, however, the next stage in the development of robotic technologies is the collective use of heterogeneous systems to perform common tasks. Simultaneous arrival at the required location or locations is one of the elements of mission execution that requires joint planning and coordination between UAVs in the group. The mission objective is that the UAV group must reach the destination simultaneously, which is achieved by creating trajectories of equal length, assuming that all UAVs fly at the same speed. The article further develops a model for planning the routes of a homogeneous group of unmanned aerial vehicles, which, unlike existing ones, takes into account the limitations and requirements for UAV safety and ensures the simultaneous arrival of the group to the final point of the mission. The planning model has three stages, taking into account the maximum curvature constraint and collision avoidance. In the first stage, permissible routes are created, in the second stage, the resulting routes are modified to obtain routes with safety constraints. In the third stage, routes of equal length are created. A route planner scheme and a flowchart of the algorithm for simultaneous arrival of UAVs at the destination are developed. Methods based on Dubins curves, hodographic curves, clothoid trajectories are used to construct UAV movement routes and coordinate the actions of the group to complete the mission. The speed of model calculations can be accelerated by using a global planner to create waypoints and/or positions or by creating intermediate waypoints.

Keywords: unmanned aerial vehicle; genetic algorithms; information technologies; trajectory optimization; planning of UAV group movement routes; air traffic control; artificial intelligence.

М. Й. Заполовський, М. В. Мезенцев, М. В. Оліфір

Національний технічний університет "Харківський політехнічний інститут", Харків, Україна

МОДЕЛЮВАННЯ ТЯГОВОГО ЕЛЕКТРОПРИВОДА ДИЗЕЛЬ-ПОЇЗДА З УРАХУВАННЯМ ВИКОРИСТАННЯ ГУМО-КОРДОВОЇ МУФТИ В ПРОЦЕСІ БУКСУВАННЯ КОЛІСНИХ ПАР

Анотація. Розглянуті питання розроблення та дослідження моделей електроприводу змінного струму дизель-поїзда. Проведено огляд літературних джерел на задану тематику та аналіз існуючих підходів до розв'язання задач, пов'язаних з проблемами розроблення систем керування, які забезпечують збільшення ефективності використання електроприводу за умов експлуатації, в тому числі при можливому буксуванні колісних пар. Розглянуті як загально прийняті підходи проведення досліджень за допомогою моделювання процесів буксування колісних пар, так і запропоновані. Розроблені комп'ютерні компоненти електромеханічної системи дизель-поїзда з урахуванням пружних зв'язків за умови забезпечення якісних показників системи керування у його елементах в процесі буксування та моделі, що описує можливий процес буксування. Приведені моделі можуть бути використані для дослідження системи автоматичного регулювання електропередачі та уточнення її параметрів, крім того для визначення параметрів гумо-кордової муфти (ГКМ) і в подальшому їх реалізації у вигляді відповідних комп'ютерних компонентів системи керування дизель-поїзда. Проведено дослідження запропонованої моделі як з точки зору її адекватності, так і її можливості застосування для цілей розроблення системи керування, що забезпечує ефективність функціонування електроприводу з ГКМ в процесі буксування. Результати досліджень надано у вигляді графіків процесів характерних фазових змінних та таблиць, в яких наведені числові значення характерних змінних системи керування в залежності від сигналів управління.

Ключові слова: комп'ютерні компоненти, моделювання процесів буксування, електропривод змінного струму, дизель-поїзд, системи керування, буксування, пружні зв'язки, гумо-кордова муфта, адекватність, частота ковзання, графіки процесів, таблиці, сигнали управління.

Вступ

У процесі роботи механічного обладнання (редуктор, полий вал) дизель-поїзда виникають динамічні навантаження, особливо в процесі буксування колісних пар, що призводять до виходу з ладу деталей, вузлів і механізмів, у зв'язку з чим знижується його довговічність та експлуатаційна надійність. З урахуванням цього виникає необхідність у пошуку рішень, пов'язаних із збільшенням терміну служби елементів приводу та зниженням негативного впливу динамічних навантажень на обладнання. Одним з таких рішень є використання компенсуючих муфт. Муфти здатні демпфувати ударні навантаження, що виникають у процесі роботи обладнання. Математичний апарат для аналітичного розрахунку еластичних муфт та їх застосування в електроприводах достатньою мірою складний. У зв'язку з цим для моделювання умов навантаження, в тому числі і для можливого процесу буксування колісних пар дизель-поїзда та на цій основі виконувати обґрунтування та розрахунки електричних параметрів муфти з еластичними елементами пропонується використання розроблених математичних моделей та їх реалізації із застосуванням програмного комплексу пакету MATLAB. Такий підхід в якій то мірі вирішує проблему при створенні локомотивів з електроприводом змінного струму, зокрема розроблення комп'ютерних компонентів для системи управління, яка забезпечує ефективне використання зчпної ваги, де в електромеханічній системі використовується гумо-кордова муфта (ГКМ). У зв'язку з цим потрібно дослідження

динаміки дизель-поїзда з метою розробки сучасних систем управління, що забезпечують запобігання буксуванню або зведенню до мінімуму енергетичних витрат, з використанням математичних моделей як самого об'єкта в цілому, так і його окремих електромеханічних компонентів, блоків і вузлів.

Постановка задачі і аналіз відомих публікацій. Огляд робіт, присвячених задачі розробки ефективних систем управління електропередачі дизель-поїзда з електроприводом змінного струму показує, що в даний час значна увага приділяється створенню систем керування на основі сучасних досягнень проектування з використанням методів оптимізації, нечіткої логіки, нейрокомп'ютерних технологій, розробленням моделей електромеханічних компонентів, блоків і вузлів та їх дослідження [1 – 3]. Проблемам розроблення та дослідження моделей електромеханічної системи електроприводу змінного струму з урахуванням пружних зв'язків та умов зчеплення колісних пар із рейками присвячено ряд робіт [2, 3]. Практичний інтерес представляє математична модель, яка враховує реалізацію моменту обертання тягового асинхронного двигуна (ТАД) при взаємодії колісних пар із рейками, зокрема математична модель реалізації сили зчеплення колісних пар дизель-поїзда при втраті та відновленні зчеплення з урахуванням пружних зв'язків. В даній роботі розглядаються питання моделювання процесів динаміки дизель-поїзда, що виникають в елементах електромеханічної системи асинхронного електроприводу в залежності від навантаженості дизель-поїзда (наведеного моменту інерції) при порушенні умов зчеплення колісних пар із

рейками за умов використання в системі електроприводу гумо-кордової муфти.

Метою даної роботи є створення комп'ютерних компонентів для системи управління дизель-поїзду на основі математичної моделі електромеханічної системи дизель-поїзда з ТАД з урахуванням пружних зв'язків у його елементах в процесі буксування та моделі, що описує можливий процес буксування. Моделі дозволяють виконувати обґрунтування і розрахунки електричних параметрів муфти з еластичними елементами та проводити дослідження на стадії проектування при розробці комп'ютерних компонентів системи керування дизель-поїзду.

Основна частина

Однією з важливих задач, які зв'язані з проектуванням систем керування об'єктами, є розроблення математичної моделі електромеханічної системи дизель-поїзда для визначення структури і параметрів регуляторів системи управління, її динамічних і статичних характеристик у різних режимах функціонування для

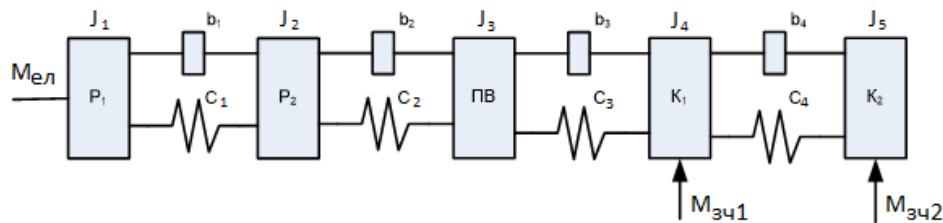


Рис. 1. Структурна схема електромеханічної системи дизель-поїзда

Для дослідження процесів буксування і впливу ГKM (параметрів муфти), приймаємо, що загальна модель системи складається з моделей: системи ГKM; еквівалентного електроприводу з небуксуючими колісними парами; електроприводу, колісні пари якого можуть бути в стані буксування; блоку формування керуючих впливів (напруги живлення ТАД і її частоти) та сигналу зворотного зв'язку обертання ротора буксуючого ТАД; блоку навантаження та формування швидкості руху дизель-поїзда. При цьому швидкість руху визначається сукупним тяговим моментом від усіх двигунів. Загальна структурна схема моделі дизель-поїзда наведена на рис. 2.

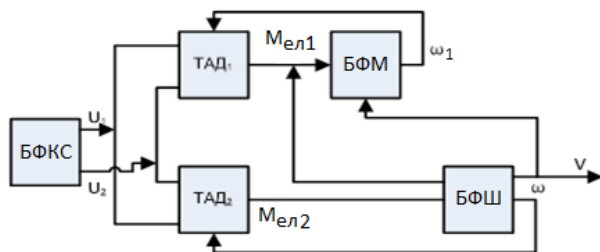


Рис. 2. Загальна схема моделі дизель-поїзда

Тут: БФКС – блок формування керуючих сигналів U_1, U_2 (напруги живлення ТАД та її частоти); ТАД₁ – тяговий асинхронний двигун (буксуючий); ТАД₂ – еквівалентний тяговий асинхронний двигун (небуксуючий); $M_{ел1}, M_{ел2}$ – електромагнітний момент відповідного ТАД; БФМ – блок формування моменту зчеплення колісних пар з урахуванням пружних зв'язків;

забезпечення надійного функціонування об'єкта в цілому. Дану інформацію можливо отримати за допомогою експериментальних досліджень або шляхом математичного моделювання.

Структурну схему електромеханічної системи дизель-поїзда наведено на рис. 1 [2].

Тут: P_1 – ротор ТАД; P_2 – редуктор; ПВ – порожнистий вал; K_1, K_2 – колеса вісі дизель-поїзда.

Зв'язок ТАД із колісною парою дизель-поїзда у загальному випадку є пружним за рахунок скручування валів редуктора та колісної пари, пружної деформації колісної пари, а також сполучних муфт. На систему діють зовнішні моменти: обертаючий момент ТАД $M_{ел}$, який прикладений до ротора і діє на колісну пару через редуктор і сполучну муфту та момент зчеплення колісної пари з рейками $M_{зч}$. З урахуванням припущень [2] окремі ланки можуть бути представлені у вигляді зосереджених матеріальних точок, що мають момент інерції ($J_1 - J_5$), а ідеалізовані зв'язки між ними характеризуються пружністю ($C_1 - C_4$) і дисипативними властивостями ($b_1 - b_4$).

БФШ – блок формування швидкості дизель-поїзда; ω , ω_1 – кругова швидкість обертання відповідної колісної пари. Залежно від цілей дослідження моделі ТАД₁ та ТАД₂ можуть бути представлені як з урахуванням середніх значень фазових змінних (в тому числі на основі методів векторного керування), так і миттєвих. Їх математичні моделі та моделі формування керуючих сигналів (блок БФКС) докладно описані у [3, 4]. В даній роботі використано математичну модель ТАД з урахуванням середніх значень фазових змінних. Формування керуючих сигналів виконувалось як з використанням частоти напруги живлення ω_e згідно розімкненого закону $U/f = \text{Const}$, так і замкненого закону $\omega_e = \omega_1 + \Delta S$ (ω_e – частота напруги живлення ТАД; ω_1 – кругова швидкість обертання небуксуючої колісної пари; ΔS – величина ковзання частоти (постійне значення або синтезоване). Значення амплітуди напруги живлення формувалось згідно заданого співвідношення $U/f = \text{Const}$.

Модель блоку формування моменту зчеплення колісних пар з урахуванням пружних зв'язків (БФМ) (згідно рис. 1) можна подати у вигляді двомасової [2, 5, 6], в якій визначальну роль відіграють параметри сполучної муфти (величина жорсткості – C , коефіцієнт в'язкого тертя – B). У цьому випадку структурна схема двомасової системи представлена зосередженими масами редуктора, валу двигуна і робочого механізму (колісними парами), між якими є пружний елемент (сполучна муфта), наведена на рис. 3.

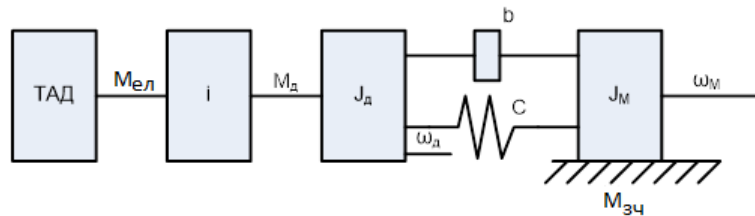


Рис. 3. Структурна схема двомасової системи

Тут вал ТАД з моментом інерції J_d пов'язаний з робочим механізмом (колісною парою) з моментом інерції J_m пружним зв'язком з коефіцієнтом жорсткості C і коефіцієнтом в'язкості тертя b . При цьому швидкість обертання ротора ТАД $\omega_d(t)$ не дорівнює швидкості механізму $\omega_m(t)$. Рівняння двомасової системи з урахуванням умов зчеплення колісних пар з рейками мають вигляд [2]:

$$\frac{d\omega_d}{dt} = \frac{1}{J_d}(M_d - M_{ne}), \quad (1)$$

$$M_{ne} = M_n + B(\omega_d - \omega_m), \quad (2)$$

$$\frac{dM_n}{dt} = C(\omega_d - \omega_m), \quad (3)$$

$$\frac{d\omega_m}{dt} = \frac{1}{J_m}(M_{ne} - M_{зч}), \text{ якщо } M_{зч} < M_{ne}, \quad (4)$$

$$\omega_m = \omega, \text{ якщо } M_{зч} \geq M_{ве}, \quad (5)$$

де $M_{не}$ – сумарний момент, що передається пружним зв'язком, яка дорівнює сумі пружного моменту M_n і моменту в'язкого тертя; J_d, J_m – відповідно момент інерції двигуна та механізму, наведені до валу двигуна; M_d – тяговий момент ТАД з урахуванням редуکتора; ω_d, ω_m – відповідно швидкості обертання ротора ТАД та колісної пари; ω – кутова швидкість колісної пари, що не буксує; $M_{зч}$ – момент зчеплення колісної пари з рейками. $M_{зч}$ може бути визначений відповідно до «Правил тягових розрахунків» [2]. Слід відмітити, що згідно цих правил модель формування $M_{зч}$ носить евристичний характер і експериментальні дані є тільки до ряду локомотивів. Для дизель-поїздів мають рекомендаційний характер.

Моделювання процесів на основі представлення електромеханічної системи у вигляді двомасової (рис. 3) пропонується наступним чином. Згідно досліджень в процесі експлуатації ГМК визначено, що процеси, які відбуваються при формуванні тягового моменту дизель-поїзда в процесі руху, використовуючи електромагнітні моменти ТАД і ГМК, можуть здійснюватися згідно певних законів – експонентного або коливального. Згідно фізичних властивостей математичну модель ГМК за певних умов можливо представити передавальною функцією вигляду:

$$W(p) = \frac{Y}{X} = \frac{k}{T_1 p^2 + T_2 p + 1}, \quad (6)$$

де Y – зображення вихідного сигналу муфти; X – зображення вхідного сигналу муфти; k – коефіцієнт передачі; T_1, T_2 – постійні часу (с).

В залежності від значень T_1, T_2 , зображення вихідної величини Y за умови постійного зображення

величини X може змінюватись згідно експоненціального або коливального закону. Таке представлення математичної моделі дозволяє проводити дослідження в певному спектрі можливих змін особистих частот (періоду коливаний) та швидкості їх затухання, що дає можливість в певному сенсі визначати електричні параметри ГМК (значення b і C). Згідно передавальної функції (6) модель ГМК представлена лінійним диференціальним рівнянням другого порядку:

$$y''(t) + a_1 y'(t) + a_2 y(t) = x(t), \quad (7)$$

де a_1 і a_2 – постійні величини, значення яких визначаються постійними часу T_1 і T_2 . Рішення рівняння (7) представляється сумою вигляду:

$$y(t) = y_o(t) + y_b(t). \quad (8)$$

Складова рівняння (8) $y_o(t)$ визначає характер перехідного процесу і визначається як:

$$y_o(t) = C_1 e^{k_1 t} + C_2 e^{k_2 t}, \quad (9)$$

де C_1, C_2 – визначаються початковими умовами рівняння (7); k_1, k_2 – корні однорідного характеристичного рівняння, яке відповідає рівнянню (7). В залежності від виду коренів (дійсні чи комплексні) отримаємо і відповідний характер перехідного процесу – експоненціальний чи коливальний. Складова $y_b(t)$, яка визначає реакцію системи в залежності від типу $x(t)$. За умови постійного значення $x(t)$, $y_b(t)$ буде визначатись як:

$$y_b(t) = x(t) / a_2. \quad (10)$$

Задання орієнтованих значень частоти (періоду) особистих коливаний в процесі досліджень з метою оцінки та вибору електричних параметрів b і C можливо на основі рішення диференціального рівняння вигляду:

$$y''(t) + \omega^2 y(t) = 0 \quad (11)$$

при заданих початкових умовах.

Аналітичне рішення рівняння (11):

$$y(t) = A \sin \omega t,$$

де значення A визначаються початковими умовами рівняння (11). В результаті орієнтоване значення частоти ω в рівнянні (11) пов'язане з коефіцієнтом a_2 рівняння (7).

Рішення однорідного диференціального рівняння (11) дає нам змогу в моделі задавати частоту (період) особистих коливаний, які характерні для ГМК, що використовується в даному електроприводі.

Згідно аналітичного рішення рівняння (9) за допомогою коефіцієнта a_1 рівняння (7) в моделі є змога задавати швидкість затухання коливань ГKM. Для оцінки параметрів ГKM використано прямі показники якості: час керування; величина перерегулювання (абсолютне значення максимального відхилення); коливальність.

При дослідженнях за умови використання ГKM практичний інтерес представляє процес формування тягового моменту в результаті буксування колісних пар дизель-поїзда та вплив ГKM на процес формування тягового моменту дизель-поїзда. При цьому береться до уваги, що процес буксування носить дискретний характер. Тобто зміна тягового моменту також носить дискретний характер. В результаті в процесі моделювання з урахуванням ГKM тяговий момент за умов стрибко-подібної зміни електромагнітного моменту згідно рівняння (10) будемо визначати як:

$$M_{\partial} = y(t)\sqrt{a_2}. \quad (12)$$

Згідно [7, 8] математичну модель ТАД представлена у вигляді:

$$\frac{d\Psi_{s1}}{dt} = \Psi_{s2}\omega_s - R_S I_{s1} - U_{s1}. \quad (13)$$

$$\frac{d\Psi_{s2}}{dt} = -\Psi_{s1}\omega_s - R_S I_{s2} - U_{s2}. \quad (14)$$

$$\frac{d\Psi_{r1}}{dt} = (\omega_s - p\omega)\Psi_{r2} - R_R I_{r1}. \quad (15)$$

$$\frac{d\Psi_{r2}}{dt} = -(\omega_s - p\omega)\Psi_{r1} - R_R I_{r2}. \quad (16)$$

де Ψ_{i1} , Ψ_{s2} , Ψ_{r1} , Ψ_{r2} , U_{s1} , U_{s2} – відповідно проекції векторів потоків статора і ротора та напруги живлення на осі координат; ω_s – кутова частота напруги живлення статора ТАД при $p = 1$; ω – кутова швидкість ротора; I_{s1} , I_{s2} , I_{r1} , I_{r2} – відповідно проекції векторів струмів обмоток статора і ротора; R_S , R_R – активні опори обмоток статора і ротора.

Рівняння для визначення електромагнітного моменту [7]:

$$M = \frac{3}{2} p \frac{k_R}{\sigma L_S} (\Psi_{r1}\Psi_{s2} - \Psi_{s1}\Psi_{r2}). \quad (17)$$

Швидкість V_t дизель-поїзда визначається по кутовій швидкості небуксуючого ТАД.

За відсутності буксування швидкість руху пропорційна кутовій швидкості обертання ротора двигуна ω . Рівняння руху має вигляд:

$$\frac{d\omega}{dt} = \frac{p}{J_T} (M_T - M_o) \quad (18)$$

де M_T – тяговий момент дизель-поїзда; J_T – момент інерції дизель-поїзда; M_o – момент опору руху; $J_T = mR_K^2$; m – маса дизель-поїзда; R_K – радіус колеса.

Тяговий момент дизель-поїзда (утворюють чотири двигуни ТАД):

$$M_T = 4 M_{\partial} i, \quad (19)$$

де i – передавальне число редуктора.

Момент опору руху визначається згідно правил тягових розрахунків:

$$M_o = R_K W_0 G_H = 9.8 R_K (1.1 + 0.12V)m, \quad (20)$$

де R_K – радіус колеса за колом кочення; W_0 – узагальнений коефіцієнт для відповідного типу поїзда згідно тягових розрахунків; G_H – вага поїзда (Т); V – швидкість поїзда (км/год).

Швидкість дизель-поїзда:

$$V = 3,6 R_K \omega_K, \quad (21)$$

де ω_K – кутова частота обертання колеса (1/с).

Оскільки тяговий момент формується на виході редуктора то і швидкість (електрична) обертання ротора ω_p в математичній моделі визначається в залежності від передатного числа редуктора.

В результаті рівняння руху має вигляд:

$$\frac{d\omega_K}{dt} = \frac{p^* i}{J_T} (4^* M_{\partial}^* i - M_o), \quad (22)$$

Адекватність моделі перевірялась на предмет відповідності експериментальних даних з даними, що отримані в процесі моделювання.

Вихідні дані визначення швидкості руху дизель-поїзда:

Склад дизель-поїзда: головних вагонів – 2; причіпних вагонів – 1. Діаметр колеса по колу кочення – 950 мм. Передатне відношення тягового редуктора (i) – 3.69. Маса порожнього спорядженого вагона: головного дизельного – 70130 кг; причіпного – 40000 кг.

Маса головного вантажного вагона: при розрахунковому завантаженні – 76030 кг; при номінальному завантаженні – 80000 кг; при максимальному завантаженні – 82000 кг. Маса причіпного вантажного вагона: при розрахунковому завантаженні – 54250 кг; при номінальному завантаженні – 59530 кг; при максимальному завантаженні – 63520 кг.

В табл. 1 наведено розрахункові значення параметрів моделі згідно вихідних даних.

Таблиця 1 – Розрахункові значення параметрів моделі

Показники	Завантаження:		
	розрахункове	номінальне	максимальне
Маса, m (кг)	260560,0	279060,0	291040,0
Момент інерції, J (кг ²)	58789,0	62956,0	65659,0
Загальний коефіцієнт, p / J	0,000051	0,000048	0,000046
Момент опору руху, M_o (Нм)	1334,2+24,9 ω_K	1428,9+26,8 ω_K	1490,0+27,9 ω_K

Дослідження проводились за наступних умов:
Завдання напруги живлення ТАД та її частоти – згідно закону керування $U/f = \text{const}$. При цьому частота напруги живлення ω_s формувалась як $\omega_s = \omega_k + \Delta S$, ω_k – кругова швидкість колеса колісної пари локомотива; ΔS – величина ковзання час-

тоти напруги живлення ТАД, що визначається різницею між частотою ω_s та електрично круговою частотою обертання ротора ТАД ω_k .

В табл. 2 наведено результати досліджень за умови: час розгону дизель-поїзда – 60 (сек); $\Delta S = 12 / \Delta S = 18$ (1/с); $U/f=14$ (В/Гц).

Таблиця 2 – Результати моделювання за певних умов

Навантаження	p / J	M_d Нм	V , км/год	S , м	a , м/с ²
Розрахункове	0.000051	2930,0/3550,0	53,18/66,0	407,1/514,6	0,23/0,29
Номінальне	0.000048	2920,0/3561,0	49,89/62,23	379,4/480,4	0,21/0,27
Максимальне	0.000046	2910,0/3550,0	47,68/59,52	360,8/457,5	0,20/0,25

Згідно результатів, що приведені в табл. 2 за різних умов процесу розгону при відповідних навантаженнях, дані, які отримані на моделі, відповідають експериментальним значенням.

Моделі для дослідження процесів, що пов'язані з буксуванням, розглядалися в двох варіантах: безпосередньо через дискретну зміну електромагнітного моменту ТАД на певну величину з використанням і без використання ГKM і зміна електромагнітного моменту при використанні ГKM в процесі буксування в залежності від впливу $M_{зч}$ – моменту зчеплення колісної пари з рейками шляхом використання специфіки функціонування ТАД при зміні частоти ковзання.

На рис. 4 приведено характер процесів електромагнітного моменту ТАД M_n (а – без ГKM, б – з ГKM) за умови його дискретної зміни на десятій секунді розгону дизель-поїзда як $M_n = 0.2 M_d$. Процес розгону відбувався в замкнутій системі керування, де

частота напруги живлення визначалась згідно співвідношення $\omega_s = \omega_k + \Delta S$. При цьому $\Delta S = 12$ (1/с). Рис. 4(б) характеризує характер перехідного процесу електромеханічної системи за умови використання ГKM з відповідними конструктивними параметрами В і С. В зоні пуску ГKM приводить до появи небажаних коливань амплітудою до $1.5 M_n$, що може приводити до появи процесу буксування. В зоні дискретної зміни M_d на заданій десятій секунді ГKM приводить до позитивного результату: електромагнітний момент стає меншим до значення $0.5 M_n$, що запобігає виникненню буксування; процес носить коливальний характер, який має плавний, а не ударний характер (негативно діє на механічну частину (редуктор) електроприводу); можливо задовольняти якісні показники перехідного процесу за допомогою вибору конструктивних параметрів ГKM.

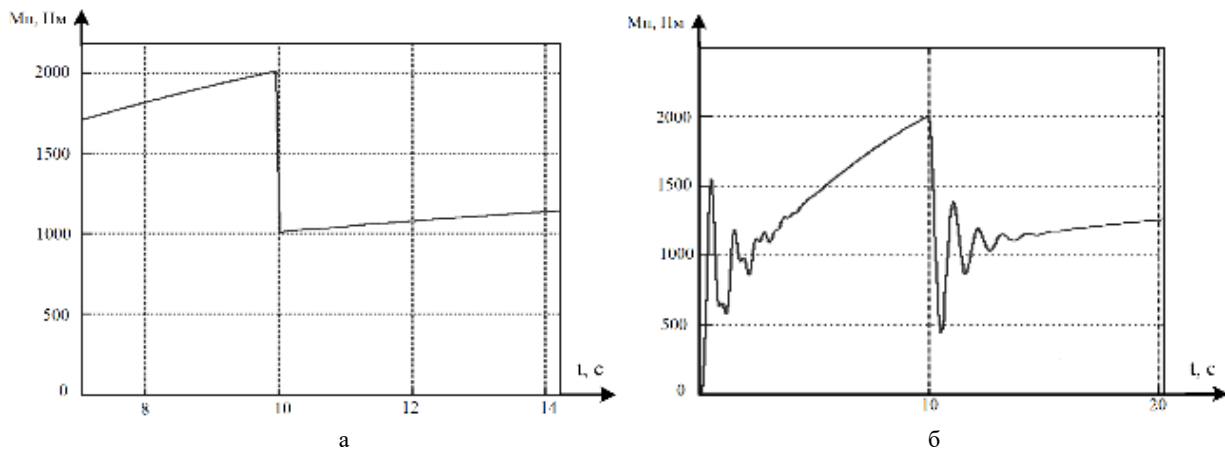


Рис. 4. Графіки зміни електромагнітного моменту ТАД

В табл. 2 приведено результати впливу конструктивних параметрів ГKM В і С в процесі досліджень за умови: час розгону дизель - поїзда – 60 (сек); $p / J = 0.000051$; $\Delta S = 12$ (1/сек); $\Delta S = 18$ (1/сек); $U/f=14$ (В/Гц).

Показники в табл. 4 M_1 , M_2 , M_3 відповідають значенням електромагнітного моменту ТАД (рис. 4, б) з ГKM в процесі розгону дизель-поїзда за умови

виникнення буксування колісної пари при різних значеннях електричних параметрів моделі муфти (В і С), напруги та частоти живлення ТАД, рівня завантаженості. Приведено відповідні якісні показники. Значення M_d відповідає процесу, що встановився.

Дані таблиці показують міру впливу на якісні показники сумісного процесу – параметрів ГKM і процесу буксування колісної пари ТАД.

Таблиця 2 – Результати дослідження

B	C	V, км/г	n	t _p , с	M ₁ , Нм	M ₂ , Нм	M ₃ , Нм	M ₄ , Нм
$\Delta S = 18$ (1/с)								
1.2	9.86	55.21	8	10	3103	-720	1360	695
1.6	9.86	55.12	6	8	3100	-455	1103	696
2.0	9.86	55.02	4	4	3092	-235	930	697
2.2	9.86	54.98	4	3	3090	-138	865	698
2.3	9.86	54.95	3	2.6	3089	-92	839	698
2.4	9.86	54.93	3	2.5	3087	-50	813	697
2.6	9.86	54.86	3	2.5	3086	33	771	697
2.2	39.44	55.38	9	5	3110	-795	1440	697
3.0	39.44	55.34	6	4	3108	-500	1157	697
5.0	39.44	55.22	3	2.5	3100	0	788	697
$\Delta S = 12$ (1/сек)								
3.0	39.44	44.24	6	4	2516	-423	935	570
4.0	39.44	44.19	4	3	2512	-192	750	570
5.0	39.44	44.13	3	2.5	2510	0	638	570

З даних досліджень виходить, що кращими результатами є такі: при формуванні частоти живлення ТАД з $\Delta S = 12$ (1/с) та $\Delta S = 18$ (1/с) – це значення $b=5.0$; $C=39.44$ ($\omega = 6.28$). Аналогічні дослідження проведено і для інших умов (рівня завантаженості, значення ΔS , моменту появи процесу буксування).

На рис. 5 приведено характер процесів електромагнітного моменту ТАД (рис. 5б)) в процесі буксування і без в залежності від впливу моменту зчеплення колісної пари з рейками (Мзч) шляхом використання специфіки функціонування ТАД при зміні частоти ковзання (рис. 5, а) за допомогою складової ($\omega_s - p\omega$) рівнянь (15) і (16).

В результаті зміни цієї складової згідно обраного закону є змога отримувати значення тягового

моменту, характерного для певного типу, який виникає в процесі буксування.

Як правило, у відомих моделях, момент інерції J_M колісної пари, що буксує, визначається у вигляді деякої константи і приймається рівним моменту інерції колісної пари.

Однак, це вірно тільки при повній втраті зчеплення колісної пари з рейкою. Якщо зчеплення $\psi_0 \neq 0$, то при утворенні моменту інерції J_M повинна враховуватися складова, яка залежить від маси поїзда, як і для випадку за відсутності буксування.

В моделі, що пропонується, цей процес враховується шляхом використання аперіодичної функції першого порядку вигляду $W(p) = \frac{1}{Tp+1}$, на вхід якої надходить сигнал ΔS_1 .

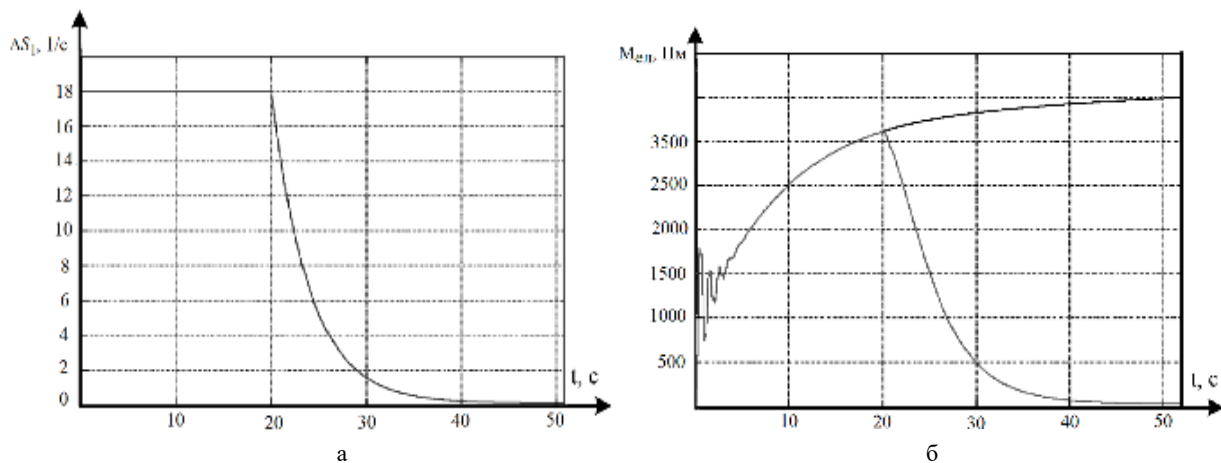
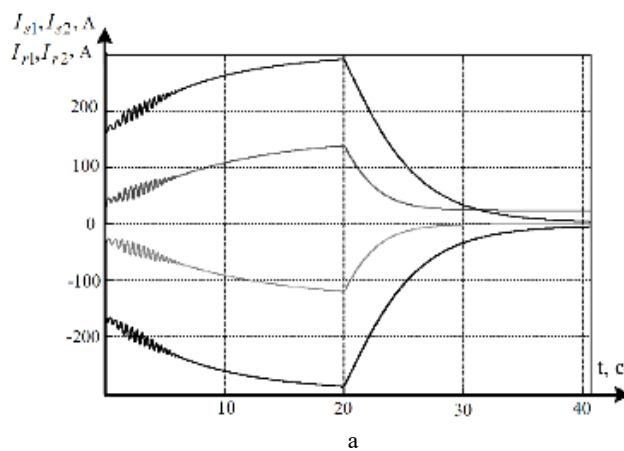


Рис. 5. Графіки зміни електромагнітного моменту ТАД

Згідно з цим формується співвідношення $\omega_{k1} = \omega_1 + \Delta S_1$ (зкореговане значення зворотного зв'язку кутової частоти обертання ротора буксуючого ТАД ω_1 ; ΔS_1 – складова, яка визначає темп зміни ω_{k1} , амплітудне значення якої в моделі обмежується заданою величиною ковзання ΔS). Це дозволяє формувати значення ковзання ΔS , яке використовується в математичній моделі ТАД в синхронній системі координат, в діапазоні від заданого до нульового.

В результаті отримуємо відповідну зміну електромагнітного моменту (рис. 5, б).

За допомогою постійної T передавальної функції можна задавати час протікання процесу буксування колісної пари дизель-поїзда в процесі моделювання.



На рис. 5, б наведено процес зміни електромагнітного моменту в процесі буксування при $T = 4$ с, який характерний при зчепленні колісної пари з рейками за умови $\psi_0 \neq 0$ відповідно до експоненційного закону. Шляхом вибору інших типів передавальних функцій можна отримувати перехідні процеси, тобто характер зміни електромагнітного моменту і в кінцевому випадку – значення швидкості дизель-поїзда.

На рис. 6 приведено графіки зміни фазного струму ТАД:

а – проекції векторів струмів статора і ротора на відповідні вісі координат;

б – фазний струм статора, що відповідають процесу формування тягового моменту при моделюванні згідно запропонованого підходу визначення та використання сигналу ковзання ΔS_1 зворотного зв'язку моделі ТАД.

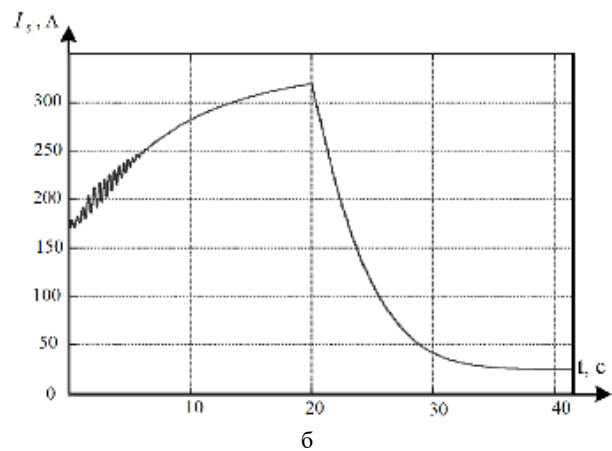


Рис. 6. Графіки зміни фазних струмів ТАД

Використовуючи дану модель є змога крім визначення електричних параметрів ГKM проводити дослідження в плані розроблення комп'ютерних компонентів для систем керування, що забезпечують ефективність функціонування електроприводу з ГKM в процесі буксування. Це в першу чергу використання можливості формування частоти ковзання відповідно до специфіки функціонування ТАД в моменти виникнення буксування, на основі інформації про характер зміни фазного струму ТАД – за значенням його похідної.

Висновки

1. Проведено аналіз математичних моделей електроприводу з ТАД з метою розроблення комп'ютерних компонентів для системи управління тяговим електроприводом за умови використання гумо-кордової муфти і можливих процесів буксування колісних пар дизель-поїзда з урахуванням ступеня завантаженості в процесі його розгону.

2. На основі аналітичного огляду систем автоматичного керування електропередачі дизель-поїзда з тяговим електроприводом змінного струму запропоновані математичні моделі електромеханічної системи дизель-поїзда з урахуванням пружних зв'язків за умови забезпечення якісних показників системи керування у його елементах в процесі

буксування та моделі, що описує можливий процес буксування.

3. Розроблені моделі дозволяють виконувати розрахунки електричних параметрів муфти з еластичними елементами та проводити дослідження на стадії проектування при синтезі системи регулювання тяговим електроприводом з використанням в процесі моделювання пакету MATLAB.

4. Приведені моделі електромеханічної системи дизель-поїзда з урахуванням зчеплення колісних пар з рейками та дисипативних сил ГKM можуть бути використані для дослідження системи автоматичного регулювання (САР) електропередачі та уточнення її параметрів, а також виконання коригування алгоритмів комп'ютерних компонентів САР, отриманих у результаті синтезу системи керування з урахуванням запобігання можливих процесів буксування.

5. На підставі проведених досліджень, як у статичних так і в динамічних режимах, за різних умов розгону дизель-поїзда встановлено, що розроблена модель з урахуванням запропонованого способу моделювання процесів буксування адекватно відображає процеси, що відбуваються в електромеханічній системі дизель-поїзда і може бути використана для розробки комп'ютерних компонентів для системи управління, що забезпечує режим ведення поїзда з

урахуванням умов зчеплення колісних пар з рейками.

6. Запропоновано моделювання ГKM при дослідженні процесів, що зв'язані з буксуванням, як двох масової системи, у вигляді лінійного диференційного рівняння другого порядку з постійними коефіцієнтами. Таке представлення математичної моделі дозволяє проводити дослідження в певному спектрі можливих змін особистих частот (періоду коливань) та швидкості їх затухання, що дає можливість в певному сенсі визначати електричні параметри ГKM.

7. Запропоновані моделі для дослідження процесів, що пов'язані з буксуванням, представлені в двох варіантах: безпосередньо через дискретну зміну електромагнітного моменту ТАД на певну величину з використанням і без використання ГKM і зміна електромагнітного моменту при використанні ГKM в залежності від впливу моменту зчеплення колісної пари з рейками шляхом використання специфіки функціонування ТАД – зміною частоти ковзання згідно заданого закону.

8. Проведено дослідження запропонованої моделі як з точки зору її адекватності, так і її можли-

вості застосування для цілей розроблення комп'ютерних компонентів для системи керування, що забезпечує ефективність функціонування електроприводу з ГKM в процесі буксування.

Результати досліджень надано у вигляді графіків процесів характерних фазових змінних та таблиць, в яких наведені числові значення характерних змінних системи керування в залежності від сигналів управління.

9. Використання сукупності розроблених моделей можливо як для визначення електричних параметрів ГKM, так і проведення досліджень в плані розроблення систем керування, що забезпечують ефективність функціонування електроприводу з ГKM в процесі буксування, використовуючи можливість формування частоти ковзання на основі показників зміни фазного струму ТАД.

10. Розроблені математичні і машинні моделі, за допомогою яких проведені дослідження з метою синтезу регуляторів і визначення значень їх параметрів, показали перспективність такого підходу при проектуванні сучасних систем керування електроприводами дизель-поїздів з тяговими асинхронними двигунами.

СПИСОК ЛІТЕРАТУРИ

1. Walrave, Ellwanger G. Hochgeschwindigkeitsverkehr in Europa // ETR – Eisenbahntechnische Rundschau. – 1995. – №10. – P. 136 – 142.
2. Shabana, A. A., Zaaza, K. E., & Sugiyama, H. (2017). Railway Vehicle Dynamics: A Computational Approach. CRC Press. Amazon.
3. Iwnicki, S. (Ed.). (2006). Handbook of Railway Vehicle Dynamics. CRC Press. CRC Press.
4. Olofsson, U., & Lewis, R. (2006). Adhesion in the Wheel-Rail Contact. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 220(4), 437–446. ResearchGate.
1. 5 Spiriyagin, M., Cole, C., & Sun, Y. (2014). Factors Affecting Wheel-Rail Adhesion. Wear, 314(1-2), 124–129. ScienceDirect.
5. Поповіч М.Г. Електромеханічні системи автоматичного керування та електроприводи: Навч. посібник / М.Г. Поповіч, О.Ю. Лозинський. – К.: Либідь, 2005. – 680 с.
6. Толочко, О.І. Моделювання електромеханічних систем. Математичне моделювання систем асинхронного електроприводу: навчальний посібник / О.І. Толочко. – Київ, НТУУ «КПІ», 2016. – 150 с.
7. Заполовський М.Й. Синтез управлінь для оптимізації динамічних процесів електроприводу змінного струму / М.Й. Заполовський, В.В. Скороделов, М.В. Мезенцев // Системи управління, навігації та зв'язку. – Вип. 4(50). Полтава. 2018. С. 38 – 41.
8. М.Й. Заполовський, М.В. Мезенцев, В.В. Скороделов. Математична модель для синтезу управлінь електроприводом змінного струму // Системи управління, навігації та зв'язку. – Випуск 5(57). Полтава. 2019. Стр. 16-21.

Received (Надійшла) 25.12.2024

Accepted for publication (Прийнята до друку) 26.02.2025

Modeling of the traction electric drive of a diesel train taking into account the use of a rubber-cord coupling during wheelset slippage

M. Zapolovskyi, M. Mezentsev, M. Olifir

Abstract. The issues of development and research of models for the AC electric drive of a diesel train are considered. A review of literary sources on the specified topic and an analysis of existing approaches to solving problems related to the development of control systems that ensure increased efficiency of the electric drive under operating conditions, including possible wheelset slippage, are conducted. Both generally accepted approaches to studying the processes of wheelset slippage through modeling and proposed approaches are examined. Computer components of the electromechanical system of the diesel train have been developed, taking into account elastic connections under the condition of ensuring quality indicators of the control system in its elements during slippage, as well as a model describing the possible slippage process. The presented models can be used to study the automatic regulation system of the electric transmission and to clarify its parameters, as well as to determine the parameters of the rubber-cord coupling (RCC) and subsequently implement them in the form of corresponding computer components of the diesel train control system. The proposed model has been investigated both in terms of its adequacy and its applicability for the development of a control system that ensures the efficiency of the electric drive with RCC during slippage. The research results are presented in the form of graphs of processes of characteristic phase variables and tables containing numerical values of characteristic variables of the control system depending on control signals.

Keywords: computer components, modeling of slippage processes, AC electric drive, diesel train, control systems, slippage, elastic connections, rubber-cord coupling, adequacy, slip frequency, process graphs, tables, control signals.

О. В. Запорожець¹, О. Р. Соболев¹, Д. В. Гриньов², І. С. Зиков²

¹ Харківський національний університет радіоелектроніки, Харків, Україна

² Національний технічний університет "Харківський політехнічний інститут", Харків, Україна

ІНТЕГРАЦІЯ СТАНДАРТІВ ISO В УПРАВЛІННЯ ПРОГРАМНИМИ ПРОЄКТАМИ

Анотація. Сучасні програмні проєкти відрізняються значною складністю та постійними змінами, що обумовлює необхідність застосування стандартизованих підходів до управління якістю та ефективністю. Міжнародні стандарти, зокрема ISO 9001, ISO/IEC/IEEE 12207 та ISO/IEC 33001, пропонують уніфіковані методики для забезпечення якості програмного забезпечення, управління життєвим циклом розробки та оцінки рівня зрілості процесів. У даній статті аналізуються значення цих стандартів у контексті програмних проєктів, їхнє впровадження в управлінські практики та можливості практичного використання. Особливо розглядається адаптація стандартів до сучасних методологій розробки, зокрема Agile, Waterfall та DevOps. Наведено приклади успішного застосування стандартів ISO у великих корпораціях та державних ініціативах, що демонструє їхню ефективність у підвищенні якості продуктів, мінімізації ризиків та збільшенні конкурентоспроможності програмних рішень.

Ключові слова: ISO, управління програмними проєктами, якість, інтеграція, ISO 9001, PDCA, ISO/IEC/IEEE 12207, ISO/IEC 33001, SPICE, Agile, Waterfall, DevOps, життєвий цикл.

Вступ

У сучасних умовах програмні проєкти стають дедалі складнішими та масштабнішими, що зумовлює необхідність застосування стандартизованих підходів до їхнього управління [1]. Якість програмного забезпечення є одним із ключових чинників, який безпосередньо впливає на комерційний успіх продукту та рівень задоволеності клієнтів. Для досягнення високої якості необхідно враховувати як чітко сформульовані, так і приховані вимоги зацікавлених сторін, а також забезпечувати ефективну комунікацію між розробниками та користувачами [2].

Стандарти ISO [3], розроблені Міжнародною організацією зі стандартизації, пропонують уніфіковані підходи до управління якістю, життєвим циклом програмного забезпечення та оцінки рівня зрілості процесів. Вони сприяють забезпеченню стабільної якості продуктів, зниженню ризиків та підвищенню ефективності роботи команд. Впровадження цих стандартів у процеси управління програмними проєктами є важливим кроком для досягнення відповідності вимогам клієнтів та нормативних вимог.

Метою даної статті є дослідження ролі стандартів ISO у управлінні програмними проєктами, визначення основних переваг їхньої інтеграції та формування практичних рекомендацій щодо їх застосування. У статті розглядаються ключові стандарти ISO, їх значення для програмних проєктів, процес впровадження та приклади практичного використання.

Стандарти ISO та їх роль у управлінні програмними проєктами

Стандарти ISO (Міжнародної організації зі стандартизації) [3] є фундаментальними інструментами для забезпечення якості, ефективності та стандартизації процесів у різних галузях, вклю-

чаючи управління програмними проєктами. Вони надають чіткі рамки, які допомагають організаціям планувати, реалізовувати, контролювати та вдосконалювати свої процеси. У контексті програмних проєктів стандарти ISO відіграють ключову роль у забезпеченні якості продукту, зниженні ризиків та підвищенні ефективності роботи команд.

ISO 9001 [4] є одним із найвідоміших стандартів, який визначає вимоги до систем управління якістю (СУЯ). Він застосовується до будь-яких організацій, незалежно від їх розміру чи галузі, і спрямований на забезпечення стабільної якості продуктів та послуг.

ISO 9001 базується на семи ключових принципах, які формують основу для ефективного управління якістю. Перший принцип – орієнтація на клієнта – підкреслює важливість розуміння потреб клієнтів та забезпечення їхнього задоволення. Другий принцип – лідерство – вимагає від керівництва активної ролі у створенні єдиної стратегії та цілей, спрямованих на якість. Третій принцип – залучення працівників – акцентує на важливості мотивації та участі всіх членів команди у досягненні спільних цілей. Четвертий принцип – процесний підхід – передбачає розгляд усіх діяльностей як взаємопов'язаних процесів, що дозволяє оптимізувати ресурси та покращувати результативність. П'ятий принцип – постійне вдосконалення – вимагає від організацій регулярного аналізу своїх процесів та впровадження змін для підвищення ефективності. Шостий принцип – прийняття рішень на основі фактів – підкреслює важливість використання даних та аналітики для обґрунтованого управління. Сьомий принцип – управління взаємовідносинами – спрямований на побудову довгострокових партнерських відносин із постачальниками та іншими зацікавленими сторонами.

Ці принципи формують основу для створення системи управління якістю, яка забезпечує стабі-

льність, прозорість та відповідність вимогам клієнтів.

Одним із ключових інструментів, який ISO 9001 пропонує для реалізації цих принципів, є метод PDCA (Plan-Do-Check-Act) [5]. Ця циклічна методика застосовується до будь-якого процесу в організації та включає чотири етапи:

1) Plan (Планування): визначення цілей, процесів та ресурсів, необхідних для досягнення бажаних результатів;

2) Do (Виконання): впровадження запланованих процесів та дій;

3) Check (Перевірка): моніторинг та аналіз результатів для оцінки їх відповідності поставленим цілям;

4) Act (Дія): впровадження коригувальних дій для покращення процесів та досягнення кращих результатів.

Цикл PDCA забезпечує постійне вдосконалення процесів, що є основним принципом ISO 9001. Він дозволяє організаціям систематично аналізувати свої дії, виявляти слабкі місця та впроваджувати покращення.

У програмних проєктах ISO 9001 відіграє ключову роль у забезпеченні якості продукту та ефективності процесів. Він допомагає командам визначати та керувати вимогами клієнтів, що особливо важливо на етапі планування проєкту. Завдяки чіткому визначенню процесів та відповідальності, ISO 9001 дозволяє уникнути недоліків у розробці, таких як невідповідність вимогам або низька якість коду.

Метод PDCA, який є основою ISO 9001, знаходить широке застосування в управлінні програмними проєктами. Наприклад, на етапі планування (Plan) команда визначає вимоги до програмного продукту, розробляє архітектуру та складає план робіт. На етапі виконання (Do) відбувається розробка програмного забезпечення, включаючи кодування та інтеграцію компонентів. На етапі перевірки (Check) проводиться тестування продукту для виявлення дефектів та оцінки відповідності вимогам. На етапі дії (Act) команда впроваджує коригувальні заходи, такі як виправлення помилок або оптимізація коду, що забезпечує постійне вдосконалення продукту.

Крім того, ISO 9001 сприяє покращенню комунікації між командами, що особливо важливо в умовах розподілених команд або при використанні методологій Agile. Завдяки чіткому документуванню процесів та результатів, всі учасники проєкту мають доступ до актуальної інформації, що зменшує ризики непорозумінь та помилок.

Стандарт ISO/IEC/IEEE 12207 [6] описує процеси життєвого циклу програмного забезпечення, включаючи планування, розробку, тестування, впровадження та підтримку. Він забезпечує структурований підхід до управління проєктами, що особливо важливо для великих та складних програмних систем.

Основні процеси ISO/IEC/IEEE 12207 [6]:

Перша категорія – процеси узгодження. Сто-

суються взаємодії між організаціями, які виступають у ролі замовників (acquirers) та постачальників (suppliers) програмних систем. Ці процеси забезпечують укладення угод, які дозволяють обох сторонам отримувати вигоду та підтримувати бізнес-стратегії. Узгоджувальні процеси можуть застосовуватися як поза межами життєвого циклу проєкту, так і під час його реалізації. Вони також можуть використовуватися всередині однієї організації для визначення відповідальностей між різними функціональними підрозділами.

Ці процеси забезпечують чітке розуміння цілей проєкту, вимог до продукту чи послуг, а також розподіл обов'язків між замовником і постачальником. Наприклад, процес придбання (acquisition process) включає визначення вимог до програмного продукту та вибір постачальника, тоді як процес постачання (supply process) забезпечує відповідність продукту вимогам замовника.

Друга категорія – організаційні процеси сприяння проєкту. Спрямовані на забезпечення ресурсів, необхідних для реалізації проєктів, які відповідають потребам та очікуванням зацікавлених сторін. Ці процеси діють на стратегічному рівні та стосуються управління бізнесом організації, розподілу ресурсів, управління ризиками та покращення процесів.

Організаційні процеси створюють середовище, в якому реалізуються проєкти. Вони включають встановлення процесів та моделей життєвого циклу, надання ресурсів (людських, фінансових тощо), а також визначення та моніторинг показників якості програмних систем та інших результатів проєктів. Ці процеси однаково важливі як для комерційних, так і для неприбуткових організацій, оскільки всі вони несуть відповідальність перед зацікавленими сторонами та стикаються з ризиками у своїй діяльності.

Третя категорія – процеси технічного управління. Стосуються управління ресурсами та активами, виділеними організацією, для виконання угод та досягнення цілей проєкту. Вони включають планування витрат, термінів та результатів, контроль виконання заходів для забезпечення відповідності планам та критеріям ефективності, а також виявлення та вибір корективних дій для усунення недоліків у прогресі та досягненнях.

Ці процеси використовуються для створення та виконання технічних планів проєкту, управління інформацією в команді, оцінки технічного прогресу щодо планів розробки програмних систем, продуктів чи послуг, а також для підтримки прийняття рішень. Процеси технічного управління застосовуються під час виконання кожного технічного процесу.

Четверта категорія – технічні процеси. Охоплюють технічні дії, які виконуються протягом усього життєвого циклу програмного забезпечення. Вони трансформують потреби зацікавлених сторін у продукт чи послугу, які забезпечують стабільну роботу та задоволення вимог клієнтів.

Технічні процеси застосовуються для ство-

рення та використання програмних систем, незалежно від їх форми (модель чи готовий продукт) та етапу життєвого циклу. Вони включають аналіз бізнес-потреб, визначення вимог, проектування архітектури, розробку, тестування, впровадження, верифікацію, валідацію та підтримку програмного продукту.

ISO/IEC 33001 [7] є основним стандартом у серії ISO/IEC 33000, яка замінила застарілий ISO/IEC 15504 (SPICE). Цей стандарт надає рамки для оцінки зрілості та результативності процесів, що дозволяє організаціям визначати слабкі місця та впроваджувати покращення.

Рівні зрілості SPICE [7]:

На нульовому рівні – неповне виконання процесів – організація не має чітко визначених процесів, і результати роботи залежать від індивідуальних зусиль працівників.

На першому рівні – виконані процеси – організація впроваджує базові процеси, але вони не є повністю керованими.

На другому рівні – керовані процеси – організація визначає стандарти та процедури для управління процесами, що дозволяє забезпечити стабільність результатів.

На третьому рівні – встановлені процеси – організація інтегрує процеси у свою діяльність та використовує метрики для їхнього моніторингу.

На четвертому рівні – передбачувані процеси – організація досягає високого рівня контролю над процесами, що дозволяє прогнозувати результати та впроваджувати проактивні покращення.

На п'ятому рівні – оптимізовані процеси –

організація постійно вдосконалює свої процеси, використовуючи інноваційні підходи та технології.

Рівні зрілості допомагають організаціям оцінити свої процеси, визначити слабкі місця та впровадити покращення для досягнення вищих рівнів ефективності.

Схема стандартів ISO у управлінні програмними проектами (рис. 1) може бути використана для структурування процесів, визначення ключових вимог та забезпечення якості на всіх етапах життєвого циклу проекту. Визначені стандарти, такі як ISO 9001, ISO/IEC/IEEE 12207 та ISO/IEC 33001, можуть стати основою для планування, моніторингу та вдосконалення процесів, забезпечуючи чіткий контроль над якістю продукту та ефективністю роботи команди.

Ця схема може слугувати контрольним переліком для оцінки поточного стану процесів, виявлення слабких сторін та визначення необхідних покращень.

Вона забезпечує основу для оцінки зусиль, спрямованих на досягнення вищих рівнів зрілості процесів, а також для забезпечення відповідності вимогам клієнтів та нормативних органів.

Крім того, схема дозволяє формувати метрики для оцінки результативності процесів, що дає можливість приймати обґрунтовані рішення щодо впровадження змін та оптимізації використання ресурсів. Використання цієї схеми забезпечує системний підхід до управління якістю, що є ключовим фактором для успішної реалізації програмних проектів.

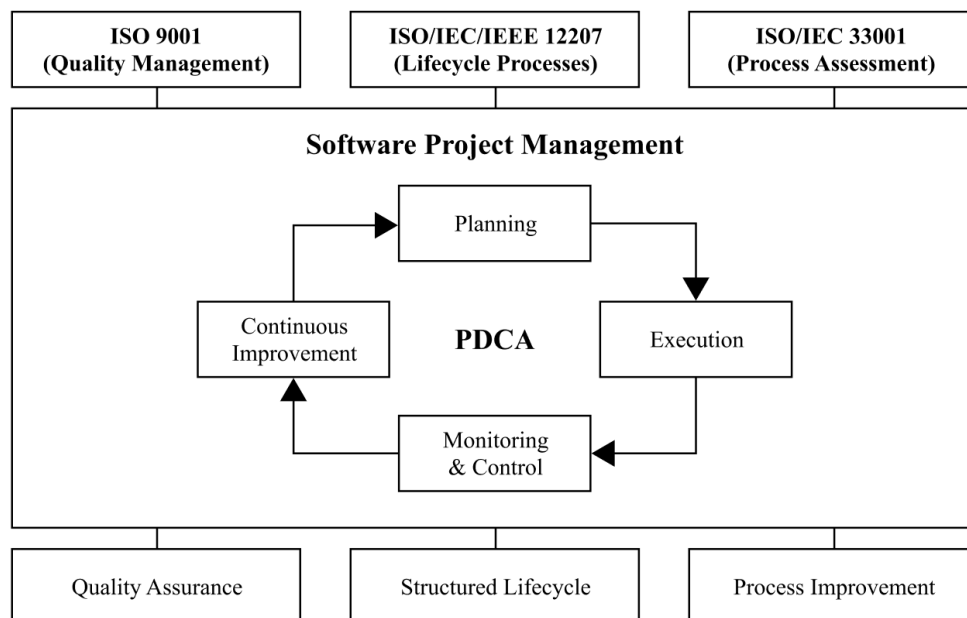


Рис. 1. Схема стандартів ISO в управлінні програмними проектами

Інтеграція стандартів ISO в управління програмними проектами

Розглянемо основні етапи інтеграції стандартів ISO [8]:

1. Аналіз вимог стандартів.

Першим кроком інтеграції є вивчення вимог стандартів ISO, які можуть бути застосовані до програмного проекту. Наприклад:

– ISO 9001 визначає вимоги до системи уп-

равління якістю, включаючи орієнтацію на клієнта, постійне вдосконалення та управління ризиками;

- ISO/IEC/IEEE 12207 описує процеси життєвого циклу програмного забезпечення, такі як планування, розробка, тестування та підтримка;

- ISO/IEC 33001 надає методику оцінки зрілості процесів, що дозволяє виявити слабкі місця та впровадити покращення.

На цьому етапі важливо визначити, які стандарти є найбільш актуальними для конкретного проекту, враховуючи його масштаб, складність та специфіку.

2. Адаптація стандартів до методологій розробки.

Стандарти ISO можуть бути інтегровані в різні методології розробки програмного забезпечення, такі як Agile, Scrum, Waterfall або DevOps. Наприклад:

- у Agile [9] принципи ISO 9001 можуть бути використані для забезпечення якості на кожному етапі розробки, включаючи регулярний зворотний зв'язок із клієнтами та постійне вдосконалення процесів;

- у Waterfall [6] стандарт ISO/IEC/IEEE 12207 може бути використаний для чіткого структурування етапів життєвого циклу програмного забезпечення, що дозволяє уникнути помилок на завершальних етапах проекту;

- у DevOps [7] стандарт ISO/IEC 33001 може допомогти оцінити зрілість процесів автоматизації та інтеграції, що дозволяє підвищити ефективність розгортання програмного забезпечення.

Адаптація стандартів до методологій розробки дозволяє забезпечити їхню відповідність конкретним умовам проекту та максимізувати їхню ефективність.

3. Впровадження інструментів управління якістю.

Для забезпечення відповідності стандартам ISO необхідно використовувати спеціалізовані інструменти управління якістю. До них належать:

- системи відстеження дефектів (наприклад, Jira, Bugzilla) для виявлення та усунення помилок на ранніх етапах розробки;

- автоматизовані тестові середовища (наприклад, Selenium, Jenkins) для забезпечення якості коду та швидкого виявлення проблем;

- системи управління документацією (наприклад, Confluence, SharePoint) для забезпечення про-

зорості та відстеження змін у вимогах і процесах.

Використання цих інструментів дозволяє автоматизувати багато аспектів управління якістю, що значно підвищує ефективність роботи команди.

4. Навчання та підготовка команди.

Успішна інтеграція стандартів ISO вимагає навчання та підготовки команди. Учасники проекту повинні розуміти вимоги стандартів та вміти застосовувати їх у своїй роботі. Наприклад:

- проведення тренінгів з основ ISO 9001 для забезпечення розуміння принципів управління якістю;

- навчання команди використанню інструментів управління якістю, таких як системи відстеження дефектів або автоматизовані тестові середовища;

- розробка внутрішніх документів та інструкцій, які описують процеси, відповідні стандартам ISO.

Навчання команди є критично важливим етапом, оскільки саме люди є основним джерелом знань та досвіду, необхідних для успішної інтеграції стандартів.

5. Моніторинг та вдосконалення процесів.

Інтеграція стандартів ISO є постійним процесом, який вимагає регулярного моніторингу та вдосконалення. Наприклад:

- проведення аудитів для оцінки відповідності процесів вимогам стандартів;

- використання метрик для оцінки результативності процесів, таких як кількість дефектів, час виправлення помилок або рівень задоволеності клієнтів;

- впровадження змін на основі результатів аудитів та аналізу метрик.

Постійне вдосконалення процесів дозволяє забезпечити їхню відповідність вимогам стандартів та підвищити ефективність роботи команди.

Схема основних етапів інтеграції стандартів ISO (рис. 2) в управління програмними проектами представляє собою структурований підхід, який дозволяє організаціям систематично впроваджувати міжнародні стандарти з метою покращення якості та ефективності своїх процесів. Ця схема включає п'ять ключових етапів: аналіз вимог стандартів, адаптацію до методологій розробки, впровадження інструментів управління якістю, навчання персоналу та моніторинг із подальшим вдосконаленням процесів.

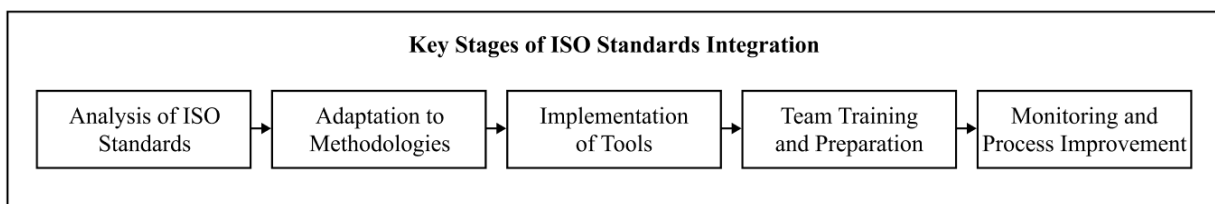


Рис. 2. Схема основних етапів інтеграції стандартів ISO

Кожен із етапів схеми є основою для визначення конкретних заходів, необхідних для успішного

впровадження стандартів ISO. Наприклад, на етапі аналізу вимог визначаються найбільш відповідні

стандарти (такі як ISO 9001, ISO/IEC/IEEE 12207, ISO/IEC 33001), які будуть застосовуватися в рамках проекту. Адаптація до методологій розробки дозволяє забезпечити відповідність стандартів специфічним умовам проекту, наприклад, при використанні Agile, Scrum або Waterfall.

Впровадження інструментів управління якістю, таких як системи відстеження помилок (Jira, Bugzilla) та автоматизовані платформи для тестування (Selenium, Jenkins), забезпечує ефективний контроль якості на всіх етапах розробки. Навчання команди сприяє засвоєнню принципів стандартів ISO та їх практичного застосування. На завершальному етапі, моніторинг і вдосконалення процесів дозволяють постійно оцінювати ефективність інтеграції та впроваджувати необхідні корективи.

Ця схема може слугувати інструментом для планування, контролю та оцінки зусиль, спрямованих на інтеграцію стандартів ISO. Вона забезпечує чітку структуру для виявлення слабких сторін, визначення пріоритетів та впровадження покращень, що є ключовим для досягнення високої якості програмних продуктів та підвищення ефективності роботи команди.

Практичне використання стандартів ISO у програмних проєктах

Стандарти ISO знайшли широке застосування у програмних проєктах завдяки їхній здатності забезпечувати якість, ефективність та відповідність вимогам клієнтів і регуляторних органів. Практичне використання цих стандартів дозволяє організаціям [10] мінімізувати ризики, покращити процеси розробки та забезпечити стабільну якість програмних продуктів.

1) Застосування ISO 9001 для забезпечення якості [4].

ISO 9001 є основним стандартом для систем управління якістю, який активно використовується у програмних проєктах. Його практичне використання передбачає визначення вимог клієнтів, контроль якості на всіх етапах розробки та постійне вдосконалення процесів.

На етапі планування проєкту команда визначає вимоги клієнтів та документує їх у вигляді техніч-

них специфікацій. Це допомагає уникнути непорозумінь і забезпечити відповідність продукту очікуванням замовників. Під час розробки програмного забезпечення ISO 9001 вимагає впровадження процесів контролю якості, таких як автоматизоване тестування для виявлення дефектів на ранніх стадіях. Після завершення кожного етапу проєкту проводиться ретроспектива, де команда аналізує помилки та визначає шляхи їх усунення.

2) Використання ISO/IEC/IEEE 12207 для управління життєвим циклом [6].

ISO/IEC/IEEE 12207 визначає процеси життєвого циклу програмного забезпечення, що робить його незамінним для великих та складних проєктів. Його практичне застосування включає планування життєвого циклу, управління конфігурацією та підтримку користувачів.

На етапі планування команда визначає етапи життєвого циклу, такі як аналіз вимог, проектування, розробка, тестування, впровадження та підтримка. Це дозволяє забезпечити чітку структуру проєкту та уникнути хаосу. Для управління конфігурацією використовуються системи контролю версій, такі як Git або SVN, що дозволяє відстежувати зміни у коді та документації. Після впровадження продукту команда забезпечує підтримку користувачів, включаючи виправлення помилок, оновлення та надання технічної допомоги.

3) Використання ISO/IEC 33001 для оцінки зрілості процесів [7].

ISO/IEC 33001 надає методіку для оцінки зрілості процесів розробки програмного забезпечення. Його практичне застосування включає оцінку поточного стану процесів, впровадження покращень та моніторинг результативності.

Організація проводить оцінку своїх процесів, використовуючи шість рівнів зрілості, визначених у стандарті. Це дозволяє виявити слабкі місця та визначити пріоритети для покращення. На основі результатів оцінки організація впроваджує зміни, спрямовані на підвищення ефективності процесів. Після впровадження змін проводиться повторна оцінка, щоб визначити ефективність змін.

У табл. 1 наведено приклади використання стандартів ISO у програмних проєктах.

Таблиця 1 – Приклади використання стандартів ISO у програмних проєктах

Стандарт	Компанія	Сфера застосування	Результат
ISO 9001	Microsoft	Управління якістю продуктів (Windows, Office)	Забезпечення стабільної якості продуктів та відповідність вимогам мільйонів користувачів.
ISO/IEC/IEEE 12207	Державні проєкти	Розробка систем електронного урядування	Забезпечення надійності та стабільності програмного забезпечення.
ISO/IEC 33001	IBM	Оцінка зрілості процесів розробки програмного забезпечення	Підвищення ефективності роботи команди та забезпечення високої якості продуктів.

Висновки

Інтеграція стандартів ISO у управління програмними проектами є важливим кроком для забезпечення якості, ефективності та відповідності вимогам клієнтів і регуляторних органів.

Стандарти ISO, такі як ISO 9001, ISO/IEC/IEEE 12207 та ISO/IEC 33001, забезпечують структурований підхід до планування, реалізації та контролю програмних проєктів.

ISO 9001 дозволяє забезпечити стабільну якість програмних продуктів через впровадження системи управління якістю, що включає:

- визначення вимог клієнтів,
- контроль якості на всіх етапах розробки,
- постійне вдосконалення процесів.

ISO/IEC/IEEE 12207 надає чітку структуру для управління життєвим циклом програмного забезпечення, що особливо важливо для великих та складних проєктів.

ISO/IEC 33001 дозволяє оцінити зрілість процесів розробки та впровадити покращення для підвищення ефективності роботи команди.

На практиці стандарти ISO широко використовуються у великих компаніях, таких як Microsoft та IBM, а також у державних проєктах, що підтверджує їхню ефективність у забезпеченні якості та конкурентоспроможності програмних продуктів.

Використання цих стандартів дозволяє забезпечити стабільну якість програмних продуктів, знизити ризики та підвищити ефективність роботи команди.

Таким чином, інтеграція стандартів ISO є невід'ємною частиною сучасного управління програмними проєктами, що дозволяє забезпечити якість продукту, знизити ризики та підвищити ефективність роботи команди.

СПИСОК ЛІТЕРАТУРИ

1. Лузан І. В., Луценко І. С. Система управління якістю як фактор підвищення конкурентоспроможності підприємства. Актуальні проблеми економіки та управління: збірник наукових праць молодих вчених. 2012. Вип. 6. С. 1-6. URL: http://nbuv.gov.ua/UJRN/eiou_2014_3-4_30.
2. Інтеграція систем управління проєктами в систему організації на різних етапах розвитку. КіберЛенінка. [Електронний ресурс]. – Режим доступу: <https://cyberleninka.ru/article/n/integratsiya-sistem-upravleniya-proektami-v-sistemu-organizatsii-na-raznyh-etapah-razvitiya>.
3. ISO: Global standards for trusted goods and services.
4. ISO 9001:2015 – Quality management systems – Requirements.
5. Цикл Демінга (The Deming Cycle), він же - Цикл Шухарта, Цикл PDCA / PDSA [Електронний ресурс]. – Режим доступу: <https://www.maxzsim.com/deming-cycle-pdca/>.
6. ISO/IEC/IEEE 12207:2017 – Systems and software engineering – Software life cycle processes.
7. ISO/IEC 33001:2015 – Information technology – Process assessment.
8. 7 Steps to Follow for ISO Management System Implementation [Електронний ресурс]. – Режим доступу: <https://quality-assurance.com.au/blog/7-steps-to-follow-for-iso-management-system-implementation/>.
9. Agile підхід до управління якістю [Електронний ресурс]. – Режим доступу: <https://www.dqsglobal.com/uk-ua/navchajtesya/centr-znan%27-dqs/agile-upravlinnya-yakistyu>.
10. Лірко Т. В. Інтеграція ризик-орієнтованого управління проєктами в сучасні бізнес-моделі. Бізнес Інформ. 2024. №1. С. 383–391. [Електронний ресурс]. – Режим доступу: https://www.business-inform.net/export_pdf/business-inform-2024-1_0-pages-383_391.pdf.

Received (Надійшла) 22.12.2024

Accepted for publication (Прийнята до друку) 05.03.2025

Integrating ISO standards into software project management

O. Zaporozhets, O. Sobol, D. Grynev, I. Zykov

Abstract. Modern software projects are characterized by significant complexity and constant changes, which necessitates the use of standardized approaches to quality and efficiency management. International standards, in particular ISO 9001, ISO/IEC/IEEE 12207 and ISO/IEC 33001, offer unified methodologies for ensuring software quality, managing the development life cycle and assessing the level of process maturity. This article analyzes the significance of these standards in the context of software projects, their implementation in management practices and the possibilities of practical use. The adaptation of standards to modern development methodologies, in particular Agile, Scrum and Waterfall, is separately considered. Examples of the successful application of ISO standards in large corporations and government initiatives are given, which demonstrates their effectiveness in improving product quality, minimizing risks and increasing the competitiveness of software solutions.

Keywords: ISO, software project management, software quality, ISO 9001, PDCA, ISO/IEC/IEEE 12207, ISO/IEC 33001, Agile, Scrum, Waterfall, life cycle, process maturity.

A. Kapiton¹, T. Franchuk², D. Tyshchenko³, A. Desiatko³, N. Sas⁴

¹ National University «Yuri Kondratyuk Poltava Polytechnic», Poltava, Ukraine

² National Academy of Statistics, Accounting and Audit, Kyiv, Ukraine

³ State University of Trade and Economics, Kyiv, Ukraine

⁴ Instituto Federal do Parana, Parana, Brazil

MODELING OF MANAGEMENT OBJECTS AND PROCESSES

Abstract. The article examines the main issues of modeling objects and management processes, analyzes and improves the methodological apparatus of the research. It has been proven that the majority of researchers study the system as an oriented and in a certain way directed set of elements, between which mutually determined and mediated connections of the studied area are established in a certain way. It was established that the system built in a certain way functions to achieve a defined and set goal, which is determined by the elements that are the components of this system and determined by the components of its theory considered in the study, using all levels of its abstract description. It is noted that at the theoretical-multiple level of its abstract description, with the optimal formation of all its components, the real state and prospects of its development and improvement are determined, which require the further perspective of formation, modernization and implementation of real systems. It is justified that, provided that business processes are correctly modeled, they provide an opportunity to increase efficiency and optimize the activities of the entire company, which is the main function of modeling and its practical utility. It is established that after the modeling process itself, an improvement plan must be developed and changes implemented to successfully increase the efficiency of all processes. The research of the software market made it possible to determine that today there is a sufficiently high level of services and platforms for modeling and designing business processes: from basic graphic editors that allow users to visualize the algorithm, to complex applications with analytical tools, constructors and other functions. Building a business process model is a useful method that allows you to effectively analyze, improve and optimize the company's business processes to achieve strategic goals and increase process automation and competitiveness. It is substantiated that the modeling of business processes is indirectly an organized approach to the construction of a business process, with its gradual analysis and interpretation.

Keywords: system, model, modeling, management processes, analysis, business processes.

Introduction

The company's activities consist of many business processes that are connected to each other into a single mechanism. Work productivity depends on the configuration and competent development of business processes, as well as the orientation of the business towards high results. Therefore, if a result is needed, you should regularly analyze, refine, and sometimes even rebuild systems and some processes within an organization or enterprise. Business process management can be improved and implemented using special techniques such as modeling. Particular attention should be paid to the methodology of business process modeling (BPM).

It will help to clarify why it is important to model business processes, what advantages it provides, what tasks and goals it allows to achieve. Types, stages, and BPM tools deserve detailed analysis. Through the modeling process, companies can identify their weaknesses, identify risks, optimize business processes and improve overall performance. The basic principle is to decompose cumbersome technological processes into small components. A detailed description of the elements and parts of various business processes allows in any situation to more deeply understand their essence, identify the potential for their improvement, and application in other areas of business.

The model is a textual, schematic, graphical representation of the internal processes of a business and the logical interdependence between them. Such

adequate models can and should reflect: the current state of processes - how it is happening now; their desired position after making modifications, adjustments - what will happen afterwards, the result.

Analysis of recent research and publications.

The problem of analyzing the evolution of the development of artificial intelligence in the direction of multimodality and transformational development in large language models (llm) has always been in the scientific focus of leading foreign and domestic scientists. It was the study of verification and interaction of information systems with the surrounding world that caused, according to scientists, this evolution of views regarding the perception of certain results that are inherently multimodal and multicomponent.

Analysis of the improvement of llm with the help of the ability to process and generate different data modalities in the field of view of a number of foreign scientists.

Poberezhny R., Kramskoi R., Pererva P. research methods of modeling the processes of management of socio-economic systems [1].

Vydrya E., Kobeleva O., Pererva R. focused on the studied structures of the enterprise's strategic innovation project and examine the content and contents of its modules [2].

Famous researchers consider the modeling processes, based on the analysis of the evolution and distribution of multimodal artificial intelligence systems and their influence on the development of the studied issues [3-6].

Others focused on the processes of modeling business processes in individual software applications, where it determines their features, advantages and disadvantages [7].

A number of modern publications are devoted to the study of best practices in the design, development and implementation of business process management systems [8-15].

Main part

Modeling is a way of researching any phenomena, processes or objects by building and analyzing their components. The term "system" has existed for more than two millennia, however, different researchers define it differently. The system can be one computer, and an automatic line or technological process in which the computer is only one of the components, and the entire enterprise or several different enterprises functioning as a single system in one industry. What one researcher defines as a system may be only a component of a more complex system for another [1-3, 5].

Based on the axiomatic statement regarding the relevance of determining that the main issues of modeling objects and management processes, there is a need for a thorough analysis and improvement of the methodological and methodical research apparatus. It should be noted that the majority of researchers study the system as an oriented and in a certain way directed collection of many elements. Note that the system functions under certain conditions, if necessary, its modification and expansion to achieve a defined and set goal, which is determined by the elements that are the components of this system, and determined by the components of its theory, which is considered in the study, using all levels of its abstract description. It is at the theoretical-multiple level of its abstract description, with the optimal formation of all its components, that the real state and prospects for its development and improvement are determined, which require further prospects for the formation, modernization and implementation of real systems.

Many researchers define a system as a purposeful set of interconnected elements of any nature. According to this definition, a system functions to achieve some goal.

This definition is quite correct for sociological and technical systems, but is not suitable for systems of the surrounding nature, the purpose of which is not always known. The definition of the concept of a system is related to an abstract theory systems, within which the following levels of abstract description are used:

- symbolic or linguistic;
- theoretical-multiple;
- abstract-algebraic;
- topological;
- logical and mathematical;
- theoretical and informational;
- dynamic;
- heuristic [4-7].

To define a certain concept, certain symbols (alphabet) are used and rules for operating them are established. A concept expressed in an abstract language means any sentence (formula) built according to the grammatical rules of this language. It is assumed that such a sentence contains selectable variables, so-called constituents, which, having only certain values, make this statement true. All statements are divided into two types: terms which point to objects, and functors that define relations between terms. The use of terms and functors makes it possible show how, based on the linguistic level, other levels can be formed abstract description of the system.

At the set-theoretic level of the abstract description of the system, it is possible to obtain fairly general information about real systems, and for specific purposes, other models are needed, which would provide an opportunity to analyze various properties of real systems in more detail. This requires lower levels of abstract description of systems, which are special cases of description of the set-theoretic level. So, if the connections between the elements of the given sets are established with the help of some single-valued functions that map the elements of the set to the original set itself, then the abstract-algebraic level of system description takes place.

In such cases, it is considered that zero, unary, binary, ternary and other relations are established between the elements of the set. If some multivalued functions are defined on these sets, then there are topological abstract models written in the language of general topology or its branches, which are called algebra topology, homologous topology, etc. p. Choosing the right level of abstract description when studying one or another real system is always the most responsible and the most difficult step in theoretical-system constructions.

This process is almost impossible to formalize and largely depends on the experience and knowledge of the researcher, his professional training, research goals, etc. [12-14].

The theory of relativity, which studies the universal physical laws that apply to the entire universe, and quantum mechanics, which studies the laws of the microcosm, are not easy to understand, and yet they deal with systems that are considered simple from the point of view of modern natural science.

Simple in the sense that they include a small number of variables, and therefore the relationship between them lends itself to mathematical processing and the derivation of universal laws. However, in addition to simple ones, there are complex systems that consist of a large number of variables and, therefore, a large number of different relationships between them.

The larger it is, the more difficult the subject of research is to achieve the final result - to deduce the regularities of the functioning of this object. The difficulties of studying these systems are also related

to the fact that the more complex the system, the more it has so-called emergent properties, that is, properties that are not present in its parts and which are a consequence of the effect of the integrity of the system.

Meteorology studies complex systems in particular, since the processes of weather formation are much less known than gravitational processes, which, at first glance, seems like a paradox.

In general, during the construction of the model, the following requirements must be taken into account:

- independence of the results of problem solving from the specific physical interpretation of the model elements; meaningfulness, that is, the ability of the model to reflect important features and properties of the real process being studied and modeled;

- deductibility;
- inductance.

Since the model is created to solve specific tasks, the model developer must be sure that he will not get absurd results, and that all the obtained results will reflect the characteristics and properties of the modeled system necessary for the researcher. System analysts use the model to make decisions and find the best ways to build or upgrade the modeled system. It is always necessary to remember that the customer is the user of the information obtained with the help of the model.

It is pointless to develop a model if it cannot be used. Moreover, working with the model should be automated for the customer to such an extent that he can work with it within his subject area. Thus, an advanced interface must be implemented between the model and the user, which is usually created using a menu system configured to use the model in a specific area. In order to implement all these methods in practice, you will need a business modeling system [8-11,14-15].

The degree of detail of the model should be chosen taking into account the goals of modeling, the possibility of obtaining the necessary input data for the model and taking into account the available resources for its creation.

The degree of detail of the model should be chosen taking into account the goals of modeling, the possibility of obtaining the necessary input data for the model and taking into account the available resources for its creation.

There are several conventional ways to create, construct and visualize a picture of processes within the framework of system modeling:

- tables,
- diagrams,
- development scenarios,
- description diagrams and so on.

Which method of describing business processes to choose depends on the preferences, objectives of the company and how many objects need to be studied.

Basic business modeling methods looks as shown in Fig. 1.

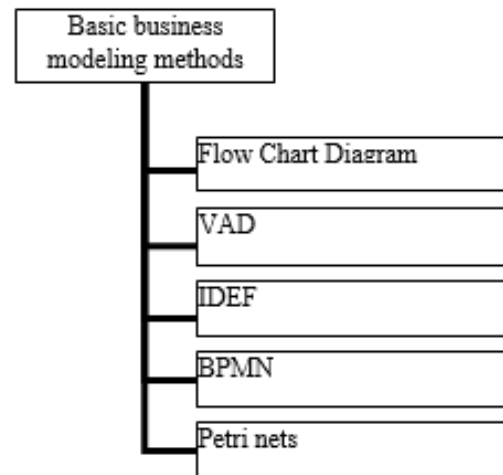


Fig. 1. Basic business modeling methods

If a correctly model a business process, it can help improve its efficiency and optimize the activities of the entire company. This is the main function of modeling and its practical benefits. If we consider a business not as something single and integral, but decompose its work into its component parts, then you will see a number of problems. At first glance, they may be understandable or even invisible, but in fact they greatly influence the work of the company. After the modeling process itself, it is necessary to develop an improvement plan and implement changes to successfully improve the efficiency of all processes. This means creating a model of existing interacting processes helps companies understand how they actually function. Therefore, under the condition of optimal and verified modeling of business processes, they provide an opportunity to increase the efficiency and optimize the activity of the entire company, which is the main function of modeling and its practical benefit. After passing the propaedeutic level of training, after the modeling process itself, there is a need to define and implement an improvement plan and implement changes to successfully improve the efficiency of all processes. Including finding solutions that help improve the work process. Processes mediated by BPM deserve analysis and more detailed research and looks as shown in Fig. 2.

Functional modeling involves dividing a large business process into functions, tasks and operations. In this approach, each function or task is represented as a separate element that is linked to other parts of the process itself. To interpret and describe this type of modeling, flowcharts are often used to observe exactly how the process works technically and how the process elements are interconnected. Simulation modeling of business processes allows you to focus on creating a model that, in theory, simulates the flow of a process under different conditions. Changes in the functions of the process and its properties are necessarily analyzed in dynamics. In the object-oriented modeling approach, business processes can be considered as a collection of objects that interact with each other to achieve certain goals.

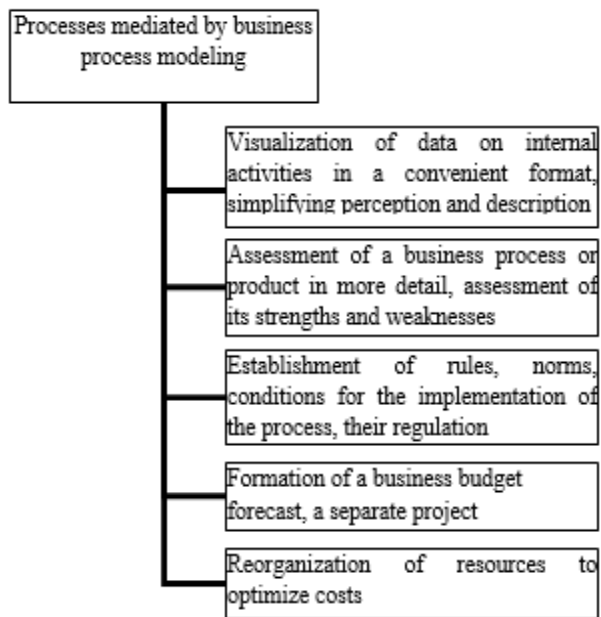


Fig. 2. Processes mediated by BPM

Each business process is a set of objects that has its own attributes and qualities.

They can be related to each other, and object-oriented modeling describes such relationships. All described types must be implemented using one of the three main methods of modeling business processes looks as shown in Fig. 3.

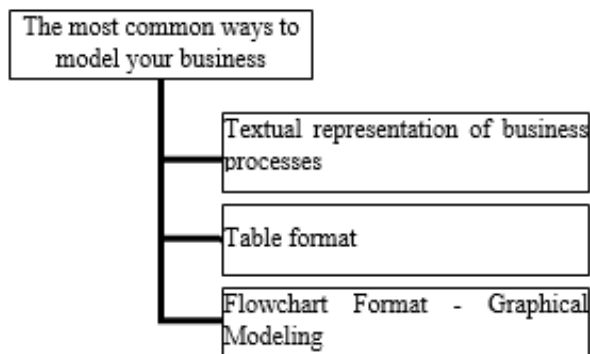


Fig. 3. The most common ways to model your business processes

The modeling process includes a number of stages, each of which must be performed sequentially, within the framework of a specific algorithm.

To make an effective business process model, you should go through the following stages:

- studying the current state of the process,
- analyzing the information received,
- developing the desired model,
- testing the updated model, finalizing the new model.

At the first stage, an analysis of the current state and stage of execution of the business process is carried out.

The stage includes searching for information: defining the goals and objectives of the process,

including the logic of its work, identifying the necessary key resources, means, and boundaries. This is the original or standard business model that will form the basis for a group of future improvements.

The second stage is the analysis of the information received, which involves the disclosure and addition of the process. It is necessary to search and identify restrictions, repetitive operations, and transitions.

Thus, at this moment we clarify, detail the “standard” model, reveal non-obvious details, functions and record the needs for changes. At the stage of developing the desired model, a model of an optimized business process is built. Based on the data obtained earlier, based on the opinions of experts and analysts, a new model is created in which the identified problems are eliminated. At the next stage, the new model is tested in real conditions, performs all functions and is introduced into the company’s activities.

After this, if necessary, it should be assessed and the results of business activities identified. If defects or errors appear during actual operation of the model, it is further refined taking into account new data, changes in external conditions, requirements and goals of the company, and a new model is finalized.

As a result of the study and analysis of the software market, it can be confidently stated that today there is a sufficiently high level of services and platforms for modeling and designing business processes: from basic graphic editors that allow users to visualize the algorithm, to complex applications with analytical tools, constructors and others functions. In our opinion, designing, developing and building a business process model is a useful method that allows you to effectively analyze, improve and optimize the company's business processes to achieve strategic goals and increase process automation and competitiveness.

As a result of the conducted analysis, it can be stated that business process modeling is an indirectly organized approach to building a business process with its gradual analysis and interpretation.

Conclusions

Today on the software market you can find dozens of services and platforms for modeling and designing business processes:

- from basic graphic editors that allow you to visualize the algorithm for users,
- to complex applications with analytics tools, designers and other features.

Each tool is focused on certain types and stages of BPM, where the model must fully comply with the goals of the project and be applicable within your business, area of activity and work system.

Building a business process model is a useful method that allows you to effectively analyze, improve and optimize a company's business processes to achieve strategic goals and increase process automation and its competitiveness.

During the modeling process, it is important to consider the organization as a single system in which various structural and organizational processes are interconnected and interdependent. To do this, you need to be able to decompose each process into its components, look for interdependencies in them, build diagrams, highlight symbols and concepts,

analyze them and operate with them. Business process modeling can help create an organized approach to building a business process and can also improve the efficiency of a company. But for this it is important to be able to handle models and to build the structure of the model, it is necessary to correctly interpret and analyze it.

REFERENCES

1. Побережний Р., Крамської Р., Перерва П. Метод моделювання процесів управління соціально-економічними системами *Слобожанські наукові читання*. Харків: ХПН, 2023. 250-255.
2. Vudrya E., Kobieliava O., Pererva P. The structure of a strategic innovation project for a transport enterprise and the content of its modules *Управління проєктами*. Дніпро : Юрсервіс, 2023. 89-93.
3. Kapiton A., Tyshchenko D., Desiatko A., Lazorenko V. Evolution and distribution analysis of multimodal artificial intelligence systems *Системи управління, навігації та зв'язку*. Полтава: ПНТУ, 2024. Т. 4(78). 75-78.
4. Курилех А., Капітон А. Використання штучного інтелекту для розвитку CRM-систем. *Стан, досягнення та перспективи інформаційних систем і технологій*. Одеса: ОНТУ, 2024. 357-359.
5. Капітон А., Сухорєбрий О., Ненич Д. Використання мультимодального штучного інтелекту в економіці, науці та транспорті. *Інформаційні технології та цифрова економіка*. Київ: ДУІТ, 2024. 83-85.
6. Kapiton A., Baranenko R., Tyshchenko D., Franchuk T. Development and implementation of customer relationship information management system *Системи управління, навігації та зв'язку*. Полтава: ПНТУ, 2022. Т. 2 (78). 24-26.
7. Gromyko O. Business process modeling in BPMN 2.0 URL: [https://pmb.com.ua/uk/obuchenie /modelirovanie-biznes-protsesov-v-notatsii-bpmn-2-0/](https://pmb.com.ua/uk/obuchenie/modelirovanie-biznes-protsesov-v-notatsii-bpmn-2-0/) (дата звернення: 29.11.2024)
8. Система управління бізнес-процесами (BPMS) URL: <https://www.creatio.com/page/uk/bpms> (дата звернення: 22.09.2024) (дата звернення: 22.12.2024)
9. PMBP URL: <https://uk.crmoz.com/blogs/post/business-process-modeling>. (дата звернення: 21.12.2024)
10. The 6 best 3D modeling software in 2024 that every designer should know URL: <https://www.rbc.ua/> (дата звернення: 25.12.2024)
11. Creating processes compatible with business process modeling notation URL: <https://support.microsoft.com/uk-ua/topic/> (дата звернення: 25.12.2024)
12. Optimization of Inventory Management in Retail Companies using Object-Centric Process Mining URL: <https://blog.rwth-aachen.de/pads/> (дата звернення: 25.12.2024)
13. Object Management Group Business Process Model and Notation URL: <https://www.bpmn.org/> (дата звернення: 25.12.2024)
14. Business/Domain logic in domain models and service objects URL: <https://stackoverflow.com/questions/> (дата звернення: 25.12.2024)
15. The Digital Media Value Chain - Modelling Core Business Objects and Processes in Digital Media Enterprises URL: <https://tech.ebu.ch/publications/tr041s> (дата звернення: 25.12.2024)

Received (Надійшла) 11.12.2024

Accepted for publication (Прийнята до друку) 19.02.2025

Моделювання об'єктів та процесів управління

А. Капітон, Т. Франчук, Д. Тищенко, А. Десятко, Н. Сас

Анотація. У статті розглянуто основні питання моделювання об'єктів та процесів управління, проаналізовано та вдосконалено методологічний апарат дослідження. Доведено, що переважна частина дослідників вивчають систему як орієнтовану та певним чином спрямовану множини елементів, між якими встановлено взаємообумовлені та опосередковані певним чином зв'язки досліджуваної області. Встановлено, що певним чином побудована система функціонує задля досягнення визначеної та поставленої мети, що обумовлено елементами, які є складовими цієї системи та обумовленими розглянутими в дослідженні компонентами її теорії, з використанням всіх рівнів її абстрактного опису. Зазначено, що на теоретико-множинному рівні її абстрактного опису при оптимальному формуванні всіх її складових визначають реальний стан та перспективи її розвитку та вдосконалення, які потребують в подальшій перспективі формування, модернізації та впровадження реальних систем. Обґрунтовано, що за умови, коли правильно змодельовані бізнес-процеси, вони надають можливість підвищити ефективність і оптимізувати діяльність всієї компанії, що є основною функцією моделювання та його практичної користі. Встановлено, що після самого процесу моделювання необхідно розробити план покращення та впровадити зміни для успішного підвищення ефективності всіх процесів. Дослідження ринку програмного забезпечення дозволило визначити, що сьогодні достатньо високий рівень сервісів і платформ для моделювання та проектування бізнес-процесів: від базових графічних редакторів, які дозволяють візуалізувати алгоритм для користувачів, до складних додатків із інструментами аналітики, конструкторами та іншими функціями. Побудова моделі бізнес-процесу є корисним методом, який дозволяє ефективно аналізувати, покращувати та оптимізувати бізнес-процеси компанії для досягнення стратегічних цілей і підвищення автоматизації процесу та його конкурентоспроможності. Обґрунтовано, що моделювання бізнес-процесів опосередковано організованим підходом до побудови бізнес-процесу, з поступовим її аналізом та інтерпретацією.

Ключові слова: система, модель, моделювання, процеси управління, аналіз, бізнес-процеси.

I. Nevlyudov, V. Yevsieiev, D. Gurin

Kharkiv National University of Radio Electronics, Kharkiv

MODEL DEVELOPMENT OF DYNAMIC REPRESENTATION A MODEL DESCRIPTION PARAMETERS FOR THE ENVIRONMENT OF A COLLABORATIVE ROBOT MANIPULATOR WITHIN THE INDUSTRY 5.0 FRAMEWORK

Abstract. The article presents a study on the development of a model for the dynamic representation the environmental description parameters for a collaborative robot manipulator within the Industry 5.0 requirements. The main focus is a mathematical model that allows the robot to quickly adapt to changes in the workspace, ensuring effective and safe interaction with humans. The proposed model takes into account data from various sensor systems, such as lidars, cameras, and ultrasonic sensors, to continuously update information about the environment. The study also considers algorithms that optimize the process of data collection and processing to improve the accuracy of prediction and response of the robot. The results of the work are aimed at increasing the efficiency of collaborative robots in production environments, improving the level of automation and ensuring harmonious cooperation between humans and machines within modern cyber manufacturing systems.

Keywords: collaborative robot, dynamic representation, environmental model, Industry 5.0, sensor systems, robot manipulator, collaboration security, automation, adaptability, cyber manufacturing systems.

Introduction

In the context of Industry 5.0 development, there is a growing need to integrate humans and robots into a joint working ecosystem, where secure collaboration, effective communication, and adaptability to changing environments are key. Collaborative robotic manipulators that interact with humans must not only perform their tasks with high accuracy but also respond to dynamic changes in the workspace, ensuring safe operations. One of the critical requirements for such systems is the ability of the robot to continuously update the environmental model, including the parameters of moving objects, human actions, and other changes in space [1].

The relevance of the study lies in the need to develop methods for dynamic environmental description for collaborative robots that take into account the current requirements of Industry 5.0. This includes not only the creation of mathematical models and algorithms for predicting changes in the environment, but also their integration with sensor systems to ensure a high level of adaptability. The development of such methods will help to increase the efficiency of robotic systems and their safe cooperation with people, which is a critical factor in modern production environments [2].

The aim of the article is to study and develop a dynamic representation of environmental parameters for robot manipulators, which will create the basis for more efficient integration of robotics into cyber manufacturing systems in accordance with the requirements of Industry 5.0.

Main part

When developing models for dynamic parameter updating in a collaborative industrial robot model, it is necessary to implement mathematical models and algorithms that allow adaptation to changes in the environment or in the system's operation. This can be

realized using various approaches, including adaptive control, machine learning, and adaptive filtering algorithms [3].

To dynamically update the workspace \mathbb{R}^3 , the model of which is presented in [X], it is proposed to use adaptive algorithms to monitor changes in the environment:

$$\mathbb{R}^3(t) = \mathbb{R}^3(t-1) + \Delta\mathbb{R}^3(t), \quad (1)$$

where $\mathbb{R}^3(t)$ - three-dimensional space at a moment in time t ;

$\mathbb{R}^3(t-1)$ - three-dimensional space at a moment in time $t-1$, describes the three-dimensional space at the previous moment in time. This is the base point from which changes are calculated.

$\Delta\mathbb{R}^3(t)$ - means the changes in the workspace at a point in time t that can be assessed by sensors or a monitoring system.

That is, it reflects the difference or changes that have occurred in three-dimensional space between the time points $t-1$ and t . These changes can be caused by object movements, changes in security perimeters, or other factors.

The assessment of $\Delta\mathbb{R}^3(t)$ can be performed using various sensors that provide data about the environment. These can be described as follows:

- 3D scanners allow collecting data on the three-dimensional geometry of the environment of a collaborative industrial robot manipulator by comparing the captured data with previous scans to determine changes in the environment, the model of such an assessment can be presented as follows:

$$\Delta\mathbb{R}^3(t) = Data_t^{3D} - Data_{t-1}^{3D}, \quad (2)$$

where $Data_t^{3D}$ - are data received from a 3D scanner at a given moment in time t

$Data_{t-1}^{3D}$ - are data at a previous moment in time $t-1$.

- Light Identification, Detection and Ranging (LIDAR) uses laser pulses to determine the distance to

objects and create accurate three-dimensional maps, allows you to analyze data from lidar scanners to determine changes in the location of objects, the model of such an assessment can be presented as follows:

$$\Delta\mathbb{R}^3(t) = Data_t^{LIDAR} - Data_{t-1}^{LIDAR}, \quad (3)$$

where $Data_t^{LIDAR}$ - are data obtained from LIDAR at a given moment in time t ;

$Data_{t-1}^{LIDAR}$ - are data at a previous moment in time $t - 1$.

- cameras (2D or 3D) record images or videos to detect objects and their changes in space, allow the use of computer vision and AI methods to track objects and estimate their movements, the model of such an estimate can be presented as follows:

$$\Delta\mathbb{R}^3(t) = Data_t^{cam} - Data_{t-1}^{cam}, \quad (4)$$

where $Data_t^{cam}$ - position of the object according to the data received from the camera at the moment of time t ;

$Data_{t-1}^{cam}$ - the position of the object obtained from the camera at the previous moment in time $t - 1$.

- distance sensors (ultrasonic, laser, etc.) ($Sens$) measure the distance to objects in real time, i.e. collect data on changes in distances to objects to assess changes in the environment, the model of such assessment can be presented as follows:

$$\Delta\mathbb{R}^3(t) = Data_t^{Sens} - Data_{t-1}^{Sens}, \quad (5)$$

where $Data_t^{Sens}$ - distance to the object according to the data received from the sensor at the moment of time t ;

$Data_{t-1}^{Sens}$ - the distance to the object obtained from the sensor at the previous moment in time $t - 1$.

On the basis of 2.45-2.48, changes in three-dimensional space can be represented as:

$$\begin{aligned} \Delta\mathbb{R}^3(t) = & \\ = (Data_t^{3D}, Data_t^{LIDAR}, Data_t^{cam}, Data_t^{Sens}) - & \\ - (Data_{t-1}^{3D}, Data_{t-1}^{LIDAR}, Data_{t-1}^{cam}, Data_{t-1}^{Sens}), & \end{aligned} \quad (6)$$

where $Data_t^{3D}, Data_t^{LIDAR}, Data_t^{cam}, Data_t^{Sens}$ - є даними, отриманими від сенсорів на момент часу t ;

$Data_{t-1}^{3D}, Data_{t-1}^{LIDAR}, Data_{t-1}^{cam}, Data_{t-1}^{Sens}$ - are data obtained at a previous point in time $t - 1$.

Combining these data allows us to adaptively update the description of three-dimensional space, taking into account changes in the environment, and ensure the accuracy and efficiency of the collaborative robot manipulator.

The working area (\mathbb{D}) can dynamically change depending on changes in objects in space or changes in security perimeters (Ω_{safe}) and can be described by the following expression:

$$\mathbb{D}(t) = \mathbb{D}(t - 1) \cup \Delta\mathbb{D}(t), \quad (7)$$

where $\mathbb{D}(t)$ - workspace at a given time t , describes the space in which the robot operates, including all objects (Ω_i) and safety perimeters (Ω_{safe}) that may affect its actions.

This can be a limited area in which the manipulator performs a task, taking into account all new objects and changes in the environment;

$\mathbb{D}(t - 1)$ - workspace at a time moment $t - 1$, describes the working area at the previous time point. This is the base point from which changes are calculated;

$\Delta\mathbb{D}(t)$ - changes in the workspace at a given time, which can be described as new or changed areas of the workspace, i.e., it reflects new objects or changes in existing objects that have appeared or been changed in the workspace between time points $t - 1$ and t .

This can be, for example, a new facility, a relocated facility, or a change in security perimeters.

A dynamic workspace ($\mathbb{D}(t)$) can be described as a system that is constantly updated in response to changes in the environment ($\Delta\mathbb{D}(t)$). Updating can occur in real time due to data from sensors and cameras [5-7] that track changes in the environment. As a result $\Delta\mathbb{D}(t)$, within the framework of these studies, it can be represented as follows [4]:

- 3D scanners and lidars collect data about the three-dimensional geometry of the environment, identifying new objects or changes in the location of objects. New information obtained from scanners or lidars can be represented as an addition to the previous zone and can be described as follows:

$$\Delta\mathbb{D}(t) = \Omega_{new}^{3D, LIDAR}, \quad (8)$$

where $\Omega_{new}^{3D, LIDAR}$ - new objects or changes in the location of objects, obtained from scanners or lidars,

- camera(s) collect images or video to detect new objects or changes in the environment. New or changed objects detected in the images are added to the workspace and can be described as follows:

$$\Delta\mathbb{D}(t) = \Omega_{new}^{cam}, \quad (9)$$

where Ω_{new}^{cam} - new objects or changes in the location of objects received from the camera;

- distance sensors measure the distance to objects, which allows detecting new objects or changes in their location, new data from distance sensors are added to the working area, and can be described as follows:

$$\Delta\mathbb{D}(t) = \Omega_{new}^{sens}, \quad (10)$$

where Ω_{new}^{sens} - new objects or changes in the location of objects received from distance sensors;

As a result, the mathematical description of the dynamic work area can be obtained by modernizing expression 7 and using expressions 8-10, the result of this solution is given below:

$$\begin{aligned} \mathbb{D}(t) = & \\ = (Data_{t-1}^{3D}, Data_{t-1}^{LIDAR}, Data_{t-1}^{cam}, Data_{t-1}^{Sens}) & \\ \cup \Omega_{new}^{3D, LIDAR}, \Omega_{new}^{cam}, \Omega_{new}^{sens} \setminus & \\ \setminus \text{Removed} (\Omega_{t-1}^{3D, LIDAR}, \Omega_{t-1}^{cam}, \Omega_{t-1}^{sens}) & \end{aligned} \quad (11)$$

where $Data_{t-1}^{3D}, Data_{t-1}^{LIDAR}, Data_{t-1}^{cam}, Data_{t-1}^{Sens}$ - is the working space of the collaborative robot manipulator at the previous moment in time, all objects and areas that were relevant at the moment $t - 1$, formally, this is an extended representation $\mathbb{D}(t - 1)$ from expression 7;

$\Omega_{new}^{3D, LIDAR}, \Omega_{new}^{cam}, \Omega_{new}^{sens}$ - are new objects that have been added to the workspace between time $t - 1$ and t ,

as a result, objects have appeared in the workspace and need to be added to the previous zone, formally, this is an extended representation $\Delta\mathbb{D}(t)$ from expression 2.50;

Removed $(\Omega_{t-1}^{3D,LIDAR}, \Omega_{t-1}^{cam}, \Omega_{t-1}^{sens})$ - are objects that were in the workspace earlier (at time $t - 1$), but are no longer part of the workspace. This can be because the objects have moved, been deleted, or are no longer relevant.

Based on 11:

$$Data_{t-1}^{3D}, Data_{t-1}^{LIDAR}, Data_{t-1}^{cam}, Data_{t-1}^{sens} \cup \Omega_{new}^{3D,LIDAR}, \Omega_{new}^{cam}, \Omega_{new}^{sens} -$$

the operation of combining two sets - the previous workspace and the new objects.

This creates an updated set of objects that are in the workspace at the time of t , including new objects, and Removed $(\Omega_{t-1}^{3D,LIDAR}, \Omega_{t-1}^{cam}, \Omega_{t-1}^{sens})$ - the set difference operation removes objects that are no longer part of the workspace.

That is, those objects that have been deleted or are no longer relevant are removed from the result of the merge (with new objects).

Here is an example of how 11 allows you to dynamically update the workspace, adapting it to changes in the environment of the collaborative robot manipulator.

Suppose that the working area of the collaborative robot manipulator at the previous time

$$\mathbb{D}(t - 1) = A, B, C$$

has the following objects, and suppose the appearance of $\Omega_{new}^{cam} = \{D, E\}$ that are the new objects that were added to the working area between $t - 1$ and t , while the object $\{B\}$ left the working area and is no longer part of it. Then, according to 11, a merge is:

$$\{A, B, C\} \cup \{D, E\} = \{A, B, C, D, E\},$$

and a delete is

$$\{A, B, C, D, E\} \setminus \{B\} = \{A, C, D, E\}.$$

Thus, at the moment t , the working area $\mathbb{D}(t)$ will include objects $\{A, C, D, E\}$ excluding objects that have been deleted. The proposed model 11 based on 7 allows us to dynamically update the working area, adapting it to changes in the environment, which is important for the accuracy and safety of collaborative industrial robots.

The model of the dynamics of objects in space is based on the model of objects in space Ω_i , and represents both areas with certain geometric shapes and sizes, the mathematical description of which is given in [5]. To dynamically update the location of objects Ω_i , it is proposed to use object tracking algorithms. Based on this, the model of the dynamics of objects in the working area of the collisional robot manipulator can be represented as follows:

$$\Omega_i(t) = \Omega_i(t - 1) + \Delta\Omega_i(t), \quad (12)$$

where $\Omega_i(t)$ - is the set of objects in space at a moment in time t , the state of objects in space at the current moment in time t , and includes all objects that are in the

working area of the collaborative robot manipulator at this moment;

$\Omega_i(t - 1)$ - is the set of objects in space at a moment of time $t - 1$, is the state of objects in space at the previous moment of time $t - 1$, is the base set of objects from which the update starts.

$\Delta\Omega_i(t)$ - change in the set of objects for the period from $t - 1$ to t , that is, new objects that have appeared or changes in existing objects. In other words, it is the difference between the current and previous state of objects. It can include new objects that have appeared or changes in the properties of existing objects (e.g., moving, resizing, changing state).

Model 12 describes how the set of objects in space is updated over time, allowing for dynamic changes in the environment, such as the addition of new objects or changes to existing ones. To give a mathematical description, the set of objects at time t can be described as $\Omega_i(t) = \{\Omega_c, \Omega_{cy}, \Omega_{co}, \Omega_{cu}, \dots, \Omega_{qp}\}$, where $\Omega_c, \Omega_{cy}, \Omega_{co}, \Omega_{cu}, \dots, \Omega_{qp}$ - objects in space that are represented as areas with certain geometric shapes and sizes. Changes in the objects $\Delta\Omega_i(t)$ can be divided into two types of changes: new objects $\Delta\Omega_i^+(t)$ and objects that have been deleted or modified $\Delta\Omega_i^-(t)$. Then:

$$\Delta\Omega_i(t) = \Delta\Omega_i^+(t) \cup \Delta\Omega_i^-(t). \quad (13)$$

Based on 12 and 13, the updated set will look like this:

$$\Omega_i(t) = \Omega_i(t - 1) \cup \Delta\Omega_i^+(t) \setminus \Delta\Omega_i^-(t), \quad (14)$$

where $\Omega_i(t - 1)$ - a set of objects in space at a given time moment $t - 1$, is the state of objects in space at a previous moment in time $t - 1$, is the base set of objects from which the update starts;

$\Delta\Omega_i^+(t)$ - new objects in the working area of the collaborative robot manipulator;

$\Delta\Omega_i^-(t)$ - objects that have been deleted or modified in the working area of the collaborative robot manipulator.

For example, let's assume that there are the following objects in our workspace, at the time $t - 1$: $\Omega_i(t - 1) = \{A, B, C\}$, new objects at time t : $\Delta\Omega_i^+(t) = \{D\}$ and objects deleted or changed at t : $\Delta\Omega_i^-(t) = \{B\}$. Then, in accordance with 14, we have the following updated set:

$$\Omega_i(t) = \{A, B, C\} \cup \{D\} \setminus \{B\} = \{A, C, D\}. \quad (15)$$

Thus, at time t , the number of objects in the working area of the collaborative robot manipulator will be $\{A, C, D\}$, where object B was deleted, and the object D was added. The proposed model of the dynamics of objects in the working area of a collaborative robot manipulator allows the system to dynamically update the list of objects in the environment, which is important for ensuring the accuracy and relevance of data in robotic systems.

The motion dynamics model $q(t)$ is based on the mathematical representation of the position function and can be implemented using adaptive control algorithms that take into account changes in the environment or in the system operation [7]:

$$\mathbf{q}(t) = \mathbf{q}(t-1) + \Delta\mathbf{q}(t) \quad (16)$$

where $\mathbf{q}(t)$ - is the state vector of the robot or manipulator at time t .

It includes the coordinates of the joints, the position of the end effector, or other system parameters that change over time. If it is represented as a vector of joint positions, it can be described as follows $\mathbf{q}(t) = [q_1(t), q_2(t), \dots, q_n(t)]^T$, Where $q_n(t)$ - is the position of the n -joint at a time moment t ;

$\mathbf{q}(t-1)$ - is the vector of the state of the robot or manipulator at the previous time moment $t-1$, and is the base state from which the update is performed;

$\Delta\mathbf{q}(t)$ - change in the system state for the period from $t-1$, and may include changes in joint position, velocity, acceleration, or other variables that characterize the dynamics of the system. Mathematical representation $\Delta\mathbf{q}(t) = [\Delta q_1(t), \Delta q_2(t), \dots, \Delta q_n(t)]^T$, where $\Delta q_n(t)$ - change in the position of the n -joint at a time moment t .

To give an example, let's say we have a robot with three joints. The state vectors of the joints at different times are described as follows:

- at the moment $t-1$: $\mathbf{q}(t-1) = [q_1(t-1), q_2(t-1), q_3(t-1)]^T = [1.0, 0.5, -0.3]$ - in radians;

- change of state for the period $\Delta\mathbf{q}(t) = [\Delta q_1(t), \Delta q_2(t), \Delta q_3(t)]^T = [0.05, -0.02, 0.1]$ - in radians.

Then the updated state vector in accordance with 16 will be as follows:

$$\mathbf{q}(t) = [1.0, 0.5, -0.3]^T + [0.05, -0.02, 0.1]^T = [1.05, 0.48, -0.2]^T. \quad (17)$$

Let's interpret the results in 17: the position of the first joint changed from 1.0 radians to 1.05 radians; the position of the second joint changed from 0.5 radians to 0.48 radians; the position of the third joint changed from -0.3 radians to -0.2 radians [6].

The proposed formula 16 describes how the system parameters (e.g., the position of the joints or the end effector) change over time.

This is important for modeling the movement of a robot or manipulator, tracking its positions, velocities, and accelerations.

Motion dynamics model $\boldsymbol{\tau}(t)$ should take into account changes in the environment and in the system operation, and is an extension of $\boldsymbol{\tau}$ (2.13) and is as follows:

$$\boldsymbol{\tau}(t) = \mathbf{M}(\mathbf{q}(t))\ddot{\mathbf{q}}(t) + \mathbf{C}(\mathbf{q}(t), \dot{\mathbf{q}}(t))\dot{\mathbf{q}}(t) + \mathbf{G}(\mathbf{q}(t)) + \Delta\boldsymbol{\tau}(t), \quad (18)$$

where $\boldsymbol{\tau}(t)$ - is a vector of force moments (torsions) applied to the joints of a robot or manipulator at time moment t . These are controlling forces or moments that are required to achieve a given state or execute a command;

$\mathbf{M}(\mathbf{q}(t))$ - the inertia matrix, which depends on the state vector $\mathbf{q}(t)$, describes how the inertia of the joints changes depending on their position;

$\ddot{\mathbf{q}}(t)$ - is the vector of joint accelerations at time t , taking into account the dynamic effects associated with accelerations;

$\mathbf{C}(\mathbf{q}(t), \dot{\mathbf{q}}(t))$ - Coriolis matrix and centrifuge forces and their speeds $\dot{\mathbf{q}}(t)$, matrix describes the forces arising from centrifugal effects;

$\dot{\mathbf{q}}(t)$ - The vector of joint velocities at a moment in time t , taking into account the Coriolis and centrifuge effects;

$\mathbf{G}(\mathbf{q}(t))$ - The vector of gravitational forces acting on the robot's joints. Depends on the position of the joints $\mathbf{q}(t)$ and describes the forces arising from gravity;

$\Delta\boldsymbol{\tau}(t)$ - changes in force moments, which can be the result of external influences or uncertain changes.

This can include noise, measurement errors, or unaccounted-for dynamic effects, and can look like this for a three-joint robot:

$$\Delta\boldsymbol{\tau}(t) = \begin{bmatrix} \text{noise}_1(t) \\ \text{noise}_2(t) \\ \text{noise}_3(t) \end{bmatrix}, \quad (19)$$

where $\text{noise}_1(t)$ - noise or uncertainty change component for the first joint or system element. It can be caused by mechanical malfunctions, measurement errors or other external factors;

$\text{noise}_2(t)$ - noise component or uncertainty change for the second joint or system element. It can affect control accuracy and system dynamics;

$\text{noise}_3(t)$ - noise component or undefined change for the third joint or system element. It can be the result of additional forces acting on the third component of the system or special operating conditions.

Within the framework of these studies, it is proposed to model the noise as a random process with a normal distribution, which is determined by the average value of μ_i and dispersion σ_i^2 .

Formally, this can be presented as:

$$\Delta\boldsymbol{\tau}(t) = \begin{bmatrix} \text{noise}_1(t) \\ \text{noise}_2(t) \\ \text{noise}_3(t) \end{bmatrix}, \quad (20)$$

where $\text{noise}_1(t)$ - noise or uncertainty change component for the first joint or system element. It can be caused by mechanical malfunctions, measurement errors or other external factors;

$\text{noise}_2(t)$ - noise component or uncertainty change for the second joint or system element. Can affect control accuracy and system dynamics;

$\text{noise}_3(t)$ - noise component or uncertainty change for the third joint or system element. It may be the result of additional forces acting on the third system component or special operating conditions.

Noise in this study is considered as a random or statistical change that may have a normal distribution. It is defined as a random process with certain parameters and will be modeled as a random process with a normal distribution.

The normal distribution is defined by the mean μ_i and dispersion σ_i^2 . Formally, it can be described as follows:

$$noise_i(t) \sim \mathcal{N}(\mu_i, \sigma_i^2) \quad (21)$$

where μ_i - is the average noise value for the i -th component;

σ_i^2 - is the noise dispersion for the i -th component.

Uncertain variables can be deterministic or random and may include the effects of environmental changes or malfunctions.

They can be described as additional or corrective components that do not have a clear statistical model.

To give an example, let's say there is a manipulator with three joints, and you need to model changes in force moments due to noise [8]:

- noise for the first joint $noise_1(t)$ has a mean of 0 and a dispersion of 0.1. This can be represented as $noise_1(t) \sim \mathcal{N}(0, 0.1^2)$;

- noise for the second joint $noise_2(t)$ has a mean of 0 and a dispersion of 0.05. This can be written as $noise_2(t) \sim \mathcal{N}(0, 0.05^2)$;

- noise for the third joint $noise_3(t)$ has a mean of 0 and a dispersion of 0.2. This can be expressed as $noise_3(t) \sim \mathcal{N}(0, 0.2^2)$.

Thus, the total vector of change in force moments can be represented as in 20, where each component $noise_i(t)$ is a normally distributed random noise with the corresponding parameters.

The mathematical representation of the vector $\Delta\tau(t)$ within the framework of the motion dynamics model $\tau(t)$ (21) is important for the accuracy of manipulator control and monitoring.

Its modeling helps in assessing and correcting errors that may occur due to uncertain or random changes in the system [9].

The safety perimeters Ω_{safe} can change depending on new data about objects in the environment or changes in the system operation at time t , so the dynamic model (updating the safety perimeters) can be represented as follows:

$$\Omega_{safe}(t) = \Omega_{safe}(t-1) \cup \Delta\Omega_{safe}, \quad (21)$$

where $\Omega_{safe}(t)$ - safety perimeters at time t , is a set of areas in space that define safety zones around the robot, which may include areas of danger to people and other objects. In this research, it is presented as a set that defines the boundaries where robots should limit their activities to ensure safety;

$\Omega_{safe}(t-1)$ - security perimeters at the previous time $t-1$, These are the values of security perimeters before the update at time t ;

$\Delta\Omega_{safe}$ - changes in security perimeters, which is a set of new zones or changes to existing security zones that need to be considered when updating the security model.

It can include new zones that are added or changes in the size of existing zones.

To give an example, suppose that at time $t-1$, the safety p

erimeters of the manipulator are defined as an area around the robot that includes a radius (r) - 2 meters. At time t , the safety perimeters may be updated due to new conditions, such as:

- new perimeters, a new safety zone with a radius of 1 meter is added around certain objects in the work area;

- changes in the existing perimeters, the average radius of the safety zone is increased by 0.5 meters.

Mathematically, this can be represented as follows:

- The safety perimeter at time $t-1$, which is equal to $r=2$ meters, can be represented as follows:

$$\Omega_{safe}(t-1) = \Omega_{safe}\{r = 2 \text{ meters}\}; \quad (22)$$

- new perimeters or changes, can be described as follows:

$$\begin{aligned} \Delta\Omega_{safe} &= \Omega_{safe}^{new} r = 1 \text{ meters} \cup \\ &\cup \Omega_{safe}^{expa} r = 2.5 \text{ meters}. \end{aligned} \quad (23)$$

To obtain the model of the updated security perimeters at time t , we substitute models 23 and 24 into 22, and we get the following model:

$$\begin{aligned} \Omega_{safe}(t) &= \Omega_{safe} r = 2 \text{ meters} \cup \\ &\cup \Omega_{safe}^{new} r = 1 \text{ meters} \cup \Omega_{safe}^{expa} r = \\ &= 2.5 \text{ meters}. \end{aligned} \quad (24)$$

Model 24 allows to take into account dynamic changes in the security zone, adapting it in accordance with new conditions or security requirements.

The safety perimeters $\Omega_{safe}(t)$ in the context of the collaborative industrial robot model, defines areas or zones around the robot where movement or the robot may be restricted to ensure the safety of people and equipment.

Adaptive communication $u(t)$ in the context of a collaborative industrial robot represents the process of exchanging information between the robot system and the operator or other systems, which can change in accordance with changes in the environment or in the system's operation.

In order to consider the function $u(t)$ in a dynamic space, it is necessary to describe the state of communication at time t , the mathematical representation of which is given below:

$$u(t) = u(t-1) + \Delta u(t), \quad (25)$$

where $u(t)$ - is the state of communication at time t , which is the current set of commands, messages, or information that the robot system exchanges with the operator or other systems. It should be noted that this parameter can include both information about the robot's state and reactions to external commands or conditions;

$u(t-1)$ - is the state of communication at the previous time $t-1$, but it is a set of commands or information that was relevant up to the time t ;

$\Delta u(t)$ - changes in communication can be: 1) These are new commands or messages that need to be added to the system to accommodate changes in the environment or in the operation of the system. 2) It may include changes in the interaction between the robot and the operator, new instructions to be followed, or new data to be transmitted.

To give an example, suppose that the robot system at time $t-1$ received and transmitted information about the current position and speed.

At time t , the system adapts to the new conditions, where the operator adds a new command to change the trajectory, and the system needs to report the current battery status.

Then the communication at time $t-1$ can be represented as follows:

$$u(t-1) = \{\text{Position Information, Speed Information}\}. \quad (26)$$

New changes in communication:

$$\Delta u(t) = \{\text{Trajectory Change Command, Battery Status Update}\}. \quad (27)$$

Then the updated communication at time t , respectively, 2.67 will be as follows:

$$u(t) = \{\text{Position Information, Speed Information, Trajectory Change Command, Battery Status Update}\}. \quad (28)$$

As can be seen from 28, adaptive communication can include mechanisms that allow the system to automatically change communication protocols depending on the context. For example:

- changes in environmental conditions, If the robot detects a new object in the work area, the system can automatically communicate information about the new object to the operator and request confirmation of further action;

- changes in the system operation, if the robot detects a decrease in the battery level, the system can initiate a notification to the operator and possibly suggest switching to an economical mode.

Such solutions allow collaborative robotic manipulators to not only respond to external changes, but also actively communicate with the operator, providing a continuous flow of information necessary for safe and efficient collaboration.

The adaptive learning model $\mathbb{M}(t)$ means that the model is constantly updated based on new data or new conditions encountered by the robot manipulator.

This allows the robot to improve its performance and adapt to changes in the environment or operating conditions.

Adaptive learning model $\mathbb{M}(t)$ is as follows:

$$\mathbb{M}(t) = \mathbb{M}(t-1) + \Delta\mathbb{M}(t), \quad (29)$$

where $\mathbb{M}(t)$ - is the state of the trained model at time t , a set of parameters or rules that determine the current level of knowledge or skills of the system. In the context of a robot manipulator, this can be a model that defines how the robot performs certain tasks, including control rules, sensor data processing, object recognition, etc.;

$\mathbb{M}(t-1)$ - is the state of the trained model at time $t-1$, which is the previous state of the model before any changes were made or additional training was performed. This is the starting point from which the model can change over time;

$\Delta\mathbb{M}(t)$ - changes in training or model updates and may be: 1. New knowledge, skills, or rules that are added to the model as a result of training, adaptation, or learning from new data; 2. Can be the result of learning from new data obtained during operation or the result of optimizing the model for better performance.

To give an example, let's say that a collaborative robot manipulator has initially learned to recognize three main types of objects in the work area.

However, in the course of its work, the robot encounters new objects (12) that need to be recognized or new situations that require changing control rules.

Accordingly, the initial state of the trained model at the time $t-1$ is as follows:

$$\mathbb{M}(t-1) = \Omega_s, \Omega_o, \Omega_t, \quad (30)$$

where Ω_s - the object is recognized as an area with a geometric shape of a cube;

Ω_o - the object is recognized as an area with a geometric shape of a cylinder;

Ω_t - the object is recognized as an area with a geometric shape of a cone.

The robot collected new data and underwent additional training to recognize a new type of object (for example, a parallelepiped (Ω_{rh})) or new rules of behavior when interacting with this object.

Then changes in learning $\Delta\mathbb{M}(t)$ will have the following form:

$$\Delta\mathbb{M}(t) = \Omega_{rh} + \text{new conduct rules}, \quad (31)$$

where Ω_{rh} - the object recognized as an area with the geometric shape of a rectangular parallelepiped.

Then the updated state of the trained model at time t , for this example, will be as follows:

$$\mathbb{M}(t) = \Omega_s, \Omega_o, \Omega_t + \Omega_{rh} + \text{new conduct rules}. \quad (32)$$

In this way, the robot manipulator constantly adapts its model based on new knowledge, which allows it to better cope with new situations or objects in the work area.

This can be realized as part of machine learning algorithms that allow the robot to "learn" while working, or through software updates based on feedback from sensors and control systems.

Conclusions

As a result of the study, a model for the dynamic representation of environmental description parameters for a collaborative robot manipulator that meets the requirements of Industry 5.0 was developed.

The developed model allows robotic manipulators to quickly respond to changes in the workspace, ensuring effective and safe cooperation with humans.

The main advantage of this model is the integration of data from various sensor systems, such as lidars, cameras, and ultrasonic sensors, which allows for a complete and accurate picture of the robot's environment in real time.

This makes it possible to work in dynamic environments where there are constant changes, such as moving objects, changing lighting, or unforeseen obstacles.

The model also takes into account the parameters of human behavior, which is a key factor in ensuring safe cooperation between the robot and the operator in a common space.

Thanks to the introduction of algorithms for processing large amounts of data in real time, the system is able to adapt to new conditions without the need to completely rebuild the entire system.

This increases the flexibility and reduces the cost of operating robotic systems in production environments.

It is proposed to implement the developed model in robotic systems used in automated production, where it is important to ensure the integration of robots with other components of the cyber production system.

The use of this model will increase the efficiency of production processes, reduce the accident rate, and ensure safe cooperation between humans and machines. In addition, the implementation of such solutions will help optimize the costs of developing and integrating robotic systems, as the model supports scalability and can be adapted to different use cases.

Promising areas for further research include improving machine learning algorithms to predict changes in the environment even more accurately, as well as integrating artificial intelligence systems to improve the interaction between the robot and the operator.

The implementation of the developed model will help to accelerate automation processes and increase the level of integration of robots into modern cyber production systems of Industry 5.0, where the priority is not only automation but also harmonious cooperation between humans and machines.

REFERENCES

1. Rahman, M. M., Khatun, F., Jahan, I., Devnath, R., & Bhuiyan, M. A. A. (2024). Cobotics: The Evolving Roles and Prospects of Next-Generation Collaborative Robots in Industry 5.0. *Journal of Robotics*, 2024(1), 2918089.
2. Gurin, D., Yevsieiev, V., Maksymova, S., & Abu-Jassar, A. (2024). Effect of Frame Processing Frequency on Object Identification Using MobileNetV2 Neural Network for a Mobile Robot. *Multidisciplinary Journal of Science and Technology*, 4(8), 36-44.
3. Palácios, R.H.C., Bertocini, J.P.S., Uliam, G.H.O. et al. Evaluation of mobile autonomous robot in trajectory optimization. *Computing* 105, 2725–2745 (2023). <https://doi.org/10.1007/s00607-023-01205-6>
4. Yevsieiev, V., & Gurin, D. (2023). Comparative Analysis of the Characteristics of Mobile Robots and Collaboration Robots Within INDUSTRY 5.0 (Doctoral dissertation, European Scientific Platform).
5. Holzinger, A., Schweier, J., Gollob, C. et al. From Industry 5.0 to Forestry 5.0: Bridging the gap with Human-Centered Artificial Intelligence. *Curr. For. Rep.* (2024). <https://doi.org/10.1007/s40725-024-00231-7>
6. Zafar, M. H., Langås, E. F., & Sanfilippo, F. (2024). Exploring the synergies between collaborative robotics, digital twins, augmentation, and industry 5.0 for smart manufacturing: A state-of-the-art review. *Robotics and Computer-Integrated Manufacturing*, 89, 102769.
7. Gurin, D., Yevsieiev, V., Maksymova, S., & Alkhalaleh, A. (2024). Using Convolutional Neural Networks to Analyze and Detect Key Points of Objects in Image. *Multidisciplinary Journal of Science and Technology*, 4(9), 5-15.
8. Bhat, F.A., Parvez, S. Emerging Challenges in the Sustainable Manufacturing System: From Industry 4.0 to Industry 5.0. *J. Inst. Eng. India Ser. C* (2024). <https://doi.org/10.1007/s40032-024-01046-y>
9. Mahmud, S., Kamarulariffin, A., Ibrahim, A.M. et al. Advancements and Challenges in Mobile Robot Navigation: A Comprehensive Review of Algorithms and Potential for Self-Learning Approaches. *J Intell Robot Syst* 110, 120 (2024). <https://doi.org/10.1007/s10846-024-02149-5>

Received (Надійшла) 18.12.2024

Accepted for publication (Прийнята до друку) 26.02.2025

Розробка моделі динамічного представлення параметрів моделі опису навколишнього середовища колоборативного робота маніпулятора в рамках індустрії 5.0

I. Ш. Невлюдов, В. В. Євсєєв, Д. В. Гурін

Анотація. У статті представлено дослідження, присвячене розробці моделі динамічного представлення параметрів опису навколишнього середовища для колоборативного робота-маніпулятора в контексті вимог Індустрії 5.0. Основна увага приділяється створенню математичної моделі, яка дозволяє роботу швидко адаптуватися до змін у робочому просторі, забезпечуючи ефективну і безпечну взаємодію з людиною. Запропонована модель враховує дані з різних сенсорних систем, таких як лідари, камери та ультразвукові датчики, для постійного оновлення інформації про навколишнє середовище. У дослідженні також розглядаються алгоритми, що дозволяють оптимізувати процес збору та обробки даних для підвищення точності прогнозування і реакції робота. Результати роботи спрямовані на підвищення ефективності колоборативних роботів у виробничих умовах, покращення рівня автоматизації та забезпечення гармонійної співпраці між людиною і машиною в межах сучасних кібервиробничих систем.

Ключові слова: колоборативний робот, динамічне представлення, модель навколишнього середовища, Індустрія 5.0, сенсорні системи, робот-маніпулятор, безпека співпраці, автоматизація, адаптивність, кібервиробничі системи.

П. Е. Решетнікова, О. Ю. Заковоротний

Національний технічний університет «Харківський політехнічний інститут», Харків, Україна

МОДЕЛЮВАННЯ ВЕРТИКАЛЬНИХ КОЛИВАНЬ ПОЇЗДА, ЩО ВИНΙΚАЮТЬ ПІД ЧАС РУХУ ЗАЛІЗНИЧНОЮ КОЛІЄЮ ЗІ СТИКОВОЮ НЕРІВНІСТЮ

Анотація. Залізничний транспорт багато років залишається одним із основних перевізників вантажів та пасажирів в Україні. Незважаючи на це, більшість поїздів та рейок основних напрямків сполучення знаходяться в незадовільному стані. В умовах бойових дій та подальшого довгого післявоєнного відновлення на перший план виходять питання подовження життя вже існуючому рухомому складу та економії енергоресурсів. Також для збереження конкурентоспроможності поїздів відносно інших видів транспорту гостро стоїть питання підвищення середніх швидкостей руху та комфорту пасажирів на залізниці. Для вирішення цих проблем необхідна розробка комплексних моделей, які би враховували фактори, які суттєво впливають на витрати палива, безпеку та комфорт пасажирів. До таких факторів можна віднести коливання рухомого складу. Роботу присвячено розробці імітаційної моделі вертикальних коливань (підстрибування та галопування) кузова та візків вагона поїзда, які залежать від нерівностей залізничної колії. Розроблена комплексна імітаційна модель руху вагона поїзда перегонном з урахуванням вертикальних коливань кузова вагона та його візків у пакеті MATLAB Simulink дозволяє досліджувати залежність цих коливань від швидкості рухомого складу, параметрів його системи підвищення та нерівностей рейкового шляху під колісними парами не тільки в режимі підтримки заданої швидкості, а й у режимі вибігу та перехідних режимах (розгоні, гальмуванні). Отримана імітаційна модель може бути використана для подальшого проведення на ній досліджень коливальних процесів поїзда на обраних ділянках шляху, для отримання на ній даних для тренування нейронної мережі для раннього визначення резонансних коливань вагона поїзда, а також може бути використана у системах підтримки прийняття рішень (СППР) бортових систем керування рухомого складу поїздів України.

Ключові слова: математична модель, система підтримки прийняття рішень, бортові системи керування, рухний склад, коливальний процес, підстрибування, галопування, резонансна швидкість.

Вступ

Постановка проблеми. Залізничний транспорт роками займає провідну роль у транспортній системі України. Абсолютна більшість вантажів, що прямують як у різні регіони нашої країни, так і закордон, перевозяться за допомогою рухомого складу залізниць. Чималою є також частка залізничного транспорту у пасажирських перевезеннях.

Незважаючи на незаперечну важливість залізниць для підтримки та подальшого розвитку економіки України, на даний час більшість поїздів знаходиться у незадовільному стані та вже давно вичерпала свій ресурс, який було закладено при їх будівництві. До того ж близько половини головних колій потребують негайного ремонту [1, 2].

В умовах бойових дій, які супроводжуються ракетними ударами, що спрямовані на знищення об'єктів критичної інфраструктури, зокрема енергетики та транспорту, та подальшого довгого післявоєнного відновлення на перший план виходять питання подовження життя вже існуючому рухомому складу та економії енергоресурсів. Також для збереження конкурентоспроможності поїздів відносно інших видів транспорту гостро стоїть питання підвищення середніх швидкостей руху та комфорту пасажирів на залізниці.

Для вирішення цих проблем необхідна розробка комплексних моделей, які би враховували фактори, які суттєво впливають на витрати палива, безпеку та комфорт пасажирів. У даній роботі пропонується звернути увагу на такі динамічні процеси, як вертикальні коливання рухомого складу. Вертикальні коливання поїзда виникають внаслідок дії вертикальних сил, які наявні у процесі руху. До них відносяться:

підстрибування, коли надресорні маси вагона переміщуються вгору і вниз, залишаючись паралельними початковому положенню, та галопування, коли надресорні маси обертаються навколо осі, що перпендикулярна осі руху поїзда. Ці коливання можуть бути викликані нерівностями рейкового шляху, пошкодженнями та зношенням ободів коліс, незбалансованою системою підвищення тощо. Коливання рухомого складу мають багатогранний негативний вплив на потяг, як на систему в цілому, так і на його частини, призводячи до пришвидшеного зносу коліс та ресорного комплексу, а також на залізничні колії, викликаючи їх різноманітні пошкодження. Підстрибування та галопування призводять до втрати зчеплення коліс, підвищенню опору середовища рухові, а отже й до збільшення витрат палива [3]. Підстрибування поїзда, особливо за підвищення швидкостей руху, можуть призводити до інерційних перенавантажень, що несуть загрозу безпеці та комфорту та можуть суттєво сказатися на здоров'ї пасажирів.

Навіть за наявності механічних гасників коливань, які використовуються у всіх сучасних пасажирських поїздах, надресорні частини візків та вагонів все ще приходять у коливальний стан. Гасіння коливань на залізницях України додатково ускладнюється й розповсюдженим використанням стикового рейкового шляху та загальним незадовільним станом шляхів. Особливо це відчувається під час перехідних процесів руху (розгоні та гальмуванні), де на пасажирів діють не тільки вертикальні, а й горизонтальні прискорення, що може призводити до підвищених дискомфорту та перенавантажень. Також дуже небезпечним є явище резонансу, коли частоти власних та вимушених коливань співпадають, що призводить до значного підвищення амплітуди цих коливань.

Аналіз останніх досліджень і публікацій. Традиційно при дослідженні коливань рухомого складу більше уваги приділяють поперечним коливанням (боковий винос, виляння), так як вони є природним процесом через особливості будови системи «колесо-рейка», а перевищення норм амплітудних значень несе загрозу виникнення аварійних ситуацій та сходів з поїзда з рейок [4–11].

Вертикальні, хоча й є не менш шкідливими та мають найбільший вплив на комфорт пасажирів [12, 13], описуються у сучасній літературі рідше та здебільшого у контексті оптимізації вибору жорсткості та коефіцієнтів демпфування для нових поїздів [14–16]. У деяких роботах, демпфуваннями навпаки нехтують та концентрують свою увагу на коливаннях саме колісних пар та можливій асиметричності нерівностей рейкового шляху [17].

Існуючі дослідження коливань у загальному випадку проводяться на заданій сталій швидкості, а тому не враховують перехідні такі перехідні процеси як розгін та гальмування [9–12, 14–18].

Моделювання збурень для вимушених коливань проводиться у різних роботах по-різному. Найрозповсюдженішим є використання так званої односторонньої нерівності, яка є найпростішою формою задання періодичної нерівності. У даній статті авторами для досягнення більш точних результатів пропонується використання двогорбої нерівності, запропонованої проф. Кудрявцевим [19].

У роботі [20] авторами була запропонована комплексна математична модель руху дизель-поїзда, яка включає в себе роботу еквівалентного двигуна поїзда, вертикальних переміщень (підстрибування) та кутів повороту (галоупування) візків та кузова вагона, які залежать від нерівності залізничної колії.

Метою статті є розробка комплексної імітаційної моделі руху вагона поїзда перегоном з урахуванням вертикальних коливань кузова вагона та його візків у пакеті MATLAB Simulink, яка б дозволяла б досліджувати залежність цих коливань від швидкості рухомого складу, параметрів його системи підвішування та нерівностей рейкового шляху під цими колесами не тільки в режимі підтримки заданої швидкості, а й у режимі вибігу та перехідних режимах (розгоні, гальмуванні), та подальший аналіз отриманих на ній результатів.

Виклад основного матеріалу

У роботі [20] представлено розрахункову схему коливань вагона у вертикальній повздовжній площині (рис. 1) та наведено систему диференціальних рівнянь коливань кузова та візків вагона з урахуванням нерівності залізничної колії.

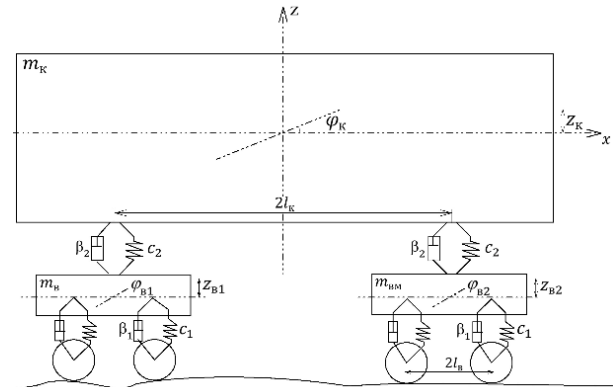


Рис. 1. Розрахункова схема коливань вагона у вертикальній повздовжній площині β_1, β_2 – коефіцієнти в'язкого тертя ресорного та центрального підвішування відповідно; c_1, c_2 – жорсткості ресорного та центрального підвішування відповідно; $z_{в1}, z_{в2}, z_k$ – підстрибування візка; $\varphi_{в1}, \varphi_{в2}, \varphi_k$ – галоупування візка та кузова; $l_в, l_к$ – половина відстані між колісними парами візка та між візками кузова; $m_{вм}, m_в, m_k$ – маси немоторного візка, моторного візка та кузова відповідно

На основі системи рівнянь, що наведено у роботі [20], складемо комплексну імітаційну модель руху вагона поїзда з осовою формулою 2o–2 та двоступінчастим ресорним підвішуванням, яка враховує вертикальні коливання кузова вагона та його візків, що викликані нерівностями колії, як за сталої швидкості, так і під час розгону та гальмування.

Опис імітаційної моделі. Моделювання виконувалося у системі MATLAB Simulink. Модель (рис. 2) складається з двох основних блоків: блоку «Задання швидкості», у якому задається швидкість за допомогою перемикачів позицій контролера машиніста поїзда, що дозволяє моделювати розгін, підтримання швидкості та гальмування поїзда, блоку «Нерівності», у якому розраховуються нерівності залізничної колії, та блоку «Коливання».

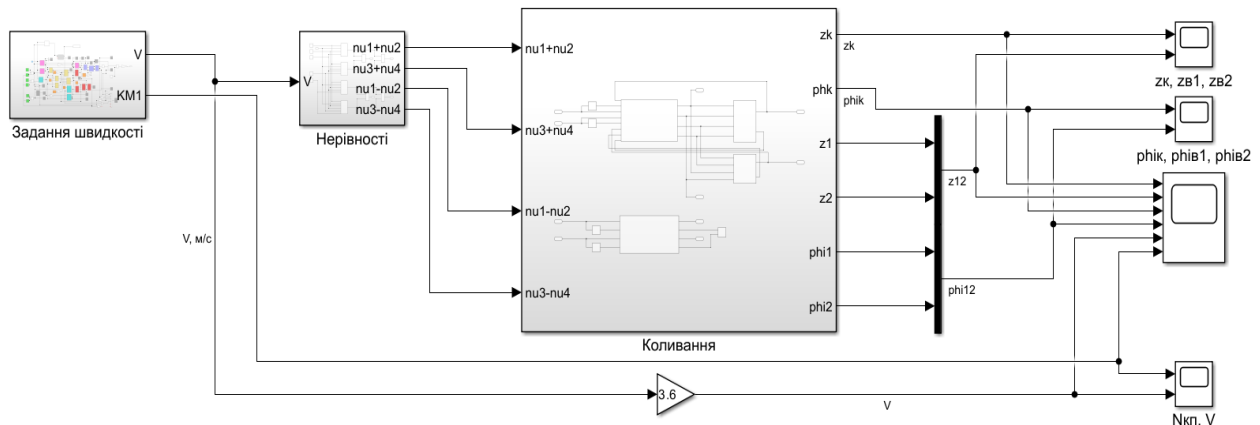


Рис. 2. Комплексна імітаційна модель коливань рухомого складу

Блок «Нерівності» приймає на вхід швидкість вагона поїзда та складається з чотирьох блоків, кожен з яких задає двогорбу нерівність під відповідною колісною парою з урахуванням фазового зміщення, зазначеного у роботі [20]. Блок формування нерівності для першої колісної пари першого візка вагона наведено на рис. 3. Оскільки стандартна формула гармонічних коливань $A \sin(\omega t + \delta)$ не дає бажаної синусоїдальної форми при частоті, що змінюється у часі, формулу двогорбої нерівності (1) було адаптовано під ці умови (2).

$$\eta = a_1 \sin(\omega t + \delta) + a_2 \sin(3\omega t + \delta) \quad (1)$$

де $\omega = 2\pi V/L_d$ – радіальна частота впливу стиків шляху за швидкості руху вагона, δ – фазове зміщення,

$$\eta = a_1 \sin(2\pi \int f(t)dt + \delta) + a_2 \sin(6\pi \int f(t)dt + \delta) \quad (2)$$

де $f = \omega/(2\pi)$ – частота впливу стиків шляху за швидкості руху вагона.

На виході блока «Нерівності» формуються загальні нерівності під кожним з візків, які використовуються у розрахунках величин вертикальних коливань.

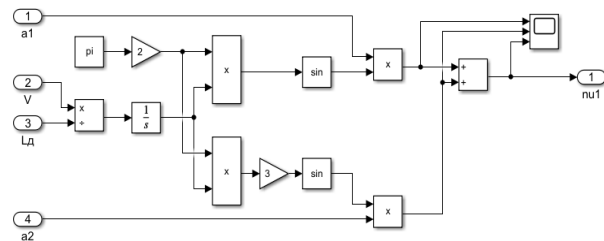


Рис. 3. Структура блоку задання двогорбої нерівності

Таблиця 1 – Параметри для розрахунків

Параметри	Умвне позначення	Прийняте значення
Маса кузова	m_K	64000 кг
Маса візка (моторний)	m_{BM}	8000 кг
Маса візка	m_B	4800 кг
База кузова	$2l_K$	17 м
База візка	$2l_B$	2,7 м
Вертикальна жорсткість першого ступеня ресорного підвішування	c_1	6567 кН/м
Вертикальна жорсткість другого ступеня ресорного підвішування	c_2	2659 кН/м
Демпфування першого ступеня ресорного підвішування	β_1	10 кНс/м
Демпфування другого ступеня ресорного підвішування	β_2	80 кНс/м
Довжина хвилі нерівності	L_d	25 м
Амплітуда першої гармоніки двогорбої нерівності	a_1	0,01 м
Амплітуда другої гармоніки двогорбої нерівності	a_2	0,008 м

Результати моделювання. Моделювання проводилося з використанням чисельного метода Рунге-Кутта із кроком, що обирається автоматично. Значення та змінних, що використовувалися під час моделювання, наведено у табл. 1.

На рис. 6 представлено результати експерименту, що проводився на сталих швидкостях (30 км/год, 60 км/год, 90 км/год) за наявності стикової нерівності з довжиною рейки $L_d = 25$ м та амплітудами двогорбої нерівності a_1 та a_2 . Із графіків видно, що зі збільшенням швидкості руху вагона частота коливань

Блок «Колівання», наведений на рис. 4, відповідає за розрахунок підстрибування та галопування кожного з візків вагона. На основі отриманих значень підстрибування візків у блоці розраховуються підстрибування та галопування вагона поїзда.

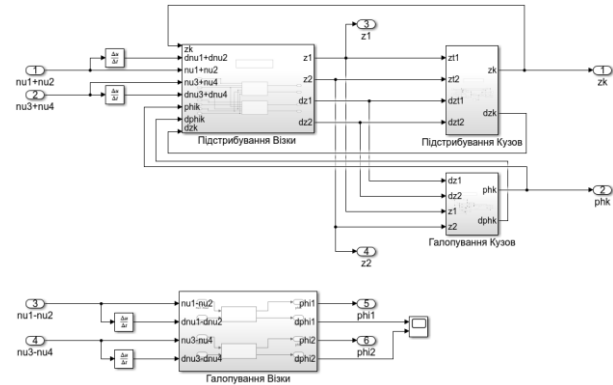


Рис. 4. Структура блоку «Колівання»

Структуру блоку розрахування підстрибування першого візка вагона наведено на рис. 5. Схожим чином реалізовані інші блоки моделі.

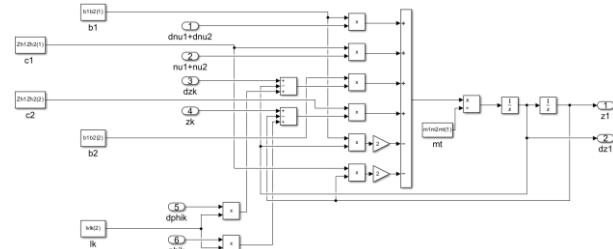


Рис. 5. Структура блоку розрахування підстрибування першого візка вагона

підстрибування також зростає. Так частоти коливань за швидкостей 30, 60 та 90 км/год дорівнюють 0.3333 Гц, 0.6667 Гц та 1 Гц відповідно. Залежність амплітуди коливань від швидкості не є лінійною через явище резонансу, яке виникає через рівність частот власних та вимушених коливань.

Так для даної моделі було визначено наявність резонансу на двох швидкостях – 40 та 118 км/год (рис. 7, 8). Амплітуди коливань за швидкості 40 та 118 км/год перевищують амплітуди коливань, наведених на рис. 5.

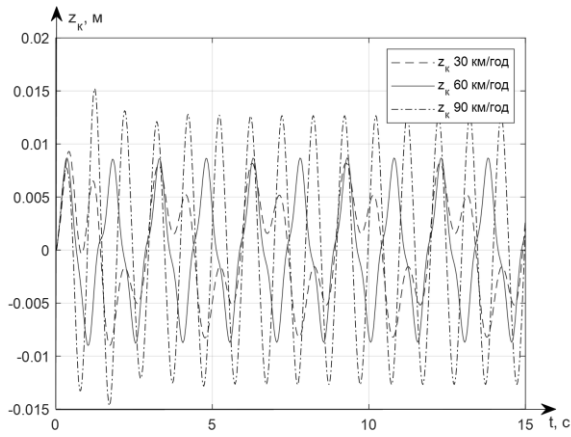


Рис. 6. Графіки підстрибування кузова вагона

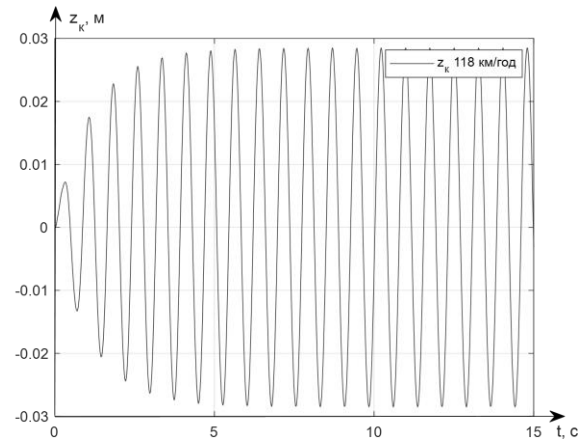


Рис.8. Графік підстрибування вагона на швидкості 118 км/год

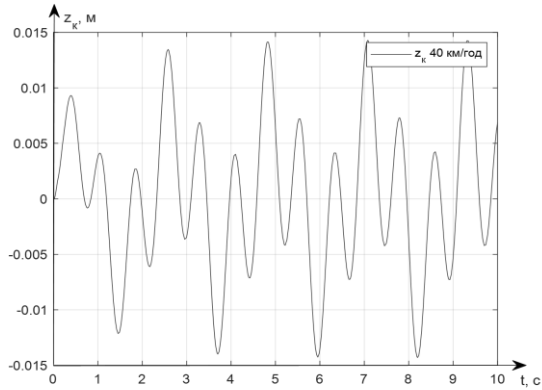


Рис. 7. Графік підстрибування вагона на швидкості 40 км/год

Промодельємо коливання вагона під час його руху перегонном, що має стикову нерівність. На рис. 9 представлені результати моделювання розгону до 130 км/год, підтримання швидкості, вибігу та гальмування та відповідних позицій контролера машиніста (N_{KM}) та гальмівного контролера машиніста ($N_{ГKM}$) на ділянці шляху, які отримані за допомогою блоку «Задання швидкості».

Графіки підстрибування кузова та візків вагона наведені на рис. 10. Спостерігається підвищення амплітуди коливань підстрибування кузова, які мають найбільші значення за швидкостей 40 та 118 км/год. Частоти підстрибування візків також підвищуються із підвищенням швидкості.

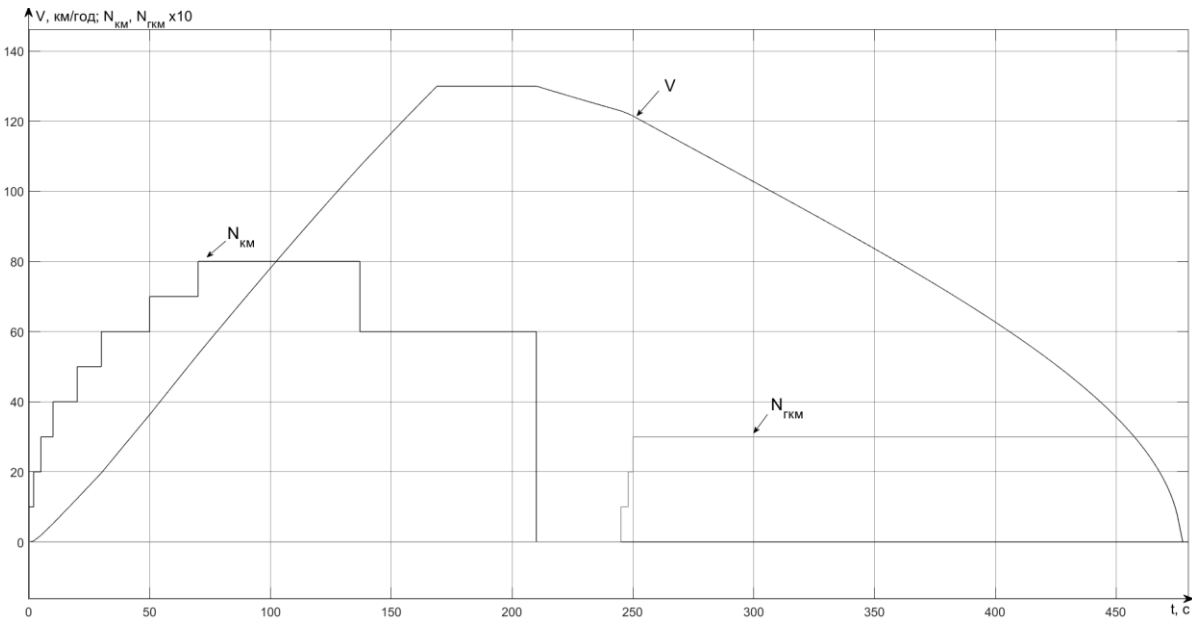


Рис. 9. Результати моделювання швидкості на ділянці шляху

На рис. 11 наведено графіки кута галопування кузова та візків вагона. Частоти цих коливань збільшуються зі збільшенням швидкості. Амплітуди коливань галопування приймають найбільші значення за швидкості 80 км/год.

Результати моделювання зі змінною швидкістю підтверджують результати, отримані за допомогою

використання сталої швидкості. Отже, отримана імітаційна модель може бути використані для подальшого проведення на ній досліджень коливальних процесів поїзда на обраних ділянках шляху, а також для отримання на ній даних для тренування нейронної мережі для раннього визначення резонансних коливань вагона поїзда.

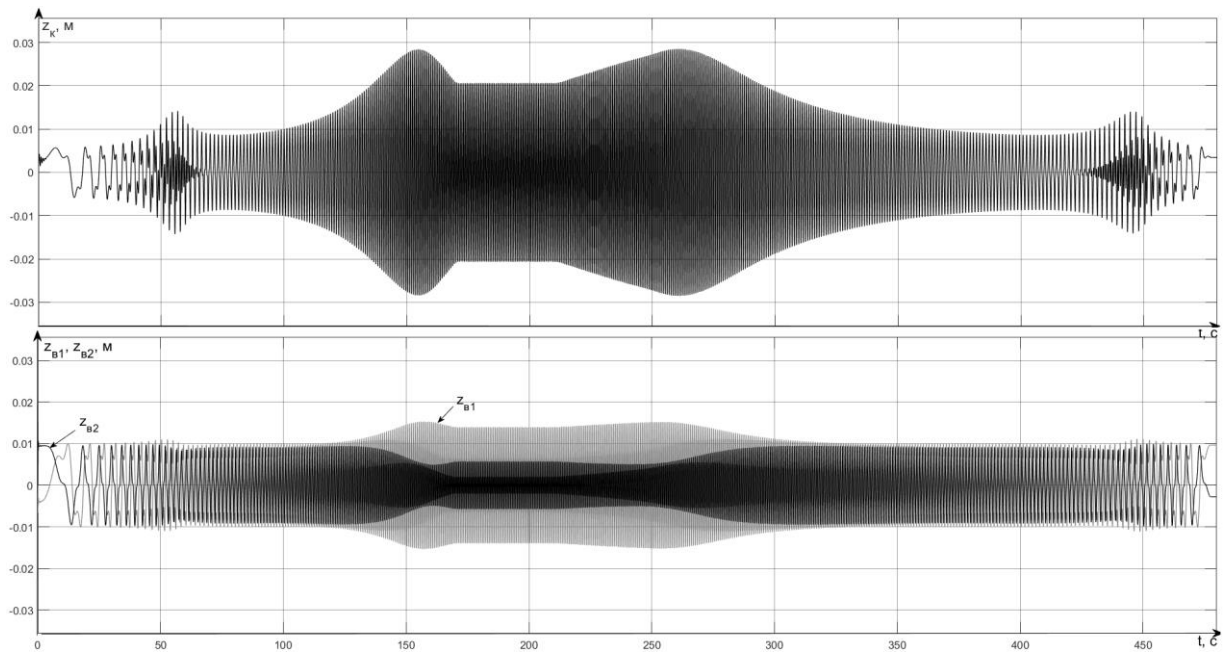


Рис. 10. Графіки підстрибування кузова та візків вагона поїзда

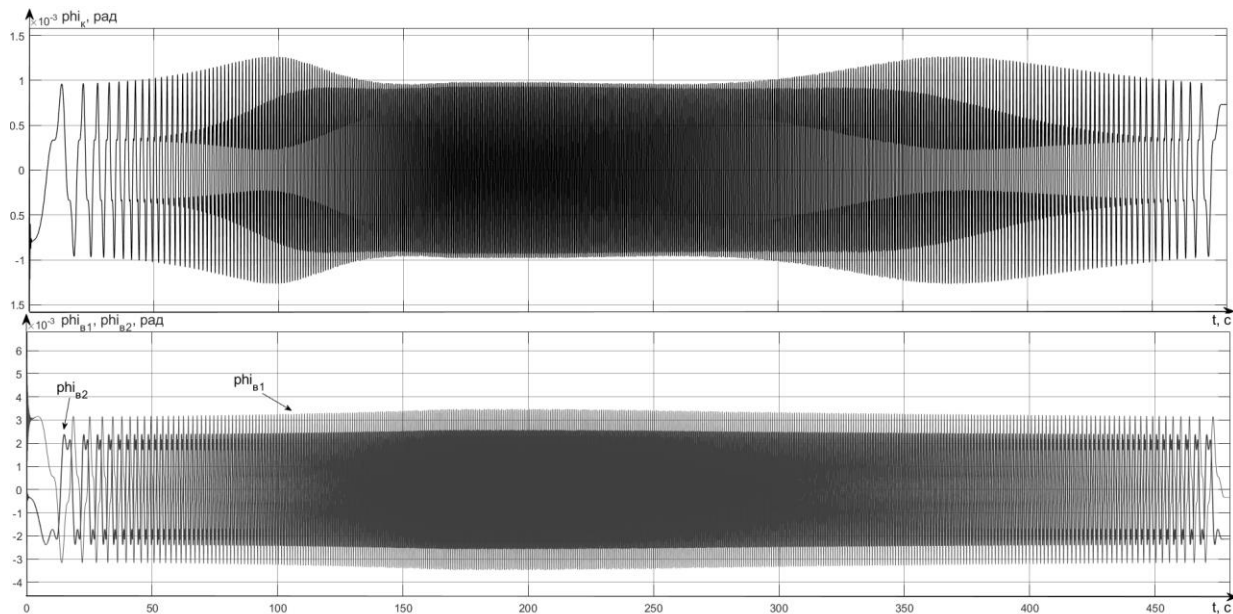


Рис. 11. Графіки підстрибування кузова та візків вагона поїзда

Висновки

У статті підкреслено необхідність досліджень, які направлені на зменшення зношення рейок та рухомого складу залізниць, а також підвищення швидкості руху та комфорту перевезення пасажирів. Розглянуто важливість урахування вертикальних коливань поїзда та причини їх виникнення.

Авторами запропоновано комплексну імітаційну модель руху вагона поїзда перегонном з урахуванням вертикальних коливань кузова вагона та його візків у пакеті MATLAB Simulink, яка дозволяє досліджувати залежність цих коливань від швидкості рухомого складу, параметрів його системи підвищення та нерівностей рейкового шляху під колісними парами не тільки в режимі підтримки заданої

швидкості, а й у режимі вибігу та перехідних режимах (розгоні, гальмуванні).

У результаті моделювання руху вагона з заданими параметрами було виявлено наявність резонансних вертикальних коливань підстрибування у кузова та першого візка вагона на швидкостях 40 та 118 км/год, а також резонансних коливань галопування кузова на швидкості 80 км/год.

Отримана імітаційна модель може бути використана для подальшого проведення на ній досліджень коливальних процесів поїзда на обраних ділянках шляху, а також для отримання на ній даних для тренування нейронної мережі для раннього визначення резонансних коливань вагона поїзда. Розроблена модель може бути використана у СППР бортових систем керування рухомого складу поїздів України.

СПИСОК ЛІТЕРАТУРИ

1. Інформаційне агентство УНІАН. Мінінфраструктури про стан потягів "УЗ": на 90% – металобрухт URL: <https://www.unian.ua/economics/transport/metalobruht-u-mininfrastrukturi-vislovilisya-pro-stan-potyagiv-uz-novini-ukrajina-11603764.html>
2. Каюкін В. О. Управління конкурентноспроможністю пасажирських перевезень залізничним транспортом в Україні: Проблеми та аналіз. Молодий вчений, 10 (62), 707-711. <https://molodyvchenyi.ua/index.php/journal/article/view/3871>
3. Lin, J.; Wang, K.; Zhai, W. Impact vibration behavior of railway vehicles: A state-of-the-art overview. Acta Mech. Sin. 2021, 37, 1193–1221.
4. Сокол Э. Н. Крушение железнодорожных поездов (Судебная экспертиза). Монография. – К.: Феникс, 2007. – 355 с.
5. Заковоротный А. Ю. Синтез автоматизированной системы управления подвижным составом на основе геометрической теории управления и нейронных сетей: дис. ... д-ра техн. наук : спец. 05.13.07 Нац. техн. ун-т "Харьков. политехн. ин-т". – Харьков, 2017. – 433 с. URL: <https://repository.kpi.kharkov.ua/handle/KhPI-Press/28330>
6. Костриця С. А. Математична модель вагона дизель-поїзда ДПКр-2 / С. А. Костриця, Ю. Г. Соболевська, А. Я. Кузишин, А. В. Батіг // Вісн. Дніпропетр. нац. ун-ту залізн. трансп. ім. акад. В. Лазаряна. 2018. – № 1 (73). – С. 15–25.
7. Кузишин А. Я. Удосконалення методів визначення основних динамічних показників моторвагонного рухомого складу на стадії проектування: дис. Доктор філософії : спец. 273 Залізничний транспорт : Дніпровський нац. ун-т. залізничного транспорту ім. ак. В.А. Лазаряна – Дніпро, 2019. – 144 с.
8. Garg, Vijay Kumar, and Rao V Dukkipati. Dynamics of Railway Vehicle Systems. New York: Academic Press, 1984.
9. João Pombo, Jorge A.C. Ambrósio, Miguel T. Silva. A new wheel-rail contact model for railway dynamic.Vehicle systems dynamics. – 2007. – P. 31-49.
10. Wang, X., Lu, Z., Wen, J. et al. Kinematics modelling and numerical investigation on the hunting oscillation of wheel–rail nonlinear geometric contact system. Nonlinear Dyn 107, 2075–2097 (2022). <https://doi.org/10.1007/s11071-021-07103-w>
11. Sun J., Jiao W., Ur Rehman A. Hunting frequency variation mechanism and its effect on carbody hunting stability for railway vehicles. Acta Mech. Sin., Vol.39, 523046 (2023). <https://doi.org/10.1007/s10409-023-23046-x>
12. Silva P., Mendes J., Seabra E, Pratas P. Railways Passenger Comfort/Discomfort: Objective Evaluation. New Research on Railway Engineering and Transportation, 2023. DOI: 10.5772/intechopen.111704
13. U.S. Department of Transportation. Federal Railroad Administration. Measuring the Displacement Environment between a Locomotive and Trailing Car. Final Report Report May 2020.
14. Лукашова Н. П. Підвищення ефективності роботи ходової частини рейкового міського транспорту з використанням електромеханічних амортизаторів: дис. ... канд. техн. наук: спец. 05.22.09 Електротранспорт: Харківський нац. ун-т міського господарства ім. О. М. Бекетова – Харків 2020 – 162 с.
15. Abood K. H. A., Khan R. A. Railway carriage simulation model to study the influence of vertical secondary suspension stiffness on ride comfort of railway carbody. Proc. IMechE Vol. 225 Part C: J. Mechanical Engineering Science, 2011 <https://doi.org/10.1177/0954406211399809>
16. Yokesh K. S., Nandakumar N. Modelling and Sensitivity Analysis of Influencing Parameters in Displacement of Dynamic Bodies. Engineering and Technology Journal, 2021, p. 879-883 DOI: 10.47191/etj/v6i5.02, I.F. – 6.39
17. Klimenda F., Skocilas J., Skocilasova B., Soukup J., Cizek R. Vertical Oscillation of Railway Vehicle Chassis with Asymmetry Effect Consideration. Sensors 2022, 22, 4033. <https://doi.org/10.3390/s22114033>
18. Yu Y., Zhao L., Zhou C. A new vertical dynamic model for railway vehicle with passenger-train-track coupling vibration. Proc IMechE Part K: J Multi-body Dynamics 2020, Vol. 234(1) p. 134–146. DOI: 10.1177/1464419319879790
19. Кудрявцев Н. Н., Белоусов В. П., Бурчак Г. П. Определение вертикальных возмущений, вызывающих колебания обрессоренных частей вагона при движении по рельсовому пути // Вестник ВНИИЖТ. 1982. № 5. С. 33–37
20. Решетнікова П. Е. Математична модель вертикальних коливань рухомого складу, викликаних нерівностями залізничної колії / Решетнікова П. Е., Заковоротний О. Ю. // Системи управління, навігації та зв'язку. Том 4 № 78, 2024, с. 42–45 <https://doi.org/10.26906/SUNZ.2024.4.042>

Received (Надійшла) 11.12.2024

Accepted for publication (Прийнята до друку) 05.03.2025

Modelling of vertical oscillations of a train arising during movement on a railway track with a joint unevenness

Polina Reshetnikova, Oleksandr Zakovorotnyi

Abstract. Rail transport has been one of the main carriers of goods and passengers in Ukraine for many years. Despite this, most trains and rails on the main routes are in poor condition. In the context of the ongoing war and the subsequent long post-war recovery, the issues of extending the life of existing rolling stock and saving energy resources are at the forefront. In addition, to maintain the competitiveness of trains relative to other modes of transport, the issue of increasing average speeds and passenger comfort on the railway is crucial. To solve these problems, it is necessary to develop comprehensive models that take into account factors that significantly affect fuel consumption, safety and passenger comfort. Such factors include rolling stock oscillations. This paper is devoted to the development of a simulation model of vertical oscillations (bouncing and galloping) of the body and bogies of a train carriage, which depend on unevenness of the railway track. The developed complex simulation model of train car movement on a track, taking into account vertical oscillations of the car body and its bogies in the MATLAB/Simulink package, allows to study the dependence of these oscillations on the speed of the rolling stock, parameters of its suspension system and unevenness of the track under the wheel pairs not only in the mode of maintaining a given speed, but also in the run-up mode and transient modes (acceleration, deceleration). The obtained simulation model can be used to further study the oscillatory processes of a train on selected sections of the track, as well as to obtain data for training a neural network for early detection of resonant vibrations of a train car, and can also be used in decision support systems (DSS) of on-board control systems for rolling stock of Ukrainian trains.

Keywords: mathematical model, decision support system, on-board control systems, rolling stock, oscillatory process, bouncing, galloping, resonant speed.

O. Fedorovich, L. Malieiev

National Aerospace University "KhAI", Kharkiv, Ukraine

MATRIX MODEL OF STRATIFIED REPRESENTATION OF THE HIGH-TECH PRODUCTS MODERNIZATION PROJECT BASED ON COMPONENT AND RISK-ORIENTED APPROACHES

Abstract. The subject of research in this article is the projects of modernization of high-tech products (HTP). The aim is to improve the quality of the HTP modernization processes by developing a stratified project representation based on the component and risk-based approaches. The methods of system decomposition, stratification, component approach and risk identification are used. The article analyzes the features of the modernization projects of the HTP, forms a model of the life cycle of the modernization project, that separates the stages of design and technological preparation of production, logistics tasks of procurement. The developed matrix model of stratified project representation allows planning work in two areas of detail: according to the main stages of the life cycle and the selected strata (structural, technological and risk strata).

Keywords: high-tech products, modernization projects, life cycle, multicomponent, risks, stratification, system model, design, 3D modeling, additive technologies.

Introduction

The main direction of development of high-tech production is the modernization of complex equipment, which allows to ensure its compliance with new functionalities and modern requirements of the sharpening process and to be competitive. The purpose of modernization of high-tech products (HTP) is to improve the technical and quality characteristics of equipment, expand its capabilities at lower costs than the purchase of new equipment [1-3].

Due to the need to ensure the competitiveness of industrial enterprises in Ukraine, the relevance of modernization in modern conditions is growing significantly.

The current level of high-tech products requires the search for effective methods of analyzing and managing projects for their modernization.

An analysis of the results of partial modernization projects of complex equipment shows that their implementation was accompanied by exceeding the planned deadlines, financial resources and, in some cases, failure to achieve the expected improvement or restoration of the functional quality of the equipment. Another complicated issue is the identification and supply of the necessary components for the modernization of sophisticated equipment, especially when the latest foreign-made components are used.

For example, in early 2022, when developing a project to modernize An-178-100R aircraft intended for the Ukrainian Air Force [4], it turned out that about a third of the components were russian-made [5]. Therefore, the basis of this project was the strategy of import substitution with the transition to Ukrainian analogues, as well as products from Canada, China, and the United States.

Such decisions require significant attention and project funds to ensure logistics processes. On the other hand, the production of new components and parts is developing at the enterprises that are part of Ukroboronprom[6].

Therefore, the article deals with the urgent task of forming a systematic view of the project for

modernization of high-tech products based on component and risk-oriented approaches. At the same time, it is necessary to highlight the innovative component of high-tech products and plan project activities for its design and subsequent manufacture.

Analysis of recent research and publications

The principles of integration in the design of complex products are implemented in the methodology of structural and parametric modeling, which consists not only in combining the stages of the life cycle, but also in the use of appropriate geometric modeling tools in the form of an appropriate mathematical apparatus. The main provisions of the method of reducing the area of design solutions are considered [7].

In automated design systems, the CALS technology has been improved to ensure the connection of the stages of building 3D models of a complex equipment object based on knowledge-based technologies during the cycle of their creation and data exchange between different components [8].

Among the generally accepted methods used in HTP system design are decomposition methods, methods for creating a block structure [9], morphological analysis, and modeling methods [10].

Recently, the importance of the modular (component) approach to the design of complex technical systems has increased, as libraries and component catalogs have been created in many areas [11].

Proven solutions are formalized in the form of reusable components and are involved in new developments, thereby reducing design risk, reducing development time and financial costs [12]. The widespread use of the component approach to the design of complex systems requires new models for describing an innovative product and the use of multiple (combinatorial) synthesis.

In the process of designing complex technical systems, a multi-level analysis is used [13].

The design object can be divided using the methods of functional and structural analysis and synthesis of technical systems [14, 15]. Based on the results of the functional analysis, the design task is divided into parts,

which is carried out in several stages: analysis of the functions of the technical system, synthesis of the functional structure, construction of a multilevel structure of the design object, identification of partial design tasks, development of a technical system design strategy.

When creating a new complex product, many risks arise. Identification and analysis of risks, especially in the early stages of technical systems development, requires special models and methods with the possibility of their automation [16].

Among the methods for assessing project risks are the following:

- statistical methods,
- methods of expert assessments,
- methods of simulation modeling,
- the method of building a "decision tree",
- sensitivity analysis, probabilistic analysis,
- fuzzy set theory [17].

Aims and objectives of the work

The modernization of complex equipment in Ukraine is carried out in the context of military aggression, and, as a result, limited funding, in the absence of workable methods and algorithms for solving this class of problems.

This fact necessitates the development of a methodological apparatus that will allow for reasonable decision-making support in the context of limited resources for the modernization of high-tech products during its planning and implementation.

From the review, it can be concluded that promising areas for improving the efficiency of HTP modernization project management are the systematic decomposition of the project product using a component approach, the use of additive manufacturing technologies for innovative components, and the assessment of relevant risks.

Thus, the purpose of the proposed study is to improve the quality of HTP modernization processes by developing a stratified project representation based on the component and risk-based approaches. The article solves the following tasks:

1. Formation of a life cycle model of the HTP modernization project.
2. Substantiation of the component approach to the management of the HTP modernization project.
3. Development of a matrix model for a stratified representation of the HTP modernization project.

Research results

Let us consider the features and form a *life cycle model* of the HTP modernization project.

Projects for the modernization of high-tech products are characterized by [18]:

- innovation of the project content;
- complex component composition,
- multilevel detailing;
- significant flows of design and control information;
- a high level of risk in the creation of high-tech products due to the innovation of projects and limited resources;

- the presence of uncertainty and a large number of internal and external random factors affecting the project.

The modernization project should achieve certain goals:

- Ensuring the advantage in properties (characteristics) of the modernized HTP compared to the prototype, which is maintained over the forecast period of time due to the introduction of the latest technologies;
- cost optimization at all stages of the project;
- applying the principles of cost minimization in the industrial production of HTPs to increase the profitability of production.

To ensure the competitiveness of products at all stages of the HTP, it is necessary to take into account the changing market environment of high-tech products consumers.

Changes in environmental parameters such as the volume of demand, advances in science and technology, and the growth of competitors' markets necessitate the modernization of existing models.

At each stage of the life cycle (LC) of HTP design or modernization, certain design activities are carried out [16]:

- analysis of external factors affecting the properties of the HTP prototype being modernized;
- development of the concept of creation and formation of the technical outline of the HTP;
- preparation and approval of organizational documents for the creation of the prototype;
- implementation of an advance design to determine the basic requirements for the prototype;
- development of the tactical and technical task;
- development of preliminary and technical designs;
- development of working design documentation;
- manufacture of a prototype and its submission for testing;
- conducting tests;
- technical and technological preparation of production;
- serial production of modernized equipment;
- commissioning;
- analysis of the functioning of the prototype (amendments to the documentation);
- checking its current scientific and technical level;
- carrying out repair or restoration work;
- making a decision on sample utilization.

Based on these standard stages, a life cycle model of the HTP modernization project was formed, taking into account the component composition and the use of additive manufacturing technologies (Fig. 1).

At the R&D stage, research and development work (RDW) is carried out to modernize products and determine the scale of distribution of new products (markets and their capacity), as well as the direction of work at the next stages.

In today's environment, the main direction of modernization of high-tech products is to improve their quality parameters during development and production. The parameters that characterize the realization of the development of any technical system change under the influence of the external environment. In most cases, this influence is stochastic.

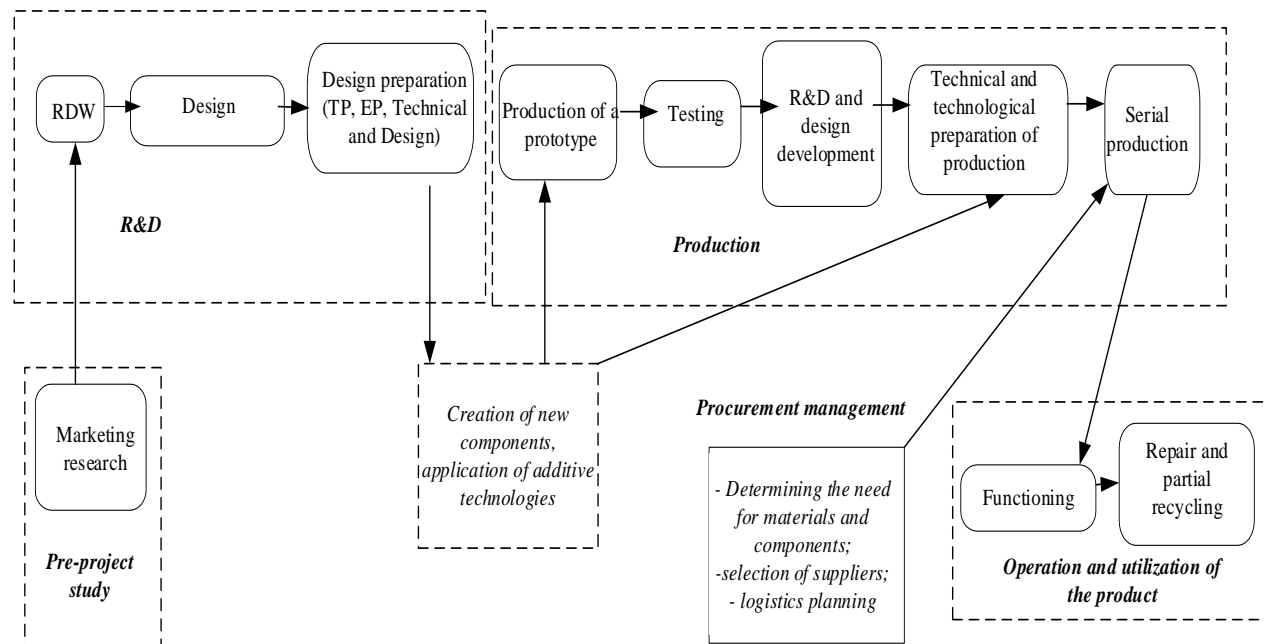


Fig. 1. HTP modernization project life cycle model

Therefore, at the initial stage of design, the technical outline of the HTP is synthesized - a set of structural and parametric data describing its most significant technical solutions and features, the component architecture of the modernized product, the composition and method of combining functionally related blocks and elements with each other are formed [19].

Typically, the initial data for creating new products is the technical specification provided by the customer.

The documentation created at the early stages of product modification contains information on the basis of which a database of many similar developments can be formed.

A preliminary design, compared to a technical design, provides a more complete picture of the product design.

Both the terms of reference and the preliminary design include textual, graphical, and quantitative (technical characteristics) information [20].

In addition, most of the information in the documents is interrelated: one information clarifies another or is an integral part of it.

Therefore, the use of ontological models in the form of semantic networks is effective for structuring information about complex modifiable products [21].

The design stage also includes experimental research, which involves the formation of a set and modeling of innovative components and ends with a technical design and preparation of working design documentation.

However, it should be noted that the first stage of modeling innovative products involves collecting various types of information:

- sketches,
- drawings,
- photographs and videos,

drawings, and often even a finished sample of an analog product (or a damaged sample of a part in equipment remanufacturing projects).

Based on the information received, a three-dimensional model is created in a 3D editor or CAD program.

Technical preparation of production involves the development of methods for manufacturing innovative products. Modern technologies in mechanical engineering have led to the use of 3D printing in the development of new parts and mechanisms. The use of 3D printing in the modernization and replacement of individual HTP elements allows solving design problems cheaper and faster. [22]. Thus, the expanded capabilities of 3D printing can increase the efficiency of HTP modernization projects at manufacturing enterprises of both general engineering and military and defense industries.

The quality of technological preparation is determined by minimizing labor resources and material costs for the implementation of technological processes, manufacturing and assembly of the product with the rational use of production assets.

At this stage, logistical tasks of procurement management arise, created to determine the needs for material and technical resources and ways to provide them [23].

This is especially important when procuring individual reusable components.

Let's consider the features of *component synthesis* in the management of high-tech product modernization projects.

Due to the multicomponent nature of HTP, a change in the characteristics of a set of interconnected elements leads to a change in the individual characteristics of their elements and, in general, may cause a decrease in the quality of performance of HTP functional tasks.

The compliance of HTP properties with the general requirements defined in the modernization project characterizes its scientific and technical level, which is assessed by the quality of performance of all tasks by the HTP sample and the level of its technical perfection. The general composition of HTP properties is a hierarchical structure. The indicators and characteristics of HTP properties are also complex in nature and have a corresponding hierarchical structure.

HTP properties are grouped into certain groups that define the relevant qualities, for example:

- the group of purpose indicators characterizes the purpose and scope of the HTP;
- the group of performance indicators and serviceability characterizes the adaptation of the HTP to reliable operation and preservation of its serviceability;
- the group of technical and economic solutions determines the feasibility of implementing a certain variant of a given modernization model, etc.

The main groups that determine the quality of the sample characterize the features (tactical and technical, technical and economic, operational and technical, production and technological) of the HTP improvement, which are used to compare its scientific and technical level with the prototype level.

The coordination of customer and contractor requirements is reflected in the technical task, which sets out the basic requirements and characteristics of the product. This includes the data necessary to form the HTP component architecture.

At the initial stages of a project, important decisions are not always made on the basis of sound reasoning regarding the choice of components of a complex product and the overall design scheme. Errors and shortcomings made in this process are difficult to eliminate at subsequent stages of the life cycle, and their elimination is associated with significant expenditures of project resources. In this regard, the pre-project work is often associated with a great deal of uncertainty in the initial data.

All project management decisions are focused on obtaining a new component architecture for the modernized product. Therefore, the project also has a component architecture, in which the project component and the product components are inextricably linked and subordinated to the global project goal. A project component is a set of managerial, organizational, technical and production actions aimed at creating a separate structural element (component) of a complex technical product [18]. From the point of view of design activity, a component is a process associated with a certain element of the architecture of a complex product, where the design process involves not only the creation of simple components, but also the combination and integration of complex ones.

The peculiarity of the component approach is that a project component simultaneously characterizes a complex product (product) and the process associated with its creation.

The project component in the information representation is an elementary information element of the project.

This information representation links the product architecture and the project structure behind its creation.

It should be noted that the component approach has the greatest impact on design decisions at the stages of general (preliminary) project planning, feasibility study, product design, and production planning for high-tech products. Therefore, we will consider these stages in the following.

Let's form a *stratified view* of the HTP modernization project.

Stratification involves representing HTP as a complex system in many aspects.

Generally, the following strata of complex system decomposition are used:

- target,
- functional,
- mathematical,
- information,
- algorithmic,
- strata of hardware and software [24].

The description of the HTP at each stratum is usually presented in the technical documentation for the creation or operation of a complex system.

To solve the problem of component synthesis of a complex product, the functional stratum and the technical stratum are important. It is on this stratum that the functional and technical requirements (quantitative characteristics) for individual, including innovative, HTP components are considered. Based on these requirements, individual components are searched for modernization.

For the HTP modernization project, the stratification is applied to other aspects:

- Structural. Component-based project management determines the structural stratum, which reflects the hierarchical structure and types of HTP components;
- Technological. Types of components with different levels of innovation require appropriate technological solutions;
- Risk. In addition to typical project risks, modernization projects require procedures for managing partial risks associated with innovative components.

The proposed matrix model of stratified project presentation has two levels of elements (Fig. 2):

- horizontal, which is based on the main stages of the project (in accordance with the LCA of innovative products) that require appropriate solutions in HTP modernization projects;
- vertical, whose elements correspond to the three strata mentioned above.

Thus, at the general planning stage and the feasibility study stage, the HTP is decomposed into components and partial parameters, taking into account functional, structural and parametric characteristics. As a result, the properties of the components to be modified are displayed. This makes it possible to subsequently search for relevant precedents (analog products) at low levels of decomposition and find precedents with a high level of similarity of innovative components.

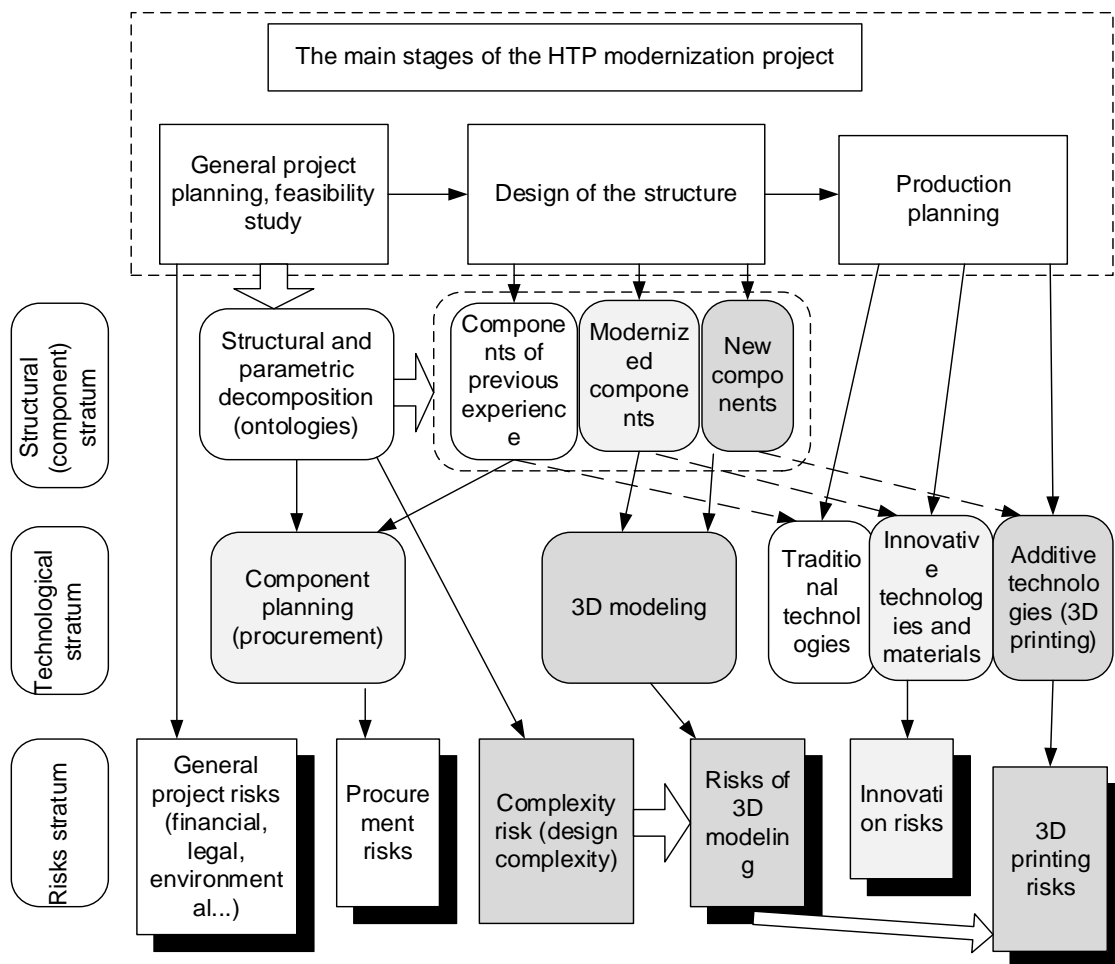


Fig. 2. Matrix model of stratified representation of the HTP modernization project based on component and risk-based approaches

Modern PLM systems implement sequential detailing during HTP design in CAD/CAM/CAE systems, which then requires systematic component synthesis.

Data decomposition and synthesis are provided by universal algorithms implemented in the design system and based on information in databases. This approach does not allow the application of standard procedures for innovative HTP components. The use of an ontological approach facilitates the process of creating a model of a complex product and allows for the coordination of parallel and sequential work of specialists on the project of its modernization.

According to the component approach, three types of components are distinguished in the product architecture based on the main information from the technical specifications [18]

- components of past experience - not subject to modernization. For them, there is a complete set of documentation and technology and experience in their production, or the project provides for the procurement of such components as finished structural elements;

- new (innovative) components - these are the components that pose the greatest risk both in design and in their production;

- components to be modernized - a combination of "old" and "new" components.

The production planning stage also requires dividing project activities according to these types of components.

Let's consider the elements of the technological stratum. At the general planning stage, the reusable components face the challenges of supply and the formation of appropriate logistics channels. At the feasibility study stage, optimization models should be used to compare two design solutions: procurement of components of a certain assortment or manufacturing of components using additive technologies.

At the design stage, for components that have a certain degree of novelty (new or modernized), the technical documentation may be incomplete or heterogeneous (data on foreign analogues, individual drawings or visual representations of the product). Modern technologies in mechanical engineering have led to the use of artificial intelligence, 3D modeling and printing in the development of new parts and mechanisms. In turn, the resulting models are the basis for further manufacturing of modeled parts using additive technologies.

Innovative technologies should also be applied at the production planning stage. The use of 3D printing allows optimizing the geometry of parts, reducing weight and increasing efficiency. 3D printing of metal components creates complex parts that are difficult to

manufacture using traditional methods, reducing waste and costs. Additive technologies are being actively implemented in the manufacture of modernized samples of high-tech products, as well as for the repair and maintenance of complex equipment. Generative design and shape optimization create parts with optimal parameters, taking into account functional, economic, and operational requirements [22]. However, despite the possibility of creating complex shapes and structures, existing additive manufacturing technologies may face limitations in the size and dimensions of products, which may limit their application in specific projects. In addition, the high cost of equipment and materials is also a significant barrier to the widespread use of additive manufacturing in the aviation and defense industries.

Consideration of risks in the management of the HTP modernization project, in particular risk assessment and management, is a mandatory attribute of the military equipment modernization system in both peacetime and wartime. Risk management should take place throughout the entire project life cycle.

Let's consider the causes and interconnection of partial project risks associated with the multicomponent composition of HTP.

The procurement processes of reusable components cause supplier selection risks, logistics risks (transportation), and communication risks [25].

At the design and production stages, scientific, technical, and production and technological factors are in effect. The scientific and technical factors should take into account the degree of novelty of the HTP, which depends on the planned degree of modernization. Production and technological factors cause the risk of integrating elements of complex equipment, which in turn is caused by both the complexity of the design and the presence of new components. Due to the innovativeness of modified HTPs and the corresponding complexity of modeling, the volume and complexity of geometric models of the structure, systems and equipment for the production and modernization of HTPs is significantly increasing. The actual characteristics formed after integration do not always coincide with the planned ones. There is a risk of assembly that affects the ability of HTP elements to perform the specified functions.

The quality of a printed part depends heavily on the digital model of the part used for printing. Therefore,

the quality of modeling a three-dimensional part before printing plays a very important role. In addition, production and technological factors include risks associated with the use of new materials when using additive manufacturing technologies.

It should be noted that these partial risks arise at successive stages of the project's LC, i.e. they have a synergistic effect, which complicates the procedures for their quantitative analysis.

Conclusions

The identified features of HTP modernization projects, such as the complexity and multicomponent nature of the project product, high level of innovation and uncertainty at the design stage, necessitate the use of system, component and risk-based approaches to managing these projects.

A life cycle model of the HTP modernization project has been formed, which, based on the typical stages of the life cycle of innovative products, separates the stages of design and technological preparation of production, logistics tasks of procurement of components.

These processes have their own peculiarities due to the multi-component nature of the HTP and determine the consideration of individual project activities from the perspective of the component approach.

The systemic approach requires consideration and management decisions to be made on the following strata: structural, technological, and risk strata.

The developed matrix model of the stratified representation of the HTP modernization project allows planning the project activities in two areas of detail: the main stages of the life cycle and the three selected strata, respectively.

The presence of components of various types of novelty and uncertainty requires the use of 3D modeling and printing technologies in the design and manufacture of innovative components of this type, which is reflected in all strata and the corresponding stages of the project's life cycle.

In the risk stratum, partial risks of the project due to the factors of multicomponent and innovative HTP are highlighted. The direction of further research is to form a model for quantifying these risks and studying their synergistic effect.

REFERENCES

1. Захарченко В. І., Єрмак С. О. Методологічні засади створення організаційно-технологічних систем у високотехнологічному виробництві // Економіка: реалії часу. – 2021. – №. 3 (55). – С. 49-60. DOI: 10.15276/ETR.03.2021.6
2. Lindgren M., Bando H. Scenario Planning: The link between future and strategy. Palgrave Macmillan UK. – 2002. – 180 p. DOI: 10.1057/9780230511620.
3. Xiao X., Joshi S. Decomposition and sequencing for a 5-axis hybrid manufacturing process // International Manufacturing Science and Engineering Conference. American Society of Mechanical Engineers. – 2020. – Vol. 84256. – DOI: 10.1115/MSEC2020-8385
4. У Києві презентували перший серійний літак Ан-178-100Р. Укрінформ. 29.12.2021 [Електронний ресурс] - Режим доступу: https://www.ukrinform.ua/rubric-other_news/3376665-u-kiyev-prezentuvali-persij-serijnij-litak-an178100r-infografika.html (дата звернення: 21.11.2024). – Назва з екрана
5. Президент ДП "Антонов": з Boeing у нас робочі стосунки. Економічна правда. 14 вересня 2020р. [Електронний ресурс] - Режим доступу: <https://www.epravda.com.ua/publications/2020/09/14/665044> (дата звернення: 21.11.2024). – Назва з екрана

6. УКРОБОРОНПРОМ | РІЧНИЙ ЗВІТ 2020, URL: https://ukroboronprom.com.ua/storage/documents/Report_ua.pdf25
7. Яблонський П. М. Деякі питання узагальнення засобів геометричного моделювання для проектування технічних об'єктів // Сучасні проблеми моделювання. – 2019. – №. 13. – С. 192-198. DOI: <https://doi.org/10.32347/0131-579x.2020.97.16-22>
8. Konotop D. I., Zinchenko V. P. 3D-models design concept of complex technical objects using knowledge-based technology // Механіка гіроскопічних систем. – 2017. – Вип. 34. – С. 5-13. DOI: 10.20535/0203-3771342017130222
9. Amaran S. et al. Simulation optimization: a review of algorithms and applications // Annals of Operations Research. – 2015. – Vol. 240. – No. 1. – Pp. 351–380. DOI: <https://doi.org/10.1007/s10479-015-2019-x>
10. Федорович О. С., Лутай Л. М., Малєєва Ю. А., Замірець Я. О., Пісклова Т. С. Моделювання комплексного формування запасів військової техніки в зоні воєнного конфлікту з використанням компонентного методу // Авіаційно-космічна техніка і технологія. – 2023. – № 2. – С. 56-66. DOI: <https://doi.org/10.32620/akt.2023.2.06>
11. Upare P., Kalatkar A., Shelke R. Review on study of design optimization for additive manufacturing of various mechanical/machine components // IOP Conference Series: Materials Science and Engineering. – 2022. – Vol. 1259, – No. 1. – Pp. 012039. DOI: <https://doi.org/10.1088/1757-899x/1259/1/01203963>
12. Кравцова Д. Ю., Зюган У. І. Пошук оптимальних рішень для технічних систем в умовах невизначеності зі комп'ютеризацією розрахунків у табличному процесорі // Гірничий вісник. – 2024. – Вип. 112. – С.63-68. doi: 10.31721/2306-5435-2024-1-112-63-68
13. Abdulsamad H., Peters J. Hierarchical decomposition of nonlinear dynamics and control for system identification and policy distillation // Proceedings of the 2nd Conference on Learning for Dynamics and Control. In: Proceedings of Machine Learning Research. – 2020. – Vol. 120. – P. 904-914.
14. Узунов О. В. Системне представлення складних технічних об'єктів в задачах аналізу та синтезу // Вісник Національного технічного університету. Серія: Машинобудування. – 2016. – №. 1. – С. 126-132. DOI: 10.20535/2305-9001.2016.76.68755
15. Dzwigol H. Research methodology in management science: Triangulation // Virtual Economics. – 2022. – Vol. 5. – No. 1. – Pp. 78-93.
16. Демідов Б. О., Величко О. Ф Кучеренко Ю. Ф., Куцак М. В. Управління проектами зі створення зразків озброєння та військової техніки в умовах прояву факторів невизначеності та ризику // Озброєння та військова техніка – 2016. – № 2(10). – С.15-19.
17. Raskin L., Sira O., Parfeniuk Y., Bazilevych K. Development of methods for supply management in transportation networks under conditions of uncertainty of transportation cost values // Eureka: Physics and Engineering. – 2021. – No. 2. – Pp. 108-123. DOI: 10.21303/2461-4262.2021.001691
18. Федорович О.Є., Яшина О.С., Белецький І.В. Компонентне проектування аерокосмічної техніки. X.: Національний аерокосмічний університет «ХАІ». – 2012. – 180с.
19. Нікул С. О., Головань В. Г., Головань А. В. Моделі забезпечення процесу прогнозування обрису складної технічної системи при її проектуванні // Збірник наукових праць. Одеса : ВА. – 2016. – Вип. 1 (5). – С. 17-22.
20. ДСТУ 3974-2000 Система розроблення та поставлення продукції на виробництво. Правила виконання дослідно-конструкторських робіт. Загальні положення. Режим доступу: https://www.dnu.dp.ua/docs/ndc/standarts/DSTU_3974-2000.pdf (дата звернення : 06.04.2024)
21. Федорович О. С., Малєєв Л. В. Онтологічна модель структури та параметрів компонент високотехнологічних виробів у проєктах їх модернізації // Сучасний стан наукових досліджень та технологій в промисловості. – 2024. – № 1(27). – С. 179–191. DOI: 10.30837/ITSSI.2024.27.179
22. Yang L., Zhang J., Yang Y. Advances and challenges in the additive manufacturing of aerospace materials // Frontiers of Materials Science. – 2020. – No. 14(3). – Pp. 295-319. DOI: 10.1088/2631-7990/ad5424
23. Fedorovych O., Kritskiy D., Malieiev L., Rybka K., Rybka, A. Military logistics planning models for enemy targets attack by a swarm of combat drones // Радіоелектронні і комп'ютерні системи. – 2024. – № 1. – С. 207-216. DOI: <https://doi.org/10.32620/reks.2024.1.16>
24. Задоров В. Б. Підхід до створення технології попереднього системного проектування КІС підприємств // Управління розвитком складних систем. – 2010. – №. 1. – С. 56-64.
25. Lytvynenko D., Malyyeyeva O. (2022). Risk management in projects of restoration the regional transport structure on the basis of participants' communication // Сучасний стан наукових досліджень і технологій в промисловості. – 2022. – № 2 (20). – С. 44–51. DOI: <https://doi.org/10.30837/ITSSI.2022.20.044>

Received (Надійшла) 26.11.2024

Accepted for publication (Прийнята до друку) 12.02.2025

Матрична модель стратифікованого уявлення проекту модернізації високотехнологічних виробів на основі компонентного та ризик-орієнтованого підходів

О. Є. Федорович, Л. В. Малєєв

Анотація. Предметом дослідження в статті є проекти модернізації високотехнологічних виробів (ВТВ). Мета - підвищення якості процесів модернізації ВТВ шляхом розробки стратифікованого уявлення проекту на основі компонентного та ризик-орієнтованого підходів. Використовуються методи системної декомпозиції, стратифікації, компонентного підходу та ідентифікації ризиків. В статті проаналізовано особливості проектів модернізації ВТВ, сформовано модель життєвого циклу проекту модернізації, яка відокремлює етапи конструкторської та технологічної підготовки виробництва, логістичні задачі закупівель. Розроблена матрична модель стратифікованого уявлення проекту дозволяє планувати роботи на двох напрямках деталізації: відповідно основних стадій життєвого циклу та виділених страт (структурної, технологічної та страті ризиків).

Ключові слова: високотехнологічні вироби, проекти модернізації, життєвий цикл, багатокomпонентність, ризики, стратифікація, системна модель, проектування, 3D-моделювання, адитивні технології.

Р. М. Царьков, Р. О. Єрмілов, О. В. Шефер

Національний університет «Полтавська політехніка імені Юрія Кондратюка», Полтава, Україна

ВИКОРИСТАННЯ ВИСОКОВОЛЬТНИХ ГЕНЕРАТОРІВ, КЕРОВАНИХ НАПРУГОЮ, ДЛЯ ПІДВИЩЕННЯ ТОЧНОСТІ СИСТЕМ ФАЗОВОГО АВТОПІДСТРОЮВАННЯ ЧАСТОТИ

Анотація. В роботі розглядаються проблеми систем фазового автопідстроювання частоти (ФАПЧ), які є одним з основних вузлів сучасної апаратури зв'язку. З аналізу останніх досліджень і публікацій авторами виявлено, що найкращі генератори, керовані напругою (ГКН) досі реалізуються здебільшого на дискретних схемах, які можуть вимагати напруги живлення до 30 В, тому перед розробниками сучасних систем ФАПЧ або радіочастотних систем постає завдання з'єднання низьковольтних мікросхем ФАПЧ та високовольтних ГКН. В роботі розглянуто структуру типової типової петлі високочастотних змішаних систем ФАПЧ, які містять як повністю цифрові блоки, так і як подільники в зворотному зв'язку та фазові детектори, так і прецизійні аналогові схеми – вузли накачки заряду та ГКН. Приведені характеристики основних блоків змішаних ФАПЧ, намічені способи їх удосконалення. Показаний приклад доступної схеми високої інтеграції. Виділено основні ознаки системи ФАПЧ, що впливають на якість її роботи – час захвату, фазовий шум та паразитні частоти, рекомендовані підходи включення елементів схем для покращення цих характеристик. Наведено приклади рекомендованих конфігурацій активних фільтрів із попереднім фільтруванням, з інвертуючим і неінвертуючим підключенням підсилювачів. Визначено переваги та недоліки інвертуючої та неінвертуючої топології схем. В роботі приділено увагу вибору операційного підсилювача для максимального використання можливостей активного фільтра в структурі ФАПЧ. Визначено, що крім смуги пропускання важливим є врахування таких основних параметрів як щільність шуму напруги, струмовий шум, вхідний струм та діапазон вхідної синфазної напруги. В результаті аналізу було виявлено, що для досягнення найкращих результатів роботи ФАПЧ потрібен операційний підсилювач з низькою щільністю шуму напруги для зниження загального фазового шуму, невеликим вхідним струмом для зменшення похибки та визначеним діапазоном вхідних напруг для можливості роботи з однополярним живленням. Наведений приклад та характеристики повністю інтегрованої високоякісної системи ФАПЧ, де широкопasmовий ГКН є інтегрованим безпосередньо у кристал мікросхеми.

Ключові слова: фазове автопідстроювання, генератор, фазовий шум, щільність шуму напруги, детектор.

Вступ

Постановка проблеми. Одним з основних вузлів сучасної апаратури зв'язку є система фазового автопідстроювання частоти (ФАПЧ). Зазвичай ця система використовується при створенні локальних генераторів у радіоприймачах або передавачах, крім того, її можна застосовувати для розподілу сигналу тактової частоти та зменшення його шумів, а також при створенні джерел тактового сигналу для високошвидкісних АЦП (аналого-цифрових перетворювачів). В наш час внаслідок розвитку технологій виробництва інтегральних схем зменшується напруга живлення пристроїв, зокрема схем ФАПЧ та інших вузлів обробки змішаного сигналу. Однак практичні технології створення генераторів, керованих напругою (ГКН) — основних вузлів ФАПЧ — змінюються не так швидко, як це потрібно, тому виникає дисбаланс елементів системи.

Аналіз останніх досліджень і публікацій. В роботах авторів [1-5] розглядається сучасний стан розробки ФАПЧ, переваги та недоліки будови різних архітектур цих систем. З оглянутої літератури можна зробити висновок, що найкращі ГКН досі реалізуються здебільшого на дискретних схемах, які можуть вимагати напруги живлення до 30 В. Тому перед розробниками сучасних систем ФАПЧ або радіочастотних систем постає завдання з'єднання низьковольтних мікросхем ФАПЧ та високовольтних ГКН.

Мета роботи – розглянути принципи створення високоякісних систем ФАПЧ та факторів, що впливають на точність таких систем, рекомендувати способи удосконалення їх параметрів.

Виклад основного матеріалу

Розглянемо структуру типової петлі ФАПЧ (рис. 1). Це система, у якій фазовий детектор керує ГКН так, щоб частота і фаза генератора точно відповідали частоті та фазі вхідного (опорного) сигналу. Система має коло зворотного зв'язку.

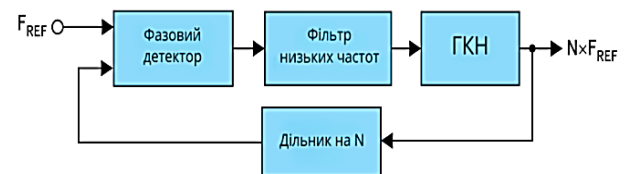


Рис. 1. Основна схема петлі ФАПЧ

Для інтеграції та згладжування позитивного або негативного сигналу помилки, а також для забезпечення стабільності петлі регулювання використовують фільтр низьких частот. Часто у зворотний зв'язок включають дільник частоти, щоб вихідна частота була кратною опорній частоті (в діапазоні реальних частот ГКН).

Оскільки ФАПЧ є системою з негативним зворотним зв'язком, сигнал помилки по частоті в стані рівноваги прагнучим до нуля. У результаті на виході ГКН буде вироблятися точна і стабільна частота. Система ФАПЧ може бути реалізована різними способами, з використанням повністю цифрових, повністю аналогових або змішаних схем, залежно від необхідного рівня шумів, розмірів і діапазону частот. Наразі високочастотні ФАПЧ містять як

повністю цифрові блоки, такі як подільники в зворотному зв'язку та фазові детектори, так і прецизійні аналогові схеми – вузли накачки заряду та ГКН. Змішані схеми ФАПЧ складаються з таких частин:

1. Фазово-частотний детектор – для формування сигналу помилки по фазі з сигналів опорної частоти і частоти зворотного зв'язку.

2. Петлевий фільтр встановлений для інтегрування імпульсів струму від схеми накачки заряду, забезпечуючи чисту напругу на керуючому вході ГКН.

3. Джерело опорної частоти. Призначене для подачі стабільної, точної частоти, під фазу якої підлаштовується радіочастотний вихідний сигнал. Зазвичай її отримують за допомогою мікросхеми температурно-компенсованого кварцового генератора.

4. Схема накачки заряду призначена для перетворення сигналу помилки в послідовність позитивних або негативних імпульсів струму, пропорційних помилці по фазі.

5. ГКН. Це генератор, у якого частота на виході залежить від напруги на його керуючому вході (V_{tune}).

ГКН має коефіцієнт перетворення, виражений в МГц/В. Основне рівняння, яке пов'язує частоту на виході ГКН з напругою на його вході, має вигляд

$$f_0 = f_c + K_V (V_{tune}), \quad (1)$$

де f_c – центральна частота коливань ГКН.

6. Дільник частоти на N . Ділить вихідну частоту так, щоб вона була рівною частоті фазового детектора або опорній частоті. Він може ділити строго на ціле число або ж на дробове. Дробовий подільник може бути реалізований шляхом перемикавання коефіцієнтів ділення в ціло-чисельному подільнику для отримання дробового середнього коефіцієнта. Наприклад, щоб отримати середній коефіцієнт ділення 4,25, потрібно три рази ділити на 4 і один раз на 5. Буде відраховано сімнадцять імпульсів і сформовано чотири імпульси, таким чином, виходить коефіцієнт ділення $17/5 = 4,25$.

На практиці кращі результати можуть бути досягнуті за допомогою техніки, яка використовується у шумоподібних перетворювачах високої роздільної здатності. Тому дробове ділення зазвичай реалізується з використанням сигма-дельта архітектури, перевагою якої є зменшення паразитних частотних складових. На рис. 2 показаний приклад доступної схеми високої інтеграції. Це функціональна схема мікросхеми ФАПЧ з дробовим коефіцієнтом ділення ADF4350 – ширококутового синтезатора з вбудованим генератором, керованим напругою (ГКН).

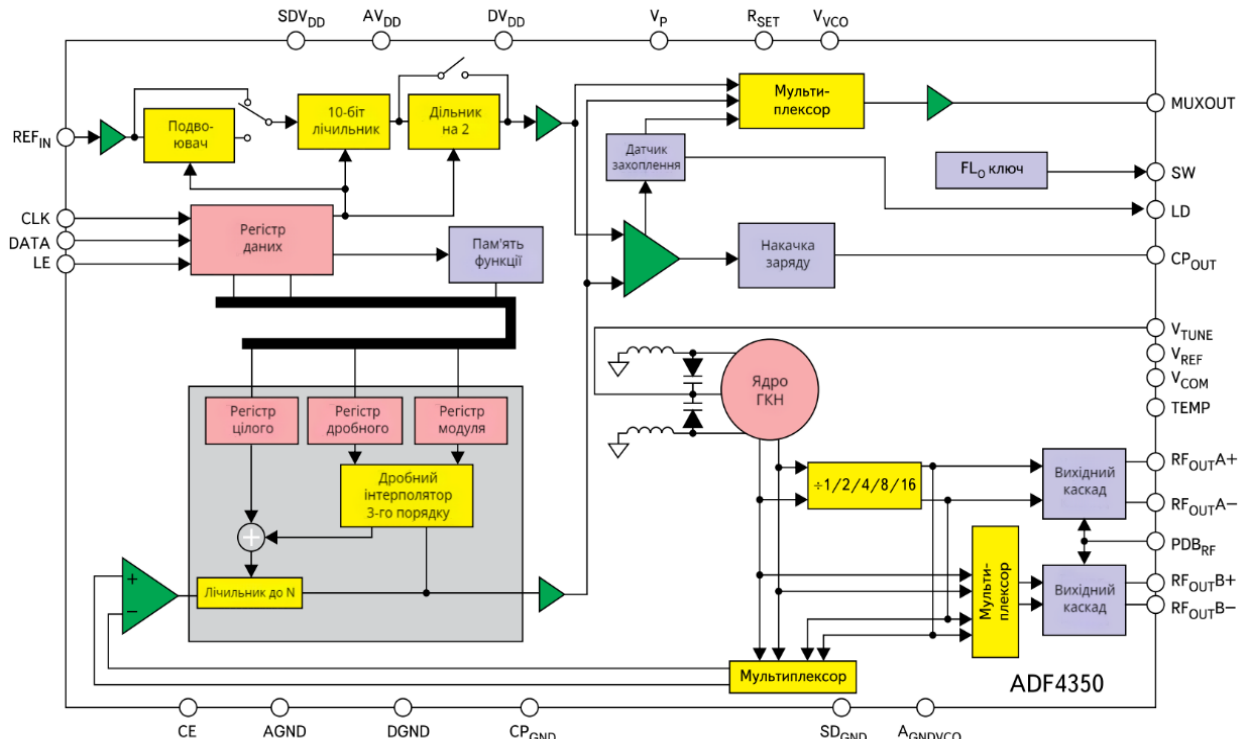


Рис. 2. Функціональна схема ширококутового синтезатора з ФАПЧ ADF4350

Діапазон вихідних частот синтезатора – від 137,5 до 4400 МГц. До основних характеристик системи ФАПЧ, що впливають на якість її роботи, можна віднести час захопту, фазовий шум та паразитні частоти. Час захопту – це час, протягом якого відновлюється ФАПЧ при стрибкоподібному переході від однієї частоти до іншої. Його можна задавати для встановлення частоти або фази. Ступінь важливості цього параметра залежить від сфери застосування.

Фазовий шум – це шум генератора або ФАПЧ, визначений для частотної області. Він являє собою середньоквадратичну суму шумів, які вносять різні компоненти ФАПЧ. У системі ФАПЧ з накачкою заряду подавляються шуми ГКН, які знаходяться в смузі петлевого фільтра. За межами смуги пропускання петлі домінують шуми ГКН. Паразитні частотні складові викликані періодичними імпульсами від схеми накачки заряду в керуючій напрузі ГКН. Вони

виглядають як зміщення частоти від несучої у фазово-частотному детекторі. У ФАПЧ з дробовим коефіцієнтом паразитні складові виникають також в результаті роботи дробового подільника.

Високоякісні ГКН залишаються одними з останніх електронних компонентів, які протистоять хвилям кремнієвої інтеграції. Лише в останні кілька років з'явилися ГКН, які повністю інтегровані в їхні радіочастотні чипсети. Однак базові станції для стільникового зв'язку, мікрохвильові системи зв'язку між двома точками, військові та аерокосмічні системи й інші складні високоякісні виробы все ще пред'являють надто високі вимоги до інтегрального кремнію, тому ці системи реалізуються на дискретних компонентах. Розглянемо причини цього явища, яких декілька.

У більшості доступних комерційних дискретних ГКН як керований елемент коливального LC-контру використовується варикап – діод зі змінною ємністю.

Якщо напруга на діоді буде змінюватися, то буде змінюватися й його ємність i , таким чином, зміниться резонансна частота коливального контуру. Будь-який шумовий сигнал на варикапі буде підсилюватися з коефіцієнтом перетворення варикапа K_V , вираженим у МГц/В, що призводить до появи фазового шуму. Щоб зберегти мінімально можливий рівень фазового шуму, необхідно робити K_V як можна найменшим. Однак, для досягнення досить великого діапазону перетворення потрібен достатньо високий K_V . Таким чином, для застосувань, де одночасно потрібні низький фазовий шум і широкий діапазон перетворення, виробники ГКН зазвичай розробляють генератори з малим коефіцієнтом перетворення і великим діапазоном вхідної напруги, щоб задовольнити ці суперечливі вимоги. Типовий діапазон зміни керуючої напруги для вузькосмугових ГКН становить від 0,5 до 4,5 В, а для широкосмугових ГКН він зазвичай лежить у межах від 1 до 14 В.

Генератор на коаксіальному резонаторі – це інший, спеціальний тип ГКН, у якому для досягнення наднизьких фазових шумів використовується дуже малий коефіцієнт перетворення і широкий діапазон керуючої напруги. Такі генератори зазвичай застосовуються у приватному мобільному радіозв'язку та наземних рухомих радіостанціях.

Більшість комерційних мікросхем синтезаторів з ФАПЧ мають вихід з накачкою заряду, максимальна напруга на якому не перевищує 5,5 В. Цього недостатньо для безпосереднього керування ГКН, який вимагає високої керуючої напруги, якщо петльовий фільтр реалізується лише на пасивних компонентах. Для отримання високих керуючих напруг необхідно використовувати активні петльові фільтри на операційних підсилювачах. Найпростіший підхід полягає в додаванні підсилювального каскаду після пасивного петльового фільтра. Однак тут є й недоліки – операційний підсилювач у інвертуючому включенні має низький вхідний імпеданс, який навантажує пасивний петльовий фільтр, що впливає на динаміку петлі.

При неінвертуючому включенні імпеданс досить високий, щоб не створювати навантаження на

фільтр, але в такому випадку підсилюється будь-який шум операційного підсилювача, який фільтр не зможе прибрати. Найкраще рішення – об'єднання підсилювального каскаду і фільтра в один активний фільтрувальний пристрій. Також тут доцільно додати попереднє фільтрування, щоб уникнути перенавантаження підсилювача короткими імпульсами зі схеми накачки заряду та обмеження діапазону вхідної напруги.

На рис. 3 показані два приклади рекомендованих конфігурацій активних фільтрів із попереднім фільтруванням, з інвертуючим і неінвертуючим підключенням підсилювача.

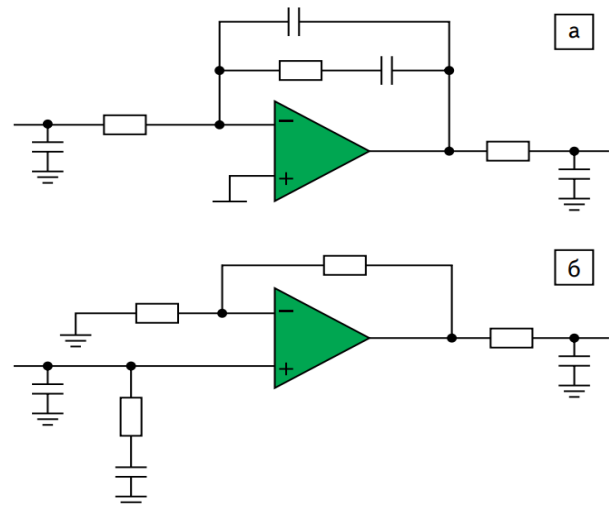


Рис. 3. Активний фільтр із попередньою фільтрацією: а – інвертуюче включення; б – неінвертуюче включення

Ці схеми підсилювачів містять інтегратори, котрі підтримують нульову розбіжність на входах фазового детектора. Напруга на виході такого підсилювача при розімкненій петлі ФАПЧ наближається до напруги одного з джерел живлення. Перевага інвертуючої топології полягає в змщенні вихідної напруги схеми накачки заряду на фіксовану величину, зазвичай рівну половині повного діапазону напруги на виході цієї схеми ($V_P/2$), що оптимально з точки зору спотворень.

У цьому випадку необхідно забезпечити чисту напругу змщення, бажано від спеціального низькошумного лінійного регулятора та адекватний розв'язок якомога ближче до вхідних виводів операційного підсилювача. Крім того, треба мати якомога меншими номінали резисторів у ланцюзі дільника напруги для мінімізації шуму. При використанні інвертуючого підключення потрібно переконатися, що мікросхема ФАПЧ дозволяє інвертувати полярність фазово-частотного детектора. Сімейство ADF4xxx має таку можливість. Неінвертуюча конфігурація петльового фільтра не потребує спеціального змщення, тому схема може бути більш компактною. Напруга накачки заряду замість змщення на фіксований рівень змінюється в межах усього робочого діапазону. Тому для цього типу фільтра дуже важливо використовувати операційний підсилювач із діапазоном вхідних напруг, що охоплює весь діапазон живлення.

Вибір операційного підсилювача — це ключовий момент для максимального використання можливостей активного фільтра. Окрім смуги пропускання тут необхідно враховувати такі основні параметри: щільність шуму напруги (в $nV/\sqrt{Гц}$), струмовий шум (в $nA/\sqrt{Гц}$), вхідний струм та діапазон вхідної синфазної напруги.

Безпосередньо впливає на частоту та фазу генерації вихідна напруга фільтра. Тому щільність шуму напруги операційного підсилювача дозволяє оцінити, який фазовий шум внесе активний фільтр.

Шум підсилювача додається як у смугі пропускання петлі ФАПЧ, так і поза нею. Його вплив найбільш помітний на частоті зрізу петлевого фільтра, особливо для підсилювачів з високою щільністю шуму напруги. Тому для досягнення цілей підсилювача і високовольтного ГКН — забезпечення мінімального фазового шуму, важливо знижувати шум підсилювача. У якісній схемі він не повинен перевищувати величини $10 nV/\sqrt{Гц}$. Струмовий шум, як правило, набагато менший у порівнянні з імпульсами струму помилки, тому його вплив значно слабший, ніж вплив шуму напруги.

Значну похибку у вихідний спектр ФАПЧ можуть вносити операційні підсилювачі з великим струмом зміщення порівняно з вихідним струмом фазово-частотного детектора. Щоб зберегти стабільну керуючу напругу ГКН та ФАПЧ у стані захвату, схема накачування заряду повинна компенсувати витік струму через вхід операційного підсилювача на кожному циклі роботи фазово-частотного детектора. Це призводить до модуляції керуючої напруги ГКН (V_{TUNE}) з частотою роботи детектора і викликає відхилення несучої частоти на цю величину. Збільшення вхідного струму зміщення підвищує модуляцію V_{TUNE} і збільшує амплітуду похибки.

Ще одним важливим параметром операційного підсилювача є діапазон синфазної напруги, або діапазон вхідної напруги. Часто на нього не звертають уваги, але він може викликати серйозні проблеми на завершальному етапі розробки.

Цей параметр визначає різницю між максимальною і мінімальною напругою на вхідних контактах і позитивним та негативним живленням відповідно. Старі операційні підсилювачі, що працювали при напрузі живлення ± 15 В, мали типовий діапазон вхідного синфазного сигналу ± 12 В. Більш пізні повільні додаткові вхідні рпр-каскади розширили діапазон синфазного сигналу до негативної напруги живлення у схемі. Незважаючи на те, що будь-який операційний підсилювач може працювати з напругою 0 В як негативною напругою живлення, необхідно обов'язково контролювати різницю між сигналом і напругами живлення. Наприклад, популярний підсилювач OP27 має діапазон вхідного сигналу $\pm 12,3$ В при напрузі живлення ± 15 В. Це означає, що напруги на входах мікросхеми не повинні підходити ближче до напруги живлення, ніж на 2,7 В. Таке обмеження на нижній стороні діапазону робить цей операційний підсилювач непридатним для роботи з великим розмахом сигналу на вході при однополярному живленні.

Можливість використання двополярного живлення значно розширює вибір операційних підсилювачів і спрощує рішення щодо вхідного струму. Якщо потрібне однополярне живлення, слід використовувати операційні підсилювачі з діапазоном вхідних напруг "від живлення до живлення", але більшість з них мають високий шум напруги. Отже, для досягнення найкращих результатів потрібен операційний підсилювач з низькою щільністю шуму напруги для зниження фазового шуму, невеликим вхідним струмом для зменшення похибки та діапазоном вхідних напруг "від живлення до живлення" для можливості роботи з однополярним живленням. Незважаючи на те, що активний фільтр додає шуми у ФАПЧ, у деяких застосуваннях його буферні властивості надають переваги порівняно з пасивним фільтром. Наприклад, якщо вхід ГКН має великий струм витоку, що спричиняє значну похибку, для її зменшення можна використовувати операційний підсилювач. Вихід операційного підсилювача з низьким імпедансом легко компенсує струм витоку входу ГКН. За нашого часу з'явилася альтернатива використанню активних фільтрів з високовольтними ГКН — це використання повністю інтегрованих високоякісних систем ФАПЧ [6]. У цьому випадку ГКН є інтегрованим безпосередньо у кристал. Компроміс між широким діапазоном перебудови та низьким фазовим шумом, що неминуче виникає, тут вирішується застосуванням широко-смугового ГКН. У системі ФАПЧ з накачкою заряду пригнічуються ті шуми ГКН, які знаходяться в смугі петлевого фільтра, а за межами смуги пропускання петлі будуть домінувати шуми ГКН.

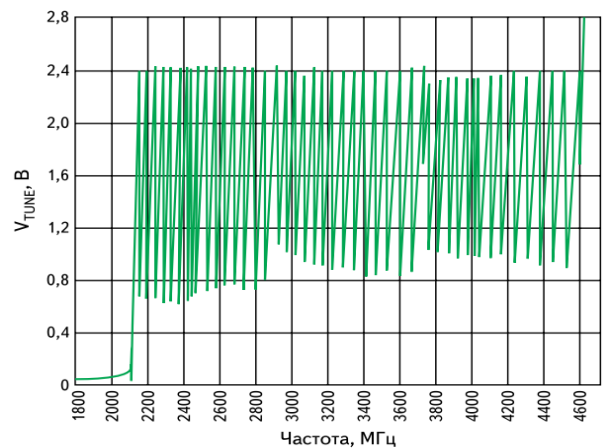


Рис. 4. Графік, що показує зв'язок керуючої напруги ГКН ADF4350 та частоти

В якості прикладу може виступати пристрій ADF4350, де на одному кристалі розміщено три окремих ГКН, кожен з яких має 16 перекриваючих піддіапазонів, що в результаті дає 48 піддіапазонів (рис. 4). При кожній зміні частоти відбувається автоматичне калібрування для вибору відповідного піддіапазону ГКН. Також вбудовано програмований вихідний дільник, який забезпечує повне перекриття діапазону частот від 137,5 МГц до частоти 4,4 ГГц. Дослідження показують, що якісні показники цього пристрою наближаються до рівня дискретних рішень за величиною фазового шуму 114 дБн/Гц при

відхиленні частоти 100 кГц і величини фазового шуму 134 дБн/Гц при відхиленні частоти 1 МГц.

Висновки

1. З аналізу літературних джерел авторами виявлено, що перед розробниками сучасних систем ФАПЧ або радіочастотних систем постає завдання з'єднання низьковольтних мікросхем ФАПЧ та високовольтних ГКН.

2. Виділено основні ознаки систем ФАПЧ, що впливають на якість її роботи – час захвату, фазовий шум та паразитні частоти, рекомендовані підходи включення елементів схем для покращення цих характеристик.

3. Визначено, що крім смуги пропускання важливим для вдосконалення ФАПЧ є врахування таких основних параметрів як щільність шуму на

пруги, струмовий шум, вхідний струм та діапазон вхідної синфазної напруги.

4. В результаті аналізу було виявлено, що для досягнення найкращих результатів роботи ФАПЧ в схемі активного фільтра потрібен операційний підсилювач з низькою щільністю шуму напруги для зниження фазового шуму, невеликим вхідним струмом для зменшення похибки та визначеним діапазоном вхідних напруг для можливості роботи з однополярним живленням.

5. Наведений приклад повністю інтегрованої системи ФАПЧ, де широкосмуговий ГКН є інтегрованим безпосередньо у кристал мікросхеми, який підходить для реалізації підвищення якісних характеристик системи. Дослідження показали, що якісні показники цього пристрою наближаються до рівня дискретних рішень за величиною фазового шуму.

СПИСОК ЛІТЕРАТУРИ

1. Zhibin Luo Jicheng Ding and Lin Zhao Adaptive Gain Control Method of a Phase-Locked Loop for GNSS Carrier Signal Tracking [Електронний ресурс] / Zhibin Luo Jicheng Ding and Lin Zhao // Hindawi. – 2018. – Режим доступу до ресурсу: <https://www.hindawi.com/journals/ijap/2018/6841285/>
2. Collins I. Phase-Locked Loop (PLL) Fundamentals [Електронний ресурс] / Ian Collins // AnalogDialogue – 2018. – URL: <https://www.analog.com/media/en/analog-dialogue/volume-52/number-3/phase-locked-loop-pll-fundamentals.pdf>
3. Bakic D., Wu J. Pll For Mmwave 5g / Daniel Bakic, Jinzhuo Wu // Lund UniversityLund University Publications. – Department Of Electrical And Information Technology Lund University – 2020. – P. 5-10.
4. Jonsson F. Design and Calibration of Integrated PLL Frequency Synthesizers / Fredrik Jonsson // Digitala Vetenskapliga Arkivet. – Royal Institute of Technology – 2008. – P. 84-87.
5. Mentze E. Understanding the basics of PLL frequency synthesis [Електронний ресурс] / Erik Mentze // PLANET ANALOG. – 2010. – URL: <https://www.planetanalog.com/understanding-the-basics-of-pll-frequency-synthesis/>
6. Царьков Р. М., Єрмілова Н.В. Підвищення точності функціонування фазового автопідстроювання за ознаковими характеристиками сигналів / Р.М. Царьков, Н.В. Єрмілова // Системи управління, навігації та зв'язку. Випуск 4 (70) – Полтава: Національний університет «Полтавська політехніка імені Юрія Кондратюка», 2022. – С. 162 – 165. – Режим доступу до ресурсу: <https://reposit.nupp.edu.ua/handle/PoltNTU/12048>
7. Basab Bijoy Purkayastha. A Digital Phase Locked Loop based Signal and Symbol Recovery System for Wireless Channel / Basab Bijoy Purkayastha, Kandarpa Kumar Sarma. — New Delhi: Springer, 2015. — 254 с.
8. Best R. E. Phase-locked loops: design, simulation, and applications (professional engineering) / Roland E. Best. — New York : McGraw-Hill Companies Inc., 2003. — 436 с.
9. Полікарівських О. І. Фазове коло як основа класифікації прямих синтезаторів частоти / О. І. Полікарівських // Вісник Хмельницького національного університету. — 2014. — № 5. — С. 133—139.

Received (Надійшла) 22.12.2024

Accepted for publication (Прийнята до друку) 26.02.2025

The usage of high-voltage VCG for increasing the accuracy of phase auto-adjustment systems of frequency

R. Tsarkov, R. Iermilov, O. Shefer

Abstract. Problems of phase-locked-frequency (PLL) systems are considered in the work, which are one of the main components of modern communication equipment. From the analysis of recent research and publications by the authors, it was found that the highest quality voltage-controlled generators (VCG) are still mostly implemented on discrete circuits that can require a supply voltage of up to 30 V, so the developers of modern PLL systems or radio frequency systems face the task of connecting low-voltage microcircuits PLL and high-voltage VCG. This work examines the structure of a typical loop of high-frequency mixed PLL systems, which contain both fully digital units, such as feedback dividers and phase detectors, as well as precision analog circuits - charge pump nodes and VCG. The characteristics of the main blocks of mixed PLLs, and methods of their improvement are outlined. An example of an available high-integration scheme is shown. The main features of the PLL system, which affect the quality of its operation, are highlighted - capture time, phase noise and parasitic frequencies, and approaches for including circuit elements to improve these characteristics are recommended. Examples of recommended active filter configurations with pre-filtering, inverting and non-inverting amplifier connections are given. The advantages and disadvantages of inverting and non-inverting scheme topology are determined. In the work, attention is paid to the selection of the operational amplifier for the maximum use of the capabilities of the active filter in the structure of the PLL. It was determined that, in addition to the bandwidth, it is important to take into account such basic parameters as voltage noise density, current noise, input current and the range of input common-mode voltage. As a result of the analysis, it was found that to achieve the best performance of the PLL, an operational amplifier with a low voltage noise density to reduce phase noise, a small input current to reduce error and a defined range of input voltages to be able to operate with a unipolar supply is required. An example and characteristics of a fully integrated high-quality PLL system is given, where the wideband VCG is integrated directly into the microcircuit crystal.

Keywords: phase self-adjustment, generator, phase noise, voltage noise density, detector.

S. Shapovalova, O. Baranichenko

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

MODELING OF FUNCTIONAL PROCESSES BETWEEN COMPONENTS OF A HIERARCHICAL CONTROL OBJECT

Abstract. The subject of the study in the article is a production knowledge representation system for representing a controlling model of a complex hierarchical object. The purpose of the work is to present a model of functioning of an automated system, which is based on a production model of knowledge representation and allows to ensure uninterrupted operation in conditions of changing the structure of the control object and/or the logic of its components. **Conclusions:** based on the production model of knowledge, a model of functioning of an automated system is proposed, which reflects the three-level organization of the hardware part of the management, allows to automatically form rules for reflecting changes in the structure of the control object and/or the logic of its components; the proposed model is implemented as a production system on the .Net Framework platform in C# and C for business logic, WPF for the graphical interface and CLIPS for logical output; the proposed model is tested on a test system of emergency lighting of a data center, which consisted of 128 components of the complex level, connected by 304 rules. A promising direction for further research is the development and creation of intelligent methods for controlling complex automated hierarchical objects.

Keywords: production systems, industrial dispatching control systems, inference system, CLIPS.

Introduction

Problem statement. Industrial dispatching control system (IDCS) is a system that performs the functions of automatic and dispatch control of an automation object. IDCS performs the following functions:

- 1) monitoring and control of the equipment status;
- 2) automatic equipment control;
- 3) dispatch control in unprotocolized situations;
- 4) storage of information about the system status, data archiving.

IDCS are widely used in energy, logistics, industry, access and control systems. An example is the power supply system of data processing centers. At the same time, the efficiency, accuracy, fault tolerance, and transparency of the operation of the automation object equipment increase.

To implement IDCS today, a three-level architecture is most often used. The field-level consists of mechanisms and devices that directly receive external information and perform actions on the equipment. This level includes sensors, switches, motors, electromagnets, relays, etc. The local-level automation consists of functional controllers (programmable logic controllers - PLC), which perform the initial processing of data received from field-level devices and implement actions on these devices. The dispatching level implements the functioning of all components of the automation object as a single system, and allows operators to coordinate the operation of the equipment in manual or automatic mode. It consists of servers and operator workstations.

There are various approaches to software implementation of automation, which are regulated by the IEC 61131-3 standard. At the local-level automation, modular programming approaches are most often used. In this case, each physical component of the field-level in the software system is represented by a high-level component that has its own properties and methods. In this case, a component-oriented approach is usually used, implemented in the Structured Text (ST) language. Another approach is, for example, cyclic-event

programming implemented using Ladder Diagram (LD) or Functional Block Diagram (FBD). However, with a large number of field devices and their interconnected operation, a problem arises, known as "spaghetti code". This is a phenomenon when a poorly designed program leads to poorly structured, confusing and difficult to understand code. Automatic control of complex systems requires monitoring of the states of components in real time. Such systems can be represented by hierarchical models, the elements of which belong to different types. As the hardware computing resources on which IDCSs are based are rapidly improving, control and management objects are becoming more complex and require processing of more data. Changing the logic of the components of the control object leads to the need to modify the corresponding software. There are cases when the replacement of the program code must occur in the operating mode without stopping the IDCS. To solve this problem, it is necessary to create methods and software for continuous analysis of current data for control in conditions of changing the IDCS structure.

Analysis of recent research and publications sources. Intellectualization of automation systems is one of the key areas of research in any sector of human activity. For example, one of the main, both in terms of importance and energy consumption, components of the food industry are refrigeration units. The safety of products and the amount of electricity consumed depend on the correctness of their operation algorithms. In works [1, 2] an approach to automating the recognition of emergency situations of these units based on neural network systems was presented and the effectiveness of this method was proven.

In work [3], the authors presented a methodology for developing and implementing an expert system in SCADA for a cable car in the Metro de Medellin (Colombia). The work presented an event-oriented structure of an expert system with fuzzy logic and provided its mathematical apparatus.

Many studies related to the use of artificial intelligence methods are carried out in the energy sector.

For example, in work [4], the feasibility of using methods for intelligent control and monitoring of power plants was substantiated. The authors gave examples of various expert models that are currently working in diagnostic systems for failures of power plants of various types. From the work, it can be concluded that the implementation and further development of production systems began in the 1980s and continues to this day. Mostly, fuzzy logic systems are used at these automation facilities. In the article [5], modeling and simulation of a SCADA system for a power grid were presented. However, this process was performed in software development environments. For such tasks, it is more rational to use specialized modeling and development environments. In the article [6], an architectural solution for building an uninterruptible power supply automation system was presented, the criteria and scenarios that must be implemented to ensure an automated process are described. Within the framework of the study, the authors argued the need to implement artificial intelligence approaches to solve the automation problem in uninterruptible power supply systems.

Task statement. The purpose of the article is to present a model of functioning of an automated system, which is based on a production model of knowledge representation and allows ensuring uninterrupted operation in conditions of changing the structure of the control object and/or the logic of its components.

To achieve the goal, the following tasks must be performed:

- 1) determine the representation of a complex hierarchical control object based on a production model;
- 2) develop a software implementation of the model of functioning of a hierarchical control object;
- 3) test the model on a probing IDCS.

Model of functioning of a hierarchical control object

The production inference system contains a resource for representing rules and a mechanism for processing them. To implement the production model, there are both standalone software tools, such as CLIPS (C Language Integrated Production System) or Drools, as well as libraries for all common programming languages, such as Jess (Java Expert System Shell). The improvement of software tools is aimed at reducing the computational resources for performing inference, so several inference engines are being developed for one programming language [7].

To implement the simulation system, it is advisable to choose the same tools that are used by real PLCs and automation servers. The basis of a significant amount of automation equipment at the local and dispatcher levels today are ST Microelectronics microcontrollers and microprocessors built on the ARM architecture. Most often, translators and compilers from the C language are used to program these devices, and therefore it is advisable to choose CLIPS for the implementation of the simulation system [8, 9].

This paper proposes a distributed three-level component-oriented production system approach (Fig. 1).

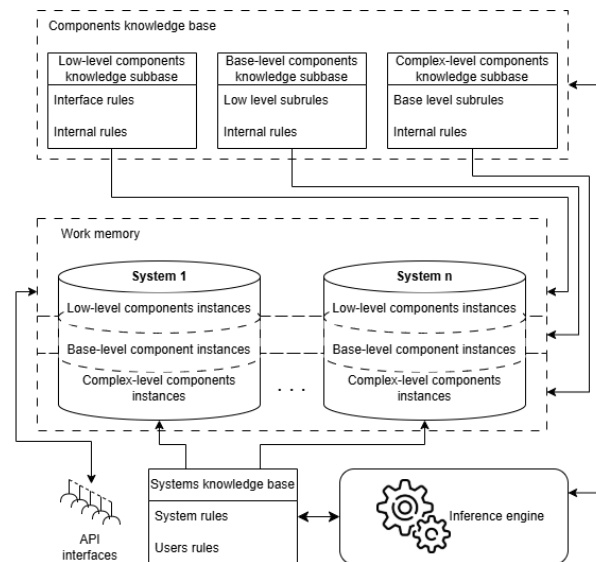


Fig. 1. Production System Architecture

The production system consists of three main components: working memory, knowledge base and inference engine. Working memory consists of a set of logically separated blocks, each of which represents a system of the automation object. Each such block is divided into three levels. The low-level components instances level contains instances of components that are responsible for a specific external hardware interface of the PLC (Fig. 2). Such interfaces include digital inputs/outputs (DI, DO), analog inputs/outputs (AI, AO) or protocol interfaces (UART, RS485, RS232, I2C etc.). In addition to standard components, this level includes timers, system interrupts, system errors, which are atypical system components.

The base-level components instances level contains instances of components that are represented at the hardware level by physical or abstract devices that have a single functional purpose and are common to any automation systems (Fig. 3). Examples of such devices include, for example, a dry contact, a button, a lamp, an electromagnet. Components of this level have a small number of states and can be fully described by a component, with no more than five interface variables. No more than five low-level components are aggregated in these components.

The complex-level component's instances level contains instances of components that are represented at the hardware level by physical devices, each of which consists of a set of base-level components. An example of such devices can be, for example, a circuit breaker with a magnetic trip and an NO/NC state contact. The components of this level do not differ significantly from the components of the base level in terms of architecture, but they have a large number of possible states that may differ depending on the specific automation tasks and have different operating logic.

Each component, in addition to its own characteristics, is inherited from a global component that contains system data, in particular a unique identifier, name, information about the logical level, and processing priority. The priority of processing components within one level consists of the following sequence of actions:

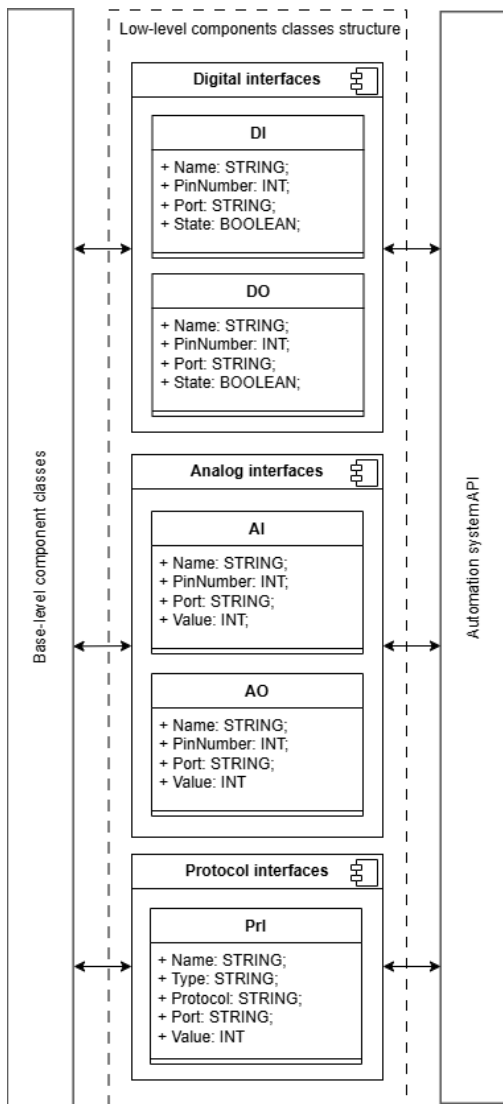


Fig. 2. Low-level Component Structure

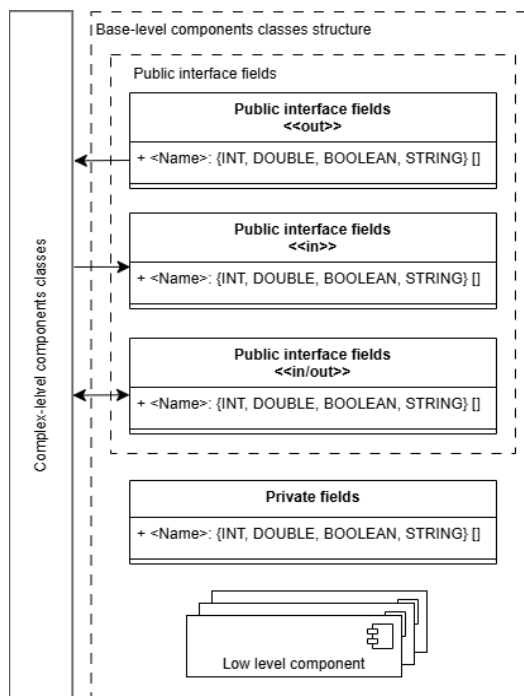


Fig. 3. Base-level Component Structure

1. Each component is assigned a priority.
2. A template is created that aggregates references to each of the components and separates the priority from them.
3. Templates are grouped into a processing queue, which is sorted during the logical inference process.
4. Before logical inference according to the rules describing the logic of the system, a system rule is launched that determines the order of their execution using the sorting method.

Production rules are presented in the format:

$$LHS \Rightarrow RHS,$$

where *LHS* (Left Hand Side) – conditional part that describes the state of the component at some moment of time; *RHS* (Right Hand Side) – the active part, which contains actions on the component and changes the value of its characteristics for further logical conclusion.

In automation systems, rules should represent the structure, characteristics and properties of equipment, rules for reacting to external events and changing the states of system components. An example of a production rule could be an expression of the following content:

```

if(
  and(
    (MagneticDecoupler.State = "On")
    (MagneticDecoupler.Error = "NoError")
    (PowerSwitch.Error = "NoError")
    (PowerSwitch.State = "1")
    (PositionSwitch.Error = "NoError")
    (PositionSwitch.State = "1")
    (ChangeStateProcess = "Start")
    (StateTime > OPERATION_TIME)
  )
)
=>
(send State "Undefined")
(send WaitState "ErrorTime")
(send TestIndicator "Off")
(send Error "ContactGroupSticking")
    
```

For correct logical inference, the following principles must be observed:

1. Each production rule of a component representation describes only one of its states.
2. For each component, a single default-production is represented - this is a rule that is executed if no other is executed.
3. The priority of production rules is set in ascending order from the low-level to the complex-level.
4. The complexity conflict resolution strategy (i.e., the one in which the most complex rule takes priority).

Software implementation

The presented modeling system was created on the .Net Framework 4.8 by C#, C++ and C languages. It consists of three main components: a graphical user interface (GUI), business logic, and a production system (Fig. 4).

CLIPS was chosen as a means of implementing the production system. This is due to the fact that CLIPS is implemented in the C language, which allows it to be easily integrated into software written in the C#, C, C++ languages, since all these languages compile in IL and executed in a single environment. Which, in turn, speeds up the work of the final software complex.

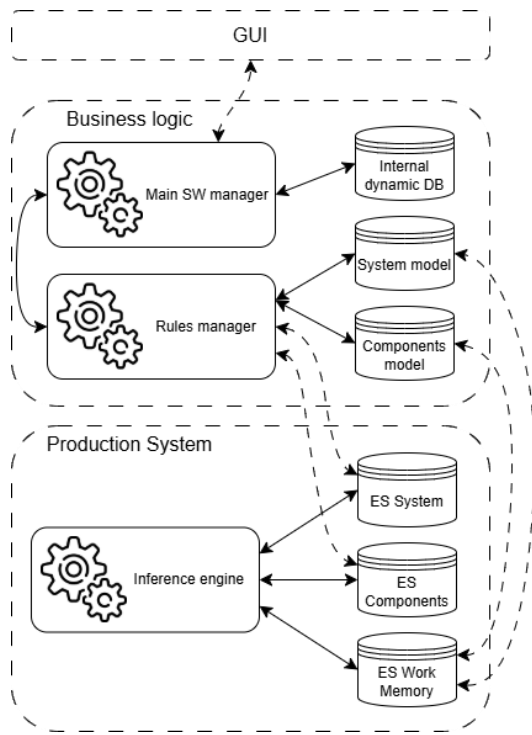


Fig. 4. Software Structure

The integration of CLIPS into the proposed modeling system was carried out through interface components written in the C language. This complicates the code, but allows to guarantee the correctness of data transfer and obtain high performance. The business logic of the system has databases that are a reflection of the working memory of the expert system, as well as a separate dynamic database for temporary storage of output chains that are necessary for automation simulation.

The graphical user interface is shown in figure 5. It is implemented using WPF technology and consists of three identical tabs, each of which is responsible for one of the three levels of abstraction. Each tab contains a workspace (1), a list of allowed components (2), and a set of rules (3).

By moving components from the list to the workspace, an automated system model is created. After that, the user specifies a set of rules that link the public interfaces of the components to each other (binding operation). After declaring the set of rules, the user can simulate the operation of the equipment. In this mode, he is given the opportunity to change external unbound interfaces through the context menu of a specific component.

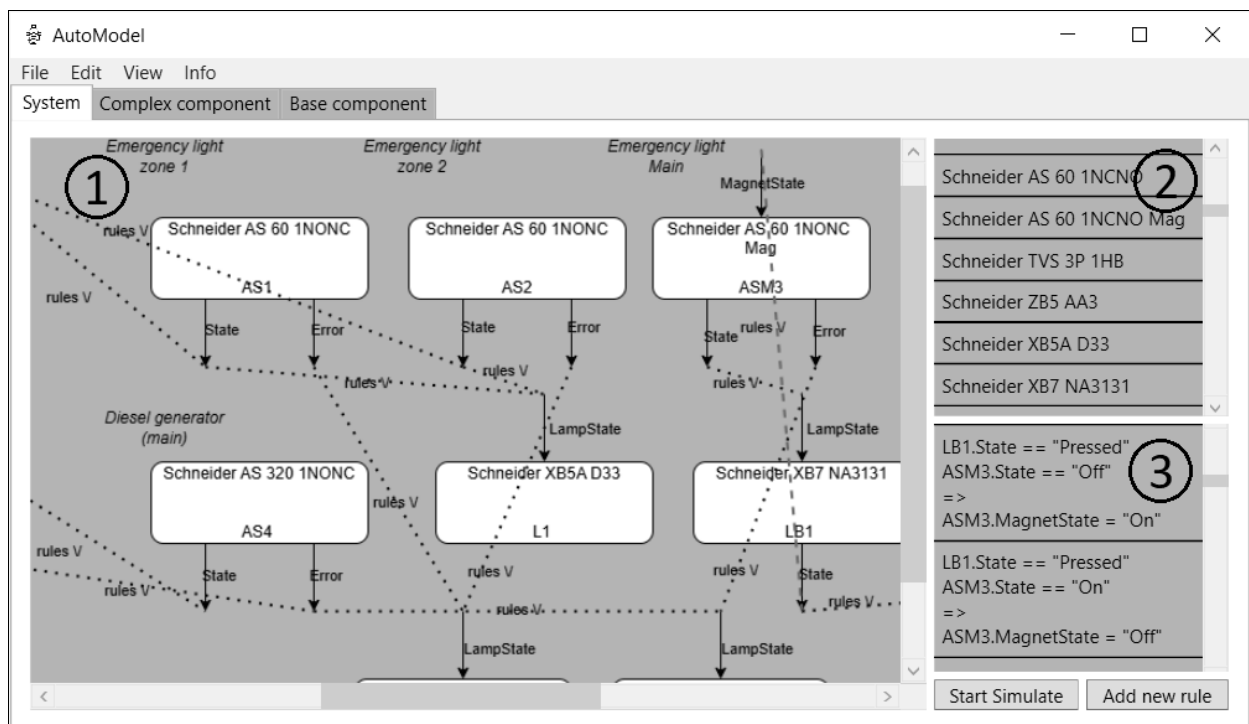


Fig. 5. Graphical User Interface

The figure also shows part of a simulated data center emergency lighting test system, consisting of a set of circuit breakers, lamps, contactors, and other field-level automation devices. The selected set of rules for emergency shutdown of the system shows the dependence of the state of the "Start/Stop" button LB1 with the main circuit breaker ASM3. When the button is pressed, the state of the circuit breaker changes, which in turn starts a chain of logical output regarding the state of the circuit breakers AS1 and AS2. In this case, depending

on their state, not only the lighting of the corresponding zones is turned on, but also other additional equipment is turned on. During the simulation, the user can interact with input devices (the LB1 button, as well as with additional contacts of the circuit breakers AS1, AS2, ASM3, AS4) via the context menu and check the correctness of the operation of the magnetic drive of the circuit breaker ASM3. In addition, during the testing process, the software sorts the current rules, so the user can view active rules and track how each of them affects

the current state of the system. At the same time, it is possible to pause the execution of some rules to monitor the system's reaction.

The low-level consists of 11 components, the work of which is implemented by 77 rules. This level includes Digital Input, Analog Input, Digital Output, Analog Output, RS232, RS485, I2C, SDIO, UART, USART, SDIO. The base-level consists of 27 components, the work of which is implemented by 254 rules. This level includes Switch NO, Switch NC, Digital Lamp, Dimmer Lamp, Magnet, Contactor, etc. The complex-level consists of 74 components, the work of which is implemented by 751 rules. This level includes various types of circuit breakers, magnetic starters, relays, frequency converters, network analyzers, fuses, control panels, data analyzers etc. At the beginning of development, the emergency power system consisted of 124 components of the complex-level, connected by 251 rules. During the development and testing process, 4 additional components and 53 rules were added. At the same time, during the testing of the system during the development process, 211 errors were detected

and fixed, of which 23 were critical and could lead to an emergency situation.

Conclusions

1. Based on the production model of knowledge representation, a model of the functioning of an automated system is proposed, which reflects the three-level (field, local and dispatcher) organization of the hardware part of the control, allows you to automatically generate rules to reflect changes in the structure of the control object and/or the logic of its components.

2. The proposed model is implemented as a production system on the .Net Framework platform in C# and C for business logic, WPF for the graphical interface and CLIPS for logical output.

3. The proposed model was tested on a system of emergency lighting of a data center, which consisted of 128 complex-level components connected by 304 rules.

A promising direction for further research is the development and creation of intelligent methods for controlling complex automated hierarchical objects.

REFERENCES

1. Khorolsky, V. et al. (2022). Kholodylni mashyny v systemi uzgodzhenoho upravlinnia elektrozpozhyvanniam kompleksu pidpriemstvo – promyslovyi kholodylnyk [Refrigeration machines in the system of consistent management of electric energy consumption of the complex enterprise-industrial refrigerator]. *Visnyk Khmelnytskoho natsionalnoho universytetu – Herald of Khmelnytskyi National University*, 313, 5, 200-212 (in Ukrainian). <https://doi.org/10.31891/2307-5732-2022-313-5-200-212>.
2. Khorolskyi, V. P. et al. (2023). Intelktualna pidsystema keruvannia pratsezdatnistiu kholodylnykh mashyn asutp promyslovykh kholodylnykv [Intelligent subsystem for control of refrigerators operating performance scada of industrial refrigerators]. *Obladnannia ta tekhnologii kharchovykh vyrobnystv – Food production equipment and technologies*, 46, 1, 80-88 (in Ukrainian). <https://doi.org/10.33274/2079-4827-2023-46-1-80-88>.
3. Quintero, O. L. et al. (2014). Design of a methodology for a SCADA expert system: massive transportation ropeway. *Memorias del XVI Congreso Latinoamericano de Control Automático*, 905-910. <https://doi.org/10.13140/2.1.4419.5842>
4. Mayadevi, N., Vinod Chandra, S. S., Ushakumari, S. A. (2014). Review on Expert System Applications in Power Plants. *International Journal of Electrical and Computer Engineering (IJECE)*, 4, 1, 116-126. <https://doi.org/10.11591/ijece.v4i1.5025>.
5. Faryal, A. et al. (2021). Modelling and Simulation of SCADA and PLC System for Power System Protection Laboratory. *A. Faryal et al. Electrical, Control and Communication Engineering*, 17, 1, 19-25. <https://doi.org/10.2478/ecce-2021-0003>.
6. Shapovalova, S., Baranichenko, O. (2021). Vyznachennia efektyvnosti mekhanizmiv lohichnoho vyvedennia [Hardware and software implementation of automatic decision support in power supply systems]. *Systemy upravlinnia, navihatsii ta zviazku – Control, navigation and communication systems*, 4, 66, 38–43 (in Ukrainian). <https://doi.org/10.26906/sunz.2021.4.038>.
7. Shapovalova, S. I. (2020). Formalizatsiia predstavlenia produktsiinykh pravyl v Erlang [Formalization of the Rules of Inference in Erlang]. *Matematychni ta kompiuterne modeliuvannia. Seriya: tekhnichni nauky – Mathematical and computer modelling. Series: Technical sciences*, 21, 125-139 (in Ukrainian). <https://doi.org/10.32626/2308-5916.2020-21.125-139>.
8. Riley, G. (2022). Adventures in Rule-Based Programming: A CLIPS Tutorial. *Kindle Edition*, 200.
9. Shapovalova, S. I., Mazhara, O. O. (2020). Vyznachennia efektyvnosti mekhanizmiv lohichnoho vyvedennia [Measuring efficiency of inference engines]. *Systemy upravlinnia, navihatsii ta zviazku – Control, navigation and communication systems*, 4, 62, 81–87 (in Ukrainian). <https://doi.org/10.26906/SUNZ.2020.4.081>.

Received (Надійшла) 11.12.2024

Accepted for publication (Прийнята до друку) 05.02.2025

Моделювання функціональних процесів між компонентами ієрархічного об'єкту управління

С. І. Шаповалова, О. М. Бараніченко

Анотація. Предметом дослідження в статті є продукційна система представлення знань для представлення моделі керування складним ієрархічним об'єктом. **Метою роботи** – представлення моделі функціонування автоматизованої системи, яка базується на продукційній моделі представлення знань і дозволяє забезпечити безперервність роботи в умовах зміни структури об'єкту управління та/або логіки роботи його компонентів. **Висновки:** на основі продукційної моделі представлення знань запропоновано модель функціонування автоматизованої системи, яка відображає трьохрівневу організацію апаратної частини управління, дозволяє автоматично формувати правила для відображення змін в структурі об'єкту управління та/або логіці роботи його компонентів; запропонована модель реалізована як продукційна система на платформі .Net Framework мовами C# та C для бізнес-логіки, WPF для графічного інтерфейсу та CLIPS для логічного виведення; запропонована модель апробована на тестовій системі аварійного освітлення центру обробки даних, яка складалась з 128 компонентів комплексного рівня, зв'язаних 304 правилами. Перспективним напрямком подальших досліджень є розробка та створення інтелектуальних методів для керування складними автоматизованими ієрархічними об'єктами.

Keywords: продукційні системи, автоматизована система диспетчерського управління, логічне виведення, CLIPS.

А. С. Янко, О. О. Крук, С. О. Гончаренко

Національний університет «Полтавська політехніка імені Юрія Кондратюка», Полтава

ТЕОРЕТИЧНЕ ОБҐРУНТУВАННЯ МАТЕМАТИЧНОЇ МОДЕЛІ БЕЗВІДМОВНОСТІ СИСТЕМИ ОБРОБКИ ІНФОРМАЦІЇ АВТОМАТИЗОВАНИХ РОБОТОТЕХНІЧНИХ ПЛАТФОРМ У СИСТЕМІ ЗАЛИШКОВИХ КЛАСІВ

Анотація. У статті розглянуто актуальну проблему забезпечення безвідмовності систем обробки інформації (СОІ) автоматизованих робототехнічних платформ (АРП) в умовах радіоелектронної боротьби (РЕБ). Обґрунтовано необхідність розробки ефективних методів захисту СОІ АРП від електромагнітних загроз. Проаналізовано існуючі підходи до підвищення надійності СОІ, зокрема використання резервування та сучасних інформаційних технологій. Зазначено, що традиційні методи захисту часто є недостатніми для протидії сучасним засобам РЕБ. У зв'язку з цим, запропоновано альтернативний підхід до створення математичного забезпечення СОІ АРП, спрямований на підвищення відмовостійкості, живучості та безвідмовності на основі використання модульності конструкції обчислювальних трактів СОІ та самокорегуючих властивостей непозиційних кодових структур системи залишкових класів (СЗК). Детально досліджено властивості СЗК, які роблять їх перспективними для застосування в СОІ АРП, зокрема можливість адаптації до вирішуваних завдань залежно від вимог до точності, швидкодії та надійності обчислень. Наголошено на можливості змінювати кількісне співвідношення між числом інформаційних та контрольних модулів у процесі вирішення задачі та гнучко використовувати резерви точності та надійності СОІ. Розроблено математичну модель безвідмовності СОІ АРП у СЗК, яка враховує особливості функціонування системи та вплив зовнішніх факторів. Описано структуру СОІ у СЗК, визначено основні фактори, що впливають на безвідмовність, та сформульовано математичні залежності між ними. Наведено алгоритм розрахунку ймовірності безвідмовної роботи СОІ. Результати дослідження, представлені в роботі, можуть бути використані для розробки та впровадження більш ефективних методів захисту АРП від РЕБ, що сприятиме підвищенню їх надійності та безпеки функціонування в складних умовах сучасного поля бою. Запропонована математична модель є ефективним інструментом для визначення надійної характеристики СОІ АРП, а саме ймовірності безвідмовного функціонування, та дозволяє оцінити вплив різних факторів на надійність СОІ та оптимізувати параметри системи для досягнення необхідного рівня захисту в умовах РЕБ.

Ключові слова: автоматизована робототехнічна платформа, перехідна ймовірність безвідмовного функціонування, показники надійності, радіоелектронна боротьба, система залишкових класів, система обробки інформації.

Вступ

В умовах сучасного поля бою, де радіоелектронна боротьба (РЕБ) набуває все більшого значення, забезпечення ефективного функціонування автоматизованих робототехнічних платформ (АРП) є критично важливим завданням. Сучасні засоби РЕБ здатні не тільки створювати потужні перешкоди, що унеможливають зв'язок та керування АРП, але й здійснювати інтелектуальні атаки, спрямовані на спотворення даних, підміну сигналів та навіть захоплення керування платформою. Враховуючи потенційні загрози порушення функціонування, втрати керування та навіть фізичного знищення АРП внаслідок впливу РЕБ, розробка та впровадження ефективних методів захисту є першочерговим завданням.

Життєздатність, мобільність та прихованість платформи в зоні виконання бойового завдання, а також її стійкість до засобів РЕБ противника, є ключовими факторами успіху. АРП повинні бути здатними не тільки протистояти безпосередньому впливу РЕБ, але й зберігати свою функціональність в умовах інформаційної ізоляції, коли зв'язок з оператором є обмеженим або відсутнім. Це вимагає розробки та інтеграції на борту АРП систем автономного керування, навігації та прийняття рішень, які базуються на захищених каналах зв'язку та інформації.

У зв'язку з цим, дослідження в області захисту АРП в умовах РЕБ, що узгоджуються зі Стратегією розвитку оборонно-промислового комплексу Украї-

ни, представляють значну наукову та практичну цінність. Стратегія розвитку ОПК України визначає пріоритетні напрями розвитку озброєння та військової техніки, серед яких забезпечення інформаційної безпеки та стійкості до впливу РЕБ займає одне з провідних місць. Дослідження в цій галузі спрямовані на створення інноваційних технологій та систем, здатних забезпечити надійний захист АРП в умовах сучасного електромагнітного поля бою, сприяючи зміцненню оборонного потенціалу країни.

Постановка проблеми. В умовах сучасних військових операцій, де домінуючим фактором є електромагнітне поле бою, забезпечення безперебійного функціонування та стійкості АРП до різноманітних електромагнітних загроз є першочерговим завданням. Ефективне керування наземними робототехнічними платформами часто ускладнюється нестабільністю зовнішніх умов, що негативно впливає на оперативність та надійність управління. Вирішення цієї проблеми потребує розробки та впровадження інноваційних підходів до керування, які базуються на надійній системі обробки інформації та забезпечують захист платформ в умовах РЕБ. Незважаючи на значний прогрес у створенні сучасних апаратно-програмних засобів для систем управління та зв'язку, низка питань у цій сфері залишається актуальною та потребує подальшого дослідження. Оскільки традиційні методи захисту часто є недостатніми для протидії сучасним засобам РЕБ. Розглянемо основні проблеми у сфері РЕБ:

1. Вразливість до ширококутних перешкод. Традиційні системи захисту часто розробляються для протидії вузькосмуговим перешкодам. Сучасні засоби РЕБ можуть створювати ширококутні перешкоди, які охоплюють широкий діапазон частот, що робить традиційні методи захисту неефективними.

2. Недостатня адаптивність. Традиційні системи захисту часто мають жорстку структуру та алгоритми, які не можуть адаптуватися до змін в умовах РЕБ. Сучасні засоби РЕБ можуть швидко змінювати свої параметри та методи впливу, що вимагає від систем захисту здатності до швидкої адаптації.

3. Залежність від частотного діапазону. Традиційні системи захисту можуть бути ефективними лише в певному частотному діапазоні. Сучасні засоби РЕБ можуть працювати в широкому діапазоні частот, що вимагає від систем захисту здатності працювати в різних частотних діапазонах.

4. Вразливість до інтелектуальних атак. Сучасні засоби РЕБ можуть використовувати інтелектуальні методи впливу, такі як імітація сигналів, створення помилкових цілей тощо. Традиційні системи захисту, які базуються на простих алгоритмах, можуть бути вразливими до таких атак.

5. Висока вартість та складність реалізації. Розробка та впровадження ефективних систем захисту від РЕБ є складним та дорогим завданням. Традиційні методи захисту можуть вимагати значних витрат на обладнання та обслуговування.

6. Обмежена ефективність проти комбінованих загроз. Сучасні засоби РЕБ можуть використовувати комбіновані методи впливу, такі як одночасне створення перешкод та використання кібернетичних атак. Традиційні системи захисту, які розроблені для протидії окремим видам загроз, можуть бути неефективними проти комбінованих загроз.

Для подолання цих проблем необхідні нові підходи до розробки систем захисту від РЕБ, які базуються на використанні сучасних технологій, таких як штучний інтелект, адаптивні алгоритми, ширококутні системи зв'язку, альтернативні математичні моделі обробки інформації та інші [1].

Критично важливим напрямом досліджень є розробка та вдосконалення надійних математичних моделей, спрямованих на мінімізацію впливу електромагнітних загроз на інформаційне середовище. Ефективним способом підвищення стійкості АРП до РЕБ є використання комбінації системи залишкових класів (СЗК) та завадостійкого кодування.

Метою роботи є обґрунтування можливості використання СЗК для підвищення рівня захисту систем обробки інформації АРП в умовах РЕБ.

Обґрунтування ефективності застосування СЗК для побудови ефективних СОІ

Складність та масштаби розв'язуваних завдань управління АРП потребує розширення функцій та можливостей засобів обчислювальної техніки, що тягне за собою збільшення обладнання АРП, ускладнення структури та математичного забезпечення систем обробки інформації (СОІ). Це, своєю чергою, викликає необхідність вжиття додаткових заходів

щодо забезпечення надійності функціонування СОІ й у першу чергу підвищення безвідмовності СОІ АРП, які мають забезпечити надійну роботу навіть за відмову її окремих елементів чи блоків, тобто. забезпечення високої живучості. Це подібно до того, як окремі ушкодження в біологічному організмі не порушують його нормальної життєдіяльності. Існує два основні шляхи підвищення надійності СОІ АРП: підвищення надійності окремих логічних елементів та запровадження різних типів надмірності (застосування різних видів резервування). Зважаючи на те, що надійність логічних елементів СОІ визначається рівнем розвитку технології, то очевидно, що введення надмірності є найбільш ефективним практичним шляхом підвищення надійності СОІ [2]. Різноманітність умов і жорсткість вимог (необхідність забезпечення високого ступеня точності, високої продуктивності функціонування СОІ АРП в реальному часі в умовах РЕБ), що накладаються на режим функціонування та експлуатації СОІ АРП, не завжди дозволяє застосовувати тимчасове та інформаційне резервування. Внаслідок цього одним із ефективних практичних методів підвищення надійності є структурне резервування на рівні дубльованої або тройованої мажоритарної структури [1].

Основний практичний шлях побудови надійних СОІ АРП, що функціонують у позиційних системах числення (ПСЧ), – створення перспективних відмовостійких архітектур, здатних до перебування за рахунок використання різних видів резервування (застосування БІС, НВІС, ПЛМ (PLD) та програмованих логічних інтегральних схем у різних функціональних блоках та вузлах СОІ) [3].

В умовах обмежень на масу, розміри та вартість бортових комп'ютерних систем та СОІ АРП, актуальним завданням є забезпечення їх надійності, живучості та стійкості до впливу електромагнітного спектра. Застосування додаткових розрядів для апаратного оперативного контролю дозволяє вирішити цю проблему, забезпечуючи гнучке керування точністю, швидкістю та надійністю обчислень. Проте, використання ПСЧ накладає певні обмеження на метод змінного масштабування, який використовується для скорочення кількості розрядів при поданні числової інформації в процесі контролю даних. Традиційні СОІ, що базуються на ПСЧ, мають низку недоліків. Зокрема, необхідність додаткових операцій зсуву перед кожним тактом виконання програми призводить до зниження швидкодії СОІ приблизно на 10%. Застосування змінного масштабування потребує значного обсягу попередніх теоретичних розрахунків для визначення оптимальних масштабних коефіцієнтів. Крім того, ефективність змінного масштабування обмежена певним класом завдань, що робить його малоприматним для СОІ, які працюють в реальному часі [4].

Отже, аналіз методів вирішення завдань АРП в умовах РЕБ показав, що на сучасному рівні розвитку технології застосування позиційних СОІ не може повністю задовольнити вимогам, що пред'являються до обчислювальних засобів реалізації систем контролю та управління даними. У зв'язку з цим виникає

проблема пошуку шляхів створення надійних, живучих та відмовостійких СОІ на основі застосування нетрадиційних систем числення.

Автори пропонують альтернативний підхід до створення математичного забезпечення СОІ АРП, спрямованими на підвищення відмовостійкості, живучості та безвідмовності на основі використання модульності конструкції обчислювальних трактів СОІ та самокорегуючі властивості непозиційних кодових структур СЗК [5], придатні для виявлення та виправлення помилок при обробці інформації в динаміці обчислювального процесу, що особливо важливо для організації відмовостійкого функціонування СОІ АРП в реальному часі [6].

Розглянемо важливу властивість системи залишкових класів (СЗК), що полягає у можливості змінювати кількісне співвідношення між числом інформаційних та контрольних (модулів) у процесі вирішення задачі та гнучко використовувати резерви точності та надійності СОІ [7]. Час, необхідний для виконання немодульної операції у СЗК, пропорційно числу інформаційних основ, тобто числу основ, що визначають точність обчислень. Переходячи до обчислень з меншою точністю, можна підвищити швидкість СОІ. Якщо деяка впорядкована СЗК розширюється шляхом додавання k основ, кожна з яких більша за основу вихідної СЗК, то мінімальна кодова відстань автоматично збільшується на величину k . Цього можна домогтися, зменшуючи кількість інформаційних основ, тобто переходячи до обчислень із меншою точністю. Отже, між коригуваними можливостями кодів СЗК і точністю обчислень існує обернено пропорційна залежність [5]. На одній і тій же СОІ можна виконувати одні й ті самі обчислення з високою точністю, але з меншим значенням надійності, інші – з меншою точністю, але з більш високим значенням надійності та швидкості.

Дана обставина вказує на можливість забезпечення необхідного рівня надійності та живучості СОІ у СЗК шляхом організації його стійкості до відмови від функціонування за рахунок застосування принципу адаптації, наприклад, на основі використання методу поступової деградації. Крім цього, апаратні засоби обробки інформації в СЗК відносяться до легко контрольованих та діагностованих. Це зумовлено специфічними особливостями подання непозиційних кодових структур у СЗК та характером прояву в них відмов та збоїв при обробці інформації. Ця особливість кодів у СЗК сприяє створенню адаптивних до відмов та збоїв архітектур СОІ, здатних до перебудови структури, що створює передумови до синтезу відмовостійких СОІ.

Таким чином, СОІУ в СЗК має властивість адаптації до розв'язуваних завдань залежно від вимог, що висуваються до точності, швидкодії та надійності обчислень. СЗК можна визначити не як систему числення, а як особливу конструкцію кодової числової структури даних, тобто спеціальним чином закодований блок числових даних.

Слід зауважити, що у пропонованому підході, СЗК не протиставляється двійковою ПСЧ, а служить як би її розширенням, що дозволяє ефективно вирі-

шувати певний клас задач. Тому, можливо, найбільш ефективним, у даному випадку, представляється підхід, що поєднує у собі комбіноване застосування СЗК та двійкової ПСЧ при побудові спеціалізованих обчислювальних комп'ютерних систем та компонентів та СОІ АРП. При цьому, наприклад, керування всією комп'ютерною системою може здійснюватися звичайними двійковими командами та блоками, а обробка даних виконується на основі модулярного представлення чисел. Таким чином, використання переваг СЗК, разом з традиційними двійковими методами побудови комп'ютерних систем та компонентів, може привести до підвищення продуктивності та надійності СОІ АРП у цілому.

Математична модель безвідмовності СОІ АРП у СЗК

Розглянемо, як можна застосувати основні математичні результати теорії марківських процесів для адекватного відображення процесу функціонування СОІ АРП у СЗК, що дозволить описати загальну модель СОІ, а потім синтезувати надійну модель у СЗК і оцінити її безвідмовність.

Розглянемо СОІ у СЗК, що складається з q інформаційних (ІОТ) та одного контрольного (КОТ) $k = 1$. Кожен ІОТ має кінцеве число різних функціональних станів. Позначатимемо кінцеве число різних функціональних станів ІОТ по модулю x_z як $\chi_z = \{0, y_z\}$. Якщо в даний момент часу t ІОТ знаходяться відповідно в станах s_1, s_2, \dots, s_q , тоді стан всіх ІОТ СОІ АРП у СЗК можна визначити вектором $S = (s_1, s_2, \dots, s_q)$. Безліч станів КОТ позначимо як $\chi_k = \{0, N_k\}$, де N_k – кількість контрольних основ СЗК. Процес функціонування СОІ АРП у СЗК можна описати за допомогою вектора станів:

$$\chi(t) = \{\chi_1(t), \chi_2(t), \dots, \chi_q(t), \tilde{\chi}(t)\}, \quad (1)$$

де $\chi_z(t)$ – стан ІОТ по модулю x_z в момент часу $t \geq 0$; $\tilde{\chi}(t)$ – стан ІОТ по модулю x_{q+1} , за умови, що розглядається СЗК лише з однією контрольною основою.

У випадку процес функціонування СОІ АРП у СЗК відбувається в такий спосіб, що у заданому інтервалі часу стан КОТ не змінюється та зберігає постійне значення $\tilde{\chi}(t) = const$.

Вектор $\chi(t)$ станів ІОТ зберігається незмінним протягом випадкового проміжку часу, що відповідає тривалості виконання елементарних арифметичних операцій t' (додавання, віднімання, множення), розподіленого за показовим законом із параметром λ_z інтенсивності відмови z – го функціонуючого обчислювального тракту (ОТ). Вектор станів $\tilde{\chi}(t)$ змінюється миттєво від значення $\tilde{\chi}(t)$ до значення $\tilde{\chi}_k(t)$. Еволюція КОТ (або резервного ОТ (РОТ)) в інтервалі часу одного циклу функціонування $[t_z, t_{z+1}]$ відбувається незалежно від процесу контролю та виправлення помилки (збою), здійснювано-

го КОТ. Отже, умовна ймовірність того, що в момент t_{z+1} вектор станів групи РОТ $\tilde{\chi}'(t_{z+1}-0)$ прийме значення v , при тривалості $(z+1)$ -го циклу функціонування, за умови, що $\tilde{\chi}(t_z) = k$, $\tilde{\chi}'(t_z) = z$, можна визначити за формулою:

$$\bar{P}_{zv}(t) = \int_0^t [P_{zv}(\theta)F]d\theta, \quad (2)$$

де $P_{zv}(\theta)$ – ймовірність того, що однорідний ланцюг Маркова перейде за час θ зі стану z у стан v .

У момент часу t_{z+1} відбувається миттєве зміна стану РОТ (КОТ) СОІ АРП у СЗК.

Розглянемо СОІ АРП у СЗК, в якій всі ОТ рівнонадійні, причому кількість ІОТ (робочих) дорівнюватиме q , два ОТ вважатимемо контрольними (КОТ), вони служать для контролю та діагностики СОІ АРП у СЗК, і r резервних (РОТ) ОТ, що грають роль резервних елементів. Для такої системи ($\lambda_z = \lambda$) сумарна робота, що виконується СОІ АРП у СЗК за час t , визначиться як:

$$W(t) = \int_0^t g[X(\theta)]d\theta, \quad (3)$$

де $g(z)$ – продуктивність СОІ АРП при одночасному працездатному стані z ІОТ.

Розглянемо набір з q ІОТ як один ІОТ по модулю

модулю $X = \prod_{z=1}^q x_z$, в момент часу t стан цього ІОТ $\chi(t)$,

а набір з r РОТ – як один РОТ по модулю

$X' = \prod_{z=q+1}^{q+2+r} x_z$, вважаючи, що в момент часу t даний

РОТ може перебувати в стані $\tilde{\chi}(t) = 0$ в залежності від станів ОТ. У цьому випадку перехідні ймовірності $P(t)$ станів СОІ у СЗК можна визначити за відомими математичними співвідношеннями [8].

Розглянемо модель СОІ АРП у СЗК, де одночасно може функціонувати $q+2$ із загальної кількості

$$\begin{aligned} P_{zv}(t) &= \int_0^t P_{(q+2)v}(t-\theta) dU_{z-q-2}[\lambda(q+2), \dots, \lambda(q+2), \theta] = \int_0^t \left\{ P_{(q+2)v}(t-\theta) \cdot e^{-\lambda(q+2)\theta} \cdot \frac{\lambda(q+2) \cdot [\lambda(q+2) \cdot \theta]^{z-q-3}}{(z-q-3)!} \right\} d\theta = \\ &= \int_0^t \left\{ P[(q+2), v, e^{-\lambda(t-\theta)}] \cdot e^{-\lambda(q+2)\theta} \cdot \frac{[\lambda(q+2) \cdot \theta]^{z-q-3}}{(z-q-3)!} \right\} d\theta = \int_0^t \left\{ C_{q+2}^v \cdot P^v (1-P)^{q+2-v} \cdot e^{-\lambda(q+2)\theta} \cdot \frac{[\lambda(q+2) \cdot \theta]^{z-q-3}}{(z-q-3)!} \right\} d\theta. \end{aligned} \quad (6)$$

Значимо, що отриманий вираз (6) є робочою формулою для визначення перехідних ймовірностей безвідмовного функціонування СОІ АРП у СЗК з q інформаційними основами та двома контрольними основами.

Приватні варіанти надійності моделей СОІ АРП у СЗК були розглянуті в рамках НДР.

Висновки

Проведене дослідження підтверджує, що для отримання та аналізу надійних характеристик СОІ АРП у СЗК доцільно використовувати відомі в тео-

$q+2+r$ ОТ, тобто активні $q+2$ ОТ, а r ОТ перебувають у резерві. У момент відмови ІОТ_z замість нього за допомогою комутатора (автомата надійності) підключається РОТ_z (зазначимо, що $x_z < x_{z+1}$). Таким чином, режим стійкості до відмови СОІ АРП у СЗК реалізується методом заміни. Якщо резерв РОТ вичерпано, тоді СОІ АРП у СЗК продовжує функціонувати в режимі живучості (режим поступової деградації).

У момент часу t для СОІ АРП у СЗК між числом $\chi(t)$ працездатних ОТ та числом $\beta(t)$ працюючих ОТ встановлюється таке співвідношення:

$$\beta(t) = \psi[\chi(t)], \quad (4)$$

де $\psi(z) = q+2$, при $q+2 \leq z \leq q+2+r$; $\psi(z) = z$, при $0 \leq z < q+2$.

Нехай набір РОТ розглядатимемо як один по модулю $X = \prod_{z=1}^q x_z$ ($\chi(t)$ – стан цього РОТ у момент часу t).

Якщо РОТ по модулю x_z знаходиться у стані $\chi_z(t)$ (тобто у працездатному стані), то за час θ РОТ по модулю x_z може перейти тільки в один із можливих станів $0 \leq v \leq z$ (відмовили $z-v$ РОТ). Якщо $0 \leq v \leq z \leq q+2$, тоді перехідну ймовірність $P_{zv}(t)$ можна визначити за відомими співвідношеннями [8].

Якщо $q+2 \leq v \leq z$, то перехідні ймовірності можна визначити за допомогою такого виразу:

$$\begin{aligned} P_{zv}(t) &= P[z-v, (q+2)\lambda t] = \\ &= e^{-(q+2)\lambda t} \cdot \frac{[(q+2)\lambda t]^{z-v}}{(z-v)!}. \end{aligned} \quad (5)$$

Якщо $0 \leq v < q+2 \leq z$, то у загальному вигляді можна визначити перехідні ймовірності $P_{zv}(t)$, використовуючи той факт, що перехід зі стану z в стан v можливий через стан $q+2$, тобто:

рії систем з частковим обслуговуванням та традиційні моделі функціонування СОІ у ПСЧ (наприклад, модель загибелі), а також математичні формули та співвідношення (1) – (6) з урахуванням властивостей та специфіки подання та обробки даних у СЗК.

Важливим аспектом є врахування стану, в якому знаходиться РОТ, надійність перемикаючих пристроїв, способу підключення обчислювальних трактів СОІ АРП, що функціонують у СЗК.

Наприклад, для аналізу безвідмовності СОІ можуть бути використані різні моделі резервування,

такі як модель ковзного резервування з навантаженим резервом без урахування впливу надійності пристроїв, що перемикають, або модель ковзного резервування з ненавантаженим резервом з урахуванням надійності перемикаючих пристроїв.

Запропонована математична модель (6) є ефективним інструментом для визначення надійної характеристики СОІ АРП, а саме ймовірності безвідмовного функціонування. Її використання дозволяє оці-

нити вплив різних факторів на надійність СОІ та оптимізувати параметри системи для досягнення необхідного рівня захисту в умовах РЕБ.

Результати дослідження, представлені в роботі, можуть бути використані для розробки та впровадження більш ефективних методів захисту АРП від РЕБ, що сприятиме підвищенню їх надійності та безпеки функціонування в складних умовах сучасного поля бою.

СПИСОК ЛІТЕРАТУРИ

1. Yanko, A., Pedchenko, N., Kruk, O. Enhancing the protection of automated ground robotic platforms in the conditions of radio electronic warfare. *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*, № 6, 2024, pp.136-142. <https://doi.org/10.33271/nvngu/2024-6/136>
2. Onyshchenko S., Yanko A., Hlushko A., Sabelnikova P. Assessment of information protection level against unauthorized access. *ScienceRise*, 2, Tallinn, Estonia, 2023, pp. 36–44. <http://doi.org/10.21303/2313-8416.2023.003211>
3. Янко А.С., Сабельнікова П.С. Метод виявлення та виправлення помилок на основі часових числових перерізів. Матеріали міжнародної науково-технічної конференції Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління, Баку-Харків-Жиліна, 2024, Т. 2, С. 14. <https://doi.org/10.32620/ICT.24.t2>
4. Янко А.С., Сабельнікова П.С. Дослідження методів контролю та корекції помилок інформації в комп'ютерних системах обробки даних, що функціонують в системі залишкових класів. Матеріали XXIV Всеукраїнської науково-технічної конференції молодих вчених, аспірантів та студентів «Стан, досягнення та перспективи інформаційних систем і технологій», Одеса, 18 – 19 квітня 2024, С. 143-144. <https://reposit.nupp.edu.ua/handle/PolNTU/15497>
5. Krasnobaev V., Kuznetsov A., Yanko A., Kuznetsova K. Correction Codes in the System of Residual Classes. *Proceedings of 2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T)*, Kyiv, Ukraine, 2019, pp. 488-492. <https://doi.org/10.1109/PICST47496.2019.9061253>
6. Янко А.С., Філь І.В. Забезпечення відмовостійкості комп'ютерних систем на основі коригувальних властивостей непозиційних кодових структур. Тези доповідей тринадцятої міжнародної наук.-техн. конф. «Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління», Баку-Харків-Жиліна, 26 – 27 квітня 2023, Т. 2, С. 51. <https://doi.org/10.32620/ICT.23.t2>
7. Краснобаев В.А., Сомов С. В., Янко. А. С. Основные свойства непозиционной системы счисления. *Системы управления, навигации та зв'язку*, 1 (25), 2013, С. 110–113.
8. Krasnobaev V., Yanko A., Fil I. Research of the possibility of fault-tolerant operation of a computer system in a non-positional number system in residual classes. *Materials of the XI International Scientific and Practical Conference "Information Control Systems & Technologies (ICST-Odessa-2023)" 21th – 23th September, 2023, Odessa.* pp. 175-177. <http://icst-conf.com/2023.pdf>

Received (Надійшла) 11.12.2024

Accepted for publication (Прийнята до друку) 26.02.2025

Theoretical substantiation of a mathematical model of failure-free for the information processing system of automated robotic platforms in the residue number system

A. Yanko, O. Kruk, S. Goncharenko

Abstract. This article addresses the critical issue of ensuring the failure-free operation of information processing systems (IPS) in automated robotic platforms (ARPs) operating in conditions of electronic warfare (EW). The increasing role of ARPs in the modern world and its vulnerability to EW influences, which can lead to loss of control, data corruption, and even physical destruction of the platform, are emphasized. The necessity of developing effective methods to protect ARP IPS from electromagnetic threats is substantiated. Existing approaches to increasing the reliability of IPS, including the use of redundancy and modern information technologies, are analyzed. It is noted that traditional protection methods are often insufficient to counter modern EW means. In this regard, an alternative approach to creating mathematical support for ARP IPS is proposed, aimed at increasing fault tolerance, survivability, and reliability based on the modular design of IPS computing paths and the self-correcting properties of non-positional code structures of the residue number system (RNS). The properties of the RNS that make them promising for use in ARP IPS are studied in detail, in particular, the ability to adapt to the tasks being solved depending on the requirements for accuracy, speed, and reliability of calculations. The possibility of changing the quantitative relationship between the number of information and control modules in the process of solving the problem and flexibly using the reserves of accuracy and reliability of the IPS is emphasized. A mathematical model of the reliability of ARP IPS in the RNS is developed, which takes into account the features of the system's functioning and the influence of external factors. The structure of the IPS in the RNS is described, the main factors affecting reliability are identified, and mathematical relationships between them are formulated. An algorithm for calculating the probability of failure-free operation of the IPS is presented. The research results presented in the paper can be used to develop and implement more effective methods of protecting ARPs from EW, which will contribute to increasing its reliability and safety of operation in the difficult conditions of the modern battlefield. The proposed mathematical model is an effective tool for determining the reliable characteristics of ARP IPS, namely the probability of failure-free operation, and allows estimating the influence of various factors on the reliability of the IPS and optimizing the system parameters to achieve the required level of protection in EW conditions.

Keywords: automated robotic platform, transition probability of failure-free operation, reliability indicators, electronic warfare, residue number system, information processing system.

Vladyslav Huts, Oleksii Gorokhovatskyi

Simon Kuznets Kharkiv National University of Economics, Kharkiv, Ukraine

EFFICIENT FAULT DETECTION IN INDUSTRIAL EQUIPMENT USING PCA AND SMOTE ENHANCED NEURAL NETWORKS

Abstract. This research addresses the challenge of fault detection in industrial equipment using high-dimensional vibration data with limited labeled examples. The goal was to develop a neural network model capable of accurately classifying measurement vectors into normal and faulty categories. The dataset consisted of 1158 samples, each with 93,752 numerical features, representing two classes: 865 normal and 293 faulty instances. A comprehensive preprocessing pipeline was employed, including standardization, dimensionality reduction using Principal Component Analysis (PCA), and Synthetic Minority Over-sampling Technique (SMOTE) for class balancing. The developed neural network achieved a baseline accuracy of 94.40% with 100 PCA components. Further experiments demonstrated that reducing the architecture and using only 50 PCA components improved accuracy to 98.81%, highlighting the effectiveness of the proposed approach. These findings emphasize the utility of combining PCA, SMOTE, and neural networks for fault detection in industrial equipment in high-dimensional, imbalanced datasets. Future research directions include exploring advanced neural network architectures, investigating the impact of PCA component count on model performance, and studying the feasibility of training effective models on synthetic data.

Keywords: fault detection; neural networks; PCA; SMOTE; dimensionality reduction.

Introduction

Ensuring the reliability of mechanical systems is a critical task in modern industrial environments. Equipment failures can cause severe disruptions, leading to costly repairs, operational delays, and, in some cases, serious safety risks. As a result, early detection of faults becomes an essential component of preventive maintenance strategies. Machine learning (ML) became a key tool in automating the fault detection process, enabling fast and accurate identification of potential issues. However, many traditional ML models require large labeled datasets, which are often difficult to obtain in real-world scenarios.

The object of study in this paper is the process of building the neural network (NN) model to classify the probable faults of equipment through the analysis of vibration data. Building and training of the model requires time and computational resources as well as the construction of the effective architecture of the model that could solve the problem. Additionally, this process is data-dependent and should utilize methods to address the unequal distribution of classes and data scarcity.

The goal of the work is to develop a robust fault detection model that can accurately identify both nominal and faulty instances in the equipment's operational data.

Related work

Fault detection in complex systems is a critical area of research, particularly in industries where early identification of equipment malfunction is essential for preventing costly breakdowns and ensuring operational safety. Traditional machine learning techniques are commonly applied in this field, yet they often face significant challenges, such as handling high-dimensional data and imbalanced class distributions. The dataset used in this study contains 1158 files, each

containing over 93,000 features—exemplifies a such scenario in industrial monitoring, where large amounts of sensor data are collected for fault detection.

High-dimensional data, such as the dataset representatives in this research, present several computational challenges, including overfitting and increased computational costs. Moreover, when working with neural networks, the risk of the “curse of dimensionality” arises, which can affect model performance by diluting the significance of individual features. To address this, principal component analysis (PCA) is commonly used as a dimensionality reduction technique. PCA transforms the high-dimensional data into a lower-dimensional space by extracting the most important features, or principal components, that explain the largest amount of variance in the dataset. PCA has been widely validated in the literature as an effective way to reduce computational complexity while retaining the most relevant information, enabling more efficient training of machine learning models, including neural networks [1 – 3].

The unequal class distribution is a common issue in fault detection problems, as there are significantly fewer instances of faulty data (class 1) compared to normal operation data (class 0). This distribution can skew model predictions, making it more likely to favor the majority class (class 0). The synthetic minority over-sampling technique (SMOTE) [2] was employed to address this. SMOTE generates synthetic samples from the minority class by interpolating between existing examples, effectively balancing the dataset. By providing a balanced dataset, SMOTE helps ensure that the model is able to detect faults as effectively as it recognizes normal operating conditions. Previous studies have demonstrated the usefulness of SMOTE in addressing class imbalance, especially in industrial fault detection scenarios where faulty conditions are rare [4, 5].

The implementation of neural network model used in this study was built using the Keras framework [6, 7]. The model architecture consists of three fully connected layers, each followed by Leaky ReLU activation functions. Leaky ReLU was chosen over the standard ReLU to prevent the issue of “dying neurons,” where neurons stop updating during training. Dropout layers were applied after each dense layer to prevent overfitting by randomly deactivating a portion of neurons during training. Dropout is a well-known regularization technique used to improve model generalization by preventing overfitting to the training data [8, 9].

The final layer of the network uses a sigmoid activation function, which is suitable for binary classification problems, distinguishing between nominal (non-faulty) and faulty data. Adam optimizer [10] was used which in an adaptive learning rate optimization algorithm that has been shown to achieve better performance on deep learning tasks due to its ability to adjust learning rates for each parameter dynamically. Binary cross-entropy was used as the loss function, which is standard for binary classification tasks. Additionally, early stopping was employed during training to prevent overfitting.

While other techniques such as prototype selection and construction methods have been applied to reduce dataset size and improve model performance, they are often computationally expensive and risk discarding important information. Prototype selection methods focus on identifying the most relevant instances from the dataset, whereas prototype construction generates new instances to represent the data more effectively. However, these approaches often face limitations in terms of processing time and accuracy, especially in high-dimensional datasets like the one used in this research. Instead, the combination of PCA and SMOTE provides a more efficient solution for managing the large dataset while retaining critical information necessary for fault detection [11, 12].

Prototype selection methods focus on identifying a representative subset of data points from the entire dataset. These methods aim to choose the most relevant instances avoiding redundancy and noise. However, in fault detection tasks, where there is a significant difference in the number of samples between classes (with fewer faulty samples compared to normal ones), traditional prototype selection techniques often face challenges. While these methods can help reduce dimensionality, they might not effectively address the issue of unequal class distribution. In this paper, the SMOTE was applied to mitigate this by generating synthetic samples, ensuring that the fault detection neural network model receives a balanced set of examples for training.

Conversely, prototype construction methods [12] focus on generating new samples that represent the underlying structure of the data. These methods employ clustering techniques or neural networks to create synthetic prototypes. While prototype construction can address some issues with the unequal distribution of classes, it may also introduce complexity and noise. PCA was utilized in this paper to perform dimensionality

reduction, effectively capturing the main variance of the dataset while eliminating redundancy. This preprocessing step enhanced memory usage (and, probably, model efficiency) by reducing the dimensionality of the vector from over 93,000 dimensions to 100.

Both prototype selection and construction methods come with trade-offs, especially in high-dimensional datasets like this. A significant drawback of these methods is the high computational cost involved in processing involved in large datasets processing the potential introduction of noise through synthetic sample generation. For our case, the leveraging PCA helped alleviate these challenges by preserving the dataset's core features, while SMOTE effectively handled class imbalance without the need for complex prototype construction methods.

The model building pipeline

In this paper, the traditional prototype selection and construction techniques were not directly implemented, as SMOTE and PCA provided a more efficient solution for high-dimensional fault detection datasets with unequal class distribution. This combined approach significantly improved the model's ability to generalize, effectively addressing both dimensionality and the challenge of having fewer faulty data samples (class 1) compared to normal data (class 0).

Data Preprocessing. To prepare the data for analysis, the first step involved the standardizing dataset. This was accomplished using the StandardScaler from the scikit-learn library, which transforms the data so that each feature has a mean of zero and a standard deviation of one. The standardization formula can be expressed as (1):

$$x' = (x - \mu) / \sigma, \quad (1)$$

where x' is the standardized value of feature x after normalization, μ – mean of the feature values, σ – standard deviation of the feature values.

Dimensionality Reduction. Given the high dimensionality of the vibration data, PCA was applied to reduce the number of features. It helps to retain the most significant variance in the data while minimizing the number of dimensions. The transformation can be described by the following equation (2):

$$Z = XW, \quad (2)$$

where Z is matrix representation of data after dimensionality reduction, X is the original feature matrix, W – weight matrix for the transformation in dimensionality reduction.

Data Balancing. Data Balancing. SMOTE generates synthetic examples for the minority class to create a more balanced dataset. The new synthetic instances are created by interpolating between existing instances of the minority class (3):

$$x_{new} = x_i + \lambda(x_j - x_i), \quad (3)$$

where x_{new} is new synthetic instance generated by SMOTE, x_i is the existing instance from the minority class, x_j is the other instance from the minority class for interpolation, and λ is the interpolation factor for generating synthetic instances.

Neural Network Architecture. Three hidden layers utilize the LeakyReLU activation function, which allows small gradients for inactive neurons. The LeakyReLU activation can be expressed as (4):

$$\text{LeakyReLU}(x) = \begin{cases} x & \text{if } x > 0; \\ ax & \text{if } x \leq 0, \end{cases} \quad (4)$$

where a is the small constant slope, typically set to 0.1.

Dropout layers are applied after each hidden layer to mitigate overfitting by randomly setting 25% of the neurons to zero during training. The output layer consists of a single neuron with a sigmoid activation function for binary classification (5):

$$y = \frac{1}{1+e^{-z}}, \quad (5)$$

where z is the input to the sigmoid function, and y is the predicted output of the model after applying the sigmoid function.

The model was trained using the Adam optimizer and binary cross-entropy as the loss function (6):

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^N \left(y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right), \quad (6)$$

where L is the loss function, calculated using binary cross-entropy, N is the total number of instances in the dataset, y_i is the actual class label for the i -th instance, and \hat{y}_i is the predicted probability for the i -th instance, representing the model's confidence that the instance belongs to the positive class.

Early stopping was implemented to monitor the validation loss and prevent overfitting.

The performance of the model was rigorously evaluated on the test dataset, focusing on metrics such as accuracy and loss. The training process was monitored across a specified number of epochs with a batch size of 32, ensuring efficient use of computational resources while tracking the total training time.

Data preprocessing

In this paper, various methods were employed to develop an effective fault detection model. The primary focus was on the use of vibration data collected from mechanical systems.

To ensure uniformity across the dataset, the data was standardized using the StandardScaler from the sklearn.preprocessing module. This normalization is crucial for optimizing the performance of the machine learning model. Given the large number of features, PCA was applied to reduce the dimensionality of the dataset. The number of principal components was set to 100, significantly reducing computational complexity while retaining essential variance in the data.

The dataset was then split into training and testing sets using train_test_split from sklearn.model_selection, with 80% allocated for training and 20% for testing. To address class imbalance—where faulty samples were significantly fewer than nominal ones—the SMOTE was utilized to generate synthetic samples for the minority class. This step was essential to ensure the model trained effectively on a balanced dataset.

The architecture of the neural network includes three layers. The first one contained 512 units, the second

layer had 256 units, and the third layer included 128 units. Each layer employed the LeakyReLU activation function to prevent inactive neurons, with a dropout rate of 0.25 applied to each dense layer to mitigate overfitting. The final layer utilized a sigmoid activation function for binary classification, distinguishing between nominal and faulty data.

The model was compiled using the Adam optimizer with a learning rate of 0.0003, and binary cross-entropy was used as the loss function. Early stopping procedure was implemented to prevent overfitting, halting training if the validation loss did not improve after two consecutive epochs. The model's performance was evaluated on the test set, yielding accuracy and loss metrics to assess its effectiveness in fault detection.

Artificial neural network modeling

Experiments were conducted on a computer with an Intel Core i5-10500H processor and 16 GB of RAM. The Python 3.8 programming environment in Jupyter was utilized, with Keras 2.4 and scikit-learn 0.24 versions, providing the necessary performance for deep learning tasks.

The dataset consists of 1158 files with vibration measurements, each containing high-dimensional information with 93752 features that needed to be processed and analyzed. There are 865 vectors representing class 0 and 293 for class 1. Each file was loaded into Python using the numpy and glob libraries, resulting in a two-dimensional array where each row represented a sample and each column corresponded to a feature. This problem was introduced by Eric Bechhoefer in the scope of Ukrainian hackathon of scientific works of young scientists in the field of intellectual information technologies [13]. The initial dataset is available in [14].

The neural network was trained over 50 epochs, with early stopping implemented to prevent overfitting. The training process recorded a significant increase in accuracy, starting from 73.48% and reaching a maximum accuracy of 100% by the 11th epoch. The loss during training decreased rapidly, beginning at 7.51 and converging close to zero. The model's performance was evaluated on the test dataset, achieving an accuracy of 94.40% and a loss of 0.2953.

The loss and accuracy curves indicate that the training and validation losses steadily decreased, while the training accuracy significantly increased, approaching 1.0. To assess the impact of the number of PCA components on model performance, tests were conducted with different values of n_components (200, 300, 400). The results showed that as the number of components increased, the model's accuracy on the test data decreased. Specifically, with n_components = 200, the model achieved an accuracy of 90.95%, while for n_components = 300, accuracy dropped to 89.66%. Further increasing the number of components to 400 resulted in a test accuracy of 84.91%. At the same time, the training accuracy remained high across all variations. Based on these results, it was decided to retain 100 components as the optimal number, as it provided the best balance between test accuracy (94.40%) and computational complexity.

Results

The conducted experiments explored the impact of different neural network configurations on fault detection accuracy. The primary goal was to determine how changes in architecture, activation functions, and the number of PCA components affect the model's accuracy and loss values. The results of experiments are shown in Table 1.

The baseline model consisted of three layers three layers with 512, 256, and 128 neurons, used the LeakyReLU activation function and 100 PCA components. It achieved a test accuracy of 94.40% with a loss value of 0.3125. This configuration served as a reference point for comparing the results of subsequent experiments. Replacing the LeakyReLU activation with regular ReLU neurons keeping the other parameters unchanged resulted in a slight improvement in test accuracy, reaching 94.83%, although the loss value increased to 0.3662. This suggests that while ReLU can improve performance, it may also introduce greater fluctuations during training.

A configuration with fewer neurons (256, 128, and 64 neurons across three layers) led to the highest test accuracy of 95.69%. The loss value for this experiment was 0.3201, suggesting that although the model was effective, it might have been overly specialized to the training data.

Increasing the number of neurons and adding a fourth layer (1024, 512, 256, 128 neurons) resulted in a slight improvement in test accuracy to 94.83%, with a lower loss value of 0.2952. This configuration helped the model generalize slightly better, as indicated by the reduced loss, though the overall accuracy did not significantly increase. Training accuracy remained at 98.63%, matching the baseline. Reducing the network to

two layers (512, 256 neurons) allowed us to reach 98.81% test accuracy with a loss value of 0.3392. This shows that a simple architecture can still perform well, potentially benefiting from the reduced complexity.

Reducing the number of PCA components to 50 showed significant improvement. This model achieved a test accuracy of 98.71% with a very low loss of 0.1225, which suggests that reducing the dimensionality helped to filter out noise and preserve critical patterns.

On the other hand, increasing the number of PCA components to 150 led to a drop in test accuracy to 92.24%, with the loss value rising to 0.5240. This suggests that adding more components introduced unnecessary complexity, making it harder for the model to learn effectively.

Finally, increasing the PCA components further to 200 reduced test accuracy to 91.81% with a loss of 0.5021. The training accuracy was 98.63%, consistent with the baseline. However, the increase in complexity from the additional components did not improve performance and, in fact, resulted in poorer generalization. Probably, it is required to add more neurons/layers in the neural network to achieve better results for 150 or 200 PCA components.

Overall, these experiments show that finding the optimal configuration of neural network architecture and PCA components is critical for achieving high performance. In particular, the reduction to 50 PCA components proved to be highly effective, significantly improving the model's ability to generalize, as evidenced by the low test loss and high accuracy. Meanwhile, the simpler architecture with two layers also demonstrated that reduced complexity can lead to better performance, highlighting that a more complex model is not always the better option. The baseline architecture of the neural network used in this study is shown in Fig. 1.

Table 1 – Model Performance Across Different Configurations

Exp.	Parameters (Number of Neurons, Layers)	Activation function	Number of PCA Components	Test Accuracy (%)
1	512, 256, 128 (3 layers)	LeakyReLU	100	94.40
2	512, 256, 128 (3 layers)	ReLU	100	94.83
3	256, 128, 64 (3 layers)	LeakyReLU	100	95.69
4	1024, 512, 256, 128 (4 layers)	LeakyReLU	100	94.83
5	512, 256 (2 layers)	LeakyReLU	100	98.81
6	512, 256, 128 (3 layers)	LeakyReLU	50	98.71
7	512, 256, 128 (3 layers)	LeakyReLU	150	92.24
8	512, 256, 128 (3 layers)	LeakyReLU	200	91.81



Fig. 1. Architecture of the neural network used for fault detection

It consists of an input layer that processes 100 principal components, followed by three dense layers

with 512, 256, and 128 neurons respectively, each using LeakyReLU activation and a dropout rate of 0.25. The

final output layer utilizes a sigmoid activation for binary classification.

The training process of this network, as shown in Fig. 2 while Fig. 3, demonstrates a smooth convergence of both training and validation loss, indicating effective learning without overfitting. Additionally, the accuracy plot reflects that the model achieves high accuracy on both training and validation data, confirming the robustness of the neural network.

Discussion

The results from the experiments highlight several interesting insights into optimizing neural network models for fault detection, particularly when dealing with high-dimensional data and imbalanced datasets. By systematically varying the network architecture, activation functions, and the number of PCA components, we were able to identify configurations that improved the model's performance, as well as those that led to less effective results. Below, we discuss the implications of these findings.

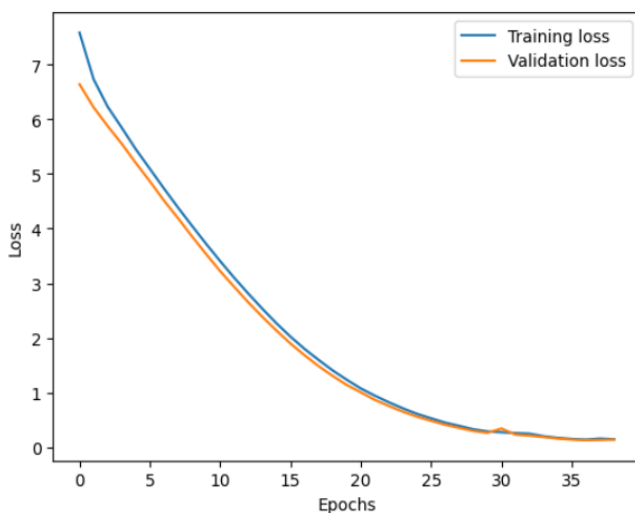


Fig. 2. Training and validation loss over epochs

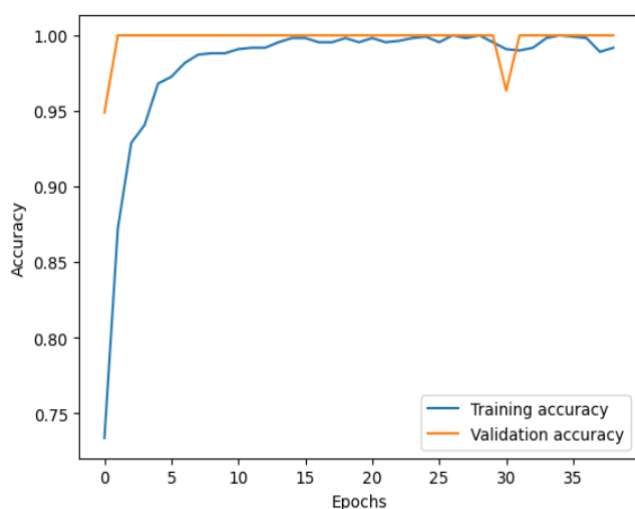


Fig. 3. Training and validation accuracy over epochs

One of the most significant observations was the effect of dimensionality reduction through PCA.

Reducing the number of components to 50 significantly improved the model's performance, leading to a test accuracy of 98.71%. This suggests that removing excess dimensions helped the model focus on the most critical features, effectively filtering out noise and reducing overfitting.

This configuration not only the accuracy but also streamlined the learning process, allowing the model to generalize better. Conversely, increasing the number of PCA components to 150 and 200 degraded the performance, with test accuracies dropping to 92.24% and 91.81%, respectively.

These results underscore the importance of finding the right balance in dimensionality; too many components can lead to unnecessary complexity and noise, hindering the model's ability to generalize. This also shows that the high-dimensional vector (with more than 93 000 dimensions) could be effectively reduced for this problem.

The experiments with different activation functions showed that replacing LeakyReLU with ReLU led to a slight increase in test accuracy to 94.83% but also increased the loss to 0.3662, suggesting a potential trade-off between stability and performance. While ReLU may boost performance under certain conditions, it can introduce more fluctuations during training, which may lead to less consistent learning outcomes. On the other hand, LeakyReLU proved to be more stable across various configurations, as reflected in the baseline model's performance.

The analysis of neural network architecture also provided valuable insights. Reducing the model to two layers with 512 and 256 neurons produced a high test accuracy of 98.81% with a loss of 0.3392. This finding suggests that simpler architectures can still achieve strong performance, particularly when the data are well-preprocessed.

Simpler models often have fewer parameters, which can help prevent overfitting and make them more efficient.

In contrast, adding complexity by increasing the network to four layers with 1024, 512, 256, and 128 neurons did not yield a significant improvement, maintaining a test accuracy of 94.83% but slightly reducing the loss to 0.2952.

This indicates that additional layers and neurons do not always lead to better generalization, and may instead complicate the learning process without adding real value.

Furthermore, the experiment that reduced the number of neurons in each layer to 256, 128, and 64 achieved high test accuracy of 95.69%.

Overall, the experiments demonstrate that simplicity and balance are the keys to optimizing model performance. Effective dimensionality reduction, as seen with 50 PCA components, can streamline the learning process and reduce the risk of overfitting by focusing the model on the most relevant features. Similarly, the success of the two-layer architecture suggests that complex, deep networks are not always necessary for high performance, especially when the data are processed effectively.

These findings highlight the importance of data preprocessing and careful architecture selection in building models that are not only accurate but also efficient and robust.

Conclusions

The paper addresses the problem of detecting equipment faults through approach that uses both machine learning techniques and advanced data processing strategies.

The contribution of this research includes the entire machine learning pipeline of using data preprocessing

methods like PCA, SMOTE to address the classification problem presented with imbalanced dataset.

The practical significance of this research provides the machine learning model effective automatic tool for predictive maintenance of the equipment.

The prospects for further research may include but are not limited with the search for more effective architecture of the neural network, deeper research of the dependency of accuracy on the quantity of elements in PCA analysis, finally, it would be interesting to investigate the possibility to train the effective model only using synthetic data.

REFERENCES

1. B. Krawczyk, "Learning from imbalanced data: Open challenges and future directions," *Progress in Artificial Intelligence*, vol. 5, no. 4, Apr. 2016. doi:10.1007/s13748-016-0094-0.
2. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002. doi: 10.1613/jair.953.
3. Y. Zhang, and S. Wang, "Deep learning for time series forecasting: A survey," *Big Data*, vol. 9, no. 1, Dec. 2020. doi: 10.1089/big.2020.0159.
4. X. Chen, D. Liu, and H. Zha, "Predictive data mining techniques for fault diagnosis of electric equipment: A review," *Applied Sciences*, vol. 10, no. 3, p. 950, Feb. 2020. doi: 10.3390/app10030950.
5. B. S. Panigrahi, T. T. M. Tamilselvi, S. B. G. Tilak Babu, P. G and B. Shaik, "Deep Learning Techniques for Fault Detection in Industrial Machinery," 2024 5th International Conference on Recent Trends in Computer Science and Technology (ICRTCST), Jamshedpur, India, 2024, pp. 221-226, doi: 10.1109/ICRTCST61793.2024.10578499.
6. Y. Zhang, et al., "Deep learning-based intelligent fault diagnosis methods toward rotating machinery," *IEEE Access*, vol. 8, pp. 9335–9346, 2019. doi: 10.1109/ACCESS.2019.2963092.
7. O. Matania, E. Bechhoefer, and J. Bortman, "Digital Twin of a Gear Root Crack Prognosis," *Sensors*, vol. 23, no. 24, p. 9883, Dec. 2023. doi: 10.3390/s23249883.
8. O. Matania, R. Cohen, E. Bechhoefer, and J. Bortman, "Anomaly Detection and Remaining Useful Life Estimation for the Health and Usage Monitoring Systems 2023 Data Challenge," *PHM Society Proceedings*, vol. 8, no. 1, Jun. 2024. doi: 10.36001/phme.2024.v8i1.4125.
9. O. Matania, E. Bechhoefer, D. Blunt, W. Wang, and J. Bortman, "Anomaly Detection and Remaining Useful Life Estimation for the Health and Usage Monitoring Systems 2023 Data Challenge," *Sensors*, vol. 24, no. 13, p. 4258, Jun. 2024. doi: 10.3390/s24134258.
10. O. Matania, L. Bachar, E. Bechhoefer, and J. Bortman, "Signal Processing for the Condition-Based Maintenance of Rotating Machines via Vibration Analysis: A Tutorial," *Sensors*, vol. 24, no. 2, p. 454, Jan. 2024. doi: 10.3390/s24020454.
11. T. Chistiakova, J. Zambrano, and B. Carlsson, "Application of machine learning methods for fault detection in wastewater treatment plants," in *Reglermöte*, 2014. doi: 10.13140/2.1.3733.4403.
12. S. Windmann, "Data-Driven Fault Detection in Industrial Batch Processes Based on a Stochastic Hybrid Process Model," in *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 4, pp. 3888-3902, Oct. 2022, doi: 10.1109/TASE.2021.3138925.
13. Ukrainian hackathon of scientific works of young scientists in the field of intellectual information technologies URL: <https://zp.edu.ua/vkiit>
14. The dataset. URL: https://pz.zp.ua/files/vkiit/zeroShot_vkiit.rar

Received (Надійшла) 30.11.2024

Accepted for publication (Прийнята до друку) 05.02.2025

Ефективне виявлення відмов у промисловому обладнанні з використанням PCA та покращених нейронних мереж на основі SMOTE

В. В. Гуць, О. В. Гороховатський

Анотація. У цьому дослідженні розглядається проблема виявлення відмов у промисловому обладнанні за допомогою високовимірних даних вібрації з обмеженою кількістю мічених прикладів. Метою було розробити модель нейронної мережі, здатну точно класифікувати вектори вимірювань на нормальні та несправні категорії. Набір даних складався з 1158 зразків, кожен із яких містив 93,752 числові ознаки, що представляли два класи: 865 нормальних та 293 несправних випадки. Було використано комплексний конвеєр попередньої обробки, який включав стандартизацію, зменшення розмірності за допомогою методу головних компонент (PCA) і техніку синтетичного збільшення меншості (SMOTE) для балансування класів. Розроблена нейронна мережа досягла базової точності 94,40% із 100 компонентами PCA. Подальші експерименти показали, що зменшення розмірності архітектури та використання лише 50 компонентів PCA підвищило точність до 98,81%, підкреслюючи ефективність запропонованого підходу. Ці результати акцентують увагу на корисності комбінування PCA, SMOTE та нейронних мереж для виявлення відмов у високовимірних незбалансованих наборах даних. Майбутні напрями досліджень включають вивчення передових архітектур нейронних мереж, аналіз впливу кількості компонентів PCA на продуктивність моделі та дослідження можливості навчання ефективних моделей на синтетичних даних.

Ключові слова: виявлення відмов; нейронні мережі; PCA; SMOTE; зменшення розмірності.

О. Ю. Заковоротний, О. А. Сапальський

Національний технічний університет "Харківський політехнічний інститут", Харків, Україна

ВИРІШЕННЯ ПРОБЛЕМ ПРОДУКТИВНОСТІ ЗА ДОПОМОГОЮ АСИНХРОННИХ МЕТОДІВ В JAVASCRIPT

Анотація. Сучасні веб-додатки часто стикаються з викликами продуктивності, пов'язаними з блокуванням інтерфейсу користувача та повільною обробкою даних. Оскільки JavaScript є однопоточною мовою програмування, браузер не здатен одночасно виконувати обчислення та оновлювати інтерфейс. Це може спричинити "лаги" — ситуації, коли додаток зависає, і користувач не може взаємодіяти з інтерфейсом. Асинхронне програмування в JavaScript надає інструменти для вирішення цих проблем, що дозволяє підвищити відгукливість додатків і ефективність виконання завдань. **Мета** цієї роботи полягає у дослідженні можливостей підвищення продуктивності веб-додатків за допомогою асинхронного програмування в JavaScript. Це включає аналіз ефективності обробки завдань та відгукливості інтерфейсу користувача, а також оцінку доцільності використання асинхронних підходів для вирішення проблем блокування інтерфейсу та запобігання "лагам", що часто виникають у сучасних однопоточних середовищах. **Отримані наступні результати:** використання асинхронного програмування в JavaScript значно підвищило продуктивність веб-додатків. Відгукливість інтерфейсу покращилася за рахунок зменшення кількості випадків блокування UI під час обробки важких завдань. У місцях, де раніше спостерігалися відчутні лаги, додаток тепер працює безперервно, забезпечуючи плавну взаємодію з користувачем. **Висновки.** Було доведено доцільність використання асинхронного програмування в JavaScript для підвищення продуктивності веб-додатків. Це дозволяє значно покращити відгукливість інтерфейсу та уникнути блокування під час виконання складних операцій, забезпечуючи плавний досвід взаємодії для користувача. Визначено перспективи подальших досліджень у напрямку оптимізації асинхронних процесів для різних типів завдань та інтеграції з багатопотоковими рішеннями.

Ключові слова: асинхронне програмування, JavaScript, продуктивність веб-додатків.

Вступ

В останні роки веб-додатки стали невід'ємною частиною нашого повсякденного життя, забезпечуючи швидкий доступ до інформації та інструментів для вирішення різних завдань. Однак зі зростанням складності сучасних додатків з'являються нові виклики, пов'язані з їх продуктивністю. Однією з основних проблем є блокування інтерфейсу користувача під час виконання ресурсоємних операцій, що значно погіршує користувацький досвід. Це особливо помітно в однопоточному середовищі, яким є JavaScript, коли браузер не може одночасно виконувати обчислення і оновлювати інтерфейс. Такі затримки в роботі програми можуть викликати "лаги" — моменти, коли додаток тимчасово не реагує на дії користувача.

Асинхронне програмування пропонує ефективні рішення для вирішення цієї проблеми. Воно дозволяє розділити завдання на менші, не блокуючи основний потік, що сприяє збереженню відгукливості інтерфейсу та загальному підвищенню продуктивності веб-додатків. Незважаючи на очевидні переваги, використання асинхронних підходів вимагає уважного планування, оскільки помилки в їх реалізації можуть призвести до непередбачуваних результатів.

Ця робота присвячена дослідженню можливостей асинхронного програмування в JavaScript для підвищення продуктивності веб-додатків і покращення взаємодії користувача з інтерфейсом.

Постановка проблеми. JavaScript, як однопоточна мова програмування, має суттєві обмеження, що впливають на продуктивність веб-додатків. Однією з основних проблем є те, що мова не може виконувати дві задачі одночасно. Це означає, що під час виконання ресурсоємних обчислень браузер не здатен оновлювати інтерфейс.

В результаті, інтерфейс працює плавно лише при виконанні задач тривалістю до ~50 мс [1] в іншому випадку виникають переривання, а при дуже тривалих операціях можливі зависання.

Варто пам'ятати, що 0,1 секунди це максимальний час, протягом якого користувачі відчувають, що вони безпосередньо взаємодіють з об'єктами в інтерфейсі. 1 секунда: — максимальний час, протягом якого користувачі можуть вільно переміщатися в командному просторі, не відчуваючи непомірного очікування.

Затримка від 0,2 до 1,0 секунди означає, що користувачі помічають цю затримку і, отже, відчувають, що комп'ютер "працює" над командою, а не що команда є безпосереднім наслідком їх дій. Велика затримка додає суттєвого ризику, що користувач може втратити увагу і, як наслідок, інтерес [2].

Таким чином, існує нагальна потреба у вирішенні цих проблем, щоб забезпечити безперервну і плавну взаємодію з веб-додатками, що є критично важливим для підтримки інтересу користувачів та забезпечення високої якості користувацького досвіду.

У зв'язку з цим, **метою дослідження** є вивчення того, як асинхронні методи в JavaScript можуть покращити продуктивність веб-додатків на практиці, та визначити підходи що дозволяють зменшити блокування інтерфейсу та підвищити відгукливість, забезпечуючи тим самим кращий користувацький досвід.

Основна частина роботи

У цьому розділі розглянемо, як асинхронне програмування може покращити продуктивність веб-додатку на прикладі системи розумного будинку. Зокрема, ми будемо розглядати проблемами, які виникають при використанні синхронного XMLHttpRequest для отримання даних.

Припустимо, у нас є система умного дому, яка періодично отримує дані з датчиків, таких як температура, вологість та стан системи безпеки.

Використовуючи синхронний XMLHttpRequest, ми можемо реалізувати запит таким чином, наведено на рис. 1.

```
1  const request = new XMLHttpRequest();
2
3  // синхронний запит
4  request.open("GET", "https://api.smart-home.com/temperature", false);
5  request.send();
6
7  if (request.status === 200) {
8    // Обробка отриманих даних
9    processTemperatureData(request.responseText);
10 }
11
```

Рис. 1. Синхронний XHR запит

Використання синхронного запиту означає, що браузер буде заблокований на час виконання запиту, що призводить до заморожування інтерфейсу. Це викликає "лаги" у користувацькому інтерфейсі, і користувач не зможе взаємодіяти з додатком, що може негативно вплинути на його досвід. Таким чином, користувач може відчувати розчарування та відмовитися від використання програми.

Натомість, асинхронні запити є кращим підходом, оскільки вони дозволяють браузеру продовжувати реагувати на дії користувача. Таким чином, не-

залежно від тривалості запиту інтерфейс залишатиметься інтерактивним. У однопоточному Javascript це досягається за рахунок того, що всі запити обробляються не самим двигуном мови, а безпосередньо браузером, що додатково дозволяє робити кілька запитів одночасно. Інструменти на кшталт "xhr" або "fetch" лише дають інструкцію браузеру запланувати запити та колбеки, які мають бути виконані після їх завершення. У потрібний момент браузер відправити ці колбеки в Event loop, де вони будуть оброблені та виконані в порядку черги після звільнення основного стека виклик (рис. 2).

```
1  const request = new XMLHttpRequest();
2  // true - асинхронний запит
3  request.open("GET", "https://api.smart-home.com/temperature", true);
4  // Обробка отриманих даних
5  request.onload = () => {
6    if (request.status === 200) {
7      processTemperatureData(request.responseText);
8    }
9  };
10 // Обробка помилки
11 request.onerror = (error) => handleError(error);
12
13 request.send();
```

Рис. 2. Асинхронний XHR запит

Ідеальним підходом для обмеження асинхронної логіки буде використання синтаксису async/await.

Підхід не збільшить продуктивність, але зробить код більш читаним і організованим (рис. 3).

```
1  async function getEnvironmentData() {
2    try {
3      const temperature = await fetch('https://api.smart-home.com/temperature');
4      const humidity = await fetch('https://api.smart-home.com/humidity');
5      // Обробка отриманих даних
6      processTemperatureData(temperature);
7      processHumidityData(humidity);
8    } catch (error) {
9      // Обробка помилки
10     handleError(error);
11   }
12 }
```

Рис. 3. Запит з використанням синтаксису async/await

Таким чином покращується робота програми. Поки запит, користувач може виконувати у додатку інші завдання, а браузер може відобразити тимчасовий компонент заглушки. Це значно покращує користувацький досвід. Інша поширена причина проблем з продуктивністю – це довгі завдання. Як вже було сказано раніше, завдання повинні бути не більше 50 мс, щоб не викликати проблем з продуктивністю. В іншому випадку користувач зможе помітити лаги. Розв'язання цього завдання виглядає очевидним: необхідно розбити одну велику задачу на безліч менших (рис. 4).



Рис. 4. Розбиття задачі

Але проблема полягає в тому, що безліч синхронних завдань, як і раніше, блокуватимуть браузер, а саме процес реакції на користувацькі події та малювання інтерфейсу [3]. Цю проблему можна вирішити за допомогою функції відкладеного виклику. Якщо дати браузеру час або вікно для виконання завдань взаємодії з інтерфейсом, то можна виключити появу лагів і зависань. Найпростіший випадок це використання `setTimeout`. Цей метод дозволяє запланувати виклик функції не раніше вказаного часу. У контексті розробки програми "розумного дому" його можна використовувати, щоб запланувати перевірку трекінгових систем раз на п'ять хвилин. Це дозволяє різко скоротити кількість викликів (рис. 5).

Іноді немає можливості вказати часовий проміжок для початку виконання функції. У цьому випадку може статись в нагоді `requestAnimationFrame`.

`requestAnimationFrame` вказує браузеру на те, що ви хочете зробити анімацію, і просить його запланувати перемальовку на наступному кадрі анімації [4]. Як параметр метод отримує функцію, яка буде викликана перед перемальовуванням.

```

1  async function checkEnvironment() {
2      // Отримання даних про навколишнє середовище
3      await getEnvironmentData();
4      // Виклик функції через 5 хвилин
5      setTimeout(checkEnvironment, 300000);
6  }
7
8  // Запуск перевірки
9  checkEnvironment();

```

Рис. 5. Виклик через `setTimeout`

Він синхронізується з частотою оновлення екрана (зазвичай 60 кадрів на секунду). Замість запуску анімації з фіксованим інтервалом (як при використанні `setInterval`), браузер сам вирішує, коли краще виконати наступний кадр, пропускаючи непотрібні оновлення та тим самим знижуючи навантаження на процесор. Так, наприклад, додаток smart home може використовувати його для малювання графіків.

Подібний метод, `requestIdleCallback`, використовується для виконання фонових завдань, які не критичні для продуктивності і можуть бути відкладені доти, доки браузер не стане "вільним" від інших пріоритетних операцій [5]. Так реальна програма може використовувати цей метод для роботи з трекінгом активності користувача та збору статистики.

Бувають випадки, коли необхідно виконати якісь складні обчислення, які не можна розділити на менші частини. У такому разі може допомогти інший підхід - використання `WebWorkers` (рис. 6).

```

1  // main.js
2  const worker = new Worker("worker.js");
3  // Обробка середньої температури
4  worker.onmessage = ({ data }) => handleAverageTemperature(data);
5  // Повідомлення для Worker
6  worker.postMessage("Обчислити середню температуру");
7  //=====
8  // worker.js
9  onmessage = ({ data }) => {
10     // Складні обчислення
11     const averageTemperature = calculateAverageTemperature(data);
12     // Відправка результату назад
13     postMessage(averageTemperature);
14 };
15
16 // Обчислення для знаходження середньої температури
17 function calculateAverageTemperature() {
18     for (let i = 0; i < 1_000_000; i++) {
19         // Складна логіка
20     }
21
22     // Повертаємо обчислене значення
23     return someCalculatedValue;
24 }

```

Рис. 6. Код `WebWorker`

Web Workers — це API для запуску JavaScript у фоновому потоці, окремо від основного потоку, який відповідає за взаємодію з DOM та обробку подій користувача.

Це дозволяє виконувати важкі обчислення або тривалі завдання без блокування інтерфейсу користувача.

Однак, Web Workers не мають доступу до DOM безпосередньо.

Це означає, що будь-які результати обчислень повинні передаватися назад до основного потоку через механізм обміну повідомленнями (`postMessage()`), що може створити невелику затримку для інтерактивних застосунків.

Так, наприклад, додаток може розраховувати різні показання датчиків без негативного впливу на продуктивність сторінки.

Висновки та перспективи подальших досліджень

На основі проведеного аналізу та використання асинхронних методів програмування в JavaScript для додатку "розумного дому" вдалося досягти значного підвищення продуктивності та покращення взаємодії з користувачем.

Перехід від синхронних запитів до асинхронних дозволив уникнути блокування інтерфейсу, що робить роботу програми плавною і безперервною. Використання `async/await` значно спрощує читабельність коду та підтримку складних запитів.

Методи відкладеного виконання, такі як

- `setTimeout``,
- `queueMicrotask``,
- `requestAnimationFrame``
- `requestIdleCallback``,

дозволяють ефективніше розподіляти навантаження між критичними і неважливими задачами, підвищуючи загальну продуктивність.

Також застосування Web Workers для фонових обчислень виводить важкі обчислення з головного потоку, що дозволяє уникнути "лагів" навіть при великій кількості даних.

Дослідження показує, що асинхронні техніки є критичними для створення відгукливих та продуктивних вебзастосунків.

Перспективи подальших досліджень можуть включати глибшу оптимізацію алгоритмів обробки даних та інтеграцію більш складних фонових завдань у Web Workers для ще більшої стабільності та ефективності роботи застосунків..

СПИСОК ЛІТЕРАТУРИ

1. Addy Osmani. A comprehensive iot attacks survey based on a building-blocked reference mode. URL: <https://web.dev/articles/long-tasks-devtools?hl=en> (дата звернення 20.10.2024).
2. Miller R. B. Response time in man-computer conversational transactions //Proceedings of the December 9-11, 1968, fall joint computer conference, part I. – 1968. – С. 271-277.
3. MDN. Window: `requestAnimationFrame()` method. URL: <https://developer.mozilla.org/en-US/docs/Web/API/Window/requestAnimationFrame> (дата звернення 20.10.2024).
1. 4 MDN. Window: `requestIdleCallback()` method. URL: <https://developer.mozilla.org/en-US/docs/Web/API/Window/requestIdleCallback> (дата звернення 20.10.2024).
4. Jeremy Wagner. Optimize long tasks. URL: <https://web.dev/articles/optimize-long-tasks?hl=en> (дата звернення 20.10.2024).

Received (Надійшла) 30.11.2024

Accepted for publication (Прийнята до друку) 05.02.2025

Solving productivity problems using asynchronous methods in JavaScript

Oleksandr Zakovorotnyi, Oleskandr Sapalskyi

Abstract. The **subject of study in the article** are the performance issues in modern web applications caused by the single-threaded nature of JavaScript. Specifically, it investigates how blocking operations, like synchronous XMLHttpRequests (XHR), can negatively impact the responsiveness of user interfaces. The study explores how asynchronous programming can improve the user experience by preventing interface lag and optimizing task handling. **The task** is to explore and demonstrate how various asynchronous techniques, including callbacks, promises, `async/await`, and methods such as `setTimeout`, `queueMicrotask`, `requestAnimationFrame`, and `requestIdleCallback`, can be employed to address performance bottlenecks in a web-based smart home application. Additionally, the use of Web Workers to offload intensive calculations to background threads is examined. **The following results were obtained:** The study shows that transitioning from synchronous XHR to asynchronous methods significantly reduces the occurrence of interface freezing. Callbacks and promises provide basic improvements in responsiveness, while `async/await` makes the code more readable and easier to manage. Methods like `setTimeout` and `requestAnimationFrame` allow deferring non-critical tasks, improving performance during active user interaction. Web Workers were effectively used to handle complex computations, preventing the main thread from being blocked and allowing the application to remain responsive even during intensive processing. **Conclusions:** Asynchronous programming techniques are essential for optimizing the performance of JavaScript-based web applications. By adopting these methods, developers can prevent interface blocking and enhance the overall user experience. The use of Web Workers for background processing further improves performance, making these approaches highly applicable to complex applications like smart home systems. Future research could focus on refining these techniques for specific use cases and exploring additional strategies for performance optimization.

Keywords: asynchronous programming, JavaScript, web application performance.

М. О. Зачепило

Національний технічний університет «Харківський політехнічний інститут», Харків, Україна

БАГАТОРІВНЕВІ ЕКОСИСТЕМИ ДЛЯ АДАПТИВНИХ АГЕНТІВ: ВІД МОДЕЛЮВАННЯ ДО ПРАКТИЧНИХ ЗАСТОСУВАНЬ

Анотація. У статті розглядається розробка багаторівневих екосистем, які забезпечують навчання автономних систем у складних і змінних умовах. Основна мета дослідження полягає у створенні середовищ із різними рівнями складності, що дозволяє автономним агентам адаптуватися до динамічних сценаріїв, долати перешкоди та ефективно використовувати ресурси. Такий підхід спрямований на вдосконалення методів навчання, які можуть бути застосовані для підготовки систем штучного інтелекту до роботи в реальних умовах. У роботі акцентується увага на моделюванні багаторівневих екосистем, що включають статичні та динамічні перешкоди, конкурентні взаємодії між агентами та обмежені ресурси. Для цього запропоновано використання алгоритмів навчання з підкріпленням, які дозволяють агентам оптимізувати свої стратегії поведінки в умовах постійно змінюваного середовища. Розроблені моделі сприяють кращому розумінню того, як агенти можуть адаптуватися до складних умов та які фактори впливають на ефективність їхньої поведінки. Особливу увагу приділено аналізу можливостей практичного застосування таких підходів у робототехніці. Зокрема, моделі багаторівневих екосистем можуть бути використані для навчання автономних роботів, які працюють у складних середовищах, наприклад, під час рятувальних операцій, дослідження незвіданих територій або виконання завдань у міських умовах. Ці моделі дозволяють створювати більш адаптивні та надійні автономні системи, здатні ефективно реагувати на змінні фактори зовнішнього середовища. Результати дослідження показують, що багаторівневі екосистеми є ефективним інструментом для підготовки автономних систем до роботи в реальних умовах. Запропоновані підходи не лише сприяють підвищенню адаптивності та ефективності агентів, але й відкривають нові можливості для їхнього застосування в різних галузях, включаючи промисловість, логістику та наукові дослідження. Використання багаторівневих середовищ забезпечує автономним системам перевагу в складних і непередбачуваних умовах, що є актуальним завданням сучасної науки та технологій.

Ключові слова: адаптивний агент, багаторівнева екосистема, навчання з підкріпленням.

Вступ

Актуальність дослідження. У сучасному світі автономні системи та штучний інтелект відіграють ключову роль у розв'язанні багатьох завдань, зокрема у сфері робототехніки, транспортних систем і автоматизації. Для забезпечення високої ефективності та адаптивності таких систем важливо розробляти методи навчання, які дозволяють агентам швидко пристосовуватися до змінюваних і багаторівневих умов середовища. Багаторівневі екосистеми, які моделюють різні рівні складності, є перспективним підходом для створення середовищ навчання, що стимулюють розвиток складних стратегій поведінки.

Актуальність дослідження зумовлена необхідністю вирішення низки викликів, які стоять перед сучасними автономними системами.

Зокрема, такі системи мають демонструвати здатність адаптуватися до нових умов, долати перешкоди та ефективно використовувати ресурси в обмежених середовищах. Імітація реальних сценаріїв у багаторівневих екосистемах дозволяє створити умови для розвитку таких здібностей, що особливо важливо для застосувань у промисловості, логістиці та наукових дослідженнях.

Запропоновані багаторівневі екосистеми дозволяють включати різні рівні складності, такі як статичні та рухомі перешкоди, конкурентні агенти та змінні ресурси. Завдяки цьому створюються умови, що стимулюють розвиток складних поведінкових стратегій у агентів. Це, у свою чергу, сприяє підвищенню ефективності систем навчання з підкріпленням, особливо у задачах, що потребують довгострокової адаптації.

Дослідження, представлене у цій роботі, зосереджене на розробці моделі навчання в умовах складних екосистем і аналізі її потенційного застосування в робототехніці. Результати моделювання демонструють, що багаторівневі екосистеми є ефективним інструментом для тренування автономних систем, які повинні працювати у складних і змінних середовищах.

Таким чином, актуальність роботи полягає у внеску в розвиток методів створення адаптивних систем, які здатні не лише вирішувати поточні задачі, але й навчатися нових стратегій у динамічних умовах. Це відкриває перспективи для впровадження новітніх технологій у галузі штучного інтелекту та автоматизації.

Мета та задачі дослідження. Метою цього дослідження є розробка багаторівневих екосистем, які створюють умови різної складності для навчання автономних систем. Такі середовища дають змогу імітувати реальні сценарії, де системи мають адаптуватися до динамічних умов, вирішувати складні завдання та ефективно використовувати ресурси. Завдяки цьому стає можливим тренування автономних агентів, здатних не лише реагувати на зовнішні подразники, але й навчатися нових стратегій поведінки, які є важливими для їхнього довгострокового функціонування.

Одним із ключових завдань є створення моделі навчання, яка враховує складність екосистем. У цій моделі агенти мають працювати в середовищах із різними рівнями складності: від простих статичних умов до динамічних середовищ із рухомими перешкодами, конкурентами та обмеженими ресурсами. Це дозволяє агентам покращувати свої адаптивні

здібності, використовуючи механізми навчання з підкріпленням, еволюційні алгоритми чи інші інтелектуальні підходи. Вивчення того, як агенти пристосовуються до змінних умов і які стратегії вони використовують для досягнення цілей, є важливим аспектом цього завдання.

Іншим важливим напрямком дослідження є аналіз можливостей застосування розроблених моделей у робототехніці. Навчання автономних роботів у середовищах із різними рівнями складності дозволяє створити системи, які зможуть працювати у реальних сценаріях, таких як навігація в міському середовищі, участь у рятувальних операціях або дослідження важкодоступних територій. Завдяки таким багаторівневим середовищам навчання роботи зможуть адаптуватися до змінних умов, приймати ефективні рішення та взаємодіяти з іншими системами чи людьми.

Запропоновані підходи мають не лише теоретичну цінність, але й практичне значення. Вони дозволяють моделювати різні ситуації, які можуть виникнути у реальному світі, і створювати навчальні алгоритми, що враховують обмеження, ризики та ресурси. Це відкриває нові горизонти для розробки автономних систем, здатних функціонувати у складних та непередбачуваних умовах.

Таким чином, результати цього дослідження сприяють розвитку інноваційних підходів до навчання автономних систем. Розробка багаторівневих екосистем забезпечує основи для створення роботів і систем штучного інтелекту, які є адаптивними, надійними та здатними працювати у реальних динамічних середовищах.

Огляд літератури та існуючих рішень. Для моделювання динаміки популяції агентів у багаторівневих екосистемах та аналізу впливу зовнішніх змін середовища на поведінку агентів у [1, 2] використовують математичний апарат теорії диференціальних рівнянь, а саме рівняння Лотки-Вольтерри для опису взаємодій між різними агентами або рівнями екосистем.

У цьому напрямку теорію марковських процесів [3, 4] застосовують для моделювання переходів між станами агентів (наприклад, виживання, навчання, відтворення) та для оцінки ймовірностей успішної адаптації агентів у середовищах різної складності.

Алгоритми нейроеволюції (NEAT - NeuroEvolution of Augmenting Topologies) [5, 6] застосовують еволюції топологій нейронних мереж агентів у багаторівневих середовищах та для адаптацій до змінних умов середовища через мутацію та схрещування.

Для аналізу стратегій взаємодії агентів у конкурентному середовищі та визначення рівноваги між кооперативними та конкурентними стратегіями використовують теорію ігор [7, 8].

Математичний апарат фрактального [9, 10] аналізу може бути застосовано для моделювання складності середовища через фрактальні структури та створення багаторівневих перешкод або ресурсів із фрактальними характеристиками. У свою чергу мультиагентні системи [11, 12] з використанням графів знайшли використання при моделюванні відносин

між агентами через графові структури та вивчення динаміки взаємодій на різних рівнях екосистем.

Інформаційну ентропію (Shannon Entropy) [13, 14] застосовують для оцінки складності та різноманітності станів системи та для аналізу адаптації агентів у багаторівневих умовах. Оптимізаційні моделі на основі рою (PSO) [15, 16] використовуються для оптимізації рішень у складних екосистемах та вивчення групових стратегій агентів.

Аналіз стійкості через власні числа матриць Якобі [17, 18] дозволяють оцінити стійкість системи при зміні параметрів. У свою чергу для навчання агентів ефективних стратегій у багаторівневих середовищах використовують Алгоритми навчання з підкріпленням (Reinforcement Learning) [19, 20], які розраховують баланс між експлуатацією та дослідженням. Також цей математичний апарат дозволить створити комплексну модель для дослідження багаторівневих екосистем та їх застосувань у робототехніці та навчанні автономних систем.

Результати досліджень

1. Формалізація моделі навчання в багаторівневих екосистемах. Модель базується на багаторівневій екосистемі, в якій агенти повинні адаптуватися до різних рівнів складності (перешкоди, ресурси, конкуренція). Навчання реалізується через алгоритм навчання з підкріпленням (Q-Learning), модифікований для умов багаторівневої адаптації.

Введемо основну формули для Q-Learning:

$$Q(S_t, A_t) \leftarrow Q(s, a) + \alpha \times [R(s, a) + \gamma \max_a Q(s', a) - Q(s, a)], \quad (1)$$

де s – поточний стан агента, a – виконана дія, $R(s, a)$ – винагорода за виконання дії a в стані s , s' – наступний стан після дії a , α – швидкість навчання, γ – коефіцієнт дисконтування (важливість майбутніх винагород).

Для реалізації багаторівневих екосистем модель екосистеми визначимо вхідними параметрами. Рівні складності задамо наступним чином (табл. 1).

Таблиця 1 – Рівні складності системи

L_1	Просте середовище (ресурси без перешкод)
L_2	Ресурси з фіксованими перешкодами
L_3	Динамічне середовище (рухомі перешкоди + конкурентні агенти)

Винагорода $R(s, a)$ розраховується за таким співвідношенням:

$$R(s, a) = \begin{cases} +10, & \text{агент успішно досягає ресурсу;} \\ -5, & \text{агент натрапляє на перешкоду;} \\ -10, & \text{якщо енергія агента вичерпана.} \end{cases} \quad (2)$$

На рівні L_3 перешкоди рухаються за правилом, яке задає таке співвідношення:

$$x_t = x_{t-1} + v \cdot \Delta t, \quad (3)$$

де v — швидкість перешкоди, Δt — час кроку.

2. Практична реалізація моделі. Задамо початкові умови системи. Агент стартує в початковому стані $s_0 = (x_0, y_0)$ в середовищі L_2 . Метою агента визначимо досягнути ресурсу, уникаючи перешкод. Причому параметрами моделі присвоюємо такі значення:

$$\begin{aligned}\alpha &= 0.1, \\ \gamma &= 0.9, \\ R_{\text{ресурс}} &= +10, \\ R_{\text{перешкода}} &= -5.\end{aligned}$$

На кроці 1 агент виконує дію a_1 (рух вперед). Тоді поточний стан системи стане $s_0 = (2,3)$, а наступний визначимо $s_1 = (3,3)$.

Очевидно, якщо агент зіткнувся з перешкодою, то $R(s_0, a_1) = -5$.

Виконуємо оновлення $Q(s_0, a_1)$ за основною формулою для Q-Learning (1):

$$\begin{aligned}Q(s_0, a_1) &= \\ &= 0 + 0.1 \cdot (-5 + 0.9 \cdot \max_{a'} Q(s_1, a') - 0).\end{aligned}\quad (4)$$

Припустимо тепер, що

$$\max_{a'} Q(s_1, a') = 2,$$

тоді

$$Q(s_0, a_1) = 0 + 0.1 \cdot (-5 + 0.9 \cdot 2) = -0.1.\quad (5)$$

На кроці 2 агент виконує дію a_2 (обхід перешкоди). Тоді поточний стан системи стане $s_1 = (3,3)$, а наступний визначимо $s_2 = (4,4)$.

Очевидно, якщо агент дістався ресурсу, то $R(s_1, a_2) = +10$.

Виконуємо оновлення $Q(s_1, a_2)$ за основною формулою для Q-Learning (1):

$$\begin{aligned}Q(s_1, a_2) &= \\ &= 0 + 0.1 \cdot (10 + 0.9 \cdot \max_{a'} Q(s_2, a') - 0).\end{aligned}\quad (6)$$

Припустимо тепер, що

$$\max_{a'} Q(s_2, a') = 0,$$

тоді

$$Q(s_1, a_2) = 0 + 0.1 \cdot (10 + 0) = 1.0.\quad (7)$$

У результаті моделювання агент навчається обирати дії, які максимізують його винагороду, уникаючи перешкод і досягаючи ресурсу.

У складних рівнях середовища (L_3) агенти демонструють адаптивну поведінку завдяки динамічним оновленням Q-значень.

У результаті моделювання агент навчається обирати дії, які максимізують його винагороду, уникаючи перешкод і досягаючи ресурсу.

У складних рівнях середовища (L_3) агенти демонструють адаптивну поведінку завдяки динамічним оновленням Q-значень.

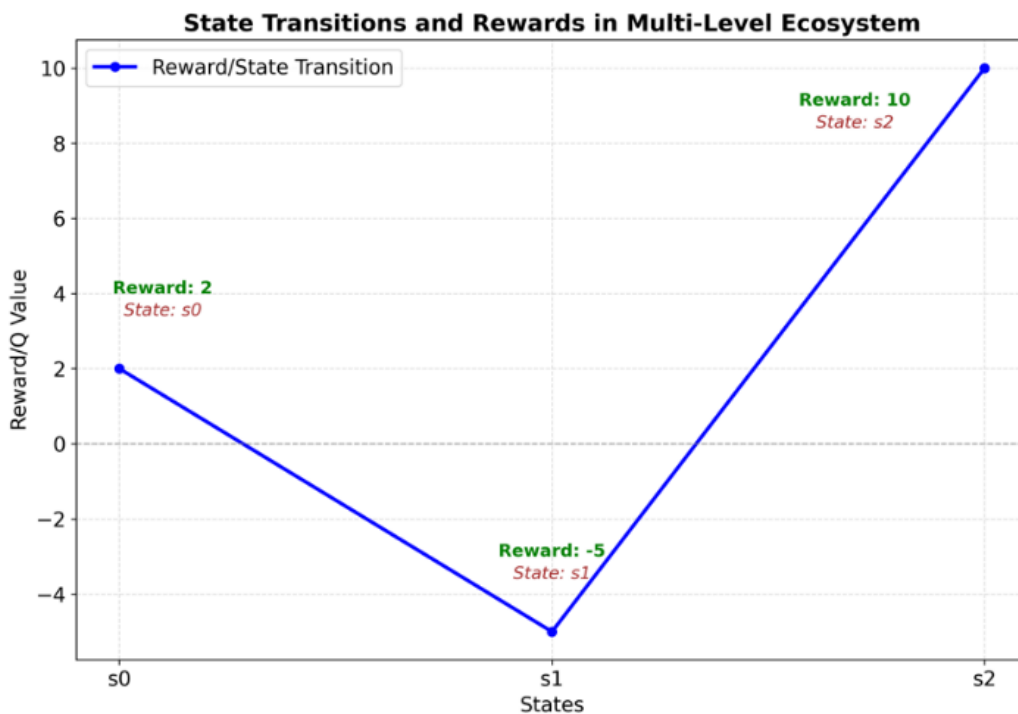


Рис. 1. Переходи між станами агентів у багаторівневій екосистемі з зазначенням винагород на кожному етапі

Висновки

Багаторівневі екосистеми демонструють високий потенціал для моделювання адаптивної поведінки агентів у складних умовах.

Використання таких систем дозволяє розробляти методики навчання, що імітують реальні сценарії, де агенти мають взаємодіяти з різними рівнями складності середовища. Результати дослідження показують, що динамічні багаторівневі середовища

стимулюють розвиток складних стратегій і сприяють ефективнішій адаптації автономних систем.

Запропонована модель, заснована на алгоритмах навчання з підкріпленням, дозволила продемонструвати, як агенти здатні навчатися ухилятися від перешкод і досягати ресурсів, отримуючи максимальні винагороди. Зміна складності середовища в реальному часі створює умови для формування поведінки, яка не лише орієнтована на виживання, але й сприяє довгостроковому покращенню адаптивних стратегій.

Це відкриває перспективи для використання таких моделей у робототехніці, зокрема для автономних транспортних систем і роботів-дослідників.

Практична значущість дослідження полягає у впровадженні багаторівневих екосистем у навчальні процеси автономних систем, що дозволяє підвищити їх стійкість та ефективність у динамічних умовах. Такий підхід може бути корисним для створення систем штучного інтелекту, здатних адаптуватися до складних і непередбачуваних середовищ, що актуально для промислових, наукових та соціальних завдань.

СПИСОК ЛІТЕРАТУРИ

1. Lawrence Perko. *Differential Equations and Dynamical Systems*: Springer, 2021. – 553 p.
2. David Betounes. *Differential Equations: Theory and Applications*: Springer, 2020. – 634 p.
3. Paul A. Gagniu. *Markov Chains: From Theory to Implementation and Experimentation*: Wiley, 2021. – 256 p.
4. Pierre Brémaud. *Discrete Probability Models and Methods*: Springer, 2020. – 556 p.
5. Sebastian Risi, Julian Togelius. *Neuro evolution: From Algorithms to Applications*: Springer, 2020. – 150 p.
6. Kenneth O. Stanley, Joel Lehman. *Why Greatness Cannot Be Planned: The Myth of the Objective*: Springer, 2020. – 141 p.
7. Kevin Leyton-Brown, Yoav Shoham. *Essentials of Game Theory: A Concise, Multidisciplinary Introduction*: Morgan & Claypool Publishers, 2020. – 88 p.
8. Martin J. Osborne. *An Introduction to Game Theory*: Oxford University Press, 2021. – 533 p.
9. Michael Frame. *Fractal Worlds: Grown, Built, and Imagined*: Yale University Press, 2021. – 312 p.
10. Christoph Bandt, Kenneth Falconer, Martina Zähle. *Fractal Geometry and Stochastics VI*: Birkhäuser, 2021. – 340 p.
11. Maria Gini, Toru Ishida, Cristiano Castelfranchi, W. Lewis Johnson. *Massively Multi-Agent Systems I*: Springer, 2020. – 300 p.
12. Adelinde M. Uhrmacher, Danny Weyns. *Multi-Agent Systems: Simulation and Applications*: CRC Press, 2020. – 566 p.
13. Myron Tribus. *Thermostatistics and Thermodynamics*: Springer, 2021. – 698 p.
14. Arieh Ben-Naim. *Entropy and the Second Law: Interpretation and Misinterpretations*: World Scientific, 2020. – 250 p.
15. Maurice Clerc. *Particle Swarm Optimization*: Wiley-ISTE, 2020. – 243 p.
16. Daniel Bratton, James Kennedy. *Computational Intelligence: The Honey Bee Swarm Approach*: Springer, 2021. – 150 p.
17. David G. Luenberger, Yinyu Ye. *Linear and Nonlinear Programming*: Springer, 2021. – 546 p.
18. Stephen Boyd, Lieven Vandenberghe. *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares*: Cambridge University Press, 2020. – 477 p.
19. Richard S. Sutton, Andrew G. Barto. *Reinforcement Learning: An Introduction*: MIT Press, 2020. – 552 p.
20. Marco Wiering, Martijn van Otterlo. *Reinforcement Learning: State-of-the-Art*: Springer, 2020. – 638 p.

Received (Надійшла) 25.11.2024

Accepted for publication (Прийнята до друку) 29.01.2025

Multilevel ecosystems for adaptive agents: from modelling to practical applications

M. O. Zachepylo

Abstract. The article considers the development of multilevel ecosystems that provide training for autonomous systems in complex and changing conditions. The main goal of the research is to create environments with different levels of complexity, which allows autonomous agents to adapt to dynamic scenarios, overcome obstacles and effectively use resources. This approach is aimed at improving training methods that can be used to prepare artificial intelligence systems for work in real conditions. The work focuses on modelling multilevel ecosystems that include static and dynamic obstacles, competitive interactions between agents and limited resources. For this purpose, the use of reinforcement learning algorithms is proposed, which allow agents to optimize their behavioural strategies in a constantly changing environment. The developed models contribute to a better understanding of how agents can adapt to complex conditions and what factors affect the effectiveness of their behaviour. Special attention is paid to the analysis of the possibilities of practical application of such approaches in robotics. In particular, multi-level ecosystem models can be used to train autonomous robots operating in complex environments, for example, during rescue operations, exploration of unexplored territories or performing tasks in urban conditions. These models allow creating more adaptive and reliable autonomous systems that are able to effectively respond to changing environmental factors. The results of the study show that multi-level ecosystems are an effective tool for preparing autonomous systems for operation in real conditions. The proposed approaches not only contribute to increasing the adaptability and efficiency of agents, but also open up new opportunities for their application in various industries, including industry, logistics and scientific research. The use of multi-level environments provides autonomous systems with an advantage in complex and unpredictable conditions, which is an urgent task of modern science and technology.

Keywords: adaptive agent, multi-level ecosystem, reinforcement learning.

Г. С. Іващенко, А. І. Кононенко, Д. О. Тимошенко

Харківський національний університет радіоелектроніки, Харків, Україна

МЕТАЕВРИСТИЧНІ МЕТОДИ ВИРІШЕННЯ ЗАДАЧІ ГІЛЬЙОТИННОГО РОЗКРОЮ

Анотація. Актуальність. Оптимізація розкрою листових матеріалів є важливою задачею в промисловості, що дозволяє мінімізувати відходи та зменшити витрати у виробництві. У цій роботі розглядаються методи вирішення задачі гільйотинного розкрою, включаючи рівневі алгоритми (зокрема, Best Fit Decreasing), генетичні алгоритми та алгоритм мурашиної колонії. **Метою даної роботи** є дослідження ефективності евристичних та метаевристичних підходів до задачі розкрою та порівняння їх за критеріями швидкодії та коефіцієнта використання матеріалу. **Об'єктом дослідження** є процеси побудови розкрою листових матеріалів на прямокутні частини з мінімізацією відходів. **Предметом дослідження** є метаевристичні алгоритми для вирішення задач гільйотинного розкрою листового матеріалу. **Результати.** У даній роботі розглядаються особливості застосування евристичних та метаевристичних підходів для вирішення задачі розкрою листового матеріалу. Представлені результати експериментальних досліджень, що демонструють переваги та недоліки кожного з запропонованих підходів для рішення поставленої задачі. **Висновок.** Найбільшу швидкодію забезпечує запропонований метод на основі алгоритму Best Fit Decreasing (BFD), проте найвищу точність демонструє підхід на основі мурашиного алгоритму.

Ключові слова: задача гільйотинного розкрою, коефіцієнт використання матеріалу, генетичний алгоритм, мурашиний алгоритм, рівневі алгоритми, евристика, метаевристика, селекція, кросовер, мутація

Вступ

Із розвитком технологій і зростанням потреб у раціональному використанні ресурсів, завдання оптимізації розкрою матеріалів набуло особливого значення у таких галузях, як будівництво, деревообробка, металургія та текстильна промисловість. Задача гільйотинного розкрою спрямована на мінімізацію відходів при розрізанні великих аркушів матеріалів на менші частини заданого розміру.

Задача прямокутного розкрою належить до класу NP-складних завдань комбінаторної оптимізації, що ускладнює її вирішення при збільшенні кількості елементів і можливих варіантів розкрою. Використання точних методів, таких як повний перебір, неможливе через експоненційне зростання обчислювальної складності, тому актуальним є використання евристичних і метаевристичних підходів, які дозволяють знаходити наближені до оптимальних рішення за прийнятний час [1].

До поширених евристичних методів вирішення задачі розкрою відносяться рівневі алгоритми (First Fit, Next Fit, Best Fit), використання яких є доцільним для випадків з обмеженим числом варіантів розкрою. У дослідженні [2] розглядаються рівневі жадібні алгоритми для двовимірної упаковки, включаючи стратегії NFDH, FFDH і BFDH, які розміщують елементи рядами (рівнями) у контейнері. Також аналізується їх ефективність, складність та найгірші випадки роботи. Проаналізовані рівневі підходи показали свою доцільність у задачах двовимірного пакування, особливо для швидкого отримання наближених рішень. Однак використання жадібної евристики у цих алгоритмах може призводити до неефективного використання простору, а найкращі результати досягаються при поєднанні рівневих методів із двофазовими чи метаевристичними підходами [2].

Генетичні алгоритми, як представники еволюційних методів, дозволяють ефективно вирішувати

задачі розкрою шляхом ітеративного покращення рішень. Вони імітують природний процес еволюції, використовуючи механізми відбору, схрещування та мутації для формування нових поколінь варіантів рішень [3]. У [4] пропонується генетичний алгоритм для вирішення як орієнтованого, так і неорієнтованого випадків завдання прямокутної упаковки. Значається проблема налаштування параметрів генетичного алгоритму, таких як розмір популяції, ймовірності кросоверу та мутації, що значно впливають на результативність алгоритму [5, 6].

Перспективним підходом є використання алгоритму мурашиної колонії (ACO, Ant Colony Optimization), що моделює поведінку мурах у природі. Цей алгоритм використовує феромонові мітки для пошуку оптимальних рішень і демонструє високу ефективність у вирішенні задач розкрою з великою кількістю варіантів [7]. Ефективність ACO залежить від налаштувань швидкості випаровування феромону та інших початкових умов [8]. У статті [9] розглядається задача гільйотинного розкрою прямокутних заготовок, запропоновано підхід на основі алгоритму мурашиної колонії та проведено чисельні експерименти для оцінки його ефективності. Розглянутий підхід показав свою доцільність у випадках розкрою змішаних, малих та довгих прямокутників, забезпечуючи коефіцієнт використання матеріалу до 95,8%. Водночас автори зазначають, що ефективність алгоритму залежить від вибору параметрів і специфіки вхідних даних, що може вимагати додаткового налаштування для інших класів задач.

Метою роботи є аналіз ефективності використання евристичних та метаевристичних підходів для вирішення задачі гільйотинного розкрою. Реалізовано підходи на основі рівневих та генетичних алгоритмів та методу мурашиної колонії. Увага приділена дослідженню таких критеріїв ефективності як швидкість виконання, коефіцієнт використання матеріалу і кількість використаних аркушів.

Постановка задачі

Задача гільйотинного розкрою формулюється наступним чином: дано прямокутний лист матеріалу розміром $W \times H$ (ширина та висота відповідно), який потрібно розрізати на менші прямокутні частини із заданими розмірами $w_i \times h_i$ для $i = 1, 2, \dots, N$ [10]. Розглядається варіант завдання з ортогональним розкром, через що допускається поворот деталей на кут 90 градусів, причому деталі розміщуються паралельно до осей координат. Необхідно знайти такий спосіб розкрою, який мінімізує відходи з кожного листа та при якому використано мінімальну кількість листів матеріалу M . У якості критерію оптимізації використовується коефіцієнт використання матеріалу (КВМ) або коефіцієнт розкрою [11]. Оскільки витрати матеріалу необхідно мінімізувати, відповідно КВМ необхідно максимізувати.

Цільова функція (1) виглядає наступним чином:

$$\frac{\sum_{i=1}^N (w'_i \times h'_i)}{M \times W \times H} \rightarrow \max. \quad (1)$$

Обмеження задачі включають:

$$0 \leq x_i + w_i \leq W, 0 \leq y_i + h_i \leq H, \forall i, \quad (2)$$

$$(x_i + w_i \leq x_j) \vee (x_j + w_j \leq x_i) \vee (y_i + h_i \leq y_j) \vee (y_j + h_j \leq y_i), \forall i \neq j, \quad (3)$$

$$\sum_{i=1}^N (w'_i \times h'_i) \leq M \times W \times H, \quad (4)$$

$$w'_i = \begin{cases} w_i, & \text{якщо } r_i = 0, \\ h_i, & \text{якщо } r_i = 1, \end{cases} \quad (5)$$

$$h'_i = \begin{cases} h_i, & \text{якщо } r_i = 0, \\ w_i, & \text{якщо } r_i = 1, \end{cases} \quad (6)$$

де x_i, y_i – координати лівого верхнього кута i -го елемента на аркуші ($i = 1, 2, \dots, N$), $r_i \in \{0, 1\}$ ознака повороту i -го елемента.

Обмеження моделі (2) забезпечує, щоб всі елементи були розміщені в межах використаного аркуша. Умова (3) встановлює, що елементи не повинні перекриватися. Згідно з (4), сумарна площа елементів не повинна перевищувати площу аркушів. Вирази (5) та (6) визначають можливість повороту елементів.

Підходи до вирішення на основі евристичних та метаевристичних методів

Евристичні методи вирішення задачі гільйотинного розкрою спираються на спрощені правила та емпіричні підходи, засновані на попередньому досвіді. Вони дозволяють знаходити прийнятні рішення за допомогою загальних евристик, таких як First Fit, Next Fit, Best Fit, Worst Fit, Smallest Piece Method тощо [2].

Метаевристичні методи використовують евристику в поєднанні з пошуковими алгоритмами для досягнення оптимального чи наближено оптимального рішення. Вони працюють у ітеративному режимі, де на кожній ітерації відбувається поліпшення поточного рішення. Метаевристичні методи для задачі розкрою можуть включати генетичні алгоритми (genetic algorithms), мурашині алгоритми (ant colony algorithms), симуляцію відпау (simulated annealing) та інші.

У роботі запропоновано методи на основі найпоширеніших евристичних та мета-евристичних алгоритмів для вирішення задачі розкрою: рівневий (Best Fit), генетичний (Genetic Algorithms) та мурашиний (Ant Colony Optimization – ACO) алгоритми.

За основу рівневих алгоритмів приймається принцип послідовного укладання елементів у прямокутну область, розташовану в межах листа [12]. Реалізація такого алгоритму передбачає кілька етапів. Спочатку деталі сортуються за площею у порядку спадання, що дозволяє спочатку розміщувати найбільші елементи. Далі відбувається послідовне розміщення деталей на наявних аркушах. Для кожної деталі перевіряються всі доступні позиції, і вибирається найкраща з точки зору мінімізації залишкового простору. Також враховується можливість повороту деталей. Якщо деталь не поміщається у стандартному положенні, вона обертається на 90 градусів і повторно перевіряється можливість її розміщення. У випадку, якщо жодна з наявних позицій не підходить, створюється новий аркуш листового матеріалу.

Описаний підхід спрямовано на оптимізацію використання простору на аркуші, максимально заповнюючи кожну область перед створенням нової. Таким чином, елементи укладаються на листі в міру їх надходження, спочатку розміщуючи більші елементи, а потім заповнюючи простір, що залишився, меншими елементами [13].

Генетичний алгоритм (ГА) – це математична модель еволюції популяції штучних особин. У контексті завдання розкрою, особиною вважається варіант розташування деталей на листі, що задовольняє умовам задачі. Хромосома, яка описує особину, є кодованим способом розподілу об'єктів на площині. Кожен ген у хромосомі представляє положення або орієнтацію прямокутників на площині. Конкретні координати розміщення предметів на листі можна отримати, розмістивши деталі в порядку їх перерахування в хромосомі за правилами декодування [14].

У запропонованому алгоритмі застосовується метод розміщення об'єктів з перевіркою на перетини, що передбачає поступове додавання об'єктів у межах допустимого простору. У разі виявлення накладання реалізується процедура коригування положення або орієнтації об'єкта, включаючи можливість його повороту на 90 градусів з метою оптимізації заповнення площі.

Робота підходу на основі генетичного алгоритму починається зі створення початкової популяції, у якій кожна особина генерується випадковим чином. Деталі розміщуються на листі у випадковому порядку з випадково обраною орієнтацією, і для кожного рішення обчислюється функція пристосованості, яка визначає ефективність використання площі. Функція пристосованості обчислюється як зворотне значення до площі, яку займають елементи на аркуші. Чим більше значення функції пристосованості, тим ефективнішим є розкрій.

Наступним етапом є відбір батьківських особин для створення нового покоління. Відбір здійснюється за допомогою турнірного методу, при якому випадково обираються декілька особин, і з них вибирається

найбільш пристосована. Таким чином, кращі рішення мають більше шансів на розмноження.

Після відбору відбувається операція схрещування, де дві батьківські хромосоми комбінуються, утворюючи дві нові особини. Використовується одноточковий кросовер, що дозволяє створювати нащадків із властивостями обох батьків шляхом поділу їх хромосом у випадковому місці та обміном відповідних генів.

На етапі мутації проводиться випадкова зміна генів особин-нащадків, що сприяє збереженню генетичного різноманіття популяції. Мутація може включати зміну координат розташування деталей, зміну орієнтації або обмін місцями двох елементів. Цей процес допомагає уникнути локальних оптимумів та покращити кінцеве рішення.

Цикл відбору, схрещування та мутації повторюється протягом заданої кількості ітерацій або поки не буде досягнуто іншої критерій зупинки.

Алгоритм мурашиної колонії імітує поведінку мурах, які залишають феромонові сліди при пошуку шляху від мурашника до джерела їжі. У задачі розкרוю при використанні підходу на основі алгоритму мурашиної колонії кожен агент (мураха) поступово розміщує деталі на аркуші, обираючи наступну деталь на основі ймовірності [15], яка визначається рівнем феромонного сліду та евристичною оцінкою

$$P_i = \frac{\tau_i^\alpha \eta_i^\beta}{\sum_j \tau_j^\alpha \eta_j^\beta}, \quad (7)$$

де τ_i – інтенсивність феромонного сліду, η_i – евристична оцінка (обернена до площі залишків), α і β – вагові коефіцієнти.

Після кожної ітерації оновлюється рівень феромонів (найкращі розміщення підсилюють свої феромонові сліди, тоді як менш ефективні шляхи розкרוю поступово зникають, що дозволяє алгоритму досліджувати нові варіанти розкרוю):

$$\tau_i = (1 - p) \tau_i + \Delta \tau_i, \quad (8)$$

де τ_i – коефіцієнт випаровування, $\Delta \tau_i$ – внесок нових рішень у феромонний слід. Завдяки цьому АСО поступово знаходить рішення з мінімальними втратами матеріалу.

Описаний процес повторюється, доки не буде досягнуто максимальної кількості ітерацій.

Таким чином, АСО алгоритм мурашиної колонії у задачі розкרוю ефективно моделює колективну інтелектуальну поведінку мурах для знаходження оптимального розкרוю матеріалу.

Результати експериментальних досліджень

Для проведення експериментів були використані набори, що складаються з 5, 10, 20, та 40 деталей. Налаштування генетичного алгоритму було здійснене наступним чином:

- для генерації початкової популяції особин використано випадкове розміщення прямокутних блоків на листі з урахуванням обмежень на перетини. Популяція містить 3000 особин;
- умова зупинки – досягнення 3000 поколінь;
- для схрещування особин застосовується одноточковий кросовер. Точка розрізу обирається випадково в межах довжини розміщення блоків;
- вибір батьківських особин здійснюється шляхом турнірного відбору з 2 учасниками;
- використовуються три типи мутації: інверсна, випадкова зміна координат блоку та обмін орієнтації блоку. Ймовірності застосування кожного типу мутації становлять 0.33, 0.33 та 0.34 відповідно.

Для дослідження мурашиного алгоритму, були встановлені такі початкові умови експерименту:

- кількість агентів (мурах) становить 500, що забезпечує високу варіативність у пошуку рішення;
- умова зупинки – досягнення 1000 ітерацій;
- коефіцієнт випаровування феромону встановлений на рівні 0.5, що забезпечує поступове зменшення впливу феромонного сліду, сприяючи пошуку нових варіантів розміщення;
- параметри α та β мають значення 1 та 2 відповідно, що налаштовує алгоритм на баланс між впливом феромону і евристичною інформацією про розміри блоків.

В табл. 1 наведені отримані результати дослідження ефективності підходів за критеріями часу виконання та коефіцієнта використання матеріалу (КВМ).

Таблиця 1 – Ефективність роботи підходів на основі рівневого, генетичного та мурашиного алгоритмів

Застосовані алгоритми	Критерії ефективності	Кількість деталей			
		5	10	20	40
Best Fit	Час, мс	54.50	46.90	98.90	115.70
	КВМ, %	53.85	53.92	63.71	72.38
	Листи, шт	1	1	1	1
Генетичний алгоритм	Час, мс	1383.60	2809.30	7363.60	34510
	КВМ, %	56.26	43.28	28.88	24.96
	Листи, шт	1	1	3	4
Мурашиний алгоритм	Час, мс	1247.0	64736.60	475345.10	1260009.03
	КВМ, %	82.07	70.51	71.18	73.20
	Листи, шт	1	1	1	1

На рис. 1 наведено графічне порівняння часу виконання, а на рис. 2 – залежність КВМ від кількості деталей у задачі.

Стратегія, що використовує рівневий алгоритм (Best Fit) демонструє найменший час виконання (рис. 1), що свідчить про його високу швидкість.

Цей метод підходить для випадків, де першочергове значення має швидкість отримання рішень.

Наприклад, при 5 деталях Best Fit працює за 54.50 мс, а при 40 деталях – за 115.70 мс, що значно швидше порівняно з підходами на основі метаевристичних алгоритмів.

Метод, що базується на генетичному алгоритмі потребує значно більші витрати часу через складність обчислювальних операцій, пов'язаних із процесом еволюційного пошуку. Час виконання збільшується від 1383.60 мс для 5 деталей до 34510 мс для 40 деталей (рис. 1).

Це обумовлено необхідністю багаторазового генерування нових рішень та їхнього оцінювання.

Рішення, побудоване на основі мурашиного алгоритму демонструє ще більші витрати часу (рис. 1), оскільки кожен агент (мураха) обчислює можливі шляхи розташування деталей, що потребує великої кількості обчислювальних ітерацій.

У випадку з 40 деталями час виконання досягає 1260009.03 мс, що значно обмежує можливість його практичного застосування в реальних умовах.

Головним критерієм ефективності розкрою є КВМ, який показує, яка частка площі аркуша була використана для розміщення деталей.

Чим вище значення КВМ, тим кращий результат (рис. 2).

Підхід на основі алгоритму Best Fit показує відносно низькі значення КВМ – від 53.85% до 72.38%, що свідчить про значні залишки матеріалу після розкрою.

Це обумовлено тим, що алгоритм не здійснює глобальної оптимізації, а лише локально розміщує деталі в найкращі доступні області.

Стратегія із використанням генетичного алгоритму демонструє зміну ефективності в залежності від складності задачі. Для 5 деталей він забезпечує КВМ 56.26%, однак при 40 деталях значення КВМ падає до 24.96%.

Це свідчить про зниження ефективності при збільшенні розміру задачі, що може бути наслідком обмеженості поколінь або потреби в оптимізації параметрів алгоритму.

Найкращі результати з точки зору використання матеріалу забезпечує метод на основі мурашиного алгоритму (рис. 2), показуючи КВМ в межах 73.20–82.07%. Це пояснюється здатністю алгоритму адаптивно змінювати стратегію розміщення та зменшувати невикористані простори. Проте, висока обчислювальна складність є значним обмеженням для його практичного застосування.

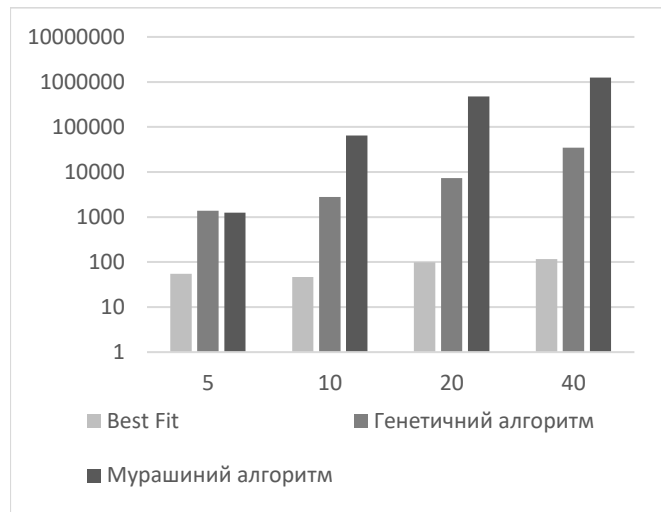


Рис. 1. Діаграма витрат часу при обчисленні розкрою з різною кількістю деталей

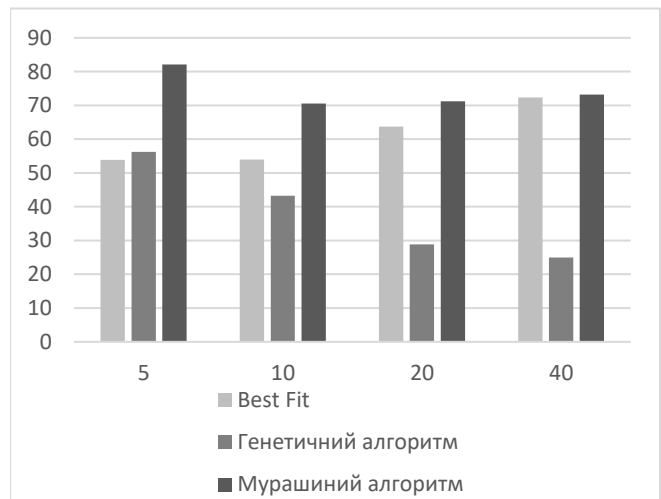


Рис. 2. Діаграма значень КВМ у результаті роботи запропонованих підходів при різній кількості деталей

Значущим фактором оцінки ефективності є також кількість використаних листів матеріалу. Найкращий результат за цим критерієм демонструють підходи на основі Best Fit і мурашиного алгоритму, які дозволяють розмістити всі деталі на одному листі.

Генетичний алгоритм у випадках підвищеної складності задачі має низьку ефективність використання матеріалу (табл. 1).

Висновки

У роботі проведено порівняльний аналіз евристичного методу на основі рівневого алгоритму, та метаевристичних підходів що базуються на генетичному алгоритму та методу мурашиної колонії для вирішення задачі гільйотинного розкрою. Експериментальні результати показали, що рівневі алгоритми демонструють найвищу швидкість, тоді як метаевристичні алгоритми забезпечують ефективніше використання матеріалу.

Найвищий коефіцієнт використання матеріалу досягнуто за допомогою підходу на основі алгоритму

мурашиної колонії, що свідчить про його здатність знаходити рішення з мінімальними залишками матеріалу.

Використання генетичного алгоритму забезпечує високі результати, але його ефективність знижується зі збільшенням складності задачі через високу обчислювальну складність.

Подальші напрями досліджень можуть включати:

- гібридизацію рівневих та метаевристичних методів для досягнення компромісу між швидкістю та ефективністю використання матеріалу;
- оптимізацію параметрів мурашиного алгоритму з метою зменшення часу обчислень;
- використання адаптивних стратегій у процесі налаштування параметрів підходу на основі генетичного алгоритму для підвищення його ефективності при великій кількості елементів розкрою.

СПИСОК ЛІТЕРАТУРИ

1. Кононенко А.І., Івашченко Г.С. (2023), “Аналіз методів вирішення задачі розкрою листового матеріалу”, III Всеукраїнська студентська наукова конференція “Науковий простір: аналіз, сучасний стан, тренди та перспективи”, С. 123-124, doi: <https://doi.org/10.36074/liga-ukr-16.06.2023>
2. Lodi A., Martello S., Vigo D. (2002), “Recent advances on two-dimensional bin packing problems”, *Discrete Applied Mathematics*, Vol. 123/124, pp. 373-380, doi: [https://doi.org/10.1016/S0166-218X\(01\)00347-X](https://doi.org/10.1016/S0166-218X(01)00347-X)
3. OuYang Q., Xu H. (2015), “The study of comparisons of three crossover operators in genetic algorithm for solving single machine scheduling problem”, *International Conference on Manufacturing Science and Engineering (ICMSE 2015)*, Atlantis Press, pp. 293-297, doi: <https://doi.org/10.2991/icmse-15.2015.55>
4. Bortfeldt A. (2006), “A genetic algorithm for the two-dimensional strip packing problem with rectangular pieces”, *European Journal of Operational Research*, Vol. 172, No 3, pp. 814-837, doi: <https://doi.org/10.1016/j.ejor.2004.11.016>
5. Kazakovtsev L.A., Stanimirovic P.S. (2013), “An Approach to the Multi-facility Weber Problem with Special Metrics”, *European Modelling Symposium (EMS)*, pp. 119-124, doi: <https://doi.org/10.1109/EMS.2013.21>
6. Kumar R., Memoria M., Thapliyal M., Kirola M., Ahmad I., Gupta A., Tyagi S., Ansari N. (2022), “Analyzing The Performance Of Crossover Operators (OX, OBX, PBX, MPX) To Solve Combinatorial Problems”, *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, pp. 817-821, doi: <https://doi.org/10.1109/COM-IT-CON54601.2022.9850689>
7. Валеева А.Ф., Петунін А.А., Файзрахманов Р.І. (2006), “Застосування конструктивних евристик у завданнях розкрою-упаковки”, Додаток до журналу «Інформаційні технології», № 11, С. 1-24.
8. Chopard B., Tomassini M. (2018), “The Ant Colony Method”, *An Introduction to Metaheuristics for Optimisation*, pp. 81-96, doi: https://doi.org/10.1007/978-3-319-93073-2_5
9. Валеева А.Ф., Петунін А.А., Файзрахманов Р.І. (2007), “Застосування конструктивної метаевристики «мурашина колонія» до завдання гільйотинного прямокутного розкрою”, *Вісник Башкирського університету*, Т. 12, № 3, С. 12-14.
10. Петунін А.А., Полезов А.В., Куреннов Д.В. (2005), “Про один підхід до вирішення завдань розкрою-упаковки”, *Конструювання та технологія виготовлення машин: Збірник наукових праць*, Ч. 2, *Вісник УДТУ – УПІ*, № 18 (70), С. 212-216.
11. Орлов А.Н., Курейчик В.В., Кудрякова Т.Ю. (2015), “Комбінований алгоритм розв'язання задачі прямокутного розкрою”, *Праці Конгресу з інтелектуальних систем та інформаційних технологій «IS&IT'15»*, С. 212-217.
12. Валиахметова Ю.І., Телицький С.В. (2012), “Застосування систем автоматизованого проектування карт розкрою в суднобудуванні”, *Вісник ВДТУ*, № 6, С. 38-43.
13. Whitwell G. (2004), “Novel Heuristic and Metaheuristic Approaches to Cutting and Packing”, *BSc Thesis Submitted to the University of Nottingham for the degree of Doctor of Philosophy, School of Computer Science and Information Technology*, pp. 68-71.
14. Подлазова А.В. (2008), “Генетичні алгоритми на прикладах вирішення задач розкрою”, *Проблеми управління*, № 2, С. 57-63.
15. Валеева А.Ф. (2005), “Застосування метаевристики мурашиної колонії до задач двомірної упаковки”, *Інформаційні технології*, № 10, С. 36-43.

Received (Надійшла) 22.12.2024

Accepted for publication (Прийнята до друку) 05.02.2025

Metaheuristic methods for solving the guillotine cutting stock problem

Heorhii Ivashchenko, Anastasiia Kononenko, Daria Tymoshenko

Abstract. Topicality. The optimization of sheet material cutting is a critical task in the industry, enabling waste minimization and cost reduction in production. This article examines methods for solving the guillotine-cutting stock problem, including level algorithms (specifically, Best Fit Decreasing), genetic algorithms, and the ant colony algorithm. **The goal of this work** is to analyze the efficiency of heuristic and metaheuristic approaches to the cutting problem and compare them based on performance speed and material utilization rate. **The object of research** is the process of cutting sheet materials into rectangular parts while minimizing waste. **The subject of research** is the algorithms used to solve the guillotine-cutting stock problem. **Results.** This paper explores the application of heuristic and metaheuristic approaches to solving the sheet material cutting problem. The results of experimental research demonstrate the advantages and disadvantages of each proposed approach for solving the given problem. **Conclusions.** The Best Fit Decreasing (BFD) algorithm-based method achieves the highest processing speed, whereas the ant colony algorithm-based method demonstrates the highest accuracy.

Keywords: guillotine cutting stock problem, material utilization rate, genetic algorithm, ant colony algorithm, level algorithms, heuristics, metaheuristics, selection, crossover, mutation.

Є. О. Ігнатюк, А. В. Попов

Національний аерокосмічний університет «Харківський авіаційний інститут», Харків, Україна

МЕТОДИКО-ІНСТРУМЕНТАЛЬНІ ЗАСОБИ АВТОМАТИЗОВАНОГО АНАЛІЗУ ДАНИХ РЕЗУЛЬТАТІВ UX-ДОСЛІДЖЕНЬ З ВИКОРИСТАННЯМ СИСТЕМ ШТУЧНОГО ІНТЕЛЕКТУ

Анотація. Стаття присвячена автоматизованому аналізу результатів досліджень взаємодії (UX-досліджень) з використанням систем штучного інтелекту, зокрема нейронних мереж та експертних систем. Метою дослідження є розробка системи методико-інструментальних засобів для автоматизації аналізу даних UX-досліджень шляхом інтеграції технологій штучного інтелекту та експертних систем. Проведено аналіз актуальності проблеми автоматизації обробки великих обсягів різномірних даних, що отримані в результаті досліджень взаємодії користувачів з цифровими продуктами. Запропоновано класифікацію методів досліджень взаємодії користувача та цифрових продуктів за типами даних та підхід до вибору відповідних архітектур нейромереж для їх аналізу. Розглянуто можливості застосування рекурентних та згорткових нейронних мереж як приклад для обробки текстових, аудіо- та відеоданих. Також запропоновано використання експертної системи для автоматизованого вибору нейромереж, оптимальних для аналізу різнотипних даних. Розроблено та протестовано пробну версію системи з використанням API штучного інтелекту, що підтверджує можливість автоматизованої обробки різних типів даних UX-досліджень за допомогою нейромереж. Окреслено перспективи подальшого розвитку системи аналізу даних результатів UX-досліджень, включаючи розробку програмного забезпечення, розширення функціональності та адаптацію для суміжних галузей. Результати дослідження мають практичне значення для підвищення ефективності аналізу даних та можуть бути використані для вдосконалення процесу проєктування інтерфейсів.

Ключові слова: UX-дослідження, автоматизація, аналіз даних, нейронні мережі, експертні системи.

Постановка проблеми

В сучасному світі якість досвіду взаємодії є критично важливим фактором успіху цифрових продуктів та послуг. Поняття «досвід взаємодії» (User Experience, UX) містить загальне представлення того, як кінцеві користувачі будуть взаємодіяти з інтерфейсами. Це поняття передбачає розуміння як поведінки користувачів, так і бізнес-процесів, для яких цей інтерфейс створюється.

Процес проєктування інтерфейсів передбачає проведення досліджень, які дозволяють приймати обґрунтовані дизайн-рішення. Однак, традиційні методи проведення досліджень взаємодії мають недоліки, такі як значні витрати людських ресурсів та часу на збір й оброблення даних. Використання технологій штучного інтелекту відкриває можливості для автоматизації та оптимізації аналізу даних досліджень взаємодії користувача з інтерфейсами. Інтелектуальні системи здатні швидко обробляти великі обсяги даних, виявляти приховані закономірності — це дозволяє суттєво прискорити дослідження та отримати глибше розуміння користувачів, їх проблем, пов'язані з задоволенням потреб (комунікаційних, навчальних, управлінських, розважальних, бізнес тощо), а також труднощів (доступності, ефективності, персоналізації послуг тощо).

Необхідним є формування та розробка плану впровадження ефективної системи та інформаційної технології досліджень взаємодії з використанням систем штучного інтелекту, що дозволить зменшити ручну обробку даних та прискорити процес проєктування інтерфейсів. Для досягнення поставленої мети були сформульовані наступні завдання дослідження:

1. Аналіз сучасного стану UX-досліджень та чинних методів їх проведення;

2. Дослідження технологій штучного інтелекту, які можуть бути використані для автоматизації;

3. Проєктування експертної системи для вибору необхідної нейронної мережі на основі типів даних;

4. Формулювання рекомендацій щодо використання методів UX-досліджень та інформаційної технології для прийняття рішень в процесі проєктування інтерфейсів.

Впровадження запропонованих методів та технологій матиме позитивний ефект на всі сфери, де проєктуються інтерфейси цифрових продуктів. Автоматизація результатів досліджень взаємодії дозволить підвищити загальний рівень зручності цифрових продуктів, зробити їх більш зрозумілими, зручними та привабливими для користувачів. Під цифровим розуміється будь-який продукт, який надається користувачеві в цифровому форматі. Наприклад, програмне забезпечення, мобільні застосунки, вебсайти, онлайн-сервіси, медіа тощо.

Таким чином, тема аналізу даних результатів UX-досліджень з використання систем штучного інтелекту має безпосередній зв'язок з важливими науковими та практичними завданнями сьогодення — підвищенням якості та ефективності процесу проєктування інтерфейсів цифрових продуктів, які максимально відповідають потребам та очікуванням користувачів. Розв'язання поставлених задач матиме значний вплив на розвиток сфери досвіду користування цифровими продуктами та цифрової економіки загалом.

Аналіз останніх досліджень і публікацій

Значна кількість публікацій, що охоплюють різні аспекти теми дослідження, включаючи процес проєктування та створення інтерфейсів, UX-дослідження, аналіз великих даних та штучний інтелект,

доводять актуальність проблеми автоматизованого аналізу UX-даних. Саме на перетині цих технологій виникає потреба в автоматизації результатів цих досліджень. Для вивчення проблеми було проаналізовано спектр різноманітних джерел.

Розпочати аналіз стану сучасних наукових робіт та матеріалів було вирішено зі світових лідерів у сфері дослідження користувацького досвіду — Nielsen Norman Group. Вони відомі своїми численними публікаціями, конференціями та навчальними програмами з теми дизайну взаємодії та проектування інтерфейсів. Під їх керівництвом було видано понад 2500 наукових статей, а також започатковано та апробовано систему сертифікації Nielsen Norman Group UX Certification System. Починаючи з автентичної праці «Usability 101: Introduction to Usability» [1] було введено поняття зручності в контексті дизайну інтерфейсів. Автори зазначають, що в сучасному світі для привернення та утримання уваги користувачів до продуктів та сервісів, вони повинні бути не лише функціональними, але й зручними та легкими у використанні. В цій статті наводяться визначення та компоненти досвіду взаємодії та важливість постійного вдосконалення цього досвіду. Зазначається, що створювані сервіси та системи мають бути орієнтовані першою чергою на кінцевого споживача. Людиноцентричний характер процесів спонукає до збільшення інвестицій робочих зусиль в процеси, пов'язані з проектуванням інтерфейсів та взаємодії з користувачем. В тому числі збільшують проведення досліджень користувацької поведінки, тестування прототипів, впровадження зручних інтерфейсних рішень та постійне вдосконалення продукту на основі зворотного зв'язку від користувачів на кожному етапі розробки програмного забезпечення.

Додатково був розглянутий інший стандарт в контексті розробки цифрових інтерфейсів — серії ISO, а конкретніше, його частина про зручність використання людино-машинних інтерфейсів «Ergonomics of human-system interaction - Part 11: Usability: Definitions and concepts» [2]. Цей стандарт розглядає поняття зручності (usability) у контексті взаємодії людини з системами. Робота висвітлює значення та складові зручності, такі як ефективність та задоволення від взаємодії з цифровим інтерфейсом. Загальна мета стандарту полягає у сприянні розвитку констант та методик оцінки зручності використання цифрових продуктів, що допоможе забезпечити створення ефективних та користувацько-орієнтованих бізнес-систем. Саме тому стандарт ISO пропонує чіткі власні визначення таких термінів, як «Зручності використання цифрових продуктів», «Дослідження взаємодії», «Методи UX-дослідження», а також «Людино-орієнтовного дизайну», де застосування людських факторів, а також знань і методів ергономіки та зручності використання є однією з обов'язкових складових дизайн-процесу. Загалом, стандарт надає концептуальний каркас для подальшого дослідження та впровадження в практику.

Досліджуючи безпосередньо напрям людиноцентричних моделей проектування та реалізації програмного забезпечення, були проаналізовані відповідні

наукові роботи та публікації. В праці «Impact of Human-Centered Design Process (HCDP) on software development process» [3] описана важливість людського фактора в процесі розробки інтерфейсу, де передбачається залучення людської перспективи на кожному етапі розробки. В цій роботі досліджується вплив застосування процесу людсько-направленого дизайну HCDP (Human-Centered Design Process) на підвищенню зручності використання та ефективності користувацького досвіду. Також в праці пропонується модель людсько-направленого дизайну HCD (Human-Centered Design) та аналіз PSC (People using Software in a Context) для досягнення високого рівня інтеграції з процесом розробки програмного забезпечення для високого рівня зручності використання створюваних цифрових продуктів.

Досліджуючи наукові праці у сфері UX-методів, було знайдено сукупність робіт, що порушують питання автоматизації оцінки зручності інтерфейсів цифрових продуктів та проведення методів UX-досліджень. Ці праці підтверджують актуальність проблеми автоматизації у сфері користувацького досвіду. Варто зауважити контекстуальну різницю — усі зазначені в цьому розділі роботи не розглядають автоматизацію обробки результатів досліджень взаємодії користувача з цифровим інтерфейсом. Саме цей процес автоматизації є основною метою майбутніх досліджень та робіт. Наразі публікації містять автоматизацію у сферах прийняття рішень, порівнянь числових об'ємів даних або загальну оцінку стану UX та зручності поточного продукту. В публікації «Detect and interpret: towards operationalization of automated user experience evaluation» [4] зазначається важливість автоматизованої оцінки досвіду взаємодії кінцевого користувача з графічними інтерфейсами. Однак зазначається, що чинні методи оцінки часто потребують високого рівня людської участі у зборі та аналізі даних. Авторки дослідження наголошують на важливості залучення експертів для проведення UX-аналізу та аналізу результатів проведених досліджень. Зазначається, що часто ціна залучення експерта для проведення та аналізу даних UX-дослідження є високою, в порівнянні з вартістю роботи інших спеціалістів в процесі розробки інтерфейсів цифрових продуктів. В роботі також зазначаються виклики автоматизованої UX-оцінки зручності використання інтерфейсами. З огляду на зростаючу кількість продуктів та користувацьку базу й пов'язані з ними дані, постійний моніторинг зручності цифрових продуктів, з якими взаємодіє користувач, стає потрібним, але водночас — складним. Це вимагає розробки систем, які здатні автоматично відстежувати UX та надавати відомості про необхідні покращення інтерфейсів. Також були розглянуті недоліки наявних підходів, в яких увага, як правило, фокусується на визначенні метрик, що відображають стан інтерфейсів, але не завжди дозволяють виявити потенційні причини проблем. В праці сформоване визначення даних поведінки користувача (User behavioral data), на основі яких відбувається аналіз метрик та UX-оцінка. З автоматичних програм та бібліотек, що колекціонують статистику, були зазначені в тому числі Google Analytics

та Hotjar. Проте ці програми згадані виключно в контексті автоматизованого збору даних користувачів і є лише однією з проміжних ланок в процесі аналізу даних UX-досліджень. Ці методи відносяться до кількісних статистичних методів досліджень користувальницького досвіду, автоматизація яких є важливою частиною оцінки зручності цифрових продуктів. На противагу — в досліджуваній роботі не вистачає частини, в якій би розглядалась автоматизація результатів якісних досліджень на основі User behavioral data. Такі методи автоматизуються найважче, ціна залучення експерта для їх аналізу є найбільш коштовною, а їх опрацювання займає більше часу.

Щоб отримати чітке розуміння чинних методів дослідження, було проаналізовано наукові роботи, що стосуються цих методів й способів їх проведення. В роботі «The past, present, and future ^[1]_{SEP} of UX empirical research» [5] прослідковуються тези необхідність переосмислення UX, аналізуючи понад 400 академічних емпіричних досліджень. Автори ставлять собі завдання визначити, як змінились аналіз та методи UX-дослідження. Було виявлено значний ріст кількісних та змішаних методологій дослідження користувальницького досвіду. Були найбільш поширені методи дослідження зручності використання цифрових продуктів, такі як опитування, огляд відеосесії користувачів та глибинні інтерв'ю.

Питання підхожих способів автоматизації даних були також досліджені в роботах сфери штучного інтелекту й нейромереж, а також вже протестованих способів аналізу даних за допомогою залучення нейронних мереж. Стаття «An integrated text analytic framework for product defect discovery» [6] пропонує інтегровану текстову аналітичну бібліотеку для виявлення дефектів у UX-продукту. Основна ідея статті полягає в розробці методології для автоматичного виявлення дефектів продукту за допомогою аналізу текстової інформації нейроною мережею. Бібліотека ефективно використовує різноманітний контент соціальних медіа та кілька автоматично сформованих сигнальних ознак, які можуть бути використані як вхідні дані для виявлення дефектів продукту. Сам підхід використовувався більше для тестування та QA-конкретних програмних систем, аніж для оцінки та покращення досвіду взаємодії. Через те, що бібліотека аналізувала метрики, які збирались автоматично та були лише текстовими, їх сфера застосування була обмежена. Автори визначають потребу у подальших дослідженнях, спрямованих на аналіз реалізації їхньої бібліотеки в інших галузях, де сфера автоматичного аналізу даних може бути масштабована з текстової до графічної та відео.

Публікація «UX Calculator: An online tool to support user testing» [7] підтверджує важливість досліджень взаємодії користувача з цифровими інтерфейсами для конкурентоспроможності бізнесу. Вони спрямовані на зручність користування та корисність, а отже стають ключовими для комерційного успіху. У статті презентується UX Calculator — вебінструмент, який дозволяє проводити статистичні тести, призначені для основних аналізів, пов'язаних з оцінкою зручності. Він дозволяє дизайнерам проводити

порівняння між двома групами та отримувати надійні висновки про продуктивність. Область застосування є обмеженою, бо він дозволяє лише порівнювати певну вибірку однорідних й лише текстових даних й не включає інтелектуальний аналіз різноманітної інформації. Цей інструмент підхожий для порівняльного аналізу результатів взаємодії з двома версіями інтерфейсу, проте також може бути використаний в подальших розробках більш самостійних та широких систем автоматизації досліджень.

В раніше розглянутих наукових роботах були визначені формати даних, процес аналізу яких може бути автоматизованим. Виходячи з типів UX-досліджень, інформація їх результатів може бути найрізноманітніша: текстова, аудіальна, візуальна, в тому числі включаючи відео. До прикладу, в результаті записів користувальницьких сесій Hotjar робочі групи отримують аудіозаписи, та відеозаписи з динамічними зображеннями теплових мап, а також та відеосесії користувальницької поведінки з сайтів. Аналіз такої інформації може забезпечити якісне проведення досліджень, що необхідно для покращення досвіду взаємодії та сфери інтерфейсного дизайну в цілому. Відповідно необхідно розглянути різноманітні технології нейромереж, які можуть бути використані для автоматизації таких даних. В роботі «Online retrainable neural networks: improving the performance of neural networks in image analysis problems» [8] автор пропонує новаторський підхід до покращення нейронних класифікаторів у завданнях розпізнавання зображень, сегментації або кодування на основі процедури повторного навчання на рівні користувача. Алгоритм навчання бере до уваги як попередні, так і поточні знання мережі з метою досягнення хорошого узагальнення. В роботі також представлені результати, які ілюструють теоретичні розвитки, а також продуктивність запропонованого підходу в реальних експериментах, які можуть бути застосовані для автоматизації досліджень. В доробку «AI in Video Analysis, Production and Streaming Delivery» [9] проаналізовані тенденцію розвитку відеотехнологій, розуміння яких необхідне для якісного аналізу інформації. А «Video content analysis using convolutional neural networks» [10] пояснює процес аналізу відеоконтенту в системах відеоспостереження, розглядається також автоматизована система аналізу відеоконтенту. Пропонується система, яка генерує текстовий файл з описом відеоконтенту, що дозволяє швидко знаходити необхідну інформацію. В роботі використовується модель згорткових нейронних мереж для виявлення та класифікації об'єктів, що з'являються на відео.

Здійснений аналіз джерел підтверджує актуальність теми UX-досліджень та їх автоматизації. Це чинна проблема, оскільки тестування і підтримка високоякісної UX-складової інтерфейсів цифрових продуктів є критичними на кожному етапі розробки інтерфейсів. У майбутньому, план автоматизації досліджень за допомогою штучного інтелекту може включати розробку спеціалізованих систем, що дозволяють здійснювати швидкий та точний аналіз великих обсягів даних.

Формулювання мети статті

Метою статті є автоматизація аналізу результатів UX-досліджень з використанням технологій штучного інтелекту, зокрема нейронних мереж та експертних систем, шляхом формування системи методико-інструментальних засобів. Дослідження взаємодії є невіддільною складовою процесу проектування інтерфейсів мобільних та настільних застосунків, вебсайт, ігор, програм змішаної та доповненої реальності тощо. Традиційно, ці дослідження проводяться вручну фахівцями з дизайну, розробки, архітектури та тестування, які використовують спеціальні програми для запису сесій користування або імітації типової поведінки користувачів. Однак, великі обсяги даних, що генеруються в результаті таких досліджень, та необхідність їх ручної обробки значно уповільнюють процес проектування інтерфейсів та вимагають залучення значних людських ресурсів. Зосередившись на дослідженнях, що аналізують вже створені інтерфейси, можна виявити можливості для їх автоматизації. Різні типи досліджень можуть бути доповнені технологіями штучного інтелекту, зокрема нейронними мережами та експертними системами, для аналізу даних та їх підготовки до прийняття рішень. Це дозволить фахівцям швидше та ефективніше опрацювати вже підготовлені результати досліджень та зосередитись на розробці концептуального. Отже, тема роботи, в якій розглядаються та вирішуються наукові питання, пов'язані з автоматизацією аналізу результатів досліджень є актуальною та потрібною. Для досягнення поставленої мети пропонується вирішення наступних завдань:

1. Дослідження чинних підходів до автоматизації проведення UX-досліджень;
2. Опис та систематизація методів проведення досліджень, придатних для автоматизації;
3. Аналіз технологій штучного інтелекту, зокрема нейронних мереж та експертних систем, які можуть бути застосовані для обробки результатів;
4. Формування схеми обробки результатів з використанням відповідних нейронних мереж та експертних систем;
5. Аналіз невирішених завдань та перспектив розвитку запропонованих методів.

Запропонована система методично-інструментальних засобів автоматизації аналізу результатів досліджень ґрунтується на поєднанні технологій штучного інтелекту, зокрема нейронних мереж та експертних систем, з наявними методами проведення досліджень.

Виклад основного матеріалу дослідження

Дослідження взаємодії кінцевого користувача з інтерфейсом цифрового продукту — це структура різноманітних наукових підходів, кожен з яких використовує свої методики та стратегії. Запропонований підхід до класифікації та стандартизації [10] є лише одним з багатьох можливих варіантів, і в ході подальших досліджень він може бути скоригований або доповнений з урахуванням нових знань та практичного досвіду в цій галузі. Оскільки ми будемо підбирати нейромережі на основі даних досліджень, потрібно обрати класифікацію за поділом за типами даних (рис. 1), які отримуються в результаті проведених досліджень. Це можуть бути текстові дані, числові або кількісні дані, фізичні артефакти, графічні дані, аудіодані, навігаційні та поведінкові дані, а також мультимедійні дані. Результати досліджень, як правило, представлені у вигляді великих обсягів різноманітних даних. Це можуть бути текстові транскрипти інтерв'ю (письмова інтерпретація усної розмови) чи коментарів користувачів, аудіо- та відеозаписи сесій тестування, теплові карти натиснення, дані відстежування очей, метрики продуктивності та інші типи інформації. Запропонований метод опрацювання результатів досліджень дозволяє автоматизувати та прискорити процес аналізу великих обсягів даних, що, своєю чергою, сприятиме більш ефективному прийняттю рішень щодо вдосконалення користувацького досвіду цифрових продуктів та інтерфейсів. З іншого боку, для реалізації автоматизованого аналізу результатів UX-досліджень пропонується застосування нейронних мереж. Для опрацювання різноманітних даних необхідно буде залучити кілька видів нейронних мереж, оскільки жодна окрема модель не зможе ефективно впоратися з усіма форматами вхідної інформації. Розглянемо приклад однієї з таких нейронних мереж, яка може бути використана в рамках даного дослідження.

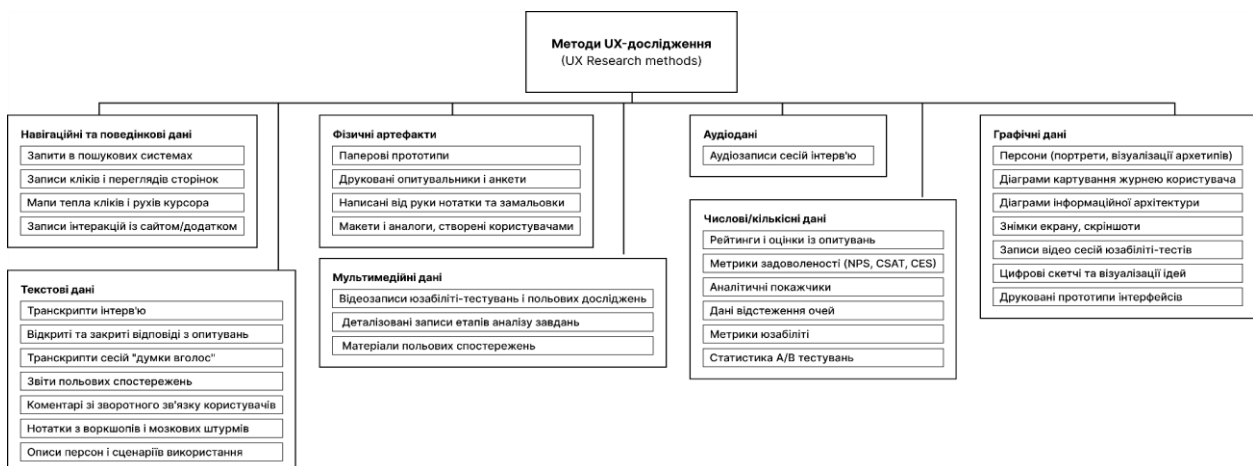


Рис. 1. Класифікація методів UX-досліджень

Повторювана або ж рекурентна нейронна мережа (RNN) [11] – це тип штучної нейронної мережі, яка використовує послідовні дані або дані часових рядів. Ці алгоритми зазвичай використовуються для перекладу, обробки природної мови, розпізнавання мовлення тощо. Ці нейромережі інтегровані в такі популярні програми, як Siri та Google Translate. Вони відрізняються своєю «пам'яттю», оскільки вони беруть інформацію з попередніх вхідних даних, щоб впливати на поточні вхідні та вихідні дані (рис. 2).



Рис. 2. Схема роботи рекурентних нейромереж

RNN можуть бути навчені переводити аудіозаписи в текст, тобто розпізнавати мовлення. А також можуть бути навчені визначати емоційний тон мовлення в аудіозаписах. Це дозволяє розуміти емоційний стан особи, що може бути корисним для подальшого



Рис. 3. Схема роботи згорткових нейромереж

Наприклад, при аналізі відеозаписів користувацьких сесій можуть виявляти елементи інтерфейсу: кнопки, меню, форми тощо, положення курсора та області взаємодії користувача [14]. Можуть бути навчені виконувати задачі класифікації та сегментації на основі виділених просторових ознак (класифікація дій користувача: визначення типу взаємодії, виконання семантичної сегментації інтерфейсу, розпізнавання та виділення різних елементів інтерфейсу) [15]. Розглянуті приклади застосування рекурентних нейронних мереж та згорткових нейронних мереж для аналізу даних UX-досліджень демонстру-

ють широкі можливості автоматизації обробки різноманітної інформації — від текстових транскриптів інтерв'ю до відеозаписів користувацьких сесій. Зазначені типи нейромереж є лише прикладами серед безлічі можливих архітектур та їх комбінацій. Для комплексної автоматизації аналізу даних можуть знадобитися гібридні підходи, які поєднують різні типи нейромереж. Завдяки поєднанню різних видів нейронних мереж з'явиться можливість створити систему (рис. 4) для автоматизованого аналізу різнотипних даних, що отримані в результаті проведення досліджень користувацького досвіду.

Окрім рекурентних нейронних мереж, існують й інші типи нейромереж, що можуть бути використані для аналізу даних досліджень. До прикладу, згорткові нейронні мережі (CNN) [13]. Такі мережі особливо ефективні для обробки візуальних даних, таких як зображення та відео; можуть автоматично виділяти просторові ознаки з відеокadrів. Ці мережі використовують операцію згортки (рис. 3) для автоматичного виділення просторових ознак та шаблонів з вхідних даних.

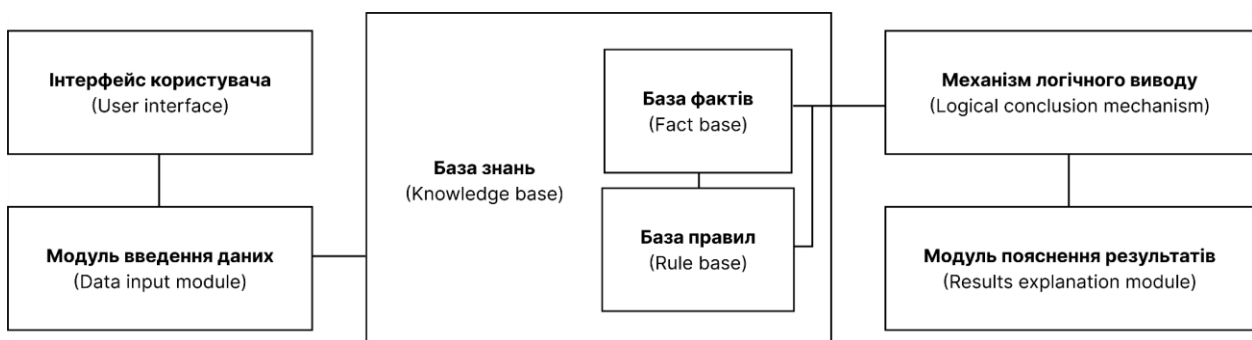


Рис. 4. Структура експертної системи

Крім того, ефективність роботи нейронної мережі залежить не лише від вибору її правильної архітектури, але й від низки інших факторів, зокрема гіперпараметрів. Моделі глибокого навчання характеризуються їх значною кількістю, таких як кількість шарів,

кількість нейронів, швидкість навчання, функції активації та оптимізатори. Наявність великої кількості можливих комбінацій цих параметрів суттєво ускладнює процес навчання нейронних мереж, а отже, і вибір оптимальної архітектури та її налаштування для кон-

кретної задачі. З огляду на це, впровадження експертної системи є необхідним кроком для автоматизації процесу вибору та конфігурації нейромережі, що дозволить спростити та прискорити розробку ефективних моделей для аналізу даних UX-досліджень. Саме тому, в рамках запропонованої системи аналізу, передбачається використання експертної системи, яка, на основі характеристик вхідних даних та цілей аналізу, зможе рекомендувати найбільш придатні архітектури нейромереж та їх гіперпараметри. Інтеграція різних типів нейромереж в рамках гнучкої та адаптивної експертної системи є перспективним напрямком для подальших досліджень та практичних реалізацій у сфері аналітики користувальницького досвіду. Раніше значалося, що наразі у сфері оцінки зручності більша частина рішень приймається експертом і навіть часткова автоматизація неможлива без створення систем, які будуть замінювати його. Пропонується розробити таку експертну систему для прийняття рішень, щодо відповідної нейромережі для обробки даних досліджень. Такі системи є одним із напрямків штучного інтелекту, що базуються на знаннях. Система містить базу знань, що акумулює досвід експертів, та механізм логічного виведення, який використовує ці знання.

База знань цієї системи складається з двох компонентів: бази фактів та бази правил. Механізм виведення обробляє факти та правила для генерації рекомендацій або рішень. Така експертна система містить модуль введення даних для забезпечення характеристик вхідних даних та цілей аналізу і механізм логічного виведення для здійснення вибору нейромереж на основі бази знань та вхідних даних. При необхідності інтерпретації результатів можуть бути обрані нейромережі з вбудованими механізмами пояснення. У контексті аналізу результатів UX-досліджень експертна система використовується для вибору найбільш придатних нейронних мереж залежно від типу вхідних даних та цілей аналізу. Така експертна система має містити базу знань, що акумулює експертний досвід у сфері машинного навчання та аналітики. База фактів може включати характеристики вхідних даних досліджень (формат, обсяг, структура) та цілі аналізу (класифікація, кластеризація, прогнозування тощо). База правил має містити евристичні правила, що пов'язують типи даних, цілі аналізу та архітектури нейронних мереж. Механізм логічного виведення, використовуючи базу знань, обиратиме найбільш придатні нейромережі для обробки конкретного набору даних досліджень та досягнення поставлених аналітичних цілей.

Для формування вимог до побудови та сучасний стан експертних систем був проведений аналіз наявних експертних систем та інструментів. До прикладу, існує інструмент Neural Network Console [16]. Це — графічний інструмент для проектування, навчання та розгортання нейронних мереж. Він надає зручний drag-and-drop інтерфейс для візуальної побудови архітектури нейромереж, включаючи шари та зв'язки між ними. Neural Network Console містить вбудовану базу знань з типових архітектур нейромереж для різних задач (класифікація зображень, обробка природної мови тощо) та дозволяє автоматично генерувати

код моделі на основі візуальної схеми. Він може бути залучений в проектуванні експертної системи для вибору інших нейронних мереж для автоматизації обробки результатів. Також існує Neural Architect [17] — вебплатформа для візуального проектування, навчання та розгортання нейронних мереж. Вона надає інтуїтивно зрозумілий інтерфейс для побудови архітектури нейромереж з використанням готових блоків. Містить вбудовану базу знань з типових архітектур для різних задач та дозволяє автоматично генерувати код моделі. Окрім готових інструментів, можливе створення спеціалізованих експертних систем на основі правил для вибору нейромережових архітектур в контексті UX-досліджень. Такі системи містять базу знань з евристичними правилами, які пов'язують характеристики даних та цілі аналізу з відповідними типами нейромереж. Ці системи можуть бути реалізовані з використанням мов програмування загального призначення (Python, Java, Kotlin) або спеціалізованих мов для створення експертних систем (CLIPS, Prolog, Lisp, Jess, Drools, RETE, PyKnow).

Аналіз інструментів показує, що вже доступні рішення для візуального проектування експертних систем вибору нейромережових архітектур, які спрощують процес розробки та адаптації моделей. Однак, попри наявність цих інструментів, не існує готового рішення, що повністю відповідає специфіці автоматизації аналізу нейромереж та даних UX-досліджень. Ці системи можуть бути використані як база, але потребують адаптації та доповнення. Незалежно від обраного підходу, використання експертних систем для автоматизованого вибору нейромережових архітектур дозволяє зменшити вплив людського фактора, прискорити процес розробки та підвищити ефективність аналізу.

Аналіз інструментів показує, що існують готові рішення для візуального проектування експертних систем. Однак перед використанням будь-якого з них, для підтвердження життєздатності запропонованої концепції необхідно провести експериментальне дослідження. Це було вирішено провести шляхом створення тестової версії системи. Ця версія, хоча й не матиме повної функціональності експертної системи, що планується реалізувати в подальших наукових працях, проте дозволить перевірити базові заплановані принципи роботи, а саме: можливість завантаження файлів UX-досліджень різних типів та їх обробку за допомогою моделей штучного інтелекту. Для реалізації тестової версії системи необхідно вирішити наступні завдання:

1. Визначити оптимальну модель штучного інтелекту для тестування, що забезпечить обробку різнотипних даних UX-досліджень;
2. Обрати архітектуру та метод створення тестової системи;
3. Підібрати відповідні технології, мову програмування та інструментальні засоби для реалізації користувацького інтерфейсу.

В рамках першого завдання було проведено порівняльний аналіз способів інтеграції моделей штучного інтелекту у систему, що розробляється. Розглянуто наступні підходи до впровадження ШІ-систем:

API-інтеграція, локальне розгортання open-source моделей, використання ML-фреймворків та використання хмарних сервісів.

API-інтеграція через сервіси, такі як OpenAI API чи Claude API забезпечує доступ до моделей через HTTP-запити. Перевагами є миттєвий доступ до технологій, відсутність витрат на обладнання та постійні оновлення. Обмеженнями виступають необхідність абонентської плати та залежність від інтернет-з'єднання. Локальне розгортання open-source моделей надає повний контроль над системою, проте вимагає значних інвестицій в обладнання, глибоких технічних компетенцій та постійного обслуговування інфраструктури. Використання ML-фреймворків для

створення власних моделей доцільне лише для вузькоспеціалізованих задач, оскільки потребує значних ресурсів на розробку, тестування та навчання моделей. Для реалізації тестової версії системи обрано API-інтеграції як оптимальний з погляду співвідношення витрат, швидкості впровадження та гнучкості розробки. Він дозволяє зосередитись на розробці функціоналу системи та її тестуванні замість обслуговування інфраструктури штучного інтелекту. Проаналізувавши доступні API-інтеграції моделей штучного інтелекту, було зосереджено увагу на двох провідних рішеннях: OpenAI API та Claude API. У табл. 1 та 2 наведено їх порівняльний аналіз з погляду технічних можливостей та комерційних умов.

Таблиця 1 – Технічні можливості API-моделей ШІ

Параметр	Claude API	OpenAI API
Максимальний розмір контексту	200K токенів	128K токенів
Підтримувані формати файлів	PDF, DOCX, XLSX, CSV, JSON, TXT	PDF, DOCX, JSON
Швидкість обробки запитів	До 3500 токенів/с	До 2000 токенів/с
Точність обробки природної мови	Висока	Висока
Багатомовна підтримка	100+ мов	95+ мов
Вбудовані аналітичні інструменти	Розширені	Базові
Обробка структурованих даних	Вбудована	Потребує додаткової розробки

Таблиця 2 – Умови використання API-моделей ШІ

Параметр	Claude API	OpenAI API
Пробний період	\$5 кредитів	Відсутній
Базовий тариф	Pay-as-you-go	Підписка
Вартість токена	Від \$0.00008	Від \$0.0002
Ліміти запитів	Гнучкі	Фіксовані

Як видно з наведених даних, Claude API демонструє кращі показники за технічними параметрами та має більш гнучкі комерційні умови використання. Особливо важливим фактором є наявність пробного періоду з кредитами на суму \$5. Доступних ресурсів пробного періоду достатньо для створення, тестування та налагодження вебінтерфейсу, а також проведення початкового аналізу даних UX-досліджень, що дозволить підтвердити життєздатність запропонованої концепції.

Наступним етапом після вибору API-інтеграції стало проектування архітектури системи для ефектвної взаємодії з API штучного інтелекту та можливість обробки різних типів даних UX-досліджень. В процесі аналізу було розглянуто основні архітектурні підходи.

Монолітна архітектура, хоча й забезпечує простоту розробки, має обмеження щодо масштабування та ускладнює процес інтеграції з API. Гібридне рішення з серверним рендерингом потребує додаткових ресурсів. Мікросервісна архітектура виявилася надмірною для тестової версії системи через складність реалізації та обслуговування. Тож для реалізації

тестової версії системи обрано класичний вебінтерфейс з розділенням на частини frontend та backend. Такий вибір обґрунтований кількома факторами: backend-частина ефективно керує взаємодією з Claude API та обробкою даних UX-досліджень. frontend — забезпечує інтерфейс для завантаження та перегляду результатів аналізу.

Додатковою перевагою є можливість незалежного масштабування компонентів системи при збільшенні навантаження.

При подальшому виборі технологічного стека проаналізовано сучасні фреймворки та мови програмування. На основі вимог до швидкості розробки та простоти інтеграції з API обрано мінімалістичний стек: HTML/CSS з JavaScript та AJAX на frontend та Python (FastAPI) на backend. Вибір технологічного стека зумовлений потребою у швидкій розробці та розгортанні системи, а також простотою інтеграції з Claude API. На основі проведеного комплексного дослідження методів UX-досліджень, аналізу технологій штучного інтелекту та вивчення підходів до автоматизації було розроблено тестову систему автоматизованого аналізу даних UX-досліджень (рис. 5).



Рис. 5. Стартовий вигляд інтерфейсу тестової системи при завантаженні файлу

Система інтегрує результати теоретичного аналізу, поєднуючи можливості нейронних мереж для обробки різних типів даних та вебтехнології для забезпечення взаємодії з користувачем.

В рамках практичної реалізації створено версію системи у вигляді вебінтерфейсу з інтеграцією Claude API.

Розроблена система втілює основні теоретичні концепції дослідження та включає такі функціональні можливості:

1. Завантаження файлів до інтерфейсу у форматах DOCX, XLSX та PDF;
2. Налаштування параметрів аналізу через уточнювальне поле запитів;
3. Візуалізація процесу обробки з індикатором прогресу;
4. Валідація та відображення вмісту завантажених файлів;
5. Структуроване представлення результатів аналізу даних UX-дослідження та попередніх висновків відповідно до запитів.

Інтерфейс системи відображає проміжні результати на кожному етапі обробки даних. При завантаженні файлу користувач бачить його вміст у структурованому вигляді:

для текстових документів — з часовими мітками та розбивкою на сегменти,

для таблиць — з коректним форматуванням рядків та стовпців.

Система автоматично виявляє проблеми зі структурою даних, такі як пошкоджені таблиці чи некоректне форматування, та формує технічні рекомендації щодо їх виправлення. Такий підхід дозволяє на тестовому етапі враховувати лише перевірені набори даних.

Для тестування системи було додатково проаналізовано можливі підходи до формування наборів даних. Розглянуто відкриті набори даних UX-досліджень (UX Researches Datasets), зокрема колекції даних на платформі Kaggle та GitHub, що містять результати масштабних користувацьких досліджень. Однак, враховуючи особливості роботи API, де обробка даних тарифікується за обсягом токенів, і

необхідність валідації результатів обробки, було прийнято рішення використовувати власні набори даних меншого обсягу з реальних UX-досліджень. На цьому етапі розробки система працює з конкретними запитом (prompts) для обробки даних, що дозволяє точно налаштувати параметри аналізу для кожного типу документів. Надалі планується розробка багатокрокового процесу автоматичного формування та уточнення запитів на основі характеристик вхідних даних.

Для валідації роботи системи було підготовлено набір тестових документів, що представляють різні типи даних UX-досліджень:

1. Таблиці результатів масових опитувань перевірки ставлення до пропонованих системою навчальних форматів на сторінці кабінету користувача;
2. Транскрипти глибоких інтерв'ю про користувацький досвід використання функції проведення онлайн-практикумів;
3. Текстові документи з аналізом змісту навчального матеріалу для подальшої обробки.

Розглянемо процес обробки та отримані результати. Перший сценарій передбачав аналіз навчального відео з метою виділення ключових тез та створення підсумку.

На першій частині рис. 6 представлено інтерфейс системи при роботі з текстовим документом, що містить транскрипт з часовими мітками. Як видно, система успішно опрацювала структурований текст, зберігаючи зв'язок з часовими мітками, та сформулила змістовний аналіз у вигляді основних тез доповіді.

Другий сценарій стосувався аналізу кількісних даних опитування студентів щодо навчальних форматів.

На другій частині рис. 6 показано роботу системи з табличними даними. При роботі з табличними даними система продемонструвала здатність виявляти проблеми з форматуванням та надавати рекомендації щодо покращення якості вхідних даних. Попри часткові обмеження вхідних даних, система змогла виділити ключові тенденції та сформулювати структуровані висновки щодо переваг користувачів.

0:06:09.800,0:06:13.840
 вакансії тут навіть знову ж таки
 говорючи про те що ви там свічер
 0:06:13.840,0:06:19.240
 наприклад да тобто преходите а однієї
 професії в іншу і в вас звісно немає ще
 0:06:19.240,0:06:24.240
 комерційного досвіду не варто
 засмучуватись тобто тут у вас вже є
 0:06:24.240,0:06:30.240
 пройдений курс у вас вже є а якісь
 напрацьовані навички практич
 0:06:30.240,0:06:36.800

Результат обробки:

Короткий зміст:
 Урок присвячений темі мотиваційного листа від досвідченої рекрутерки Ірини, яка має понад 5 років досвіду в рекрутменті. Вона пояснює суть, призначення та важливість мотиваційного листа при пошуку роботи.

3 найцікавіших думки:

1. Мотиваційний лист – це короткий документ обсягом 200–250 слів, який доповнює резюме та є способом самопрезентації.
2. На відміну від резюме, яке містить структуровану інформацію про освіту та досвід роботи, мотиваційний лист дозволяє розповісти про особисту мотивацію та зацікавленість у конкретній вакансії.
3. Не всі компанії вимагають мотиваційний лист, але в деяких він може бути обов'язковим або суттєво підвищити шанси кандидата на працевлаштування.

2024-06-07 13:57:08.228000	Так, звісно!	5
2024-06-07 13:58:10.136000	Так, звісно!	5
2024-06-07 13:58:35.547000	Так, звісно!	5
2024-06-07 14:00:26.383000	Так, звісно!	5
2024-06-07 14:06:42.761000	Так, звісно!	5
2024-06-07 14:06:46.214000	Так, звісно!	5
2024-06-07 14:09:14.847000	Так, звісно!	5
2024-06-07 14:21:57.472000	Так, звісно!	5

Результат обробки:

[TextBlock(text='На жаль, наданий зміст виглядає як частково пошкоджена або некоректно сформована таблиця, що ускладнює її повноцінний аналіз. Однак, базуючись на заголовках та частково видимих даних, можна зробити наступні попередні висновки:
 Ключові спостереження:
 1. Опитування стосується студентів онлайн-школи дизайну та креативних професій
 2. Дослідження охоплює кілька важливих аспектів:
 Інтерес до різних форматів навчальних подій
 Зручний час для відвідування заходів
 Оцінка професійних напрямків
 Цікавість до інструментальних майстер-класів
 3. Методологія опитування передбачає:
 Шкалу оцінювання від 1 до 5
 Шкалу оцінювання від 1 до 10
 Можливість множинного вибору
 Для більш детального та точного аналізу потрібна повна, коректно сформована таблиця даних.
 Рекомендація: Перевірте цілісність та коректність вихідного файлу перед подальшим аналізом.', type='text')]

Рис. 6. Знімки з екрана зі змістом та результатом обробки даних UX-досліджень

Результати тестування підтвердили працездатність обраних технічних рішень та дозволили виявити особливості роботи системи з різними типами даних. Зокрема, при обробці текстових документів система ефективно працює з форматованим текстом, зберігаючи структуру документа та метадані, такі як часові мітки. При роботі з табличними даними реалізовано механізми валідації, що дозволяють виявляти потенційні проблеми з форматуванням та структурою даних ще до початку аналізу. Водночас тестування виявило напрямки для подальшого вдосконалення системи:

1. Автоматизація формування запитів на основі типу та структури вхідних даних;
2. Розширення підтримки форматів файлів для різних типів UX-досліджень;

3. Впровадження додаткових механізмів валідації та попередньої обробки даних.

Оптимізація обробки великих даних;

Особливістю розробленої системи є її гнучкість щодо типів аналізу даних UX-досліджень. Система може бути застосована для різних завдань: від простого узагальнення текстових даних до складного аналізу результатів користувацьких опитувань. Хоча поточна версія використовує базову модель штучного інтелекту, що може потребувати додаткових ітерацій для оптимізації результатів, вона демонструє принципову можливість автоматизованої обробки даних UX-досліджень. На основі проведеного дослідження та результатів тестування системи була сформована загальна концепція роботи системи методико-інструментальних засобів автоматизованого аналізу даних

UX-досліджень. Схема роботи системи, представлена на рис. 7 й описує концепцію, що полягає у застосуванні чинних методів UX-досліджень, типів даних та наявних на сьогодні нейронних мереж, а також їх інтеграції в єдину систему. Ключовою ідеєю є використання необхідних нейромереж залежно від типів даних для забезпечення ефективного аналізу та обробки цих даних. Схема демонструє, що нейромережі можуть бути залучені для аналізу певних типів даних, зібраних під час конкретних методів досліджень. Таким чином, запропонований загальний

шлях автоматизованого аналізу із застосуванням відповідних нейромереж. Існує ряд невідомих елементів, які потребують уточнення та додаткових досліджень. З одного боку, це стосується методів UX-досліджень, які на цей час не узгоджені, не класифіковані та не мають стандартизованого формату представлення результатів.

В статті пропонується лише один із можливих способів класифікації методів UX-досліджень та визначення типів даних, що генеруються в результаті їх проведення.

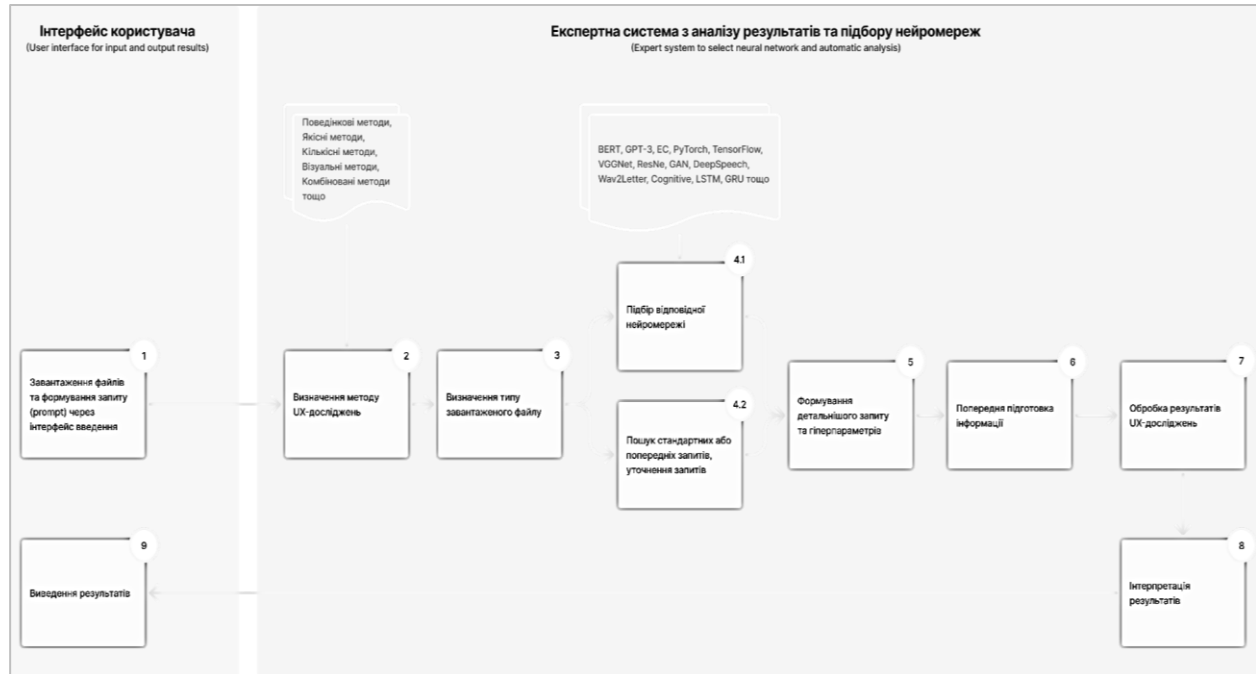


Рис. 7. Схема роботи системи методико-інструментальних засобів автоматизованого аналізу UX-досліджень

Така систематизація є необхідною для подальшої інтеграції цих методів з відповідними нейромережами в єдину автоматизовану систему.

Подальші дослідження будуть спрямовані на аналіз чинних та розробку нових версій експертної системи. Наступними кроками є деталізація загального плану побудови автоматизованої системи аналізу UX-даних та опис практичної реалізації такої системи з використанням обраних нейромережових технологій. Для цього планується проведення дослідження нейромереж та формування та базу знань з описом їх архітектур та гіперпараметрів. Окрім нейронних мереж, до системи можуть бути додані експертні компоненти для поглибленого аналізу та інтерпретації результатів. Лише за умови ретельної розробки всіх складових елементів можна досягти поставленої мети — створення ефективної методології автоматизованого аналізу даних на основі інтелектуальних систем.

Висновки

Проведений аналіз літературних джерел виявив певні недоліки традиційних методів UX-досліджень, пов'язані з високими витратами часу та ресурсів на збір і обробку даних, що зумовлює необхідність

автоматизації цього процесу з використанням технологій штучного інтелекту;

Запропоновано концептуальну схему системи методико-інструментальних засобів автоматизованого аналізу результатів з використанням нейронних мереж та експертної системи, яка передбачає класифікацію типів досліджень за форматами даних та вибір відповідних архітектур нейромереж для їх обробки.

Запропоновано використання експертної системи для автоматизованого вибору оптимальних нейромережових моделей залежно від типу та характеристик вхідних даних досліджень;

Наведено можливості інтеграції системи, що розробляється, з сучасними платформами штучного інтелекту, що дозволяє розширити функціональність системи автоматизованого аналізу;

Отримані результати свідчать про значний потенціал розробленої системи для практичного застосування та підвищення ефективності процесу проектування інтерфейсів. Подальші дослідження можуть бути спрямовані на практичну реалізацію системи, розробку відповідного ПЗ та експериментальну оцінку ефективності автоматизованого аналізу різних типів даних.

Проведене дослідження відкриває широкі перспективи для подальшого розвитку та вдосконалення методів автоматизованого аналізу результатів UX-досліджень з використанням технологій штучного інтелекту. Реалізація завдань дослідження дозволяє суттєво удосконалити процес проектування користу-

вацьких інтерфейсів, забезпечивши його більшу швидкість, ефективність та орієнтованість на реальні потреби та очікування користувачів. Це, своєю чергою, сприятиме підвищенню якості та конкурентоспроможності цифрових продуктів та послуг в умовах стрімкого розвитку інформаційних технологій

СПИСОК ЛІТЕРАТУРИ

- Nielsen J. Usability 101: Introduction to Usability [Електронний ресурс] / J. Nielsen. – 2012. – Режим доступу: <https://www.nngroup.com/articles/usability-101-introduction-to-usability/> – Дата звернення: 01.02.2023.
- ISO 9241-11:2018(en), Ergonomics of human-system interaction - Part 11: Usability: Definitions and concepts [Електронний ресурс]. – Режим доступу: <https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:vl:en>
- Farooqui T. Impact of Human-Centered Design Process (HCDP) on software development process / T. Farooqui, T. Rana, F. Jafari // Proceedings of the 2nd International Conference on Communication, Computing and Digital systems (C-CODE 2019). – Institute of Electrical and Electronics Engineers, 2019. – P. 110-114. – Режим доступу: <https://doi.org/10.1109/C-CODE.2019.8680978>
- Tsui A.S.M., Kuzminykh A. Detect and interper: towards operationalization of automated user experience evaluation. Design User Experience and Usability. 2020. P. 82-96. URL: https://dl.acm.org/doi/10.1007/978-3-031-35702-2_6
- Robinson J. The past, present, and future of UX empirical research / J. Robinson, C. Lanus, R. Weber // Commun. Des. Q. Rev. – 2018. – Vol. 5. – P. 10-23. – Режим доступу: <https://doi.org/10.1145/3188173.3188175>
- Abrahams A. S. An integrated text analytic framework for product defect discovery / A. S. Abrahams, W. Fan, G. A. Wang, Z. J. Zhang, J. Jiao // Prod. Oper. Manag. – 2015. – Vol. 24. – P. 975-990. – Режим доступу: <https://doi.org/10.1111/poms.12303>
- Wang R. UX Calculator: An online tool to support user testing / R. Wang, S.-L. Chen, C. Labbé, M. Fredette, A. Abdessemed, F. Courtemanche, C. K. Coursaris, S. Sénécal, P.-M. Léger // [Б. м. : б. в.].
- Doulamis A. D. On-line retrainable neural networks: improving the performance of neural networks in image analysis problems / A. D. Doulamis, N. D. Doulamis, S. D. Kollias. URL: <https://ieeexplore.ieee.org/abstract/document/822517>
- Jayanthiladevi A. AI in Video Analysis, Production and Streaming Delivery [Електронний ресурс] / A. Jayanthiladevi, A. G. Raj, R. Narmadha, S. Chandran, S. Shaju, K. K. Prasad. – Режим доступу: <https://iopscience.iop.org/article/10.1088/1742-6596/1712/1/012014/meta>
- Fleury, S., & Chaniaud, N. (2023). Multi-user centered design: acceptance, user experience, user research and user testing. Behaviour & Information Technology, 42(3), 209-224. URL: <https://www.tandfonline.com/doi/full/10.1080/1463922X.2023.2166623>
- Salem, F. M. (2023). Recurrent Neural Networks: From Simple to Gated Architectures. Springer International Publishing.
- Okut, H. (2021). Deep Learning: Long-Short Term Memory. University of Kansas Medical Center. https://www.researchgate.net/publication/352383391_Deep_Learning_Long-Short_Term_Memory
- Khodambashi, S., Øvrelid, E. A., Pedersen, S. A., & Jørstad, I. (2020). User involvement and experimentation for technology adoption in hospitals: A conceptual study and research agenda. SN Applied Sciences, 2(1), 76. — Режим доступу: <https://doi.org/10.1007/s13748-019-00203-0>
- Aljarrah I. Video content analysis using convolutional neural networks [Електронний ресурс] / I. Aljarrah, D. Mohammad. – Режим доступу: <https://ieeexplore.ieee.org/abstract/document/8355453>
- Kumar A. Image Sentiment Analysis Using Convolutional Neural Network [Електронний ресурс] / A. Kumar, A. Jaiswal. – Режим доступу: https://link.springer.com/chapter/10.1007/978-3-319-76348-4_45
- Sony Corporation. (2023). Neural Network Console [Електронний ресурс]. Режим доступу: <https://dl.sony.com/>
- Microsoft Research. (2023). Archai: Platform for Neural Architecture Search [Електронний ресурс]. Режим доступу: <https://www.microsoft.com/en-us/research/project/archai-platform-for-neural-architecture-search/>

Received (Надійшла) 24.12.2024

Accepted for publication (Прийнята до друку) 26.02.2025

System of methodological and instrumental tools for automated analysis of UX research data using artificial intelligence systems

Y. Ihnatiuk, A. Popov

Abstract. The article is dedicated to the automated analysis of user experience (UX) research results using artificial intelligence systems, particularly neural networks and expert systems. The aim of the research is to develop a system of methodological and instrumental tools for automating UX research data analysis through the integration of artificial intelligence technologies and expert systems. The relevance of automating the processing of large volumes of heterogeneous data obtained from user interaction research with digital products has been analyzed. A classification of user interaction research methods and digital products by data types and an approach to selecting appropriate neural network architectures for their analysis are proposed. The possibilities of using recurrent and convolutional neural networks as examples for processing text, audio, and video data are examined. The use of an expert system for automated selection of neural networks optimal for analyzing various types of data is also proposed. A trial version of the system using artificial intelligence API has been developed and tested, confirming the possibility of automated processing of various types of UX research data using neural networks. Prospects for further development of the UX research data analysis system are outlined, including software development, functionality expansion, and adaptation for related fields. The research results have practical significance for improving the efficiency of data analysis and can be used to enhance the interface design process.

Keywords: UX research, automation, data analysis, neural networks, expert systems, web-interface.

С. І. Клівець¹, О. В. Кулешов¹, Р. В. Зубко²

¹ Харківський національний університет Повітряних Сил імені Івана Кожедуба, Харків, Україна

² Національний університет оборони України, Київ, Україна

МЕТОД ФОРМУВАННЯ ПІДСИСТЕМИ ОБРОБКИ ОПЕРАТИВНИХ ТРАНЗАКЦІЙ ІНТЕРНЕТУ РЕЧЕЙ

Анотація. **Актуальність.** Транзакції Інтернету речей зазвичай оброблюються у хмарних центрах обробки даних. Але при необхідності обробляти оперативні транзакції виникають часові затримки, що пов'язані з передачею даних у хмарне середовище. Дану проблему можна вирішити за рахунок перенесення обчислень до шарів, що наближені до датчиків IoT. **Мета статті** – розробити метод формування підсистеми обробки оперативних транзакцій Інтернету речей, орієнтований на граничний та туманний шари хмарного середовища. **Результати дослідження.** Запропонована математична модель процесу обробки оперативних транзакцій. На основі моделі розроблений метод, котрий дозволяє врахувати специфічні особливості нижніх шарів хмарного середовища підтримки Інтернету речей та обрати найбільш прийнятний варіант побудови підсистеми обробки оперативних транзакцій Інтернету речей. **Висновок.** Розроблений метод дозволяє виконати вимоги QoS для оперативних транзакцій Інтернету речей.

Ключові слова: Інтернет речей, оперативна транзакція, туманний шар, мережний вузол.

Вступ

На сьогодні парадигма Інтернету речей набула бурхливого розвитку у багатьох галузях [1–4]. Відповідно розвивається і оточуюча інфраструктура Інтернету речей [5, 6]. Зокрема, тепер пристрої граничного та туманного шарів інфраструктури можуть проводити обробку інформаційних потоків. Це надало змогу перенести обробку транзакцій Інтернету речей, які потребують швидкої відповіді, тобто оперативних транзакцій [7]. Але при цьому для виконання часових вимог QoS потрібно враховувати особливості таких мережних вузлів, які мають обмежені обчислювальні ресурси та орієнтовані на розподілену обробку [8]. **Мета роботи** – розробити метод формування підсистеми обробки оперативних транзакцій Інтернету речей, орієнтований на граничний та туманний шари хмарного середовища.

Результати досліджень

Для визначення ефективності виконання оперативних транзакцій Інтернету речей (IoT) у розподіленому туманному шарі хмарного середовища (XC) застосовують такі технічні критерії:

показники завадостійкості (відношення сигнал/шум, імовірність помилкового прийому та ін.), пропускна здатність, надійність, комбінації різних технічних параметрів.

Вони характеризують якість системи, її технічний рівень. Однак, використовуючи тільки технічні показники, ми тільки локально оцінюємо ефективність процесу проходження транзакції, не враховуючи особливості мережних вузлів як граничного, так і туманного шарів, які, зазвичай, мають досить обмежені обчислювальні ресурси. Тому при формуванні підсистеми обробки оперативних транзакцій варто враховувати й наявні ресурси мережних вузлів нижніх шарів.

Виходячи з цього, сформулюємо принципи мінімуму ресурсних витрат і максимуму ефекту, на основі яких можуть формуватися критерії ефективності

виконання оперативних транзакцій IoT [9]. Принцип мінімуму витрат використовують у тих випадках, коли при заданому ефекті потрібно визначити варіант $z \in Z$, при якому забезпечуються мінімально можливі ресурсні витрати. Якщо необхідно досягти цілі при виконанні існуючих обмежень на ресурсні витрати обчислювальних вузлів, є прийнятним другий принцип. Висока швидкість обробки інформації досягається за рахунок застосування розподілених обчислень, що є характерним для нижніх шарів хмарного середовища підтримки IoT. Використання в такому випадку тільки першої вирішальної схеми пов'язано із значними ресурсними витратами, необхідними для забезпечення необхідного значення обраних показників завадостійкості. Тому цю задачу краще вирішити, застосувавши другу вирішальну схему, що гарантує необхідні показники якості при виконанні низки обмежень [10].

Методологічною основою рішення задач оптимізації підсистеми обробки оперативних транзакцій (ПООТ) є теорія дослідження операції. Опишемо ПООТ і представимо задачу її оптимізації в узагальненому вигляді [11]:

$$\text{extr } v(X) = (x_1, x_2, \dots, x_n); \quad (1)$$

$$\phi_{x_k}(X) = 0, \phi_{x_l}(X) \leq 0; \quad (2)$$

$$\phi_{y_k}(X) = 0, \phi_{y_i}(X) \leq 0; \quad (3)$$

Сформулюємо задачу: знайти вектор внутрішніх параметрів X_{opt} , що задовольняє рівностям і нерівностям (2), (3) і доставляючим мінімум цільової функції $v(X)$.

Якщо на змінні накладені обмеження типу функціональних нерівностей, точку екстремального значення функції знаходимо методом їхнього вилучення або методом множників Лангранжа, коли перемінні неможливо або важко представити в вигляді

$$\text{extr } v(X), X = (x_1, x_2, \dots, x_n); \quad (4)$$

$$\phi_i(X) = 0, i = 1, 2, \dots, m; m < n. \quad (5)$$

При виборі параметрів для оцінки ефективності ПООТ відповідно до системного підходу в загальному випадку варто було б враховувати усі види витрат, що мають місце при обробці інформації. Так як проектується оптимальна підсистема, що забезпечує потрібний час та необхідну завадостійкість P_e^{nomp} при мінімальних витратах C_{nomp} , показником порівняльної ефективності всіх допустимих підсистем буде час отримання відповіді на транзакцію $T1$. Очевидно, у цільову функцію повинен входити тільки цей параметр, оскільки у будь-якому разі для оперативної транзакції бажано його зменшувати. У результаті маємо таке завдання вибору оптимальної підсистеми в математичному вигляді:

задані множини векторів внутрішніх і зовнішніх параметрів плануємої підсистеми:

$$X = (x_1, x_2); Y = (y_1, y_2),$$

компоненти яких $x_1 = m$, де m – кількість незадіяних обчислювальних вузлів при розпаралелюванні завдання; $m = 2, 3, 4, \dots, M$; x_2 – ймовірнісний параметр; y_1 – ресурсний параметр, y_2 – параметр, що характеризує завадостійкість підсистеми.

Сформуємо компоненти за допомогою системи таких обмежень:

$$\Phi_x \rightarrow \begin{cases} 2 < x_1 < M; \\ 0 < x_2 < 1; \end{cases}$$

$$\Phi_y \rightarrow \begin{cases} y_1 \leq C_{nomp}; \\ y_2 \leq P_e^{nomp}, \end{cases}$$

де M – максимально можлива кількість мережних вузлів нижніх шарів хмарного середовища підтримки IoT, які можуть бути задіяні для функціонування підсистеми обробки оперативних транзакцій.

Розглянемо рівняння зв'язку зовнішніх і внутрішніх параметрів:

$$y_1 = A_1 (\bar{h}_i)^\alpha + A_2 \gamma_j(m); i = 1, 2; j = 1, 2, 3; \quad (6)$$

$$y_2 = \beta_k(m) P_0^m; k = 1, 2; \quad (7)$$

$$C_1 = A_1 (\bar{h}_i)^\alpha = d P^\alpha = D \left(\frac{(4\pi R)^2}{G_1 G_2 \lambda^2 F^2} h_i P_{mp} \right)^\alpha, \quad (8)$$

де d , α – постійні коефіцієнти, що характеризують тип мережних вузлів, P_{mp} – узагальнена характеристика оперативної транзакції, \bar{h}_i – середнє значення відношення сигнал-перешкода [12].

У загальному вигляді визначимо завадостійкість ПООТ, використовувачи формулу для середньої ймовірності помилки при n -кратному розпаралелюванні завдання [12]:

$$P_0(\bar{h})_n = \left(n! / 2(1 + \bar{h}/2)(2 + \bar{h}/2) \dots (n + \bar{h}/2) \right).$$

При $n = 1$, тобто при задіянні тільки двох вузлів ПООТ отримаємо:

$$P_0 = 1/(\bar{h} + 2); \bar{h}_1 \cong 1/P_0;$$

при $n = 2$;

$$P_0 = 1/(1 + \bar{h}/2)(2 + \bar{h}/2); h_2 \cong 2/\sqrt{P_0}.$$

Підставивши у вираз (8) значення для \bar{h}_i отримаємо при $\alpha = 1$ такі значення:

$$C_1 = A_1 (1/P_0)^\alpha;$$

$$C_1' = A_1 (2\sqrt{P_0})^\alpha.$$

У той же час

$$C_2 = A_2 \gamma_j(m),$$

де A_1, A_2 – нормовані коефіцієнти, а $\gamma_j(m)$ – функція, що визначає складність другої вирішальної схеми,

$$\gamma_1(m) = 2(m - 1);$$

$$\gamma_2(m) = \log_2(m + 1);$$

$$\gamma_3(m) = 2 + \log_2(m + 1);$$

$$\beta_1(m) = c \frac{m}{2^{m-1}} \geq$$

$$\geq 1/2\sqrt{2} \left(m^{-m-0,5} (1/2m-1)^{-m} \times \right. \\ \left. \times ((m-1)/(2m-1))^{-m+0,5} \right);$$

$$\beta_2(m) = 3^{m-2} + 1.$$

Введемо дані рівняння в цільову функцію й обмеження і запишемо задачу оптимального математичного програмування з нелінійним функціоналом, одним нелінійним обмеженням у виді рівності і двома обмеженнями у виді нерівностей:

$$y_1 = A_1 (\bar{h}_i)^\alpha + A_2 \gamma_j(m) \rightarrow \min; \quad (9)$$

$$\beta_k(M) P_0^m \leq P_e^{nomp}; \quad (10)$$

$$2 \leq m \leq M; \quad (11)$$

$$0 < P_0 < 1. \quad (12)$$

Таким чином, розглянута задача зводиться до знаходження значень змінних m і P_0 , що задовольняють рівнянням (10) – (12) і мінімізують функціонал (9).

Пропонується вирішення задачі класичним методом множників Лагранжа.

Підставляємо функцію

$$\varphi(X) = \left(\beta_k(m) P_0^m - P_e^{nomp} = 0 \right)$$

з невідомим множником λ у цільову функцію трьох змінних x_1, x_2, λ :

$$v(X, \lambda) = v(X) + \lambda \varphi(X)$$

одержуємо задачу оптимального математичного програмування без обмежень

$$\begin{aligned} & \min_{X, \lambda} v(X, \lambda) = \\ & = \min_{X, \lambda} [A_1 \bar{h}_i + A_2 \gamma_j(m) + \lambda \varphi(X)], \end{aligned} \quad (13)$$

Варіанти рішення цієї системи для різних значень $\bar{h}_i, \gamma_j(m), \beta_k(m)$ зведені в табл. 1, де

$$\Psi(m) = \frac{m^2 - m + 0,5}{m(m-1)(2m-1)};$$

$$K(m) = \frac{3^{m-2} \ln 3}{3^{m-2} + 1}.$$

звідки оптимальні значення незалежних перемінних знаходимо рішенням системи з трьох рівнянь:

$$\frac{\partial [v(X, \lambda)]}{\partial x_1} = 0;$$

$$\frac{\partial [v(X, \lambda)]}{\partial x_2} = 0;$$

$$\varphi(X) = 0.$$

Шукане оптимальне рішення X_{opt} одержуємо шляхом ітераційного процесу. Підставляючи у формулу (7) значення $m_{opt-i}, P_{o.opt-i}$, знаходимо мінімальне значення цільової функції.

На рис. 1 показана залежність цільової функції від P_0 при значеннях m , що задовольняють умові (7).

Таблиця 1 – Варіанти рішень оптимізаційної задачі (9) – (12)

\bar{h}_i	γ_j	$\beta_k(m)$	Система рівнянь для X_{opt}
1/P ₀	2(m-1)	C _{2m-1} ^m	ln P ₀ = -[Ψ(m) + (2A ₂ mP ₀)/A ₁]; C _{2m-1} ^m P ₀ ^m = 10 ⁻¹⁰
	log ₂ (m+1)	C _{2m-1} ^m	ln P ₀ = -[Ψ(m) + (A ₂ mP ₀)/A ₁ (m+1)ln 2]; C _{2m-1} ^m P ₀ ^m = 10 ⁻¹⁰
	2 + log ₂ (m+1)	3 ^{m-2} + 1	ln P ₀ = -[K(m) + (A ₂ mP ₀)/A ₁ (m+1)ln 2]; (3 ^{m-2} + 1)P ₀ ^m = 10 ⁻¹⁰
2/√P ₀	2(m-1)	C _{2m-1} ^m	ln P ₀ = -[Ψ(m) + (2A ₂ mP ₀ ^{1/2})/A ₁]; C _{2m-1} ^m P ₀ ^m = 10 ⁻¹⁰
	log ₂ (m+1)	C _{2m-1} ^m	ln P ₀ = -[Ψ(m) + (A ₂ mP ₀ ^{1/2})/A ₁ (m+1)ln 2]; C _{2m-1} ^m P ₀ ^m = 10 ⁻¹⁰
	2 + log ₂ (m+1)	3 ^{m-2} + 1	ln P ₀ = -[K(m) + (A ₂ mP ₀ ^{1/2})/A ₁ (m+1)ln 2]; (3 ^{m-2} + 1)P ₀ ^m = 10 ⁻¹⁰

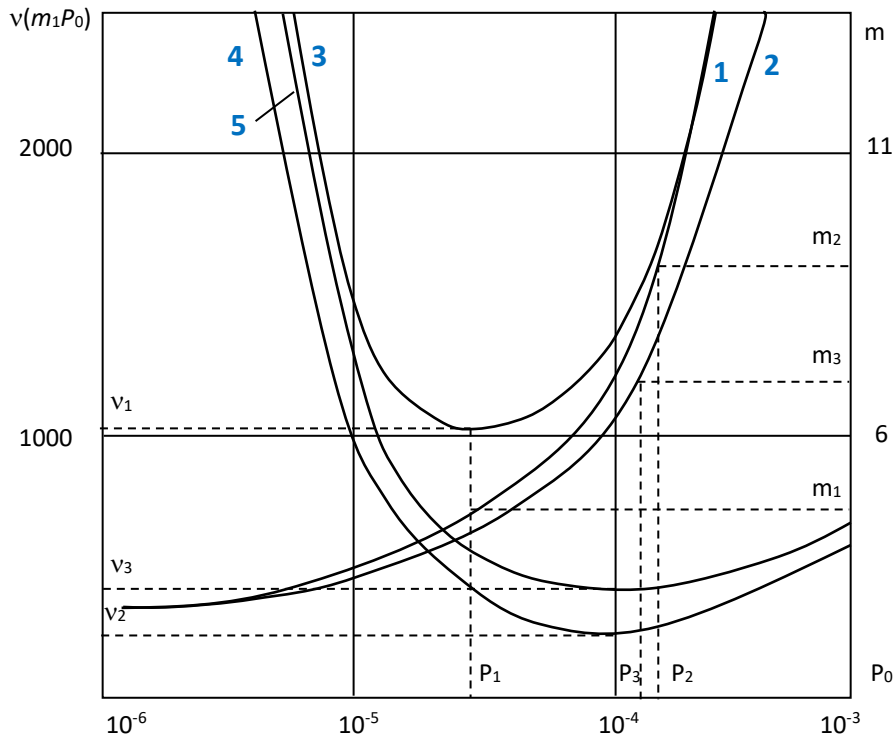


Рис. 1. Залежність цільової функції від P₀ при різних значеннях m

Аналітичні записи побудованих кривих зведені в табл. 2.

Таблиця 2 – Аналітичні записи кривих на рис. 1

Функція	Аналітичний запис	№ кривої
Y_2	$C_{2m-1}^m P_o^m = 10^{-10}$	1
Y_2	$(3^{m-2} + 1) P_o^m = 10^{-10}$	2
Y_1	$A_1(1/P_0)^\alpha + A_2 2(m-1)$	3
Y_1	$A_1(1/P_0)^\alpha + A_2 \log_2(m+1)$	4
Y_1	$A_1(1/P_0)^\alpha + A_2[2 + \log_2(m+1)]$	5

Відхилення значень параметрів від оптимальних приводить до помітного збільшення обсягу потрібного обчислювального ресурсу мережних вузлів граничного і туманного шарів.

У той же час застосування в другій вирішальній схемі розподіленої схеми підвищує ефективність ПООТ (рис. 1, криві 4, 5), що працює на обмежених ресурсах мережних вузлів та каналах низької якості. У результаті порівняльного аналізу можливих варіантів розподіленої схеми можна обрати найбільш зручний варіант формування ПООТ, зменшуючи час доведення інформації до одержувача.

Висновки

У статті запропонована математична модель процесу обробки оперативних транзакцій.

На основі моделі розроблений метод, котрий дозволяє врахувати специфічні особливості нижніх шарів хмарного середовища підтримки Інтернету речей та обрати найбільш прийнятний варіант побудови підсистеми обробки оперативних транзакцій Інтернету речей.

Розроблений метод дозволяє виконати вимоги QoS для оперативних транзакцій Інтернету речей.

СПИСОК ЛІТЕРАТУРИ

1. Fabre, W., Haroun, K., Lorrain, V., Lepecq, M. and Sicard, G. (2024), "From Near-Sensor to In-Sensor: A State-of-the-Art Review of Embedded AI Vision Systems", *Sensors*, vol. 24(16), 5446, doi: <https://doi.org/10.3390/s24165446>
2. Zhang, Z. (2023), "A computing allocation strategy for Internet of things' resources based on edge computing", *International Journal of Distributed Sensor Networks*, vol. 17(12), doi: <https://doi.org/10.1177/15501477211064800>
3. Sharma, A. and Singh, N. (2022), "Sensors, Embedded Systems, and IoT Components", *Mathematical Modeling for Intelligent Systems: Theory, Methods, and Simulation*, pp. 1–15, doi: 10.1201/9781003291916-1
4. Aburukba, R.O., Landolsi, T. and Omer, D. (2021), "A heuristic scheduling approach for fog-cloud computing environment with stationary IoT devices", *Journal of Network and Computer Applications*, vol. 180, no. 102994, doi: <https://doi.org/10.1016/j.jnca.2021.102994>
5. Li, W., Zhao, B., Zhu, L., Yixuan W., Zhong, Q. and Yu, S. (2024), "TCEC: Integrity Protection for Containers by Trusted Chip on IoT Edge Computing Nodes", *IEEE Sensors Journal*, doi: <https://doi.org/10.1109/JSEN.2024.3445576>
6. Nechausov A., Mamusuc I., Kuchuk N. Synthesis of the air pollution level control system on the basis of hyperconvergent infrastructures. *Сучасні інформаційні системи*. 2017. Т. 1, № 2. С. 21 – 26. DOI: <https://doi.org/10.20998/2522-9052.2017.2.04>
7. Кучук Н. Г., Мерлак В. Ю., Скороделов В. В. Метод зменшення часу доступу до слабкоструктурованих даних. *Сучасні інформаційні системи*. 2020. Т. 4, № 1. С. 97-102. doi: <https://doi.org/10.20998/2522-9052.2020.1.14>
8. Kovalenko, A. and Kuchuk, H. (2022), "Methods to Manage Data in Self-healing Systems", *Studies in Systems, Decision and Control*, Vol. 425, pp. 113–171, doi: https://doi.org/10.1007/978-3-030-96546-4_3
9. Kuchuk, N., Kovalenko, A., Ruban, I., Shyshatskyi, A., Zakovorotnyi, O. and Sheviakov, I. (2023), "Traffic Modeling for the Industrial Internet of NanoThings", *2023 IEEE 4th KhPI Week on Advanced Technology*, KhPI Week 2023 - Conference Proceedings, 2023, doi: 194480. <http://dx.doi.org/10.1109/KhPIWeek61412.2023.10312856>
10. Krishnan, S. and Ilmudeen, A. (2023), "Internet of Medical Things in Smart Healthcare: Post-COVID-19 Pandemic Scenario", *Imprint Apple Academic Press*, New York, doi: <http://dx.doi.org/10.1201/9781003369035>
11. Kuchuk G., Kovalenko A., Komari I.E., Svyrydov A., Kharchenko V. Improving big data centers energy efficiency: Traffic based model and method. *Studies in Systems, Decision and Control*, vol 171. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (Eds.). Springer Nature Switzerland AG, 2019. Pp. 161-183. DOI: http://doi.org/10.1007/978-3-030-00253-4_8
12. Dotsenko, N., Chumachenko, I., Galkin, A., Kuchuk, H. and Chumachenko, D. (2023), "Modeling the Transformation of Configuration Management Processes in a Multi-Project Environment", *Sustainability (Switzerland)*, Vol. 15(19), 14308, doi: <https://doi.org/10.3390/su151914308>

Received (Надійшла) 25.10.2024

Accepted for publication (Прийнята до друку) 11.12.2024

Method for forming a subsystem for processing operative transactions of the IoT

Sergii Klivets, Alexander Kuleshov, Roman Zubko

Abstract. Topicality. Internet of Things transactions are usually processed in cloud data centers. However, when it is necessary to process operational transactions, there are time delays associated with data transfer to the cloud environment. This problem can be solved by transferring calculations to layers close to IoT sensors. The purpose of the article is to develop a method for forming a subsystem for processing operational transactions of the Internet of Things, focused on the boundary and fog layers of the cloud environment. Research results. A mathematical model of the process of processing operational transactions is proposed. Based on the model, a method has been developed that allows taking into account the specific features of the lower layers of the cloud environment supporting the Internet of Things and choosing the most appropriate option for building a subsystem for processing operational transactions of the Internet of Things. Conclusion. The developed method allows you to meet the QoS requirements for operational transactions of the Internet of Things.

Keywords: Internet of Things, operational transaction, fog layer, network node.

С. Ю. Леонов, Д. А. Тиртишний

Національний технічний університет “Харківський політехнічний інститут”, Харків, Україна

РОЗРОБКА ПРОГРАМНОЇ ПЛАТФОРМИ ДЛЯ ТЕСТУВАННЯ ПРОДУКТИВНОСТІ КЛІЄНТСЬКОЇ ЧАСТИНИ ВЕБЗАСТОСУНКУ

Анотація. У даній статті були вивчені методи та використання оригінально розробленого фреймворку для тестування продуктивності клієнтської частини вебзастосунків. Було проаналізовано завдання та виклики, які стоять перед розробниками при оптимізації взаємодії користувачів та рендерингу сторінок. Проведено моделювання взаємодії користувачів з вебзастосунками та створено фреймворк, що включає інструменти для створення сценаріїв користувацької взаємодії, моніторингу швидкодії браузера, аналізу часу завантаження сторінок, та керування тестовим процесом. Експерименти проведено на реальних вебзастосунках, що використовувалися як тестове середовище. Результати показують, що застосування запропонованого фреймворку призводить до оптимізації продуктивності клієнтської частини вебзастосунків, що значно поліпшує їхню швидкість та надійність. В ході досліджень були визначені та підтверджені ознаки подальшого покращення взаємодії користувачів з вебзастосунками.

Ключові слова: комп'ютерна система; програмна платформа; програмне забезпечення; тестування продуктивності; клієнтська частина; вебзастосунок; фреймворк; оптимізація; експеримент; розробка.

Вступ

Тестування продуктивності клієнтської частини вебзастосунків є критично важливим аспектом забезпечення якості, який включає в себе використання автоматизованих тестів, що імітують взаємодію реальних користувачів з інтерфейсом застосунку, таким як браузер. Цей вид тестування є невід'ємною частиною процесу підвищення якості кінцевого продукту, оскільки будь-які збої або проблеми, пов'язані з низькою продуктивністю, можуть призвести до втрати клієнтів, які відмовляться від використання вашого вебзастосунку.

Уявімо ситуацію: під час великого онлайн-розпродажу, такого як Чорна п'ятниця, інтернет-магазин очікує на значне збільшення трафіку. І раптом, через проблеми з продуктивністю клієнтської частини, сайт починає зависати, сторінки завантажуються повільно, товари додаються у корзину з затримками, оплата не проводиться з першого разу - користувачі починають залишати сайт та переходити до конкурентів. Це може призвести до величезних фінансових втрат для компанії, як, наприклад, Amazon чи eBay, якщо їхні вебзастосунки не витримують навантаження у такий критичний момент.

Тестування продуктивності клієнтської частини дозволяє переконатися, що вебдодаток ефективно працює під час пікових навантажень і забезпечує користувачам безперервне та плавне керування інтерфейсом. Це допомагає виявити та оптимізувати елементи інтерфейсу, які можуть сповільнювати роботу або створювати незручності для користувачів.

Мета статті: виконати дослідження та розробити фреймворк для тестування продуктивності клієнтської частини вебзастосунків.

Основна частина

Тестування продуктивності клієнтської частини вебзастосунків є критично важливим для забезпечення оптимального користувацького досвіду. Продуктивність клієнтської частини відображає швидкість застосунку з точки зору кінцевого користувача, включаючи доставку контенту, його трансформацію та рендеринг у браузері або нативному застосунку. Важливо розуміти, що це відрізняється від часу обробки сервером, і не показує рівень навантаження, який сервер може витримати.

Для оцінки продуктивності клієнтської частини можна використовувати різні метрики, такі як час відповіді, час повного завантаження сторінки, розмір сторінки, кількість HTTP-запитів, розмір зображень, час виконання Javascript та інші. Ці метрики допомагають зрозуміти, як додаток веде себе з точки зору користувача, що є ключовим для оптимізації продуктивності. Однією з основних причин, чому тестування продуктивності клієнтської частини є таким важливим, є те, що вимірювання на стороні сервера не можуть адекватно відображати поведінку з точки зору кінцевого користувача. Наприклад, час відповіді сервера не дорівнює часу повного завантаження сторінки на стороні клієнта. Різні місця розташування клієнтів, різні конфігурації машин та використання сторонніх сервісів можуть значно вплинути на продуктивність.

Типи тестування продуктивності клієнтської частини включають мережеве тестування, тестування з різних місць, тестування користувацького досвіду, крос-браузерне тестування та мобільне тестування. Кожен з цих типів має свої особливості та допомагає виявити потенційні проблеми, які можуть вплинути на користувацький досвід. Наприклад, мережеве тестування шляхом емуляції різних пропускових здатностей та типів мереж може виявити специфічні вузькі місця для користувачів з мобільними або повільними з'єднаннями. Тестування з різних місць допомагає зрозуміти вплив мережевої швидкості та затримки, а також ефективність CDN. Тестування користувацького досвіду імітує поведінку кінцевих користувачів, дозволяючи оцінити, як додаток веде себе під час нормального та підвищеного навантаження. Крос-браузерне тестування важливе для забезпечення сумісності застосунку в різних браузерах, кожен з яких має свої особливості в обробці Javascript та рендерингу.

Загалом, тестування продуктивності клієнтської частини є невід’ємною частиною розробки вебзастосунків, оскільки воно допомагає забезпечити високий рівень задоволеності користувачів та ефективність вебзастосунку.

Для реалізації фреймворку тестування продуктивності клієнтської частини потрібно спочатку дослідити різні інструменти для тестування продуктивності, вивчити їх взаємодію та розробити архітектурну діаграму (рис. 1).

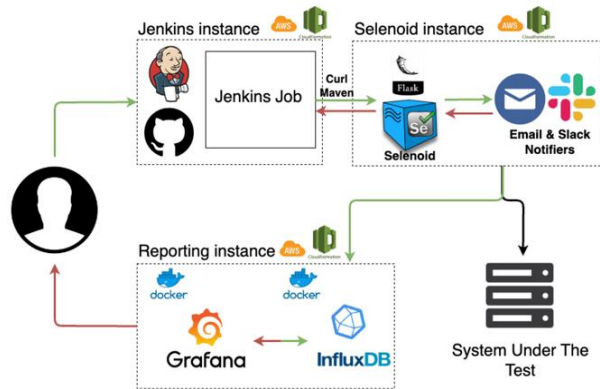


Рис. 1. Високорівнева архітектура фреймворку тестування продуктивності

Опираючись на архітектуру фреймворку тестування продуктивності клієнтської частини вебзастосунків, почнемо реалізацію поточного рішення. За основу візьмемо автотести написані на мові Java з використанням Selenium. Єдина відмінність тестів продуктивності від звичайних автотестів є те, що ми додаємо лише 1 стрічку коду для кожного методу (рис. 2) і вона виконує «магію», яка дозволяє автоматично дістати метрики продуктивності сторінки через Navigation Timing API та Resource Timing API, потім підрахувати усі важливі для візуалізації метрики, та надіслати їх до InfluxDB – в якій зберігаються всі метрики тестування продуктивності.

```
public LaptopPage openFirstLaptop() {
    String firstLaptopInTheListName = laptopProductList.get(0).getText();
    log.info("Let's grab first laptop: " + firstLaptopInTheListName);
    laptopProductList.get(0).click();
    wait.until(ExpectedConditions.elementToBeClickable(productDescriptionAndPricesButton));
    Assert.assertTrue(openedLaptopTitle.getText().contains(firstLaptopInTheListName), "Wrong laptop was opened!");
    wait.untilPageIsFullyLoaded(wait);
    getNavigationTiming.writeToInflux(pageName: "ProductPage");
    log.info("Asserted that the laptop we wanted to open and actually opened laptop are the same product!");
    return this;
}
```

Рис. 2. Приклад методу з додаванням автоматичного підрахунку метрик продуктивності клієнтської частини вебзастосунку

Для того щоб досягти такого ефекту та легкості використання розробленого методу, був створений Java клас – NavigationTiming, в якому виконується декілька важливих дій:

- «Перехоплення» метрик продуктивності з браузера, запущеного за допомогою «WebDriver».
- Підрахунок усіх важливих для візуалізації метрик.
- Надсилання метрик до InfluxDB.

Почнемо з першого пункту - «Перехоплення» метрик продуктивності з браузера, запущеного за допомогою «WebDriver» (рис. 3).

```
private final String javascriptForPerformance = "var timings = performance.timing || {};return timings;";
...
public Map<String, Object> getAllTimings() {
    JavaScriptExecutor jsExecutor = (JavaScriptExecutor) getDriver();
    if (!jsExecutor.executeScript("return performance.timing;")) {
        Map<String, Object> timings = (Map<String, Object>) jsExecutor.executeScript(javascriptForPerformance);
        timings = parseNavigationTimingDataFromJs(timings);
    } else {
        timings = (Map<String, Object>) jsExecutor.executeScript(javascriptForPerformance);
    }
    return timings;
}
```

Рис. 3. Перехоплення метрик продуктивності з браузера

За допомогою кількох команд на мові JavaScript робимо запит до браузера на повернення NavigationTiming об’єкту (рис. 4) для поточної сторінки, який потім приводимо до Java об’єкту Map<String, Object>, де String – це назва події, а Object – це timestamp. Зазвичай timestamp відображається у вигляді числа, що представляє кількість мілісекунд, які пройшли від певної епохи часу. Наприклад, у багатьох мовах програмування або базах даних використовується Unix timestamp, який показує кількість секунд, які пройшли від 1 січня 1970 року до певного моменту. Використання timestamp дозволяє легко порівнювати події, встановлювати порядок подій і виконувати інші операції, пов’язані з часом, у різних програмних системах.

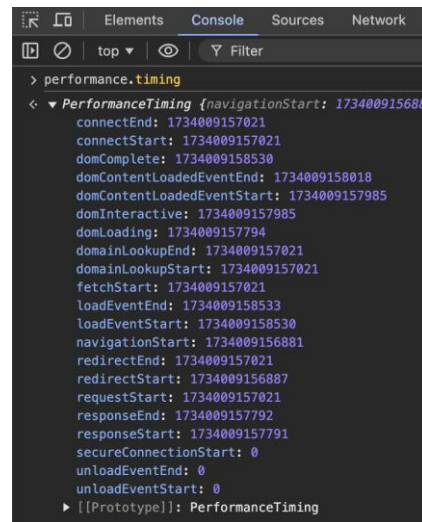


Рис. 4. Приклад Navigation Timing об’єкту

Наступним етапом є «Підрахунок усіх важливих для візуалізації метрик». Для цього підійдуть стандартні методи мови програмування Java. Достатньо перетворити отриманий timestamp у формат long, та зробити елементарні операції віднімання для того щоб отримати потрібні для візуалізації метрики. (рис. 5). Наприклад: для отримання важливої метрики «timeToInteract» достатньо відняти час події з назвою «domInteractive» від «domLoading», а для того щоб отримати повний час завантаження сторінки достатньо відняти час останньої події об’єкту Navigation timing для поточної сторінки, від першої події. А саме «Load Event End» - «Navigation Start». Останнім етапом є відправка підрахованих метрик до бази даних, в якій зберігаються результати тестування продуктивності. Для цього була використана бібліотека influxdb-java. Спочатку була піднята база даних InfluxDB, після цього я встановив зв’язок з цією базою.


```

235 //results
236 2 usages 4 Dmytro Tyrtshynyi
237 public long getLatency() {return getresponseStart() - getnavigationStart();}
238
239
240 1 usages 4 Dmytro Tyrtshynyi
241 public long getBackend_response() {
242     return getresponseEnd() - getresponseStart();
243 }
244
245 2 usages 4 Dmytro Tyrtshynyi
246 public long getTimeToInteract() {
247     return getdomInteractive() - getdomLoading();
248 }
249
250 2 usages 4 Dmytro Tyrtshynyi
251 public long getTimeToLoad() {return getdomComplete() - getdomInteractive();}
252
253 2 usages 4 Dmytro Tyrtshynyi
254 public long getOnload() {return getloadEventEnd() - getloadEventStart();}
255
256 2 usages 4 Dmytro Tyrtshynyi
257 public long getTotal_time() {return getloadEventEnd() - getnavigationStart();}
258
259

```

Рис. 5. Підрахунок важливих для візуалізації метрик.

Після цього потрібно сформувати об'єкт класу Point, в якому потрібно вказати measurement (таблицю для зберігання результатів), .time – час запису результатів в базу даних, .tag – теги для фільтрації результатів, .field – поля для підрахованих раніше метрик. Після цього кожен такий Point додається в батч, та відправляється асинхронно в базу даних. (рис. 6).

```

64 InfluxDB.setLogLevel(InfluxDB.LogLevel.BASIC);
65 Point point = Point.measurement("clientSide") // Формуємо об'єкт для запису в базу даних
66     .time(System.currentTimeMillis(), TimeUnit.MILLISECONDS)
67     .tag("projectName", PROJECT_NAME)
68     .tag("scenario", SCENARIO_NAME)
69     .tag("env", ENV_NAME)
70     .tag("browser", BROWSER_NAME)
71     .tag("page", pageName)
72     .tag("periodicity", PERIODICITY)
73     .tag("periodicity.comment", VERSION)
74     .tag("buildId", BUILD_ID)
75     .addField("latency", this.getLatency())
76     .addField("backend_response", this.getBackend_response())
77     .addField("titi", this.getTimeToInteract())
78     .addField("ttl", this.getTimeToLoad())
79     .addField("onload", this.getOnload())
80     .addField("total_time", this.getTotal_time())
81     .build();
82 batchPoints.point(point);
83 influxDB.write(batchPoints); // записуємо об'єкти в базу даних

```

Рис. 6. Запис результатів в базу даних

І найголовніше – усі ці операції займають лише декілька мілісекунд, тому на метрики та результати це ніяк не впливає.

Після розробки тестового сценарію, його можна запускати. Дуже важливою частиною розробленого інструменту є візуалізація отриманих результатів за допомогою Grafana. Для того щоб візуалізувати результати в Графані треба створити свій власний Dashboard (дошку), на яку можна додавати потрібні панелі для детальної візуалізації результатів, як в загальному на весь тестовий сценарій, так і на завантаження кожної сторінки, або виконання кожної дії окремо. Grafana dashboard - це інтерактивний інтерфейс, який використовується для візуалізації та моніторингу даних з різних джерел. Grafana - це популярний інструмент для створення і відображення dashboardів у реальному часі.

Під час побудови власної дошки, я керувався принципами того, що треба відобразити усі найважливіші метрики разом для побудови загального тренду швидкості роботи вебзастосунку, який тестується (рис. 7). Також, були враховані можливості Графани для відображення деталей та усіх важливих метрик по кожній сторінці або дії окремо. Що дозволяє знаходити проблемні ділянки в нашому застосунку і сфокусуватись на їх виправленні.

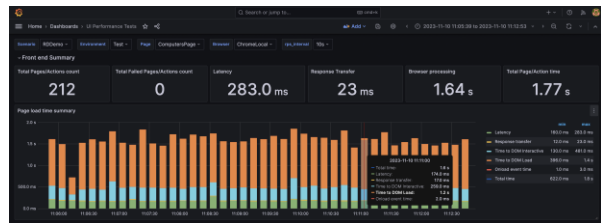


Рис. 7. Загальний тренд продуктивності по сценарію

Важливо додати, що за допомогою тегів, які ми додавали до наших записів метрик, тепер ми можемо фільтрувати відображення результатів на нашій дошці. Можна фільтрувати за різними сценаріями, тестовими середовищами, сторінками, результати для яких будуть відображатись. Також, можна фільтрувати, а також порівнювати результати між собою за різними тегами, як наприклад "browser". Завдяки цьому тегу можна порівняти швидкість роботи одних і тих самих сторінок або дій в різних браузерах, що допоможе надавати однаково найкращу швидкість різним користувачам. Для демонстрації роботи фільтрів, пропоную спочатку обрати одну сторінку для відображення метрик по ній (рис. 8), а потім ще кілька і порівняти результат відображення у Графані (рис. 9).

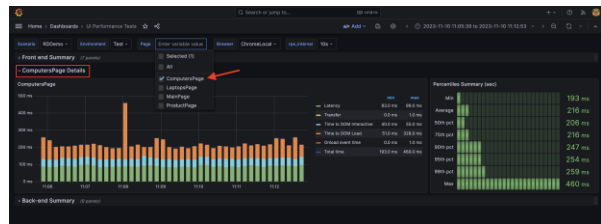


Рис. 8. Відображення панелей лише для 1 вибраної сторінки

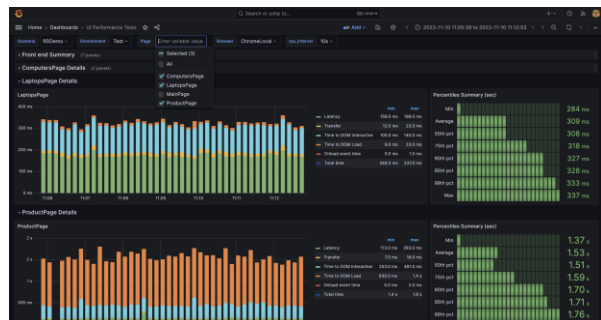


Рис. 9. Відображення панелей для кількох обраних сторінок

Як ви можете бачити на скріншоті вище, ProductPage сторінка працює приблизно в 5 разів повільніше, ніж LaptopsPage, це вже є ознакою того, що треба більш детально розглянути контент цієї сторінки та зрозуміти природу більш повільної роботи та відображенню.

Далі розглянемо більш детально кожен панель для відображення метрик по окремих сторінках. Для візуалізації метрик по окремій сторінці було створено 2 панелі, які автоматично реплікуються для кожної обраної сторінки. Це було досягнуто завдяки використанню InfluxQL – мова запитів до бази даних InfluxDB, та можливостей візуалізації в Grafana.

Перша панель відображає швидкість відкриття сторінки на часовому проміжку. Тобто на цьому графіку можна порівняти швидкість та стабільність відкриття сторінок на великому проміжку часу. Кожен стовпець з результатом відповідає за окреме відкриття сторінки та візуалізує різними кольорами швидкість виконання різних операцій з завантаження та відображення сторінки: Latency, Time to Dom Interactive, Time to DOM Load, Total Time (рис. 10).

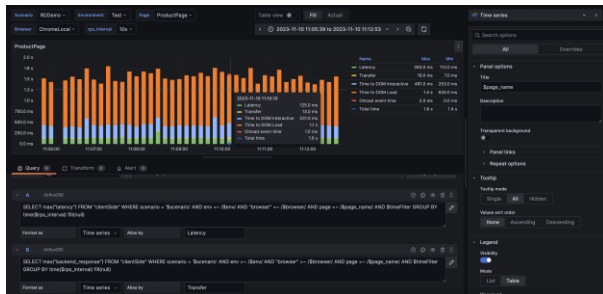


Рис. 10. Приклад налаштування панелі зі стовпцями в Grafana

Друга панель відображає результати в різних перцентиліях, та була побудована по тому ж самому принципу, за виключенням різних налаштувань та типу візуалізації в Графані. Перша панель має тип «Time Series», коли друга панель – це «Bar Gauge», яка має трохи іншу візуалізацію. Головна задача другої панелі візуалізувати агреговані результати за обраний проміжок часу з цілю зрозуміти як швидко сторінка буде відображатись з різними рівнями вірогідності. Для цього в навантажувальному тестуванні використовуються перцентилі. Перцентиль (в контексті Графани та аналізу даних) - це статистичний термін, що вказує на значення в наборі даних, нижче якого певний відсоток спостережень. Наприклад, якщо ви маєте 90-й перцентиль для швидкості завантаження вебсайту, це означає, що 90% користувачів мають швидкість завантаження, яка менше або рівна даному значенню, а лише 10% користувачів мають швидкість завантаження вище цього значення. У Графані, при аналізі метрик та візуалізації даних, використовуються перцентилі для того, щоб отримати уявлення про розподіл даних і виявити екстремальні значення або аномалії. Це може бути корисно при моніторингу продуктивності, де важливо зрозуміти, які частини користувачів можуть відчувати найкращий чи найгірший досвід (рис. 11).

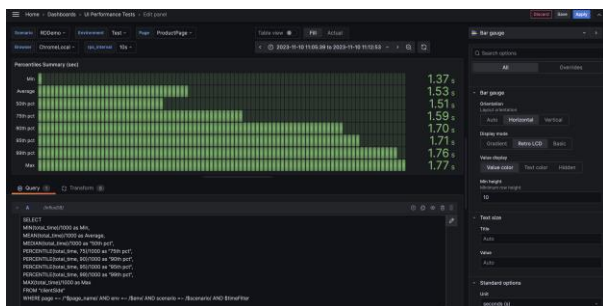


Рис. 11. Приклад налаштування панелі з перцентиліями в Grafana

Як бачимо, на скріншоті вище, перцентилі рахуються завдяки можливостям InfluxQL, в якому є метод PERCENTILE, що підраховує перцентилі для потрібних нам метрик на обраному проміжку часу. Фільтрування метрик за тегами та часом відбувається завдяки ключовому слову WHERE, де ми можемо вказати теги для фільтрації та часовий проміжок, наприклад: WHERE page =~ /^\$page_name/ AND env =~ /\$env/ AND scenario =~ /\$scenario/ AND \$timeFilter.

Також, завдяки інтеграції з системою неперервної інтеграції Jenkins в моєму інструменті тестування продуктивності, ми можемо запускати тести через дуже зручний інтерфейс, який дозволяє робити автоматичний запуск тестів за розкладом, наприклад вночі (Див. Додаток Б). Користувач має лише задати параметри запуску в зручному користувацькому інтерфейсі і натиснути кнопку Build, яка почне виконання тестів (рис. 12). Передавати можна такі параметри:

- назва тестового сценарію;
- браузер;
- тестове середовище (для розробки, для тестування);
- кількість ітерацій запуску тестового сценарію;
- чекбокс для автоматичного запису відео у разі виникнення помилок під час тесту;
- користувачі, які отримають автоматичну імейл нотифікацію після тесту.

This build requires parameters:

environment

browser

scenario
Choose name of a test scenario

loop_count
Whole scenario iterations

enable_video
Check if you want to receive the video of test execution in case of failure

periodicity

version

email_recipients
coma separated list of email recipients

slack_notification
Receive Slack notification?

Рис. 12. Інтерфейс Jenkins

Після автоматичного запуску тестів, усі користувачі, імейли яких були вказані при запуску, отримають автоматичний звіт про тест, в якому можна буде побачити статус тесту, отримати відео для ітерацій в яких виникли помилки, та отримати посилання на відображення результатів в Графані, або відображення логів у Jenkins (рис. 13).

Обговорення результатів

В ході дослідження були виконані дослідження теоретичної частини продуктивності клієнтської частини вебзастосунків, включаючи огляд особливостей цього тестування, його переваги, та основні види.

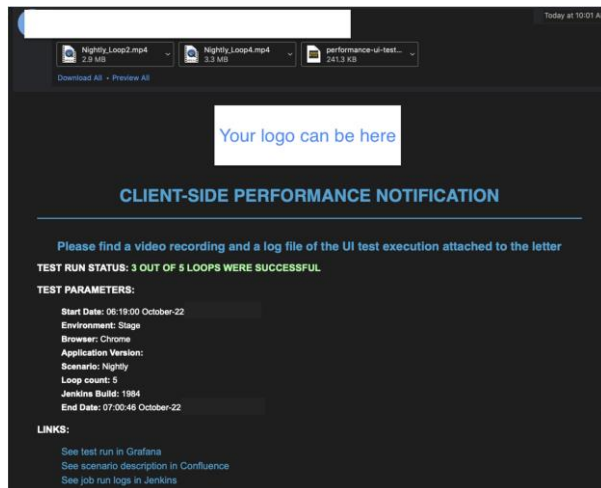


Рис. 13. Приклад автоматичного звіту на імейл

Після того були досліджені різні інструменти для розробки фреймворку тестування продуктивності клієнтської частини вебзастосунків. Усі ці інструменти були застосовані на практиці та продемонстровані у дії. Після розробки фреймворку, був розроблений тест для прикладу, та завдяки цьому тестовому сценарію був протестований інтернет-магазин.

Тести продуктивності клієнтської частини виявили найповільніші для завантаження сторінки та вказали на найпроблемніші місця, які потребують поглибленого аналізу з використанням більш потужних та деталізованих інструментів для аналізу та

оптимізації проблем з продуктивністю клієнтської частини вебзастосунків.

Висновки

У результаті проведеного дослідження тестування продуктивності серверної частини вебзастосунків та розробки фреймворку потрібно зазначити, що впровадження цього фреймворку дозволяє підвищити ефективність роботи та стабільність клієнтської частини вебзастосунків.

Були детально досліджені особливості тестування продуктивності клієнтської частини вебзастосунків, продемонстровані варіанти рішень, які включають в себе набір спеціалізованих інструментів для досягнення мети дослідження.

Розроблений фреймворк включає інструменти для збору метрик продуктивності з браузера клієнта, моніторингу, аналізу даних, керування тестовим процесом, а також модулі для вдосконалення продуктивності. Він є зручним та гнучким інструментом для ефективного аналізу та вдосконалення продуктивності вебзастосунків.

Експерименти, проведені в рамках дослідження, підтвердили, що використання розробленого фреймворку допомагає в дослідженні з метою значного покращення продуктивності клієнтської частини вебзастосунків.

Отже, результати цього дослідження можуть бути використані розробниками вебзастосунків для підвищення продуктивності та надійності своїх продуктів.

СПИСОК ЛІТЕРАТУРИ

1. «Проблеми інформатики та моделювання (ПІМ-2023), Харків, 2023», матеріали міжнародної науково-технічної конференції, стор.69 «КЛЮЧОВІ АСПЕКТИ ТЕСТУВАННЯ ПРОДУКТИВНОСТІ»
2. "Web performance warrior" book by Andy Still.
3. "MicroCAD-2024, 9.4 ІНФОРМАТИКА І МОДЕЛЮВАННЯ", с.1407, Леонов С.Ю, Тиртишний Д.А, Використання сучасних методів тестування та аналізу клієнтської частини вебзастосунків
4. «Проблеми інформатики та моделювання (ПІМ-2024), Харків, 2024», матеріали міжнародної науково-технічної конференції, стор.77 «Стратегія кешування контенту для оптимізації LCP у динамічних вебзастосунках»
5. Джеймс Бах та Джейсон Теслер: "Автоматизоване тестування ПЗ: Вступ до професійної практики".
6. Мартін Аббот: "Тестування продуктивності програмного забезпечення".
7. Джеймс Петер: "Тестування продуктивності та навантаження".
8. «Інформатика, управління та штучний інтелект», матеріали п'ятої міжнародної науково-технічної конференції, стор.65 «ТЕСТУВАННЯ ПРОДУКТИВНОСТІ WEB-САЙТУ ДЛЯ ЗРОСТАННЯ ЯКОСТІ ПРОДУКТУ»
9. WebDev [Електронний ресурс]: Understanding time to first byte metric : [https://web.dev/articles/ttfb]
10. WebDev [Електронний ресурс]: Understanding first contentful paint (FCP): [https://web.dev/articles/fcp]
11. WebDev [Електронний ресурс]: Understanding time to interactive metric : [https://web.dev/articles/tti]

Received (Надійшла) 08.11.2024

Accepted for publication (Прийнята до друку) 12.02.2025

Development of a software platform for testing the performance of the client part of a web application

Sergey Leonov, Dmytro Tyrtshnyi

Abstract. This article explores methods and the use of an originally developed framework for testing the client-side performance of web applications. The tasks and challenges faced by developers in optimizing user interactions and page rendering were analyzed. Simulation of user interactions with web applications was conducted, and a framework was created that includes tools for creating user interaction scenarios, monitoring browser performance, analyzing page load times, and managing the testing process. Experiments were carried out on real web applications used as a testing environment. The results show that applying the proposed framework leads to optimizing the performance of the client-side of web applications, significantly improving their responsiveness and reliability. During the research, signs of further improvement in user interactions with web applications were identified and confirmed.

Keywords: computer system; software platform; software; performance testing; client-side; web applications; framework; optimization; experiment; development.

С. В. Носко, С. С. Бульба, О. В. Коломійцев, Д. О. Лисиця, Г. І. Молчанов

Національний технічний університет «Харківський політехнічний університет», Україна, Харків

ПРОПОЗИЦІЇ ЩОДО АВТОРИЗАЦІЇ В САЙДКАР КОМПОНЕНТІ МІКРОСЕРВІСНОЇ АРХІТЕКТУРИ

Анотація. У статті розроблено пропозиції щодо доцільності реалізації авторизації у сайдкар компоненті мікросервісу, що дозволяє відокремити бізнес-логіку від завдань авторизації, ведення журналів, кешування тощо. Така реалізація дозволяє основним бізнес-модулям залишатися сфокусованими виключно на своїй бізнес-логіці, змінюючись лише тоді, коли є оновлення у бізнес-процедурах, а задачі, які є загальними майже для будь-якого сучасного рішення, передавати допоміжним компонентам таким як сайдкар. Правильна реалізація авторизації виступає ключовим аспектом будь-якої системи, є дуже важливим та вимагає врахування великої кількості особливостей та використання передових практик для забезпечення збереження даних від несанкціонованого доступу. Розглянуто можливості open-source рішення Keycloak. Таке рішення є дуже популярним сервісом для аутентифікації/авторизації, має підтримку стандартних протоколів ідентифікації, таких як: OpenID Connect, OAuth 2.0, SAML 2.0. Рішення підтримує багатфакторну аутентифікацію, підтримує контейнеризацію та може бути легко розподілений у середовищі мікросервісів як Kubernetes та інтегруватися з різними зовнішніми сервісами, включаючи Google, Facebook тощо. Розроблено архітектуру та пропозиції, що необхідні для виконання інтеграції з сайдкар компонентом. Проведено дослідження щодо продуктивності розробленого рішення. Отримано числові значення, що наведено у таблицях та на графіку.

Ключові слова: сайдкар, мікросервіс, авторизація, Keycloak, продуктивність, кеш, контейнер, Kubernetes, масштабованість, open-source, Gatling, розподіл ресурсів, балансування навантаження, веб-сервер, інформаційна система.

Вступ

Постановка проблеми. В сучасних розподілених системах, що базуються на архітектурі мікросервісів, важливо забезпечити високу продуктивність, безпеку та масштабованість. Одним із викликів є ефективне управління авторизацією та аутентифікацією. Використання окремих мікросервісів для обробки авторизації може призвести до зниження продуктивності через дублювання зусиль та збільшення обсягу мережевих взаємодій.

Відповідальність мікросервісів за авторизацію додатково збільшує навантаження на них, а, отже, сповільнює реалізацію бізнес процесів. Підхід щодо авторизації має бути уніфікованим та централізованим, щоб спростити архітектуру рішення. Для реалізації таких комплексних завдань використовується сайдкар дизайн паттерн.

Авторизація та аутентифікація є одним із важливих процесів, де сайдкар знаходить ефективне застосування. До переваг такого рішення можливо віднести наступні:

- реалізація аутентифікації та авторизації у сайдкарі дозволяє управляти безпекою централізовано для усієї системи та усуває необхідність реалізації цих функцій у кожному мікросервісі окремо;
- використання сайдкару забезпечує однорідне застосування контролю доступу та політик безпеки для усіх мікросервісів, що усуває необхідність у додатковій конфігурації окремих мікросервісів, сприяє як зменшенню кількості помилок, так і полегшенню підтримки;
- масштабованість мікросервісів проходить простіше, оскільки усі налаштування вже інтегровані у сайдкар компонентах, а не у мікросервісах.

Авторизація у розподілених системах, особливо у архітектурах мікросервісів, є однією з найбільш складних та критичних задач, яка вимагає не тільки

технічного, але й експертного дослідження у сфері керування безпекою. Складність полягає у забезпеченні надійної та безпечної авторизації без негативного впливу на продуктивність системи. Вибір неправильного рішення може призвести до вразливостей у безпеці та зниження ефективності системи.

Таким чином, необхідно обрати ефективний інструмент, який дозволить вирішити висвітлене наукове завдання. Одним із шляхів вирішення може бути застосування Keycloak, який активно використовується у індустрії для управління ідентичностями та авторизацією у розподілених системах.

Аналіз останніх досліджень і публікацій. Авторизація у мікросервісних архітектурах є ключовим компонентом, що впливає на безпеку та ефективність розробки і впровадження розподілених сервісів. У [1] проведено дослідження питання щодо безпеки мікросервісів. Визначено авторизацію як основний механізм для забезпечення дотримання принципів обмеженого доступу та ізоляції служб.

Документація Keycloak [2] надає інструкції щодо розгортання потужного серверу аутентифікації та авторизації у контексті мікросервісних архітектур, до якого сайдкар звертається для валідації токенів. Keycloak підтримує стандарти ID та доступу, такі як: OpenID Connect і OAuth 2.0, що дозволяє гнучко і безпечно управляти авторизацією користувачів.

У [3] проведено аналіз використання облікових даних API у мікросервісах, де пропонується підхід до ефективної авторизації та аутентифікації через API шлюзи, що дозволяє покращити забезпечення безпеки у таких системах. Використовується API шлюз Kong, але його участь у процесах авторизації мінімальна, оскільки майже увесь об'єм роботи бере на себе сайдкар.

У [4] проведено дослідження різних підходів до механізмів авторизації у мікросервісах. Акцентуєть-

ся увага на використанні Role-Based Access Control (RBAC) як ефективного засобу управління доступом на різних рівнях. Їх робота підкреслює потребу у стандартизації взаємодій між незалежними сервісами, щоб забезпечити безпечний обмін інформацією. Такий підхід підтримується сервером Keycloak, який вибраний для реалізації у даній роботі.

У [5] проведено дослідження, як сайдкар може використовуватись для забезпечення безпечної авторизації, розділивши основні обов'язки між ядром сервісу та сайдкар проксі. Такий підхід дозволяє масштабувати авторизаційні контролі ефективніше та зменшити складність управління. Даний аналіз використовує схожі підходи у авторизації з використанням сайдкару як і у даній роботі.

У [6] проведено детальний аналіз використання Envoy Proху як сайдкар рішення для управління авторизацією та моніторингу у мікросервісних архітектурах. Envoy дозволяє відстежувати та керувати забезпеченням безпеки на рівні кожного мікросервісу, подавати детальні логи доступу для аналізу та виявлення можливих атак на сервіс. Envoy Proху це альтернатива рішенню, представленого у даній статті, але її архітектура не дозволяє кастомізувати та додавати нових можливостей таких як інтеграція зі штучним інтелектом.

Мета статті – розробка пропозицій щодо реалізації авторизації у сайдкар компоненті мікросервісу для відокремлення бізнес-логіки мікросервісу від завдань авторизації та підвищення продуктивності рішення.

Виклад основного матеріалу

Застосування підходу з сайдкарками і Keycloak у якості авторизаційного сервера дозволяє в цілому підвищити безпеку та управління доступом у мікросервісних архітектурах.

Keycloak має відкритий вихідний код, що надає обширні можливості для кастомізації та розширення функціональності для потреб бізнесу. Платформа адаптована для роботи у великих організаціях та здатна легко масштабуватися зі зростанням системи.

При розробці використовувалась версія, яка має інтеграцію з OpenID Connect Provider, що представлено на рис. 1.

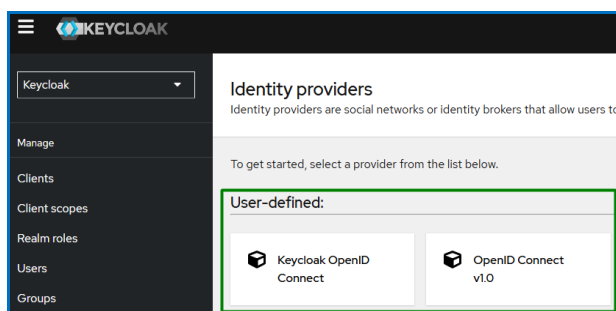


Рис. 1. Вибір OpenID Identity провайдера

Недоліком рішення є ускладнення управління системою при впровадженні додаткового компонента, що може виявитися складним у налаштуванні для людей без відповідного досвіду та вимагати до-

даткових обчислювальних ресурсів. Також, Keycloak є відносно комплексною системою, що потребує регулярних оновлень безпеки та підтримки.

Незважаючи на можливі ускладнення в управлінні та вимоги до ресурсів, переваги використання Keycloak як рішення для розумного сайдкара, особливо у контексті забезпечення безпеки, автентифікації та централізованого управління доступом, значно перевищують потенційні недоліки. Таке рішення забезпечує високий рівень безпеки та ефективності для мікросервісних архітектур [7, 8].

Використання Keycloak як окремого контейнера у Kubernetes середовищі є одним із оптимальних способів розгортання для досягнення високої доступності, масштабованості та управління. Таке рішення надає ряд важливих переваг:

- розгортання Keycloak у окремому контейнері забезпечує ізоляцію від інших частин системи, що підвищує безпеку та стабільність;

- Kubernetes підтримує високу доступність шляхом автоматичного перезапуску контейнерів у випадку їх падіння або помилок, забезпечуючи таким чином високу відновлюваність.

У якості типу контролера у Kubernetes для Keycloak обрано StatefulSet замість Deployment, який використовується для stateless мікросервісів. Таким чином зберігається стан сесій, що забезпечує відновлення після збою без втрати даних як представлено на рис. 2.

У процесі реалізації системи управління доступом та ідентифікацією за допомогою Keycloak важливо враховувати інтеграцію та синхронізацію даних користувачів та їх ролей. Варіанти рішення цього завдання залежать від архітектурних особливостей вже існуючої системи.

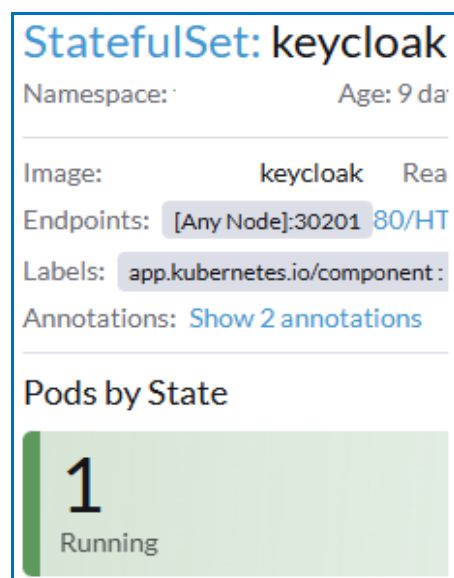


Рис. 2. Використання StatefulSet контролера

У контексті мікросервісної архітектури, де вже існує повнофункціональне зберігання користувачів та їх ролей, ефективним рішенням є використання Keycloak REST API для забезпечення синхронізації

даних. Такий підхід передбачає, що при зміні інформації про користувача у основній мікросервісній системі, такий як модуль mod-users, потрібно виконувати відповідні POST, PUT або DELETE HTTP запити до Keycloak API, що дозволяє забезпечити консистентність даних між різними компонентами системи.

На рис. 3 представлено список користувачів, зареєстрованих у визначеному Realm системи Keycloak. У даному контексті, Realm є логічним простором у системі, що містить окремі користувачькі ідентичності, облікові дані, ролі та інші конфігураційні параметри.

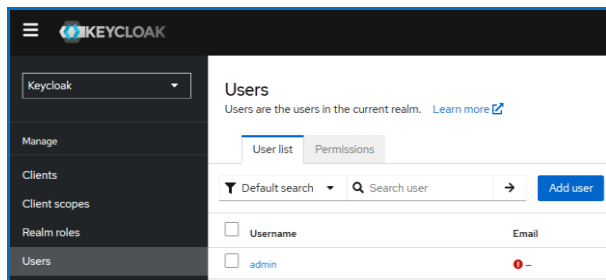


Рис. 3. Користувачі у системі

На рис. 4 представлені ролі, які доступні у системі для подальшого присвоєння користувачам. Ролі у контексті Keycloak необхідні для управління правами доступу, вони можуть бути як глобальними, так і визначеними для конкретного Realm.

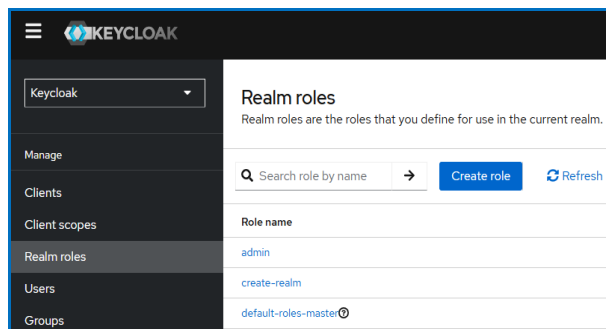


Рис. 4. Ролі у системі

Запропоновано у реалізації розумного сайдкару авторизацію здійснювати у сайдкар компонентах. Виклик між модулями здійснювати через відповідні сайдкари.

До інших варіантів реалізації запропонованих пропозицій можливо віднести наступне:

- централізований API шлюз. У даному випадку усі перевірки авторизації відбуваються перед тим, як трафік досягне мікросервісів. Даний шлюз може керувати аутентифікацією, шифруванням, логуванням та іншими аспектами безпеки для усіх вхідних запитів;

- обробка JWT токенів на рівні мікросервісів. Мікросервіси можуть самостійно обробляти JWT токени, декодувати їх та перевіряючи валідність. При такій реалізації кожен мікросервіс повинен мати логіку для валідації токена, що призводить до дублювання логіки та ускладнення впровадження змін у способі авторизації;

- використання зовнішньої системи для оцінки політик, контролю та безпеки доступу таких як Open Policy Agent. Такі системи дозволяють централізувати правила авторизації та легко їх оновлювати.

У сучасних мікросервісних архітектурах, де досягнення високого рівня безпеки доступу є одним із ключових викликів, дослідження засобів авторизації критичне для забезпечення загальної безпеки системи. Використання «розумного» сайдкара для обробки авторизації, зокрема через інтеграцію із Keycloak, надає певні переваги порівняно з авторизацією, виконаною безпосередньо у мікросервісі.

Перш за все, такий підхід сприяє відокремленню зон відповідальності. Сайдкари беруть на себе повну відповідальність за обробку авторизації, звільняючи основний контейнер мікросервісу від необхідності включати розгалужені перевірки безпеки, які можуть з часом змінюватись або оновлюватись.

Додатково, використання уніфікованої логіки авторизації у сайдкарі сприяє виключенню дублювання логіки авторизації у кожному мікросервісі, значно спрощуючи процес розробки та підтримки. При змінах у правилах авторизації, потрібно здійснити оновлення лише в одному місці – сайдкарі, що зменшує комплексність управління оновленнями, порівняно з необхідністю різних адаптацій у кожному мікросервісі.

Використання сайдкара, також, підвищує загальний рівень безпеки, оскільки зменшує ймовірність помилок у реалізації авторизації через меншу кількість точок входу та стандартизацію підходів.

Кроки взаємодії Keycloak з мікросервісною архітектурою з використанням "smart sidecar" для обробки авторизації виглядають наступним чином:

1. Користувач взаємодіє з інтерфейсом користувача (UI), наприклад, веб-сторінка, який викликає відповідний API ендпоінт. HTTP-запит, здійснений користувачем, містить у своєму заголовку токен авторизації (JWT токен).

2. Запит спочатку обробляється Kong, який виступає у ролі API Gateway. Kong аналізує HTTP заголовок, визначає URI запити та, використовуючи внутрішню схему бази даних, де збережено відповідності URI до специфічних мікросервісів, перенаправляє запит до відповідного мікросервісу.

3. Запит досягає відповідного мікросервісу та перехоплюється сайдкар, пов'язаним із цим мікросервісом. Сайдкар відіграє ключову роль у обробці запиту на цьому етапі.

4. Сайдкар здійснює запит до Keycloak, передавши необхідні дані про користувача (зазвичай у вигляді вищезгаданого JWT токена). Запит включає важливі контекстні дані.

5. Keycloak перш за все перевіряє чинність самого токена – зокрема, правильність його формату, підпису, а також час дії (expiration).

6. Після успішної валідації JWT токена Keycloak додатково перевіряє, чи має користувач достатні права авторизації для виконання запиту. Це включає перевірку заданих у токені ролей та дозволів користувача порівнюючи з обмеженнями доступу, визначених у Keycloak для даного ендпоінту або ресурсу.

На рис. 5 показана перша частина діаграми авторизації, де запит оброблюється компонентом Kong, що виступає у ролі Арі Gateway. На рис. 6 показана друга частина діаграми, де сайдкар викликає сервер авторизації Keycloak для перевірки прав користувача перед викликом свого мікросервісу.

Використання сайдкар-компонентів забезпечує можливість локалізації процесів авторизації на рівні окремого мікросервісу, що підвищує продуктивність системи, оскільки після успішної авторизації в Keycloak, сайдкар мікросервісу 1 може безпосередньо комунікувати з мікросервісом 2, оминувши повторні перевірки авторизації.

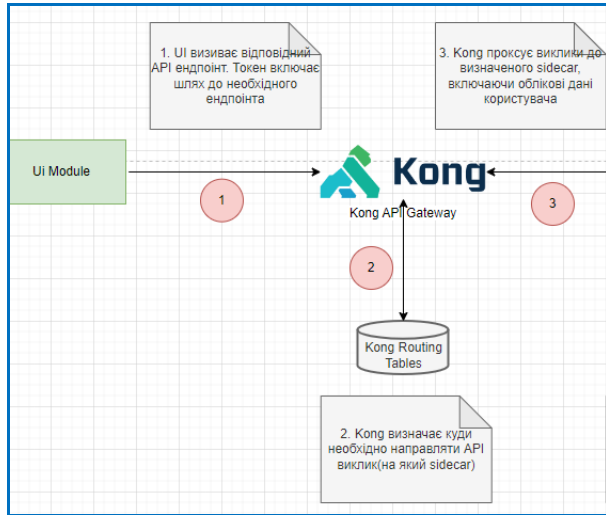


Рис. 5. Обробка запитів Арі Gateway

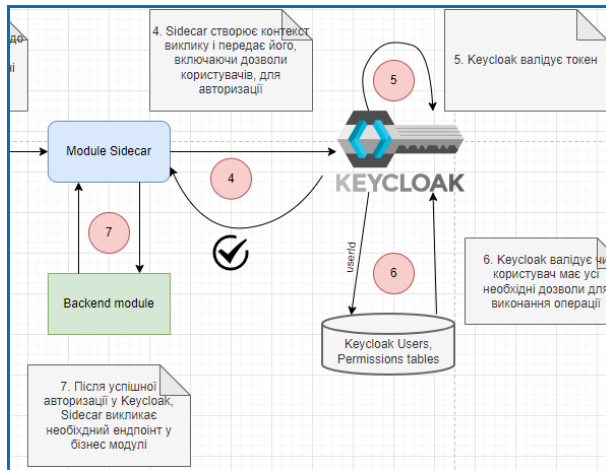


Рис. 6. Комунікація між сайдкар і Keycloak

Однак, необхідність виконання авторизації для кожного АРІ-виклику всередині сайдкара може створити значне навантаження на Keycloak, особливо у системах з високою частотою запитів. Таке навантаження може призвести до зниження загальної продуктивності авторизаційного Keycloak сервісу, що, відповідно, негативно вплине на ефективність роботи мікросервісної архітектури.

На рис. 7 представлено кількісні дані логування, які ілюструють звернення мікросервісів до відповідних сайдкарів для валідації токенів у Keycloak протягом визначеного 17-хвилинного інтервалу.

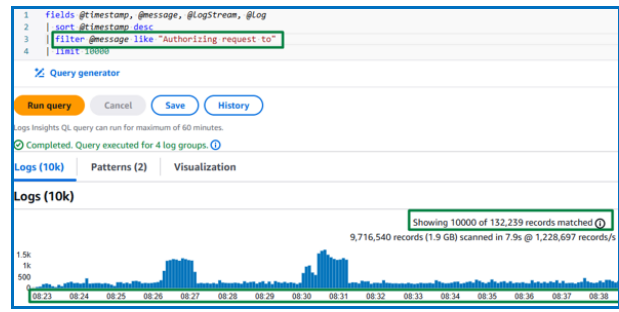


Рис. 7. Кількість звернення сайдкару до Keycloak під час пікових навантажень

Дані отримані з проведених випробувань стійкості системи, яка була розроблена.

З рис. 7 можливо стверджувати, що протягом вказаного періоду сайдкари здійснили 132,239 звернень до сервера автентифікації Keycloak. Дана статистика свідчить про значне навантаження, що накладається на Keycloak, що наочно демонструє його роль як потенційного вузького місця у системній архітектурі, особливо під час періодів не особливо високої навантаженості на інфраструктуру. Спостереження за системою під час пікових навантажень вказує на те, що кількість звернень до Keycloak може зростати кратно.

Тому, критично важливим є розробка та впровадження механізму, який дозволить зменшити кількість звернень до Keycloak із сайдкар-компонентів. Один із можливих підходів – є використання кешування токенів, що дозволить знизити частоту безпосередніх запитів до авторизаційного сервісу.

У даній роботі використовується локальний кеш на рівні кожного сайдкара для зберігання токенів, які часто запитуються, що, у свою чергу, мінімізує необхідність здійснення зовнішніх HTTP-запитів при перевірці авторизації.

Діаграма компонентів такого метода надана на рис. 8 та включає наступні кроки:

1. Користувач ініціює запит до мікросервісу 1. При цьому, запит перехоплюється відповідним сайдкардом.

2. Сайдкар 1 аналізує, чи містить локальний кеш значення вхідного токена. У випадку знаходження відповідного токена у локальному кеші, валідація вважається успішною, і звернення до служби Keycloak може бути повністю уникнуте.

3. За умови, якщо локальний кеш не містить вказаного токена, сайдкар ініціює HTTP-запит до АРІ Keycloak для проведення валідації токена.

4. Після перевірки токена сайдкар 1 здійснює виклик іншого модуля 2 для реалізації відповідної бізнес-логіки. Такий процес перехоплюється сайдкардом 2, який повторює логіку, подібну до застосованої сайдкардом 1. Важливо підкреслити, що кожен сайдкар має власний локальний кеш. Тому, присутність токена у кеші одного сайдкара не гарантує його наявності у кеші іншого сайдкара.

Зазначені кроки інтеграції вимагають особливої уваги до деталей реалізації, адже некоректне управління кешем може створити потенційні вразливості,

які дозволяють зловмисникам обходити процедуру валідації кешу. Одним з аспектів безпеки, що вимагає уваги, є установка часу життя (time to live, TTL) для записів у кеші. Правильне налаштування TTL є критично важливим для інвалідації застарілих або недійсних токенів. Також, важливо забезпечити вилучення токенів з кешу при видаленні користувачів, щоб запобігти несанкціонованому доступу до системи з використанням старих токенів.

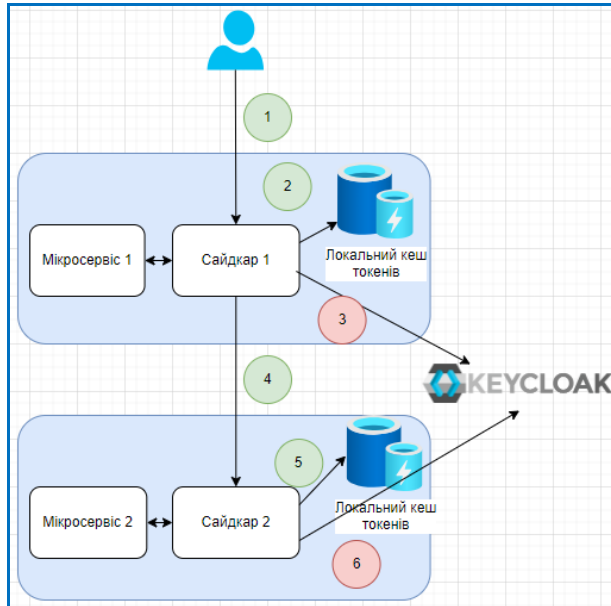


Рис. 8. Діаграма кешування авторизаційних токенів

У процесі реалізації локального кешування токенів важливо визначити необхідний розмір об'єктів, які будуть зберігатися у пам'яті. Розмір токена у Keycloak може варіюватися у залежності від кількості клеймів (claims) та інших атрибутів, які включені до токена. Зазвичай, JWT (JSON Web Tokens), які використовує Keycloak для авторизаційних токенів, мають розмір від кількох сотень байтів до кількох кілобайтів.

Припустимо, що середній розмір токена становить приблизно 2 КБ (2048 байтів). При цьому, розмір локального кешу для 100 токенів буде складати 200 кб. Цей показник є порівняно невеликим, що свідчить про мінімальний вплив на загальні рекомендації з виділення пам'яті для сайдкара.

Слід зазначити, що локальні кеші також знижують ризики збоїв системи, що можуть виникнути у випадку виходу з ладу централізованих кеш-систем, оскільки сайдкару не має необхідності звертатися до сторонніх систем кешування. Порівнюючи представлене рішення з централізованими системами кешування такими як Redis або Memcached, локальний кеш простіший для реалізації, оскільки не потребує додаткових компонентів, які потім треба ще додатково підтримувати.

Також, використання ізольованих кешів у рамках кожного сайдкара зменшує шанси на несанкціонований доступ до даних з інших частин системи, таких як новий компонент з розподіленим кешом, який містить чутливі дані.

Таким чином, реалізація локального кешу у кожному окремому екземплярі сайдкара без необхідності їх синхронізації представляє собою ефективне та доцільне рішення у високопродуктивній реалізації сайдкара.

Щодо вибору бібліотеки локального кешу слід звернути увагу на Caffeine. Вона оптимізована для високої продуктивності та мінімальних затримок доступу до кешованих даних. Функціонал Caffeine включає можливість управління кешем з використанням алгоритмів LRU (Least Recently Used), які виконуються ефективно завдяки оптимізаціям на рівні компілятора JVM. Ефективне управління пам'яттю та асинхронне API дозволяють уникнути блокувань і конкуренції між потоками, що мінімізує вплив кешу на загальну чутливість системи до затримок.

Згідно з дослідженням [9] Caffeine демонструє значні переваги у швидкості вставлення даних порівняно з іншими популярними бібліотеками кешування, такими як Guava або ConcurrentHashMap.

На рис. 9 показано порівняльний графік запису у кеш між Caffeine, ConcurrentHashMap та Guava.

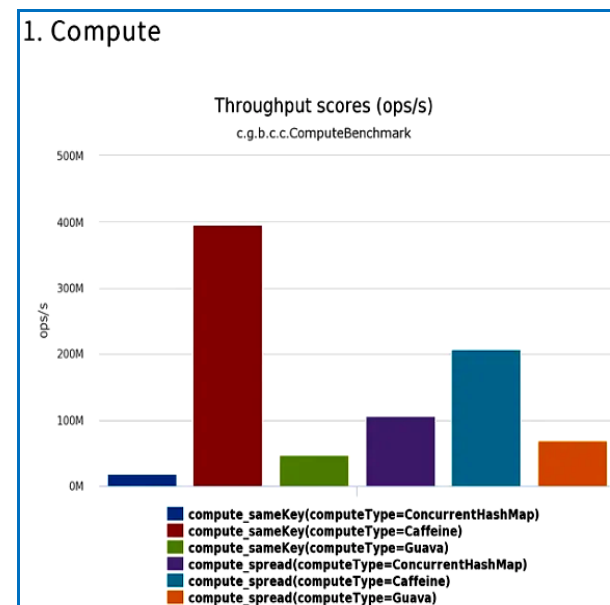


Рис. 9. Переваги Caffeine при запису у кеш

Інтеграція локального кешу Caffeine у сайдкар для зберігання токенів авторизації викликає значне підвищення загальної продуктивності системи. Використання Caffeine кешу дозволяє суттєво зменшити затримки, що пов'язані з постійними зверненнями до сервера автентифікації таких як Keycloak та суттєво зменшити навантаження на них, забезпечуючи відчутне підвищення швидкодії при обробці запитів.

Для перевірки продуктивності реалізованого рішення використано Gatling фреймворк [10]. Проведено оцінку продуктивності під різними видами навантажень:

- вхідні запити з JWT токеном користувача;
- вхідні запити з системним JWT токеном;
- неавторизовані вихідні запити.

Кожен тест можливо налаштувати, використовуючи профіль у файлі .conf.:


```

ingress-10u-300s-50s {
  baseUrl = "http://localhost:19021"
  tracingEnabled = true
  rampUpUsers = 10
  rampUpDuration = 1s
  testDuration = 300s
  tokenRefreshInterval = 60s
  config {
    type = "ingress"
    includeUserJwt = true
    includeSystemJwt = true
    requestsFile = "ingress-requests.json"
  }
}

```

Для тесту сайдкар був максимально ізольованим від інших системних служб. Усі запити на зовнішні служби та сервіси були імітованими за допомогою інструменту Wiremock. Це необхідно, щоб точно оцінити саме продуктивність окремого компоненту, сайдкара у нашому випадку, від оцінки всієї системи в цілому. Для даного тесту Keycloak містив тенант з 100 користувачами та 10 випадковими ролями.

Ресурси сайдкар компонента для різних тестів виділялися у таких діапазонах:

CPU Reservation: 128 / 256 / 512 / 1024 (0.125 / 0.25 / 0.5 / 1.0 vCPU);

CPU Limit: 128 / 256 / 512 / 1024 (0.125 / 0.25 / 0.5 / 1.0 vCPU);

Memory Reservation: 192 MB;

Memory Limit: 224 MB.

Для мікросервісу, з яким пов'язаний сайдкар:

CPU limit: 1024 (1.0 vCPU);

Memory Limit: 400 MB;

Response Delay: 50 / 200 / 1000 ms;

Request Body: (jsonObject with key: 10 random chars, value: 100 random chars), number of keys = 50 / 500 / 5000.

Для сервера авторизації Keycloak:

CPU Limit: 3072 (3.0 vCPU);

CPU Reservation: 2048 (2.0 vCPU);

Memory Limit: 1200 MB;

Memory Reservation: 1200 MB.

Тест проводився протягом 600 секунд з вико-

ристанням 100 віртуальних користувачів (V_USERS). У якості параметру RAMP_UP використовувалося значення у 550 секунд. Це означає, що V_USERS будуть поступово додаватися до тесту протягом 9 хв. та 10 с.

У даному випадку крива навантаження буде рівномірно зростати, доки не буде досягнуто повного об'єму одночасних користувачів. Таке плавне збільшення навантаження може допомогти виявити потенційні проблеми зі стійкістю та продуктивністю системи при поступовому зростанні користувачів, як представлено на рис. 10.

Параметри першого тесту Gatling, виконаного у рамках дослідження, включали затримку у 50 мілісекунд (ms), виділення обчислювальних ресурсів на рівні 0,1 віртуального центрального процесора (vCPU), налаштування одного обробника подій (Event Loop) та використання п'яти робочих потоків (Thread Workers). Такі параметри були обрані для тестування впливу конфігурації сервісу на його здатність обробляти навантаження та забезпечувати стабільну роботу в умовах різного рівня запитів.

У табл. 1 представлено результати вхідних викликів, які оброблялися сайдкарком для проведеного тесту 1.

Загальна кількість запитів склала 27,713, з яких усі були успішно оброблені без помилок (0% помилок). Середня інтенсивність запитів становила 46,19 запитів за секунду.

Мінімальний час на відгук склав 19 мілісекунд. Медіанний (50-й перцентиль) час відгуку становить 5360 мс, а час відгуку для 75-го перцентилля – 8230 мс. Час відгуку для 95-го перцентилля дорівнював 11596 мс, а для 99-го перцентилля – 12309 мс.

Максимальний зафіксований час на відгук склав 13851 мс, середній час відгуку – 5779 мс, зі стандартним відхиленням 3293 мс.

Отримані результати свідчать про високу продуктивність та стабільність сайдкару при обробці HTTP запитів.

Для наступного тесту 2 були збільшені процесорні потужності для сайдкару з 0,1 vCPU до 0,5 vCPU (табл. 2).

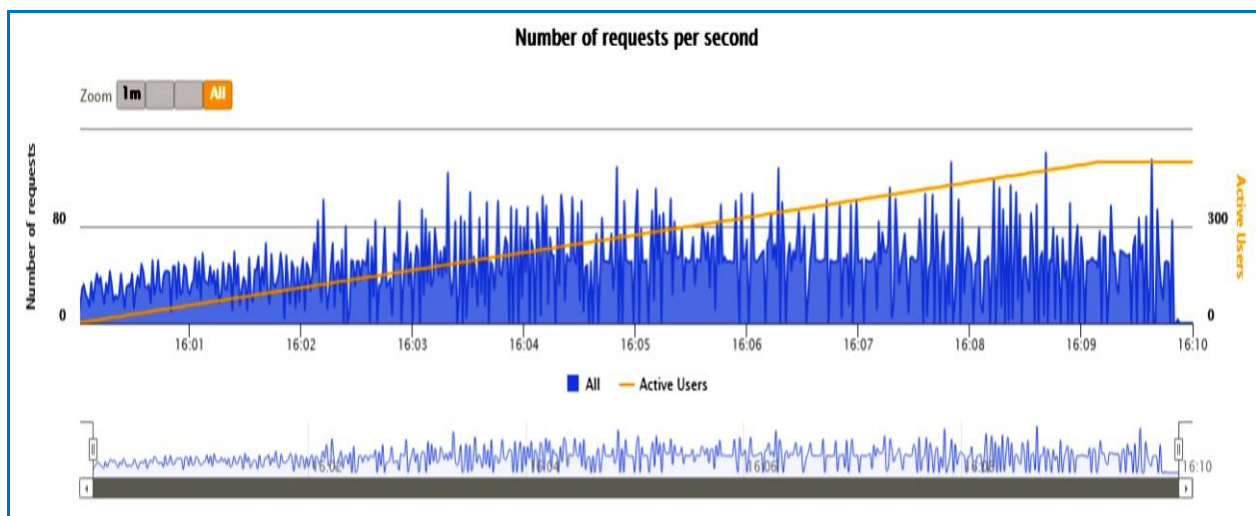


Рис. 10. Графік збільшення кількості одночасних запитів під час тесту

Таблиця 1 – Тест №1 з 0,1 vCPU

Requests	Executions					Response Time (ms)							
	Total	OK	KO	% KO	Cnt/s	Min	50th pct	75th pct	95th pct	99th pct	Max	Mean	Std Dev
All Requests	27713	27713	0	0	46,19	19	5360	8230	11596	12309	13851	5779	3293
Authorize user	1141	1141	0	0	1,9	19	25	39	70	123	164	34	20
GET request	6096	6096	0	0	10,16	63	5611	8511	11605	12241	13354	6072	3149
POST request	4821	4821	0	0	8,04	56	5577	8385	11580	12204	13305	6013	3139
DELETE request	10703	10703	0	0	17,84	55	5534	8436	11508	12165	13851	6004	3132
PUT request	4952	4952	0	0	8,25	55	5616	8335	11607	12156	13305	6030	3123

Таблиця 2 – Тест №2 з 0,5 vCPU

Requests	Executions					Response Time (ms)							
	Total	OK	KO	% KO	Cnt/s	Min	50th pct	75th pct	95th pct	99th pct	Max	Mean	Std Dev
All Requests	176426	176426	0	0	294,04	19	707	1407	2292	2608	3411	920	704
Authorize user	1291	1291	0	0	2,15	19	31	48	110	185	271	44	32
GET request	40164	40164	0	0	66,94	52	713	1403	2310	2618	3313	927	702
POST request	31714	31714	0	0	52,86	53	706	1408	2288	2584	3190	924	705
DELETE request	70577	70577	0	0	117,63	52	700	1402	2286	2587	3411	925	702
PUT request	32680	32680	0	0	54,47	52	715	1405	2229	2580	3206	930	701

Після аналізу отриманих результатів другого Gatling тесту сайдкару з конфігурацією 0,5 vCPU, можливо стверджувати, що збільшення процесорних ресурсів має значний вплив на продуктивність сайдкару.

Істотні зміни спостерігаються у наступних параметрах:

1. Загальна кількість запитів за той самий час значно зросла з 27,713 до 176,426.

2. Інтенсивність запитів збільшена з 46,19 до 294,04 запитів за секунду.

3. Середній час на відповіді значно покращився з 5779 мс до 920 мс.

4. Мінімальний, медіанний (50-й перцентиль), 75-й перцентиль, 95-й перцентиль та 99-й перцентиль часу на відповіді значно покращились, вказуючи на швидшу обробку запитів.

Дані зміни демонструють, що сайдкар збільшив свою продуктивність та здатність обробляти більшу кількість запитів за одиницю часу при збільшенні віртуальних центральних процесорів. Таке свідчить про те, що сайдкар ефективно використовує додаткові процесорні ресурси. Це стає ключовим аспектом для оптимізації продуктивності у виробничих середовищах, де вимоги до швидкодії та масштабованості постійно зростають.

За результатами аналізу другого Gatling тесту сайдкару з конфігурацією 0,5 віртуального центрального процесора (vCPU), можливо стверджувати про наступні ключові пікові значення, які вказують на оптимальну продуктивність:

- максимальна кількість запитів за секунду: загальна інтенсивність запитів досягла 294,04 запитів за секунду. Таке значення вказує на те, наскільки ефективно сайдкар може обробляти велику кількість вхідних запитів;

- мінімальний час на відповіді зафіксовано для авторизації користувача – 19 мс.

Висновки

Таким чином, використання open-source рішення Keycloak для делегації авторизації у сайдкарі є ефективним рішенням як з погляду безпеки, так і з точки зору оптимізації продуктивності.

Централізація авторизації у сайдкарі дозволила зменшити навантаження на мікросервіси, звільняючи їх від необхідності обробляти авторизаційну логіку та дозволяючи сфокусуватися на виконанні бізнес-логіки.

Також, розміщення авторизації у сайдкарі сприяє консолідації безпеки, оскільки уся інформація про авторизацію та її логіка управління концент-

руються у одному місці. Використання Keycloak дає змогу легко інтегрувати сторонні постачальники ідентичності через стандартизовані протоколи, такі як OpenID Connect або SAML. Це надає більшу гнучкість та розширює можливості щодо забезпечення безпеки без значних зусиль з розробки на стороні мікросервісів.

Завдяки комплексному підходу щодо оптимізації у конфігурацію та архітектуру сайдкару, можливо досягти значного підвищення продуктивності. Отримані дані тестувань Gatling підтвердили результати.

Одним з ключових внесків у оптимізацію є рішення про кешування токенів авторизації Keycloak.

Використання бібліотеки Caffeine для таких цілей дозволяє значно скоротити затримки, що пов'язані з частими запитами на авторизацію.

За результатами впровадження розроблених пропозицій з боку оптимізації встановлено, що сайдкар не тільки значно покращує свою продуктивність, але і демонструє здатність щодо ефективного масштабування для задоволення зростаючих вимог сучасних веб-додатків та сервісів.

СПИСОК ЛІТЕРАТУРИ

1. K. Salah, R. N. Calheiros, and R. Buyya, "Security challenges in microservice architectures: A comprehensive survey," *IEEE Trans. Cloud Comput.*, vol. 9, no. 3, pp. 1185-1206, Jul. 2019.
2. Документація Keycloak: <https://www.keycloak.org/documentation.html>
3. S. Nkomo, "Managing API credentials for microservices security," *J. Network and Systems Management*, vol. 28, no. 2, pp. 345-367, Apr. 2020.
4. K. Ueda, T. Fujibayashi, and H. Suzuki, "Role-Based Access Control in Microservice Architectures," *Journal of Information Security and Applications*, vol. 56, pp. 102-114, Mar. 2021.
5. D. Wang, H. Jiang, and L. Meng, "Secure Authorization Mechanism Using Sidecar Architecture in Microservices," in *Proceedings of the Symposium on Applied Computing*, pp. 1522-1529, March 2021.
6. Y. Zhang, P. Li, and J. Xu, "Enhancing Microservices Security via Sidecar Proxies: A Case Study with Envoy," *IEEE Access*, vol. 10, pp. 23452-23463, 2022.
7. С. С. Бульба, О. В. Коломійцев, О. І. Соловійова, С. В. Носко. Засоби побудови додаткового рівня системи комунікацій у мікро-сервісній архітектурі. Грааль науки : міжнар. наук. журнал. – Вінниця : ГО «Європейська наукова платформа»; НУ «Інститут науково-технічної інтеграції та співпраці», 2024. – No 46. – 651-659 с. DOI 10.36074/grail-of-science.
8. Kuchuk, N., Shiman, A., Filonenko, A. and Bulba, S. 2021. Розрахунок ефективності використання обчислювальних ресурсів самовідновлювальної комп'ютерної системи. Системи управління, навігації та зв'язку. Збірник наукових праць. 3, 65 (Вер 2021), 92-95.
9. DevOps блог: "Caffeine Cache". URL: <https://blog.devops.dev/easy-to-use-caffeine-cache-1-3db5861f6f39>
10. Gatling фреймворк: веб-сайт. URL: <https://gatling.io/>

Received (Надійшла) 29.11.2024

Accepted for publication (Прийнята до друку) 19.02.2025

Suggestions for authorization in the sidekick component of microservice architecture

S. Nosko, S. Bulba, O. Kolomiitsev, D. Lysytsia, H. Molchanov

Abstract. The article develops proposals on the feasibility of implementing authorization in the sidekick component of a microservice, which allows separating business logic from authorization tasks, logging, caching, etc. Such an implementation allows the main business modules to remain focused exclusively on their business logic, changing only when there are updates to business procedures, and tasks that are common to almost any modern solution to be transferred to auxiliary components such as a sidekick. Proper implementation of authorization is a key aspect of any system, it is very important and requires taking into account a large number of features and using best practices to ensure the safety of data from unauthorized access. The possibilities of the open-source Keycloak solution are considered. This solution is a very popular service for authentication/authorization, it supports standard identification protocols, such as: OpenID Connect, OAuth 2.0, SAML 2.0. The solution supports multi-factor authentication, supports containerization, and can be easily distributed in a microservices environment like Kubernetes and integrated with various external services, including Google, Facebook, etc. The architecture and proposals necessary for integration with the sidekick component are developed. A study on the performance of the developed solution was conducted. Numerical values were obtained and presented in tables and graphs. Additionally, it can integrate with various external services, including Google, Facebook, and others. The article presents the architecture and method required for integration with a sidecar component and includes a performance analysis of the proposed solution. To enhance the performance further, the sidecar architecture incorporates a local caching mechanism, utilizing Caffeine cache — known for its high performance and effectiveness. This cache stores authorization tokens, significantly decreasing the volume of calls required to Keycloak by reusing the tokens stored locally until they expire. This mechanism reduces reliance on central Keycloak servers and minimizes latency, providing a performance boost through decreased network traffic and accelerated token retrieval processes. Extensive performance testing was conducted using the Gatling framework to validate the integration and caching strategy. These tests demonstrated that the sidecar configuration, equipped with the local Caffeine cache, maintained consistent performance under varying loads and was capable of horizontal scaling without degradation in response times. Moreover, the reduced load on the Keycloak servers showcased the effectiveness of the caching approach in minimizing backend calls for token validations. The approach confirms the robustness and scalability of the sidecar, poised to handle increased loads efficiently while safeguarding sensitive authorization data.

Keywords: Sidecar, microservices, authorization, Keycloak, performance, cache, containers, Kubernetes, scalability, open-source, Gatling, resource allocation, load balancing, web server, information system.

Artem Protsenko, Volodymyr Fedorchenko

Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

MOBILE APPLICATION SECURITY ANALYSIS MODEL BASED ON ARTIFICIAL INTELLIGENCE

Abstract. The article considers the possibilities of using neural networks to ensure a secure environment for using devices. It reviews several neural network architectures that are already used to prevent attacks by attackers, the main areas of attack on mobile applications, and learning algorithms. It describes the features of using recurrent neural networks to analyze the dangerous space. The results of the article show that neural networks can be an effective tool for preventing data loss and hacker attacks. However, further research is needed to optimize the architecture and parameters of neural networks to improve the accuracy of threat detection.

Keywords: neural network, ML, DL, ANN, mobile application, cyber threat, firewall.

Introduction

Statement of the problem. The use of mobile devices has recently become an integral part of our lives. Today, they contain a lot of private information, such as banking data or documents. Many applications do not have proper data protection systems that can give the user confidence in the confidentiality of data. One of the possible solutions may be neural networks. The development of the field of artificial intelligence can help in the development of security methods for any programs used everywhere.

Analysis of recent research and publications. In the work of Dragutsov R., Rabchun D., Brzhevska Z. [1], the main vectors of attacks on information systems where Android OS applications are used as client interfaces are considered, compared and analyzed. This analysis is carried out in order to obtain initial material for the creation of practical methods for ensuring security at the architecture level of such systems. The article divides into categories possible attacks and vulnerabilities that underlie them, in the context of Android application security and taking into account the security model of the operating system itself and the environment.

In the work of Sosnovy V.O. and Zamriy I.V. [3], the effectiveness of recurrent neural networks for protecting information systems in cyberspace was investigated. The experiment showed that RNN works much better than basic machine learning algorithms. This is possible due to the built-in memory of RNN, where it is possible to remember several previous states and implicitly separate characteristic features, hidden complex structure and complex sequential communication within the data, which helps to achieve the best accuracy.

In the work of Ivanichenko Yu., Sablina M., Kravchuk [4], the main machine learning technologies that can be implemented in the cybersecurity organization are formulated. The main type of artificial neural network that will be used to prevent and detect cyber threats is also described, and it is established that the main machine learning technology for general application is artificial neural networks based on a multilayer perceptron with backpropagation of errors.

In the work of Lakhno V., Erbolat K., Bagdat Yu., Kryvoruchko O., Desyatko A., Tsyutsyura S., Tsyutsyura M. [7] a new approach to improving the information security of the educational institution network is proposed.

The proposed approach is structured and systematic. It allows to assess the degree of security of the educational institution network as a whole, as well as its individual subsystems and individual components that ensure the information security of the educational institution. Heuristic, expert, statistical and other indicators are used to assess the security of the system. The proposed model will allow to describe the process of ensuring the information security of the university network. The authors proposed a balanced system of IS indicators, which allows to assess the effectiveness of methods for ensuring the security of the university network in real time.

Purpose of the article. Creating a model for analyzing the security of a mobile application based on artificial intelligence technology. Review of methods for improving vulnerability recognition using neural network tools. Review of the use of neural networks of various architectures to prevent data leaks through unverified services, and the destruction of software design due to viruses entering the system.

Presentation of the main material

Every app store user data. It is this data that users try to hide from attackers when using their devices. This data should be properly protected so that other apps cannot access it, but this protection is not always provided.

JNI. Java Native Interface (JNI) is a mechanism in Java that allows you to interact with code written in other programming languages, such as C or C++. This is especially useful in cases where you need to call functions written in a low-level language or use libraries written in C languages.

Security flaws in the Java Native Interface (JNI) in Android applications can have devastating consequences, leading to vulnerabilities, attacks, or unauthorized access to some critical device resources. Here are some types of issues and how to fix them:

- Security vulnerabilities in the code C/C++:

Impact: The threat of memory leaks, buffer overflows, and unsafe function calls may lead to uncontrolled application behavior, including the possibility of malicious code execution.

- Incorrect handling of pointer data:

Impact: Faulty pointers or incorrect memory operation may lead to confidential data leakage or device damage.

- Insufficient permission checking:

Impact: Using insufficient permission checks may lead to unauthorized access to functions or resources, resulting in security vulnerabilities.

- Leakage of confidential data:

Impact: Incorrect handling of sensitive data during transfer between Java and C/C++ code may lead to data leakage, with serious implications for user privacy.

- Insecure call from Java to C/C++ code:

Impact: Lack of proper control and validation when calling functions from Java may lead to unpredictable behavior or vulnerability attacks.

- Using untrusted libraries:

Consequence: Using libraries with security vulnerabilities can compromise the entire application.

To minimize risks and increase security when using JNI, it is important to follow security programming best practices, carefully review C/C++ code, use secure functions, and implement access control and permission checks.

Danger areas of mobile applications. When analyzing mobile applications, first of all, it is necessary to outline the main areas of attack that can be carried out on the following systems:

- on server-side processes.
- on the client application.
- on the communication channel.

Server-side attack vectors. The information system, which involves the use of a mobile application as a client interface, contains a server part that implements the main part of the functionality free of charge.

The processes of the server part of the program are understood as any processes in the information system that are provided by functional, implemented server software. Such elements include:

- authentication, authorization and access delimitation.
- session control.
- business logic.
- data validation.
- error handling.

Attack vectors on the client application. Attacks carried out on a client mobile application within the studied architecture can primarily affect data that is processed locally. The main functional elements specific to the client application include:

- user interface.
- interprocess communication.
- local data storage.

Attack vectors on the communication channel. Many modern client-server applications for Android OS require HTTP communication protocol. Using this protocol without additional protection can be considered

a serious vulnerability of the system, as there is no way to ensure the confidentiality and integrity of information during transmission.

To provide these features, a secure connection protocol - HTTPS is used. However, given the fact that mobile devices mostly operate in an untrusted environment that can be controlled by a potential attacker, the possibilities of using the regular HTTPS protocol are limited.

From this, four possible types of attacks on the communication channel can be distinguished:

- MitM attack on an unsecured communication channel (HTTP).

- MitM attack with DNS server substitution on the network on a secure communication channel (HTTPS).

- exploitation of vulnerabilities in the implementation of SSLPinning technology to attack a secure channel.

- cryptographic attacks on a secure communication channel.

Neural networks. Machine learning (ML) is the process by which machines learn from information they are given, building logic and predicting a certain output for a given input. There are three types of ML: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning uses a dataset labeled with a correct answer symbol to learn from. These labels define characteristics for each dataset. Once the model has completed its self-learning, it can begin to make predictions or decisions about new data or situations that are subsequently fed to it.

In unsupervised learning, there is no need for such a defined data set. When a model is given a data set, it automatically finds patterns and relationships, creating clusters within them. Unfortunately, this type of learning cannot predict anything. When new data is added, the model assigns it to one of the existing clusters or creates a new one. Reinforcement learning is the ability of a system to interact with its environment and determine the best results.

The system is “rewarded” or “punished” with a score for correct or incorrect answers, and based on the positive reward results obtained, a model is automatically formed. Similarly, after training, an artificial neural network is prepared to predict new data that is given to it. Deep learning (DL) is a class of ML algorithms that uses multiple layers to gradually extract higher-level features from the original input.

The main differences between ML and DL are ML algorithms almost always require structured data, while DL networks rely on an artificial neural network (ANN) layer. Quite often, ML requires human intervention to produce further results with a larger data set, while DL does not. One of the core concepts of DL is ANN.

ANN is a model that is based on the principle of organization and functioning of the human brain (i.e., a network of nerve cells in a living organism). In other words, a neural network algorithm attempts to create a function that maps input data to desired results. Neural networks (NNs) are typically organized into layers (Figure 1). Layers consist of a set of interconnected

“nodes” that contain an “activation function.” Patterns enter the network through an “input layer,” which passes data to one or more “hidden layers,” where the actual processing is done through a system of weighted “connections.” The hidden layers are then fed to an “output layer,” where the response is the output.

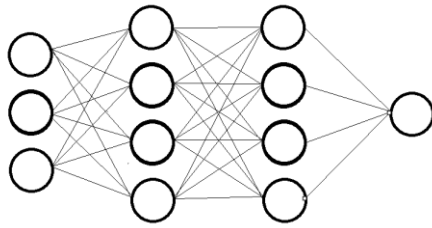


Fig. 1. Example of a neural network

Neural networks in cybersecurity. The scope of ML in cybersecurity is vast: from detecting anomalies and suspicious or unusual behavior to identifying your own vulnerabilities and fixing known ones.

For example, Reathi and Malathi published a set of ML algorithms trained on the NSL-KDD intrusion detection dataset for detecting abuse. Meanwhile, Buchak et al. focused on network intrusion detection using ML [6].

Meliher et al. proposed using NN for password strength verification. They compressed their model to hundreds of kilobytes and developed a JavaScript-like tool on the client side. For the active phase of password strength verification, they used neural networks such as Multilayer Perceptron (MLP) and Single Layer Perceptron (SLP). It should be noted that MLP provides better results than SLP when testing datasets. In addition, the number of layers is 10, and thus a better result is obtained. User and Subject Behavior Analysis (UEBA) uses the capabilities of ML to analyze behavioral patterns and network traffic in real time and respond instantly to attacks. This process can be accomplished by forcing the user to log in again, blocking the attack, or assessing the risk level and alerting the company's information security officers so that they can take the necessary measures.

Most ML and DL methods, such as clustering and decision tree learning, are used to detect abuse, anomalies, and hybrid cyber intrusions.

DARPA is working with BAE Systems to develop a system that will allow sensors to be configured and defenses to be deployed “at machine speed.” The initiative, called CHASE, stands for Cyber Hunting at Scale. It aims to create automated tools to detect and profile new attack vectors, gather the right contextual data, and distribute defenses both within and across companies [8].

The cyberattacks considered by hackers are related to the general idea of big news. Information that can be gathered from a media source can help predict similar incidents using NLP and ML methods.

We can also use ML to identify developers' programs. Rachel Greenstadt and Eileen Kaliskan have created a system that can “de-anonymize” programmers by analyzing source code or compiled binaries.

Another method of monitoring a system and network for malicious activity or policy violations is an

intrusion detection system (IDS). Another method is an intrusion prevention system (IPS) - this system works in conjunction with an IDS. These systems detect intrusions and stop the detected fraudulent activity.

Both systems use supervised and unsupervised ML techniques to detect anomalies: point, contextual, and collective. The main task of a firewall is to ensure the network security of a system by controlling incoming and outgoing network traffic. Firewalls allow or block traffic by comparing its characteristics with predefined patterns (i.e., firewall rules).

In their paper, Ukar and Ozhan presented the results of automatic anomaly detection in firewall pattern repositories based on ML and high-performance computing methods such as Naive Bayes, kNN, table solutions, and HyperPipes. Firewalls filter content between servers and are also a method specifically designed for web application content. A web application firewall (WAF) is deployed in front of web applications. Thus, it analyzes bidirectional web interface (HTTP) traffic, and then shows and blocks any malicious or illegal content. To implement such functionality in WAF, developers choose regular expressions, tokens, behavioral analysis, reputation analysis, and ML technologies.

Among ML techniques, special predictive techniques can be used for data loss prevention (DLP) to minimize the risk of hacking or leakage. Practical DLP solutions allow us to set special rules that classify confidential and non-confidential information so that it cannot be disclosed fraudulently or accidentally by unauthorized users. This process can be created using supervised learning algorithms and two types of examples: positive examples (i.e. data that needs to be protected) and counterexamples (i.e. documents like the positive set but do not need protection).

Using ML in cyberattacks. This section describes how the process of finding cyberattack results can be achieved using ML. Automated vulnerability scanning is one of the most obvious and challenging tasks in cybersecurity. For example, CSRF attack files contain only 5% of the additives, as reported in the OWASP Top 10 2017, while most frameworks include CSRF protection.

Accordingly, Calzavara et al. introduced Mitch, the first ML-based tool for detecting the same CSRF black box, which identifies 35 new CSRF vulnerabilities in 20 websites out of 10,000 Alexa websites and three more tools for previously undetected CSRF vulnerabilities in working software that has already been tested. analyzed using another state-of-the-art tool. Mitch is a binary classifier that labels sensitive or insensitive queries using a random forest algorithm in a 49-dimensional feature space. Compared to the heuristic classifiers BEAP and CsFire, Mitch demonstrated the best F1 scores and accuracy (Table 1). Trustwave released an open-source scanning tool that uses facial recognition to automatically track topics on social media. Image scanning simplifies this process, removing false positives in search results and speeding up data analysis for the human operator.

Using the data obtained about the target, the attacker can attach pre-designed fake news to the victim.

Table 1 – Validity measures of tested classifiers (BEAP, CsFire, Mitch)

Classifier	Accuracy	Response time	F1
BEAP	0.29	0.88	0.44
CsFire	0.24	0.95	0.32
Mitch	0.79	0.65	0.71

ML tools help recognize fake news, but the researchers say that the best way for ML to do this is to learn to create such fake news on its own. So, they developed a supervised text generation model called Grover. The testing process used four classes of articles: human news, machine news, human propaganda, and machine propaganda. Amazon Mechanical Turk experts rated the skin condition, including their overall reliability. In the case of propaganda, the score increased from 2.19 for the manually created article to 2.42 for the machine-generated article.

SNAP_R was introduced at DEFCON 24. SNAP_R is the world's first automated depth generator for deep-sea fishing on Twitter [6]. While previous tools are based on Markov chain models, SNAP_R is based on a finite NN with an LSTM architecture. Using Twitter as a medium offers some advantages for automated text generation. For example, it limits the length of a message, which reduces the likelihood of grammatical errors. Despite this, Twitter links are often shortened, making it very easy to mask malicious domains. This, in turn, has significantly increased the success rate of detecting such attacks from 5–14% on Markov chain-based tools to 30–66%, which supports 45% for manual fishing. In this case, attackers do not know the exact algorithm for detecting the malware but can understand certain features it uses by using developed test algorithms in the black box algorithm. MalGAN selection is a competitive network generative algorithm that generates examples of rogue malware that can mitigate black-box ML modeling. This element can minimize the level to almost zero, making it difficult for a retraining-based defense method to work against such examples. The architecture of MalGAN is shown in Fig. 2.

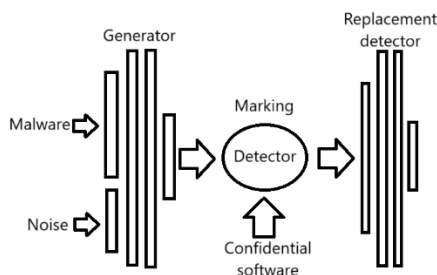


Fig. 2. MalGAN architecture

The generator uses the malware feature vector and the noise vector to transform the former into its generic version. The surrogate detector is needed to tune the black box detector and obtain gradient information to train the generator. Both networks are represented as multi-layer ANNs with direct input. The competing examples used against the black box detector come from

a variety of machine learning methods trained on 160-dimensional binary vector functions representing API system calls, logistic regression, decision trees, support vector machines, and multilayer perceptrons, and voting on the ensemble of these algorithms. Malware developers retrain detectors after learning about such undetected examples, but MalGAN requires that this entire retraining epoch reach a 0% true positive rate.

Another example of the application of GANs in cybersecurity is password attack. There is a new method for generating passwords based on DL and generative adversarial networks, known as PassGAN. The key difference of this method is that DL does not require deep knowledge of the password structure, unlike methods based on templates, Markov models, and FLA. PassGAN helps improve the training of GAN Wasserstein (IWGAN) Gularajani et al. using the ADAM optimizer. The generator and discriminator in PassGAN are based on ResNets. The architecture of the generator and discriminator is shown in Fig. 3 and 4, and the residual block image is shown in Fig. 5.

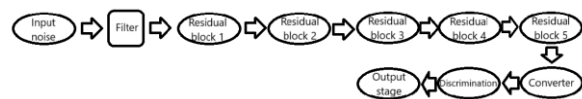


Fig. 3. Architecture of the PassGAN generator

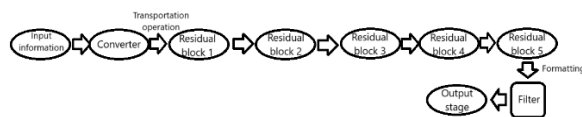


Fig. 4. PassGAN discriminative architecture

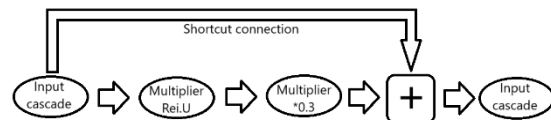


Fig. 5. Architecture of the residual block in PassGAN

To achieve maximum efficiency, attackers often choose multiple password cracking tools, such as HashCat, John the Ripper, PCFG, OMEN, and FLA, to combine various attack methods. For example, by combining raw PassGAN data with HashCat Best64 results, usability researchers found 51–73% more unique passwords than HashCat.

Traditional botnets wait for commands from a C&C, but now crooks are using automation to make decisions on their own. Fortinet developers predicted that cybercriminals would replace simple botnets with intelligent AI clusters of compromised devices, which have become a type of attack that can use self-learning to shape consumption of vulnerable systems with minimal oversight.

In the early stages of an attack, fraudsters often face the problem of missing a captcha. Suphanee et al. developed a simple attack that uses DL technology to semantically annotate an image. The system needs about 19 seconds per call to solve the tasks, with an accuracy of 70.78% for reCaptcha and 83.5% for Facebook image captcha. The system automatically selects which of the provided images are semantically like the sample images.

The system first obtains information for all images using Google ReverseImages Search (GRIS); Clarifai [6], which is built on deconvolutional networks; TDL, which is based on artificial Boltzmann machines or NeuralTalk and Caffe. Next, if no clue is found, the system searches the labeled dataset for images to obtain something like it, if possible.

Analysis model. Developing an artificial intelligence (AI)-based mobile application security analysis model is an important task due to the increasing number of mobile applications and the increasing threats to their security. Such a model can use various neural network techniques to identify potential threats and vulnerabilities in mobile applications. Here are some key aspects that can be considered when developing such a model:

- Source code analysis.
- Detection of abnormal behavior.
- Network activity monitoring.
- Malware detection.
- Vulnerability analysis in third-party libraries.
- Monitoring changes in system calls.
- Detecting social engineering attacks.
- Regular updates of rules and algorithms.

When choosing an AI training architecture, two options were considered:

- Tutored learning (perceptron).
- Unsupervised learning (adaptive resonance networks).

Studying with a teacher. Supervised learning is not suitable for this domain analysis because the perceptron has limited effectiveness when dealing with complex class boundaries and is not capable of solving nonlinear problems. This method is more suitable for simple projects. Blended learning is not suitable due to

the high complexity of the training and very long computation time. This method is more suitable for very large projects where immense data sets need to be processed.

Studying without a teacher. Adaptive Resonance Theory (ART) networks are a type of neural network that can be used in cases where the data has a nonlinear structure or when it is necessary to consider the dynamic behavior of the data. ART networks can be used for clustering and timely detection of leaks and outliers.

Conclusions

The article examines the use of neural networks to ensure security using artificial intelligence. The latest research in this field is analyzed. It is demonstrated that neural networks can be an effective tool for solving security problems. The article examines the areas of danger of mobile applications, the use of neural networks in the field of cybersecurity. The results of existing AIs that cope well with the tasks assigned to them in detecting cyber threats are considered.

Based on the research conducted, the following conclusions can be drawn neural networks can be an effective tool for detecting and preventing threats in mobile applications; it is important to choose the right neural network architecture and configure its parameters.

In further research in this area, it is advisable to study the following issues: creating our own practical model for analyzing mobile application threats based on AI. The training will be carried out in Python using the TensorFlow library. The unsupervised learning method was chosen for security analysis because it meets the requirements for creating the AI we need.

REFERENCES

1. Drahuntsov R., Rabchun D. and Brzhevska Z. (2020) "PRINCIPLES OF ENSURING SECURITY OF INFORMATION SYSTEM ARCHITECTURE BASED ON CLIENT APPLICATIONS FOR ANDROID OS". Electronic professional scientific publication "Cybersecurity: education, science, technology". No. 4 (8). P. 49-60.
2. Enck W., Ongtang M. and McDaniel P. (2009) "Understanding android security". IEEE security & privacy. No. 1. P. 50-57.
3. Sosnovy V.O. and Zamriy I.V. (2022) "NETWORK SECURITY USING A RECURRENT NEURAL NETWORK". Electronic specialist scientific publication "Word of a Scientist". No. 5. P. 21-24.
4. Ivanichenko, Y., Sablina, M. and Kravchuk, K. (2021) "USING MACHINE LEARNING IN CYBER SECURITY". Electronic professional scientific publication "Cybersecurity: education, science, technology". No. 4 (12). P. 132-142.
5. Attack indicators based on artificial intelligence allow you to predict and stop threats as quickly as possible: website. URL: <https://iitd.com.ua/news/>
6. Sharma B., Mangrulkar R. (2019) "Deep learning applications in cyber security: a comprehensive review, challenges and prospects". International Journal of Engineering Applied Sciences and Technology. No. 4(8). P. 148-159.
7. Lakhno V., Yerbolat K., Bagdat Y., Kryvoruchko O., Desiatko A., Tsiutsiura S. (2022). "Local network protection model of educational institution server virtualization system". Cybersecurity: education, science, technology. No. 2 (18). P. 6-23.
8. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin and A. Courville. (2017) "Improved training of Wasserstein GANs". In Proc. of the 31st Int. Conf. on Neural Information Processing Systems. P. 5769-5779.

Received (Надійшла) 08.11.2024

Accepted for publication (Прийнята до друку) 12.02.2025

Модель аналізу безпеки мобільних застосунків на основі штучного інтелекту

А. С. Проценко, В. М. Федорченко

Анотація. Проводиться огляд декількох архітектур нейромереж, що вже використовуються для запобігання атак зловмисників, основних напрямів атак на мобільні застосунки, алгоритмів навчання. Описуються особливості використання рекурентних нейронних мереж для аналізу небезпечного простору. Результати статті показують, що нейронні мережі можуть бути ефективним інструментом запобігання втрат даних та хакерських атак. Однак необхідні подальші дослідження для оптимізації архітектури та параметрів нейронних мереж для покращення точності виявлення загроз.

Ключові слова: нейромережа, ML, DL, ANN, мобільний застосунок, кіберзагроза, брандмауер.

O. Skakalina, A. Holub

National University «Yuriy Kondratyuk Poltava Polytechnic», Poltava, Ukraine

LARGE LANGUAGE MODELS: BUSINESS APPLICATIONS AND DEVELOPMENT PROSPECTS

Abstract. The article explores the potential of large language models (LLMs) in business. The main areas of their use are identified, including customer service, marketing and sales, internal operations, product development and innovation. The benefits of implementing LLMs are analyzed, such as increasing efficiency, reducing costs, improving service quality and stimulating innovation. The challenges and limitations associated with the use of LLMs are considered, in particular, issues of data quality, cost, security, ethics and the need for human control. The prospects for the development of LLMs and their impact on the future business environment are outlined, including hyper-personalization, automated decision-making, employee empowerment, the creation of new business models and multimodality. The conclusion is made about the significant potential of LLMs for business transformation and the importance of further research in this area.

Keywords: large language models, artificial intelligence, business, efficiency, innovation, automation, development prospects.

Introduction

The rapid development of artificial intelligence technologies, in particular large language models (LLMs), opens up broad prospects for their application in various areas of public life, including business. LLMs demonstrate impressive capabilities in understanding and generating natural language, which stimulates the emergence of innovative business models based on automation, personalization and intelligent data analysis. At the same time, there is a need to understand the potential benefits and risks associated with the implementation of LLMs, their impact on the labor market and the economy as a whole. Insufficient understanding of the limits of applicability of these technologies can lead to ineffective investments and potential losses. Large Language Models (LLM) are of great importance to large businesses, providing a wide range of opportunities to optimize processes, improve customer interactions, and make informed decisions. This article discusses the main benefits and use cases of LLM for large businesses.

Analysis of recent research and publications. The idea of creating artificial intelligence originated in the middle of the 20th century. The fundamental article in this direction was Alan Turing's "Computing Machinery and Intelligence" (1950), where the concept of thinking machines was first proposed and the Turing test was formulated. Further development of the field is associated with the work of Arthur Samuel "Some Studies in Machine Learning Using the Game of Checkers" (1959), in which the term "machine learning" was introduced and a checkers program capable of learning from its own experience was described. A revolutionary step was the creation of the first mathematical model of a neural network in the article "A Logical Calculus of the Ideas Immanent in Nervous Activity" (1943) by Warren McCulloch and Walter Pitts.

The development of reinforcement learning is associated with the work of Richard Sutton "Learning to Predict by the Methods of Temporal Differences" (1988), which introduced the temporal difference method, which became one of the main algorithms for reinforcement learning. The Transformer architecture, presented in the

article "Attention is All You Need" (2017) [1], revolutionized the field of natural language processing and became the basis for most modern LLMs.

Recently, much attention has been paid to the development of models capable of "reasoning" such as ChatGPT o1-preview from OpenAI, which can generate a chain of thoughts before issuing a response, which significantly improves their performance[2][3].

Modern scientific research pays considerable attention to studying the potential of LLMs in various business sectors, as well as analyzing the challenges that arise in the process of their integration.[4][5][6][7][8] Separately, the economic feasibility of implementing these technologies is investigated, including an assessment of the costs and benefits that accompany their deployment, and a proposal, CEBench (Fig. 1) is introduced as a method for assessing [9]. In addition, considerable attention is paid to the issues of overcoming existing limitations of LLMs, in particular, improving their productivity, adaptability and efficiency in specific business scenarios [10]. Since LLMs are systems that operate with large amounts of data, the safety of their use remains a critically important aspect. In this regard, threats associated with possible vulnerabilities of models and risks of confidential data leakage, as well as measures to minimize the impact of these risks on information security are considered [11].

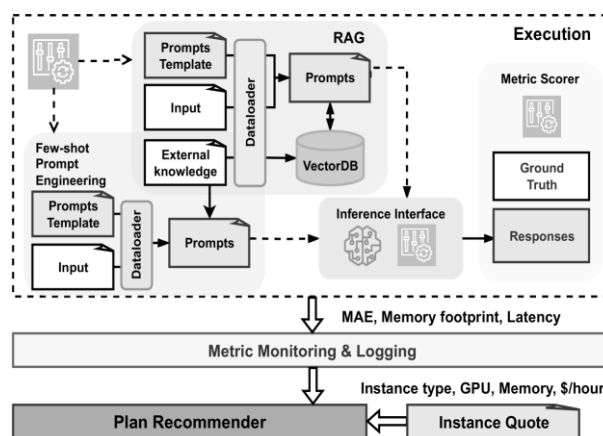


Fig. 1. Workflow of benchmarking LLM pipelines using CEBench

Research objective. The purpose of this study is to analyze the potential applications of large language models (LLMs) in business, identify the main areas of their use, as well as determine the advantages, challenges and limitations associated with their implementation. In addition, the work aims to outline the prospects for the development of LLMs and their impact on the future business environment.

Presentation of the main material

Main areas of application of LLMs in business. LLMs are increasingly being used in various aspects of business, transforming traditional approaches and opening up new opportunities for process optimization and efficiency improvement. It is useful to consider the application of LLMs in the context of key business functions.

Customer service. LLMs are revolutionizing the customer service industry by automating and personalizing interactions at various stages.

Chatbots and virtual assistants. The use of LLMs allows you to create intelligent chatbots and virtual assistants capable of conducting meaningful dialogues with customers, simulating a live conversation. These systems can not only provide reference information, but also solve complex problems, provide qualified support, place orders and perform other operations 24/7. This approach allows you to significantly reduce the load on contact center operators, optimize service costs and increase its availability and efficiency.

Feedback analysis. LLMs provide the ability to effectively analyze large amounts of text data containing customer feedback collected from a variety of sources: social networks, surveys, e-mail, forums, etc. Through deep semantic analysis, LLMs identify key topics, detect trends, determine the emotional coloring of statements (positive, negative, neutral) and isolate problematic aspects that concern customers. This allows companies to obtain valuable information for improving products and services, increasing the quality of service and responding promptly to complaints.

Personalization of offers. Based on the analysis of customer data (purchase history, views, behavioral factors, demographics, preferences), LLMs generate personalized recommendations for products, services, and special offers. This approach takes into account the individual needs and interests of each customer, which contributes to increased conversion, increased average transaction value, and strengthened brand loyalty.

Marketing and sales. LLMs open up new opportunities for automating and optimizing marketing activities and sales processes.

Content generation. LLMs are able to automatically create a variety of high-quality marketing texts: product and service descriptions, articles for blogs and websites, posts for social networks, advertisements, video scripts, email newsletters, etc. This allows content marketers to significantly save time and resources, automate routine tasks, and focus their efforts on solving strategic issues, developing creative concepts, and analyzing content effectiveness.

SEO optimization. LLMs help optimize content for search engines. They analyze current search queries, the semantic core, and create text content that is as relevant as possible to these queries, which improves the website's ranking in search results, increases organic traffic, and, as a result, increases the company's visibility in the online environment.

Market and competitor analysis. LLMs are able to analyze large amounts of data about the market, the competitive environment, the target audience, and consumer trends. They identify new market niches, determine promising development directions, identify the strengths and weaknesses of competitors, and also form forecasts about future trends. Such analytics allow companies to make strategically sound decisions regarding product development, pricing, positioning, promotion, and other aspects of marketing activities.

Internal operations. LLMs help optimize internal business processes, automate routine operations, and increase employee efficiency.

Document workflow automation. LLMs allow you to automate the processing of various documents, such as invoices, contracts, acts, reports, etc. They are able to recognize text, extract structured data from it (for example, names of counterparties, dates, amounts, product nomenclature), classify documents, automatically fill in the corresponding fields in information systems and databases, and detect errors and discrepancies. This speeds up the document processing process, minimizes the number of manual operations, and reduces the likelihood of errors associated with the human error.

Knowledge management. LLMs can serve as the basis for creating corporate knowledge bases that aggregate information from various internal sources. Thanks to semantic search capabilities, LLMs allow employees to quickly find the necessary information, instructions, regulations, answers to frequently asked questions, as well as share experience and best practices. This helps to increase productivity, reduce the time for adapting new employees and ensure the effective use of the company's accumulated experience.

Decision support. LLMs are capable of analyzing large volumes of internal company data and providing analytical insights that can be used by managers to make more informed and effective decisions. For example, LLMs can detect hidden patterns in financial indicators, predict risks, identify potential problems, and suggest solutions.

Product development and innovation. LLMs stimulate innovative activities of companies and contribute to the development of new products and services.

Idea generation. LLMs can generate new ideas for products, services, business models, marketing campaigns, etc. Using machine learning methods and analyzing large amounts of data about the market, competitors, consumer trends, and technological innovations, LLMs offer creative and unconventional solutions that can become the basis for breakthrough innovations and help companies stay ahead of their competitors.

Prototyping. LLMs can be used to speed up the prototyping process. For example, in the field of software development, LLMs are able to generate parts of code or even entire software modules based on a textual description of their functionality. In the field of design, LLMs can create sketches, mockups, and other visual elements. This allows you to reduce the time and cost of the prototype development phase and move faster to testing and implementing new products.

Testing and feedback. LLMs can be involved in the testing process of new products. They can generate test scenarios, analyze test results, and collect and process user feedback in the early stages of development. This allows them to identify potential problems and shortcomings of the product before it is released to the market, as well as take into account user wishes for its improvement.

Examples of practical application of LLMs. Business Process Management. Research shows that LLMs can extract imperative and declarative business process models from textual descriptions with accuracy that surpasses traditional approaches, opening up new

opportunities for automating business process analysis and optimization.[12]

Financial analysis. Models like GPT-4 demonstrate high performance in predicting changes in company profitability based on standardized financial statements, even outperforming professional financial analysts in forecast accuracy.

Fig. 2 compares the forecasting performance of GPT and human analysts. The naive model is based on extrapolating the change in earnings to the current change in earnings. Analyst 1m (3m, 6m) denotes the median analyst forecast issued one (three, six) months after the earnings release. GPT (without CoT) denotes GPT forecasts without any chain-of-thought prompts.

In this case, the model is provided with only structured and anonymized financial statement information. GPT (with CoT) denotes the model with financial statement information and detailed chain-of-thought prompts. Accuracy (percentage of correct forecasts out of total forecasts) for each method (left) and F1 score (right) [13].

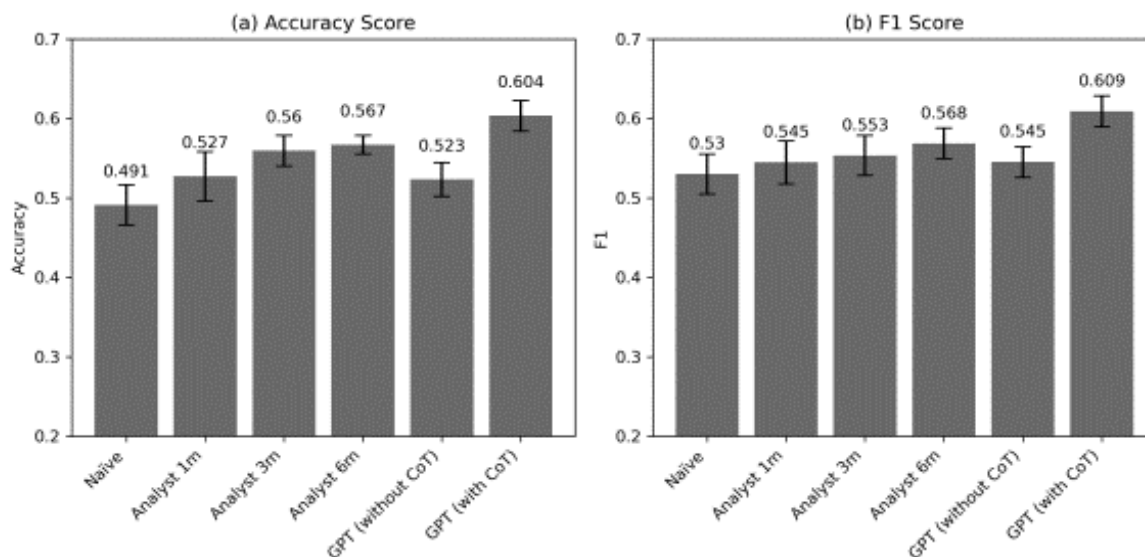


Fig. 2. GPT vs. Human Analysts

Text analysis. LLMs such as ChatGPT are effectively used for sentiment analysis in texts, for example, in customer reviews, as well as for automatic summarization of large amounts of text information and content generation [14]

Translation. LLMs provide high quality and naturalness of text translation, which is critically important for companies operating in global markets, outperforming traditional machine translation tools in these indicators.

Code generation. LLMs significantly simplify and accelerate the software development process by automating code generation and allowing you to obtain analytical data from databases using SQL queries.

Data analysis. LLM's ability to identify patterns, trends, and anomalies in large data sets opens up vast opportunities for in-depth data analysis, which contributes to making informed business decisions.

Productivity improvement. Research confirms that using LLM in consulting allows you to complete 12.2%

more tasks, and do it 25.1% faster. The quality of task performance increased by more than 40% compared to the control group (Fig. 3). [15]

Benefits of Using an LLMs in Business. Improving efficiency. Automating routine and time-consuming tasks, such as document processing, responding to customer inquiries, and generating content, allows employees to focus on more complex, creative, and strategically important tasks, which contributes to overall productivity growth.

Cost reduction. Process automation and resource optimization using LLMs lead to a reduction in operational costs associated with manual information processing, maintaining a large staff of employees, and other cost items.

Improving customer service quality. LLMs provide fast and accurate information to customers, personalized support 24/7, which leads to increased customer satisfaction and strengthening their loyalty to the company.

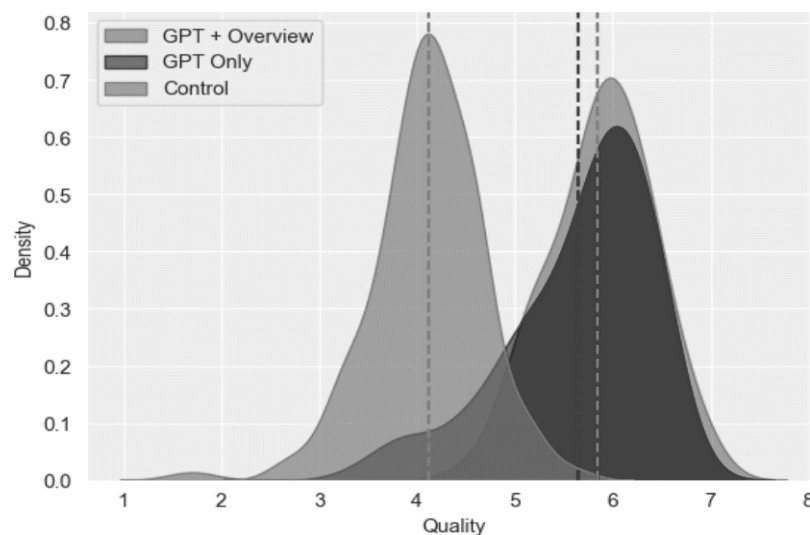


Fig. 3. Performance distribution

Stimulating innovation. Thanks to the ability to generate new ideas, accelerate the process of product development and testing, LLMs contribute to the creation of innovative solutions, which allows companies to remain competitive in a dynamic market environment.

Making more informed decisions. The analytical capabilities of LLMs, their ability to process and interpret large volumes of data, allow managers to gain a deeper understanding of the business environment, identify hidden patterns, and make decisions based on objective data rather than intuition.

Challenges and limitations. *Data quality.* The effectiveness and accuracy of LLMs is directly dependent on the quality of the data they are trained on. Using poor-quality, incomplete, outdated, or biased data can lead to inaccurate results, erroneous conclusions, and poor decision-making. Therefore, it is critical to pay attention to the quality of the data used to train and operate LLMs.

Cost. Developing, implementing, and maintaining LLMs can require significant financial investments, especially for small and medium-sized businesses. This is due to the costs of computing resources, developing and training models, attracting qualified specialists, etc.

Security and privacy issues. The use of LLMs is associated with risks of leakage of confidential information and cyberattacks. Comprehensive measures must be taken to ensure the security of data processed by LLMs and to protect systems from unauthorized access.

The need for human control. Despite the high level of automation, LLMs still require human control and oversight to ensure the accuracy, relevance, and consistency of their work results with their intended objectives. Relying entirely on LLMs to make important decisions is unacceptable.

Ethical aspects. The use of LLMs raises a number of ethical issues related to algorithmic bias, decision-making transparency, responsibility for consequences, possible misuse of technology, etc. These issues require careful consideration and regulation.

The Future of LLMs in Business. The role of LLMs in business will steadily grow in the coming years. This will be facilitated by the development of technology, the reduction in the cost of computing resources, and the accumulation of experience in the application of LLMs in various fields.

Promising directions of development and application of LLMs. *Hyperpersonalization.* LLMs will allow for even more personalized customer experiences, taking into account not only their past preferences, but also real-time contextual information such as location, mood, current needs, and intentions.

Automated decision-making. LLMs will play an increasingly important role in business decision-making, automating not only the analysis of data, but also the process of formulating recommendations and, in some cases, even taking certain actions based on those recommendations.

Employee empowerment. LLMs will become indispensable assistants for employees, providing them with quick access to the necessary information, automating routine tasks, helping to solve complex problems, and contributing to increased productivity.

Creating new business models. LLMs will open up opportunities for creating fundamentally new business models based on automation, artificial intelligence, and deep personalization, which will lead to the emergence of new markets and the transformation of existing industries.

Multimodal LLMs. Further development of LLMs will be related to the ability to process not only textual information, but also other types of data, such as images, audio and video. This will allow LLMs to understand context even better, provide more accurate and relevant results, and expand the scope of its applications.

Impact on the labor market. The introduction of LLMs will undoubtedly have a significant impact on the labor market. Some professions related to the performance of routine, repetitive tasks may be partially or fully automated. This primarily applies to professions related to the processing of large volumes of text

information, simple data analysis, providing basic customer support, etc.

However, at the same time, new professions related to the development, implementation, maintenance, and monitoring of LLMs will emerge. Machine learning specialists, data analysts, prompting engineers, AI ethicists, and other professionals with the knowledge and skills necessary to work with LLMs will be needed. Additionally, LLMs are likely to transform many existing professions, requiring workers to acquire new skills related to the use of AI tools in their work. Thus, the impact of LLMs on the labor market will be ambiguous, and it is important to prepare for these changes now by investing in education and retraining of personnel.

Conclusions

This paper analyzes the possibilities and prospects for the application of large language models in business. The application of LLMs in the following areas is considered: customer service, marketing and sales, internal operations, product development and innovation, and outlines the methods of application. Examples of practical application are also given and promising areas of development and implementation of LLMs are investigated. The speed of development of artificial intelligence necessitates continuous scientific research aimed at analyzing its functional capabilities and potential areas of application.

REFERENCES

1. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser L., Polosukhin, I. Attention Is All You Need. arXiv preprint. (2017). <https://doi.org/10.48550/arXiv.1706.03762>
2. Ahmed El-Kishky, Daniel Selsam, Francis Song, Giambattista Parascandolo, Hongyu Ren, Hunter Lightman, Hyung Won Chung, Ilge Akkaya, Ilya Sutskever, Jason Wei, Jonathan Gordon, Karl Cobbe, Kevin Yu, Lukas Kondracyuk, Max Schwarzer, Mostafa Rohaninejad, Noam Brown, Shengjia Zhao, Trapit Bansal, Vineet Kosaraju, Wenda Zhou. OpenAI. Learning to reason with LLMs. (2024) [Online]. Available:<https://openai.com/index/learning-to-reason-with-llms/>.
3. OpenAI. Introducing OpenAI o1-preview. (2024). Available:<https://openai.com/index/introducing-openai-o1-preview/>.
4. Subramanyam Sahoo, Kamlesh Dutta, Boardwalk Empire: How Generative AI is Revolutionizing Economic Paradigms(2024) arXiv preprint<https://doi.org/10.48550/arXiv.2410.15212>
5. Maxim Vidgof, Stefan Bachhofner, Jan Mendling. Large Language Models for Business Process Management: Opportunities and Challenges. (2023) arXiv preprint<https://doi.org/10.48550/arXiv.2304.04309>
6. Son The Nguyen, Theja Tulabandhula. Generative AI for Business Strategy: Using Foundation Models to Create Business Strategy Tools.(2023) arXiv preprint<https://doi.org/10.48550/arXiv.2308.14182>
7. Sai Krishnan Mohan. Management Consulting in the Artificial Intelligence – LLM Era(2024) Management Consulting Journal DOI:10.2478/mcj-2024-0002
8. Jean Kaddour, Joshua Harris, Maximilian Mozes, Herbie Bradley, Roberta Raileanu, Robert McHardy. Challenges and Applications of Large Language Models. (2023) arXiv preprint<https://doi.org/10.48550/arXiv.2307.10169>
9. Wenbo Sun, Jiaqi Wang, Qiming Guo, Ziyu Li, Wenlu Wang, Rihan Hai. CEbench: A Benchmarking Toolkit for the Cost-Effectiveness of LLM Pipelines. (2024) arXiv preprint<https://doi.org/10.48550/arXiv.2407.12797>
10. Mahei Manhai Li, Irina Nikishina, Özge Sevgili, Martin Semmann. Wiping out the limitations of Large Language Models – A Taxonomy for Retrieval Augmented Generation. (2024) arXiv preprint<https://doi.org/10.48550/arXiv.2408.02854>
11. Ayush RoyChowdhury, Mulong Luo, Prateek Sahu, Sarbartha Banerjee, Mohit Tiwari. ConfusedPilot: Confused Deputy Risks in RAG-based LLMs (2024) arXiv preprint<https://doi.org/10.48550/arXiv.2408.04870>
12. Michael Grohs, Luka Abb, Nourhan Elsayed, Jana-Rebecca Rehse. Large Language Models can accomplish Business Process Management Tasks (2023) arXiv preprint<https://doi.org/10.48550/arXiv.2307.09923>
13. Alex G. Kim, Maximilian Muhn, Valeri V. Nikolaev. Financial Statement Analysis with Large Language Models (2024) WORKING PAPER NO. 2024-65 Becker Friedman Institutehttps://papers.ssrn.com/sol3/papers.cfm?abstract_id=4835311#
14. Ming Cheung. A reality check of the benefits of LLM in business. (2024). arXiv preprint<https://doi.org/10.48550/arXiv.2406.10249>
15. Fabrizio Dell'Acqua, Saran Rajendran, Edward McFowland III, Lisa Krayner, Ethan Mollick, François Candelon, Hila Lifshitz-Assaf, Karim R. Lakhani, Katherine C. Kellogg. Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. (2023) Harvard Business School Technology & Operations Mgt. Unit Working Paper No. 24-013.https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4573321#

Received (Надійшла) 29.11.2024

Accepted for publication (Прийнята до друку) 26.02.2025

Великі мовні моделі: бізнес-застосування та перспективи розвитку

О. В. Скакаліна, А. Ю. Голуб

Анотація. У статті досліджено потенціал застосування великих мовних моделей (LLM) у бізнесі. Визначено основні сфери їх використання, серед яких обслуговування клієнтів, маркетинг та продажі, внутрішні операції, розробка продуктів та інновації. Проаналізовано переваги впровадження LLM, такі як підвищення ефективності, зниження витрат, покращення якості обслуговування та стимулювання інновацій. Розглянуто виклики та обмеження, пов'язані з використанням LLM, зокрема питання якості даних, вартості, безпеки, етики та необхідності людського контролю. Окреслено перспективи розвитку LLM та їх вплив на майбутнє бізнес-середовища, включаючи гіперперсоналізацію, автоматизоване прийняття рішень, розширення можливостей співробітників, створення нових бізнес-моделей та мультимодальність. Зроблено висновок про значний потенціал LLM для трансформації бізнесу та наголошено на важливості подальших досліджень у цій галузі.

Ключові слова: великі мовні моделі, штучний інтелект, бізнес, ефективність, інновації, автоматизація, перспективи розвитку.

А. В. Шостак

Національний аерокосмічний університет «ХАІ», Харків, Україна

ПРО ОСОБЛИВОСТІ ФОРМУВАННЯ ТРИПЛЕТІВ ДЛЯ НАВЧАННЯ СІАМСЬКОЇ НЕЙРОННОЇ МЕРЕЖІ

Анотація. Предмет дослідження – процеси розпізнавання зображень із застосуванням нейронних мереж. Додаток для розпізнавання ґрунтується на архітектурі сіамської мережі з триплетною функцією втрат і зі згортовою нейронною підмережею. **Мета статті** – обґрунтування вибору квазівипадкової N -вимірної послідовності векторів як вкладень для навчання сіамської нейронної мережі з триплетною функцією втрат та оцінка отриманих після навчання характеристик кластерів вкладень зображень цифр. **Завдання:** експериментальна оцінка отриманих після навчання характеристик кластерів вкладень зображень цифр. **Методи досліджень:** метод прямого пошуку для функцій з кількома змінними для визначення оцінок N -вимірних векторних представлень вхідних зображень. **Результати досліджень.** Проведено дослідження квазівипадкової N -вимірної послідовності векторів як вкладень для навчання сіамської нейронної мережі з триплетною функцією втрат та її тестування. Показано, що запропоновані методи визначення N -вимірних векторних представлень вхідних зображень є робастними та значно зменшують обсяг обчислень під час навчання. Під час тестування використовувалися зображення рукописних цифр із тестового набору MNIST. Визначено, що використання квазівипадкової N -вимірної послідовності векторів як вкладень зображень під час навчання сіамської нейронної мережі з триплетною функцією втрат дає змогу значно поліпшити характеристики отриманих кластерів вкладень зображень. **Висновки.** Результати, отримані в роботі, можуть бути використані при порівняльній оцінці та визначенні N -вимірних векторних представлень різних класів вхідних зображень з метою їх розпізнавання з використанням архітектури сіамської нейронної мережі з триплетною функцією втрат.

Ключові слова: сіамська нейронна мережа, триплет, вкладення, прототип вкладень кластера, тестування нейронної мережі.

Вступ

Сіамська нейронна мережа (СНМ) – це один із видів нейронної мережі, яка широко використовується в системах розпізнавання обличчя та інших графічних образів, у системах перевірки підпису, для порівняння текстів тощо [1–3].

Однією з найефективніших під час навчання СНС є триплетна функція втрат такого вигляду [4]:

$$\text{Loss} = \sum_{i=1}^k L((A, P, N)_i), \quad (1)$$

причому

$L((A, P, N)_i) = \max\{d(A, P)_i - d(A, N)_i + \text{margin}, 0\}$ – величина втрат для i -го триплету; k – розмір набору триплетів використовуваних для навчання мережі; $(A, P, N)_i$ – i -й триплет зображень, що складається з якірного зображення A , позитивного зображення P (зображення A і P належать до одного класу зображень) і негативного зображення N (зображення A і N належать до різних класів зображень), $d(A, P)_i$ – евклідова відстань між вкладеннями i -ї пари зображень A та P , $d(A, N)_i$ – евклідова відстань між вкладеннями i -ї пари зображень A та N , $\text{margin} > 0$ – параметр.

Якірні зображення A – це дані деякого класу, які визначають, на якому класі триплет буде навчати СНС модель.

Мета використання триплетної функції втрат – мінімізувати відстань між вкладеннями зображень A і P , водночас максимізуючи відстань між вкладеннями A і N , тобто зробити кластери вкладень зображень більш компактними, такими, що не перетинаються, та щоб кластери різних класів були на максимальній відстані один від одного.

Відповідно до виразу для триплетної функції втрат виділяють три види триплетів [4]:

– легкі триплети, які мають втрату, що дорівнює 0, тому що $d(A, P)_i + \text{margin} < d(A, N)_i$;

– жорсткі триплети, в яких вкладення негативного зображення N ближче до вкладення якоря A , ніж вкладення позитивного зображення P , тобто $d(A, N)_i < d(A, P)_i$;

– напівжорсткі триплети, в яких вкладення негативного зображення N не ближче до вкладення якоря A , ніж вкладення позитивного зображення P , але значення втрати все ще залишається позитивним $d(A, P)_i < d(A, N)_i < d(A, P)_i + \text{margin}$.

Зазвичай, для оптимізації якості та швидкості навчання СНС, набори триплетів із заданими властивостями (y, x, z) (де y – кількість жорстких триплетів у наборі, x – кількість легких триплетів, z – кількість напівжорстких триплетів, k – розмір набору триплетів, причому $x + y + z = k$) і вкладення для їхніх складових формують безпосередньо під час навчання. Так, наприклад, у [4] $k = 256$, $y = k/2$, $x+z = k/2$ (причому співвідношення x і z між собою чітко не задається). У [5] $k = 32$, $y = k/2$, $x+z = k/2$ (співвідношення x і z між собою також чітко не задають), причому набір із k триплетів формують із випадкового набору з 200 триплетів. У [6] $k = 256$ і гарантується, що щонайменше $z = k/2$, а решта триплетів можуть бути довільними. Тобто відповідно до [4, 5, 6] жорсткі та напівжорсткі триплети забезпечують більш якісне навчання СНС. Такий підхід до структури і складу триплетів видається не цілком ефективним з погляду обсягу обчислень, якості та швидкості навчання СНС.

Основна частина

Оскільки у функції втрат (1) використовують евклідові відстані між вкладеннями пари зображень A та P та пари зображень A та N , то під час навчання СНС обчислюватимемо вкладення якірних

зображень A та використовуватимемо оцінки прототипів вкладень класів зображень EP та EN . Тоді вираз для триплетних втрат матиме вигляд

$$L((A, EP, EN)_i) = \max\{d(A, EP)_i - d(A, EN)_i + \text{margin}, 0\}.$$

Тут зображення A і вкладення EP належать до одного класу, зображення A і вкладення EN належать до різних класів, а вкладення EP і EN є оцінками прототипів вкладень різних класів зображень.

Як оцінки прототипів вкладень EP і EN під час навчання СНС було використано квазівипадкову n -вимірну ($n=10$) послідовність із 10 векторів, генерованих за методом Соболя [3, 7].

Структуру підмережі СНС [5], що формує вкладення для вхідних зображень, наведено на рис. 1. На вхід підмережі СНС подається з набору даних MNIST [8] одноканальне зображення у градаціях сірого кольору з рукописною цифрою розміру 28×28 пікселів

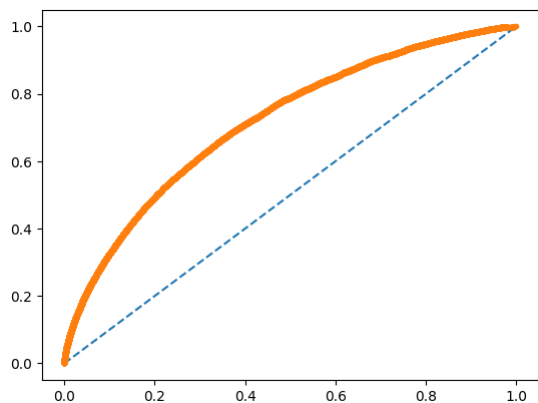
- $\text{Input}(28 \times 28 \times 1)$. Далі йде згортковий шар $\text{Conv2D}_1(128, (7, 7), \text{relu})$, який формує за допомогою ядер згортки розміру $(7, 7)$ 128 карт ознак і використовує функцію активації relu . Після першого шару згортки слідує шар підвибірки MaxPooling2D , що замінює дані у вікні їхнім максимальним значенням. Далі - другий шар згортки $\text{Conv2D}_2(128, (3, 3), \text{relu})$ з 128 картами ознак, шар підвибірки MaxPooling2D і третій шар згортки $\text{Conv2D}_3(256, (3, 3), \text{relu})$. Потім шар Flatten , на виході якого з його вхідних даних формується одновимірний вектор. Далі повнозв'язний шар $\text{Dense}(4096, \text{relu})$ з 4096 вузлів з використанням функції активації relu і повнозв'язний шар $\text{Dense}(10)$ з $n=10$ вузлів без використання функції активації для забезпечення повного діапазону значень вкладення. Вихід підмережі формується Lambda шаром, який виконує операцію L_2 -нормалізації значень вкладення з попереднього шару $\text{Dense}(10)$.

Input($28 \times 28 \times 1$) \rightarrow Conv2D_1($128, (7, 7), \text{relu}$) \rightarrow MaxPooling2D \rightarrow Conv2D_2($128, (3, 3), \text{relu}$) \rightarrow MaxPooling2D \rightarrow Conv2D_3($256, (3, 3), \text{relu}$) \rightarrow Flatten \rightarrow Dense($4096, \text{relu}$) \rightarrow Dense(10) \rightarrow Lambda(L_2 -normalize)

Рис. 1. Структура підмережі СНС

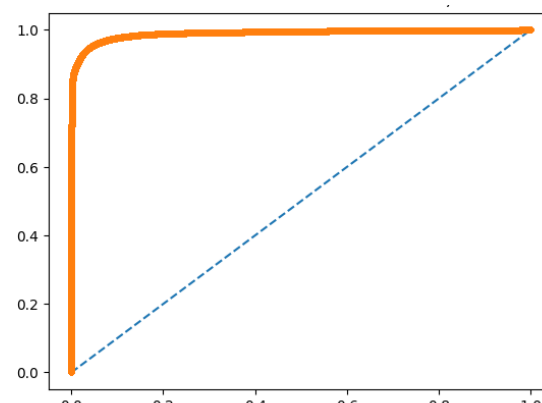
На кожній ітерації навчання формувалася набір із 32 триплетів у такий спосіб. Спочатку на підставі тренувального набору і вкладень для десяти цифр генерувався набір із 200 випадкових триплетів виду $(A, EP, EN)_i$. Потім цей набір сортувався за зменшенням різниці між відстанями $d(A, EP)_i$ і $d(A, EN)_i$ і з відсортованого набору вибирали 16 перших триплетів. Далі з решти набору з 200 триплетів вибирали ще 16

довільних триплетів. Модель було навчено на 9000 ітерацій. Для оцінювання якості моделі СНС на рис. 2 представлено розраховану на підставі вкладень тестового набору зображень ROC-криву [9] а) до навчання, б) після навчання, а також значення площі під кривою (AUC), чутливості (Sensitivity) і порогу (Threshold), який забезпечує частоту хибнопозитивних результатів (FPR) не більше ніж $10e-3$.



AUC = 0.717, Sensitivity = 1.9%, Threshold = 0.0448

а



AUC = 0.990, Sensitivity = 80.1%, Threshold = 0.2216

б

Рис. 2. ROC-крива: а – до навчання, б – після навчання

На рис. 3 показано візуалізовані з використанням алгоритму зниження розмірності t -SNE [10] десять кластерів вкладень рукописних цифр тестового набору зображень MNIST а) до навчання та б) після навчання.

Бачимо, що після навчання з використанням триплетної функції втрат, триплетів вигляду $(A, EP, EN)_i$ і підмережі СНС (рис. 1), усі десять кластерів вкладень практично не перетинаються, і відстані між вкладеннями кожного кластера стали істотно меншими.

Прототипи вкладень кластерів вкладень зображень цифр є узагальненою характеристикою кластерів вкладень відповідних цифр і використовуються для розрахунку ступеня схожості вхідного зображення із зображенням відповідної цифри.

Для пошуку прототипів вкладень зображень цифр тестового набору даних було використано метод Пауелла прямого пошуку для функції з кількома змінними з Python-бібліотеки SciPy. Прототипи вкладень зображень цифр для підмережі СНС (рис. 1) після навчання мають вигляд:

- 0) [-0.2824016 -0.27252399 -0.37384685 -0.33512156 -0.29360956 -0.30970557 -0.27883729 -0.33485432 -0.33369731 -0.3317796];
 1) [-0.05223146 -0.39773643 0.1311696 -0.11012863 -0.00831585 -0.24040615 0.46429436 0.28060366 0.45396905 0.13724699];
 2) [0.29550691 -0.24569951 -0.32850522 -0.30957372 0.33073034 0.32076124 -0.20231111 0.33361188 0.34647749 0.36148253];
 3) [-0.29281085 0.31785711 0.28628017 0.36124768 -0.35524327 -0.27347703 0.31099376 -0.32188501 -0.32414977 -0.25594271];
 4) [-0.2123403 -0.13414584 0.13346918 0.29032724 -0.09325196 -0.50832853 -0.15217541 0.50219026 0.4942752 0.1325437];
 5) [0.37764855 0.40513021 -0.38194006 -0.13066819 0.39994588 0.11103943 0.40234225 -0.10447713 -0.10416116 -0.40455939];
 6) [0.10466758 -0.38080509 0.42403376 0.16230785 0.11301583 0.3711326 -0.39282424 -0.39239505 -0.3939226 -0.12383774];
 7) [-0.40467949 0.18668864 -0.1246697 -0.40202019 -0.38693982 -0.13544805 0.19201151 0.25637364 0.26939899 0.48835121];
 8) [-0.35404521 -0.19320955 0.4521353 -0.07188427 0.08048192 -0.25373991 -0.09305534 0.48344564 0.49281425 -0.21215699];
 9) [0.17273745 0.34933134 -0.0630242 0.45769747 -0.45803124 0.2907892 0.46623946 -0.04591382 -0.03500779 0.32947503].

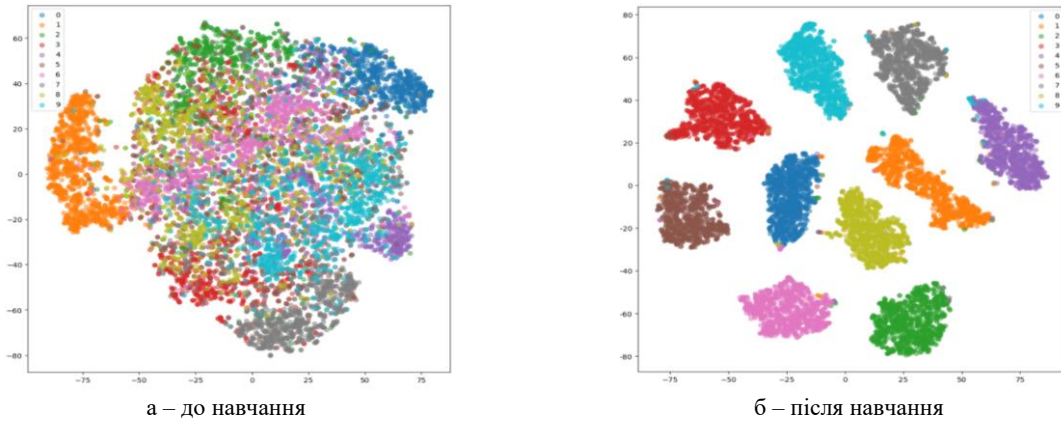


Рис. 3. Візуалізація десяти кластерів вкладень рукописних цифр тестового набору зображень MNIST

У табл. 1 наведено розраховані для тестового набору даних розміру 10000 з використанням прототипів вкладень характеристики кластерів вкладень зображень цифр - кількість зображень різних цифр у тестовому наборі (k), мінімальна відстань (r), максимальна відстань (радіус кластера) між вкладеннями кластера і прототипом вкладення (R), середня внутрішньокластерна відстань (SSR) і середнє квадратичне відхилення (sD). У табл. 1 перший рядок для кожної цифри відповідає стану СНС до навчання, другий рядок - після 9000 ітерацій навчання.

Таблиця 1 – Характеристики кластерів вкладень зображень цифр

Цифра	k	r	R	SSR	sD
0	980	0.0769	0.7303	0.3171	0.0961
		0.0043	1.1589	0.0287	0.0517
1	1135	0.0889	0.8783	0.2924	0.1082
		0.1453	1.4764	0.5182	0.2003
2	1032	0.1900	0.8640	0.4255	0.1078
		0.0454	1.4558	0.2137	0.1666
3	1010	0.09336	0.7289	0.3342	0.0905
		0.04274	1.8494	0.2048	0.1668
4	982	0.1466	0.8356	0.3705	0.1088
		0.0292	1.3727	0.2093	0.1615
5	892	0.1239	0.6659	0.3461	0.0913
		0.0218	1.5957	0.1306	0.1513
6	958	0.1645	0.7195	0.3818	0.0970
		0.0192	1.7079	0.1220	0.1641
7	1028	0.1058	0.7845	0.3298	0.0991
		0.0439	1.1062	0.2406	0.1536
8	974	0.1344	0.7767	0.3628	0.0930
		0.0344	1.4964	0.1966	0.1710
9	1009	0.0883	0.6607	0.3067	0.0874
		0.0274	1.4920	0.1363	0.1614

Відповідно до даних із табл. 1 до навчання найбільш некомпактний за показником R кластер вкла-

день зображень цифри 1 і кластер вкладень зображень цифри 2 за показником SSR . Найкомпактніший за показником R кластер вкладень зображень цифри 9 і кластер вкладень зображень цифри 1 за показником SSR .

Після навчання найбільш некомпактний за показником R кластер вкладень зображень цифри 6 і кластер вкладень зображень цифри 1 за показником SSR . Найкомпактніший за показником R кластер вкладень зображень цифри 7 і кластер вкладень зображень цифри 0 за показником SSR .

Після навчання значення мінімальної відстані r для кластерів вкладень зображень усіх цифр, крім цифри 1, зменшилися від 2,2 до 17,9 разів. Для вкладень цифри 1 значення r після навчання збільшилися в 1,6 раза. Значення максимальної відстані R для всіх кластерів вкладень цифр збільшилися від 1,4 до 2,5 разів. Значення середньої внутрішньокластерної відстані SSR для всіх кластерів вкладень цифр, крім цифри 1, зменшилися від 1,4 до 11,0 разів. Для вкладень цифри 1 значення SSR після навчання збільшилися в 1,8 разів. Значення середнього квадратичного відхилення sD для всіх кластерів вкладень цифр, крім вкладень цифри 0, збільшилися від 1,5 до 1,9 раза. Для вкладень цифри 0 значення sD після навчання зменшилися в 1,9 раза.

У табл. 2 для трьох 10-вимірних наборів вкладень розміру $n=10$ (перший набір - вектори квазівипадкової послідовності Соболя, другий набір - прототипи вкладень кластерів рукописних цифр для моделі СНС (рис. 1) до навчання і третій набір - прототипи вкладень кластерів рукописних цифр для моделі СНС (рис. 1) після навчання) наведені відповідно мінімальна відстань (rE) між вкладеннями в наборі, максимальна відстань (RE) і середня відстань ($SSRE$).

У результаті навчання на підставі розрахунків на тестовій вибірці (рядки 2 і 3 табл. 2) мінімальна відстань rE між прототипами вкладень кластерів рукописних цифр збільшилася в 3,26 рази, максимальна

відстань RE - у 2,63 раза і середня відстань SSRE - у 4,00 раза.

Таблиця 2 – Характеристики трьох наборів вкладень

N	rE	RE	SSRE
1	1.4252	4.3768	2.6435
2	0.2112	0.7458	0.3536
3	0.6885	1.9595	1.4129

Висновки

У роботі досліджено спосіб використання розрахованої за методом Соболя квазिवипадкової n-вимірної послідовності векторів для формування триплетів під час навчання СНС.

Використання квазिवипадкової n-вимірної послідовності векторів як вкладень EP і EN триплету (A, EP, EN) усуває необхідність застосування позитивного і негативного зображень P і N та обчислення підмережею СНС вкладень для них. Тобто завдяки використанню триплету виду (A, EP, EN) значно скорочується обсяг обчислень під час навчання СНС. При цьому в результаті навчання СНС на триплетах виду (A, EP, EN) за 9000 ітерацій площа під ROC-кривою

збільшилася з 0,717 до 0,99, чутливість - з 1,9% до 80,1% і поріг, що забезпечує частоту хибнопозитивних результатів не більше $10e^{-3}$, з 0,045 до 0,222.

Ефект від навчання СНС із триплетною функцією втрат і з триплетом виду (A, EP, EN) полягає в значному відокремленні та стисненні кластерів вкладень зображень рукописних цифр. Так після навчання значення мінімальної відстані для кластерів вкладень зображень усіх цифр тестової вибірки, окрім цифри 1, зменшилися від 2,2 до 17,9 разів. Значення середньої внутрішньокластерної відстані для всіх кластерів вкладень цифр, крім цифри 1, зменшилися від 1,4 до 11,0 разів. Також у результаті навчання мінімальна відстань між прототипами вкладень кластерів цифр збільшилася в 3,26 рази, максимальна відстань - у 2,63 рази і середня відстань - у 4,00 рази.

Подальші дослідження слід спрямувати на вибір адаптивних алгоритмів перерахунку векторів вкладень для формування триплетів під час навчання з метою прискорення навчання і збільшення його якості, а також для поліпшення характеристик кластерів вкладень зображень.

СПИСОК ЛІТЕРАТУРИ

1. Chicco D. Siamese Neural Networks: An Overview. Artificial Neural Networks. MIMB, vol. 2190, 2020, pp. 73-94. URL: https://link.springer.com/protocol/10.1007/978-1-0716-0826-5_3
2. Шостак А. В. Про особливості формування дескрипторів у сіамській нейронній мережі. Системи управління, навігації та зв'язку, Полтава: НУ ПП, 2021, вип. 4(66). С. 91-96. DOI: <https://doi.org/10.26906/SUNZ.2021.4.079>
3. Шостак А. В. Про особливості формування вхідних даних у сіамській нейронній мережі. Системи управління, навігації та зв'язку, Полтава: НУ ПП, 2024, вип. 3(77). С. 193-195. DOI: 10.26906/SUNZ.2024.3.193
4. Schroff F., Kalenichenko D., Philbin J. FaceNet: A unified embedding for face recognition and clustering. Proceedings of the IEEE CSC on CVPR, 2015, pp. 815-823.
5. Craeymeersch E. One Shot learning, Siamese networks and Triplet Loss with Keras. URL: <https://medium.com/@crimy/one-shot-learning-siamese-networks-and-triplet-loss-with-keras-2885ed022352>
6. Trotter C. How To Train Your Siamese Neural Network. URL: <https://towardsdatascience.com/how-to-train-your-siamese-neural-network-4c6da3259463>
7. Owen A.B. On Dropping the First Sobol' Point. In: Keller, A. (eds) Monte Carlo and Quasi-Monte Carlo Methods. MCQMC 2020. Springer Proc. in Mathematics & Statistics, vol 387. Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-98319-2_4
8. The Mnist database of handwritten digits. URL: <http://yann.lecun.com/exdb/mnist/>
9. Hernandez-Orallo J. ROC curves for regression. Pattern Recognition. 2013. vol. 46, no. 12. pp. 3395–3411. doi: 10.1016/j.patcog.2013.06.014
10. Van der Maaten L.J.P. Accelerating t-SNE using Tree-Based Algorithms. Journal of Machine Learning Research 15(Oct). – 2014. - С. 3221-3245.

Received (Надійшла) 15.12.2024

Accepted for publication (Прийнята до друку) 05.03.2025

On the features of triplet's formation for Siamese neural network training

A. Shostak

Анотація. Summary. The subject of research – image recognition processes using neural networks. The recognition application is based on a Siamese network architecture with a triplet loss function and a convolutional neural subnetwork. **The purpose of the article** – to justify the choice of a quasi-random N-dimensional sequence of vectors as embeddings for training a Siamese neural network with a triplet loss function and to evaluate the characteristics of the clusters of digit image embeddings obtained after training. **Objective:** an experimental evaluation of the characteristics of clusters of embeddings of digit images obtained after training a Siamese neural network with a triplet loss function. **Research methods:** a direct search method for functions with several variables to determine estimates of N-dimensional vector representations of input images. **Research results.** A study of a quasi-random N-dimensional sequence of vectors as embeddings for training a Siamese neural network with a triplet loss function and its testing is carried out. It is shown that the proposed methods for determining N-dimensional vector representations of input images are robust and significantly reduce the amount of computation during training. During testing, images of handwritten digits from the MNIST test set were used. It has been determined that the use of a quasi-random N-dimensional sequence of vectors as image embeddings in training a Siamese neural network with a triplet loss function can significantly improve the characteristics of the obtained image embedding clusters. **Conclusions.** The results obtained in this work can be used for comparative evaluation and determination of N-dimensional vector representations of different classes of input images for their recognition using the architecture of a Siamese neural network with a triplet loss function.

Ключові слова: Siamese neural network, triplet, embedding, prototype cluster embedding, neural network testing.

M. Shupyliuk, V. Martovytskyi

Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

ANALYSIS OF PERSONALITY DETECTION AND WRITER IDENTIFICATION METHODS

Abstract. Handwritten text as multi-sensory activity can show one's personality and at the same time can serve as one's biometric identifier. Handwriting analysis is used in various fields including history, forensic, education, security, personnel matters etc. In this article handwriting analysis methodologies were considered and categorized in four groups highlighting advantages and disadvantages of each group. Also, this article depicts various problems associated with developing handwriting analysis systems such as improper feature extraction, overfitting, underfitting, unreliable training data, picking model for assessing personality types, etc. Both methods for robust offline writer identification and methods for prediction of human personality that are used in state-of-the-art handwriting analysis systems are presented. In addition, current studies and common approaches for performance measurement and database selection in both writer identification and personality detection fields were analyzed. Also, perspective development directions of modern handwriting analysis systems are presented.

Keywords: handwriting, handwriting analysis, personality, handwriting features, graphology, personality detection, writer identification.

Introduction

Handwriting analysis systems aim to automatically extract important information from handwriting, therefore research in the field of graphology has a wide range of applications, such as writer identification, personality detection etc. M. Hengl [1], Hemlata et al. [2], Wirmanto et al. [15], Samsuryadi et al. [31] presented an overview of modern methods of handwriting analysis that are used in various fields. This paper presents some challenges in handwriting analysis systems, the motivation for further research, application areas, design approaches and issues, writer identification and personality detection methods. In addition, perspective development directions are presented.

Formulation of the problem and its connection with important scientific or practical tasks

Digital transformation is a continuous activity and affects various aspects of human life. Despite that fact handwriting remains a core human activity and considered as one of the most complex human activities. Everyone's handwriting is unique and has characteristics that are typical of the given individual which creates a basis for handwriting analysis [1]. Handwriting analysis is a scientific method for recognizing, assessing and understanding a writer's personality and identity through the shapes and word patterns in the handwriting [2].

Handwriting analysis has three major branches: graphology, handwriting expert and forensic linguistics [1]. The main fields of application for handwriting analysis are criminalistic, education, history, mental hygiene, medical diagnosis, public life, personnel matters and personality profiles. In the criminalistic field handwriting analysis is associated with crimes that involve written documents such as kidnapping, defamation, false witness, false accusations, violation of rules for accounting and forgery of public documents.

Detecting an individual based on handwriting is a kind of behavior biometric identification well acknowledged by psychologists, neurologists, paleographers, forensic analysts, document analysts, and computer science researchers. Similarly, the existence of

a relationship between handwriting and different demographic attributes of writers, such as gender, handedness, and age, is also confirmed by psychologists and neurologists [3]. So, handwriting is also called brain writing. Each personality trait has a neurological brain pattern in the human brain. Each neurological brain pattern design delivers one of a kind neuromuscular movement which is the same for each individual who has that specific personality trait. Each stroke or movement in handwriting reveals a particular personality trait [2].

Writer identification and personality detection have been studied for many years but they remain an open research problem for such reasons as diversity, similarity and high intra-variability among authors in handwriting. Digital era brings new challenges into writer identification and personality detection because now text can be handwritten not only on paper but also on digital devices. This leads to the necessity to differentiate between online and offline. In online writer identification and personality detection data contains temporal information about the text formation [4]. In contrast, offline writer identification and personality detection deals only with the handwritten text itself without any additional information [4]. This paper mainly focuses on the offline handwriting analysis process to detect the personality of a person and perform writer identification based on their handwritten documents. Section III describes research results, where subsection I describes categorized handwriting analysis methodologies, subsection II describes the related works in the field of handwriting analysis for writer identification, subsection III describes related works for personality detection, subsection IV provides discussion topics and future scope, section IV gives conclusion.

Research results

Handwriting analysis methodologies. Handwriting analysis is a complex problem so there have been many studies proposed. To categorize these studies and identify advantages and disadvantages different aspects can be used.

Considering the type of features as a functional context, the methods reported in the literature can be

categorized into four main groups: structural-based, textural-based, grapheme-based and auto-learned based methods [17].

Structural-based methods consider the allograph shapes and then apply a grapheme-emission probability distribution. For structural-based systems cost of runtime is high due to complexity of preprocessing steps which include segmentation, binarization and edge detection. Strong reliance on preprocessing is a limitation of these methods, since preprocessing step failure means feature extraction and classification steps are prone to failure. Concentration of such methods mainly on allograph shapes means that additional information drawn in the same word between allographs is omitted. Additional drawback of such methods is vulnerability to any variations in the character or allograph characteristics (i.e. slant, aspect ratio).

Textural-based (transformation-based) methods work with digitized image of handwriting which are considered as special textures and extract features (after some transformation techniques are applied) from the whole document, Regions of Interest (or ROIs like blocks, grid cells, connected-components, words, etc.) or Writing Fragments (WFs). Textural-based methods are more efficient runtime-wise, as they do not need additional preprocessing steps such as segmentation and binarization. Another advantage is parameter-free learning. On the other hand, textural-based methods usually need more data to extract reliable and highly discriminative features, which might not be the case (i.e. forensic experts to predict the authenticity have to deal with small queried handwriting samples). Additional drawback of such methods is a need for high contrast images.

Grapheme-based (bags of features (BOF)-based) methods rely on codebook generation for all character types. In such methods, handwritten text segmented into character segments and graphemes are generated for each character, after that a codebook is generated using any clustering algorithm, for each codebook histogram distance scores are calculated and the resulting histogram used for classification. Such methods are considered as effective for the writer's identification because they do not require an excessive amount of the writer's original handwriting sample to predict the authenticity of the queried handwriting sample. However, this type of methods have some drawbacks such as the need to spend much time extracting and comparing grapheme details. Additionally, such methods require a substantial amount of memory due to the large number of grapheme features, especially for methods that apply a single clustering algorithm.

Auto learned (model-based) methods rely on the usage of deep learning techniques to extract automatic features learned by deep models. Such methods are robust and report high identification rates, especially for word-level identification. Another advantage is that models automatically learn representations from data (i.e. automatic feature extraction) that's why Convolutional Neural Networks have been widely used to extract features from the images (i.e., words, pages, patches). The main drawback of model-based methods is training because a large number of samples is needed, the training could be very time consuming and models tend to overfit.

Another important drawback of model-based methods is the difficulty of selecting the best values for a large number of parameters. Additionally, this type of methods require very high computational resources and time.

Writer identification. During handwriting analysis to perform a writer identification the method of individual personal identification is used, the essence of which is the full exploration of the individual characteristics of the document being examined and the comparison of these with the known manifestations of these characteristics in other documents (comparative material) [1].

The writer identification methodologies share common/standard steps of preprocessing, feature extraction and writer identification (frequently referred as classification) as shown in Figure 1.

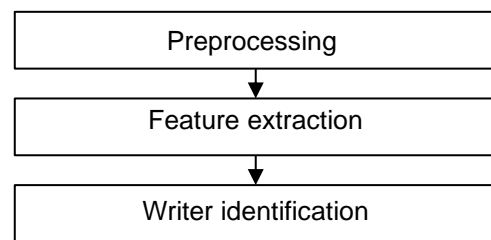


Fig. 1. Writer identification methodologies common steps

Studies in the field of writer identification also fall into four handwriting analysis methodologies categories.

In textural-based category Christlein et al. [4] proposed to use local descriptor RootSIFT to capture local properties of handwritten documents, use GMM supervectors as encoding method and use Exemplar-support vector machines (SVM) for similarity measure.

In auto learned category Christlein et al. [5] proposed use of Convolutional Neural Network (CNN) to train a powerful patch representation using cluster memberships as targets and a global image descriptor is created by means of Vector of Locally Aggregated Descriptors (VLAD) encoding. Chen et al. [6] proposed deep CNN (ResNet-50) and the weighted label smoothing regularization (WLSR) for data augmentation to allow more discriminative features to be learned to represent the properties of different writing styles. Helal et al. [8] experimented CNN DL with an additional dissimilarity approach with an SVM classifier. Sulaiman et al. [9] proposed a local binary pattern (LBP) as a hand-crafted feature descriptor and AlexNet structure for extracting deep description as a second descriptor, both descriptors then assembled into a data matrix and encoded using VLAD. A SEGmentation-free model presented by Kumar and Sharma [10] where they apply the CNN to characterize the writer of a given sample. Koepf et al. [12] presented a novel method based on vision transformer (ViT) with K-Nearest Neighbors (KNN) classifier. Semma et al. [13] presented a deep CNN trained to capture deep features in small patches which extracted from handwritten images using the FAST key points and Harris Corner (HC) detector and the deep learned features are encoded using VLAD and the Triangulation Embedding with a KNN classifier. At first He and Schomaker [7] proposed deep adaptive CNN based on multi-task adaptation of the AlexNet structure, after that He and Schomaker [11]

proposed a deep neural network (FragNet, inspired by Fraglets and BagNet) based on fragments segmented on the input document image and feature pyramid maps in the CNN, in more recent work He and Schomaker [14] proposed global-context residual recurrent neural network (GR-RNN) where spatial relationship between the sequence of fragments is modeled by the recurrent neural network (RNN) to strengthen the discriminative ability of the local fragment features and the complementary information between the global-context and local fragments is leveraged. Wirmanto et al. [15] proposed to segment data into two inputs and use Xception as feature extractor in the Siamese network, the last one generates encoded representation and extracts identification label.

State-of-the-art writer identification methodologies usually measure performance based on publicly available datasets (i.e CVL, IAM ect.) and metrics used for comparison are Top-1 Accuracy and Top-5 Accuracy. Where Top-1 accuracy means expected result is exactly the expected answer (i.e. the one with highest probability) and Top-5 accuracy means expected result is among first five highest probability answers. State-of-the-art writer identification methods with high recognition rates on public CVL dataset with Top-1 accuracy are shown in Table 1.

Table 1 – State-of-the-art writer identification methods accuracy with CVL dataset

Author	Method	Acc (%)
Christlein et al. [4]	GMM super vector + E-SVM	99.4
Christlein et al. [5]	CNN + VLAD	99.5
Chen et al. [6]	ResNet-50	99.2
He and Schomaker [7]	Adaptive CNN	79.1
Helal et al. [8]	CNN	99.80
Sulaiman et al. [9]	CNN + LBP	97.55
Kumar and Sharma [10]	SEG WI model	99.35
He and Schomaker [11]	FragNet-64	99.1
Koepf et al. [12]	ViT	99.0
Semma et al. [13]	Resnet-34	99.5
He and Schomaker [14]	GR-RNN	99.3
Wirmanto et al. [15]	Xception+siamese	99.88

Personality detection. During handwriting analysis to perform personality detection one investigates the differences of the written samples from standard writing and on the basis of this the personality of the person providing the writing sample is deduced [1].

Similar to writer identification, personality detection methodologies share the common/standard steps of

preprocessing, feature extraction and personality detection (frequently referred as classification) as shown in Fig. 2.

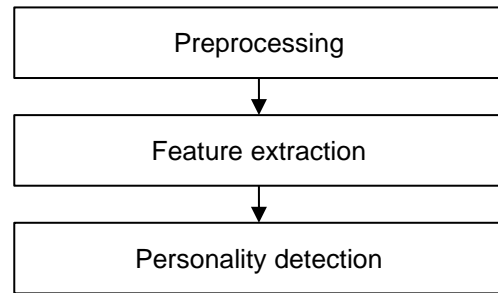


Fig. 2. Personality detection methodologies common steps

To perform personality detection one need to pick a personality psychology measurement model. There are two popular models for assessing personality types: Big Five and Myers Briggs Type Indicator (MBTI). The Big Five model formalizes personality types as scores while MBTI as classes. Article [32] performed a comparison and concluded that even though algorithms trained on MBTI could have better performance, the Big Five is much more informative and has great variability in performance depending on the algorithm.

Studies in the field of personality detection show a wide variety of approaches and they, for the most part, fall into the auto learned category.

So in auto learned category Gavrilescu [18] proposed 3-level neural network architecture with a ANN, SVM and KNN classification on 3-d level. Polap and Wozniak [19] proposed a flexible neural network where some neurons adapt to classification purposes. Topaloglu and Ekmekci [20] proposed to use decision tree and data mining techniques in conjunction with ID3 and J48 algorithms. Gavrilescu and Vizireanu [21] continued the 3-level architecture approach and this time base layer includes normalization and segmentation, intermediary layer generates Handwriting Map (HM) using binary code, top layer performs classification using feed-forward neural network(FFNN). Joshi et al. [22] developed a classification framework based on the SVM and used template-matching for letter extraction. Wijaya et al. [23] compared left, right, top and bottom margins segmentation and used SVM for classification. Fatimah et al. [24] proposed CNN based classification with SGD and AdaDelta optimization methods. Chitlangia and Malathi [25] proposed to perform feature extraction using Histogram of Oriented Gradient (HOG) technique and classification using SVM. Thomas et al. [26] collected texts written by ten subjects using a CNN model and logistic regression, and selected seven features for the classification task. Pathak et al. [27] proposed deep neural network architecture based on CNN with Long Short-Term Memory (LSTM) and Connectionist Temporal Classification (CTC). Elngar et al. [28] proposed a novel Personality Trait Level Detection Model (PTLDM) where they use Personality Analyzing Network (PAN) and PersonaNet for classification. Bernardo et al. [29] proposed as a feature extractor to use SqueezeNet lightweight architecture model optimized with hyper-parameter optimization method and SVM for

classification. Rahman et al. [30] proposed extracting features using graph-based writing representation approach and performing classification using Semi-supervised Generative Adversarial Network (SGAN). Samsuryadi et al. [31] performed a complex comparison of SVM (three variations of the kernel: linear, RBF, and polynomial), KNN, and decision tree approaches.

Personality detection researchers, unlike writer identification researchers, usually instead of publicly available datasets measure performance based on private handwriting datasets and metrics used for the comparison are accuracy, precision, recall, F1 score, true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Personality detection literature methods with high recognition rates on datasets with accuracy metric are shown in Table 2.

Discussion. Handwriting analysis systems aim to automatically extract important information from handwriting. They include such disciplines as personality detection, writer identification and many others.

For further development of such systems, it is necessary to overcome many problems, for example, improper feature extraction, overfitting, underfitting, unreliable training data, picking model for assessing personality types etc. The current directions in the development of modern handwriting analysis systems are the study and use of various architectures of neural networks and deep learning. The current high accuracy rates also suggest the need for larger datasets. This would also widen the scope for techniques relying on more training data. Since writer identification and personality detection fields have so much in common it might be reasonable to use techniques from one field to improve the other one.

Conclusions

Thus, this article presents four groups of methods for handwriting analysis categorized by the type of features as a functional context. In addition the actual advantages and disadvantages with the use and development of such methods are presented. Both methods for identifying the personality of a person and writer identification methods which are used in the state-of-the-art handwriting analysis systems are presented.

Table 2 – Personality detection literature methods.

Author	Method	Acc (%)
Gavrilescu [18]	ANN, SVM, and KNN	88.6
Polap and Wozniak [19]	Flexible neural network	93
Topaloglu and Ekmekci [20]	Decision tree	93.75
Gavrilescu and Vizireanu [21]	FFNN and template-matching	84.4
Joshi et al. [22]	SVM and template-matching	97
Wijaya et al. [23]	SVM	82.73
Fatimah et al. [24]	CNN	97.31
Chitlangia and Malathi [25]	HOG+SVM	80
Thomas et al. [26]	CNN	65
Pathak et al. [27]	Deep NN	97.7
Elngar et al. [28]	ANN + PersonaNet	65
Bernardo et al. [29]	Hybrid two-stage SqueezeNet and SVM	91.26
Rahman et al. [30]	SSL, SGAN	91.30
Samsuryadi et al. [31]	SVM, KNN, and decision tree	99

The studies in these fields and comparative analysis are presented. Also, discussion about performance measurements, databases and typical metrics are presented.

Based on analysis, conclusions about future development and possible directions of handwriting analysis systems were presented.

REFERENCES

- Hengl M. Comparison of the Branches of Handwriting Analysis, Часопис Національного університету "Острозька академія". Серія: Право. – 2014. – № 2(10)
- Hemlata S. Personality detection using handwriting analysis: Review / Hemlata S., Singh S. // In Seventh International Conference on Advances in Computing, Electronics and Communication. – 2018. – 18-19 August – P. 85-89.
- Alaei F. Review of age and gender detection methods based on handwriting analysis / Alaei F., Alaei A. // Neural Computing and Applications. – 2023. – September – P. 23909-23925.
- Christlein V. Writer Identification Using GMM Supervectors and Exemplar-SVMs / Christlein V., Bernecker D., Höning F., Maier A., Angelopoulou E. // Pattern Recognition. – 2017. – Vol. 63 – March – P. 258-267.
- Christlein V. Unsupervised Feature Learning for Writer Identification and Writer Retrieval / Christlein V., Gropp M., Fiel S., Maier A., // In 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). – 2017. – Vol. 13 – 9-15 November.
- Chen S. Semi-supervised Feature Learning For Improving Writer Identification / Chen S., Wang Y., Lin C., Ding W., Cao Z. // Information Sciences. – 2019. – Vol. 482 – May – P. 156-170.
- He S. Deep adaptive learning for writer identification based on single handwritten word images / He S., Schomaker L. // Pattern Recognition. – 2019. – Vol. 88 – April – P. 64-74.
- Helal L. G. Representation Learning and Dissimilarity for Writer Identification / Helal L. G., [et al.]. // 2019 International Conference on Systems, Signals and Image Processing (IWSSIP). – 2019. – June – P. 63-68.

9. Sulaiman A. Length Independent Writer Identification Based on the Fusion of Deep and Hand-Crafted Descriptors / Sulaiman A., Omar K., Nasrudin M. F., Arram A. // *IEEE Access*. – 2019. – Vol. 7 – June – P. 91772–91784.
10. Kumar P. Segmentation-free writer identification based on convolutional neural network / Kumar P., Sharma A. // *Computers & Electrical Engineering*. – 2012. – Vol. 85 – June.
11. He S. FragNet: Writer Identification Using Deep Fragment Networks / He S., Schomaker L. // *IEEE Transactions on Information Forensics and Security*. – 2020. – Vol. 15 – March – P. 3013–3022.
12. Koepf M. Writer Identification and Writer Retrieval Using Vision Transformer for Forensic Documents / Koepf M., Kleber F., Sablatnig R., // *Document Analysis Systems*. – 2022. – May – P. 352–366.
13. Semma A. Writer Identification using Deep Learning with FAST Keypoints and Harris corner detector / Semma A., Hannad Y., Siddiqi I., Djeddi C., El Youssfi El Kettani M. // *Expert Systems with Applications*. – 2021. – Vol. 184 – Dec – P. 115473.
14. He S. GR-RNN: Global-Context Residual Recurrent Neural Networks for Writer Identification / He S., Schomaker L. // *Pattern Recognition*. – 2021. – Vol. 117 – Apr.
15. Wirmanto S. Offline Handwriting Writer Identification using Depth-wise Separable Convolution with Siamese Network / Wirmanto S., Agustini D.A.R., Atmanto D.A., // *International Journal On Informatics Visualization*. – 2024. P. 535–541.
16. Purohit N. State-of-the-Art: Offline Writer Identification Methodologies / Purohit N., Panwar S., // *International Conference on Computer Communication and Informatics (ICCCI)*. – 2021. – Jan – P. 1–8.
17. Ahmed B. Q. Offline text-independent writer identification using a codebook with structural features / Ahmed B. Q., Hassan Y. F., Elsayed A. S., // *PLOS ONE* 18. – 2023. – Vol. 18(4) – April – P. 1–31.
18. Gavrilescu M. Study on determining the Myers-Briggs personality type based on individual's handwriting / Gavrilescu M., // *In Proceedings of the 5th IEEE International Conference on E-Health and Bioengineering*. – 2015. – Nov – P. 1–6.
19. Połap D. Flexible neural network architecture for handwritten signatures recognition / Połap D., Wozniak M., // *International Journal of Electronics and Telecommunications*. – 2016. – Vol. 62(2) – April – P. 197–202.
20. Topaloglu M. Gender detection and identifying one's handwriting with handwriting analysis / Topaloglu M., Ekmekci S., // *Expert Systems with Applications*. – 2017. – Vol. 79 – March – P. 236–243.
21. Gavrilescu M. Predicting the big five personality traits from handwriting / Gavrilescu M., Vizireanu N., // *EURASIP Journal on Image and Video Processing* – 2018. – Vol. 2018(1) – July.
22. Joshi P. A machine learning approach to employability evaluation using handwriting analysis / Joshi P., Ghaskadbi P., Tendulkar S., // *In Proceedings of the Communications in Computer and Information Science ICAICR 2018*. P. 253–263.
23. Wijaya W. Personality analysis through handwriting detection using android based mobile device / Wijaya W., Tolle H., Utaminingrum F., // *International Journal of Information Technology and Computer Science* – 2018. – Vol. 2(2). P. 114–128.
24. Fatimah S. H. Personality features identification from handwriting using convolutional neural networks / Fatimah S. H., Djamel E. C., Ilyas R., Renaldi F. // *In Proceedings of the 4th International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE* – 2019. – Nov. – P. 119–124.
25. Chitlangia A. Handwriting analysis based on histogram of oriented gradient for predicting personality traits using SVM / Chitlangia A., Malathi G., // *Procedia Computer Science* – 2019. – Vol. 165 – Jan. – P. 384–390.
26. Thomas S. A framework for analyzing financial behavior using machine learning classification of personality through handwriting analysis / Thomas S., Goel M., Agrawal D., // *Journal of Behavioral and Experimental Finance*, 2020. Vol. 26(2).
27. Pathak A. R. Personality analysis through handwriting recognition / Pathak A. R., Raut A., Pawar S., Nangare M., Abbott H. S., Chandak P., // *Journal of Discrete Mathematical Sciences and Cryptography* – 2020. – Vol. 23(1) – Jan. – P. 19–33.
28. Elngar A. A. A deep learning based analysis of the big five personality traits from handwriting samples using image processing / Elngar A. A., [et al.], // *Journal of Information Technology Management* – 2021. – Vol. 12 – P. 3–35.
29. Bernardo L. S. A hybrid two-stage SqueezeNet and support vector machine system for Parkinson's disease detection based on handwritten spiral patterns / Bernardo L. S., Damasevicius R., De Albuquerque V. H. C., Maskeliunas R., // *International Journal of Applied Mathematics and Computer Science* – 2021. – Vol. 31(4) – Dec. – P. 549–561.
30. Rahman A. U. Predicting the big five personality traits from hand-written text features through semi-supervised learning / Rahman A. U., Halim Z., // *Multimedia Tools and Applications* – 2022. – Vol. 81(23) – Sep. – P. 1–17.
31. Samsuryadi S. A Framework for Determining the Big Five Personality Traits Using Machine Learning Classification through Graphology / Samsuryadi S., [et al.], // *Journal of Electrical and Computer Engineering*. 2023. Vol. 2023(1). Jan. P. 1–15.
32. Lepri F. Is big ve better than MBTI? / Lepri F., Lepri. B., // *Proceedings of the Fifth Italian Conference on Computational Linguistics CLiC-it* – 2018. – Jan. – P. 93–98.

Received (Надійшла) 15.12.2024

Accepted for publication (Прийнята до друку) 29.01.2025

Аналіз методів визначення особистості та ідентифікації письменника

М. В. Шупилок, В. О. Марговицький

Анотація. Рукописний текст як мультисенсорна діяльність може відображати особистість і водночас може слугувати біометричним ідентифікатором. Аналіз почерку використовується в різних сферах, включаючи історію, криміналістику, освіту, безпеку, кадрові питання тощо. У цій статті методології аналізу почерку було розглянуто та розділено на чотири групи, виділивши переваги та недоліки кожної групи. Також у цій статті описуються різні проблеми, пов'язані з розробкою систем аналізу рукописного тексту, як-от неправильне виділення ознак, перенавчання, недонавчання, ненадійні навчальні дані, вибір моделі для оцінки типів особистості тощо. Представлено як методи надійної ідентифікації письменника поза мережею, так і методи прогнозування людської особистості, які використовуються в найсучасніших системах аналізу почерку. Крім того, були проаналізовані поточні дослідження та загальні підходи до вимірювання продуктивності та вибору бази даних як для ідентифікації письменника, так і для визначення особистості. Також представлено перспективні напрями розвитку сучасних систем аналізу рукописного тексту.

Ключові слова: почерк, аналіз почерку, особистість, риси почерку, графологія, визначення особистості, ідентифікація письменника.

O. Barkovska, Ya. Ni, A. Havrashenko, Ye. Peretiaka, A. Romanenko

Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

SYSTEM FOR DETECTING CRITICAL HUMAN HEALTH CONDITIONS BASED ON THE ANALYSIS OF PHYSIOLOGICAL INDICATORS

Abstract. Relevance. The modern increase in cardiovascular diseases, diabetes, and psychological disorders, particularly post-traumatic stress disorder (PTSD), necessitates the implementation of intelligent health monitoring systems. WHO statistics indicate 15 million premature deaths annually, with 32% attributed to cardiovascular diseases. Additionally, the war in Ukraine has significantly impacted stress levels among the population, increasing mortality risks. Traditional monitoring methods do not ensure timely detection of critical conditions, making the adoption of AI-based automated solutions essential. **The object** of this study is a system for detecting critical human health conditions based on the analysis of biometric indicators and their dynamics using machine learning methods. The aim of the article is to develop and evaluate the effectiveness of an automatic system for detecting critical health conditions that operates using wearable devices and artificial intelligence algorithms. To achieve this goal, a stress level classifier based on physiological indicators was implemented, and a comparative analysis of two algorithms, MLPClassifier and RandomForestClassifier, was conducted. **As a result** of the research, an architecture for a continuous health monitoring system was proposed, an algorithm for stress level assessment using ECG, EDA, BCP, and breathing patterns as input parameters were developed, and MLP and Random Forest classifiers were trained, and tested on a dataset of 65 participants. MLPClassifier demonstrated higher classification accuracy (91.3%), confirming its effectiveness for monitoring critical health conditions.

Keywords: health monitoring system, critical conditions, wearable sensors, electrocardiogram, machine learning, classification, MLPClassifier, RandomForestClassifier, telemedicine, biophysiological indicators.

Introduction

Problem Statement. Modern trends indicate that the global population is experiencing an increasing tendency toward premature deaths. According to WHO data, approximately 15 million people aged 30–70 die prematurely each year. Of these, 32% are due to cardiovascular diseases, 17% to cancer, 8% to injuries and accidents, and 7% to respiratory diseases and infections (Fig. 1).

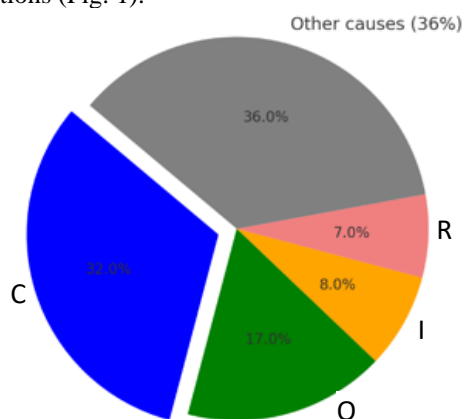


Fig. 1. Diagram of Causes of Human Mortality:
C – Cardiovascular diseases (32%);
R – Respiratory and infectious diseases (7%);
I – Injuries and accidents (8%);
O – Oncological diseases (17%)

Statistics show that 85% of these deaths occur in low- and middle-income countries. This is linked to risk factors such as a sedentary lifestyle, increased stress levels, environmental degradation, and the rising prevalence of diabetes and obesity. In addition to the growing number of psychological disorders (around 3–4 million Ukrainians show signs of PTSD), the war in Ukraine has led to a 15–20% increase in premature mortality [1].

Moreover, over 400,000 public infrastructure facilities (approximately 2.7% of all buildings in Ukraine) have been damaged by shelling, including more than 57,086 residential buildings and 126 medical institutions. This has further deteriorated living conditions and healthcare systems. According to research, the highest levels of loneliness among adults are observed in Eastern European countries. There, 7.5% of young people and 9.6% of middle-aged individuals report feelings of loneliness. Among older adults over 60, this figure rises to 21.3%. In Northern Europe, loneliness levels are lower: 2.9% among young people, 2.7% among middle-aged individuals, and 5.3% among the elderly.

It is crucial to understand that loneliness has serious health consequences. Scientific studies have identified a significant link between loneliness and an increased risk of severe illnesses. Loneliness and social isolation are associated with a higher probability of premature death, cardiovascular diseases, stroke, and type 2 diabetes. In particular, loneliness increases the risk of premature death by approximately 30%.

Remote health monitoring and telemedicine are highly relevant today, requiring the development of new approaches and methods [2], as well as the creation of assistants and support systems for people with disabilities [3]. Looking at such negative statistics, as well as social and economic risks, the relevance of these problems leads humanity to the idea of a comprehensive approach to preventing and detecting critical health conditions. The main tools for reducing premature mortality include:

- medical and personal health monitoring;
- analysis and mitigation of risk factors;
- promotion of a healthy lifestyle;
- reduction of environmental hazards;
- vaccination and infection control.

The further development of a comprehensive approach to preventing critical health conditions (see Table 1) will be driven by technological trends (AI in

medical diagnostics and telemedicine) and adherence to key system principles:

- continuous monitoring;
- personalized approaches;
- integration of multidisciplinary data;
- proactive response.

These improvements will enhance the accuracy of disease prediction and, as a result, significantly reduce mortality rates among the population. The following physiological markers of critical health conditions have been selected for this study: ECG, Blood Circulation Pulse (BCP), Breathing Pattern, and Electrodermal Activity (EDA).

Table 1– Development of Critical Health Condition Prevention

Diagnosis Level	Treatment Methods	Disease Detection Technologies	Health Condition Indicators	Expected Outcome
Primary Prevention	<ul style="list-style-type: none"> – Healthy lifestyle – Regular medical check-ups – Proper nutrition 	<ul style="list-style-type: none"> – Mobile monitoring apps – Fitness trackers – Genetic testing 	<ul style="list-style-type: none"> – Biometric indicators – Heredity – Age and gender 	<ul style="list-style-type: none"> – Prevention of disease occurrence
Secondary Prevention	<ul style="list-style-type: none"> – Early diagnosis – In-depth examination – Monitoring of risk groups 	<ul style="list-style-type: none"> – Artificial intelligence – MRI/CT diagnostics – Biomarker analysis 	<ul style="list-style-type: none"> – Deviations in tests – Risk symptoms – Changes in the body 	<ul style="list-style-type: none"> – Stopping disease progression
Tertiary Prevention	<ul style="list-style-type: none"> – Rehabilitation – Correction of complications – Psychological support 	<ul style="list-style-type: none"> – Personalized medicine – Telemedicine – Neural networks 	<ul style="list-style-type: none"> – Treatment dynamics – Quality of life – Psycho-emotional state 	<ul style="list-style-type: none"> – Minimization of disease consequences
Predictive Level	<ul style="list-style-type: none"> – Risk prediction – Individual forecasting – Preventive measures 	<ul style="list-style-type: none"> – Big Data analysis – Genomic medicine – Machine learning 	<ul style="list-style-type: none"> – Genetic profile – Behavioral patterns – Social factors 	<ul style="list-style-type: none"> – Proactive health management

ECG is a critical indicator for assessing heart function and diagnosing and monitoring conditions such as arrhythmia, heart failure, and stress. BCP is a circulation indicator reflecting changes in the pulse wave in blood vessels. The Breathing Pattern is used to assess both the physiological and emotional state of a person. Additionally, analyzing the breathing pattern is utilized in HRV (Heart Rate Variability) to evaluate the balance of the autonomic nervous system [4]. Electrodermal Activity refers to the electrical activity of the skin, which changes due to the function of sweat glands. It is used to evaluate emotional state, stress, and the activity of the nervous system.

Analysis of Recent Research and Publications.

The article [5] provides a review of smart wearable systems used for monitoring human biological parameters. The authors discuss hardware, sensor technologies, data analysis algorithms, and methods for integrating these devices into healthcare systems. Special attention is given to artificial intelligence and machine learning, which help improve the processing of acquired signals and enhance the accuracy of medical condition prediction. The results demonstrate that AI helps reduce errors in biosignal analysis, while the use of flexible nanomaterials allows for the creation of highly sensitive, wearable sensors and autonomous power sources, which hold promise for improving the independence of devices. However, the study remains largely theoretical, with practical experiments still limited.

The work [6] examines modern optical sensors for non-contact measurement of vital signs. The principles of operation of fluorescent, spectroscopic, and surface plasmon resonance (SPR) sensors are described. Their advantages over traditional electrochemical and mechanical sensors, particularly in terms of accuracy and speed, are analyzed. The results indicate that optical sensors show high precision in measuring biomarkers in blood and saliva and integrating these sensors into smartphones and wearable devices allows for at-home

health monitoring. SPR methods enable non-contact analysis of biological fluids with minimal intrusion. However, real-world experiments on the long-term use of these sensors with patients are not included in the study.

The article [7] focuses on the use of wearable devices for monitoring the health of elderly individuals. The primary devices and sensors that monitor heart rate, blood oxygen levels, blood pressure, and physical activity in real time are described. A separate section is dedicated to the use of IoT and cloud technologies, which enable doctors to receive data remotely. The conclusion highlights that wearable devices significantly reduce the need for hospital monitoring, allowing patients to remain at home, while IoT use enables doctors to respond quickly to critical changes in patient status. However, the greatest challenge remains ensuring data security and the energy autonomy of devices.

In the work [8] an overview of current wearable devices for diagnosing and monitoring patients is presented. Biometric sensors, electrochemical sensors, and artificial intelligence in disease prediction are analyzed. Significant attention is given to the future of wearable devices, including their integration with medical databases and augmented reality (AR) technologies. However, the study lacks practical data on the effectiveness of these devices in real-world conditions.

The work [9] discusses the mathematical model of remote photoplethysmography (rPPG), based on optical and physiological properties of light reflection from the skin. The article analyzes key rPPG methods for extracting pulse signals from video recordings, including Blind Source Separation (PCA, ICA), CHROM, PBV, and 2SR. The main contribution of the paper is a detailed theoretical analysis of rPPG principles, enabling the development of new algorithms for specific tasks. The authors propose a new approach called "Plane-Orthogonal-to-Skin" (POS), which uses the projection of a color vector onto a plane orthogonal to skin tone to

improve pulse signal extraction. However, the proposed method is more focused on theoretical analysis than practical clinical application, and some methods may have errors under different lighting conditions. Furthermore, the POS method does not fully account for the variability in light reflection from skin of people with different skin types.

The analysis shows that modern health monitoring systems significantly expand the possibilities for controlling the physical condition of individuals. Future developments in health monitoring should focus on the integration of wearable and stationary systems, improving device autonomy, enhancing measurement accuracy, and expanding telemedicine capabilities.

After analyzing the characteristics and features of the wearable devices described for health monitoring, it can be concluded that these devices have a high cost, do not provide a comprehensive approach to health monitoring, have a limited monitoring duration, require a continuous Internet connection, and are not sufficiently accurate. However, the accumulation of primary biometric data through wearable sensors is insufficient to obtain valuable clinical information. A comprehensive development of methods for processing and interpreting the obtained data is needed, allowing for a transition from quantitative indicators to a qualitative assessment of the human functional state. This approach reflects the effectiveness of performing certain activities and the level of regulatory system tension in the body. Classifying functional states allows for a transition from a continuous stream of various biometric data to discrete categories with clear physiological and clinical interpretations. Several main functional states are identified: physiological norm (optimal state), functional tension, overstrain (distress), exhaustion, and pathological state. Each of these states is characterized by specific changes in the functioning of physiological

systems and can be identified by a set of objective indicators.

The main problem in classifying functional states is the selection of informative features. Collected biometric data contains a significant amount of redundant information and noise, which complicates their direct use for classification and is further complicated by the high individual variability of physiological indicators.






Artificial intelligence offers the necessary tools for solving the problem of classifying functional states. Unlike traditional statistical methods, machine learning algorithms are capable of working with large amounts of unstructured data, identifying hidden patterns, and adapting to individual user characteristics. Among the classical machine learning methods, the most effective for classifying functional states are the Support Vector Machine (SVM), Random Forest, and Multi-Layer Perceptron (MLP) [10, 11].

The goal of the work is to develop a model of a system for detecting critical health conditions and stress levels in humans based on the analysis of physiological indicators, which will provide continuous remote monitoring of the patient's condition using portable devices.

To achieve the stated goal, the following tasks must be solved:

- development of a model for detecting critical human states;
- development of an algorithm for the stress level sensor based on the medical data obtained from the patient;
- implementation of classifiers based on machine learning methods (MLP and Random Forest) for classifying stress levels based on the patient's biophysical data;
- comparative analysis of classification accuracy based on MLP and Random Forest.

Table 2 – Wearable Devices for Home Use

System Name	Description of Main Functions of the System	System Drawbacks	Demonstration
Apple Watch	Measurement of pulse rate, ECG, blood oxygen saturation (SpO2)	Inaccuracy of measurements compared to medical devices, low battery life, dependency on the Apple ecosystem.	
BioIntelliSense BioSticker	Continuous monitoring of heart rate, body temperature, body movement tracking, remote monitoring capability.	Limited duration of operation for a single sticker (up to 30 days), measurement inaccuracies during movement, system dependence on a stable internet connection.	
CarePredict Tempo	Tracking the daily activity of elderly people, communication function with caregivers.	High cost of the system, limited device autonomy, dependency on wearing the device.	
Withings BPM Core	Measurement of blood pressure, ECG, heart sound recording	Difficulty of use, dependence on mobile app and internet connection, external noise or body movements may affect measurement results.	
Elderly Care Systems (Vayyar Home, KamiCare, Walabot)	Fall detection without wearable sensors, tracking movement and human activity, tracking breathing at high frequencies through walls	Dependence of accuracy on proper placement of sensors, recognition issues in low light conditions, high installation and maintenance costs.	

Main Content

This work proposes a model of a system for detecting critical health conditions and stress levels in humans based on the analysis of physiological indicators, designed to trigger medical services in case one or more health indicators exceed critical thresholds (fig. 2).

The model includes the following modules – data collection module (physiological indicators), data analysis module, and data storage module. The purpose of each module is as follows:

- the data collection module ensures the collection of data using a wearable device that contains sensors for measuring specific physiological parameters of the patient;

- the data analysis module receives and processes the biometric data of the patient using artificial intelligence. The system also includes response mechanisms for critical situations, which are implemented through the "decision-making" module. This module contains a closed-loop learning cycle for improving the system based on the processed data received from the data storage module.

- the data storage module ensures the reception and storage of information received from the data analysis module, including biometric data, medical history, age, and full name of the patient. This information is later forwarded for use in the closed-loop learning and system improvement process.

Within the proposed model, an algorithm for detecting stress levels has also been implemented. This algorithm involves the analysis of collected data from an electro-optical heart activity sensor (ECG), an electrical skin activity sensor (EDA), a blood circulation

measurement sensor (BCP) based on optical photoplethysmography (PPG) technology, and a combination of accelerometers and optical PPG sensors for tracking breathing patterns (Breathing pattern) using machine learning methods (Fig. 3). A key requirement for the algorithm's functioning is the preliminary accumulation of data to determine the normal state of the body, followed by sensor calibration before directly assessing stress levels based on deviations from the normal state of the indicators. The proposed algorithm is based on a comprehensive approach to analyzing the patient's biophysiological data and consists of several sequential stages of information processing. The first stage is the collection of physiological data, namely: ECG, BCP, EDA, and Breathing pattern. These indicators are the most informative for assessing the psycho-physiological state of a person and the level of stress, as they reflect the autonomic nervous system's response to stress factors.

The next stage involves the preliminary processing of the collected signals, including noise filtering and artifact removal. To improve the accuracy of the analysis, digital filters and adaptive filtering algorithms are used, effectively separating the useful signal from noise components. Special attention is given to removing motion artifacts that could distort the analysis results.

Based on the comprehensive analysis of the obtained parameters, the stress level is determined. For stress level classification, a machine learning algorithm is used, trained on a dataset consisting of data from 65 participants. The system classifies the current stress level as either low or high. To improve classification accuracy, individual user characteristics are taken into account, such as the baseline level of indicators at rest and the dynamics of their changes under stress.

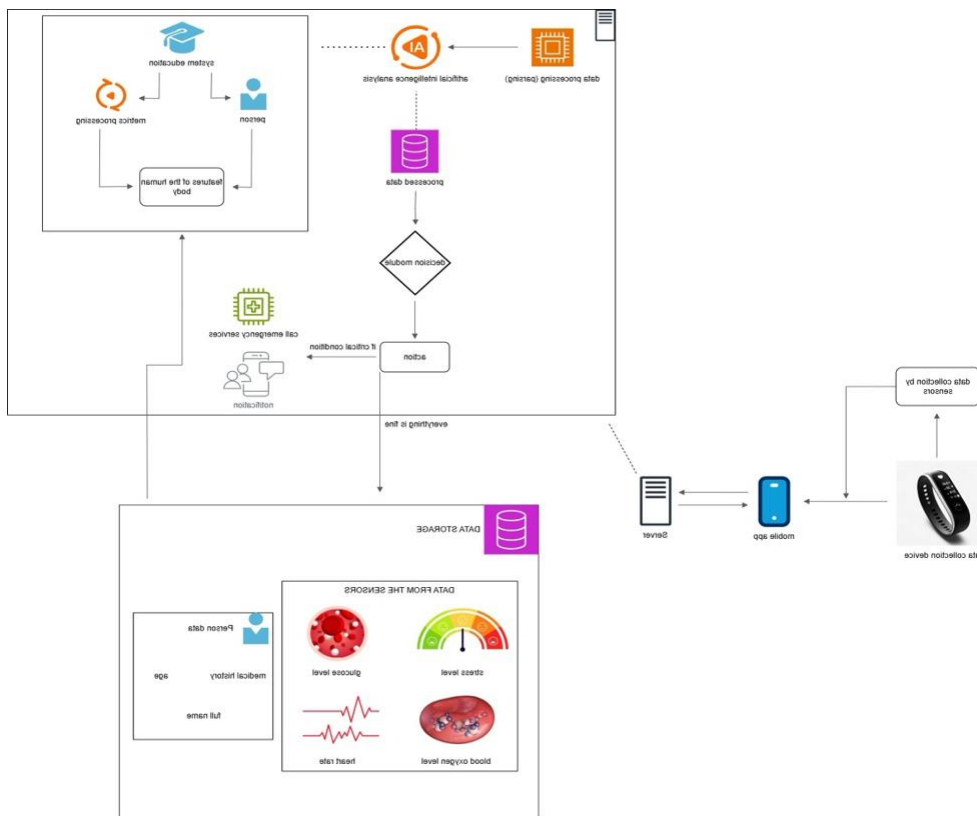


Fig. 2. Diagram of the system model for detecting critical health conditions based on the analysis of physiological indicators

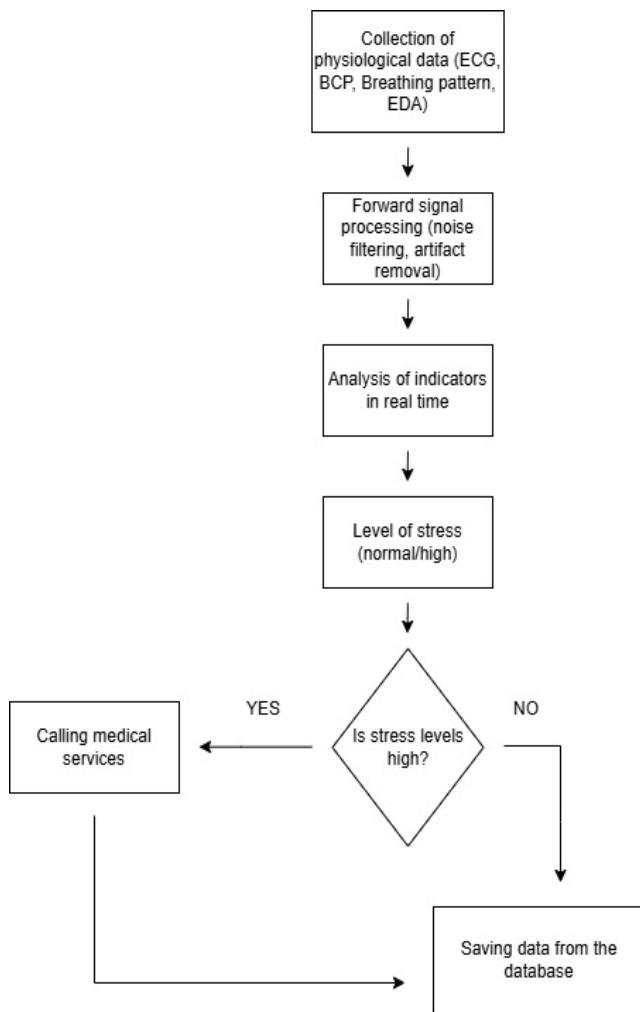


Fig. 3. Algorithm of the Stress Level Sensor Operation

In the case of detecting a high-stress level, the system activates the medical service call mechanism. Simultaneously, the data is saved in a database for further analysis and the formation of recommendations. If the stress level is determined to be normal, the data is also

stored in the database to track the dynamics of the indicators over time.

In this study, a dataset based on data from 65 participants is used to train classifiers. This dataset includes facial video, audio, and physiological signals, from which the physiological indicators are considered for analysis. A graph based on these data is shown in Figure 4. The physiological indicators used in the study include:

- ECG (electrical activity of the heart): The ECG line is demonstrated with an amplitude range of -0.5 to 1.5, displaying the typical pattern with R-peaks corresponding to ventricular contractions. During stress (indicated in red), changes in the heart rate and the amplitude of the R-peaks can be observed as part of the physiological response to stress;

- Electrodermal Activity (EDA): A physiological indicator for measuring changes in the electrical properties of the skin, demonstrated on a scale from 2.0 to 2.4 microsiemens (μS) along the Y-axis. During stress, there is a decrease in the EDA values (from 2.3 μS to 2.1 μS), indicating a change in skin conductivity;

- Breathing Pattern: A measure of the rhythm, depth, frequency, and overall structure of breathing movements, shown by a curve with an amplitude range from -1.5 to -0.5. During stress, a deeper and more controlled breathing pattern is often observed, which is a common response to stress;

- Blood Circulation Pulse (BCP): This indicator reflects periodic changes in blood volume and pressure in the blood vessels associated with heartbeats. The curve is shown with clear pulsatile waves ranging from 60-120 beats per minute. During stress (red zones), an increase in the amplitude of the pulse waves is observed.

The lower part of the graph (Predicted stress level) shows the reference values for stress level distribution into two classes: red – high-stress level, green – normal stress level. The scale ranges from 0 to 1, where all values up to 0.5 represent normal levels, and those above 0.5 indicate high-stress levels.

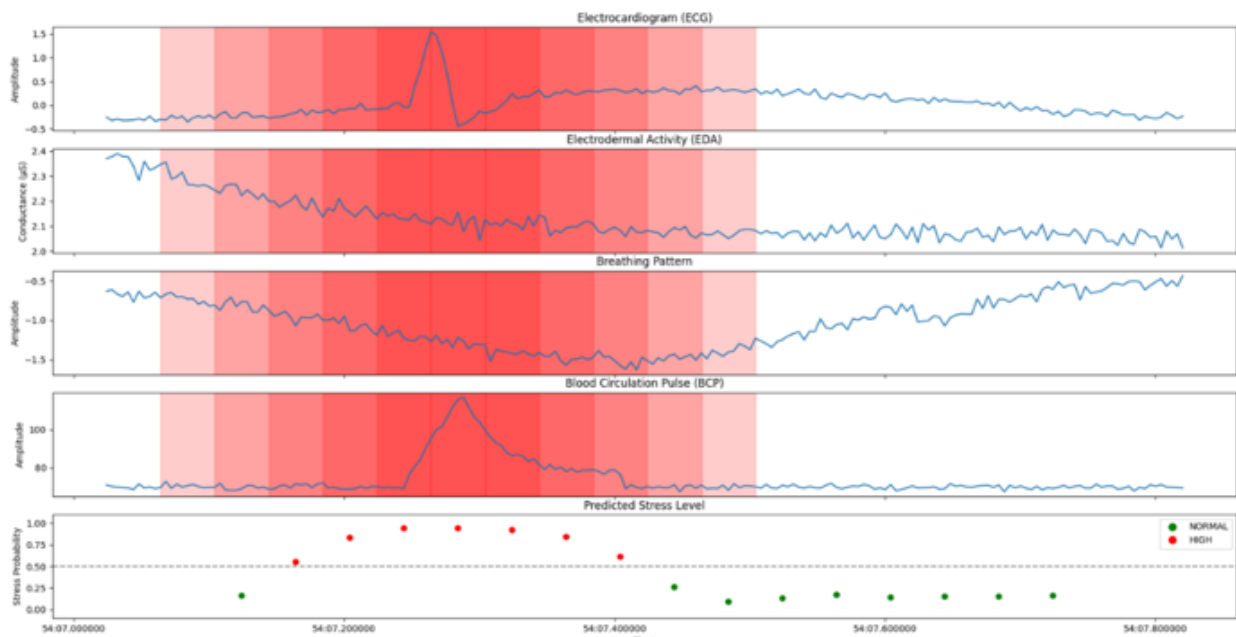


Fig. 4. Dataset for tracking human physiological parameters

The dataset described above serves as input for training the selected machine learning methods, specifically RandomForestClassifier and MLP, whose effectiveness is the subject of research in this work.

Table 3 presents a comparison of the confusion matrix structure for the two machine learning algorithms: Multi-layer Perceptron (MLPClassifier) and Random Forest (RandomForestClassifier).

The results of the study show that MLPClassifier outperforms in all the metrics considered. Specifically, the number of true negatives (TN) for MLPClassifier is 285 compared to 270 for RandomForestClassifier, indicating more accurate identification of negative cases in favor of MLP (+15). MLP also shows fewer false positives (FP) and false negatives (FN) (by 15 and 20, respectively), confirming its higher accuracy and lower error rate in classification. Furthermore, the number of true positives (TP) for MLPClassifier is 20 higher, further emphasizing its effectiveness in identifying positive cases.

Additional performance metrics analysis, shown in Table 4, indicates that the MLPClassifier has a higher accuracy (91.3%) compared to the RandomForest Classifier (82.5%), with an increase of 8.8%. Its precision also significantly outperforms that of the RandomForestClassifier (84.2% vs. 66.7%, a 17.5% improvement). Similarly, the MLPClassifier shows higher recall (80.0% vs. 60.0%, +20.0%), which highlights the model's ability to correctly identify positive cases. The F1-score, combining precision and recall, for the MLPClassifier is 82.1%, while for the RandomForestClassifier it is 63.2% (+18.9%), further confirming the overall advantage of the MLPClassifier.

Based on the two comparative tables (3 and 4) on error structure and metrics, a visualization of the obtained results was created in the form of confusion matrices (Fig. 5) for the RandomForestClassifier and MLP

classifiers, which classify data into 2 classes – NORMAL and HIGH. Each matrix shows the number of correct and incorrect predictions.

Table 3– Confusion Matrix Structure

Metric	MLPClassifier	RandomForestClassifier	Improvement in MLPClassifier
Accuracy	91.3%	82.5%	+8.8%
Precision	84.2%	66.7%	+17.5%
Recall	80.0%	60.0%	+20.0%
F1-Score	82.1%	63.2%	+18.9%

Table 4– Comparison of the performance of selected classifiers based on metrics

Metric	MLPClassifier	RandomForestClassifier	Difference
True negatives (TN)	285	270	+15 in favor of MLP
False Positives (FP)	15	30	-15 in favor of MLP
False Negatives (FN)	20	40	-20 in favor of MLP
True Positives (TP)	80	60	+20 in favor of MLP

For the MLPClassifier model, there are 285 instances where the model correctly identified the NORMAL class and 80 instances where it correctly identified the HIGH class. However, some errors occurred: the model mistakenly classified NORMAL as HIGH 15 times and HIGH as NORMAL 20 times.

In the case of the RandomForestClassifier, 270 instances of the NORMAL class and 60 instances of the HIGH class were correctly identified. However, this model made more errors: it mistakenly predicted NORMAL as HIGH 30 times and HIGH as NORMAL 40 times.

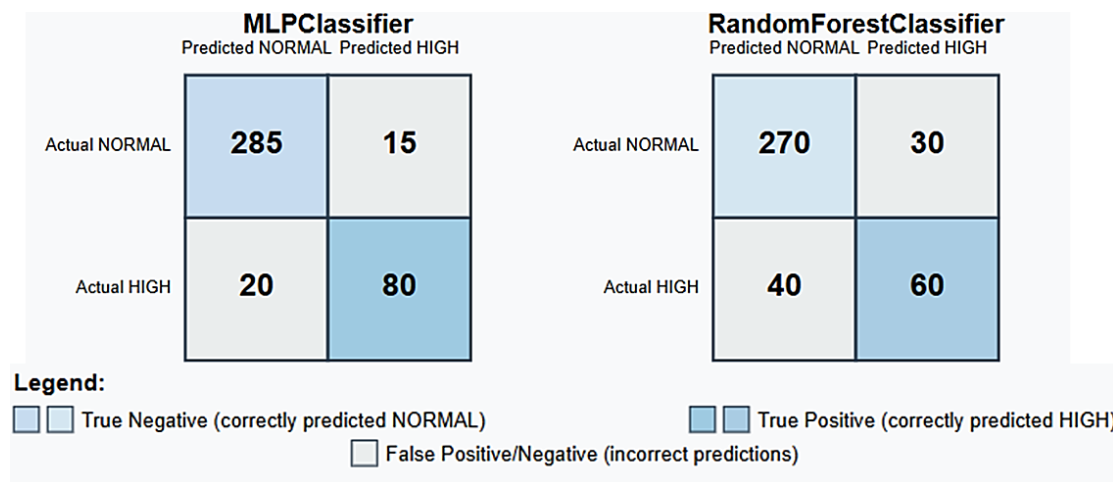


Fig. 5. Confusion matrix of two classifiers: MLP and Random Forest

Compared to the RandomForestClassifier, the MLPClassifier model demonstrates higher accuracy, as it has fewer errors and a greater number of correctly classified objects.

For stress level classification tasks based on biometric data, MLPClassifier demonstrated the best

results among the models considered. It provides high accuracy (91.3%) and a good balance between precision and recall for the "HIGH" class (stress), which is crucial for minimizing both false positives and false negatives when determining elevated stress levels.

However, if the goal is to identify which biometric indicators are most significant for determining stress

levels, it would be worth considering the use of RandomForestClassifier, which provides a feature importance metric.

Conclusions

The paper presents a system for detecting critical health conditions based on the analysis of physiological indicators. A novel approach is proposed, utilizing wearable sensors for continuous monitoring of vital parameters such as electrocardiogram (ECG), blood circulation pulse (BCP), breathing pattern, and electrodermal activity (EDA).

Using machine learning methods, specifically MLPClassifier and RandomForestClassifier, the system achieves high accuracy in detecting deviations in physiological indicators that may indicate critical conditions. A comparative analysis of the algorithms' performance was conducted, with MLPClassifier

demonstrating higher accuracy (91.3%) compared to RandomForestClassifier (82.5%). The results confirm the effectiveness of the proposed approach for remote health monitoring, which is particularly relevant in the context of telemedicine.

The developed system enables the automatic detection of critical health conditions based on physiological parameters, significantly improving the response time to threatening situations. The use of machine learning methods, particularly MLPClassifier, ensures high classification accuracy, making the system promising for implementation in telemedicine services, rehabilitation programs, and personal monitoring systems. Future research will focus on optimizing algorithms, expanding the dataset, and improving the system's adaptation to individual users' physiological characteristics.

СПИСОК ЛІТЕРАТУРИ

1. Dietrich, O., et al. (2025). An Open-Source Tool for Mapping War Destruction at Scale in Ukraine Using Sentinel-1 Time Series. arXiv:2406.02506, arXiv, 20 Feb. 2025. arXiv.org. <https://doi.org/10.48550/arXiv.2406.02506>.
2. Barkovska, O., Oliinyk, D., Sorokin, A., Zabroda, I., & Sedlaček, P. (2024). A system for monitoring the progress of rehabilitation of patients with musculoskeletal disorder. *Advanced Information Systems*, 8(3), 13–24.
3. Barkovska, O., & Serdechnyi, V. (2024). Intelligent Assistance System for People with Visual Impairments. *INNOVATIVE TECHNOLOGIES AND SCIENTIFIC SOLUTIONS FOR INDUSTRIES*, (2)28, 6–16. DOI.org (Crossref). <https://doi.org/10.30837/2522-9818.2024.28.006>.
4. Kunczik, J., et al. (2022). Breathing pattern monitoring by using remote sensors. *Sensors*, 22(22), 8854.
5. Deng, Z., et al. (2023). Smart Wearable Systems for Health Monitoring. *Sensors*, 23(5), 2479. <https://doi.org/10.3390/s23052479>.
6. Kaur, B., et al. (2023). Novel Wearable Optical Sensors for Vital Health Monitoring Systems—A Review. *Biosensors*, 13(2), 181. <https://doi.org/10.3390/bios13020181>.
7. Olmedo-Aguirre, J. O., et al. (2022). Remote Healthcare for Elderly People Using Wearables: A Review. *Biosensors*, 12(2), 73. <https://doi.org/10.3390/bios12020073>.
8. Adeghe, E. P., et al. (2024). A Review of Wearable Technology in Healthcare: Monitoring Patient Health and Enhancing Outcomes. *Open Access Research Journal of Multidisciplinary Studies*, 7(1), 142–148. <https://doi.org/10.53022/oarjms.2024.7.1.0019>.
9. Wang, W., et al. (2016). Algorithmic principles of remote PPG. *IEEE Transactions on Biomedical Engineering*, 64(7), 1479–1491.
10. Tirink, C., et al. (2023). Estimation of body weight based on biometric measurements by using random forest regression, support vector regression and CART algorithms. *Animals*, 13(5), 798.
11. Di, H., Shafiq, M., & AlRegib, G. (2018). Patch-level MLP classification for improved fault detection. SEG technical program expanded abstracts 2018. Society of Exploration Geophysicists, 2211–2215.

Received (Надійшла) 06.01.2025

Accepted for publication (Прийнята до друку) 05.03.2025

Система виявлення критичних станів здоров'я людини на основі аналізу фізіологічних показників

О. Ю. Барковська, Я. С. Ні, А. О. Гаврашенко, Є. О. Перетяка, А. О. Романенко

Анотація. Актуальність. Сучасний ріст захворювань серцево-судинної системи, діабету та психологічних розладів, зокрема посттравматичного стресового розладу (ПТСР), вимагає впровадження інтелектуальних систем моніторингу стану здоров'я. Статистика ВООЗ вказує на 15 мільйонів передчасних смертей щороку, причому 32% припадає на серцево-судинні захворювання. Крім того, війна в Україні суттєво вплинула на рівень стресу серед населення, що підвищує ризик смертності. Традиційні методи моніторингу не забезпечують оперативного виявлення критичних станів, що робить необхідним застосування автоматизованих рішень на основі штучного інтелекту. **Об'єктом дослідження** є система виявлення критичних станів здоров'я людини, що базується на аналізі біометричних показників та їхньої динаміки за допомогою методів машинного навчання. **Метою статті** є розробка та оцінка ефективності системи автоматичного виявлення критичних станів здоров'я, що працює на основі носимих пристроїв і алгоритмів штучного інтелекту. Для досягнення цієї мети реалізовано класифікатор рівня стресу на основі фізіологічних показників та проведено порівняльний аналіз двох алгоритмів: MLPClassifier та RandomForestClassifier. **В результаті** проведених досліджень запропоновано архітектуру системи безперервного моніторингу критичних станів здоров'я, розроблено алгоритм оцінки рівня стресу, що використовує ECG, EDA, BCP та Breathing pattern як вхідні параметри, навчено та протестовано класифікатори MLP та Random Forest на датасеті із 65 учасниками. MLPClassifier продемонстрував вищу точність класифікації (91.3%), що підтверджує його ефективність для моніторингу критичних станів здоров'я.

Ключові слова: система моніторингу здоров'я, критичні стани, носимі сенсори, електрокардіограма, машинне навчання, класифікація, MLPClassifier, RandomForestClassifier, телемедицина, біофізіологічні показники.

І. В. Ільїна, К. Д. Костін

Харківський національний університет радіоелектроніки, Харків, Україна

СУЧАСНІ МЕТОДИ OLAP У ПАРТНЕРСЬКОМУ МАРКЕТИНГУ: ОПТИМІЗАЦІЯ ПРИБУТКУ ЗА ДОПОМОГОЮ БАГАТОВИМІРНОГО АНАЛІЗУ ДАНИХ

Анотація. У статті досліджено можливості та переваги використання OLAP-технологій у партнерському маркетингу, зокрема їхній вплив на оптимізацію прибутку за допомогою багатовимірного аналізу даних. Розглянуто сучасні методи OLAP, включаючи MOLAP, ROLAP, HOLAP, а також новітні технологічні тенденції, такі як In-Memory обробка, хмарні рішення та інтеграція зі штучним інтелектом. Значну увагу приділено практичним аспектам використання OLAP, зокрема сегментації партнерів, аналізу ефективності маркетингових кампаній та виявленню шахрайства. Також представлено реальні кейси впровадження OLAP у бізнес, включаючи аналіз успішних компаній, що використовують цю технологію для підвищення ефективності своїх партнерських програм. Окремо розглянуто виклики та перспективи розвитку OLAP у маркетинговій аналітиці, включаючи питання безпеки даних, інтеграції з іншими аналітичними системами та потенційні напрями розвитку у найближчі роки. Дослідження підтверджує, що OLAP є важливим елементом сучасного цифрового маркетингу, який сприяє швидкому прийняттю обґрунтованих рішень та підвищенню прибутковості компаній.

Ключові слова: OLAP, партнерський маркетинг, багатовимірний аналіз даних, MOLAP, ROLAP, HOLAP, In-Memory технології, хмарні рішення, штучний інтелект.

Вступ

Актуальність теми. У сучасному бізнес-середовищі, де дані відіграють ключову роль, партнерський маркетинг стає все більш залежним від ефективного аналізу великих обсягів інформації. Зростання цифрових платформ та онлайн-каналів збільшує кількість даних, які необхідно обробляти для прийняття обґрунтованих рішень. Технології OLAP (Online Analytical Processing) забезпечують багатовимірний аналіз даних, що дозволяє компаніям швидко отримувати необхідну інформацію для оптимізації маркетингових стратегій. Використання OLAP у партнерському маркетингу стає особливо актуальним через потребу в оперативному аналізі ефективності партнерських програм та швидкому реагуванні на зміни ринку.

Партнерський маркетинг стикається з низкою викликів. По-перше, зростання кількості партнерів і клієнтів ускладнює обробку та аналіз даних традиційними методами, що потребує використання спеціалізованих технологій для ефективного управління великими масивами інформації. По-друге, важливим завданням є оцінка ефективності партнерів: компаніям необхідно швидко визначати, які з них приносять найбільшу цінність, і відповідно оптимізувати співпрацю. По-третє, значну проблему становить виявлення шахрайства, оскільки деякі партнери можуть використовувати недобросовісні схеми, що впливають на прибутковість. Нарешті, одним із ключових викликів є прогнозування тенденцій – бізнесу потрібно передбачати зміни в поведінці споживачів і адаптувати маркетингові стратегії відповідно до нових трендів. OLAP-технології допомагають вирішувати ці завдання, забезпечуючи швидкий та гнучкий аналіз багатовимірних даних, що сприяє прийняттю обґрунтованих рішень у партнерському маркетингу.

Метою цього дослідження є аналіз сучасних методів OLAP та їхнього впливу на оптимізацію прибутку. У межах роботи розглядаються основні OLAP-технології та їх застосування в партнерському

маркетингу, аналізується їхній вплив на покращення ефективності партнерських програм, а також визначаються переваги та обмеження використання цих методів у сучасних компаніях.

Для проведення дослідження використано аналіз наукової літератури, що включає статті, монографії та дослідження, присвячені застосуванню OLAP у маркетинговій аналітиці та управлінні даними. Додатково здійснено огляд практичних кейсів, які демонструють реальні приклади впровадження OLAP-технологій у компаніях, що працюють у сфері партнерського маркетингу. Також враховано результати вторинних досліджень, що містять статистичні дані щодо ефективності партнерських програм до та після впровадження OLAP. Окрім цього, у процесі аналізу залучено думки експертів із маркетингової аналітики та інформаційних технологій для отримання практичних інсайтів. Такий підхід дозволяє комплексно оцінити роль OLAP у партнерському маркетингу та його вплив на оптимізацію прибутковості компаній.

Виклад основного матеріалу

Огляд літератури. У сучасній науковій літературі питання застосування OLAP-технологій у партнерському маркетингу досліджено недостатньо. Однак існують роботи, які розглядають використання OLAP у суміжних сферах, що можуть бути корисними для розуміння його потенціалу в партнерському маркетингу.

Зокрема у [1] представлено загальну інформацію про сутність, створення та практичне застосування технологій інтелектуального аналізу даних, включаючи OLAP. Ці технології можуть бути застосовані для аналізу великих обсягів даних у маркетингових дослідженнях.

У [2] розглядаються теоретичні аспекти маркетингових комунікацій, включаючи приклади практичної діяльності компаній. Хоча посібник не фокусується безпосередньо на OLAP, він надає контекст для розуміння, як аналітичні інструменти можуть підтримувати маркетингові стратегії.

У дослідженні [3] розглядаються можливості застосування OLAP для підтримки бізнес-процесів та прийняття ефективних рішень на різних рівнях управління організацією, включаючи маркетингові рішення. Автори підкреслюють, що впровадження OLAP-технологій може сприяти успіху організації шляхом покращення аналізу даних та підтримки прийняття рішень.

У статті [4] досліджується система оперативного опрацювання Інтернет-даних на основі OLAP-технологій. Автори зазначають, що OLAP надає підприємству зручні та швидкі засоби доступу, перегляду та аналізу ділової інформації, що є важливим для прийняття обґрунтованих управлінських рішень.

У роботі [5] досліджуються методи Data Mining для оцінки маркетингових заходів торговельних мереж. Автори акцентують увагу на технології Text Mining, яка, будучи розробленою на основі базових технологій Data Mining, дозволяє шукати тренди, закономірності та взаємозв'язки в неструктурованих текстах для прийняття управлінських рішень.

У статті [6] представлено огляд основних застосувань OLAP-технологій для аналізу інформаційних мережевих даних. Автори підкреслюють, що OLAP дозволяє ефективно аналізувати великі обсяги даних, що є критично важливим для прийняття обґрунтованих управлінських рішень.

У дослідженні [7] розглядається питання, де і як штучний інтелект та великі дані можуть бути корисно впроваджені в партнерський маркетинг. Автор аналізує, як ці технології можуть покращити ефективність партнерських програм.

У статті [8] досліджується забезпечення інформацією маркетингової інформаційної системи та технічні нюанси даного процесу. Автори підкреслюють, що OLAP-технології можуть бути використані для збору та аналізу даних, що сприяє прийняттю ефективних маркетингових рішень.

У роботі [9] представлено огляд технологій OLAP та їх вплив на прийняття рішень. Автори досліджують різні технології OLAP, їх застосування та дизайн, що може бути корисним для розуміння їх впливу на процес прийняття рішень у маркетингу].

У [10] було проведено порівняння різних алгоритмів машинного навчання для виявлення шахрайства в партнерському маркетингу. Автори аналізують, як різні фактори, такі як розмір набору даних, вибір ознак та параметри алгоритму, впливають на продуктивність алгоритмів. У статтях [11, 12] обговорюються інструменти для збору та аналізу даних у партнерському маркетингу. Автори статті [11] підкреслюють важливість використання таких інструментів, як дослідження ключових слів, аудит сайтів та аналіз конкурентів, для оптимізації стратегій SEO у партнерському маркетингу. Автори у статті [12] зазначають, що більшість партнерських бізнесів використовують комбінацію бізнес-аналітичних додатків, таких як статистичні інструменти, інструменти прогнозного моделювання та інструменти добування даних, для підвищення продуктивності та прибутковості.

У роботі [13] обговорюються техніки аналізу даних для покращення партнерського маркетингу.

Автори підкреслюють важливість аналізу демографічних та психографічних даних, відстеження та аналізу джерел трафіку партнерів, проведення A/B тестування та використання прогнозової аналітики для стратегічного планування.

У статті [14] обговорюються статистичні дані, які вимірюють ефективність, тенденції та зростання партнерського маркетингу. Автори зазначають, що ці статистичні дані можуть включати такі показники, як кількість партнерів, коефіцієнти конверсії, продажі, зароблені комісії та темпи зростання галузі.

Таким чином, хоча прямих досліджень щодо застосування OLAP у партнерському маркетингу обмежено, існуючі роботи в суміжних сферах підкреслюють потенціал цих технологій для покращення аналізу даних та підтримки прийняття рішень у маркетингових стратегіях, і OLAP серед них займає важливе місце.

OLAP у маркетинговій аналітиці: сучасний стан та тенденції. OLAP – це технологія, що дозволяє проводити багатовимірний аналіз даних, і вона є ключовою для маркетингової аналітики (рис. 1).

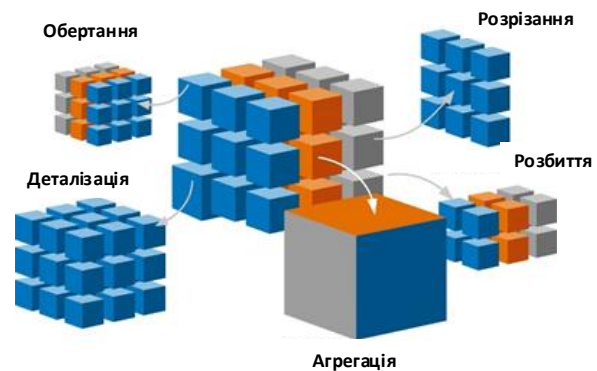


Рис. 1. Принцип роботи з даними в OLAP структурах [15]

Сучасні тенденції розвитку OLAP спрямовані на підвищення продуктивності та гнучкості аналітичних процесів, що виконуються при обробці великих обсягів даних. Одним із важливих напрямків є перехід до In-Memory обробки, що дозволяє зберігати та обробляти дані безпосередньо в оперативній пам'яті, значно прискорюючи аналітику. Іншою важливою зміною є перехід до хмарних рішень, які дозволяють компаніям масштабувати аналітичні можливості без необхідності інвестувати у власну інфраструктуру.

Також все більше значення набуває використання штучного інтелекту та машинного навчання, що дозволяє автоматизувати процеси аналізу партнерських даних, прогнозувати результати маркетингових кампаній та виявляти шахрайство.

Важливо розрізнити OLAP від інших аналітичних підходів, таких як Big Data, машинне навчання та BI-платформи. На відміну від Big Data, яке працює з неструктурованими даними та великими обсягами інформації, OLAP забезпечує швидкий доступ до агрегованих показників для стратегічного аналізу. Машинне навчання дозволяє виявляти складні закономірності та прогнозувати ефективність маркетингових кампаній, тоді як OLAP використовується для швидкої оцінки їхньої продуктивності. BI-платформи, такі

як Tableau чи Power BI, є інструментами для візуалізації даних, які часто інтегруються з OLAP-системами для створення аналітичних звітів.

Зміни у вимогах бізнесу до аналітики в партнерському маркетингу зумовлені зростаючою конкуренцією та необхідністю приймати швидкі рішення на основі даних. Компанії прагнуть до автоматизації процесів аналізу, що можливо завдяки AI-driven OLAP, де штучний інтелект допомагає визначати тренди, формувати звіти та пропонувати оптимальні стратегії. Збільшення масштабів партнерських програм також вимагає ефективних рішень для боротьби з шахрайством, і OLAP може бути використаний для моніторингу аномальних патернів у транзакціях та автоматичного виявлення підозрілих дій.

OLAP таким чином відіграє критичну роль у маркетинговій аналітиці, забезпечуючи швидкий доступ до структурованої інформації, інтеграцію з сучасними технологіями та можливість автоматизації процесів ухвалення рішень.

Багатовимірний аналіз у партнерському маркетингу. У партнерському маркетингу аналітика відіграє ключову роль у вимірюванні ефективності кампаній, оптимізації співпраці з партнерами та прогнозуванні майбутніх результатів. Використання OLAP дозволяє швидко аналізувати великі обсяги даних, отримувати багатовимірні зрізи інформації та приймати обґрунтовані рішення. Основними метриками, які застосовуються для оцінки ефективності партнерських програм, є ROI (повернення інвестицій), коефіцієнти конверсій та продуктивність окремих партнерів. ROI визначає, наскільки вигідною є конкретна кампанія, зрівнюючи витрати на партнерську рекламу до отриманого доходу. Аналіз конверсій допомагає зрозуміти, який відсоток відвідувачів, що перейшли за партнерським посиланням, завершили цільову дію, наприклад покупку або реєстрацію. Крім цього, OLAP дозволяє детально оцінювати ефективність окремих партнерів, виділяючи найприбутковіших та визначаючи фактори їхнього успіху.

Багатовимірний аналіз у партнерському маркетингу також дозволяє проводити сегментацію партнерів, що допомагає оптимізувати взаємодію з різними групами афіліатів. OLAP дає змогу розподіляти партнерів за рівнем продуктивності, джерелами трафіку, географічними регіонами або поведінковими показниками. Наприклад, компанія може визначити, які партнери залучають найбільш лояльних клієнтів або які рекламні канали приносять найвищий коефіцієнт конверсії. Завдяки цьому маркетингові бюджети можуть бути ефективніше розподілені, а комісійні винагороди адаптовані залежно від якості залучених потенціальних клієнтів, які можуть здійснити покупку (лідів). Ще одним важливим аспектом використання OLAP у партнерському маркетингу є виявлення шахрайства та маніпуляцій. Оскільки партнерські програми нерідко стикаються з фрод-активністю (підроблені кліки, фальшиві конверсії, використання ботів для генерації трафіку тощо), необхідні надійні механізми моніторингу аномальних патернів. OLAP дозволяє аналізувати дані за численними параметрами, такими як, порівняння середнього часу до

конверсії між різними партнерами або виявляти неприродні сплески активності. Використовуючи історичні дані, система може автоматично визначати партнерів із високим ризиком шахрайських дій та вживати відповідних заходів.

Крім оцінки поточних показників, OLAP допомагає у прогнозуванні результатів маркетингових кампаній. Завдяки аналізу історичних даних та виявленню трендів компанії можуть передбачати, які партнерські канали будуть найефективнішими в майбутньому, як зміниться поведінка аудиторії та які фактори вплинуть на загальну рентабельність. Наприклад, якщо аналіз показує, що певний сегмент партнерів демонструє стабільне зростання конверсій у певний сезон, компанія може заздалегідь скоригувати маркетинговий бюджет, щоб максимізувати прибуток.

Отже, OLAP у партнерському маркетингу виконує одразу кілька важливих функцій:

- забезпечує глибокий аналіз основних метрик;
- дозволяє сегментувати партнерів для кращого управління програмами;
- допомагає виявляти шахрайство та створює можливість прогнозування майбутніх маркетингових трендів.

Це дозволяє компаніям не лише оцінювати поточну ефективність своїх партнерських програм, а й приймати більш стратегічні рішення для підвищення прибутковості.

Методи та інструменти OLAP для оптимізації прибутку. Сучасний партнерський маркетинг неможливо уявити без аналітики великих обсягів даних. OLAP-технології дозволяють ефективно обробляти інформацію та ухвалювати стратегічні рішення, які безпосередньо впливають на прибутковість бізнесу. Існують різні підходи до OLAP-аналізу, серед яких найпоширенішими є MOLAP, ROLAP та HOLAP.

MOLAP використовує багатовимірні бази даних для швидкої обробки агрегованих показників, що особливо корисно для оцінки загальних тенденцій у партнерському маркетингу, наприклад, відстеження змін у коефіцієнті конверсії чи загальному прибутку за сегментами партнерів.

ROLAP, навпаки, працює безпосередньо з реляційними базами даних, що забезпечує більшу гнучкість і дозволяє аналізувати детальні показники, наприклад, розподіл клієнтів за джерелами трафіку або ефективність конкретних рекламних кампаній.

HOLAP об'єднує обидва підходи, дозволяючи поєднувати швидкість MOLAP та деталізацію ROLAP, що робить його оптимальним варіантом для складних партнерських мереж із великими обсягами інформації.

Розвиток OLAP-технологій йде в напрямку підвищення продуктивності та масштабованості аналітичних систем. Сучасні компанії все частіше використовують In-Memory технології, які зберігають та обробляють дані безпосередньо в оперативній пам'яті, що суттєво прискорює виконання складних запитів. Це особливо корисно для швидкого аналізу ефективності партнерських кампаній у режимі реального часу, коли важливо оперативно реагувати на зміни на ринку.

Іншим важливим трендом є перехід до хмарних рішень, що дозволяє компаніям відмовитися від локальних серверів і використовувати масштабовані ресурси AWS Redshift, Google BigQuery або Microsoft Azure Analysis Services. Завдяки цьому маркетингові команди можуть отримувати доступ до аналітики з будь-якої точки світу та працювати з великими обсягами даних без обмежень у продуктивності.

Інтеграція штучного інтелекту та машинного навчання значно розширює можливості OLAP-аналізу в партнерському маркетингу. Використовуючи алгоритми AI, компанії можуть автоматично визначати найефективніших партнерів, прогнозувати зміну рентабельності рекламних кампаній та навіть виявляти шахрайські дії, такі як фальшиві кліки або підроблені ліди. Наприклад, аналізуючи патерни поведінки партнерів та аномалії у звітах, AI-моделі можуть сигналізувати про підозрілі дії, що дозволяє оперативно блокувати зловживання та мінімізувати втрати компанії. Крім того, інтеграція OLAP із BI-системами на зразок Tableau, Power BI чи Looker забезпечує зручну візуалізацію аналітичних даних, що спрощує процес ухвалення рішень і допомагає краще розуміти ефективність партнерських програм.

Прикладом успішного використання OLAP у партнерському маркетингу є міжнародна e-commerce компанія, яка використовує OLAP-аналіз для моніторингу ефективності своїх афіліатів. Завдяки поєднанню In-Memory обробки та AI-аналітики компанія змогла зменшити час аналізу ключових показників ефективності (KPI) з декількох годин до лічених хвилин. Це дозволило їй швидко виявляти партнерів із низькою продуктивністю, перенаправляти бюджети на більш ефективні кампанії та запобігати шахрайству, що безпосередньо вплинуло на зростання прибутку.

Такі підходи демонструють, як сучасні OLAP-інструменти можуть не лише покращити аналітику, а й підвищити конкурентоспроможність компаній у партнерському маркетингу.

Таким чином, OLAP на даний момент залишається одним із найефективніших інструментів для оптимізації прибутку у партнерських програмах. Завдяки комбінації класичних та новітніх технологій, таких як In-Memory обробка, хмарні рішення, AI-аналітика та інтеграція з BI-системами, компанії отримують змогу значно підвищити швидкість та точність прийняття рішень. Це відкриває нові можливості для зростання бізнесу та оптимального використання маркетингових ресурсів.

Кейси та реальні дослідження. OLAP є потужним інструментом для аналізу даних, який широко використовується в різних галузях, включаючи партнерський маркетинг. Хоча конкретні приклади застосування OLAP у партнерському маркетингу можуть бути обмеженими в публічному доступі, існують відомі компанії, які активно використовують OLAP для аналізу даних та оптимізації своїх маркетингових стратегій.

Реальні приклади впровадження OLAP у партнерському маркетингу демонструють, як багатовимірний аналіз даних сприяє підвищенню ефективності бізнес-процесів, оптимізації витрат та покращенню

взаємодії з партнерами.

Корпорація «Оболонь» здійснила цифрову трансформацію, впровадивши OLAP для вдосконалення управління фінансами та партнерськими відносинами. Це дозволило компанії швидко та ефективно аналізувати великі обсяги даних у багатовимірному форматі, виконувати складні запити та проводити інтерактивний аналіз даних з різних перспектив, що значно прискорило процес прийняття рішень [16].

ВІТ Impulse, українська компанія, розробила інструмент для бізнес-аналітики, який спрощує роботу з великими обсягами даних. Ця система швидко впроваджується в компаніях, що використовують технології Microsoft, такі як MS OLAP, MS Dynamics та MS SQL. Вона формує звіти у зрозумілому вигляді (таблиці, графіки, діаграми) і забезпечує доступну візуалізацію аналітики, що дозволяє компаніям ефективніше аналізувати дані та приймати обґрунтовані рішення [17].

За неформальними оцінками деяких спеціалістів компанії, які використовують OLAP для аналізу маркетингових даних - скорочують час ухвалення рішень на 30-50% порівняно з традиційними методами аналізу даних, такими як Excel-звіти або стандартні реляційні бази даних без попередньо агрегованих даних, а використання OLAP в рамках інших налаштованих підсистем управління дозволяє підвищити точність прогнозування продажів на 25-40% у порівнянні з методами без використання OLAP, де аналітика будується на статичних звітах або ручних моделях без інтерактивного аналізу даних.

Ще приклад корисного використання OLAP - до впровадження OLAP маркетинговий відділ компанії міг використовувати окремі Excel-таблиці для аналізу партнерських каналів, що займало до 10 годин на підготовку щомісячного звіту, а після впровадження OLAP цей процес скорочується до 30 хвилин (при добре налаштованих інших суміжних процесах), а точність аналізу покращилася за рахунок автоматизованої обробки даних.

Також за оцінками спеціалістів компанії, які використовують OLAP для аналізу партнерських транзакцій, можуть зменшити рівень шахрайських дій у партнерських програмах на 20-30%. Це відбувається за рахунок автоматичного виявлення аномалій та аналізу патернів поведінки підозрілих афіліатів.

Інші відомі компанії також використовують або OLAP або OLAP пов'язані технології.

Наприклад, компанія Amazon, відома своєю партнерською програмою Amazon Associates, використовує аналітичні інструменти для відстеження та аналізу ефективності партнерських посилань, що дозволяє оптимізувати маркетингові стратегії та підвищувати прибутковість [18].

Ці аналітичні платформи можуть включати OLAP-підходи, для аналізу ефективності партнерських посилань. Це дозволяє компанії швидко оцінювати продуктивність афіліатів, сегментувати партнерів за рівнем залучення трафіку та перерозподіляти маркетингові бюджети.

Booking.com використовує потужності Amazon Web Services (AWS) та генеративний штучний інтелект для обробки понад 150 петабайт даних, покращуючи

сервіси бронювання. Співпрацюючи з AWS Professional Services та використовуючи технології, такі як Amazon SageMaker та Amazon Bedrock, компанія побудувала масштабовану платформу машинного навчання, що прискорює експерименти та інновації [18]. Booking.com, що має масштабну партнерську програму, працює з великими обсягами даних та аналітичними інструментами для управління комісіями та оптимізації конверсій. Використання OLAP-підходів тут може сприяти швидкому аналізу поведінкових факторів, що впливають на продажі. Ще одним потенційним прикладом є Uber, що активно використовує аналітику для оцінки ефективності своєї реферальної програми. Інтеграція складних аналітичних моделей у маркетингові системи дозволяє компанії коригувати винагороди для партнерів, аналізувати залучення нових водіїв і пасажирів, а також адаптувати стратегії для підвищення лояльності користувачів. Подібний підхід, з використанням багатовимірного аналізу даних, забезпечує компанії конкурентні переваги у сфері цифрового маркетингу.

Uber інтегрувала свої локальні операції з хмарними сервісами AWS, використовуючи такі рішення, як Amazon Athena та Amazon S3. За підтримки AWS Enterprise Support, Uber покращила операційну безпеку та знизила витрати, що сприяло оптимізації бізнес-процесів та підвищенню ефективності [19].

І таких прикладів можна навести чимало. Загалом, навіть якщо компанії не завжди відкрито повідомляють про використання OLAP, практика багатовимірного аналізу даних у партнерському маркетингу є загальноприйнятною. OLAP дозволяє маркетологам швидко оцінювати ефективність каналів залучення трафіку, прогнозувати майбутні результати та оптимізувати витрати. Впровадження подібних технологій дає змогу підвищити рентабельність маркетингових зусиль, зменшити втрати через неефективні партнерські програми та зробити процес прийняття рішень більш обґрунтованим. OLAP-системи самі по собі не є автономним джерелом прибутку, а виступають інструментом для підвищення ефективності управлінських процесів. Вони сприяють оптимізації окремих підсистем підприємства, що, у свою чергу, через взаємодію з більш масштабними управлінськими структурами компанії покращує її загальну продуктивність та фінансові показники. Саме тому дослідження рівня впливу OLAP на прибутковість бізнесу є перспективним напрямом подальших аналітичних розвідок. Як бачимо, можливості компаній розширюються не лише завдяки використанню OLAP, а й завдяки впровадженню суміжних технологій. І OLAP тут належить важливе місце. Це пояснюється тим, що OLAP-системи вже містять структуровані дані, які легко аналізувати та використовувати в режимі онлайн у різних розрізах. Вони є важливим кроком до максимальної ефективної аналітики, що підвищує стратегічні перспективи компаній.

Виклики та перспективи розвитку. Попри значні переваги OLAP у партнерському маркетингу, його впровадження супроводжується певними викликами. Однією з головних проблем є складність інтеграції OLAP із існуючими маркетинговими платформами та базами даних. Багато компаній працюють із розрізненими джерелами інформації, які можуть містити

неструктуровані або застарілі дані, що ускладнює створення єдиної аналітичної моделі. Крім того, для належного використання OLAP потрібні високі технічні компетенції, оскільки налаштування багатовимірних кубів і написання складних запитів може вимагати залучення фахівців із бізнес-аналітики та баз даних.

Безпека та конфіденційність даних також є критично важливими аспектами впровадження OLAP у партнерський маркетинг. Оскільки ці системи аналізують великі обсяги інформації про клієнтів, партнерів і транзакції, компанії мають дотримуватися стандартів захисту даних, таких як GDPR у Європі або CCPA у США. Незахищені OLAP-системи можуть стати ціллю для хакерських атак або витоку чутливої інформації, що загрожує не лише фінансовими втратами, а й репутаційними ризиками. У відповідь на ці загрози все більше компаній використовують хмарні рішення з вбудованими механізмами шифрування та багаторівневого доступу, що підвищує рівень безпеки даних.

У найближчі п'ять років OLAP-системи продовжуватимуть розвиватися, адаптуючись до зростаючих вимог бізнесу та технологічних трендів. Основним вектором змін стане перехід до реального часу обробки даних, що дозволить компаніям миттєво реагувати на зміни у поведінці партнерів і клієнтів. Іншою важливою тенденцією є інтеграція OLAP із штучним інтелектом та машинним навчанням, що допоможе автоматизувати аналіз великих обсягів даних і знаходити складні закономірності, які неможливо було б виявити вручну.

Розвиток хмарних OLAP-рішень також відіграватиме важливу роль, оскільки компанії все більше переходять на serverless-аналітику, що дозволяє масштабувати ресурси відповідно до навантаження без значних інвестицій у власну інфраструктуру. Використання візуалізаційних BI-інструментів, таких як Power BI, Tableau та Looker, спростуватиме аналіз даних та зробить його доступним не лише для технічних спеціалістів, а й для маркетологів та керівників.

Загалом, майбутнє OLAP у партнерському маркетингу виглядає перспективним. Збільшення потужностей обробки даних, автоматизація аналітики та покращення безпеки сприятимуть активному розвитку цієї технології, роблячи її ще більш ефективним інструментом для оптимізації бізнес-процесів і підвищення прибутковості маркетингових кампаній.

Висновки

Таким чином, дослідження показало, що OLAP-технології відіграють ключову роль у партнерському маркетингу, забезпечуючи глибокий аналіз великих обсягів даних у режимі реального часу. Вони дозволяють бізнесу не лише оцінювати ефективність партнерських програм, а й прогнозувати тенденції, оптимізувати бюджети та оперативно виявляти шахрайство. Використання багатовимірного аналізу забезпечує компаніям швидкість у прийнятті рішень, точність у сегментації партнерів і можливість адаптації до змін ринку.

З практичної точки зору бізнесу слід розглядати OLAP не як окремих інструмент, а як частину загальної системи бізнес-аналітики. Його ефективність зростає при інтеграції з CRM, BI-системами та штучним інтелектом. Компаніям, які працюють з партнерським

маркетингом, варто інвестувати у хмарні OLAP-рішення, що дозволяють масштабувати аналітику та забезпечують гнучкість у роботі з великими масивами даних. Важливим аспектом залишається безпека та конфіденційність інформації, особливо у зв'язку з ростом вимог щодо захисту персональних даних.

Подальші дослідження у цій сфері можуть зосередитися на оцінці реального економічного ефекту від використання OLAP у партнерському маркетингу,

зокрема на вимірюванні зростання прибутковості компанії після впровадження таких технологій.

Також перспективним напрямом є розширення можливостей OLAP через інтеграцію з машинним навчанням та штучним інтелектом для автоматизованого виявлення закономірностей у даних. Майбутній розвиток ринку OLAP-технологій сприятиме ще більшій персоналізації маркетингових стратегій і підвищенню їхньої ефективності.

СПИСОК ЛІТЕРАТУРИ

1. Черняк О. І., Захарченко П. В. Інтелектуальний аналіз даних: підр. Київ: Знання, 2010. 837 с. URL: <https://moodle.znu.edu.ua>
2. Братко О.С. Маркетингова політика комунікацій: Навчальний посібник. - Тернопіль: Карт-бланш, 2006. - 275 с. URL: <https://lib.kart.edu.ua/bitstream/123456789/8436/1/Навчальний%20посібник.pdf>,
3. Яцько О., Ватаманіца Е, Горський М. Особливості застосування OLAP-моделювання в освіті, Опт-ел. інф-енерг. техн., вип. 44, вип. 2, с. 5–12, Січ 2023 URL: <https://oeipt.vntu.edu.ua/index.php/oeipt/article/view/619>
4. Іванов Сергій. Моделювання вітрини даних розповсюдження – замовлення – продаж із використанням OLAP-технологій // Галицький економічний вісник, № 5 (72) 2021 с. 85-94 URL: <https://galicianvisnyk.tntu.edu.ua/pdf/72/1020.pdf>
5. Грибков С. В., Ханбабаєв Р. Р., Харкянен Є. В. Аналіз маркетингових заходів торговельної мережі методами Text Mining. // ХАРЧОВА ПРОМИСЛОВІСТЬ № 28, 2020. с.149-157 URL: <https://dspace.nuft.edu.ua/server/api/core/bitstreams/7f887d80-7566-47c9-9ef3-00d04984f597/content>
6. Paulo Orlando Queiroz-Sousa, Ana Carolina Salgado. A Review on OLAP Technologies Applied to Information Networks. 13 December 2019. URL: <https://dl.acm.org/doi/10.1145/3370912>
7. Fabian Maile. Artificial Intelligence and Big Data in Affiliate Marketing. 2018 URL: <https://opus-htw-aalen.bsz-bw.de/frontdoor/index/index/docId/400>
8. Ivanov S. Modeling of Digital Marketing Systems Construction Using OLAP-Technologies.2021 URL: https://www.researchgate.net/publication/356299631_Modeling_of_digital_marketing_systems_construction_using_olap_-_technologies
9. Dr. Rajeev Shrivastavaa, Dr. Rajesh Tiwari, Dr. Kamal Mehtac, Sarifta Bano. Various OLAP Technologies and Their Impact on Decision Making. 23 Apr 2021 URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3832711
10. Oskar Ahlqvist. Detecting Fraud in Affiliate Marketing. 2023 URL: <https://www.diva-portal.org/smash/get/diva2%3A1784368/FULLTEXT01.pdf>
11. Rosalee Wells. Affiliate Marketing: How to Master Data Collection URL: <https://azuramagazine.com/articles/affiliate-marketing-how-to-master-data-collection>
12. Elizabeth Sramek. Data Analysis Tools For Affiliate Marketing Business. April 13, 2021 URL: <https://www.scaleo.io/blog/data-analysis-tools-for-affiliate-marketing-business/>
13. Affiliate Marketing Programs: Data Analysis, by FasterCapital. 15 Jun 2024 URL: <https://www.fastercapital.com/content/Affiliate-marketing-programs--Data-Analysis--Data-Analysis-Techniques-for-Improving-Affiliate-Marketing.html>
14. Matt Diggity. 7 Affiliate Marketing Statistics. 2024. URL: <https://diggitymarketing.com/affiliate-marketing/statistics/>
15. Online Analytical Processing – das Multitalent für die mehrdimensionale Datenanalyse URL: <https://www.disy.net/de/beitraege/was-ist-online-analytics-processing-olap/>
16. Пшеничний О.«Оболонь» – в ногу з часом. Як диджиталізація змінила стратегію управління фінансами одного з найбільших виробників напоїв в Україні URL: <https://forbes.ua/company/obolon-v-nogu-z-chasom-yak-didzhitalizatsiya-zminila-strategiyu-upravlinnya-finansami-odnogo-z-naybilshikh-virobnikiv-napoiv-v-ukraini-09092024-23323>
17. Бережна Вікторія. Як перетворити складну роботу з великим обсягом даних на легкий і доступний процес: інструмент для бізнес-аналітики від української компанії URL: <https://itc.ua/ua/articles/yak-peretvoryty-skladnu-robotu-z-velykum-obsyagom-danyh-na-legkyj-i-dostupnyj-protseks-instrument-dlya-biznes-analitiky-vid-ukrayinskoyi-kompaniyi>
18. Booking.com on AWS URL: <https://marchenko.marketing/partnerskiy-marketing-vid-a-do-ya>
19. How Uber Optimizes Using AWS Enterprise Support URL: <https://aws.amazon.com/solutions/case-studies/uber-support-case-study/>

Received (Надійшла) 11.01.2025

Accepted for publication (Прийнята до друку) 05.03.2025

Modern OLAP methods in affiliates marketing: optimization of profits using multidimensional data analysis

I. Iliina, K. Kostin

Abstract. This article explores the possibilities and benefits of using OLAP technologies in affiliate marketing, in particular their impact on profit optimization through multidimensional data analysis. Modern OLAP methods are considered, including MOLAP, ROLAP, HOLAP, as well as the latest technological trends, such as In-Memory processing, cloud solutions and integration with artificial intelligence. Considerable attention is paid to practical aspects of using OLAP, in particular, partner segmentation, analysis of marketing campaign effectiveness and fraud detection. The article also presents real cases of implementing OLAP in business, including an analysis of successful companies that use this technology to improve the effectiveness of their affiliate programs. The challenges and prospects for the development of OLAP in marketing analytics are separately considered, including data security issues, integration with other analytical systems and potential development directions in the coming years. The study confirms that OLAP is an important element of modern digital marketing, which contributes to the rapid adoption of informed decisions and increasing the profitability of companies.

Keywords: OLAP, affiliate marketing, multidimensional data analysis, MOLAP, ROLAP, HOLAP, In-Memory technologies, cloud solutions, artificial intelligence.

А. А. Коваленко, Т. Г. Куценко, А. С. Шаповал, О. В. Ситник, Я. С. Ні

Харківський національний університет радіоелектроніки, Харків, Україна

ОГЛЯД МЕТОДІВ АНАЛІЗУ БЕЗПЕЧНОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

Анотація. У статті проведений детальний огляд методів аналізу безпечного програмного забезпечення. **Метою дослідження** є детальний огляд існуючих методів статичного і динамічного аналізу програм, які є базою для забезпечення безпеки. Для цього проводиться порівняльний аналіз функціональних можливостей різних методів і інструментів та виявляються основні недоліки. **Результати дослідження.** Представлений детальний огляд і порівняльний аналіз існуючих методів зіставлення вихідних і бінарних файлів, аналізу змін між версіями програмного забезпечення, пошуку витоків динамічної пам'яті і помилок використання звільненої пам'яті. Доведено, що незважаючи на велику кількість доступних методів, існують серйозні обмеження на клас завдань, для яких вони можуть бути ефективно застосовані. **Висновок.** Необхідно приборати існуючі обмеження, розвиваючи запропоновані методи. Доцільно розглядати методи не окремо, а в сукупності. При цьому треба враховувати вклад, який по окремоті приносять методи, в загальну картину забезпечення безпеки програм. Такий спільний розвиток та сумісне використання розглянутих у статті методів дозволить отримати більш якісний аналіз.

Ключові слова: безпечне програмне забезпечення, вразливість програми, витік інформації, зміни версії, бінарний код.

Вступ

Актуальність дослідження. Програмне забезпечення (ПЗ) швидко інтегрується в усі сфери життя. Воно використовується в персональних пристроях і критичних системах. З розвитком ПЗ зростають загрози безпеки. Обсяг і складність систем постійно збільшуються. Хакери використовують складні методи атак. Вони експлуатують вразливості програмних систем. Атаки включають фішинг і вразливості «нульового дня».

Згідно зі звітами NIST, кількість вразливостей зростає. За десять років вона зростає з 8 000 до 28 000 [1]. Це відбувається попри розвиток інструментів аналізу. Кількість і якість таких інструментів постійно зростають. Виявлення та усунення загроз стає пріоритетним завданням. Тому необхідні нові методи захисту програмного забезпечення. Для цього необхідно провести аналіз існуючих методів та виявити їх недоліки. Це і зумовлює актуальність даного дослідження.

Аналіз джерел. Великі компанії розуміють важливість розробки безпечного програмного забезпечення, тому активно створюють спеціалізовані рішення та платформи для аналізу коду відкритого доступу. Одним із найвідоміших прикладів є CodeQL від компанії GitHub, який дозволяє виконувати запити до коду та виявляти потенційні вразливості [2]. CodeQL обробляє програмний код як базу даних, що дає можливість моделювати вразливості та помилки у вигляді запитів. Цей підхід забезпечує використання як стандартних запитів, розроблених дослідниками GitHub і спільнотою, так і створення власних запитів для вузькоспеціалізованого аналізу. Бази даних CodeQL містять абстрактне синтаксичне дерево, графі потоку даних і управління, що надає всебічне представлення коду та є основою для глибокого аналізу його безпеки.

Google також активно розвиває напрям безпеки програмного забезпечення, створивши проєкт OSS-Fuzz, який дозволяє виконувати фазинг відкритого

коду на кластерних потужностях [3]. Цей проєкт було запущено після виявлення критичної вразливості Heartbleed в бібліотеці OpenSSL – одному з найпопулярніших інструментів для шифрування веб-трафіку [4]. Вразливість могла зачепити мільйони користувачів Інтернету та була викликана помилкою переповнення буфера пам'яті, яку можна було б виявити за допомогою фазингу. Однак на той момент цей метод тестування не був достатньо автоматизований і не отримав широкого розповсюдження серед розробників. Щоб змінити ситуацію, Google створив OSS-Fuzz, який автоматизував процес фазингу та зробив його доступним для відкритого програмного забезпечення.

OSS-Fuzz працює, запускаючи фазери для проєктів із відкритим вихідним кодом та сповіщаючи розробників про виявлені помилки, що дозволяє оперативно усувати уразливості.

З моменту свого запуску OSS-Fuzz став одним із найважливіших інструментів для спільноти розробників відкритого коду. Спочатку його було розроблено для мов C та C++, але згодом підтримка розширилася на інші безпечні для пам'яті мови, такі як Go, Rust і Python. Завдяки цьому проєкту вдалося виявити понад 22 000 критичних помилок у програмному забезпеченні, що свідчить про його ефективність та значення для забезпечення безпеки цифрових систем.

Окрім широкого спектру інструментів, необхідно також впроваджувати правильну методологію забезпечення безпеки програмного забезпечення, починаючи ще з етапу проєктування та розробки. Для цього застосовуються методи безпечного циклу розробки ПЗ, які дозволяють виявляти та усувати потенційні вразливості ще до того, як програмний продукт буде впроваджений у реальне використання. Одним із ключових елементів забезпечення комплексної безпеки є використання якісних інструментів аналізу коду, які допомагають розробникам знаходити та виправляти критичні помилки.

Однак, незважаючи на значну кількість доступних інструментів, що охоплюють різні аспекти

аналізу ПЗ, існує безліч невирішених питань, які суттєво впливають на рівень безпеки.

Методи статичного аналізу, що використовуються для пошуку вразливостей у програмному забезпеченні, не завжди враховують наявність відомих уразливостей у сторонніх бібліотеках та відкритих проектах. Це створює ситуації, коли вразливий код може бути включений у проєкт без відома розробників, що піддає систему потенційним загрозам. Кожного року виявляються десятки критичних вразливостей «нульового дня», які стають причиною мільйонів атак зловмисників на користувачів та організації. Згідно зі звітом Google Project Zero за 2023 рік, було знайдено понад 50 таких уразливостей у різних операційних системах і браузерах, що потенційно зачепили мільйони користувачів [5].

Різні інструменти аналізу коду широко застосовуються у багатьох сферах, включаючи виявлення вірусів, пошук застарілих бібліотек, ідентифікацію клонів коду з помилками та інші важливі завдання. Однак методи фазингу, які використовуються для тестування безпеки, мають серйозні обмеження, зокрема при генерації структурованих даних та побудові ланцюжків викликів системних або бібліотечних функцій. Це призводить до недостатнього покриття коду під час тестування. За результатами аналізу ефективності проєкту OSS-Fuzz, в середньому він охоплює менше 40% коду, що вказує на необхідність удосконалення підходів до фазингу. Крім того, наразі не існує централізованої платформи, яка б дозволяла об'єднувати кілька інструментів аналізу в єдину систему, а також забезпечувала ефективне збирання, збереження та зручне використання великих обсягів артефактів відкритого програмного забезпечення.

Мета та задачі дослідження. Метою цього дослідження є детальний огляд існуючих методів статичного і динамічного аналізу програм, які є базою для забезпечення безпеки. Для цього проводиться порівняльний аналіз функціональних можливостей різних методів і інструментів та виявляються основні недоліки.

Результати досліджень

1. Огляд методів аналізу виправлених помилок та вразливостей. У роботі [6] автори визначають, яке конкретне змінення призвело до виникнення помилки в програмі. Для цього вони аналізують систему управління помилками, скануючи зареєстровані проблеми. Далі для кожної помилки вони намагаються встановити, яке саме внесене зміна є її виправленням. Для цього аналізується історія змін у системі керування версіями програмного забезпечення, а також проводиться детальний розгляд логів комітів. У статті описуються два основні методи аналізу – синтаксичний та семантичний, які застосовуються у певній послідовності. Синтаксичний аналіз виконується першим і здійснює пошук шаблонів у вигляді номерів помилок із системи управління помилками, використовуючи регулярні вирази. Семантичний аналіз, у свою чергу, шукає додаткову інформацію про виправлення помилки в логах комітів на основі таких факторів:

- наявність номера помилки у логах комітів;
- опис помилки міститься безпосередньо у повідомленні коміту;
- автор коміту також є відповідальним за її виправлення;
- файли, прикріплені до завдання, також присутні у відповідному коміті.

Після виявлення коміту, що містить виправлення, дослідники намагаються визначити, яке саме попереднє зміна спричинило появу цієї помилки в проєкті. Головним недоліком даного методу є залежність від якості описів змін, які залишають розробники у коментарях до комітів. Це не завжди є достовірним джерелом інформації, оскільки розробники можуть не включати детальні відомості про помилку у логи комітів або навіть не реєструвати її у системі управління помилками. У підсумку, це призводить до численних пропущених зв'язків між помилками та відповідними змінами. У середньому, 54% виправлених помилок не можуть бути однозначно пов'язані з відповідними виправленнями у коді. Подібні проблеми спостерігаються і в інших дослідженнях. Наприклад, інструмент VagCache [7], що використовує аналогічні метрики, також стикається з подібними труднощами.

Окремий напрям досліджень присвячений пошуку відомих вразливостей у вихідному коді. У роботі [8] автори використовують базу відомих вразливостей CVE для пошуку схожих помилок у заданому програмному проєкті. Для цього застосовуються три підходи: текстове порівняння, аналіз лексем та порівняння синтаксичних абстрактних дерев. У статті [9] представлений метод пошуку вразливостей із використанням технологій глибокого навчання. У цьому підході пошук вразливостей розглядається як задача класифікації тексту, а для навчання використовується тестовий набір SARD, що містить великий обсяг даних про помилки. Дані витягуються з цього набору, після чого застосовується глибоке навчання для автоматичного розпізнавання потенційних вразливостей у коді.

У роботі [10] описано інструмент VCCFinder, який знаходить вразливості, використовуючи класифікатори SVM. Навчання цього інструменту здійснюється на основі аналізу змін у репозиторіях понад 60 проєктів. Для виявлення змін, які могли усунути вразливість, реалізовано два підходи. Перший алгоритм здійснює пошук у базах даних вразливостей CVE та перевіряє, чи містяться у них посилання на відповідні коміти виправлення. Другий алгоритм аналізує історію комітів у різних проєктах та шукає вказівки на ідентифікатори CVE. На основі отриманих комітів з виправленнями реалізовано алгоритм для автоматичного пошуку подібних помилок у коді.

2. Методи відстеження змін між версіями програмного забезпечення. Відстеження змін у програмному забезпеченні має широкий спектр практичних застосувань. Аналізуючи внесені зміни, можна визначити, чи не містять вони нові помилки або потенційні вразливості. У випадку програм із відкритим вихідним кодом отримати інформацію про зміни можна за допомогою простого синтаксичного

порівняння двох версій файлів. Одним із найбільш відомих інструментів для такого порівняння є утиліта diff, що використовується в операційних системах Linux. Аналогічний підхід застосовується у всіх сучасних системах керування версіями програмного забезпечення. Проте синтаксичне порівняння файлів надає лише інформацію про додані чи видалені рядки у файлах і не враховує структурні (семантичні) зміни, що можуть впливати на функціональність програми.

Існує багато досліджень, присвячених удосконаленню методів аналізу змін між версіями програмного забезпечення.

Наприклад, інструмент change distler [11] використовує підхід, заснований на порівнянні синтаксичного абстрактного дерева програмного коду. Ця технологія базується на попередніх дослідженнях, у яких зміни у програмному коді представляються у вигляді модифікацій в абстрактному синтаксичному дереві. Алгоритм, який застосовується в цьому інструменті, дозволяє знаходити різні типи змін, зокрема додавання, видалення, переміщення або оновлення окремих фрагментів коду.

Ще одним ефективним інструментом для аналізу змін у програмному забезпеченні є LSDIFF, що був представлений у дослідженні [12]. Основною перевагою цього інструменту є його здатність виявляти систематичні зміни у коді та групувати їх у відповідні категорії.

Такий підхід значно спрощує перегляд та аналіз змін, оскільки базується на дослідженнях, які доводять, що сегменти коду зі схожими структурними характеристиками, як правило, зазнають аналогічних модифікацій. Подібний підхід застосовується в інструменті ClDiff [13], який, як і LSDIFF, спрямований на згруповане представлення змін у програмному коді для полегшення їхнього подальшого аналізу.

Аналіз існуючих методів показує, що більшість рішень для пошуку змін між версіями програмного забезпечення засновані на аналізі логів комітів та систем управління помилками. Однак такий підхід має значні обмеження, оскільки він не дозволяє виявити всі можливі виправлення вразливостей у коді. Це створює потенційні ризики, оскільки вразливі компоненти можуть залишитися непоміченими та продовжувати використовуватися в майбутніх версіях програмного забезпечення.

3. Методи зіставлення вихідного і бінарного коду. Інструмент Pigaio [14] призначений для зіставлення символів вихідного коду із символами відповідного бінарного коду. Важливо відзначити, що цей інструмент не вимагає, щоб вихідний код попередньо компілювався або компонувався, що значно спрощує процес аналізу. Для виконання зіставлення Pigaio проходить через кілька основних етапів:

- розбір вихідного і бінарного коду;
- вилучення артефактів для кожної функції вихідного та бінарного коду;
- зіставлення функцій вихідного коду з функціями бінарного коду на основі отриманих артефактів і визначення коефіцієнта схожості для кожного зіставлення.

Для розбору вихідного коду використовується компілятор Clang. Артефакти, які витягуються із коду, включають:

- імена функцій;
- константні рядки;
- кількість циклів, умов, викликів функцій, зовнішніх і глобальних змінних;
- кількість операторів "switch" та кількість випадків для них;
- наявність рекурсивних викликів у функціях;
- список викликаних функцій.

Після попереднього зіставлення функцій вихідного і бінарного коду на основі артефактів застосовуються додаткові евристичні методи у разі, якщо не всі функції були зіставлені:

- зіставлення викликаних та викликаючих функцій: якщо дві функції були зіставлені, порівнюються також функції, які вони викликають.
- зіставлення найближчих функцій: якщо функції F1, F2 і F3 знаходяться в одному файлі вихідного коду, і функції F1 і F3 вже були зіставлені, то F2 порівнюється з функціями, адреси яких знаходяться між цими функціями у бінарному коді.
- зіставлення на основі рідко зустрічаючихся констант: для рядкових констант, що зустрічаються менше трьох разів, знаходяться функції у вихідному та бінарному коді, що їх використовують, після чого виконується їх порівняння та зіставлення.

Визначення коефіцієнта схожості для кожного зіставлення ґрунтується на технологіях машинного навчання. База для навчання була сформована вручну на основі аналізу спрацьовувань інструмента.

Узагальнюючи результати, можна сказати, що середній відсоток помилок першого роду становить 17,9%, тоді як помилки другого роду досягають 72,8%.

Додатково варто зазначити, що коли у бінарному файлі доступні імена функцій (тобто не було виконано команду "strip"), відсоток правильно зіставлених функцій значно зростає. Якщо ж інформація про імена функцій у бінарному файлі відсутня, середній відсоток помилок першого роду збільшується до 32%, а другого роду – до 80,2%.

Інший інструмент, RE-Source [158], також використовується для зіставлення бінарного коду з вихідним.

На вході він отримує лише бінарний файл, виконує пошук у спеціалізованих онлайн-сховищах (список наведено далі) і повертає інформацію про знайдені зіставлення.

Основні етапи онлайн-аналізу включають:

- вилучення характерних особливостей бінарного файлу, таких як константні значення операндів, імпортовані бібліотеки та виклики функцій, значення константних рядків;
- створення запиту на основі отриманих особливостей та його відправка у пошуковій системі для аналізу вихідного коду.

Ще одним підходом є використання інструмента CodeBin [15], що застосовує новий метод ідентифікації бінарних функцій шляхом порівняння їх із базою попередньо обробленого вихідного коду.

Головна ідея цього інструмента полягає у вилученні характеристик вихідного коду, які зберігаються навіть після компіляції та складання, і, як правило, не залежать від конкретної платформи, компілятора або рівня оптимізації.

Характеристики вилучаються як із вихідного коду, так і з бінарних файлів, після чого вони використовуються для зіставлення відповідних функцій.

Порівняння функцій здійснюється шляхом генерації унікальних відбитків, що базуються на наступних характеристиках:

- виклики функцій у вихідному коді;
- виклики API-функцій та стандартних бібліотек;
- кількість аргументів у функціях;
- складність потоку управління, що визначається за цикломатичною складністю;
- цілочисельні та рядкові константи.

Для зіставлення бінарного коду з вихідним використовується спеціалізований анований граф викликів функцій.

Кожна вершина цього графа містить додаткову інформацію про виклики функцій, використання змінних та інші важливі особливості коду. Процес зіставлення проходить через наступні кроки:

- витягуються характеристики вихідного коду;
- витягуються характеристики бінарного коду;
- виконується пошук бінарного анованого графа викликів функцій у базі вихідних графів для знаходження відповідностей.

Результати тестування CodeBin показують, що середній відсоток помилок першого роду становить 4,2%, а другого роду – 79,3%.

Водночас аналіз проведених досліджень виявив кілька недоліків цього підходу:

- підтримуються лише певні бібліотеки з фіксованими версіями;
- для підтримки нових бібліотек чи конкретних версій необхідно вручну додавати унікальні ідентифікатори;
- визначення версії здійснюється на основі спеціального рядка, у якому вказана версія, але такий рядок може бути відсутній у деяких випадках;
- метод не враховує потік даних, що може призвести до неточностей у зіставленні;
- метод має труднощі з правильною обробкою випадків, коли функції були вбудовані в інший код.

У ході дослідження було проведено порівняльний аналіз інструментів для зіставлення вихідного і бінарного коду, серед яких Pigaiois, RESource, CodeBin і Karta.

Початковий код CodeBin є закритим, що обмежує можливість його модифікації та інтеграції у власні розробки. Інструмент RESource має відкритий вихідний код, проте наразі не підтримується і потребує підключення до мережі для виконання онлайн-аналізу.

Важливо відзначити, що деякі кодові бази, з якими працював RESource, більше не доступні, що суттєво знижує його ефективність.

Інструмент Karta дозволяє зіставляти бінарний код із бібліотеками, при цьому його вихідний код є відкритим, що дає можливість розширення

функціоналу. Також Karta може відновлювати версії бібліотек, які були статично злиновані у бінарному файлі.

Проте середній відсоток помилок першого роду в Karta становить 4,2%, а помилок другого роду – 79,3%, що свідчить про його обмеження у точності зіставлення.

Інструмент Pigaiois приймає як вихідний, так і бінарний код, після чого здійснює зіставлення функцій. Його початковий код є відкритим, що дозволяє його вдосконалення.

Водночас середній відсоток помилок першого роду у Pigaiois складає 17,9%, а другого роду – 72,8%.

Основними недоліками інструменту є неможливість коректного аналізу вбудованих функцій та відсутність підтримки аналізу потоку даних. Попри це, Pigaiois продовжує активно розвиватися та залишається одним із найкращих методів зіставлення серед доступних рішень.

Аналіз отриманих результатів показує, що завдання якісного зіставлення вихідного та бінарного коду залишається актуальним і потребує подальших досліджень для розробки більш точних та ефективних методів.

4. Методи пошуку витоків динамічної пам'яті.

Інструмент SMOKE [16] використовує два основних етапи для досягнення високої точності та масштабованості.

На першому етапі застосовується легкий, але менш точний аналіз для ідентифікації всіх можливих шляхів витоків пам'яті, при цьому відсіюються шляхи, які не можуть призвести до реального витоку. Для цього будується спеціальне програмне подання, яке називається графом потоку використання динамічної пам'яті (use-flow graph). Цей граф містить всю необхідну інформацію про потік управління для об'єктів купи, включаючи послідовність їх використання. Кожне ребро графа ановане умовами, що визначають можливість переходу. SMOKE також моделює життєвий цикл покажчиків динамічної пам'яті, створюючи спеціальний вузол "out-of-score", який позначає момент, коли об'єкт купи більше не використовується.

На другому етапі застосовується більш точний аналіз. Спочатку інструмент ідентифікує всі шляхи, які ведуть від точки виділення пам'яті до вузла "поза областю видимості" без виконання операції її звільнення.

Потім для кожного з виявлених шляхів застосовується вирішувач обмежень Z3 [17], щоб перевірити їхню здійсненність та відфільтрувати хибні спрацьовування. SMOKE забезпечує чутливість до потоку, шляхів і контексту програми, але має кілька недоліків:

- відсутня чутливість до полів об'єктів;
- аналіз покажчиків у графі потоку використання пам'яті є неточним;
- не враховує арифметичні операції над покажчиками;
- не підтримує аналіз бібліотечних функцій.

Інструмент PCA [18] базується на платформі LLVM [19] і виконує аналіз у кілька етапів. Спочатку

вихідні файли компілюються у проміжне подання LLVM. Потім, використовуючи компонувальник LLVM Gold, усі проміжні файли об'єднуються в єдине подання, після чого виконується аналіз показників за алгоритмом Андерсена. Отримані дані використовуються для побудови графа викликів, який включає як прямі, так і непрямі виклики функцій. На завершальному етапі створюється міжпроцедурний граф залежностей даних. Головний недолік PCA – він забезпечує лише часткову чутливість до потоку управління.

Інструмент SVF [20], також побудований на базі LLVM, виконує подібні кроки: компіляція у проміжне подання, об'єднання проміжних файлів за допомогою компонувальника LLVM Gold і запуск модуля аналізу показників (підтримуються різні алгоритми, зокрема аналіз Андерсена). SVF дозволяє розбивати пам'ять на окремі області, що значно покращує масштабованість і дає змогу аналізувати великі програми, а також застосовувати спеціалізовані детектори на основі графів потоку значень (ГПЗ).

Незважаючи на переваги, цей інструмент має такі обмеження:

- відсутня чутливість до полів та шляхів виконання програми;
- погано масштабується для аналізу великих програм.

Інструмент Fastcheck [21] застосовує міжпроцедурний алгоритм для виявлення витоків пам'яті в програмах мовою Сі. Він аналізує потік значень від точок виділення динамічної пам'яті до точок їхнього звільнення, використовуючи розріджене представлення програми у вигляді графа потоку значень. Ребра графа анотовані умовами розгалуження, що визначають можливість присвоєння значень.

Інформація про виклики та повернення функцій додається до ребер, що забезпечує контекстно-чутливий аналіз.

Пошук витоків пам'яті зводиться до задачі досяжності у графі потоку значень.

Інструмент Clang Static Analyzer (CSA) [183] використовує метод символічного виконання для пошуку різноманітних помилок, включаючи витoki пам'яті. Він забезпечує міжпроцедурний аналіз і чутливість до шляхів виконання. В основі аналізу лежить розібраний граф (РГ), який містить інформацію про можливі варіанти виконання програми. Для пошуку витоків пам'яті використовуються спеціалізовані детектори, які перевіряють, чи існують шляхи виконання, на яких не виконується операція звільнення пам'яті. Якщо такі шляхи виявлені, детектор формує відповідне повідомлення про помилку. Однак CSA має два критичні обмеження:

- цикли виконуються обмежену кількість разів (за замовчуванням 4), після чого РГ не розширюється, і подальші інструкції не аналізуються;
- кількість шляхів зростає експоненційно залежно від кількості інструкцій розгалужень, що робить аналіз великих програм малоефективним.

Інструмент Infer [22] використовується для аналізу програм, написаних мовами Java і Сі/Сі++, і дозволяє виявляти витoki пам'яті та використання нульових показників. Його особливістю є висока чутливість до потоку, шляхів виконання, полів та контексту програми, що робить його ефективним у виявленні помилок.

Інструмент PML Checker [23] працює шляхом лексичного та синтаксичного аналізу вихідного коду програми, після чого будується абстрактне синтаксичне дерево (АСД). Для остаточного аналізу використовується символічний вирішувач, що дозволяє виявляти потенційні витoki пам'яті. Цей інструмент забезпечує чутливість до потоку, шляхів виконання та полів програми.

У табл. 1 наведено результати тестування інструментів.

Таблиця 1– Результати порівняння чутливості та здатності до виявлення витоків інструментів

Інструмент	Чутливість до потоку	Чутливість до шляхів	Чутливість до контексту	Чутливість до полів	Витік-1	Витік-2
CSA	Так	Так	Так	частково	Ні	Так
Infer	Так	Так	Так	частково	Ні	Так
SMOKE	Так	Так	Так	Ні	Ні	частково
PCA	Ні	Ні	Ні	Ні	Ні	Ні
Fastcheck	Так	Так	частково	Ні	частково	Так
SVF	Так	-	Так	Ні	Ні	Ні
PML Checker	Так	Так	Так	Так	Ні	частково

Аналіз даних, отриманих в результаті тестування, показує, що жоден із доступних інструментів не здатен виявляти витoki пам'яті з достатньо високою точністю та не володіє максимально можливою чутливістю.

Це підкреслює актуальність подальшої розробки вдосконалених методів пошуку витоків пам'яті.

Висновки

У статті представлений детальний огляд і порівняльний аналіз існуючих методів зіставлення вихідних і бінарних файлів, аналізу змін між версіями програмного забезпечення, пошуку витоків динамічної пам'яті і помилок використання звільненої пам'яті.

Незважаючи на велику кількість доступних методів, існують серйозні обмеження на клас завдань, для яких вони можуть бути ефективно застосовані.

Необхідно прибирати ці обмеження, розвиваючи запропоновані методи та доцільно розглядати їх

не окремо, а в сукупності, враховуючи вклад, який по окремоті привносять методи, в загальну картину забезпечення безпеки програм. Такий спільний розвиток та сумісне використання розглянутих у статті методів дозволить отримати нову якість аналізу.

СПИСОК ЛІТЕРАТУРИ

1. NIST CVE report. URL: https://nvd.nist.gov/vuln/search/statistics?form_type=Basic&results_type=statistics&search_type=all&isCpeNameSearch=false
2. Github. URL: <https://github.com>
3. Oss-fuzz. URL: <https://github.com/google/oss-fuzz>
4. Openssl. URL: <https://www.openssl.org>
5. Google-zero. URL: <https://googleprojectzero.blogspot.com/2022/04/the-more-you-know-more-you-know-you.html>
6. Sliwinski J., Zimmermann T., Zeller A. When Do Changes Induce Fixes, *ACM SIGSOFT Software Engineering Notes*, vol. 3, no. 4. 2005. DOI: <https://doi.org/10.1145/1082983.1083147>
7. F. Rahman, D. Posnett, A. Hindle, E. Barr, P. Devanbu, "BugCache for Inspections: Hit or Miss?", *Proceedings of the 19th ACM SIGSOFT symposium and the 13th European conference on Foundations of software engineering*, 2011. DOI: <https://doi.org/10.1145/2025113.2025157>
8. J. Yang, X. Song, Y. Xiong, Y. Meng, "An Open Source Software Defect Detection Technique Based on Homology Detection and Pre-identification Vulnerabilities", *International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, 2018. DOI: https://doi.org/10.1007/978-3-319-93554-6_94
9. A. Xu, T. Dai, H. Chen, Z. Ming, W. Li, "Vulnerability Detection for Source Code Using Contextual LSTM", *5th International Conference on Systems and Informatics*, 2018. DOI: <https://doi.org/10.1109/ICSAI.2018.8599360>
10. H. Perl, S. Dechand, M. Smith, D. Arp, F. Yamaguchi, K. Rieck, S. Fahl, Y. Acar, "VCCFinder: Finding Potential Vulnerabilities in Open-Source Projects to Assist Code Audits", *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications*, 2015. DOI: <https://doi.org/10.1145/2810103.2813604>
11. Change Distiller. URL: <https://bitbucket.org/sealuzh/tools-changedistiller/src/master>
12. M. Kim, D. Notkin, "Discovering and representing systematic code changes", *International Conference on Software Engineering (ICSE)*, pp. 309-319, 2009. URL: <https://web.cs.ucla.edu/~miryung/Publications/icse09-lsdiff.pdf>
13. Kaifeng Huang, Bihuan Chen, Xin Peng, Daihong Zhou, Ying Wang, Yang Liu, Wenyun Zhao, "CIDiff: Generating Concise Linked Code Differences", *Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering*, 2018. DOI: <https://doi.org/10.1145/3238147.3238219>
14. Pigaios. URL: <https://github.com/joxeankoret/pigaios>
15. ON MATCHING BINARY TO SOURCE CODE. <https://users.encs.concordia.ca/~mmannan/student-resources/Thesis-MASc-Shahkar-2016.pdf>
16. G. Fan, R. Wu, Q. Shi, X. Xiao, J. Zhou, C. Zhang, "SMOKE: Scalable Path-Sensitive Memory Leak Detection for Millions of Lines of Code", *International Conference on Software Engineering (ICSE)*, 2019. URL: https://gangfan.github.io/assets/papers/gang_smoke_icse2019_preprint.pdf
17. Z3. URL: <https://github.com/Z3Prover/z3>
18. W. Li, H. Cai, Y. Sui, D. Manz, "PCA: memory leak detection using partial call-path analysis", *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 2020. DOI: <https://doi.org/10.1145/3368089.3417923>
19. Andersen pointer analysis. URL: <https://github.com/grievetia/andersen>
20. SVF. URL: <https://github.com/SVF-tools/SVF>
21. S. Cherem, L. Princehouse, R. Rugina, "Practical memory leak detection using guarded value-flow analysis", *Proceedings of the 28th ACM SIGPLAN Conference on Programming Language Design and Implementation*, 2007. DOI: <https://doi.org/10.1145/1250734.1250789>
22. Fbinfer. URL: Available: <https://fbinfer.com>
23. X. Sun, S. Xu, C. Guo, J. Xu, N. Dong, X. Ji, S. Zhang, "A Projection-Based Approach for Memory Leak Detection", *IEEE 42nd Annual Computer Software and Applications Conference*, 2018. <https://doi.org/10.1109/COMPSAC.2018.10271>

Received (Надійшла) 12.01.2025

Accepted for publication (Прийнята до друку) 05.03.2025

Overview of secure software analysis methods

A. Kovalenko, T. Kutsenko, A. Shapoval, O. Sytnyk, Ya. Ni

Abstract. The article provides a detailed review of methods for analyzing secure software. The purpose of the study is to provide a detailed review of existing methods for static and dynamic program analysis, which are the basis for ensuring security. For this purpose, a comparative analysis of the functional capabilities of various methods and tools is carried out and the main shortcomings are identified. Research results. A detailed review and comparative analysis of existing methods for comparing source and binary files, analyzing changes between software versions, searching for dynamic memory leaks and errors in using freed memory are presented. It is proved that despite the large number of available methods, there are serious limitations on the class of tasks for which they can be effectively applied. Conclusion. It is necessary to remove existing limitations by developing the proposed methods. It is advisable to consider the methods not separately, but in combination. In this case, it is necessary to take into account the contribution that the methods individually bring to the overall picture of ensuring program security. Such joint development and joint use of the methods considered in the article will allow for a higher-quality analysis.

Keywords: secure software, program vulnerability, information leak, version changes, binary code.

Я. І. Бірук, А. В. Климчук

Київський національний університет будівництва і архітектури, Київ, Україна

ДОСЛІДЖЕННЯ УМОВ МАКСИМАЛЬНОЇ ЕФЕКТИВНОСТІ ПЕРФОРОВАНИХ ЕЛЕКТРОМАГНІТНИХ ЕКРАНІВ

Анотація. Визначені основні співвідношення для розрахунку ефективності екранування електромагнітних полів у дальній зоні металевими матеріалами і конструкціями з отворами круглої та прямокутної форми. Враховане відбиття електромагнітних хвиль від перфорованої поверхні. Показано, що на отворах з кінцевою товщиною стінки можливі виникнення частотних смуг прозорості конструкції внаслідок резонансних явищ. Це впливає з теорії хвилеводів. Товщину стінки екрана вважати хвилеводом такої довжини. Проведено дослідження залежності ефективності екранування від частоти екранованого електромагнітного поля. Встановлено наявність частотних смуг прозорості екрана на частотах, нижчих за частоти зрізу для даних геометричних характеристик отворів. Отримані співвідношення для розрахунку внеску у екранування суцільних ділянок перфорованої конструкції. Співвідношення враховують електрофізичні властивості матеріалу екрана, частоту екранованого електромагнітного поля та глибину проникнення у матеріал. Враховуючи емпіричний характер цих співвідношень і можливі додаткові чинники можливих змін екрануючих властивостей (вплив регулярно розташованих вентиляційних отворів) доцільне тестування спроектованих конструкцій у реальних умовах експлуатації. Перспективним напрямом досліджень є удосконалення форми екранів шляхом подовження глибини отворів.

Ключові слова: екранування, перфорація, хвилевод, електрофізичні властивості.

Вступ

На сьогодні найбільш поширеними є металеві і композиційні матеріали для екранування електромагнітних полів та випромінювань. При цьому застосування суцільних металевих екранів обмежене через високі коефіцієнти відбиття електромагнітних хвиль. Усі сучасні екрануючі матеріали мають різні ступені неоднорідності. Композиційні матеріали, принаймні двокомпонентні – діелектрична матриця з екрануючим електропровідним або магнітним наповнювачем. Металеві структури, як правило, сітчасті і розраховані на екранування електромагнітного випромінювання певної частоти та амплітуди. Але у реальних умовах експлуатації доцільно використовувати екрануючі поверхні з перфорацією різної форми та щільності розташування. Наприклад, для феромагнітних матеріалів це може забезпечити високі коефіцієнти екранування магнітного поля наднизької частоти і знижує коефіцієнти відбиття електромагнітних випромінювань ультрафіолетових і вищих частот. Крім того, у багатьох випадках отвори потрібні для прокладання комунікацій, охолодження обладнання тощо. Тому доцільно дослідити ефективність екранування перфорованих структур і визначити закономірності змін коефіцієнтів екранування у залежності від частоти випромінювання.

Огляд літературних джерел. Розробленню неоднорідних матеріалів і конструкцій приділяється багато уваги. В основному це композиційні матеріали на основі мікро- та наноструктур [1]. Недоліками таких матеріалів є висока вартість та схильність до деградації в процесі експлуатації [2]. Деградація полімерних матриць можлива не тільки внаслідок внутрішніх процесів, а й під впливом високих температур у разі екранування промислового обладнання. Це стосується і облицювання зовнішніх поверхонь будівель [3].

У дослідженні [4] показано можливість екранування магнітної складової електромагнітного поля промислової частоти сітчастими конструкціями. Але така конструкція повинна мати великі розміри і бути надійно заземленою. У роботі [4] узагальнено існуючі підходи до застосування металевих екранів усіх конструкцій та визначено недоліки. Зокрема показано, що сітчасті структури фактично монохромні. А широка смуга поглинання електромагнітних випромінювань досягається за рахунок ускладнення конструкції. Методологія проектування неоднорідних електромагнітних екранів показана у [5]. Але частина результатів немає достатнього обґрунтування. Також відсутнє експериментальне підтвердження теоретичних даних. У роботі [6] показано, що заходи з екранування електромагнітних полів і випромінювань необхідно здійснювати на основі електромагнітного моніторингу, що обумовлено змінами амплітудно-частотних характеристик цього фактора у просторі у часі. Дослідження [7] свідчить, що ефективність екранування перфорованої металеві структури можна розрахувати з прийнятною похибкою, але не наведено верифікації отриманих результатів. Дослідження щодо неоднорідності великих розмірів, арматури й діелектричних каркасів будівельних конструкцій, [8] свідчить що, для достатньої ефективності такі структури повинні мати велику товщину що не завжди прийнятно. У статті [9] наведено результати теоретичних досліджень щодо екранування електромагнітних випромінювань неоднорідними металевими панелями. Зокрема показано, що зручні у застосуванні співвідношення строго випливають з теорії хвилеводів. Але відомо, що у хвилеводах можливе виникнення резонансних явищ, що вимагає експериментальних досліджень щодо частотної залежності коефіцієнтів екранування. Наведене обумовлює актуальність дослідження.

Викладення основного матеріалу

Врахування наявності отворів у екрануючій конструкції доцільне для частот, нижчих за частоту «зрізу», тобто для електромагнітних хвиль, для яких екран непроникний. При цьому розглядається дальня зона електромагнітного поля. Частоти зрізу строго розраховуються, виходячи з теорії хвилеводів [9]. Для прямокутного отвору $f_3 = 1,5 \cdot 10^8 / b$ Гц, де b – більший розмір отвору. Для отвору круглої форми $f_3 = 1,75 \cdot 10^8 / D$ Гц, де D – діаметр отвору.

При цьому втрати на поглинання для частот, нижчих за f_3 приблизно дорівнюють:

Для прямокутного отвору

$$A = 27 \frac{d}{b}, \text{ дБ},$$

де d – товщина стінки екрануючої конструкції;

Для круглого отвору

$$A = 32 \frac{d}{D}, \text{ дБ}.$$

Обов'язковим є врахування відбиття електромагнітних хвиль R .

Для круглого отвору:

$$R = 102 - 20 \lg(Df), \text{ дБ},$$

Для прямокутного отвору:

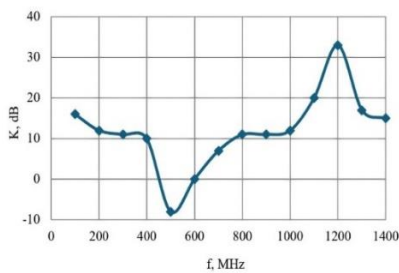
$$R = 100 - 20 \lg(bf) + 20 \lg\left(1 + \ln \frac{b}{a}\right), \text{ дБ},$$

де b, a – ширина та висота прямокутного отвору.

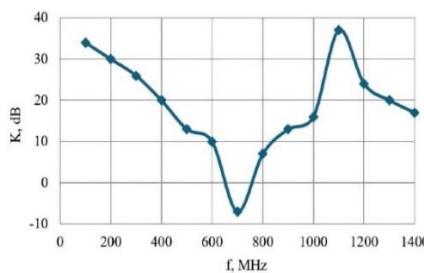
Але, як показано у [10], проходженню електромагнітної хвилі крізь хвилевод притаманні резонансні явища. Це обумовлює те, що хвилевод стає понадмежним за умови $m\lambda / 2a > \sqrt{\epsilon}$, де ϵ – діелектрична проникність середовища, a – висота хвилевода, λ – довжина електромагнітної хвилі, m – натуральне число. Саме різним значенням m відповідають коливання інтенсивності випромінювання.

Для визначення можливих вікон прозорості електромагнітного екрана з отвором на певних частотах були проведені випробування кількох конструкцій з отворами.

Досліджувалися замкнені металеві конструкції, що виключало проникнення електромагнітного поля поза екраном. На рис. 1, а наведено залежність коефіцієнта екранування конструкції зі щілиною розміром 15×5 см від частоти екранованого поля.



а – 15×5 см



б – 10×1 см

Рис. 1. Залежність коефіцієнта екранування металеві конструкції з отвором розміром від частоти екранованого електромагнітного поля

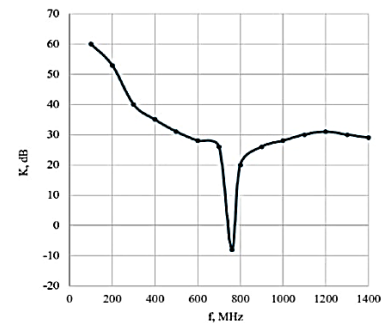


Рис. 2. Результати випробувань конструкції з отвором розмірами 20×3 см

Як видно з наведених даних, конструкція є радіопрозорою на частоті приблизно 400–600 МГц.

Аналогічні дослідження проведені для отвору розміром 10×1 см (рис. 1, б).

На рис. 3 наведено результати випробувань конструкції з отвором розмірами 20×3 см.

Аналізуючи отримані дані, що за різних розмірів отворів спостерігаються частотні смуги прозорості конструкції для електромагнітного поля. Це, очевидно пов'язане з резонансними явищами. Визначити строгу залежність смуги пропускання електромагнітних хвиль у залежності від розмірів отвору складно й недоцільно. Значною мірою цей показник залежить від товщини матеріалу, яку можна вважати довжиною хвилевода. Найбільш раціональним є попередні випробування конструкцій з урахуванням амплітудно-частотних характеристик електромагнітного випромінювання, яке потребує екранування. Ефективність екранування електромагнітного поля може знижуватися внаслідок проникнення поля безпосередньо крізь екрануючий матеріал. При цьому цей показник істотно залежить від частоти поля. Підвищення частоти

знижує проникність поля у провідний матеріал, тобто зменшується товщина скін-шару.

Для регулярно розташованих отворів поправочний коефіцієнт на проникнення поля через матеріал між отворами можна визначити з емпіричного співвідношення, наведено у [11]:

$$K = -20 \lg\left(1 + 35 p^{-2,3}\right), \text{ дБ}; \quad p = \frac{l}{\delta},$$

де l – довжина матеріалу між отворами, δ – товщина скін-шару.

Товщина скін-шару (глибина проникнення поля у матеріал) визначається співвідношенням:

$$\delta = c / \left(2\pi \sqrt{\sigma f}\right),$$

де c – швидкість розповсюдження електромагнітних хвиль, σ – питома провідність матеріалу екрана.

Таким чином, поправочний коефіцієнт визначається як:

$$K = -20 \lg\left(1 + 35 \left(2\pi l \sqrt{\sigma f} / c\right)^{-2,3}\right), \text{ дБ}.$$

Наведені співвідношення є емпіричними, тому їх застосування доцільне для попереднього оцінювання ефективності екранування з урахуванням амплітудно-частотних характеристик екранованого поля. Слід врахувати, що вони виконуються тільки для дальньої зони електромагнітного поля.

У процесі проектування захисної конструкції можуть виникнути додаткові проблеми, наприклад, вплив великої кількості регулярно розташованих отворів для забезпечення вентиляції та охолодження обладнання. Тому для забезпечення нормативних умов перебування людей та забезпечення стабільної роботи електронного обладнання необхідне випробування дослідного зразка екрануючого матеріалу або конструкції у реальних умовах експлуатації. Враховуючи, що для перфорованих захисних конструкцій можливо застосувати теорію хвилеводів, перспективним напрямом досліджень є визначення можливості розроблення конструкцій з певною формою отворів, яка збільшить довжину хвилевода. Такі конструкції сприятимуть підвищенню ефективності екранування без зниження сумарного перерізу отворів.

Висновки

1. Визначено співвідношення для розрахунків ефективності екранування електромагнітних полів у хвильовій зоні з використанням металевих матеріалів отворами. Показано, що за різних геометричних конфігурацій отворів можливі наявності смуг прозорості екрануючих конструкцій для електромагнітних випромінювань.

2. Проведені дослідження частотної залежності ефективності екранування для отворів різної довжини та ширини. Встановлено немонотонність таких залежностей з наявністю зон прозорості, що обумовлене резонансними явищами у отворах матеріалів кінцевої товщини.

3. Наведено співвідношення для розрахунків внеску поглинання електромагнітної енергії металевим матеріалом з урахуванням його електрофізичних властивостей. Наголошено, що враховуючи емпіричність співвідношень, доцільне випробування спроектованих матеріалів і конструкцій у реальних умовах експлуатації.

СПИСОК ЛІТЕРАТУРИ

- Kolcunová I., Zbojovský J., Pavlík M., Bucko S., Labun J., Hegedus M., Vavra M., Cimbala R., Kurimský J., Dolník B., Petráš J., Džmura J. Shielding Effectiveness of Electromagnetic Field by Specially Developed Shielding Coating. *Acta Physica Polonica*, 2020, Vol. 137 Is. 5, p. 711-713. <http://doi.org/10.12693/APhysPolA.137.711>
- Caramitu A.R., Ion I., Bors A.M., J. Pinte, Caramitu A.-M.D. Preparation and Spectroscopic Characterization of Some Hybrid Composites with Electromagnetic Shielding Properties Exposed to Different Degradation Factors. *MATERIALE PLASTICE*. 2023. 59. 82-94 <https://doi.org/10.37358/MP.22.4.5627>
- Glyva V., Levchenko L., Panova O., Tykhenko O., Radomska M. The composite facing material for electromagnetic fields shielding. *Innovative Technology in Architecture and Design (ITAD 2020): IOP Conf. Series*. 2020. Vol. 907.
- Спосіб виготовлення композиційного матеріалу для екранування електромагнітного поля. Пат. 144972, Україна МПК G12B 17/00 (2020.01); (2006.01) G21F 1/02. № 2020 03227; заявл. 28.05.2020; опубл. 10.11.2020, Бюл. № 21. 4 с.
- Глива В.А., Тихенко О.М., Ходаковський О.В. Методологія проектування неоднорідних електромагнітних екранів. Системи управління, навігації та зв'язку. Полтава, 2019. Вип. 4(56). С. 122–125.
- Касаткіна Н.В., Левченко Л.О., Панова О.В., Тихенко О.М., Ченчевой В.В. Оптимізація параметрів екранування електромагнітних полів різномірних джерел у виробничих будівлях. *Вісті ДГП*. 2020. Вип. № 1(46). С. 181–188.
- Касаткіна Н.В., Тихенко О.М., Фурсенко О.М. Розрахункові методи проектування електромагнітних екранів із заданими захисними властивостями. *Системи управління, навігації та зв'язку*. 2020. Вип. 2(620). С. 118–121.
- Burdeina, N., Levchenko, L. and oth.. (2024). Applying heterogeneous building materials for the protection of people against electromagnetic radiation. *EEJET*, 5(10 (131)), 45–52. <https://doi.org/10.15587/1729-4061.2024.313629>
- Biruk, Y., & Klymchuk, A. (2024). Засади проектування неоднорідних електромагнітних екранів. *Системи управління, навігації та зв'язку. Збірник наукових праць*, 3(77), 199–202. <https://doi.org/https://doi.org/10.26906/SUNZ.2024.3.199>
- Пенкин Ю.М., Бердник С.Л., Катрич В.А., Нестеренко М.В., Пшеничная С.В. Управление энергетическими характеристиками волноводных сочленений с диэлектрическим включением. *Прикладная радиоэлектроника*, 2017, Т. 6, № 1,2, с. 51-60.
- Демковський Д.В., Фомин И.А., Марченко М.В. Автоматизация расчёта эффективности экранирования» *Технологии ЭМС*, №1, 2013, №1, с. 44-54.

Received (Надійшла) 28.11.2024

Accepted for publication (Прийнята до друку) 26.02.2025

Study of conditions for maximum efficiency of perforated electromagnetic screens

Y. Biruk, A. Klymchuk

Abstract. The basic relations for calculating the efficiency of shielding electromagnetic fields in the far zone by metal materials and structures with round and rectangular holes are determined. The reflection of electromagnetic waves from the perforated surface is taken into account. It is shown that at the holes with a finite wall thickness, frequency bands of transparency of the structure may occur due to resonant phenomena. This follows from the theory of waveguides. The thickness of the screen wall is considered to be a waveguide of this length. The dependence of the shielding efficiency on the frequency of the shielded electromagnetic field is investigated. The presence of frequency bands of screen transparency at frequencies lower than the cutoff frequencies for these geometrical openings is established. The relations for calculating the contribution to the shielding of continuous sections of a perforated structure are obtained. The relations take into account the electrophysical properties of the shield material, the frequency of the shielded electromagnetic field, and the depth of penetration into the material. Taking into account the empirical nature of these relations and possible additional factors of possible changes in shielding properties (the influence of regularly spaced ventilation holes), it is advisable to test the designed structures in real operating conditions. A promising area of research is to improve the shape of the screens by lengthening the depth of the holes.

Keywords: shielding, perforation, waveguide, electrophysical properties.

Н. Б. Бурдейна¹, Л. А. Зозуля², Т. Б. Петруньок¹

¹ Київський національний університет будівництва і архітектури, Київ, Україна

² Державний університет «Київський авіаційний інститут», Київ, Україна

ЗАСТОСУВАННЯ КОМПОЗИЦІЙНОГО ЗАЛІЗОВМІСНОГО МАТЕРІАЛУ ДЛЯ ЕКРАНУВАННЯ ЕЛЕКТРОМАГНІТНИХ ПОЛІВ УЛЬТРАВИСОКИХ І ВИЩИХ ЧАСТОТ

Анотація. Досліджено зміни коефіцієнтів проходження, відбиття та поглинання електромагнітних випромінювань композиційними матеріалами із масовим вмістом карбонільного заліза 40 % та 60 % у діелектричній матриці. Встановлено, що коефіцієнти проходження електромагнітного випромінювання у частотному діапазоні 2–12 ГГц знижуються з 0,80–0,85 до 0,10–0,30. Відповідно коефіцієнти відбиття електромагнітних хвиль підвищуються від 0,05–0,08 до 0,22–0,28. При цьому на частотах 8–10 ГГц спостерігаються вищі коефіцієнти відбиття для зразків з меншим вмістом карбонільного заліза. У подальшому цей показник збільшується і стабілізується. Зниження коефіцієнта відбиття електромагнітних хвиль при вмісті провідної субстанції у діелектричному матеріалі можна пояснити досягненням перколяційного ефекту – підвищенням провідності матеріалу. Це відбувається при утворенні у матеріалі кіл провідностей через значні значення електромагнітних констант частинок наповнювача. Тому при проектуванні композиційних матеріалів слід керуватися об'ємним вмістом наповнювача у діелектричній матриці. Отримані дані дозволяють розрахувати комплексні діелектричні та магнітні проникності матеріалів. Це у подальшому спростить проектування композиційних матеріалів з потрібними захисними властивостями. У процесі проектування захисних матеріалів і конструкцій слід враховувати різні залежності тангенса кута втрат для діелектричних і провідних матеріалів. Дана межа для композицій дещо умовна, тому попередньою умовою для розроблення захисних матеріалів є визначення амплітудно-частотних характеристик полів, які потребують екранування.

Ключові слова: композиційний матеріал, екранування електромагнітного випромінювання, коефіцієнт відбиття, електрофізичні властивості.

Вступ

Основними напрямками розроблення і впровадження композиційних матеріалів у галузях електромагнітної безпеки та електромагнітної екології є забезпечення високих коефіцієнтів екранування за рахунок поглинання енергії електромагнітних хвиль. Тобто, за високої ефективності коефіцієнти відбиття електромагнітних хвиль повинні бути мінімальними. Це є складною задачею, оскільки базується на фундаментальних положеннях електродинаміки суцільних середовищ. На сьогодні більшість таких досліджень стосується нанокомпозитів, але такі матеріали мають високу вартість, що обумовлює обмеженість сфери їх застосування. Багатошарові структури з мінімальною відбивальною здатністю не завжди прийнятні через складність технологій їх використання. Тому доцільно провести комплекс експериментальних досліджень з пошуку компромісного рішення щодо забезпечення достатнього коефіцієнта загального екранування з прийнятним значенням коефіцієнта відбиття.

Стан питання

Складність проблематики екранування електромагнітних полів обумовлена значними розбіжностями амплітудно-частотних характеристик електромагнітної обстановки на окремих територіях, в будівлях та визначених виробничих умовах. Допустимі рівні полів регламентуються міжнародним нормативом [1] та національними нормативами [2, 3]. Особливістю застосування екранування електромагнітних полів ультрависоких і вищих частот є одночасний захист людей від електромагнітних впливів та забезпечення стабільної роботи усіх засобів бездротового

зв'язку. Крім того, такі екрани повинні мати малі коефіцієнти відбиття електромагнітних хвиль для уникнення непередбачуваних просторових перерозподілів потоків випромінювань. Найвищі ефективності мають матеріали на основі наночастинок [4, 5]. Але їх вартість як при отриманні наноструктур, так і при виготовленні матеріалів дуже висока. Тому їх застосування є доцільним тільки для забезпечення електромагнітної сумісності електронного обладнання – захист від сторонніх електромагнітних впливів. Дослідження щодо розроблення й випробування захисних властивостей композицій на основі мікроструктур стосуються екранування електромагнітних полів частотами до 3 ГГц. У роботі [6] визначено зміни коефіцієнтів екранування електромагнітного поля частотою 1,8 ГГц у залежності від вмісту та дисперсності магнетиту. Показано, що при досягненні прийнятних загальних коефіцієнтів екранування коефіцієнт відбиття набуває високих значень. У дослідженні [7] для підвищення коефіцієнта екранування матеріалами малих товщин у вихідну суміш додається графітизована сажа. Вона має високу вартість, а технологія виготовлення композиції ускладнюється. У роботі [8] показано застосування двокомпонентних матеріалів на основі залізородного концентрату для створення градієнту електрофізичних властивостей по глибині матеріалу. При цьому зовнішній шар має відносно малий коефіцієнт відбиття електромагнітних хвиль, а внутрішні – високі поглинальні властивості. Нанесення на поверхню кількох шарів захисного матеріалу також збільшує час створення захисного покриття і знижує його стійкість до механічних та інших впливів. Дослідження [9] стосуються застосування карбонільного заліза у якості наповнювача матриці.

Показано, що на такому матеріалі можна отримати малі коефіцієнти відбиття і високі загальні коефіцієнти екранування. Товщини досліджуваних зразків великі – 5 мм, що не завжди зручно. Матеріал матриці модельний і потребує багатоступінчастого виготовлення, застосування кількох шарів лаку. Але висока відносна магнітна проникність заліза, прийнятна електропровідність та хімічна стійкість частинок, захищених шаром вуглецю робить цей матеріал дуже перспективним. Доцільним є дослідити застосування дрібнодисперсного заліза у якості наповнювача полімерної матриці з точки зору ефективності екранування електромагнітних полів ультрависоких і вищих частот.

Мета дослідження – визначення коефіцієнтів поглинання й відбиття електромагнітних полів ультрависоких і надвисоких частот та створення засад застосування таких матеріалів для захисту людей від електромагнітних впливів.

Викладення основного матеріалу

Для проведення досліджень було виготовлено зразки захисних матеріалів на основі латексу

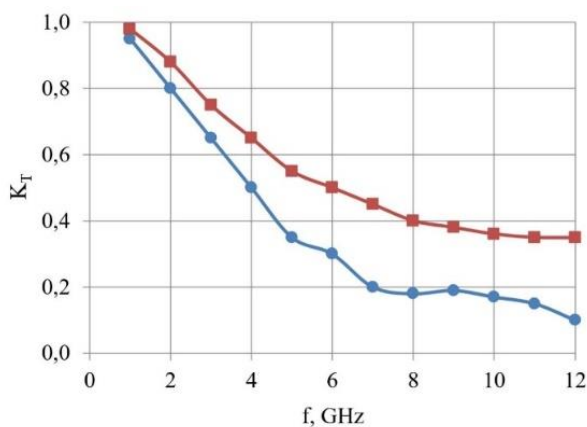


Рис. 1. Залежність коефіцієнтів проходження електромагнітного випромінювання крізь матеріал від частоти,
 — для кількості наповнювача 40 % за масою;
 — для кількості наповнювача 60 % за масою

Одержаний результат свідчить, що відбиття електромагнітних хвиль має більш складну частотну залежність. На частотах, близьких до 6 ГГц зразок з вмістом 60 % заліза має чіткий максимум відбиття, а на частотах 8–10 ГГц цей показник має мінімальні значення. При цьому для частот, більших за 7 ГГц коефіцієнти відбиття більші для матеріалів з меншим вмістом металевої субстанції 40 %. Цей результат не очікуваний й потребує з'ясування.

На рис. 3 наведені результати вимірювань коефіцієнта поглинання отриманих матеріалів.

Як видно з рис. 3, на частотах 8–10 ГГц спостерігається певний максимум поглинання, що відповідає частотам зниження коефіцієнта відбиття K_A . Також деякі відхилення від монотонності спостерігаються на кривій проходження електромагнітних хвиль. Це можна пояснити явищами резонансу при певних

(матриця) та карбонільного заліза (наповнювач). Кількість наповнювача складала 40 % та 60 % за масою. Дисперсність карбонільного заліза – 6–8 мкм. Товщина зразків складала приблизно 2 мм. Для випробувань захисних властивостей використовувався хвильовий метод.

Вимірювання здійснювалися у діапазоні частот 1–12 ГГц, що відповідає робочим частотам більшості високочастотних джерел електромагнітного випромінювання, які використовуються у виробничих та інших процесах. Здійснювалося вимірювання коефіцієнтів проходження електромагнітного випромінювання K_T , коефіцієнта відбиття K_R та коефіцієнта поглинання K_A . В усіх випадках ефективність екранування оцінювалася за зміною потужності випромінювання.

На рис. 1, 2 наведено частотну залежність коефіцієнта проходження електромагнітного випромінювання.

Як видно з рис. 1 для обох зразків коефіцієнт проходження практично монотонно знижується зі збільшенням частоти випромінювання. Деякі відхилення спостерігаються у смузі частот 8–10 ГГц.

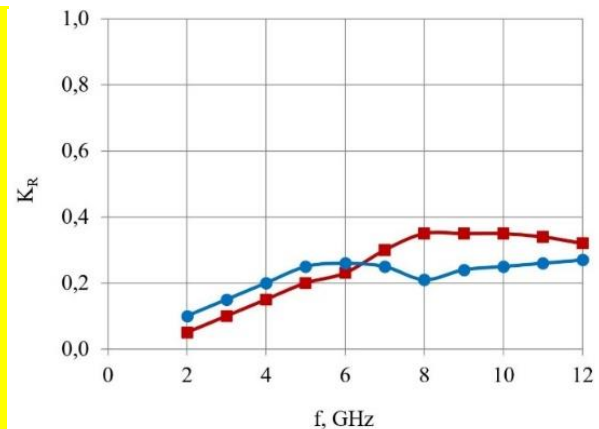


Рис. 2. Залежність коефіцієнтів відбиття електромагнітного випромінювання крізь матеріал від частоти,
 — для кількості наповнювача 40 % за масою;
 — для кількості наповнювача 60 % за масою

умовах вимірювань та властивостями феромагнітного резонансу. Але загальна тенденція підвищення ефективності екранування зі зростанням частоти випромінювання відповідає положенням електродинаміки суцільних середовищ. При розробленні захисних матеріалів слід враховувати, що коефіцієнти відбиття мають тенденцію до зростання, що може бути неприйнятним для вирішення деяких прикладних задач.

Якщо узагальнити отримані дані, то має виконуватися співвідношення:

$$K_T + K_R + K_A = 1,$$

Але у результатах є деякі відхилення. Це є результатом певної похибки вимірювань. Але загальні тенденції зміни трьох показників свідчать про коректність даних.

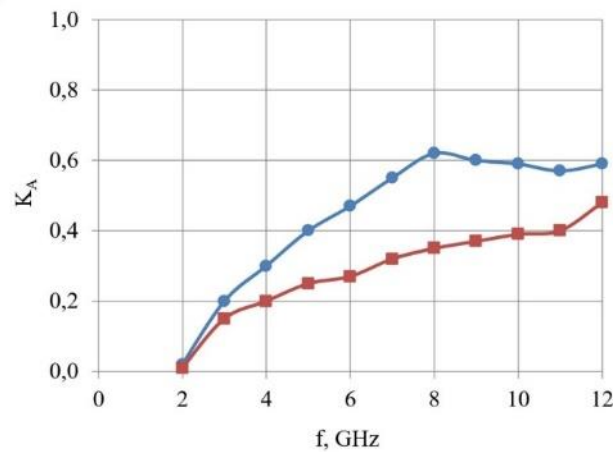


Рис. 3. Залежність коефіцієнтів поглинання електромагнітного випромінювання крізь матеріал від частоти,

— для кількості наповнювача 40 % за масою;
 — для кількості наповнювача 60 % за масою

На основі отриманих даних застосування методики Беккера-Джарвіса можна розрахувати значення комплексних діелектричної та магнітної проникності матеріалів.

Зокрема, дійсні складові діелектричної проникності матеріалу з вмістом карбонільного заліза 80 % має значення 9,0–8,5 в усьому діапазоні, а з вмістом 40 % – 6,2–6,0. При цьому уявна складова має невеликі значення, якими можна нехтувати. Магнітна проникність за дійсною складовою знижується з 2,5 та 2,0 до 1,5–1,6 на частотах 8–10 ГГц. При цьому уявні складові також незначні. Така стабілізація й обумовлює вихід коефіцієнта відбиття на практично сталі значення.

Коефіцієнт відбиття електромагнітних хвиль від поверхні матеріалу визначається співвідношенням:

$$K_R = \frac{Z_M - 1}{Z_M + 1},$$

де Z_M – поверхневий імпеданс матеріалу.

$$Z_M = \sqrt{\frac{\mu\mu_0}{\varepsilon\varepsilon_0(1 - itg\delta)}},$$

де $tg\delta$ – тангенс кута втрат.

Для провідних матеріалів:

$$tg\delta = \frac{\sigma}{\omega\varepsilon\varepsilon_0},$$

де σ – питома електрична провідність матеріалу.

Для діелектриків:

$$tg\delta = \frac{\varepsilon''}{\varepsilon'}.$$

Слід враховувати, що тангенс кута магнітних втрат складається з кількох компонентів – втрат на гістерезис, вихрові струми та залишкову намагніченість.

Межа між провідними та непровідними матеріалами дуже умовна. Це ілюструє зміни коефіцієнтів відбиття для різних вмістів карбонільного заліза

(рис. 2). На певній частоті коефіцієнт відбиття матеріалу з більшим вмістом заліза стає нижчим за коефіцієнт матеріалу з меншим вмістом заліза. Це може відбуватися на межі перколяційного ефекту – зростання провідності матеріалу. Тобто, частина ефектів, суттєвих з точки зору захисту від електромагнітних впливів, пов'язаних саме з цим показником. Слід враховувати, що визначення вмісту провідного наповнювача за масою не дає уявлення про межу провідності. Різке підвищення провідності забезпечується утворенням кіл провідності у діелектричній матриці за наявності контакту між частинками наповнювача. Тому у процесі проектування композиційного матеріалу слід керуватися об'ємним вмістом провідного наповнювача у матриці. Складність забезпечення поверхневих імпедансів захисних матеріалів близьких до хвильового опору повітря обумовлена проблематичністю врахування розподілу екрануючих домішок у тілі діелектричної матриці. У процесі збільшення магнітної проникності невідворотно змінюється діелектрична проникність. Застосування відомих емпіричних співвідношень (Оделевського, Максвелла-Гарнетта, Бруггемана) для оцінювання електрофізичних властивостей спроектованої композиції не дає бажаного ефекту. Тому доцільно провести серію експериментальних досліджень для композицій з різним вмістом магнітної та провідної речовини для отримання графічних залежностей змін поверхневих імпедансів за різних умов. Доцільно розглянути можливості отримання на основі експериментальних даних емпіричних співвідношень, які описують зміни електрофізичних та магнітних властивостей композицій за великого об'ємного вмісту (більше 30 %) екрануючої субстанцією у діелектричній матриці. Такі співвідношення повинні враховувати й морфологію екрануючих включень – відхилення від сферичної форми співвідношення довжини й ширини тощо.

Висновки

1. Досліджено зміни коефіцієнтів проходження, відбиття та поглинання електромагнітних

випромінювань у композиційних матеріалах з різним вмістом карбонільного заліза великої дисперсності.

Встановлено частотну залежність цих показників у діапазонах ультрависоких та надвисоких частот.

Показано, що ці залежності мають немонотонний характер у певних смугах частот, що обумовлено резонансними явищами.

2. Показано можливість визначення за результатами експерименту комплексних діелектричної та магнітної проникностей композиційних матеріалів. Це спрощує процеси проектування матеріалів

з потрібними захисними властивостями. Слід врахувати відмінності показників втрат для діелектричних та провідних матеріалів.

3. Показано, що найбільші зміни у відбивальних властивостях композиційних матеріалів відбуваються при концентраціях екрануючого наповнювача у діелектричній матриці, які відповідають межі різкого зростання провідності матеріалу.

Для коректного визначення межі провідності у процесі проектування матеріалів слід керуватися об'ємним вмістом екрануючого наповнювача у діелектричній матриці.

СПИСОК ЛІТЕРАТУРИ

1. Глива В. А., Халмурадов Б. Д., Занько С. М., Подобєд І. М. Дослідження електромагнітного навантаження на виробничі середовища аеропортів та головні напрями його зниження. Проблеми охорони праці в Україні. 2013. Вип. 27. С. 44–48.
2. Про затвердження Змін до Державних санітарних норм і правил захисту населення від впливу електромагнітних випромінювань [Чинний від 2017-05-23]: наказ М-ва охорони здоров'я України від 13.03.2017 № 266
3. ДСНІП 3.3.6.096-2002. Державні санітарні норми і правила при роботі з джерелами електромагнітних полів [Чинний від 2003-03-13]: затв. наказом М-ва охорони здоров'я України від 18.12.2002 р. № 476. Київ, 2003. 16 с. (Державні санітарні норми України).
4. Glyva V.A., Podoltsev A.D., Bolibrukh B.V., Radionov A.V. A Thin Electromagnetic Shield of a Composite Structure Made On the Basis of a Magnetic Fluid. *Tekhnichna elektrodynamika*. 2018. № 4. P.14–18. <https://doi.org/10.15407/tech-2018.04.014>.
5. Tudose I.V., Mouratis K., Ionescu O.N., Romanitan C., Pachiou C., Popescu M., Khomenko V., Butenko O., Chernysh O., Kenanakis G., Barsukov V.Z., Sucheana M.P., Koudoumas E. Novel Water-Based Paints for Composite Materials Used in Electromagnetic Shielding Applications. *Nanomaterials*. 2022, 12(3). P. 487. <https://doi.org/10.3390/nano12030487>
6. Glyva V., Lyashok J., Matvieieva I., Frolov V., Levchenko L., Tykhenko O., Panova O., Khodakovskyy O., Khalmuradov B., Nikolaiev K. Development and investigation of protective properties of the electromagnetic and soundproofing screen. *Eastern-European Journal of Enterprise Technologies*. 2018. Iss. 6/5 (96). P. 54–61. <https://doi.org/10.15587/1729-4061.2018.150778>
7. 198. Senyk I., Kuryptia Y., Barsukov V., Butenko O., Khomenko V. Development and application of thin wide-band screening composite materials. *Physics and Chemistry of Solid State*. 2020. 21(4). Pp. 771–778
8. Бурдейна Н.Б., Бірук Я.І., Ніколаєв К.Д. (2023). Розроблення матеріалів багатошарової структури градієнтного типу на основі рідких композицій для екранування електромагнітних полів. Екологічна безпека та природокористування. 45 (1). С. 68–75. <https://doi.org/10.32347/2411-4049.2023.1.68-75>.
9. Burdeina, N., Levchenko, L., Korduba, I., Shamanskiy, S., Biruk, Y., Dovhanovskiy, M., Zozulya, S., Klymchuk, A., Nikolaiev, K., & Osadchyi, D. (2024). Applying heterogeneous building materials for the protection of people against electromagnetic radiation. *Eastern-European Journal of Enterprise Technologies*, 5(10 (131)), 45–52. <https://doi.org/10.15587/1729-4061.2024.313629>

Received (Надійшла) 28.11.2024

Accepted for publication (Прийнята до друку) 26.02.2025

Application of composite iron-containing material for shielding electromagnetic fields of ultra-high and higher frequency

N. Burdeina, L. Zozulya, T. Petrunok

Abstract. Changes in the transmission, reflection and absorption coefficients of electromagnetic radiation by composite materials with a mass content of carbonyl iron of 40% and 60% in the dielectric matrix were investigated. It was found that the transmission coefficients of electromagnetic radiation in the frequency range of 2–12 GHz decrease from 0.80–0.85 to 0.10–0.30. Accordingly, the reflection coefficients of electromagnetic waves increase from 0.05–0.08 to 0.22–0.28. At the same time, higher reflection coefficients are observed at frequencies of 8–10 GHz for samples with a lower content of carbonyl iron. In the future, this indicator increases and stabilizes. The decrease in the reflection coefficient of electromagnetic waves with the content of a conductive substance in the dielectric material can be explained by the achievement of the percolation effect – an increase in the conductivity of the material. This occurs when conductivity circles are formed in the material due to significant values of the electromagnetic constants of the filler particles. Therefore, when designing composite materials, one should be guided by the volumetric content of the filler in the dielectric matrix. The obtained data allow one to calculate the complex dielectric and magnetic permeability of materials. This will further simplify the design of composite materials with the required protective properties. In the process of designing protective materials and structures, one should take into account the different dependences of the loss tangent angle for dielectric and conductive materials. This limit for compositions is somewhat conditional, therefore, a prerequisite for the development of protective materials is the determination of the amplitude-frequency characteristics of the fields that require shielding.

Keywords: composite material, electromagnetic radiation shielding, reflection coefficient, electrophysical properties.

В. А. Глива, М. О. Довгановський

Київський національний університет будівництва і архітектури, Київ, Україна

ПРОЄКТУВАННЯ ОДЯГУ ДЛЯ ЗАХИСТУ ВІД ТЕРМІЧНИХ ВПЛИВІВ ТА ЗБЕРЕЖЕННЯ ТЕПЛА

Анотація. На основі фундаментальних співвідношень теплопровідності розроблено модель теплопередачі крізь декілька шарів текстильного та захисного матеріалу. Враховано наявність повітряних прошарків між окремими шарами матеріалів. Отримано зміни температур окремих шарів матеріалу, що дозволяє визначити номенклатуру текстильних матеріалів та захисних шарів у залежності від поставлених задач термоізоляції. Розроблена методологія дозволяє проектувати одяг для захисту людей від зовнішніх термічних впливів та запобігати негативних змін у термічному режимі тіла людини. Отримані залежності щодо зміни температури зовнішньої поверхні захисного матеріалу з часом. Це дозволяє визначити проміжки часу, за які захисний одяг виконує свої функції на прийнятному рівні. Проведено верифікацію результатів досліджень. Реальний матеріал на основі скловолокна обстежувався за допомогою стандартного тепловізора у реальних умовах. Показано, що принаймні за градієнту температури у 20 К матеріал повністю ізолює людину від впливу зовнішнього середовища. Зазначено, що частинки скловолокна шкідливі для людей, тому у процесах виготовлення та експлуатації скловолокно повинне бути ізолювано за допомогою декоративного покриття.

Ключові слова: теплоізоляція, захисний одяг, моделювання, скловолокно.

Вступ

Не дивлячись на велику увагу, що приділяється розробленню спеціального одягу для працівників служб надзвичайних ситуацій, металургійної промисловості тощо, такі розробки стосуються або одного конкретного впливу, або мають загальний характер не враховуючи можливі значні коливання температур та фізичне навантаження людини. На сьогодні розроблені загальні засади проектування багатшарових теплоізолюючих конструкцій. Але ці розробки мають дещо абстрактний характер, оперуючи набором констант без прив'язки до конкретних матеріалів та градієнтів температур. У багатьох випадках необхідний одяг, який буде захищати людину від впливів високих температур та забезпечувати термоізоляцію в умовах низьких температур зовнішнього середовища. Для цього необхідно дослідити процеси проходження теплової енергії крізь шари одягу з різними теплофізичними властивостями, і визначити ефективність теплового захисту. Це дозволить раціонально обрати теплофізичні параметри звичайного одягу та матеріалу, який забезпечує термоізоляцію людини. Такий підхід може забезпечити захист від впливу високих та низьких температур, а також проектувати засоби інфрачервоного камуфляжу.

Огляд літературних джерел. Аналіз доступних досліджень з розроблення захисного одягу, зокрема для працівників служб з надзвичайних ситуацій, свідчить, що їх недоліком є застосування застарілих текстильних матеріалів [1]. Значною мірою це обумовлене обмеженнями на публікації таких матеріалів через їх застосування при розробленні інфрачервоного камуфляжу. Дослідження щодо створення такого камуфляжу [2] на основі наноструктур мають високу ефективність але складні у виготовленні і мають високу вартість. Дослідження [3] розглядає захист від теплового випромінювання з використанням рідкої суміші на основі серійного лаку з наповнювачами з нітриду бору та оксиду хрому.

Такі наповнювачі мають великі теплоємності і забезпечують поглинання теплової енергії на певний час, але вони наносяться на тканину, яка стає жорсткою і незручною для виготовлення одягу. Більш придатні для виготовлення захисного шару матеріали з напиленнями на поверхні сплавів германію, сурми, телуру. Але такі матеріали занадто дорогі [4]. Це ж стосується захисного матеріалу з монокульок полістиролу [5]. У роботі [6] показано можливість застосування скловолокна у якості засобу поглинання інфрачервоного випромінювання. Теоретичні міркування свідчать, що такий підхід перспективний. Скловолокна мають поглинання властивості у довжинах хвиль 9–14 мкм, що відповідає більшості частот інфрачервоного випромінювання. Загальний недолік більшості досліджень – розроблення і визначення захисних властивостей одного визначеного матеріалу. Це обумовлене тим, чинні міжнародні нормативи, наприклад [7], вимагають вимірювання теплового потоку крізь захисний шар без врахування інших шарів, які можуть впливати на цей показник. В той же час важливими є зміни цього показника внаслідок проходження тепла крізь шкіру людини, звичайний одяг та наявність певних повітряних проміжків між шарами. Це обумовлює актуальність дослідження.

Викладення основного матеріалу

У загальному випадку для розроблення засобів термоізоляції за умови захисту людини від зовнішніх термічних впливів та захист від втрат тепла використовуються однакові підходи. Ці задачі є прямою та зворотною, а кінцевий результат обумовлений виключно спрямованістю градієнту температури.

Для дослідження процесу теплопередачі (теплового потоку) від тіла людини до зовнішньої поверхні матеріалу спеодягу використовується наступне рівняння теплопровідності:

$$\frac{\partial T}{\partial t} = \left(\lambda_x \frac{\partial^2 T}{\partial x^2} \right), \quad (1)$$

де T – температура, λ – коефіцієнт теплопровідності, t – час.

В моделі приймається, що область дослідження складається з різних шарів, а саме: тіло людини, білизна, спецодяг (комплект одягу), куртка, що відрізняються теплофізичними властивостями.

Розрахункова схема показана на рис. 1.

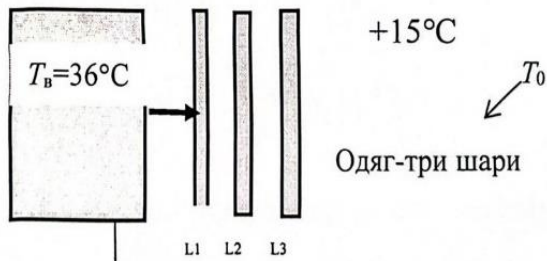


Рис. 1. Розрахункова схема шарів моделі

Таким чином, фізична модель складається з трьох шарів одягу та трьох шарів повітря:

джерело теплогенерації – тіло людини має такі фізичні параметри:

товщина шару шкіри тіла $l = 3,2$ мм;

теплоємність $C_T = 3350$ Дж/(кг*К);

щільність $\rho = 1056$ кг/м³;

коефіцієнт теплопровідності $\lambda_T = 0,48$ Вт/(м*К).

Між тілом людини та трьома шарами одягу враховуємо параметри повітря:

коефіцієнт теплопередачі $\lambda_p = 0,02$ Вт/м*К;

теплоємність $C_p = 1000$ Дж/(кг*К);

щільність $\rho_p = 1,226$ кг/м³.

$L_1 = 1$ мм – відстань між тілом людини та першим шаром одягу (натільна білизна);

$d_1 = 1,0$ мм – товщина матеріалу білизни;

$\lambda_1 = 0,03$ Вт/м*К – коефіцієнт теплопередачі;

$C_1 = 1$ кДж/кг*К – теплоємність тканини білизни;

$\rho_1 = 100$ г/м² – щільність;

$L_2 = 2$ мм – відстань першим та другим шаром одягу (камуфляж);

$d_2 = 1,5$ мм – товщина матеріалу спецодягу;

$\lambda_2 = 0,04$ Вт/м*К – коефіцієнт теплопередачі;

$C_2 = 1,9$ кДж/кг*К – теплоємність тканини камуфляжу;

$\rho_2 = 220$ г/м² – щільність;

$L_3 = 5$ мм – відстань між другим шаром (камуфляж) та підкладкою куртки;

$d_3 = 3,5$ мм – товщина матеріалу спецодягу;

$\lambda_3 = 0,05$ Вт/м*К – коефіцієнт теплопередачі;

$C_2 = 2,1$ кДж/кг*К – теплоємність тканини верху куртки;

$\rho_3 = 350$ г/м² – щільність;

Температура зовнішнього середовища $+15$ °С.

Тобто, постає задача визначення:

часу досягнення температури на поверхні матеріалу верху куртки $\geq +15,1$ °С при навантаженні 120 Вт;

параметрів температурного режиму в підкостюмному просторі;

часу досягнення температури на поверхні матеріалу верху куртки $\geq +15,1$ °С при навантаженні 400 Вт.

Коефіцієнт температуропровідності для кожного шару визначається як:

$$a_n = \lambda_n / (\rho_n C_n), \quad (2)$$

Для моделюючого рівняння теплопровідності (1) ставляться такі граничні умови (рис. 1):

– на лівій межі (зі сторони навколишнього середовища) задається значення температури $T = 36$ °С.

– на правій межі (на поверхні матеріалу верху верхнього одягу) задається значення температури $T_0 = 15$ °С. При проведенні розрахунків гранична умова з часом має вигляд:

$$T_{n-1} = T_n, \quad (3)$$

де T_n – температура в останній розрахунковій комірниці, T_{n-1} – температура в попередній розрахунковій комірниці. Початкова умова при $t = 0$ в усій розрахунковій області приймається $T = 36$ °С.

Для проведення розрахунків використовується прямокутна різницева сітка. Значення температури визначається в центрах різницевих комірок. Для чисельного розв'язання моделюючого рівняння (1) використовується явна різницева схема:

$$T_i^{n+1} = T_i^n + \Delta t \cdot a_1 \frac{T_{i+1}^n - T_i^n}{\Delta x^2} + \Delta t \cdot a_2 \frac{-T_i^n - T_{i-1}^n}{\Delta x^2}, \quad (4)$$

$$a_1 = \frac{2(a_{i+1} \cdot a_i)}{a_{i+1} + a_i}, \quad (5)$$

$$a_2 = \frac{2(a_{i-1} \cdot a_i)}{a_{i-1} + a_i}. \quad (6)$$

Залежності (4) та (5) використовуються для розрахунку значення коефіцієнта температуропровідності на границі шарів з різними теплофізичними параметрами. Усього 50 розрахункових вузлів по осі Ox , крок сітки 7.7/50 мм. Було розраховано температури кожного шару захисного одягу та температуру зовнішньої поверхні одягу за різні проміжки часу (рис. 2).

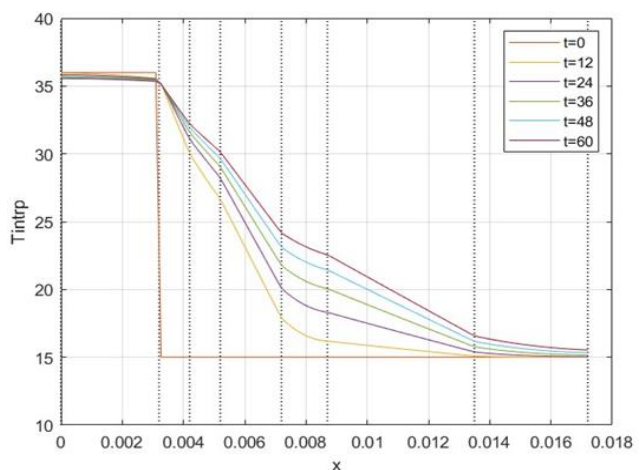


Рис. 2. Температури окремих шарів та зовнішньої поверхні захисного одягу у залежності від часу передачі тепла

Розрахунки виконані за умови наявності прошапків повітря завтовшки 1–2 мм між усіма шарами

одягу. Встановлено, що принаймні за обраних умов зміна фізичного навантаження людини суттєво не впливає на параметри теплоізоляції. Було проведено верифікацію результатів розрахунків з використанням матеріалу, який у межах похибки товщин окремих шарів відповідає параметрам модельного захисного матеріалу. Вимірювання здійснювалися з використанням тепловізора OTS-XLT 160 2-8X. Результати вимірювань наведено на рис. 3.

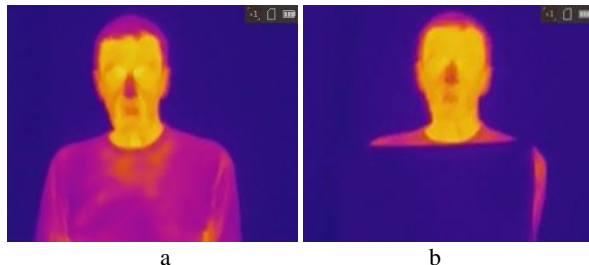


Рис. 3. Перевірка ефективності захисного одягу у реальних умовах експлуатації:
а – ображення об'єкта без захисту;
б – об'єкт захищений теплоізоляційним матеріалом;
синій колір відповідає температурі 16°C

Отриманий результат свідчить, що скловолокно є достатньо ефективним матеріалом для виготов-

лення захисного одягу. Слід враховувати, що у процесі його виготовлення у повітря потрапляють частинки скловолокна. Тому навіть у процесі експлуатації скловолокно необхідно покривати захисним шаром.

Висновки

1. Виконані розрахунки щодо проектування захисного термоізоляційного одягу. Показано, що наведений підхід придатний для визначення ефективності захисту від зовнішніх термічних впливів та підтримання теплового режиму людини. Отримані зміни температури у окремих шарах одягу з урахуванням наявності повітряних зазорів між окремими шарами. Це дозволяє обрати найбільш придатні матеріали для забезпечення необхідних параметрів теплоізоляції у залежності від умови експлуатації.

2. Проведено натурні випробування матеріалу на основі скловолокна. Встановлено, що принаймні за градієнту температури у 20 К розроблений матеріал надійно теплоізолює людину. Враховуючи малу товщину матеріалу, він може бути використаний для забезпечення інфрачервоного камуфляжу. Зазначено, що у процесі вироблення і експлуатації матеріалу на основі скловолокна його частинки можуть негативно впливати на людей, скловолокно повинне бути ізолюваним від людей захисним декоративним шаром.

СПИСОК ЛІТЕРАТУРИ

1. Токарський О. Методичні засади визначення оптимального часу роботи рятувальника. ВІСТІ Донецького гірничого інституту. Донецький національний технічний університет, м. Покровськ., 2022. Т. 1, № 50. С. 14–20.
2. Diao, Z., Kraus, M., Brunner, R., Dirks, J. H., & Spatz, J. P. (2016). Nanostructured Stealth Surfaces for Visible and Near-Infrared Light. *Nano letters*, 16(10), 6610–6616. <https://doi.org/10.1021/acs.nanolett.6b03308>
3. Glyva, V., Krasnianskyi, G., Dovhanovskyi, M., & Krasnianskyi, T. (2024). Розроблення і дослідження захисних властивостей матеріалів для блокування електромагнітних випромінювань інфрачервоного діапазону. *Системи управління, навігації та зв'язку. Збірник наукових праць*, 3(77), 203-205. <https://doi.org/https://doi.org/10.26906/SUNZ.2024.3.20>
4. Qu, Y., Li, Q., Cai, L. et al. Thermal camouflage based on the phase-changing material GST. *Light Sci Appl* 7, 26 (2018). <https://doi.org/10.1038/s41377-018-0038-5>
5. Ho Kun Woo, Kai Zhou, Su-Kyung Kim, Adrian Manjarrez, Muhammad Jahidul Hoque, Tae-Yeon Seong, Lili Cai (2022). Visibly Transparent and Infrared Reflective Coatings for Personal Thermal Management and Thermal Camouflage. *Advanced Functional Materials*. V. 32, (38). <https://doi.org/10.1002/adfm.202201432>
6. Burdeina, N., Levchenko, L., Korduba, I., Shamanskyi, S., Biruk, Y., Dovhanovskyi, M., Zozulya, S., Klymchuk, A., Nikolaiev, K., & Osadchyi, D. (2024). Applying heterogeneous building materials for the protection of people against electromagnetic radiation. *Eastern-European Journal of Enterprise Technologies*, 5(10 (131), 45–52. <https://doi.org/10.15587/1729-4061.2024.313629>
7. ISO 15831:2007 Clothing — Physiological effects — Measurement of thermal insulation by means of a thermal manikin. (EN ISO 15831:2004), ID.

Received (Надійшла) 28.11.2024

Accepted for publication (Прийнята до друку) 26.02.2025

Designing clothing for protection against thermal effects and heat retention

V. Glyva, M. Dovhanovskyi

Abstract. Abstract. A model of heat transfer through several layers of textile and protective material is developed based on the fundamental relations of thermal conductivity. The presence of air gaps between individual layers of materials is taken into account. The temperature changes of individual material layers were obtained, which allows us to determine the range of textile materials and protective layers depending on the thermal insulation tasks. The developed methodology makes it possible to design clothing to protect people from external thermal influences and prevent negative changes in the thermal regime of the human body. Dependencies on the temperature change of the outer surface of the protective material over time were obtained. This makes it possible to determine the time intervals during which protective clothing performs its functions at the required level. The research results were verified. A real fibreglass-based material was examined using a standard thermal imager in real-world conditions. It is shown that at least at a temperature gradient of 20 K, the material completely isolates a person from the influence of the external environment. It is noted that glass fibre particles are harmful to humans, so in the processes of manufacturing and operation, glass fibre should be isolated by means of a decorative coating.

Keywords: thermal insulation, protective clothing, modelling, glass fibre.

Д. В. Резнік¹, О. О. Ченчева¹, Є. Є. Лашко¹, І. С. Петренко¹, Д. С. Гаврилець²

¹ Кременчуцький національний університет імені Михайла Остроградського, Кременчук, Україна

² Вище професійне училище № 7, Кременчук, Україна

ДОСЛІДЖЕННЯ СТАНУ ПОВІТРЯНОГО СЕРЕДОВИЩА В ПРИМІЩЕННЯХ ЗАКЛАДІВ ПЕРЕДВИЩОЇ ТА ПРОФЕСІЙНОЇ (ПРОФЕСІЙНО-ТЕХНІЧНОЇ) ОСВІТИ

Анотація. Мета. Дослідження динаміки респірабельної фракції пилу ($PM_{2.5}$ та PM_{10}) та аналіз її впливу на здоров'я учнів в навчальних аудиторіях і виробничих майстернях закладів передвищої та професійної (професійно-технічної) освіти м. Кременчук. **Методика.** Дослідження якості повітря у навчальних приміщеннях здійснювалося на основі емпіричного методу під час натурного експерименту, статистичного аналізу отриманих даних та апроксимації результатів. **Результати.** Дослідження показало, що концентрація респірабельної фракції пилу ($PM_{2.5}$, PM_{10}) у навчальних майстернях значно варіюється залежно від типу приміщення та виконуваних робіт. Концентрація пилу змінюється протягом дня: вона зростає на початку занять, знижується під час перерв і суттєво підвищується під час активної роботи. На робочих місцях, які досліджувалися, зафіксовано перевищення пилового навантаження, а найбільший рівень запиленості зафіксовано на робочому місці зварювальника. Оцінка вентиляційних систем показала, що швидкість руху повітря під час занять збільшилася у 5–15 разів, що свідчить про ефективність їх роботи, особливо на робочому місці зварювальника. Дослідження показали, що для покращення рівня пилового забруднення в робочій зоні необхідно вдосконалити та автоматизувати системи вентиляції й аспірації, а також впровадити використання засобів індивідуального захисту. **Наукова новизна.** Проведені дослідження вперше враховують реальні умови навчально-виробничого процесу у закладах фахової передвищої та професійної (професійно-технічної) освіти. Запропоновано комплексний підхід до аналізу якості повітря, що включає моніторинг забруднення ($PM_{2.5}$, PM_{10} , виявлення закономірностей зміни концентрації пилу залежно від етапів навчання, математичний опис динаміки забруднення та оцінку ефективності вентиляційних систем. Отримані результати можуть стати основою для вдосконалення заходів контролю якості повітря в навчальних майстернях та розробки нормативних актів, що забезпечать безпеку освітнього процесу. **Практична цінність.** Результати дослідження дозволяють: оптимізувати вентиляційні системи шляхом визначення необхідної швидкості повітряного потоку на основі реальних даних про концентрацію забруднюючих речовин у повітрі навчально-виробничих майстерень; розробити адаптивні системи фільтрації повітря для зниження рівня респірабельної фракції пилу в виробничих приміщеннях; обґрунтувати необхідність систематичного моніторингу повітряного середовища в навчальних майстернях, зокрема з використанням сучасних технологій (бездротові сенсорні модулі); запровадити заходи захисту для учнів вікової категорії 14–19 років, які є найбільш чутливими до впливу виробничого пилу. Застосування отриманих результатів сприятиме підвищенню рівня безпеки навчального процесу, зниженню професійних ризиків для учнів та покращенню умов праці в навчальних майстернях.

Ключові слова: респірабельна фракція пилу, навчальні майстерні, концентрація пилу, динаміка запилення, вентиляційна система.

Вступ

Здоров'я людини значною мірою залежить від якості повітря, яким вона дихає. За даними Всесвітньої організації охорони здоров'я (ВООЗ), понад 90% світового населення дихає забрудненим повітрям, в якому перевищені допустимі норми [1].

Однією з найбільш вразливих груп населення є молодь у віці 14–19 років, яка перебуває у процесі активного фізіологічного розвитку та водночас стикається з високими інтелектуальними навантаженнями під час навчання. В умовах забрудненого повітря у навчально-виробничих приміщеннях значно знижується когнітивна активність, підвищується ризик розвитку респіраторних захворювань, таких як бронхіальна астма, хронічний бронхіт та алергічні реакції. Наявність у повітрі таких шкідливих речовин, як респірабельних фракцій пилу ($PM_{2.5}$, PM_{10}), діоксид азоту (NO_2), діоксид сірки (SO_2), чадний газ (CO) та сірководень (H_2S), створює загрозу не лише для загального самопочуття, а й для довгострокового здоров'я молодих людей [1].

Зважаючи на це, важливим є дослідження якості повітря у навчальних закладах, особливо у майстернях та аудиторіях, де використовується технічне обладнання та матеріали, що можуть

сприяти забрудненню повітряного середовища. У місті Кременчук, яке має розвинену систему закладів передвищої та професійної (професійно-технічної) освіти та готує спеціалістів з технічних робітничих спеціальностей, необхідно приділяти особливу увагу моніторингу та покращенню умов навчання, рис. 1. Впровадження ефективних систем вентиляції та фільтрації повітря дозволить знизити рівень забруднення та забезпечити комфортне перебування учнів і викладачів у навчальних приміщеннях.



Рис. 1. Навчально-виробничі майстерні м. Кременчука

Метою дослідження є визначення концентрацій виробничого пилу респірабельної фракції ($PM_{2.5}$, PM_{10}) та шкідливих газів (NO_2 , SO_2 , CO , H_2S) в навчальних аудиторіях і виробничих майстернях закладів передвищої та професійної (професійно-технічної) освіти м. Кременчука.

Постановка завдання та його вирішення

В Україні відсутній єдиний нормативний документ, який би чітко регламентував вимоги до якості повітря у навчально-виробничих приміщеннях закладів передвищої та професійної (професійно-технічної) освіти. Це створює правову прогалину, що може негативно впливати на здоров'я підлітків у віці 15-19 років, які навчаються за технічними спеціальностями та проходять практичне навчання у токарних, фрезерних та зварювальних майстернях.

Натомість для закладів загальної середньої освіти діє Санітарний регламент [2], який встановлює вимоги до мікроклімату, вентиляції та провітрювання класних кімнат. Подібні вимоги прописані й для дошкільних навчальних закладів у Санітарному регламенті для дошкільних навчальних закладів [3]. Проте, жоден з цих документів не містить норм щодо концентрацій пилу чи хімічного складу повітря в приміщеннях, що використовуються для навчально-виробничої діяльності. Крім того, Наказ Міністерства освіти і науки України № 510 [4], регламентує організацію освітнього процесу в закладах фахової передвищої освіти, включаючи питання практичної

підготовки. Однак цей документ також не містить конкретних вимог до якості повітря в навчально-виробничих приміщеннях.

Нормативна база України у сфері виробничої діяльності, регламентує гранично допустимі концентрації (ГДК) шкідливих речовин у повітрі робочої зони. Так [5] встановлює вимоги до безпеки роботи з металами, включаючи параметри вентиляції та концентрації забруднюючих речовин. Проте цей документ призначений для промислових підприємств і не враховує особливостей освітнього процесу.

Європейське законодавство містить нормативи, що регулюють якість атмосферного повітря [6]. В цій директиві встановлено граничні значення для $PM_{2.5}$, PM_{10} , NO_2 та SO_2 у зовнішньому середовищі. Однак ці вимоги не стосуються повітря всередині приміщень, включаючи навчальні майстерні. Тобто, навіть у країнах ЄС відсутні окремі нормативи, які б регламентували стан повітря у навчально-виробничих приміщеннях передвищої та професійної (професійно-технічної) освіти.

ВООЗ надає рекомендації щодо якості повітря у приміщеннях, встановлюючи допустимі рівні концентрації пилу та інших забруднювачів [1], табл. 1.

Таблиця 1 – Нормативні показники забруднюючих речовин у повітрі та їх вплив на людину

Показник	Допустимі рівні концентрації у повітрі	Наслідки перевищення концентрацій забруднюючих речовин
$PM_{2.5}$	Річне середнє значення не повинно перевищувати 5 $мкг/м^3$, середньодобове значення – 15 $мкг/м^3$	Підвищений ризик серцево-судинних та респіраторних захворювань
PM_{10}	Річне середнє значення не повинно перевищувати 15 $мкг/м^3$, середньодобове – 45 $мкг/м^3$	Виникнення респіраторних хвороб, посилення астматичних симптомів
NO_2	Річне середнє значення обмежене 10 $мкг/м^3$, а максимальна погодинна концентрація – 200 $мкг/м^3$ (не більше 18 перевищень на рік)	Подразнення дихальних шляхів, сприяння утворенню озону та вторинних аерозолів, негативний вплив на легеневу функцію
SO_2	Максимальна погодинна концентрація – 40 $мкг/м^3$, середньодобове значення – 20 $мкг/м^3$	Погіршення стану здоров'я людей, особливо астматиків та осіб із захворюваннями дихальної системи
CO	Максимальне 8-годинне середнє значення – 4 $мг/м^3$	Отруєння організму, зменшення здатності крові транспортувати кисень, що може призвести до серйозних неврологічних та серцево-судинних наслідків
H_2S	Граничне середньодобове значення ≤ 7 $мкг/м^3$	Токсична дія на нервову систему

Отже, відсутність нормативного документа, що регулює якість повітря у навчально-виробничих приміщеннях закладів передвищої та професійної (професійно-технічної) освіти, є суттєвою проблемою, яка може негативно впливати на здоров'я учнів. Запровадження чітких норм сприятиме створенню безпечних умов праці та навчання, зменшенню впливу шкідливих факторів і підвищенню загального комфорту освітнього процесу.

Особливу увагу слід приділити віковій категорії 14–19 років, оскільки в цей період відбувається активний ріст і формування організму, а вплив забрудненого повітря, зокрема виробничого пилу, може мати довгострокові наслідки для здоров'я.

Важливим є системний моніторинг якості повітря та розробка спеціальних заходів безпеки, спрямованих на захист учнів.

Оптимальним рішенням є розроблення уніфікованого нормативного документа, який поєднав би

сучасні санітарні стандарти ВООЗ, норми промислової безпеки та найкращі практики ЄС щодо якості повітря у приміщеннях, забезпечуючи належний рівень охорони здоров'я та безпеки для всіх учасників навчального процесу.

Результати дослідження

Матеріально-технічна база закладів фахової передвищої та професійної (професійно-технічної) освіти Кременчука забезпечує учням доступ до технічного обладнання та засобів, які відповідають вимогам сучасного виробництва.

В навчальних аудиторіях та майстернях використовуються верстати, комп'ютери, 3D-принтери та електротехнічне обладнання.

Це, все дозволяє учням отримувати практичні навички для роботи на підприємствах регіону та України, однак одночасно важливим є забезпечення безпечних умов навчання.

Важливо під час проведення занять в аудиторіях та навчально-виробничих майстернях застосовувати адаптивні системи вентиляції та фільтрації повітря. Це пов'язано з тим, що обробляються різноманітні матеріали, застосовуються мастила й охолоджуючі рідини можуть призвести до зміни якісних показників повітря. У випадку, коли вентиляційна система не здатна вчасно видалити виробничий пил повітря, накопичення респірабельних фракцій пилу може стати потенційною небезпекою для здоров'я учнів.

Дослідження повітря у закладах фахової передвищої та професійної (професійно-технічної) освіти здійснювалося на базі приміщень ВПУ 7, що містить усі типи виробничих майстерень та обладнання. З метою отримання достовірних даних, у кожному приміщенні проводилося серія вимірів параметрів повітряного простору, результати яких оброблялися з використання сучасних програмних продуктів.

Дослідження якості повітря в навчальних приміщеннях закладів освіти здійснювався за допомогою модулю Vaisala AQT530, який дозволяє аналізувати рівні забруднення повітря, таких як $PM_{2.5}$, PM_{10} , NO_2 , CO , H_2S , SO_2 .

Модуль AQT530 поєднує електрохімічні сенсори для аналізу газових забруднювачів (NO_2 , CO , H_2S , SO_2) та оптичний метод розсіювання світла для визначення концентрації твердих частинок ($PM_{2.5}$, PM_{10}). Електрохімічні сенсори реагують на гази шляхом окисно-відновних процесів, створюючи

електричний сигнал, який точно відображає концентрацію забруднюючих речовин. Для аналізу твердих частинок використовується лазерна спектроскопія: частинки в потоці повітря розсіюють світло, а фотодетектори аналізують його інтенсивність, що дозволяє класифікувати частинки за розміром. Основними перевагами модулю AQT530 є – автоматична температурна компенсація, корекція перехресної чутливості, цифровий вихід для інтеграції та низьке енергоспоживання. Модуль підтримує передачу даних через бездротові технології (Wi-Fi, Ethernet, GSM), а також дозволяє відображати дані в реальному часі на інтерактивному інтерфейсі та зберігати їх для подальшого аналізу.

План досліджень був складений таким чином, щоб зафіксувати початковий рівень показників якості повітряного середовища до занять та протягом навчального дня, навчального тижня.

Використовуючи можливості модулю Vaisala AQT530 здійснено аналіз хімічного та пилового складу повітря в процесі занять.

За результатами дослідження, протягом робочого часу, у майстернях не було виявлено перевищення концентрацій NO_2 , SO_2 , CO та H_2S , табл. 2. Показники повітря за хімічним складом відповідають вимогам національних і міжнародних стандартів для такого типу приміщень [1, 6]. Це стосується як фонових рівнів забруднення перед початком занять, так і складу повітря під час самих занять.

Таблиця 2 – Середні показники хімічного забруднення повітря на робочих місцях у навчально-виробничих майстернях протягом робочого часу

Робоче місце:	фрезерувальника				токаря				зварювальника				електромонтера			
	NO_2	SO_2	CO	H_2S	NO_2	SO_2	CO	H_2S	NO_2	SO_2	CO	H_2S	NO_2	SO_2	CO	H_2S
Показник																
Середнє знач. за робочий час, мкг/м ³	2,7	9,3	650	1,7	2,6	8,8	783	1,8	4,2	10,5	1039	3,5	1,8	6,7	578	1,1

Основними джерелами хімічного забруднення повітря (NO_2 , SO_2 , CO та H_2S), є різноманітні антропогенні та природні фактори.

Викиди цих забруднювачів зумовлені діяльністю транспорту, промислових підприємств, спалюванням викопних видів палива, а також природними процесами, такими як розкладання органічних матеріалів.

Такі процеси в навчально-виробничих майстернях м. Кременчука не здійснюються.

Дослідження показників пилу здійснювалися у тих же самих навчально-виробничих майстернях у яких проводився аналіз хімічного забруднення повітряного середовища. За результатами побудовані залежності зміни концентрації пилу на робочих місцях у відповідних майстернях протягом навчального тижня, рис. 2.

Аналіз отриманих закономірностей динаміки зміни концентрації респірабельних фракцій пилу ($PM_{2.5}$ та PM_{10}) корегувався в залежності від графіку навчального процесу, а саме навчання та відпочинку.

Таким чином, на рис. 2, а-г, спостерігається коливання концентрацій, а саме:

– з 8.50 до 9.45 – початкове зростання концентрації частинок, коли учні приходять на заняття, переодягаються та розташовуються за партами для проведення теоретичної підготовки до практичних занять;

– з 9.45 до 10.00 – триває перерва, тому концентрація пилу спадає;

– з 10.00 до 13.30 – спостерігається активна робота у майстернях, коли застосовуються технологічне обладнання.

Кожне робоче місце учня характеризується відповідними значеннями максимуми та мінімуми наявності пилу, табл. 3.

Так у певні періоди робочого дня спостерігаються просадки концентрації частинок пилу, що свідчить про наявність перерв у роботі учнів та активну роботу вентиляційних установок.

Крім того, у табл. 3 наведено отримане середньодобове значення концентрації респірабельних фракцій пилу. Аналіз результатів дослідження концентрацій пилу на робочих місцях виявив, що найбільший рівень запиленості спостерігається на робочому місці зварювальника.

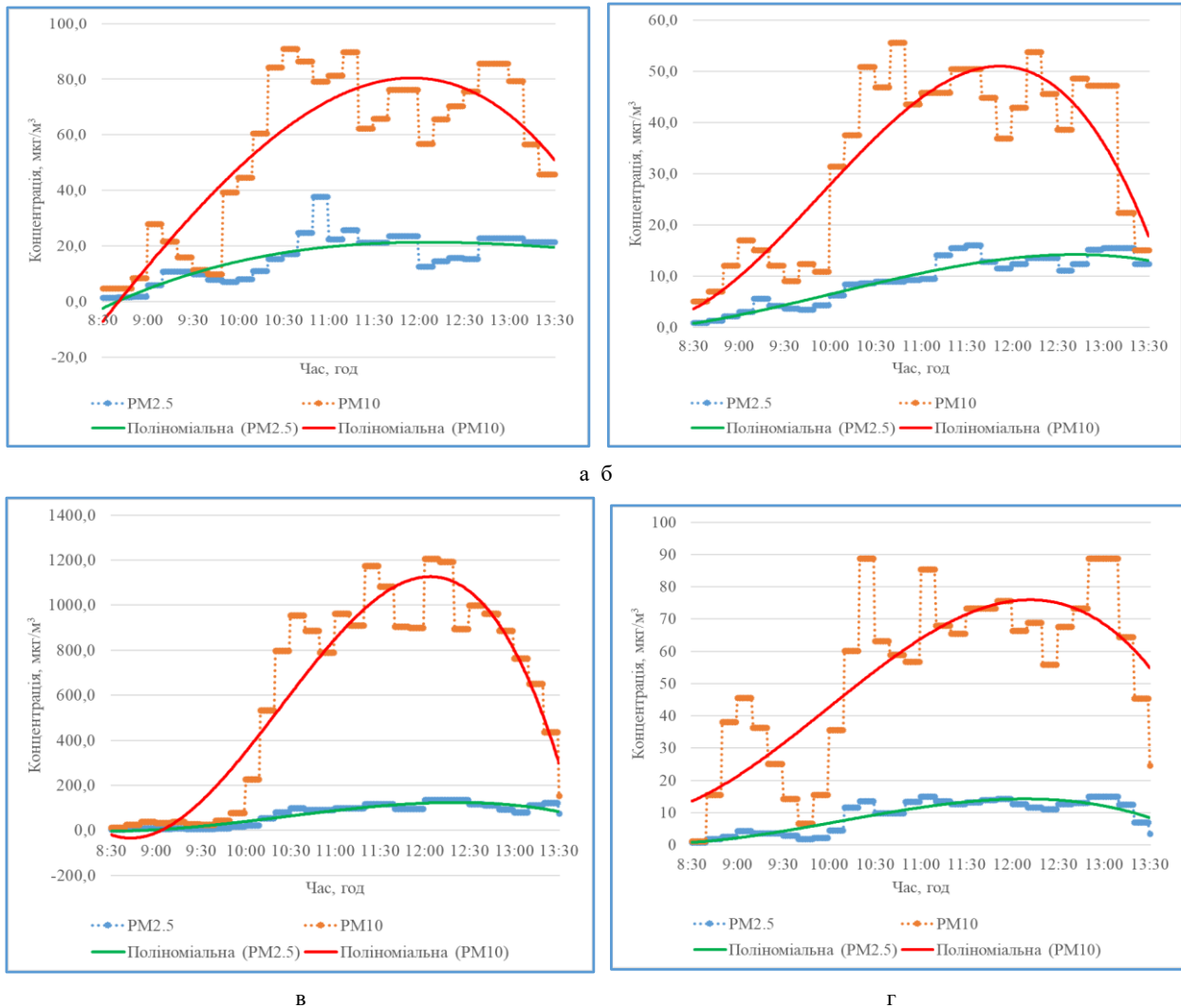


Рис. 2. Концентрації респірабельних фракцій пилу та їх апроксимація в залежності від типу робочого місця: а – робоче місце фрезерувальника; б – робоче місце токаря; в – робоче місце зварювальника; г – робоче місце електрика

Таблиця 3 – Зведені значення дослідження концентрації респірабельних фракцій пилу (PM_{2.5} та PM₁₀)

Робоче місце	PM _{2.5}			PM ₁₀		
	min, мкг/м ³	max, мкг/м ³	середньодобове значення, мкг/м ³	min, мкг/м ³	max, мкг/м ³	середньодобове значення, мкг/м ³
Фрезерувальника	1,3	37,7	4,4	4,7	90,9	15,3
Токаря	0,9	16,0	3,8	5,0	55,6	12,9
Зварювальника	2,6	134,5	16,9	12,0	1205,3	137,4
Електрика	0,8	14,9	2,7	1,0	88,8	12,1

Зокрема, середньодобова концентрація пилу за PM_{2.5} перевищує встановлені нормативи в 1,1рази, досягаючи значення 16,9 мкг/м³, а за PM₁₀ перевищення у три рази і становить 137,4 мкг/м³.

Інші робочі місця характеризуються меншою запиленістю, однак на них відзначаються випадки, коли миттєві значення концентрацій пилу перевищують середньодобові норми. Використовуючи математичні пакети виконано апроксимацію та отримано матема-

тичний опис зміни концентрації респірабельних фракцій пилу протягом заняття на робочих місцях.

У загальному вигляді зміну концентрації PM_{2.5} та PM₁₀ можна описати полігармонійною залежністю третього порядку:

$$PM_{2.5(10)} = a + b \cdot t + c \cdot t^2 + d \cdot t^3, \quad (1)$$

Коефіцієнти виразу, а також коефіцієнти детермінації й кореляції занесені до табл. 4.

Таблиця 4 – Зведена таблиця коефіцієнтів рівняння зміни концентрації респірабельних фракцій пилу (PM_{2.5} та PM₁₀)

Робоче місце	Розмір респірабельних фракцій пилу	Коефіцієнти рівняння				Коеф. детермінації R ² , %	Коеф. кореляції R, %
		a	b	c	d		
Фрезерувальника	PM _{2.5}	-2,662	0,259	-9*10 ⁻⁴	8*10 ⁻⁷	62,5	79,1
	PM ₁₀	-7,862	0,684	-4*10 ⁻⁴	-4*10 ⁻⁶	76,1	87,2
Токаря	PM _{2.5}	0,707	0,047	3*10 ⁻⁴	-9*10 ⁻⁷	88,6	94,1
	PM ₁₀	3,420	0,156	1,9*10 ⁻³	-8*10 ⁻⁶	78,3	88,5
Зварювальника	PM _{2.5}	-3,12	-0,019	7,2*10 ⁻³	-2*10 ⁻⁵	88,8	94,2
	PM ₁₀	-17,535	-2,529	0,098	-3*10 ⁻⁴	90,6	95,2
Електромонтера	PM _{2.5}	0,725	0,031	0,001	-2*10 ⁻⁶	79,9	89,4
	PM ₁₀	13,324	0,206	0,002	-7*10 ⁻⁶	63,8	79,8

Так, коефіцієнти детермінації й кореляції мають відносно велику розбіжність, що можна пояснити високою варіантністю даних та різними режимами роботи в майстернях.

З метою оцінки роботи системи вентиляції було досліджено швидкість руху повітря на найбільш запылених робочих місцях. Дослідження виконувалися у два етапи, а саме до початку занять і під час навчання.

Вимірювання швидкості повітря здійснювалося за допомогою анемометра GM8903, який здатний фіксувати швидкість повітря в діапазоні 0,001 – 45 м/с та має похибку ±3%. Отримані результати занесені до табл. 5.

Таблиця 5 - Дослідження швидкості руху повітря у приміщеннях виробничих майстерень

Робоче місце	Швидкість повітря, м/с	
	перед заняттям	під час заняття
Фрезерувальник	0,027	0,15
Токаря	0,036	0,18
Зварювальника	0,03	0,45
Електромонтера	0,019	0,22

На рис. 3 наведено результати вимірів протягом 100 с на робочому місці зварювальника.

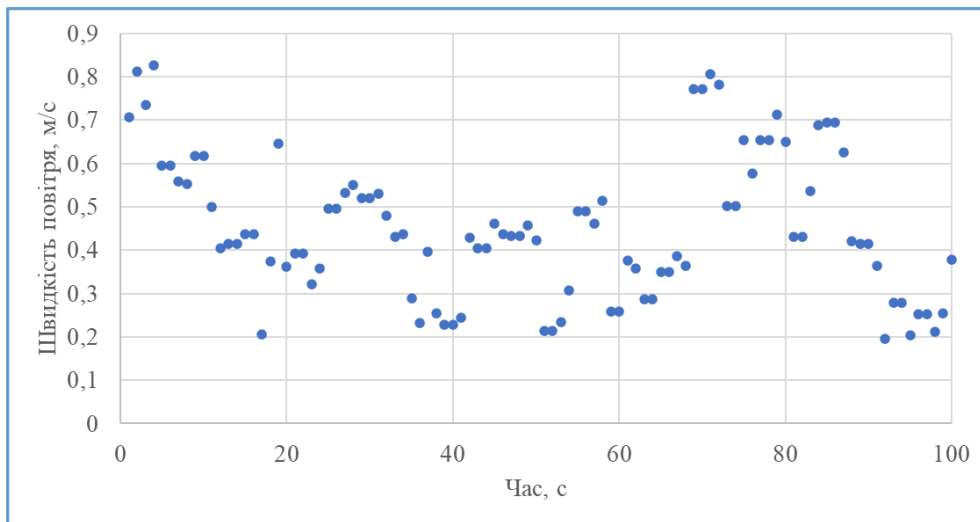


Рис. 3. Вибірка зміни швидкості повітря на робочому місці зварювальника

Аналізуючи отримані результати за табл. 4, очевидно, що швидкість руху повітря під час занять змінюється на усіх робочих місцях порівняно з початковими умовами.

Найменше зростання спостерігається на робочому місці фрезерувальника (з 0,027 м/с до 0,15 м/с) та токаря (з 0,036 м/с до 0,18 м/с), що обумовлено відносною статичністю виконуваних операцій. Найбільша динаміка прослідковується на робочому місці зварювальника (з 0,03 м/с до 0,45 м/с).

Зміни у швидкості повітря пов'язані з режимами роботи системи вентиляції, рухом частин технологічного обладнання та людей, які створюють конвекційні потоки, що підвищують швидкість руху повітря.

Найбільші рівні руху повітря відбувається в майстерні зварювальників та електромонтерів, що пов'язано з потребою виведення шкідливих газів та продуктів технологічних процесів з повітряного простору виробничих майстерень.

Загалом отримані результати виміру швидкості повітря відповідають нормативним актам.

Висновки

За результатами дослідження показників якості повітря у навчальних майстернях закладів передвищої та професійної (професійно-технічної) освіти м. Кременчук встановлено, що концентрація респірабельних фракцій пилу ($PM_{2.5}$ та PM_{10}) значно варіюється залежно від виду навчального приміщення та характеру виконуваних робіт. Найбільший рівень запиленості виявлено на робочому місці зварювальника:

– середньодобове значення $PM_{2.5}$ становить 16,9 мкг/м³,

– середньодобове значення PM_{10} – 137,4 мкг/м³, що перевищує встановлені норми [1].

Дослідження показало, що на робочих місцях токаря, фрезерувальника й електромонтера також спостерігається підвищення показників виробничого пилу респірабельної фракції.

Динаміка зміни концентрації твердих часток свідчить про зростання концентрацій виробничого пилу на початку занять, зниження під час перерв та значне підвищення під час активної роботи учнів. Оцінка роботи вентиляційних систем показала, що швидкість руху повітря під час занять зросла у 5–15 разів порівняно з початковими значеннями. Найбільший приріст зафіксовано на робочому місці зварювальника – з 0,03 м/с до 0,45 м/с (збільшення у 15 разів), що свідчить про ефективність вентиляції у видаленні шкідливих часток та газів.

Отримані результати підкреслюють необхідність подальшого вдосконалення вентиляційних систем, розробки заходів із зниження запиленості робочої зони учня. Запровадження превентивних заходів з унормування показників повітряного середовища та використання сучасних і ефективних засобів індивідуального захисту органів дихання дозволить зменшити негативний вплив перевищених показників якості повітря виробничих майстерень на здоров'я учнів.

СПИСОК ЛІТЕРАТУРИ

1. WHO Air Quality Guidelines, 2021.
2. Про затвердження Санітарного регламенту для закладів загальної середньої освіти : наказ Міністерства охорони здоров'я України № 2205 від 25.09.202 р. URL: <https://surl.li/spwxv>
3. Про затвердження Санітарного регламенту для дошкільних навчальних закладів : наказ Міністерства охорони здоров'я України № 234 від 24.03.2016 р. URL: <https://surl.li/diknmr>
4. Про затвердження Типового положення про організацію освітнього процесу в закладах фахової передвищої освіти та Положення про практичну підготовку здобувачів фахової передвищої освіти : наказ Міністерства освіти і науки України № 510 від 02.05.2023 р. URL: <https://surl.li/aohqck>
5. Про затвердження Правил охорони праці під час холодного оброблення металів : наказ Міністерства енергетики та вугільної промисловості України № 749 від 16.10.213 р. URL: <https://surl.li/kyvpiv>
6. Директива (ЄС) 2024/2881 Європейського Парламенту та Ради від 23 жовтня 2024 року про якість атмосферного повітря та чистіше повітря для Європи. URL: <https://surl.li/brpccq>

Received (Надійшла) 08.01.2025

Accepted for publication (Прийнята до друку) 05.03.2025

Research of the state of the air environment in the premises of primary and vocational (vocation-technical) education institutions

Dmytro Rieznik, Olga Chencheva, Yevhenii Lashko, Ivan Petrenko, Dmytro Havrylets

Abstract. Purpose. Study of the dynamics of the respirable fraction of dust ($PM_{2.5}$ and PM_{10}) and analysis of its impact on the health of students in classrooms and training workshops of institutions of higher and vocational education in the city of Kremenchuk. **Method.** The study of air quality in educational premises was carried out on the basis of the empirical method during a field experiment, statistical analysis of the obtained data and approximation of the results. **The results.** The study showed that the concentration of the respirable fraction of dust ($PM_{2.5}$, PM_{10}) in training workshops varies significantly depending on the type of premises and the work performed. The concentration of dust changes during the day: it increases at the beginning of classes, decreases during breaks, and increases significantly during active work. Exceeding the dust load was recorded at the investigated workplaces, and the highest dust level from the purpose of the article was recorded at the welder's workplace. The assessment of ventilation systems showed that the speed of air movement during classes increased by 5-15 times, which indicates the efficiency of their work, especially at the welder's workplace. Studies have shown that in order to improve the level of dust pollution in the working area, it is necessary to improve and automate ventilation and aspiration systems, as well as to implement the use of personal protective equipment. **Scientific novelty.** For the first time, the conducted research takes into account the real conditions of the educational and production process in institutions of higher and vocational education. A comprehensive approach to air quality analysis is proposed, which includes pollution monitoring ($PM_{2.5}$, PM_{10}), detection of patterns of changes in dust concentration depending on the stages of training process, mathematical description of pollution dynamics, and evaluation of the efficiency of ventilation systems. The obtained results can become the basis for improving air quality control measures in educational workshops and developing regulations that will ensure the safety of the educational process. **Practical value.** The results of the research allow: to optimize ventilation systems by determining the required speed of air flow based on real data on the concentration of pollutants in the air of educational and production workshops; develop adaptive air filtration systems to reduce the level of the respirable fraction of dust in production facilities; justify the need for systematic monitoring of the air environment in training workshops, in particular using modern technologies (wireless sensor modules); introduce protective measures for students aged 14-19, who are most sensitive to industrial dust exposure. The application of the obtained results will contribute to increasing the level of safety of the educational process, reducing professional risks for students and improving working conditions in educational workshops.

Keywords: respirable dust fraction, training workshops, dust concentration, pollution dynamics, ventilation system.

О. В. Третьяков, Б. Д. Халмурадов, Н. М. Кічата, А. В. Ремська

Державний університет «Київський авіаційний інститут», Київ, Україна

ПІДХІД ДО КІЛЬКІСНОЇ ОЦІНКИ СТІЙКОСТІ ОБ'ЄКТІВ КРИТИЧНОЇ ІНФРАСТРУКТУРИ

Анотація. Стійкість об'єктів критичної інфраструктури визначається як здатність готуватися до мінливих умов та адаптуватися до них, а також витримувати збої й швидко відновлюватися після них, включно з навмисними атаками, аваріями або природними загрозами. Забезпечення високого рівня безпеки і стійкості об'єктів критичної інфраструктури для України в умовах агресії РФ є край актуальним питанням. **Мета досліджень** полягала в розробці методологічного підходу для кількісної оцінки рівня стійкості об'єктів критичної інфраструктури не залежно від секторів критичної інфраструктури, до якого вони належать та усіх видів проектних загроз. **Об'єкт досліджень** – безпека і стійкості об'єктів критичної інфраструктури. **Предмет досліджень** – методологія кількісної оцінки рівня стійкості об'єктів критичної інфраструктури. **Отримані результати.** Початкова стадія після небезпечної події це різновид прояву катастрофи «складка», визначеної у теорії катастроф. Такий підхід дозволяє визначити втрати обсягу надання послуги об'єктом критичної інфраструктури внаслідок небезпечної події, що визначають вразливість об'єкту. Стійкість об'єкта критичної інфраструктури можна визначити, як добуток часу на повне відновлення на витрати, пов'язані з відновленням обсягу послуг до вихідного рівня. Чим більше цей показник тим менше стійкість об'єкту критичної інфраструктури. Застосування імітаційної моделі для каскадних ефектів, дає можливість отримати ймовірнісні оцінки розвитку подій за визначеними сценаріями і дозволяє здійснити оцінювання загроз для об'єкта критичної інфраструктури за величиною ймовірності настання подій і переходів між ними. На основі отриманих значень ймовірності настання небезпечних подій для усіх елементів об'єкта критичної інфраструктури визначаємо найбільш вразливі і для них проводимо кількісну оцінку стійкості. Це дає змогу оцінити необхідні ресурси (фінансові, матеріальні, енергетичні, людські, транспортні тощо) для підвищення стійкості. Визначити необхідні елементи резервування для уникнення каскадних ефектів і небажаних наслідків. **Висновки.** На основі теорії катастроф розроблено єдиний методологічний підхід для кількісної оцінки рівня стійкості об'єктів критичної інфраструктури не залежно від секторів критичної інфраструктури, до якого вони належать. Запропонований підхід дозволяє проводити аналіз стійкості по усіх елементах об'єкту критичної інфраструктури, проводити співставний аналіз вразливості і стійкості об'єктів сектору, оцінювати розміри необхідних додаткових інвестицій для зниження вразливості і підвищення стійкості елементів об'єкту, розробляти секторальні програми підвищення стійкості об'єктів сектору, визначити необхідні територіальні ресурси резервування та їх обсяги.

Ключові слова: об'єкт критичної інфраструктури, безпека, стійкість, теорія катастроф.

Вступ

Основні базові поняття у сфері безпеки критичної інфраструктури, що визначені в нормативних документах і наведені у [1], які відображають сутність поняття «безпека», можуть бути визначені як такі: безпека критичної інфраструктури, цілісність, самодостатність, стійкість, загрози критичній інфраструктурі, небезпека критичній інфраструктурі, технології попередження загрози, технології виявлення загрози, технології ліквідації загроз, технології відновлення діяльності, відшкодування завданих збитків. Зокрема стійкість об'єкта критичної інфраструктури визначена як наявність необхідних умов, елементів та системних зв'язків між ними на рівні, достатньому для стримування дії загроз та відновлення після їх дії [1].

Концепція стійкості розроблялася і застосовувалася в різноманітних напрямках діяльності (психологія, психіатрія, екологія, соціальні науки, економіка та техніка) протягом кількох десятиліть [2, 3], останнім часом вона привертає дедалі більшу увагу у сфері управління ризиками. Зокрема, спільнота критичної інфраструктури еволюціювала від першочергового акценту на захисті безпеки у 1990-х роках до більш широкого акценту на безпеці та стійкості.

У сфері національної безпеки визначати національну політику щодо зміцнення та підтримки безпечної, функціональної та стійкої критичної інфраструктури в секторах, які є важливими для національної

безпеки, громадського здоров'я та безпеки, економічної життєздатності та загальної якості життя. При цьому стійкість визначається як здатність готуватися до мінливих умов та адаптуватися до них, а також витримувати збої й швидко відновлюватися після них, включно з навмисними атаками, аваріями або природними загрозами [4].

Перехід від захисту критичної інфраструктури до забезпечення її стійкості має на меті врахувати ключові зміни в ландшафті ризиків, що характеризуються зростанням невизначеності. Для того, щоб краще інтегрувати складність, взаємозалежність і взаємопов'язаність критичної інфраструктури, прийняття системного підходу до стійкості критичної інфраструктури забезпечує додаткові перспективи [5].

Підвищення стійкості національної та європейської критичної інфраструктури визначено одним із пріоритетів безпекової політики ЄС та закріплено в рішеннях Ради ЄС, спрямованих на посилення заходів із підвищення стійкості критичної інфраструктури [6]. Таким чином, забезпечення високого рівня безпеки і стійкості об'єктів критичної інфраструктури для України в умовах агресії РФ є край актуальним питанням.

Аналіз літературних даних та постановка проблеми. Людство захищає критичну інфраструктуру з часів винайдення колеса, однак за останні 20 років більшість національних політик і стратегій у сфері критичної інфраструктури на Заході

еволюціонували від зосередження виключно на захисті критичної інфраструктури до підвищення її безпеки та стійкості. Цей зсув в першу чергу пов'язаний з тим, що перед зацікавленими сторонами стоїть надскладне завдання захистити усі системи критичної інфраструктури від зростаючої кількості факторів ризику, з якими вони стикаються.

Відповідно до концепції Безпеки та стійкості критичної інфраструктури (БСКІ), терміни «безпеки» і «стійкості», безумовно, підтримують ідею захисту, але вони мають специфічні значення. *Безпека* означає зменшення ймовірності успішних атак на критичну інфраструктуру або наслідків природних чи техногенних катастроф шляхом застосування фізичних засобів або оборонних заходів кібербезпеки. *Стойкість* – це здатність критичної інфраструктури протистояти, поглинати, відновлюватися або успішно адаптуватися до мінливих умов. Відмовостійка інфраструктура є надійною, гнучкою, адаптивною, здатною протистояти і швидко відновлюватися після збоїв, навмисних атак,

аварій або природних загроз чи інцидентів. З огляду на зростання природних і техногенних загроз і вразливостей, з якими стикаються сучасні суспільства та які викривають обмеженість традиційної оцінки ризиків і зусиль зі зниження ризиків, концепція БСКІ видається особливо корисною для формування політики, спрямованої на пом'якшення наслідків таких подій, і вказує на життєво важливу потребу країн у розробці та впровадженні всеосяжної стратегії управління ризиками [7]. Основні елементи цього визначення – здатність готуватися до мінливих умов і адаптуватися до них, а також витримувати і швидко відновлюватися після збоїв – можна охарактеризувати чотирма структурними елементами: готовність, заходи з пом'якшення наслідків, реагування і заходи з відновлення [8].

Разом ці чотири складники можуть допомогти фахівцям розкласти концепцію стійкості на практичні кроки та, зрештою, якісно оцінити прогрес у підвищенні стійкості з плином часу. У табл. 1 описані ці складники і наведені приклади для розгляду [9].

Таблиця 1 – Складники стійкості

Складники	Опис	Приклади
Підготовленість	Діяльність, спрямована на передбачення відповідних загроз/небезпек і можливих наслідків від їх виникнення, включно із заходами з попередження та захисту; свідчить про адаптивність інфраструктурних систем і процес інтеграції та врахування здобутого досвіду.	Утримання сил безпеки. Установлення/моніторинг фізичного контролю доступу. Розробка планів безперервності, планів на випадок надзвичайних ситуацій і планів кібербезпеки; навчання персоналу щодо планів. Проведення регулярних навчань для перевірки планів Створення механізмів обміну інформацією.
Пом'якшення наслідків	Діяльність, спрямована на протистояння та/або поглинання негативних наслідків події, зменшення тяжкості або наслідків загрози; свідчить про надійність інфраструктури.	Модернізація підприємств для пом'якшення наслідків різних природних загроз (наприклад, протипаводкове обладнання, протипаводкові бар'єри). Модернізація обладнання, яке буде протистояти передбачуваним небезпекам. Підвищення надійності/резервування систем підтримки інфраструктури. Створення альтернативного резервного майданчика, який може продовжити роботу після інциденту й сприяти відновленню. Розуміння міжгалузевих залежностей від ключових зовнішніх ресурсів (наприклад, електроенергія, паливо, вода, зв'язок). Завбачлива підготовка додаткових запасів (наприклад, палива, резервних генераторів, резервного зв'язку).
Реагування	Заходи та програми, що здійснюються або розробляються для реагування та адаптації до негативних наслідків події; свідчить про винахідливість власників та операторів інфраструктури в управлінні кризовими ситуаціями	Підтримання можливостей реагування на місці на ключові небезпеки (наприклад, розливи хімічних речовин, пожежі, вибухові речовини, збройні напади, надзвичайні ситуації медичного характеру). Побудова відносин з місцевими службами швидкого реагування та міжсекторальними партнерами; наявність можливостей для управління позаштатними ситуаціями на місці, включно з навченим персоналом, функціональним оперативним центром і розумінням міжгалузевих проблем.
Відновлення	Діяльність і програми, спрямовані на те, щоб допомогти організаціям повернути умови роботи до прийняттого рівня та відновитися після події; свідчить про здатність швидко відновити надання послуг.	Укладання угод про першочергове відновлення з ключовими постачальниками послуг; оцінка часу й заходів, необхідних для відновлення повноцінної роботи організації після збою. Стратегії швидкої заміни/ремонту критично важливих компонентів (наприклад, сертифіковані постачальники. Підтримка запасів на випадок надзвичайних ситуацій).

З наведеної в таблиці інформації видно, що стійкість об'єктів критичної інфраструктури є інтегруючим показником, який охоплює практично усі складові функціонування об'єктів і це ускладнює його пряму кількісну оцінку.

Барамі Б. підкреслює складний і багатогранний характер стійкості критичної інфраструктури. Барамі Б. застосовує багаторівневий підхід, що ґрунтується на оцінці ризиків і враховує взаємозалежності складних інфраструктур, розглядаючи при цьому потенційні рішення, які можна застосувати протягом життєвого циклу інфраструктурної системи (тобто, проектування, будівництво та експлуатація). Таким чином, стійкість визначається не як єдиний результат або виключно здатність до відновлення після катастрофи, а скоріше як динамічний процес, який застосовує метод, заснований на ризиках і життєвому циклі, для усунення вразливостей систем критичної інфраструктури, роблячи системи більш відмовостійкими, ефективними, розумнішими і здатними краще адаптуватися до несподіваних викликів [10].

Таким чином, коли відбувається інцидент, цілі стійкості критично важливої інфраструктури можуть бути виміряні у двох вимірах: обмеження масштабу пошкоджень та обмеження тривалості перерви у наданні послуг, спричиненої пошкодженнями. Важливо зазначити, що відновлення не обов'язково означає повернення до попереднього стану, який існував до інциденту, але може передбачати зміну, адаптацію до нових умов та покращення функціональності систем з плином часу.

Проведена якісна оцінка рівня стійкості об'єктів критичної інфраструктури за запропонованим підходом дійсно допомагає фахівцям розкласти концепцію стійкості на практичні кроки, але не дає кількісної оцінки цього показника, що не дозволяє проводити порівняльний аналіз стійкості об'єктів критичної інфраструктури, тим більше, якщо вони належать до різних секторів критичної інфраструктури.

На сьогодні в Україні визначено 24-и сектора критичної інфраструктури, чи мало об'єктів в яких потребують додаткових інвестицій для підвищення стійкості [11]. Саме через це державним уповноваженим органом у сфері захисту критичної інфраструктури України затверджені Методичних рекомендацій щодо розроблення та затвердження місцевих програм забезпечення безпеки та стійкості критичної інфраструктури, програм підвищення стійкості територіальних громад до кризових ситуацій, викликаних припиненням або погіршенням надання важливих для їх життєдіяльності послуг чи для здійснення життєво важливих функцій [12]. Для розробки місцевих програм забезпечення безпеки та стійкості критичної інфраструктури та програм підвищення стійкості територіальних громад до кризових ситуацій необхідно серед цих об'єктів визначити пріоритетні, а це без кількісної оцінки стійкості неможливо. Тому край нагальним є розробка методики кількісної стійкості об'єктів критичної інфраструктури та їх елементів.

Мета та задачі дослідження. Мета досліджень полягає в розробці методологічного підходу для кількісної оцінки рівня стійкості об'єктів критичної

інфраструктури не залежно від секторів критичної інфраструктури, до якого вони належать та усіх видів проєктних загроз.

Для досягнення мети були поставлені наступні завдання:

- знайти підхід для не прямого кількісного оцінювання стійкості об'єктів критичної інфраструктури;
- оцінити його адекватність з позицій теорії катастроф.

Результати досліджень

Задля подолання складнощів при розгляді складових стійкості і концептуалізації їх у контексті інфраструктурних операцій з точки зору часової перспективи можна розглянути функціонування об'єкту критичної інфраструктури як залежність обсягу надання послуги в часі за різних умов, особливо при дії небезпечного фактору (природного, техногенного, терористичного, воєнного походження), як це означено на рисунку 1.

До моменту прояву небезпечної події об'єкт критичної інфраструктури функціонує у сталому режимі і надає послуги у проєктному обсязі. З моменту настання небезпечної події: природного катаклізму (землетрус, зсув, повень тощо), техногенних аварій, несанкціонованого втручання, кібератаки, терористичного акту, воєнного нападу, тощо обсяг надання послуги об'єктом критичної інфраструктури різко знижується, або зовсім припиняється (t_1). Після чого настає період підготовки до відновлення функціонування об'єкту (t_2) (проєктні роботи, зосередження необхідних матеріальних ресурсів, залучення підрядників тощо), що передує відновлюваним роботам, після чого розпочинається відновлення потужностей об'єкту з поступовим виходом на сталий режим надання послуги у проєктному обсязі.

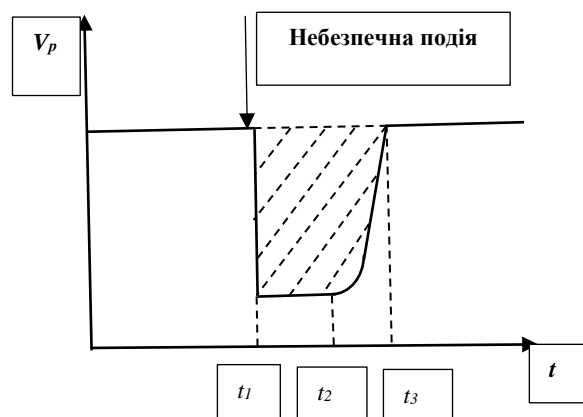


Рис. 1. Залежність обсягу надання послуги об'єктом критичної інфраструктури від часу при дії небезпечного фактору

Початкова стадія після небезпечної події це різновид прояву катастрофи у теорії катастроф [13]. Катастрофа типу «складка» ($x^3 + ax$) – одна з найпростіших катастроф. Стандартна деформація (падіння рівня об'єму послуг) в цьому випадку задається формулою:

$$V_p(t) = t^3/3 + a \cdot t, \quad (1)$$

де V_p – обсяг надання послуги; t – час.

Числовий коефіцієнт введено, щоб спростити подальші розрахунки. Багаторізнovidність M такої катастрофи визначається рівнянням:

$$0 = \frac{d}{dt} V_p(t) = t^2 + a. \quad (2)$$

Втрати обсягу надання послуги об'єктом критичної інфраструктури внаслідок небезпечної події буде визначатися:

$$W = \int_{t_1}^{t_3} V_p(t) dt, \quad (3)$$

і характеризувати вразливість об'єкту критичної інфраструктури.

Вивчення елементарних катастроф починається з перелічення основних структур, пов'язаних з катастрофами [13]. Для цього розглянемо самеїство функції:

$$V: S \times C \rightarrow R, \quad (4)$$

де S – деяка багаторізнovidність R^n , яка має назву простір станів; C – багаторізнovidність R^r , що визначається як простір управління, а число r розмірність деформації.

Багаторізнovidністю катастрофи M називається підмножина в $R^n \times R^r$, що визначається рівнянням:

$$D \cdot V_c(t) = 0, \quad (5)$$

де $V_c(t) = V(t, c)$ – множина усіх критичних точок потенціалів V_c з сімейства V .

Відзеркаленням катастрофи S називається обмеження M натуральної проєкції:

$$\pi: R^n \times R \rightarrow R^r, \quad (6)$$

$$\pi(t, C) = C. \quad (7)$$

Особливою множиною S називається підмножина в M , яка складається з особливих точок відображення χ , де ранг похідної D_χ менше ніж r . Образ особливої множини $\chi(S) \in C$ називають біфуркаційною множиною B .

Існує множина тих точок $(x, c) \in M$, в яких $V_c(x)$ має вирожену критичну точку. Значить, B являє собою місце, де змінюється кількість і природа критичних точок. У зв'язку з структурною усталеністю морсовських функцій така зміна може відбутися лише за умови переходу крізь вирожену критичну точку. У більшості застосувань найбільш важливим є саме біфуркаційна множина, так як вона лежить в просторі управління.

Рівняння (2) показує, що в якості карти для M треба взяти координату t .

Загальна точка багаторізнovidності M записується у вигляді:

$$(t, a) = (t - t^2). \quad (8)$$

Розкладення в ряд Тейлора, яке відповідає цій точці на S має наступний вигляд

$$V_p(t + T) = \frac{1}{3}(t + T)^3 + (-t^2)(t + T) = \frac{1}{3}T^3 + tT^2 + 0T - \frac{2}{3}t^3. \quad (9)$$

Таким чином, багаторізнovidність катастрофи являє собою параболу, а біфуркаційна множина складається з однієї точки.

Гладкість цієї поверхні не гарантує, що при плавній зміні однієї змінної обидві інші також змінюються плавно.

Виходячи з такого підходу стійкість об'єкта критичної інфраструктури (або його частки, підрозділу тощо) можна визначити, як добуток часу на повне відновлення на витрати, пов'язані з відновленням обсягу послуг до вихідного рівня:

$$S_i = \Delta t \cdot \sum E_i, \quad (10)$$

де Δt – час на повне відновлення об'єкту критичної інфраструктури (або його частки, підрозділу тощо); $\sum E_i$ – усі витрати на відновлення (фінансові, матеріальні, енергетичні, людські, транспортні тощо).

Чим більше цей показник тим менше стійкість об'єкту критичної інфраструктури ($S_i \rightarrow \min$).

Для зручності витрати на відновлення об'єкту критичної інфраструктури (або його частки, підрозділу тощо) можна брати не як абсолютну величину, а як частку від проєктної вартості об'єкту.

Якщо кількісна оцінка ризику реалізації небезпечних подій здійснюється на основі імітаційної моделі для оцінювання загрози виникнення каскадних ефектів для різних сценаріїв розвитку подій у зоні об'єкта критичної інфраструктури, яка передбачає виконання таких процедур:

- визначення подій в сценарії розвитку ситуації (складові елементи сценарію, що здійснюють потенційний вплив на реалізацію загрози);
- визначення множини можливих станів подій, що впливають на рівень загрози;
- формування сценаріїв розвитку загрози (визначення ланок, що складаються з пар: «подія – перехід в заданий стан»), що призводять до реалізації загрози, представлення структурно-логічної моделі розвитку кризової ситуації, що має складну структуру за різними варіантами розвитку сценарію на об'єкті критичної інфраструктури;
- формування оргграфу сценаріїв загроз (структурно-логічна модель, що включає всі сценарії реалізації загрози);
- оцінка ймовірностей станів подій та їх переходів;
- оцінювання ймовірності реалізації сценаріїв загроз.

Застосування такої імітаційної моделі для каскадних ефектів, дає можливість отримати ймовірнісні оцінки розвитку подій за визначеними сценаріями і дозволяє здійснити оцінювання загроз для об'єкта критичної інфраструктури за величиною імовірності настання подій і переходів між ними.

Приклад:

Нехай на трансформаторній підстанції стався вибух з середнім рівнем правдоподібності реалізації загрози. Тоді оцінка наслідків впливу загроз (стійкості) буде мати наступний вигляд:

– час на повне відновлення об'єкту критичної інфраструктури $\Delta t \approx 30$ днів;

- витрати на відновлення об'єкту $E_i = 2$ млрд грн;
- чисельність населення, яке зазнало ризику втрати – 200 осіб;
- зниження рівня надання життєво важливих функцій/послуг, тривалість зниження рівня здоров'я (фатальність) – 3 дні;
- шкода довкіллю – територія району на термін до 1 року;
- зниження рівня виробництва / надання основної послуги (продукції) (електрозабезпечення) – втрата послуги (електро- забезпечення) для понад 300 тис. абонентів; чисельність ураженого населення (вимушена міграція чи потреба прихистку/допомоги);
- вплив на 100 тис. осіб на термін до 30 днів;

– публічне сприйняття ситуації, спричинене впливом загрози – занепокоєння відчуває 5% населення [14].

Розрахунок стійкості:

$$S_i = \Delta t \cdot \sum E_i = 30 \text{ днів} \cdot 2 \cdot 10^9 \text{ грн} = 6 \cdot 10^{10} (\text{днів} \cdot \text{грн}). \quad (11)$$

Це значення відображає інтегральну оцінку впливу загрози на об'єкт та масштаби витрат, необхідних для його повного відновлення.

Наявність резервної трансформаторної підстанції дозволить скоротити час на повне відновлення надання послуги об'єктом КІ до декількох хвилин, що суттєво забезпечить підвищення стійкості об'єкту (рис. 2).

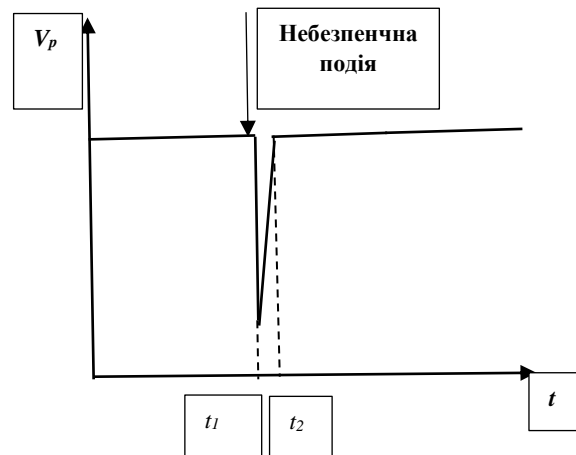


Рис. 2. Залежність обсягу надання послуги об'єктом критичної інфраструктури від часу при наявності резервного елемента об'єкту КІ

На основі отриманих значень імовірності настання небезпечних подій для усіх елементів об'єкта критичної інфраструктури визначаємо найбільш вразливі і для них проводимо кількісну оцінку стійкості. Це дає змогу оцінити необхідні ресурси (фінансові, матеріальні, енергетичні, людські, транспортні тощо) для підвищення стійкості.

Визначити необхідні елементи резервування для уникнення каскадних ефектів і небажаних наслідків.

Такий підхід для об'єкта критичної інфраструктури дозволяє:

- провести аналіз стійкості по усіх елементах об'єкту;
- визначити вразливість і стійкість кожного при реалізації будь яких загроз у кількісному вимірі;
- означити найбільш вразливі і найменш стійкі елементи об'єкту;
- оцінити розмір необхідних додаткових інвестицій для зниження вразливості і підвищення стійкості елементів об'єкту;
- визначити необхідні ресурси резервування та їх обсяг.

Для секторального органу у сфері захисту критичної інфраструктури:

- проводити співставний аналіз вразливості і стійкості об'єктів сектору;

- визначити найбільш вразливі і найменш стійкі;

- розробити секторальну програму підвищення стійкості об'єктів сектору;

- визначити пріоритети інвестування для підвищення стійкості об'єктів сектору.

Для територіальних громад:

- провести аналіз стійкості по усіх об'єктах критичної інфраструктури;

- визначити найбільш вразливі і найменш стійкі на території громади;

- розробити територіальну програму підвищення стійкості об'єктів критичної інфраструктури;

- визначити необхідні територіальні ресурси резервування та їх обсяг.

Висновки

На основі теорії катастроф розроблено єдиний методологічний підхід для кількісної оцінки рівня стійкості об'єктів критичної інфраструктури не залежно від секторів критичної інфраструктури, до якого вони належать.

Запропонований підхід дозволяє:

- проводити аналіз стійкості по усіх елементах об'єкту критичної інфраструктури,

- проводити співставний аналіз вразливості і стійкості об'єктів сектору,

оцінювати розміри необхідних додаткових інвестицій для зниження вразливості і підвищення стійкості елементів об'єкту,

розробляти секторальні програми підвищення стійкості об'єктів сектору та визначити необхідні територіальні ресурси резервування та їх обсяги.

СПИСОК ЛІТЕРАТУРИ

1. Франчук В.І., Пригунов П.Я., Мельник С.І. Безпека об'єктів критичної інфраструктури в Україні: організаційно-нормативні проблеми та підходи. *Соціально-правові студії*. 2021. Випуск 3 (13). С. 142-148.
2. Реншлер, Кріс С., Емі Е. Фрейзер, Люсі А. Арентт, Джан-Паоло Чімелларо, Андрій М. Рейнхорн та Мішель Брюно, «Структура для визначення та вимірювання стійкості на рівні громади: концепція стійкості людей», Національний інститут стандартів і технологій. 2010 р. 73 р. Дата перегляду 13 лютого 2020 р. www.hSDL.org/?view&did=790013.
3. Rose, A., 2009, Economic Resilience to Disasters, CARRI Research Report 8. Available at http://www.resilientus.org/Library/Research_Report_8_Rose_1258138606.pdf, accessed on November 2, 2010.
4. АПКБІ, «Національний план захисту інфраструктури (НПЗІ) 2013: партнерство заради безпеки та стійкості критичної інфраструктури». 2013 р. Дата перегляду: 13 лютого 2020 року. www.cisa.gov/national-infrastructure-protection-plan
5. OECD (2019), Належне врядування для забезпечення стійкості критичної інфраструктури, Огляди політики управління ризиками OECD, Публікація OECD, Париж. <https://doi.org/10.1787/02f0e5a0-en>.
6. Council Recommendation on a Union-wide coordinated approach to strengthen the resilience of critical infrastructure. Brussels, 9 December 2022 (OR. en) 15623/22. URL: <https://data.consilium.europa.eu/doc/document/ST-15623-2022-INIT/en/pdf>
7. Сприяння колективній обороні НАТО: Безпеката стійкість критичної інфраструктури. Посібник НАТО COE-DAT/ К. Андерсон, М. Бейкер, Р. Бірс та ін. 2022. 469 р.
8. Methodology for assessing regional infrastructure resilience (CISA). May 2021. Version 1.0.
9. Карлсон, Дж. Лон, Ребекка А. Хаффенден, Гілберт В. Бассет, Вільям А. Берінг, Майкл Д. Коллінз, III, Стивен М. Фолга, Фредерік Петі, Джулія А. Філіпс, Дуетін Р. Вернер і Рональд Вітфілд, «Стійкість: Теорія та застосування». 2012 р. США. doi:10.2172/1044521. www.osti.gov/biblio/1044521-resilience-theory-application.
10. Барамі Б. Відмовостійкість інфраструктури: структура на основі ризиків Департамент США Транспорт, https://www.volpe.dot.gov/sites/volpe.dot.gov/files/docs/Infrastructure%20Resiliency_A%20Risk-Based%20Framework.pdf (дата доступу: 25 лютого 2019 р.).
11. Деякі питання об'єктів критичної інфраструктури. Постанова КМ України від 09.10.2020 р. № 1109 (зі змінами) <https://zakon.rada.gov.ua/laws/show/1109-2020-%D0%BF#Text>.
12. Про затвердження Методичних рекомендацій щодо розроблення та затвердження місцевих програм забезпечення безпеки та стійкості критичної інфраструктури, програм підвищення стійкості територіальних громад до кризових ситуацій, викликаних припиненням або погіршенням надання важливих для їх життєдіяльності послуг чи для здійснення життєво важливих функцій. Наказ Адміністрації Держспецзв'язку України видала наказ від 30.11.2023 № 997. <https://www.cip.gov.ua/ua/docs/nakaz-administraciyi-derzhspetsvvyazku-vid-30-listopada-2023-roku-997>.
13. Thompson, J. Michael T. Instabilities and Catastrophes in Science and Engineering. New York: Wiley, 1982. 256 р.
14. Суходоля О. М. Стійкість критичної енергетичної інфраструктури та життєдіяльності громад. Нац. ін-т стратег. дослідж., 2024. – с.160. URL: <https://doi.org/10.53679/niss-analytrep.2024.04>

Received (Надійшла) 11.12.2024

Accepted for publication (Прийнята до друку) 26.02.2025

Approach to quantitative assessment of the resilience of critical infrastructure facilities

O. Tretyakov, B. Khalmuradov, N. Kichata, A. Remska

Abstract. The resilience of critical infrastructure facilities is defined as the ability to prepare for and adapt to changing conditions, as well as to withstand and recover quickly from failures, including deliberate attacks, accidents or natural threats. Ensuring a high level of security and resilience of critical infrastructure facilities for Ukraine in the face of Russian aggression is an extremely urgent issue. **The purpose of the research** was to develop a methodological approach for quantifying the level of resilience of critical infrastructure facilities, regardless of the critical infrastructure sectors to which they belong and all types of project threats. **The object of research** is the safety and resilience of critical infrastructure facilities. **The subject of research** is the methodology for quantifying the level of resilience of critical infrastructure facilities. **Results obtained.** The initial stage after a dangerous event is a type of manifestation of the "fold" catastrophe defined in the theory of catastrophes. This approach allows you to determine the losses in the volume of service provision by a critical infrastructure facility as a result of a dangerous event, which determine the vulnerability of the object. The higher this indicator, the lower the stability of the critical infrastructure facility. The use of the simulation model for cascading effects makes it possible to obtain probabilistic estimates of the development of events according to certain scenarios and allows to assess threats to a critical infrastructure facility by the probability of occurrence of events and transitions between them. Based on the obtained values of the probability of occurrence of dangerous events for all elements of the critical infrastructure facility We identify the most vulnerable and quantify resilience for them. This makes it possible to assess the necessary resources (financial, material, energy, human, transport, etc.) to increase resilience. Determine the necessary redundancy elements to avoid cascading effects and unintended consequences. **Conclusions.** On the basis of the theory of disasters, a single methodological approach has been developed for quantifying the level of resilience of critical infrastructure facilities, regardless of the critical infrastructure sectors to which they belong. The proposed approach allows to conduct a stability analysis for all elements of a critical infrastructure facility, to conduct a comparative analysis of vulnerability and resilience of sector facilities, to assess the amount of additional investments needed to reduce vulnerability and increase the resilience of facility elements, to develop sectoral programs to increase the resilience of sector facilities, to determine the necessary territorial resources of reservation and their volumes.

Keywords: critical infrastructure facility, safety, resilience, theory of disasters.

Зв'язок, телекомунікації та радіотехніка

УДК 528.854

doi: 10.26906/SUNZ.2025.1.184-187

Ю. Р. Зоураб, Н. В. Єрмілова

Національний університет «Полтавська політехніка імені Юрія Кондратюка», Полтава, Україна

ІДЕНТИФІКАЦІЯ ОБ'ЄКТІВ ПОШТОВОГО ЗВ'ЯЗКУ ДЛЯ ОПТИМІЗАЦІЇ ОПРАЦЮВАННЯ ПОШТОВОЇ КОРЕСПОНДЕНЦІЇ

Анотація. У статті розглянуто актуальну проблему ідентифікації об'єктів поштового зв'язку як важливого елемента оптимізації процесів обробки поштової кореспонденції. Зростання обсягів відправлень, зумовлене розвитком електронної комерції та глобалізацією, створює значні виклики для поштових служб, зокрема щодо забезпечення швидкості, точності та економічної ефективності обробки. Традиційні методи ручного сортування є ресурсомісткими, схильними до помилок та не відповідають сучасним вимогам до продуктивності, особливо в пікові періоди. Дослідження фокусується на впровадженні автоматизованих систем ідентифікації, які базуються на використанні технологій RFID, систем комп'ютерного зору та штрих-кодів. Проведено аналіз їх ефективності за трьома основними показниками: швидкість обробки, точність ідентифікації та економічна ефективність. Встановлено, що автоматизація дозволяє значно підвищити швидкість обробки: RFID забезпечує обробку до 500 відправлень на годину, системи комп'ютерного зору — до 450 відправлень, а штрих-кодові технології — до 400 відправлень. Це значно перевищує швидкість ручного сортування, яка становить лише 200 відправлень на годину. Щодо точності ідентифікації, RFID досягає 99%, що робить цю технологію найбільш надійною, особливо в умовах високого навантаження. Системи комп'ютерного зору забезпечують точність 98%, а штрих-коди, через чутливість до пошкоджень маркувань, мають точність на рівні 95%. Економічна оцінка показала, що автоматизація дозволяє суттєво скоротити витрати на обробку відправлень. Впровадження RFID у поштовому центрі з навантаженням 10 000 відправлень на день дало змогу скоротити витрати на оплату праці на 40%, що еквівалентно \$20 000 щомісячної економії. Системи комп'ютерного зору дозволяють заощадити \$15 000, знижуючи потребу у ручній перевірці. Експериментальне впровадження RFID у поштовому центрі Києва підтвердило практичну ефективність автоматизації. Стаття також висвітлює перспективи розвитку систем автоматизації, зокрема інтеграцію штучного інтелекту та Інтернету речей. Ці технології мають потенціал покращити моніторинг та прогнозування, забезпечуючи подальше вдосконалення логістичних процесів. Впровадження таких рішень дозволить поштовим операторам відповідати сучасним викликам, зберігаючи високу якість послуг та конкурентоспроможність.

Ключові слова: ідентифікація об'єктів, поштовий зв'язок, машинний зір, автоматизовані сортувальні системи, оптимізація процесів, поштові відправлення, логістика.

Вступ

Постановка проблеми. Сучасні поштові служби стикаються з масштабними викликами, пов'язаними з обробкою значних обсягів поштової кореспонденції.

Швидке зростання кількості відправлень, зумовлене розвитком електронної комерції та глобалізацією, вимагає від поштових операторів удосконалення технологічних процесів для забезпечення своєчасної доставки та мінімізації помилок. Традиційні методи ручного сортування є ресурсомісткими, схильними до помилок та значно сповільнюють загальну продуктивність [1].

Однією з ключових проблем є ідентифікація об'єктів поштового зв'язку: листів, посилок, бандеролей тощо. Процес ідентифікації включає розпізнавання типу відправлення, його ваги, об'єму, адреси відправника та отримувача. Помилки на етапі ідентифікації спричиняють затримки у доставці, дублювання роботи та підвищення операційних витрат. Зокрема, у пікові періоди, такі як свята чи акційні розпродажі, навантаження на поштові служби зростає в кілька разів, що унеможливає ефективну роботу без використання систем автоматизації [2].

Актуальним рішенням є впровадження автоматизованих систем ідентифікації, що базуються на сучасних технологіях, таких як комп'ютерний зір, RFID-технології, штрих-коди та машинне навчання. Ці системи дозволяють швидко й точно ідентифікувати відправлення, оптимізуючи процеси сортування та доставки [3,4].

Таким чином, постає необхідність дослідження методів ідентифікації об'єктів поштового зв'язку для оптимізації їх обробки. Це дослідження допоможе розробити ефективні рішення для підвищення продуктивності поштових операторів, зниження витрат і поліпшення якості обслуговування клієнтів

Аналіз останніх досліджень і публікацій. Дослідження оптимізації поштових процесів та ідентифікації об'єктів активно розвивається в останні роки завдяки впровадженню новітніх технологій. Вчені та практики фокусуються на розробці кількісних і якісних методів аналізу для підвищення ефективності роботи поштових служб [5, 6].

Зокрема, кількісні методи включають використання математичних моделей, статистичного аналізу та симуляційних моделей для прогнозування обсягів поштових відправлень та оцінки часу їх обробки. Дослідження показують, що автоматизація

сортувальних ліній дозволяє підвищити продуктивність на 30-40% порівняно з ручною обробкою [5].

Якісні методи, такі як інтерв'ю та спостереження, застосовуються для вивчення досвіду працівників поштових служб. Зокрема, було встановлено, що перехід на автоматизовані системи сортування знижує кількість помилок, але вимагає додаткових витрат на навчання персоналу та технічне обслуговування обладнання [1].

Окремо слід відзначити дослідження, що стосуються впровадження RFID-технологій та систем на основі комп'ютерного зору. Такі технології забезпечують швидке розпізнавання об'єктів, їх ідентифікацію за вагою, об'ємом та призначенням. Проте висока вартість впровадження та необхідність модернізації інфраструктури залишаються стримуючими факторами [2].

Незважаючи на успіхи, існує потреба у подальших дослідженнях, які поєднують в собі кількісні методи, сучасні технології ідентифікації та практичну реалізацію для підвищення ефективності обробки поштової кореспонденції [5].

Мета роботи. Метою статті є дослідження методів і технологій ідентифікації об'єктів поштового зв'язку для оптимізації процесів обробки поштової кореспонденції. Зокрема, стаття спрямована на оцінку ефективності автоматизованих систем, що використовують комп'ютерний зір, RFID-технології та штрих-кодування для підвищення швидкості, точності та продуктивності роботи поштових служб.

Виклад основного матеріалу

Дане дослідження спрямоване на аналіз впливу автоматизації на процеси ідентифікації об'єктів поштового зв'язку. Для цього ми розглянули три ключові технології: RFID, системи комп'ютерного зору та штрих-коди. Основні метрики для оцінки включали швидкість обробки, точність ідентифікації та економічну ефективність.

Швидкість обробки є критичним показником для оптимізації поштових процесів. У середньому поштовому центрі, який обробляє 10 000 відправлень на день, традиційне ручне сортування дозволяє обробляти не більше 200 відправлень на годину. Це створює вузькі місця в пікові періоди, такі як святкові дні.

RFID (радіочастотна ідентифікація) є однією з найсучасніших технологій, що використовується для автоматизації процесів у поштовій логістиці. Основний принцип її роботи полягає в тому, що кожен об'єкт, який підлягає ідентифікації, оснащується RFID-міткою. Ця мітка передає дані на приймач через радіохвилі, забезпечуючи швидку й безконтактну передачу інформації. Завдяки цьому RFID виключає необхідність прямого контакту або видимості між міткою та зчитувачем, що є особливо важливим у випадках, коли об'єкти упаковані в непрозорі матеріали або перебувають у складних умовах, наприклад, під час транспортування.

Однією з головних переваг технологій RFID є її масштабованість і здатність зчитувати дані з кількох міток одночасно. Це дозволяє одночасно обробляти великі партії відправлень, що значно прискорює

операції на сортувальних лініях. Крім того, RFID-мітки можуть містити значно більший обсяг інформації порівняно зі штрих-кодами, зокрема додаткові дані про відправлення, такі як умови зберігання чи пріоритет доставки. Водночас, вартість впровадження RFID-технологій залишається доволі високою, що іноді обмежує їх використання в невеликих поштових центрах.

Системи комп'ютерного зору базуються на використанні камер та програмного забезпечення для аналізу отриманих зображень. Ці системи дозволяють ідентифікувати об'єкти за їхніми візуальними характеристиками, включаючи текстову інформацію, графічні символи та зображення. Однією з ключових функцій комп'ютерного зору є здатність розпізнавати навіть пошкоджені або частково стерті маркування, що недоступно для традиційних штрих-кодів.

Незважаючи на це, у технології комп'ютерного зору є свої виклики. Наприклад, її точність залежить від якості зображень, що вимагає належного освітлення та регулярного обслуговування обладнання. Крім того, обробка зображень потребує значних обчислювальних ресурсів, що може уповільнювати процеси в умовах високого навантаження. Проте розвиток штучного інтелекту дозволяє поступово оптимізувати ці системи, знижуючи потребу в людському втручанні.

Штрих-коди залишаються найбільш поширеною технологією завдяки їхній простоті, низькій вартості та доступності обладнання. Вони є ефективним рішенням для невеликих поштових центрів або організацій із середнім навантаженням. Проте штрих-коди мають певні обмеження, включаючи залежність від якості друку та необхідність прямої видимості між сканером і маркуванням. Це робить їх менш ефективними в складних умовах, таких як пил, бруд або механічні пошкодження.

Порівнюючи ці три технології, можна зробити висновок, що кожна з них має свої унікальні переваги.

RFID виділяється високою точністю та швидкістю, але вартість методу є суттєвим бар'єром. Комп'ютерний зір забезпечує гнучкість і широкий спектр можливостей, хоча і потребує значних інвестицій у програмне забезпечення та обладнання. Штрих-коди залишаються найбільш економічно вигідним рішенням, особливо для малих та середніх компаній, однак їх функціональність значно обмежена в складних експлуатаційних умовах.

У комплексному підході до автоматизації поштових процесів часто використовується комбінація цих технологій, що дозволяє максимально ефективно оптимізувати роботу поштових центрів.

Автоматизація процесів значно підвищує швидкість:

- Штрих-кодові системи збільшують швидкість до 400 відправлень на годину.
- RFID забезпечує обробку до 500 відправлень на годину.
- Системи комп'ютерного зору демонструють швидкість 450 відправлень на годину, оскільки додатковий час витрачається на аналіз зображень.

Автоматизація дозволяє зменшити кількість помилок, які виникають через людський фактор або пошкодження маркувань.

Проведений аналіз даних показав, що є такі результати:

- RFID досягає точності 99%, оскільки дані зчитуються безконтактно, навіть у складних умовах.
- Комп'ютерний зір має точність 98%, однак інколи виникають проблеми з нерозбірливими адресами.
- Штрих-коди забезпечують точність 95%, але схильні до помилок через пошкодження або забруднення етикеток.

Переваги автоматизації виходять за межі швидкості та точності.

У дослідженні розраховано економічний ефект від впровадження технологій у поштовому центрі з навантаженням у 10 000 відправлень на день.

Використання технологій RFID дозволяє скоротити витрати на оплату праці на 40%, що відповідає економії близько \$20 000 на місяць.

Системи комп'ютерного зору забезпечують економію близько \$15 000 щомісячно, зменшуючи потребу в ручній перевірці.

Штрих-кодові технології мають найнижчий економічний ефект, але залишаються вигідними завдяки низькій вартості впровадження.

У одному з поштових центрів міста Київ було проведено експеримент із впровадженням RFID. За три місяці продуктивність зросла на 35%, а час обробки одного відправлення скоротився з 2,5 хвилин до 1 хвилини.

Таблиці отриманих результатів (табл. 1, 2) та економічного ефекту, а також графіки порівняння швидкості і точності ідентифікації (рис. 1, 2) наведено для візуалізації результатів:

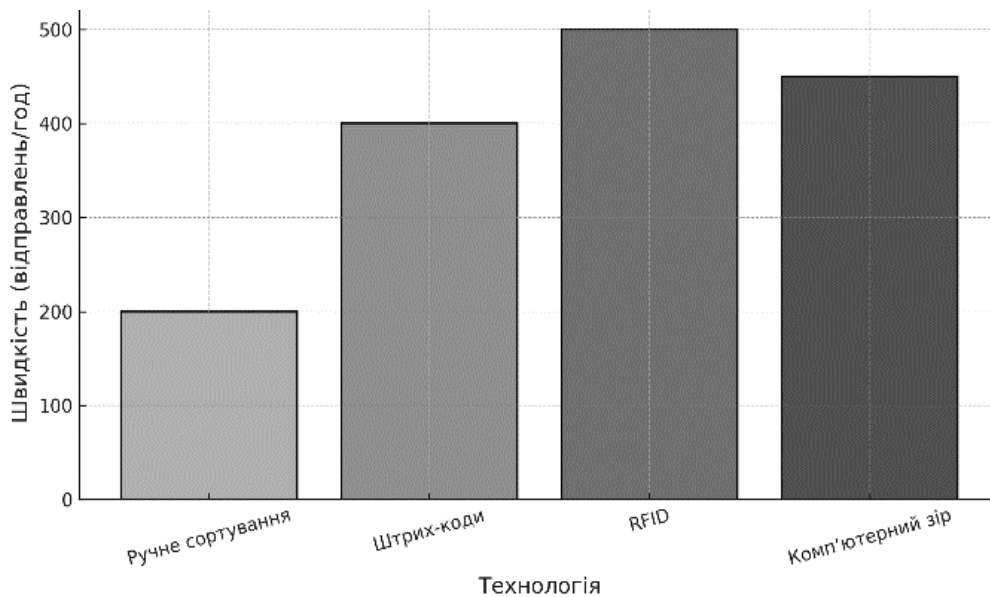


Рис. 1. Порівняння швидкості обробки відправлень за різними технологіями

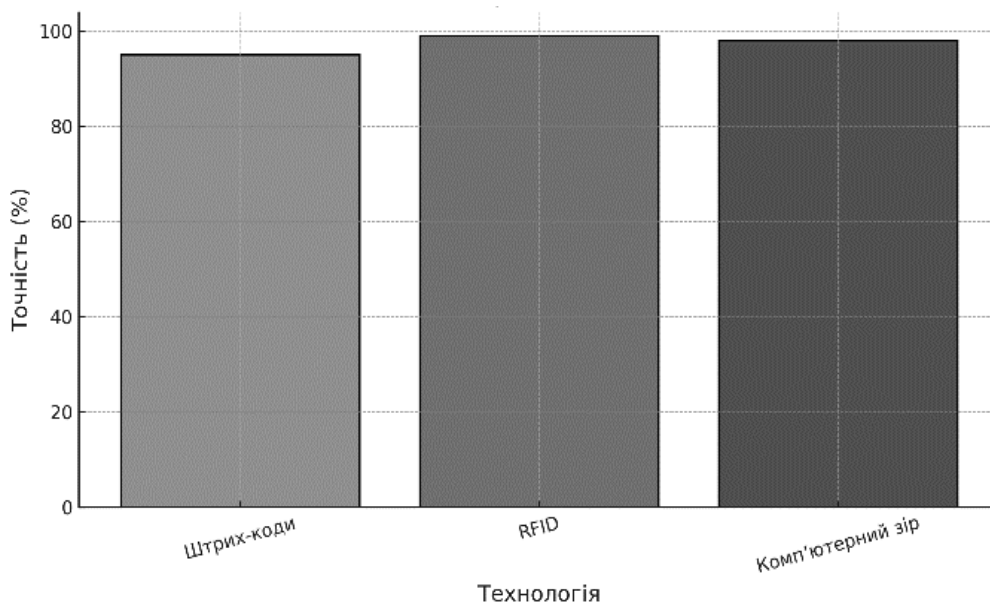


Рис. 2. Порівняння точності ідентифікації відправлень за різними технологіями

Таблиця 1 – Швидкість обробки відправлень

Технологія	Швидкість (відправлень/год)
Ручне сортування	200
Штрих-коди	400
RFID	500
Комп'ютерний зір	450

За рівнем точності технології розташовані наступним чином:

- RFID — 99%;
- Комп'ютерний зір — 98%;
- Штрих-коди — 95%.

Висновки

1. Проведені дослідження показують, що автоматизація ідентифікації об'єктів поштового зв'язку сут

Таблиця 2 – Економічний ефект від автоматизації

Технологія	Зниження витрат (%)	Економія (\$/місяць)
RFID	40	20 000
Комп'ютерний зір	30	15 000
Штрих-коди	20	10 000

тєво підвищує швидкість і точність обробки, забезпечуючи економію витрат та зростання продуктивності.

2. Використання технологій RFID і комп'ютерного зору довели свою ефективність, скоротивши час обробки відправлень удвічі.

3. Подальший розвиток технологій, таких як ШІ та IoT, відкриває нові можливості для оптимізації логістичних процесів і зміцнення конкурентоспроможності поштових служб.

СПИСОК ЛІТЕРАТУРИ

1. Бережний А.І., Петров В.В. Автоматизація логістичних процесів: перспективи і виклики. — Київ: Логістика України, 2021. — 312 с.
2. Коваленко М.С. Технології RFID у логістиці: теоретичні та практичні аспекти. Журнал "Інноваційні рішення", 2020. — №3. — С. 45–51.
3. Smith J., Brown T. Advances in Computer Vision for Postal Services. International Journal of Logistics, 2022. — Vol. 15, Issue 2. — P. 123–137.
4. Захарченко О.М., Іванова Н.В. Використання штрих-кодових технологій у поштової галузі. Вісник транспортних систем, 2019. — №7. — С. 67–72.
5. Макодзєб В. М. Автоматизовані системи поштового зв'язку: поштооброблювальні машини, комплекси обладнання, автоматизовані лінії: Навчальний посібник, 2-ге видання перероблене і доповнене /В.М. Макодзєб – Одеса: ОНАЗ ім. О. С. Попова, 2012. — 304 с.
6. Jones R., Smith A. RFID Technology: Implementation and Challenges in Postal Systems. Logistics Research Journal, 2021. — Vol. 12, Issue 3. — P. 89–101.
7. Мельник О.Л. Аналіз економічної ефективності автоматизованих систем у поштових відправленнях. Економіка і логістика, 2021. — №5. — С. 112–118.
8. Brown M., Wilson T. IoT and Artificial Intelligence in Modern Logistics. Technology & Logistics Journal, 2022. — Vol. 10, Issue 4. — P. 200–215.

Received (Надійшла) 27.12.2024

Accepted for publication (Прийнята до друку) 12.02.2025

Identification of Postal Communication Objects for Optimizing Postal Correspondence Processing

Y. Zourab, N. Yermilova

Abstract. The article examines the pressing issue of postal object identification as a crucial element in optimizing mail processing operations. The increase in shipment volumes, driven by the development of e-commerce and globalization, poses significant challenges for postal services, particularly in ensuring speed, accuracy, and cost-effectiveness of processing. Traditional manual sorting methods are resource-intensive, prone to errors, and fail to meet modern productivity demands, especially during peak periods. The research focuses on the implementation of automated identification systems based on RFID technologies, computer vision systems, and barcodes. Their effectiveness was analyzed according to three key indicators: processing speed, identification accuracy, and cost efficiency. It has been established that automation significantly enhances processing speed: RFID enables processing of up to 500 items per hour, computer vision systems—up to 450 items, and barcode technologies—up to 400 items. This far exceeds the speed of manual sorting, which is only 200 items per hour. In terms of identification accuracy, RFID achieves 99%, making it the most reliable technology, particularly under high workloads. Computer vision systems ensure 98% accuracy, while barcodes, due to their susceptibility to label damage, have an accuracy rate of 95%. The economic evaluation revealed that automation substantially reduces processing costs. The implementation of RFID in a postal center handling 10,000 items per day reduced labor costs by 40%, equivalent to 20,000 in monthly savings. Computer vision systems allow for savings of 15,000 by reducing the need for manual verification. Experimental implementation of RFID in a Kyiv postal center confirmed the practical effectiveness of automation. The article also highlights the prospects for the development of automation systems, including the integration of artificial intelligence and the Internet of Things. These technologies have the potential to enhance monitoring and forecasting, further improving logistics processes. The adoption of such solutions will enable postal operators to address modern challenges while maintaining high service quality and competitiveness.

Keywords: object identification, postal communication, computer vision, automated sorting systems, process optimization, postal shipments, logistics.

Nina Kuchuk, Roman Udyansky, Vladyslav Usichenko, Pavlo Buslov, Artem Huk

Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

MOBILE APPLICATION SECURITY ANALYSIS MODEL BASED ON ARTIFICIAL INTELLIGENCE

Abstract. This article analyzes and studies holographic services, which established that the growth of traffic generated by holographic services will increase several times in the foreseeable future. The main features inherent in the holographic type of communication are identified. A study was conducted of innovative technologies for recording holographic copies, data compression methods to ensure holographic communication, and the transmission of holographic copies to the end user with high-quality reproduction. An analysis of the traffic of multimedia and holographic services, as well as Internet of Things services, was conducted, traffic models were developed, it was determined that the traffic of these services is a mixture of various distributions, and also that the traffic of holographic services has the property of self-similarity.

Keywords: holographic communication, traffic, Internet of Things, self-similarity, bandwidth.

Introduction

Today, the transmission of holographic information to provide the effect of the presence of the interlocutor is of great interest to users and researchers around the world.

In practice, this is confirmed by the emergence of a huge number of diverse and different types of equipment that allows both to form a holographic flow and to reproduce it on the other side in real time [1]. Obviously, this type of communication places high demands on the network bandwidth, which for existing communication networks vary in the range from 10 Mbit/s to 4.2 Tbit/s [2].

Statement of the problem. Trends in the implementation of holographic services and holographic communication today already require a revision of the principles of planning, designing and building existing communication networks, as well as approaches to the implementation of sixth-generation 6G networks, which are based on the integration of various technologies and communication networks into a single network [3, 4].

A separate issue is the assessment of the quality of service and the quality of perception of holographic services by both objective and subjective assessment methods [5].

Analysis of recent research and publications. There are practically no criteria for assessing the quality of a holographic image, including scales and methods for subjective assessment of the quality of holographic services [6, 7].

Moreover, the properties of the holographic flow are poorly studied, and even less so its influence on communication networks and requirements for network parameters, which makes the tasks of studying traffic characteristics and assessing the quality of service of holographic services very relevant. [8]

Purpose of the article. The object of the article is the holographic type of communication, and the subject is the characteristics of traffic and quality of service of the holographic type of communication.

The purpose of the article is to study the models of traffic of the holographic type of communication in networks of the sixth and subsequent generations.

Presentation of the main material

The relevance of the article on the development of holographic technologies is due to the transition to a new stage of the implementation of modern innovative technologies, which involves the active use of holography in the future.

There is a general global trend in the work of specialists with three-dimensional images in various fields of human activity.

Holography, displaying an object in three dimensions, is one of the most promising areas of visualization of three-dimensional objects, which justifies the emergence of a certain scientific interest in this area of research.

Displaying real reality, the hologram exhibits a unique property: a holographic image reproduces an exact three-dimensional copy of the original object, unlike photography, which creates a flat image. A holographic image with a large number of angles that change with the movement of the observation point can be so realistic that it is often indistinguishable from a real object.

In the context of the introduction of modern technologies, the holographic method of recording information has received the opportunity to reveal its potential. The exponential increase in the number of scientific publications on the results of research into the theory of holography, its practical application indicates interest in the topic, its undoubted prospects and relevance.

Theoretical developments in the field of digital holography have made it possible to create devices with unique characteristics - a holographic TV, a digital holographic lensless microscope, holographic glasses, a holographic control system for unmanned aerial vehicles (drones). Holographic network applications are emerging, which is a direct development of the digital stage. Holographic assistants and holographic communication are becoming the most popular among the images of the future for young people.

One of the most promising areas is digital pictorial holography. The effect of presence in real time is achieved when computer programs synthesize color volumetric images of objects.

Considering the trajectory of the development of communication tools in accordance with the increase in the technological level of the communication system, it is possible to graphically represent from a landline telephone to a holographic type of communication

through the stage of using mobile communication tools and sequentially through the stage of video conferencing.

Fig. 1 schematically shows the development of the communication system.

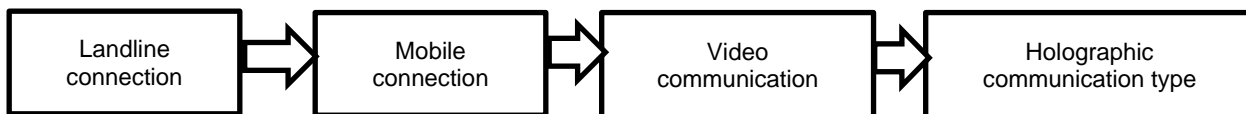


Fig. 1. The development of the communication system

The holographic method of recording information, which underlies holographic technologies and holographic communications, is just beginning to reveal its potential.

One of the most important scientific problems in the implementation of holographic communication is the problem of transmitting holographic copies to the end user with high quality reproduction.

The widespread implementation of holographic copies will depend on how communication networks can support high-quality interactive transmission of video holograms in real time.

Research of 3D video stream traffic servicing processes in various transmission environments. Transmission of various types of traffic is directly related to the study of the problem of the influence of the data transmission environment on the quality of service. Three types of data transmission environment can be distinguished, namely:

- electrical (copper) cables based on twisted pair wires;
- wi-fi;
- fiber optic.

Let's consider the objectively inherent features of the data transmission environment. Let's focus on the three most important of them:

- bandwidth (frequency range of signals) and signal attenuation. These characteristics are directly related to each other. Let's emphasize that with an increase in the signal frequency, its attenuation increases;

- noise immunity of the transmission environment;
- signal propagation speed.

Traffic, as a process of moving data through a network, can also, depending on the environment in which the transmission occurs, be wired and wireless.

Various technologies and protocols are used (IEEE 802.3 (Ethernet), TCP/IP). Wired networks provide high data transfer speeds (from 10 Mbit/s to 100 Gbit/s and higher) and a reliable connection.

Fiber optic cable, twisted pair (Cat5e, Cat6, Cat6a, etc.), and coaxial cable are used to transmit traffic.

Research of 3D video stream traffic transmission using wireless environment (Wi-Fi). The modern trend of expanding the scope of use (Wi-Fi) as a data transmission medium has necessitated the study of 3D video stream traffic and provision of innovative holographic communication services.

Indeed, mobility, ease of deployment, low installation cost, the ability to create public access points and connect Internet of Things (IoT) devices confirm the prospects for further development of this technology.

When studying the transmission of holographic copies in various environments, it is also necessary to study the transmission of holographic content using Wi-Fi.

For this stage of the study, a structural model of a full-scale experiment was developed.

The diagram of the laboratory stand is shown in Fig. 2.

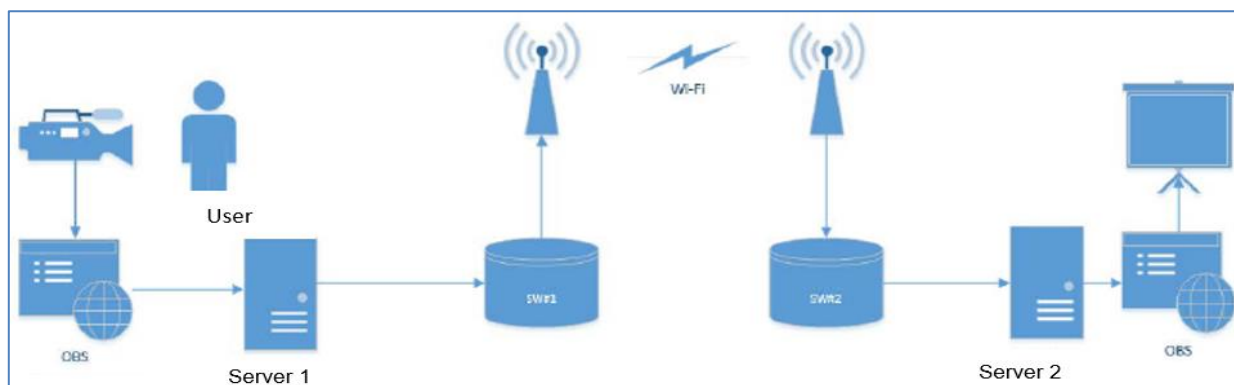


Fig. 2. Laboratory setup diagram

The necessary equipment for the laboratory stand was assembled on a modern technical base. The interaction between the switches was carried out via a wireless channel.

Below are graphic materials of traffic monitoring on the switches. The results of the 3D video stream transmission on the transmitting side are shown in Fig. 3. The packet transmission intensity during the observation

session is shown in Fig. 4. Thus, when using a wireless data transmission medium (Wi-Fi) between switches, the quality of holographic content reproduction may deteriorate.

Therefore, in order to implement this technology into mass use, it is necessary to solve the problem of

increasing the network's resistance to peak loads and the influence of the external environment.

Transmission of 3D video stream traffic using Ethernet technology. To conduct the experiment, a full-scale model of the network was built, the structure of which is shown in Fig. 5.

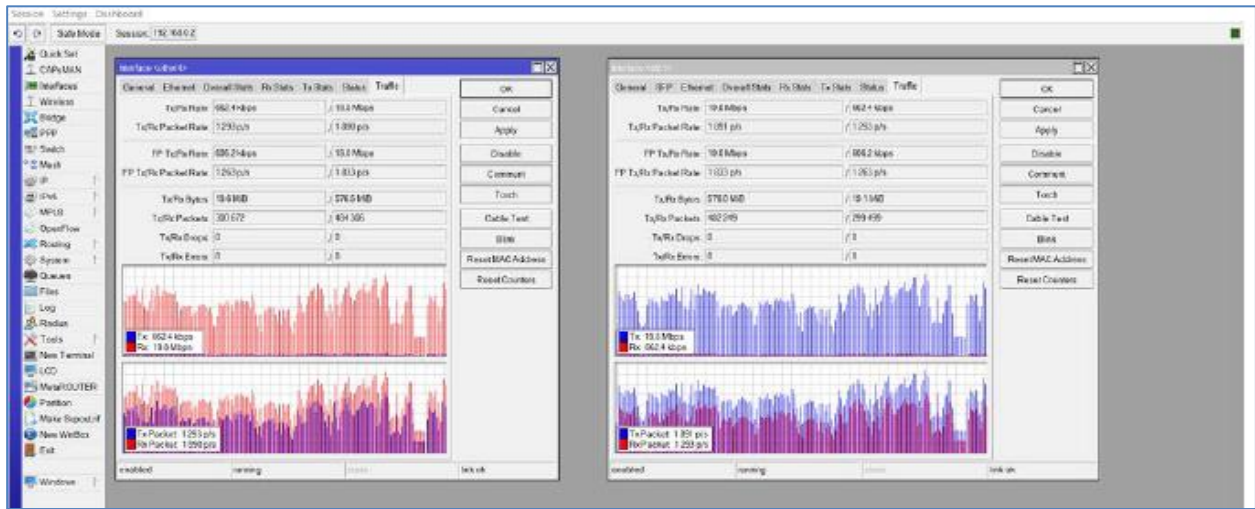


Fig. 3. Results of monitoring the transmission of 3D video stream on switches

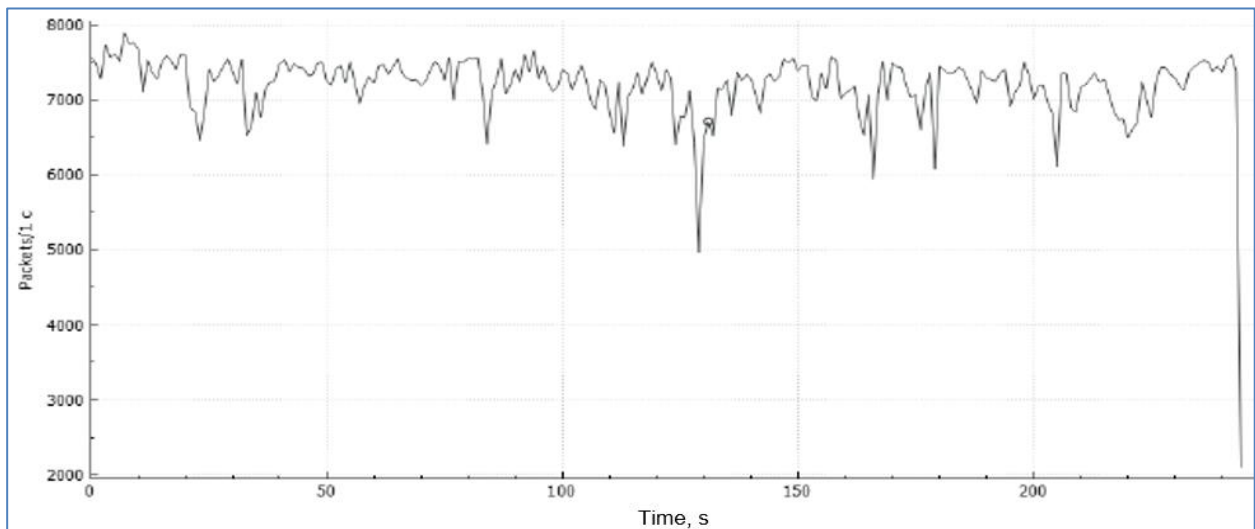


Fig. 4. Packet transmission intensity during the observation session

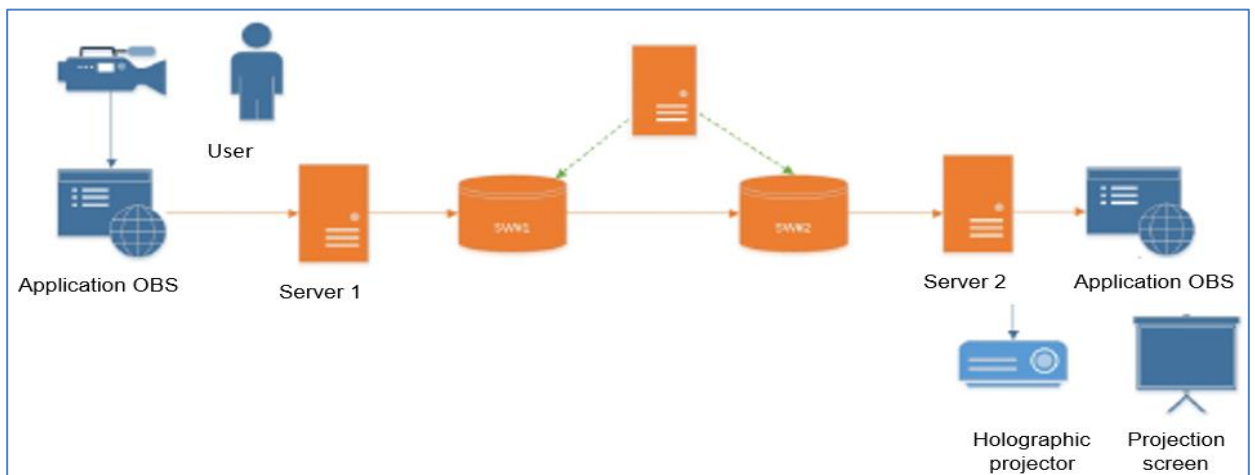


Fig. 5. Model network structure

It includes the following equipment: a Microsoft Kinect camera sensor, which allows recording a moving object in real time, two servers, two Mikrotik switches (SW#1, SW#2) and a holographic laser projector LG with 3D support for reproducing a three-dimensional image.

The data transmission medium between the switches is fiber optics, between the switch and the server - an SC-SC patch cord or a four-pair cat.5e 3.0m UTP patch cord, since twisted pair is used in access networks.

Video traffic was obtained using a Microsoft Kinect 3D camera (Full HD (1920×1080)).

The 3D video stream was generated using a special software tool – OBS Studio (Open Broadcaster

Software). Wireshark was used to intercept and analyze network traffic. Web interfaces of switches were launched on the virtual server to monitor traffic. An SDN controller was implemented to virtualize the network management level.

Fig. 6 shows a graph demonstrating the dynamics of packet intensity changes during an observation session (350 s – total traffic recording time).

The traffic was obtained by recording human movements with a 3D camera and transmitting the resulting video image through the model network to a laser holographic projector for playback.

Fig. 7 shows a histogram illustrating the distribution of packet lengths.

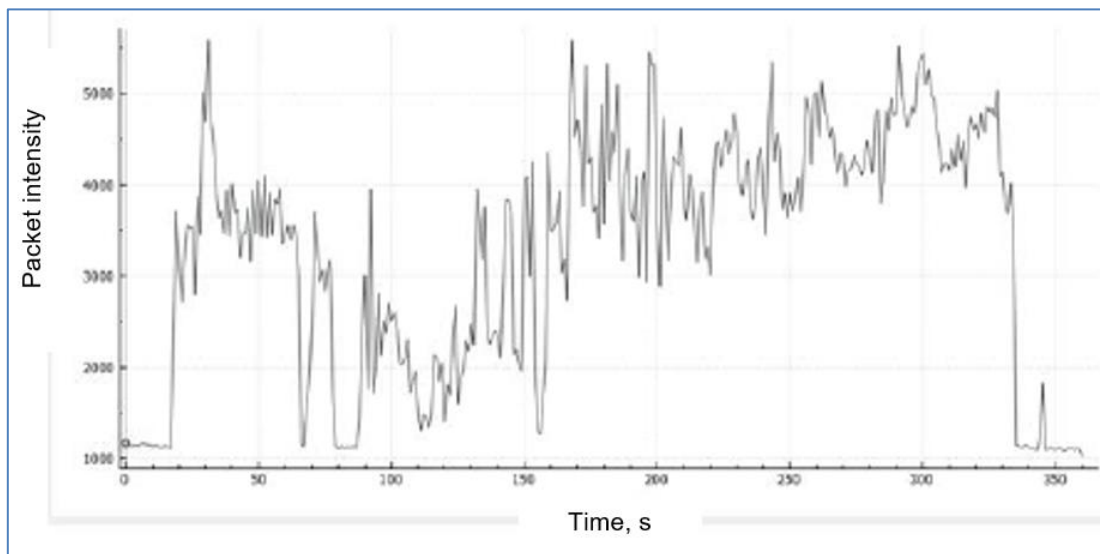


Fig. 6. Packet transmission intensity during the observation session

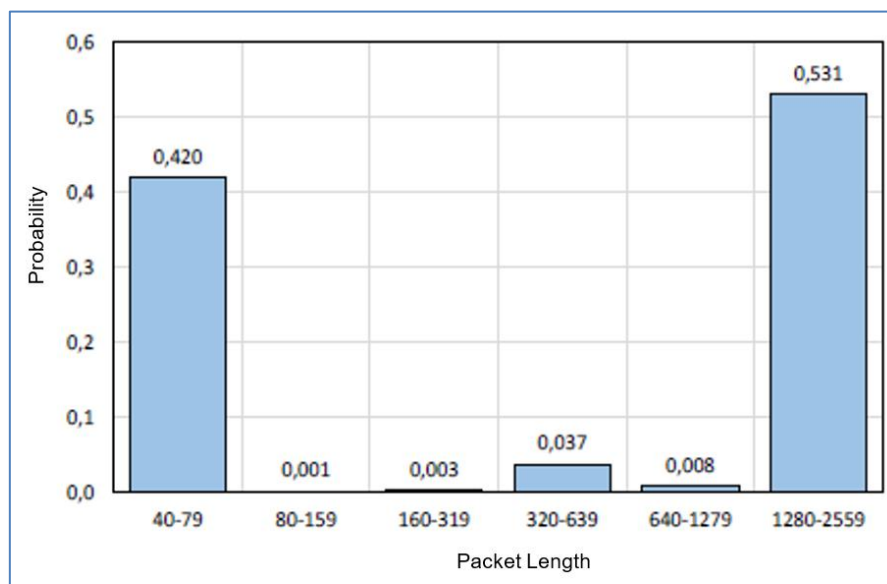


Fig. 7. Packet Length Distribution

The main parameters of the traffic under study are given in Table 1.

The histogram above shows that the largest share of packets (over 95%) are packets from 40 to 80 bytes in size and packets from 1280 to 2560 bytes in size. The

average packet length and the standard deviation from it were 1226 bytes and 544 bytes, respectively. This result is important for describing the model of servicing a flow by a network as a mass service system in terms of choosing a model of packet transmission time.

Table 1 – Numerical values of traffic parameters

Parameter	Value
Packet intensity	3290,7 ± 113 Packet /s
Traffic intensity	32,3 ± 1,1 Mbit /s
Average packet size	1226 ± 1 byte
Total number of observed packets	1151745 number
Observation duration	350 s
Total traffic volume	1412,04 Mbyte

Conclusions

The current trend of exponential increase in traffic volume and, accordingly, an increase in the network load, indicates a serious problem in the field of providing network resources.

The distribution of time intervals between packets has a multimodal distribution and can be described by a mixed distribution.

The obtained traffic characteristics allow us to judge it as a random process, as well as to obtain numerical estimates of its main parameters. The data set is sufficient for choosing models and calculating the main parameters of traffic servicing.

High bandwidth, low latency, connection reliability and standardization of communication networks are

necessary conditions for the successful implementation of holographic communication.

The above results can be used to model the 3D video traffic service in a communication network. The results of the study of the distribution of intervals between packets and packet lengths make it possible to use approximate models from the theory of queuing to describe the time parameters of the functioning of the communication network.

Also, the result of the study was the developed analytical model of 3D video stream traffic, obtained on the basis of the data of a full-scale experiment and the results of statistical processing, which can be used in solving traffic research problems, as well as planning new networks and communication services taking into account the requirements for quality of service.

A promising direction of research may be a more comprehensive study of holographic traffic and problems of transmitting holographic copies.

The need to introduce new methods of visualizing objects based on holographic technologies creates a need to study the properties of a new type of traffic transmitting a 3D video stream in real time. The importance and relevance of the problem is emphasized by a significant number of studies devoted to the study of the characteristics of video traffic, multimedia traffic, 3D image traffic, and traffic modeling in the context of the development of communication networks.

REFERENCES

1. Qin, R., Zhang, C., Li, X., Wang, H. (2024), "Research on Coordinated Control Method of Trunk Line in Holographic Traffic Perception Environment", Wuhan Ligong Daxue Xuebao (Jiaotong Kexue Yu Gongcheng Ban) / *Journal of Wuhan University of Technology*, vol. 48(4), pp. 622–627, doi: <https://doi.org/10.3963/j.issn.2095-3844.2024.04.003>
2. Luevano, L., Lopez de Lara, E. and H. Quintero (2019), "Professor Avatar Holographic Telepresence Model", *Holographic Materials and Applications*, IntechOpen, Sep. 25, 2019. doi: <https://doi.org/10.5772/intechopen.85528>
3. Kuchuk, H. and Malokhvii, E. (2024), "Integration of iot with Cloud, Fog and Edge computing: a review", *Advanced Information Systems*, vol. 8, no. 2, pp. 65–78, doi: <https://doi.org/10.20998/2522-9052.2024.2.08>
4. Kuchuk, N., Kashkevich, S., Radchenko, V., Andrusenko, Y. and Kuchuk, H. (2024), "Applying edge computing in the execution IoT operative transactions", *Advanced Information Systems*, vol. 8, no. 4, pp. 49–59, doi: <https://doi.org/10.20998/2522-9052.2024.4.07>
5. Petrovska I., Kuchuk, H. And Mozhaiev M. (2022), "Features of the distribution of computing resources in cloud systems", *2022 IEEE 3rd KhPI Week on Advanced Technology*, KhPI Week 2022 - Conference Proceedings, 03-07 October 2022, doi: <https://doi.org/10.1109/KhPIWeek57572.2022.9916459>
6. Li, G., Liu, Y., Wu, J., Lin, D. and Zhao, Sh. (2019), "Methods of Resource Scheduling Based on Optimized Fuzzy Clustering in Fog Computing", *Sensors*, vol. 19(9), doi: <https://doi.org/10.3390/s19092122>
7. Kuchuk, N., Kovalenko, A., Ruban, I., Shyshatskyi, A., Zakovorotnyi, O. and Sheviakov, I. (2023), "Traffic Modeling for the Industrial Internet of NanoThings", *2023 IEEE 4th KhPI Week on Advanced Technology*, KhPI Week 2023 - Conference Proc., 194480, doi: <http://dx.doi.org/10.1109/KhPIWeek61412.2023.10312856>
8. Wang, Y., Chen, Y., Li, G., ... Yu, Z., Sun, W. (2023), "City-scale holographic traffic flow data based on vehicular trajectory resampling", *Scientific Data*, vol. 10(1), 57, doi: <https://doi.org/10.1038/s41597-022-01850-0>

Received (Надійшла) 08.01.2025

Accepted for publication (Прийнята до друку) 05.03.2025

Трафік, що забезпечує голографічний тип комунікацій

Н. Г. Кучук, Р. О. Удяньський, В. Д. Усиченко, П. В. Буслов, А. С. Гук

Анотація. У статті проведено аналіз та дослідження голографічних послуг, в ході якого встановлено, що зростання трафіку, що виробляється голографічними послугами, збільшиться в рази в найближчому майбутньому. Позначено основні особливості, властиві голографічному типу комунікації. Проведено дослідження інноваційних технологій запису голографічних копій, методів стиснення даних для забезпечення голографічного зв'язку, передачі голографічних копій кінцевому користувачеві з високою якістю відтворення. Проведено аналіз трафіку мультимедіа та голографічних послуг, а також послуг Інтернету речей, розроблено моделі трафіку, визначено що трафік даних послуг є сумішно різними розподілів, а також, що трафік голографічних послуг має властивість самоподібності.

Ключові слова: голографічний зв'язок, трафік, Інтернет речей, самоподібність, пропускна здатність.

Л. І. Леві, О. Є. Зима

Національний університет «Полтавська політехніка імені Юрія Кондратюка», Полтава, Україна

КОНЦЕПТУАЛЬНА МОДЕЛЬ ІЄРАРХІЧНОЇ ОРГАНІЗАЦІЇ ІНЖЕНЕРНИХ МЕРЕЖНИХ СИСТЕМ

Анотація. Через спільності ряду характерних властивостей та особливостей, якими, зокрема, є функціональне призначення, топологічна структура, моделі потокорозподілів, а також задачі, методи та алгоритми управління, системи сільськогосподарського водопостачання, а також зрошувальні системи можуть бути віднесені до класу ієрархічно організованих інженерних мережних систем (ІМС). На основі аналізу існуючих та проєктованих ІМС можна зробити висновок, що такі об'єкти можуть бути декомпозовані на певним чином взаємозалежні та взаємодіючі між собою в процесі функціонування окремі локальні мережі (ЛМ) різного рівня ієрархії. При цьому реальні ІМС на найвищому рівні ієрархії як ЛМ містять магістральні мережі (ММ), до яких підключені ЛМ нижчого рівня ієрархії, – розподільні мережі (РМ). До РМ передостаннього рівня ієрархії підключені РМ найнижчого рівня ієрархії – мережі споживачів цільового продукту (МСЦП). До них безпосередньо підключені споживачі цільового продукту (СЦП), що здійснюють у підсумку споживання цільового продукту (ЦП), – води, із розглядуваної ієрархічно організованої ІМС [1, 2]. Функціональним призначенням ММ є споживання ЦП із зовнішніх по відношенню до ІМС джерел, його транспортування та перерозподіл між РМ, підключеними до розглядуваної ММ. Функціональним призначенням РМ є споживання ЦП з ММ, до якої підключена дана РМ, його транспортування та перерозподіл між МСЦП, підключеними до цієї РМ. Функціональним призначенням МСЦП є споживання ЦП з РМ, до якої підключена дана МСЦП, його транспортування та перерозподіл між СЦП, підключеними до цієї МСЦП. Внаслідок стохастичних змін режимів споживання ЦП окремими СЦП, для створення припустимих умов їхнього функціонування є необхідною реалізація процесів оперативного управління ІМС.

Ключові слова: інженерні мережні системи, локальні мережі, магістральні мережі, розподільні мережі, мережі споживачів цільового продукту, цільовий продукт, споживачі цільового продукту.

Вступ

Постановка проблеми. Відповідно до [1, 2], розглянемо зміну наступних найважливіших характерних властивостей та особливостей локальних мереж (ЛМ) залежно від рівня ієрархії в інженерній мережній системі (ІМС).

1. Функціональне призначення.
2. Геометричні розміри.
3. Конструктивно-технологічні особливості реалізації.
4. Топологічна структура моделюючих графів.
5. Енергетична характеристика потокорозподілу.
6. Режимі споживання цільового продукту (ЦП).
7. Ступінь стохастичності потокорозподілу.
8. Елементний склад дуг.
9. Можливість практичної реалізації різних видів ідентифікації.

Формулювання мети статті. На підставі [1, 2], що містять матеріали та результати з різних аспектів дослідження ІМС як об'єктів управління, розглянемо найважливіші властивості та особливості, характерні для окремих ЛМ різного рівня ієрархії.

Основна частина

Конструктивно магістральні мережі (ММ) складаються з великих трубопроводів великого перерізу, здатних витримати великі напори та пропускати великі потоки ЦП. ММ мають велику довжину. Моделюючи графи ММ або є деревами, або мають відносно невелику кількість циклів, тому цикломатичні числа таких графів дорівнюють або не набагато перевищують нульове значення. Конструктивно

розподільчі мережі (РМ) складаються з трубопроводів меншого діаметра, розрахованих на передачу менших потоків ЦП з напорами, меншими щодо відповідних величин в ММ.

Реальні РМ характеризуються меншою проти ММ протяжністю. Моделюючи графи РМ містять більшу порівняно з моделюючими графами ММ кількість циклів, тому відповідні цикломатичні числа набувають більших значень, ніж для моделюючих графів ММ.

Конструктивно мережі споживачів ЦП (МСЦП) представляють вельми розгалужені та закільцьовані мережі дрібних трубопроводів низького тиску, що мають невелику довжину. Цикломатичні числа, відповідні моделюючим графам реальних МСЦП, набувають більших значень, ніж для моделюючих графів РМ.

Режими споживання ЦП із вищих ЛМ нижчими ЛМ чи споживачами цільового продукту (СЦП) визначаються величинами потоків ЦП, а також значеннями напорів, за яких споживаються ці потоки. Для реальних СЦП має місце стохастичний характер режимів споживання, обумовлений цілим рядом випадкових чинників, до яких, зокрема, можна віднести погодні умови, біологічні потреби організмів, і навіть суб'єктивні міркування [3, 4].

Відповідно до математичної моделі потокорозподілу в розглядуваній ЛМ визначається режимами споживання ЦП підключених до неї нижчих ЛМ або СЦП. Тому стохастичність режимів споживання СЦП породжує стохастичність потокорозподілів окремих ЛМ різного рівня ієрархії.

Однак у силу тієї ж стохастичності режимів споживання СЦП чи нижчестоящих ЛМ різні випадкові зміни (збурення) таких режимів взаємно компен-

суються, вирівнюються, згладжуються при їхньому впливі на вищі ЛМ. Тому поточкорозподіли у вищих ЛМ та режими споживання ними ЦП є більш детермінованими, ніж в нижчих ЛМ.

Особливістю елементного складу дуг реальних ММ – це наявність в них потужних АЕ та відсутність або мінімальна кількість ПЕ. Доцільність розміщення великих АЕ в дугах ММ обумовлена наявністю в них поточкорозподілів з високими енергетичними параметрами (потоками та напором), що мають висой ступінь детермінованості, що дозволяє застосувати більш потужні, а отже і більш економічні конструкції та типорозміри АЕ, створивши при цьому для них більш сприятливі умови функціонування. Розміщення ж ПЕ в дугах ММ є менш доцільним, оскільки в умовах потужних поточкорозподілів це призводить до істотних енергетичних витрат [1, 2].

У дугах реальних РМ можуть міститися як АЕ, так і ПЕ. Проте відносна кількість АЕ тут менша, ніж у дугах ММ, а потужність їх нижча. Це пояснюється тим, що енергетичні параметри поточкорозподілів в РМ нижчі, ніж у ММ, а ступінь їх стохастичності – вищий. У таких умовах важко створити припустимі умови для функціонування реальних АЕ, особливо для тих, які мають велику потужність. Наявність же тут ПЕ припустима в більшій мірі, ніж в ММ, оскільки відповідні енергетичні параметри поточкорозподілів тут нижчі, а отже нижчі й енергетичні втрати в ПЕ [1, 2].

У дугах же реальних МСЦП внаслідок високої стохастичності та низьких енергетичних параметрів поточкорозподілів містяться лише ПЕ [1, 2].

Таким чином, ієрархічність структури реальних ІМС обумовлена необхідністю узгодження невеликого числа в основному детермінованих процесів споживання ЦП із зовнішнього середовища, що мають високі енергетичні параметри, з дуже великим числом стохастичних процесів його споживання СЦП, що характеризуються низькими енергетичними параметрами.

Для реалізації процесів управління ІМС велике значення має можливість їх структурної та параметричної ідентифікації, а також ідентифікації стану таких об'єктів [1, 2]. На підставі розглянутих властивостей ММ, РМ та МСЦП визначимо можливість практичної реалізації різних видів ідентифікації цих ЛМ. Зокрема, ММ характеризуються відносно невеликим числом відносно стійких змінних і параметрів, взаємозв'язаних між собою у відносно меншому ступені, ніж аналогічні величини в РМ та МСЦП, та в достатньому ступені доступних для виміру. Тому можна зробити висновок, що для ММ зазначені види ідентифікації є практично реалізованими [1, 2].

РС характеризуються значно великим числом змінних та параметрів, стохастичність яких істотно вища, ніж для аналогічних величин в ММ. Крім цього, зазначені величини в РМ взаємозв'язані між собою в більшому ступені, ніж у ММ, та у меншому ступені доступні для виміру.

Тому можна зробити висновок, що в РМ практична реалізація розглянутих видів ідентифікації

зустрічає певні труднощі [1, 2]. У зв'язку з цим припуститимемо, що в РМ можуть бути ідентифіковані лише істотні структурні взаємозв'язки, а також найважливіші змінні та параметри.

МСЦП характеризуються дуже великим числом змінних та параметрів, ступінь стохастичності яких занадто високий.

В силу структурно-топологічних особливостей моделюючих графів та конструктивно-технологічних особливостей реалізації МСЦП, ці величини дуже сильно взаємозв'язані між собою та важкодоступні для виміру.

Тому можна зробити висновок, що вказані види ідентифікації в МСЦП реалізувати практично неможливо [1, 2].

Як приклад, що конкретизує основні положення даної роботи, розглянемо мережну систему сільськогосподарського водопостачання, що складається з ММ, до якої підключені РМ [1, 2].

ММ, що входить в розглядувану систему, характеризується наступними основними властивостями та параметрами. Її довжина знаходиться в межах від декількох сотень метрів до декількох кілометрів. За своєю конфігурацією ММ довжиною до 200 м є тупиковими, а ММ більшої довжини мають декілька циклів.

Ділянками МС являються сталеві, чавунні або азбестоцементні труби діаметром від 100 мм до 500 мм В якості АЕ в розглядуваній ММ використовуються відцентрові насоси типу К та КМ з подачею від 0,45 м³/год до 360 м³/год та напорами від 8,8 м до 98 м, а також багатоступінчасті секційні насоси типу ММ з подачею від 22 м³/год до 400 м³/год та напорами від 30 м до 1230 м. Розглядувана ММ в якості ПЕ містить декілька засувок, що знаходяться на ділянках, інцидентних АЕ. Добова нерівномірність процесу водоспоживання з ММ, що має певну тенденцію, представлена відповідним погодинним графіком на рис. 1. Він має два максимуми, що досягають 10 м³/год, а також мінімум, який відповідає 1 м³/год.

Ідентифікація розглядуваної МС як об'єкту управління здійснюється за допомогою вимірювальних перетворювачів тисків та потоків. Як перетворювачі тисків використовуються перетворювачі, що мають похибку виміру 1,5% та час встановлення сигналу від 0,5 с до 2,5 с.

Вимір потоків в розглядуваній ММ реалізується за допомогою електромагнітного перетворювача потоків, похибка виміру якого складає 1%, а типовий час встановлення вихідного сигналу – від 0,6 с до 2,3 с.

Кожна з РМ, підключених до розглядуваної ММ, характеризується наступними основними властивостями та параметрами.

Довжина розглядуваної РМ знаходиться в межах від декількох десятків до декількох сотень метрів. Ділянками РМ є сталеві труби діаметром від 50 мм до 100 мм.

Дана РМ не містить АЕ, у якості ПЕ використовуються вентилі, що знаходяться на ділянках, інцидентних як ММ, так і СЦП.

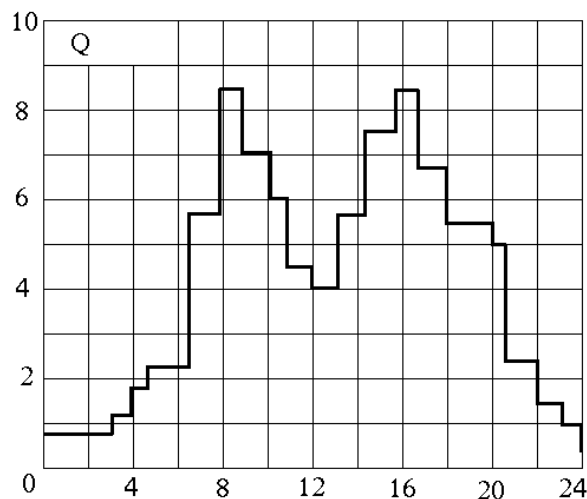


Рис. 1. Погодинний графік водоспоживання з ММ системи сільськогосподарського водопостачання.

При цьому швидкості руху води на ділянках РМ складають від 0,7 м/с до 1,4 м/с, а напори можуть досягати 60 м.

Через складність процесів водоспоживання з РМ, обумовленої, зокрема, суб'єктивними чинниками, пов'язаними із споживачами цільового продукту, їхнє статистичне моделювання у вигляді відповідних графіків не робиться.

Також не робиться ідентифікація даних РМ як об'єктів управління.

Висновки

Зазначені найважливіші характерні властивості та особливості окремих ЛМ різного рівня ієрархії можуть бути використані для постановки відповідних локальних задач (ЛЗ) та вибору адекватних локальних алгоритмів (ЛА) їхньої реалізації.

Зокрема, отримані результати використані при синтезі керуючих алгоритмів ММ зрошувальних систем.

СПИСОК ЛІТЕРАТУРИ

1. Леві Л.І. Оперативне управління інженерними мережними системами: Монографія. – Луганськ: Вид-во СНУ, 2001. – 176 с.
2. Леві Л.І. Інтелектуальні інформаційні технології в ідентифікації і керуванні складними технічними об'єктами в умовах невизначеності: [монографія]. – Полтава: Національний університет «Полтавська політехніка імені Юрія Кондратюка», 2021. – 194 с.
3. Айрапетян Т.С. Конспект лекцій з дисципліни «Міські інженерні мережі» (для студентів 1–3 курсів денної та заочної форм навчання спец. 192 – Будівництво та цивільна інженерія). – Харків: ХНУМГ ім. О.М. Бекетова, 2017. – 97 с.
4. Автоматизація проектування та розрахунків водогосподарсько-меліоративних об'єктів: навч. посібник / за ред. проф. А.М. Рокочинського. [Електронне видання]. – Рівне: НУВГП, 2020. – 257 с.

Received (Надійшла) 27.12.2024

Accepted for publication (Прийнята до друку) 26.02.2025

Conceptual model of hierarchical organization of engineering network systems

L. Lievi, O. Zyma

Abstract. Due to the commonality of a number of characteristic properties and features, which, in particular, are the functional purpose, topological structure, flow distribution models, as well as management tasks, methods and algorithms, agricultural water supply systems, as well as irrigation systems, can be attributed to the class of hierarchically organized engineering network systems (ENS). Based on the analysis of existing and designed ENS, it can be concluded that such objects can be decomposed into separate local networks (LN) of different hierarchy levels, which are interdependent and interact with each other in the process of functioning. At the same time, real ENS at the highest level of the hierarchy as LN contain backbone networks (BN), to which LNs of lower hierarchy levels are connected - distribution networks (DN). The LNs of the lowest hierarchy level - networks of consumers of the target product (NCTP) are connected to the LN of the penultimate hierarchy level. They are directly connected to the consumers of the target product (CTP), which ultimately consume the target product (TP), water, from the considered hierarchically organized ENS [1, 2]. The functional purpose of the BN is the consumption of TP from sources external to the ENS, its transportation and redistribution between the DNs connected to the considered BN. The functional purpose of the DN is the consumption of TP from the BN to which this DN is connected, its transportation and redistribution between the NCTPs connected to this DN. The functional purpose of the NCTP is the consumption of TP from the DN to which this NCTP is connected, its transportation and redistribution between the CTPs connected to this NCTP. Due to stochastic changes in the modes of TP consumption by individual NCTP s, the implementation of the operational management processes of the ENS is necessary to create acceptable conditions for their functioning.

Keywords: engineering network systems, local networks, backbone networks, distribution networks, networks of consumers of the target product, target product, consumers of the target product.

В. М. Почерняєв¹, М. С. Магомедова², Н. М. Сивкова¹

¹ Національна академія Служби безпеки України, Київ, Україна

² Київський фаховий коледж зв'язку, Київ, Україна

ПІДВИЩЕННЯ ЗАВАДОСТІЙКОСТІ МОБІЛЬНИХ ЦИФРОВИХ КОМБІНОВАНИХ РАДІОТЕХНІЧНИХ СИСТЕМ НВЧ ДІАПАЗОНУ

Анотація. У статті оцінюється завадостійкість спільної системи псевдовипадкової перебудови робочої частоти – ширококутний сигнал по відношенню до використання тільки псевдовипадкової перебудови робочої частоти та окремо ширококутний сигнал. Відомо, що у мобільних комбінованих радіотехнічних системах використовуються як фазоманіпульовані ширококутні сигнали, так і псевдовипадкова перебудова робочої частоти. У мобільній цифровій комбінованій тропосферно–космічній системі також використовуються фазоманіпульовані ширококутні сигнали та псевдовипадкова перебудова робочої частоти. Застосування спільної системи псевдовипадкової перебудови робочої частоти – ширококутних сигналів обумовлена вимогою стійкого функціонування будь-якої радіотехнічної системи за умов складної перешкодової обстановки, створюваної станціями радіопридушення супротивника. У статті проведено аналіз відповідних літературних джерел. Наведено схеми передавача та приймача щодо формування та прийому спільної псевдовипадкової перебудови робочої частоти – ширококутний сигнал. Особливістю даних схем є те, що передавальний та приймальний НВЧ тракту реалізуються на частково заповнених діелектриком прямокутних хвилеводах. В роботі досліджено вплив загороджувальної шумової перешкоди на стійкість до перешкод такого приймача. Показано можливість помилки на біт для QPSK модуляції в умовах впливу білого гаусівського шуму і загороджувальної перешкоди. Вираз для ймовірності помилки на біт записано через функцію Крампа. Ймовірність помилки на біт знайдена за умов, що спектр перешкоди перевищує спектр ширококутний сигнал і за умови, що спектр перешкод більш вузький ніж спектр ширококутний сигнал. У роботі розраховані залежності ймовірності помилки на біт від відношення потужності сигналу до сумарної потужності шуму і перешкоди при різних значеннях коефіцієнтів розширення спектрів для випадків застосування тільки псевдовипадкової перебудови робочої частоти, ширококутний сигнал і спільної псевдовипадкової перебудови робочої частоти - ширококутний сигнал. Зроблено висновки, що з урахуванням наявності пристрою захисту в приймачі ширококутний сигнал в умовах впливу перешкоди в частині смуги перешкодостійкість спільної системи вища, ніж перешкодостійкість системи псевдовипадкової перебудови робочої частоти або системи застосування окремо ширококутний сигнал.

Ключові слова: псевдовипадкова перебудова робочої частоти; фазоманіпульований сигнал; ширококутний сигнал; система із розширенням спектру.

Вступ

Мобільна цифрова комбінована тропосферно-радіорелейна система складається з тропосферної та радіорелейної компонент. Розроблено два варіанти таких мобільних комбінованих радіотехнічних систем [1, 2]. Ці наземні мобільні комбіновані радіотехнічні системи мають загальний тракт частото формування, що досягається за рахунок схематехнічних можливостей збудника-гетеродина та введення електронної комутації в передавальний НВЧ тракт. У цих системах використовуються як фазоманіпульовані (ФМ) ширококутні сигнали (ШСС), так і псевдовипадкова перебудова робочої частоти (ППРЧ). У мобільній цифровій комбінованій тропосферно-космічній системі [3] також використовуються ФМ ШСС та ППРЧ.

У різних мобільних радіотехнічних системах використовуються системи з розширенням спектра, наприклад, як ППРЧ так і ШСС. З погляду завадостійкості застосування розширення спектра як методом ППРЧ, так і методом ШСС, явних переваг щодо одне одного немає. Спільну систему ППРЧ-ШСС можна розглядати як ШСС на кожному стрибку частоти, що еквівалентно ШСС з несучою, що періодично змінюється за псевдовипадковим законом. Бажаність застосування спільної системи ППРЧ – ШСС обумовлена вимогою стійкого функціонування будь-якої радіотехнічної системи в умовах складної завадової обстановки, створюваної станціями радіопо-

давлення супротивника. Спільна система ППРЧ – ШСС за рахунок стрибків частоти дозволяє частину часу не потрапляти в ту частину спектра, яка піддається дії перешкод. Але, якщо все ж таки стрибок частоти потрапив на перешкоду, то така перешкода може бути ослаблена шляхом розширення спектру з подальшим стисненням і фільтрацією у більш вузькій смузі, як це відбувається в системах зв'язку з ШСС. Однак, у цьому випадку смуговий фільтр, щоб пропустити ШСС, що залишився після зняття стрибків частоти, повинен мати більшу ширину смуги, ніж тільки для однієї ППРЧ.

Аналіз літературних джерел. Системи з розширенням спектру використовують у радіотехнічних системах різного призначення. Розширення спектру використовується для ширококутний зв'язку між підводними човнами [4], в мережах безпілотних літальних апаратів [5,6], в радарній системі [7], в системах зв'язку з просторовим рознесенням сигналів [8], у розрахованих на багато користувачів системах зв'язку [9], у безпроводних сенсорних мережах [10]. Існують різні математичні моделі для систем із розширенням спектру [11, 12]. Ці системи успішно застосовуються для боротьби зі структурними та хаотичними перешкодами [13-16]. Зазначимо, що у роботах [17, 18] описані й інші різновиди систем із розширенням спектру.

Мета роботи – оцінити завадостійкість спільної системи ППРЧ – ШСС.

Основна частина

Структурна схема передавача спільної системи ППРЧ – ШСС показано на рис. 1. У передавачі використовується генератор псевдовипадкової послідовності (ПВП), як для генерування розширюючого спектру послідовності $c(t)$, так і для формування послі-

довності частот ППРЧ $f(t)$. Стрибок частоти відбувається періодично після надходження фіксованого числа елементів ПВП. У приймачі після зняття стрибків частоти і послідовності, що розширює спектр, залишається несуча, модульована повідомленням. У спільних системах може використовуватися як когерентна, так і некогерентна модуляція.

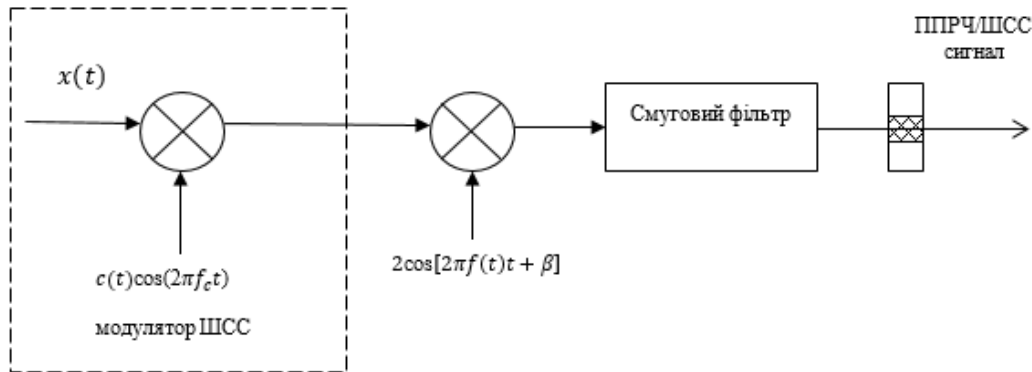


Рис. 1. Передавач спільного ППРЧ/ШСС сигналу

Наприклад, для квадратурної фазової модуляції QPSK сигнал на виході ШСС модулятора передавача можна записати наступним чином:

$$y_m(t) = \sqrt{2P}x(t)c(t) \cos(2\pi f_c t), \quad (1)$$

де P – середня потужність сигналу; $c(t)$ – функціональна залежність, що розширює спектр ПВП; $x(t)$ – вхідна послідовність даних; f_c – частота несучої. Зазначимо, що кожен біт даних модулюється K_c елементами ПВП: $K_c = \frac{T_b}{\tau_c}$, де T_b та τ_c – тривалості

інформаційних бітів та елементів ПВП відповідно. Після перетворення в змішувачі та проходження через смуговий фільтр випромінюваний спільний сигнал запишемо у вигляді:

$$y_c(t) = \sqrt{2P}x(t)c(t) \cos[2\pi(f_c + f(t))t + \varphi_0], \quad (2)$$

де φ_0 – початкова фаза опорного сигналу змішувача.

Структурна схема приймача спільної системи ППРЧ-ШСС показано на рис.2, де СФ - смуговий фільтр; ВП - вирішальний пристрій.

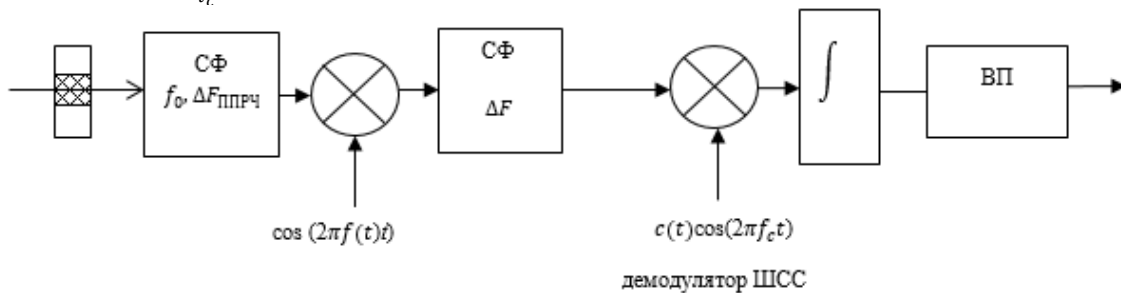


Рис. 2. Приймач спільного ППРЧ/ШСС сигналу

У приймачі відбувається зняття стрибків частоти та перехід на проміжну частоту. Смуговий фільтр зі смугою пропускання ΔF відфільтровує перешкоди і пропускає ШСС з такою самою смугою частот без перешкод. Пройшовши через смуговий фільтр сигнал множиться на ПВП і інтегрується, а вирішальний пристрій (ВП) визначає, який був переданий біт.

На рис.1 і рис.2 показано, що передавальний та приймальний НВЧ тракти реалізуються на частково заповнених діелектриком прямокутних хвилеводах (ЧЗДПХ). Для мобільних цифрових комбінованих радіотехнічних систем розроблено набір пристроїв та елементів на ЧЗДПХ [19-30].

Розглянемо вплив загороджувальної шумової перешкоди на завадостійкість приймача на рис. 2.

У цьому випадку у всій області допустимих частот впливають адитивно білий гаусівський шум зі

спектральною щільністю потужності N_0 і загороджувальна перешкода із спектральною щільністю потужності N_3 . Ймовірність помилки на біт для QPSK модуляції в умовах впливу білого гаусівського шуму та загороджувальної перешкоди при зазначених умовах визначається за формулою:

$$P_{E_b} = \Phi\left(\sqrt{2E_b/N_E}\right), \quad (3)$$

де E_b – енергія біт; $\Phi(\alpha)$ – функція Крампа; $N_E = N_0 + N_3$. Функцію Крампа запишемо у вигляді:

$$(\alpha) = \sqrt{2/\pi} \int_0^\alpha e^{-t^2/2} dt, \quad (4)$$

Як відомо, можна використовувати інтеграл ймовірності або Q-функцію, які від функції Крампа будуть відрізнятися коефіцієнтом перед інтегралом.

Перетворюємо аргумент функції Крампа:

$$\frac{2E_b}{N_0 + N_3} = \frac{2P_c k_c}{P_T + P_3}, \quad (5)$$

де P_T і P_3 – потужності теплового шуму та загороджувальної перешкоди відповідно; $k_c = K_\Delta K_C$. Тоді формула (3) з урахуванням формули (5) набуває вигляду:

$$P_{Eb} = \Phi\left(\sqrt{2P_c k_c / (P_T + P_3)}\right). \quad (6)$$

З формули (6) випливає, що в умовах впливу загороджувальної шумової перешкоди у всій виділеній смузі частот стійкість до перешкод спільної системи ППРЧ-ШСС еквівалентна завадостійкості системи ШСС з коефіцієнтом розширення спектра, рівним добутку коефіцієнта розширення за рахунок ППРЧ і ШСС, або системи ППРЧ з коефіцієнтом розширення спектра, що дорівнює k_c .

Ймовірність помилки на біт може бути знайдена в такий спосіб за умови, що спектр перешкод перевищує спектр ШСС. Припускаємо, що: H_0 – біт не уражений завадою; H_{01} – біт уражений завадою частково; H_{02} – біт уражений завадою повністю. Ймовірність помилки на біт може бути знайдена за формулою повної ймовірності:

$$P_{Eb} = P(E/H_0)P(H_0) + P(E/H_{01})P(H_{01}) + P(E/H_{02})P(H_{02}), \quad (7)$$

де ці ймовірності визначаються як геометричні при рівномірному розподілі положення перешкоди в області допустимих частот.

Умовні можливості помилок визначаються з урахуванням формули (6). Для випадку, коли спектр

перешкод у частині смуги повністю перекриває спектр ШСС, маємо:

$$P(E/H_0) = \Phi\left(\sqrt{2P_c k_c / P_T}\right), \quad (8a)$$

$$P(E/H_{01}) = \Phi\left(\sqrt{\frac{2E_b \Delta F_{\text{ППРЧ}}}{(N_0 + h\tau_c P_3) \Delta F_{\text{ППРЧ}}}}\right), \quad (8b)$$

$$P(E/H_{02}) = \Phi\left(\sqrt{\frac{2E_b \Delta F_{\text{ППРЧ}}}{(N_0 + N_3) \Delta F_{\text{ППРЧ}}}}\right), \quad (8c)$$

де $\Delta F_{\text{ППРЧ}}$ – область частот, виділена для ППРЧ; h – параметр перешкоди. Після підстановки формул (8a) - (8c) у формулу (7) отримаємо вираз для ймовірності помилки на біт у випадку спільного ППРЧ/ШСС сигналу при впливі перешкоди в частині смуги частот, що перевищує спектр ШСС:

$$P_{Eb} = \Phi\left(\sqrt{2P_c K_C / (P_T + P_3)}\right), \quad (9)$$

Формула (9) враховує, що $\Delta F_3 > \Delta F_{\text{ШСС}}$, $\Delta F_{\text{ППРЧ}} \gg \Delta F$, $\Delta F = \Delta F_{\text{ШСС}}$.

На рис. 3, а наведено залежність ймовірності помилки на біт від відношення $P_c / (P_T + P_3)$ при $K_\Delta = 16$, $K_C = 16$, $k_c = 256$. На рис. 3б показані аналогічні криві при значеннях $K_\Delta = 4$, $K_C = 64$, $k_c = 256$. Як випливає з рис. 3, а і рис. 3, б завадостійкість спільної системи ППРЧ – ШСС при $\Delta F_3 > \Delta F$ перевищує і стійкість до перешкод ППРЧ і перешкодостійкість ШСС при відповідних значеннях K_Δ та K_C .

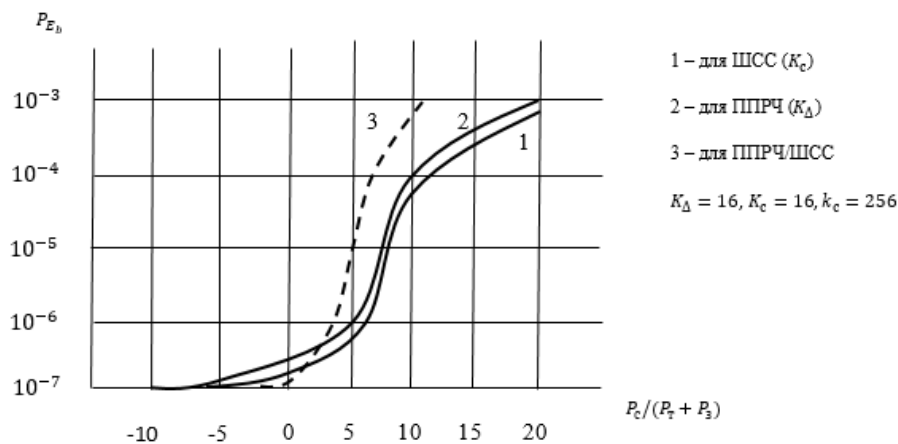


Рис. 3, а. Залежність ймовірності помилки на біт від відношення $P_c / (P_T + P_3)$ при $K_\Delta = K_C = 16$, $k_c = 256$

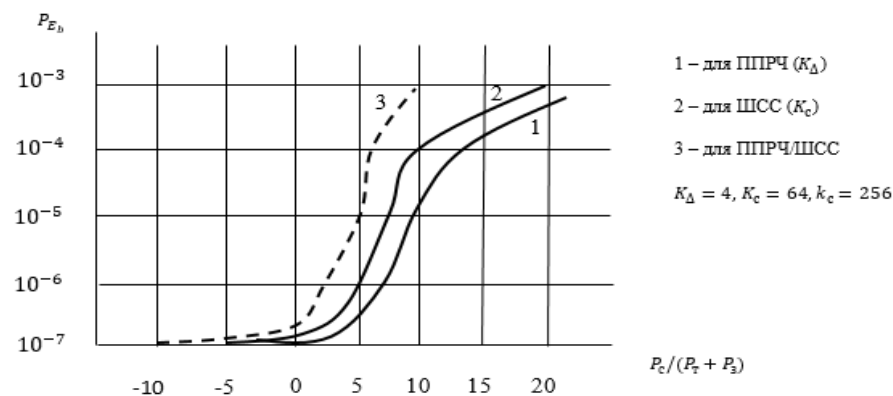


Рис. 3, б. Залежність ймовірності помилки на біт від відношення $P_c / (P_T + P_3)$ при $K_\Delta = 4$, $K_C = 64$, $k_c = 256$

Зазначимо, що стійкість до перешкод сигналу ППРЧ/ШСС збільшується з розширенням спектра за рахунок ШСС.

Коли спектр завади більш вузький ніж спектр ШСС, то в цій ситуації спектр перешкод або частково перекриває спектр сигналу, або взагалі не перекриває. Умовні ймовірності помилки визначаються з урахуванням формули (7), і середня ймовірність помилки набуває вигляду:

$$P_{E_b} = P(H_0)\Phi\left(\sqrt{2P_c K_\Delta K_c / P_T}\right) + P(H_1)\Phi\left(\sqrt{2P_c K_c / K_\Delta / (P_T + hP_3 K_\Delta)}\right). \quad (10)$$

При розширенні спектра за рахунок ШСС ймовірність ураження біта збільшується, і середня ймовірність помилки спільної системи ППРЧ – ШСС при потужній перешкоді може перевершити ймовірність помилки ППРЧ.

У цьому випадку вплив вузькосмугової перешкоди істотно може бути ослаблений шляхом загороження за допомогою загороджувального фільтру.

Висновки

Практично у всіх компонентах мобільних цифрових комбінованих радіотехнічних систем застосовується сигнал ППРЧ/ШСС. Розглянуто варіанти реалізації спільних ППРЧ – ШСС приймачів, коли на кожному стрибку частоти відбувається розширення спектра за допомогою ШСС. При дії загороджувальної шумової перешкоди у всій смузі частот перешкодостійкість спільної системи трохи вище ППРЧ або ШСС систем. При дії широко поширеної на практиці перешкоди в частині смуги з шириною спектра, що перевищує ширину спектра ШСС, перешкодостійкість спільної системи вище, ніж перешкодостійкість ППРЧ або ШСС з мультиплікативним коефіцієнтом розширення спектра. Таким чином, можна зробити висновок, що з урахуванням наявності пристрою захисту в приймачі ШСС в умовах впливу перешкоди в частині смуги стійкість до перешкод спільної системи вище, ніж перешкодостійкість системи ППРЧ або системи ШСС.

СПИСОК ЛІТЕРАТУРИ

1. Патент №112217 Україна. С2. Мобільна цифрова тропосферно-радіорелейна станція / Почерняєв В.М., Повхліб В.С.; заявл. 12.09.2014; опубл. 10.08.2016 // Бюл.№ 15. <https://iprop-ua.com/?qi=112217>
2. Патент №120288 Україна. Мобільна цифрова тропосферно-радіорелейна станція / Почерняєв В.М., Повхліб В.С., Зайченко В.В.; заявл. 29.08.2017; опубл. 11.11.2019// Бюл.№21. <https://iprop-ua.com/?qi=120288>
3. Патент №126203 Україна. С2. Мобільна цифрова тропосферно-космічна станція/ Почерняєв В.М., Повхліб В.С.; заявл. 17.01.2020; опубл. 10.08.2020 <https://iprop-ua.com/?qi=126203>
4. Yaohui Hu, Shuping Han, Heng Zhao, Yubo Han, Jingfeng Xu, Gang Yang. Extension of the Covert Performance Evaluation System for Underwater Spread Spectrum Communication//2023 3rd International Conference on Electronic Information Engineering and Computer (EIECT), 17-19 November 2023, Shenzhen, China, <https://doi.org/10.1109/EIECT60552.2023.10442269>
5. Xiaopeng Tan, Shaojing Su, Xiaoyong Sun. Research on Narrowband Interference Suppression Technology of UAV Network Based on Spread Spectrum Communication//2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIS), 20-22 March 2020, Dalian, China, <https://doi.org/10.1109/ICAIS49377.2020.9194891>
6. D. S. Chirov, O. G. Chertova, E. M. Lobov, M. V. Bazylev. Construction of a Communication Channel with UAVs Based on Direct Sequence Spread Spectrum Signals//2024 Systems of Signals Generating and Processing in the Field of on Board Communications, 16 April 2024, <https://doi.org/10.1109/IEEECONF60226.2024.10496771>
7. Adrian Tang, Yanghyo Kim, Gabriel Virbila, M.-C. Frank Chang. A 205 GHz Serial Direct-Sequence Spread Spectrum (DS/SS) Radar System-on-Chip in 28nm CMOS//2019 IEEE MTT-S International Microwave Symposium (IMS), 25 July 2019, Boston, MA, USA, <https://doi.org/10.1109/MWSYM.2019.8700823>
8. Tarak Arbi, Benoit Geller, Oudomsack Pierre Pasquero. Direct-Sequence Spread Spectrum with Signal Space Diversity for High Resistance to Jamming//MILCOM 2021 - 2021 IEEE Military Communications Conference (MILCOM), 30 December 2021, San Diego, CA, USA, <https://doi.org/10.1109/MILCOM52596.2021.9652967>
9. Kheriya. F. Alhaddar, Hager Bishi, Somya. H. Alshepane, Mohamed M. Elfituri. Direct Sequence Spread Spectrum Technique for Multi-User Communication System Application//2021 IEEE 1st International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering MI-STA, 29 June 2021, Tripoli, Libya, <https://doi.org/10.1109/MI-STA52233.2021.9464450>
10. Daniel Alexandru Visan, Mariana Jurian, Ioan Lita, Laurentiu Mihai Ionescu, Alin Gheorghita Mazare. Direct Sequence Spread Spectrum Communication Module for Efficient Wireless Sensor Networks//2019 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), 23 April 2020, Pitesti, Romania, <https://doi.org/10.1109/ECAI46879.2019.9041979>
11. Fengyuan Zhu, Juhong Tong. Simulation of Spread Spectrum Communication System Based on MATLAB//2024 4th International Conference on Neural Networks, Information and Communication Engineering (NNICE), 19-21 January 2024, Guangzhou, China, <https://doi.org/10.1109/NNICE61279.2024.10498676>
12. Olanrewaju Bola Wojuola. Mathematical Model for Bit-error-rate of a Spread-Spectrum System//2019 21st International Conference on Advanced Communication Technology (ICTACT), 02 May 2019, PyeongChang, Korea, <https://doi.org/10.23919/ICTACT.2019.8701934>
13. Wenzhun Huang, Shanwen Zhang, Robert Hao Yan. Novel Spread Spectrum Communication Theory and the Anti-jamming Applications//2021 6th International Conference on Inventive Computation Technologies (ICICT), 26 February 2021, Coimbatore, India, <https://doi.org/10.1109/ICICT50816.2021.9358748>
14. I. V. Malugin, A. S. Luchinin, V. A. Surgutskaya, Yu. V. Kozlov. One of the Ways to Protect a Spread Spectrum Communication System from such Structural Interference//2020 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO), 13 August 2020, <https://doi.org/10.1109/SYNCHROINFO49631.2020.9166090>

15. Guomin Tang, Lidong Zhu, Qi Wu, Qihui He, Le Yu. A Hybrid Spread Spectrum Communication Method Based on Chaotic Sequence//2021 International Symposium on Networks, Computers and Communications (ISNCC), 25 November 2021, Dubai, United Arab Emirates, <https://doi.org/10.1109/ISNCC52172.2021.9615817>
16. Md Rakibur Rahman, Sathesh Bojja- Venkatakrishnan, Elias A. Alwan, John L. Volakis. Spread Spectrum Techniques for Interference Mitigation in Large Bandwidth//2019 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting, 31 October 2019, Atlanta, GA, USA, <https://doi.org/10.1109/APUSNCURSINRSM.2019.8888648>
17. Jingtian Xu, Zhixuan Zhang. FPGA Spread Spectrum Communication Method Based on M Sequence// 2021 3rd International Conference on Intelligent Control, Measurement and Signal Processing and Intelligent Oil Field (ICMSP), 20 August 2021, Xi'an, China, <https://doi.org/10.1109/ICMSP53480.2021.9513383>
18. Avinash Kumar, Ye Zhu. Extending Direct Sequence Spread-Spectrum for Secure Communication//2021 International Symposium on Networks, Computers and Communications (ISNCC), 25 November 2021, Dubai, United Arab Emirates, <https://doi.org/10.1109/ISNCC52172.2021.9615783>
19. Почерняев В.Н. Устройства на частично заполненных диэлектриком волноводах. Киев: УКНИПСК, 2000. 224с.
20. Почерняев В.Н., Цибизов К.Н. Теория сложных волноводов. Науовий світ, 2003. 223с.
21. Pochernyaev V., Syvkova N. Broadband switch on partially filled by dielectric rectangular waveguide». The scientific heritage, 2021. № 60 (60). с. 49-52. <https://doi.org/10.24412/9215-0365-2021-60-1-49-52>
22. Почерняев В.М., Сивкова Н.М. Пристрій управління потужністю НВЧ на частково заповненому діелектриком прямокутному хвилеводі. Інфокомунікаційні та інформаційні технології, 2021. №1 (01). с.81-89. <https://visn-icct.uu.edu.ua/index.php/icct/article/view/28/7>
23. Почерняев В.М., Сивкова Н.М. Фільтр на хвилеводних трійниках, що частково заповнені діелектриком. The scientific heritage, 2021. № 75 (75). с. 42-47.
24. Почерняев В.М., Магомедова М.С. Сивкова Н.М. Пристрій регулювання потужністю НВЧ на частково заповнених діелектриком прямокутних хвилеводах. Інфокомунікаційні та інформаційні технології, 2021. №2 (02). с. 161-171. <https://visn-icct.uu.edu.ua/index.php/icct/article/view/49/36>
25. Почерняев В.М., Сивкова Н.М. Перемикач на хвилеводному трійнику, частково заповнений діелектриком. The scientific heritage, 2022. № 84 (84). с.53-57. DOI:10.24412/9215-0365-2022-84-1-53-57
26. Почерняев В.М., Магомедова М.С. Сивкова Н.М. Обмежувач потужності НВЧ на частково заповнених діелектриком прямокутних хвилеводах. Інфокомунікаційні та комп'ютерні технології, 2022. №1(03). С.90-101. <https://doi.org/10.36994/2788-5518-2022-01-03-06>
27. Pochernyaev V., Syvkova N., Mahomedova M. Switch-filter on a rectangular waveguide partially filled by dielectric. Informatyka, Automatyka, Pomiaru W Gospodarce I Ochronie Środowiska, 2022. №12(3). р.8-11. <https://ph.pollub.pl/index.php/iapgos/article/view/3052/2728>
28. Почерняев В.М., Магомедова М.С., Сивкова Н.М. Фазо-частотний пристрій на частково заповненому діелектриком прямокутному хвилеводі. Системи управління, навігації та зв'язку, 2022. №4 (70). С.158-161 <https://doi.org/10.26906/SUNZ.2022.4.158>
29. Pochernyaev V., Mahomedova M., Syvkova N. Комутаційний фазообертач на частково заповненому діелектриком прямокутному хвилеводі. Системи управління, навігації та зв'язку. Полтава: ПНТУ, 2023. Т. 1 (71). С. 171-176. <https://doi.org/10.26906/SUNZ.2023.1.171>
30. V Pochernyaev, M Mahomedova, N Syvkova. Плоска фазована антенна решітка для мобільних цифрових станцій зв'язку «точка-багаточка» НВЧ діапазону. Системи управління, навігації та зв'язку, 2024. №2, С.187-190. <https://doi.org/10.26906/SUNZ.2024.2.187>

Received (Надійшла) 27.12.2024

Accepted for publication (Прийнята до друку) 05.02.2025

Increasing the interference immunity of mobile digital combined microwave radio engineering systems

V. Pochernyaev, M. Mahomedova, N. Syvkova

Abstract. In the article evaluates the interference immunity of the joint system pseudo-random frequency hopping – wideband signals. It is known that mobile digital combined microwave radio engineering systems use phase-shift keyed wideband signals and pseudo-random frequency hopping. The mobile digital combined troposcatter-space system also uses phase-shift keyed wideband signals and pseudo-random frequency hopping. The use of a joint system of pseudo-random frequency hopping - wideband signals is due to the requirement for stable operation of any radio engineering system in the conditions of a complex interference environment created by enemy radio suppression stations. The article analyzes the relevant literature. The schemes of the transmitter and receiver for the formation and reception of a joint pseudo-random frequency hopping - a wideband signal are given. The peculiarity of these schemes is that the transmitting and receiving microwave paths are implemented on rectangular waveguides partially filled by a dielectric. The influence of barrage noise interference on the interference immunity of such a receiver is investigated in the article. The probability of error per bit for QPSK modulation under the influence of white Gaussian noise and barrage interference is shown. The expression for the probability of error per bit is written through the Crump function. The probability of error per bit is found under the conditions that the interference spectrum exceeds the spectrum of the wideband signal and under the condition that the interference spectrum is narrower than the spectrum of the wideband signal. The article calculates the dependences of the probability of error per bit on the ratio of the signal power to the total power of noise and interference for different values of the spectral expansion coefficients for the cases of using only pseudo-random frequency hopping, only a wideband signal and joint pseudo-random frequency hopping – wideband signal. It is concluded that taking into account the presence of a protection device in the receiver of a wideband signal, under conditions of interference in part of the band, the noise immunity of the joint system is higher than the interference immunity of a pseudo-random frequency hopping system or a system using a wideband signal.

Keywords: pseudo-random frequency hopping; phase-shift keyed signal; wideband signal; spread spectrum system.

М. В. Савченко, М. В. Шиман

Національний технічний університет «Харківський політехнічний інститут», Харків, Україна

МЕТОД АНАЛІЗУ ЗАВАНТАЖЕННЯ ВУЗЛІВ КЛАСТЕРУ MESH-МЕРЕЖІ НА ОСНОВІ МАТЕМАТИЧНОЇ МОДЕЛІ МЕРЕЖ ДЖЕКSONA

Анотація. У статті представлено метод аналізу завантаження кластеру вузлів у MESH-мережах на основі математичного апарату мереж Джексона. Розглянуто математичну модель, що описує взаємодію вузлів у кластері, зокрема їхню здатність обробляти вхідний трафік та перерозподіляти потоки між собою. Модель враховує ймовірності переходу заявок між вузлами, швидкість обслуговування кожного вузла та зовнішній вхідний трафік. Основну увагу приділено аналізу завантаження вузлів та умовам стабільності системи. Проведено розрахунок коефіцієнтів завантаження вузлів, ймовірності відсутності заявок, а також середньої кількості заявок у кожному вузлі. Показано, що для забезпечення стабільної роботи системи необхідно, щоб коефіцієнт завантаження кожного вузла був меншим за одиницю. Визначено вплив параметрів системи, таких як інтенсивність вхідного трафіку та швидкість обслуговування, на продуктивність кластеру. Запропонований підхід є ефективним інструментом для оптимізації MESH-мереж, оскільки дозволяє ідентифікувати вузли з критичним навантаженням та оцінити стабільність системи в цілому. Модель може бути використана для проектування нових мереж та вдосконалення існуючих, особливо в умовах різноманітного трафіку та динамічних змін параметрів системи.

Ключові слова: мережа Джексона, кластер, математична модель, MESH-мережа, трафік.

Вступ

Актуальність дослідження. MESH-мережі є однією з найважливіших технологій сучасних телекомунікаційних систем, які знаходять широке застосування у бездротових мережах, сенсорних системах та інтернеті речей. Завдяки їхній адаптивній структурі та високій гнучкості вони дозволяють створювати розподілені та самовідновлювальні мережі, здатні ефективно функціонувати навіть за умов втрати окремих вузлів. Однак у таких мережах виникають складні задачі аналізу та оптимізації трафіку, оскільки взаємодія вузлів кластеру супроводжується нерівномірним завантаженням і перерозподілом заявок, що може призводити до перевантаження окремих елементів системи.

Актуальність дослідження полягає в необхідності розробки ефективних методів моделювання та аналізу завантаження вузлів у MESH-мережах для забезпечення їхньої стабільності та продуктивності. Одним із перспективних підходів є застосування математичного апарату мереж Джексона, який дозволяє моделювати складні системи черг із взаємодіючими вузлами та оцінювати ключові характеристики, такі як коефіцієнт завантаження, середній час перебування заявок та ймовірність їх втрат. Такий підхід забезпечує можливість адаптації мережі до змін у трафіку та оптимізації її роботи.

У статті пропонується метод аналізу завантаження кластеру вузлів у MESH-мережах на основі мереж Джексона. Дослідження спрямоване на оцінку стабільності системи, виявлення вузлів із критичним навантаженням та визначення умов, за яких система залишається ефективною. Запропонований підхід може бути застосований для оптимізації існуючих MESH-мереж, планування нових систем та розробки інструментів для автоматичного управління трафіком у розподілених мережах.

Мета та задачі дослідження. Метою статті є розробка методу аналізу завантаження кластеру вуз-

лів у MESH-мережах на основі математичного апарату мереж Джексона, що дозволяє оцінювати стабільність системи, виявляти критично завантажені вузли та оптимізувати розподіл трафіку. Дослідження спрямоване на удосконалення методів моделювання телекомунікаційних систем з метою підвищення їх продуктивності та адаптивності в умовах змінних характеристик трафіку.

Одним із завдань є побудова математичної моделі, яка описує взаємодію вузлів кластеру, враховуючи зовнішній трафік, швидкість обслуговування заявок та ймовірності переходу між вузлами. Така модель має відображати основні характеристики системи черг та дозволяти проводити аналіз її параметрів, таких як коефіцієнти завантаження вузлів, середня кількість заявок та ймовірність відсутності заявок у кожному вузлі.

Наступним завданням є дослідження умов стабільності системи на основі розрахунку коефіцієнтів завантаження вузлів. Важливо визначити критичні значення параметрів, при яких система залишається у стабільному режимі роботи, а також оцінити вплив змін зовнішнього трафіку та швидкості обслуговування на продуктивність кластеру.

Ще однією важливою задачею є виявлення вузлів із критичним навантаженням та розробка рекомендацій щодо їхнього розвантаження. Це передбачає аналіз можливих шляхів оптимізації, таких як зміна розподілу трафіку, збільшення швидкості обслуговування або перерозподіл заявок між вузлами.

Останнім завданням є застосування отриманих результатів для практичної оптимізації MESH-мереж, зокрема, їхнього планування, управління трафіком та розробки адаптивних рішень для роботи в умовах різноманітного трафіку. Результати дослідження можуть бути використані як інструмент для підвищення ефективності існуючих мереж та вдосконалення їхньої архітектури.

Огляд літератури. У сучасних дослідженнях моделювання завантаження вузлів MESH-мереж із

використанням мереж Джексона привертає значну увагу. Зокрема, у статті [1] автори досліджують середній вік інформації в мережах Джексона з обмеженим розміром буфера. Перевагою цього підходу є можливість оцінки ефективності інформаційних систем, однак обмеженням є складність у застосуванні до великих мереж через обмеження буферів.

Інше дослідження, представлене у статті [2], аналізує узагальнені мережі Джексона з урахуванням можливих відмов. Перевагою є здатність моделі враховувати різні режими відмов та стратегії відновлення, проте недоліком є складність у застосуванні до реальних мереж через необхідність точного визначення параметрів відмов.

У книзі [3] автори детально розглядають теоретичні основи та практичні аспекти віку інформації в мережах, що дозволяє глибше зрозуміти динаміку оновлення даних у складних системах. Перевагою є комплексний підхід до аналізу.

Підхід на основі мереж Джексона має суттєві переваги порівняно з класичними марківськими

моделями [4, 5], зокрема в контексті моделювання завантаження вузлів у складних системах, таких як MESH-мережі.

На відміну від марківських моделей, які переважно орієнтовані на аналіз станів окремих вузлів з безпам'ятністю, мережі Джексона дозволяють моделювати взаємодію множинних вузлів і враховувати перерозподіл трафіку між ними. Це забезпечує більш реалістичне відображення динаміки системи, особливо для розподілених і адаптивних мереж. Проте мережі Джексона мають вищу обчислювальну складність і вимагають точного визначення ймовірностей переходів між вузлами, що може бути викликом для великих і гетерогенних систем.

Результати досліджень

1. Математична модель кластеру мережі з трьома вузлами. Побудуємо математичну модель частини MESH-мережі з трьома вузлами, де вузли взаємопов'язані, використовуючи математичний апарат мереж Джексона (рис. 1).

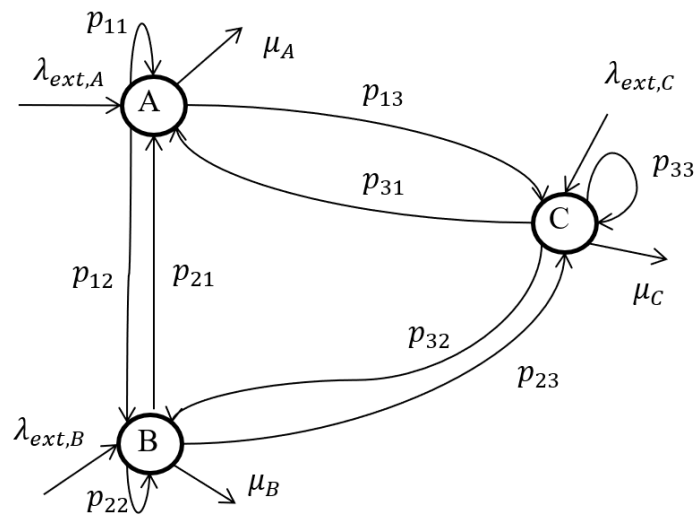


Рис. 1. Модель мережі з трьома вузлами

Опишемо структуру мережі за допомогою стандартних позначень теорії графів. Є три вузли А, В, С, які можуть обмінюватися трафіком один з одним та з зовнішнім середовищем (іншими вузлами великої мережі, які не потрапили у кластер, що розглядається).

Між вузлами встановлені ймовірності переходів p_{ij} , де i, j — індекси вузлів.

У свою чергу задано наступні параметри системи, інтенсивність зовнішнього потоку для кожного вузла $\lambda_{ext,i}$ для $i = A, B, C$, швидкість обслуговування заявок у кожному вузлі μ_i для $i = A, B, C$, матриця ймовірностей переходу між вузлами $P = \{p_{ij}\}$, де p_{ij} — ймовірність переходу заявки з вузла i до j .

Використаємо рівняння потокового балансу. Для кожного вузла сума вхідних потоків дорівнює сумі оброблених та переданих потоків. Запишемо це співвідношення.

$$\lambda_i = \lambda_{ext,i} + \sum_{j \neq i} \lambda_j p_{ji}, \quad (1)$$

де λ_i — сумарна інтенсивність потоку, що обробляється вузлом i .

Отримана система рівнянь для потоків λ_i вирішується, враховуючи матрицю переходів P .

2. Аналіз показників якості обслуговування у вузлах мережі на базі запропонованої моделі. Побудуємо модель кластеру мережі з трьома вузлами. Інтенсивність потоку пакетів, що надходять до цих вузлів відповідно

$$\lambda_{ext,A} = 5, \lambda_{ext,B} = 3, \lambda_{ext,C} = 2.$$

У свою чергу швидкість обслуговування кожного вузла дорівнює $\mu_A = 25, \mu_B = 20, \mu_C = 15$, а матриця переходів задана співвідношенням

$$P = \begin{bmatrix} 0 & 0.3 & 0.2 \\ 0.4 & 0 & 0.3 \\ 0.3 & 0.5 & 0 \end{bmatrix}, \quad (2)$$

де діагональні елементи $p_{ij} = 0$, оскільки вузол не пересилає пакети сам собі.

Складемо систему рівнянь для потоків:

$$\begin{aligned}\lambda_A &= \lambda_{ext,A} + \lambda_B p_{BA} + \lambda_C p_{CA}, \\ \lambda_B &= \lambda_{ext,B} + \lambda_A p_{AB} + \lambda_C p_{CB}, \\ \lambda_C &= \lambda_{ext,C} + \lambda_A p_{AC} + \lambda_B p_{BC}.\end{aligned}\quad (3)$$

Підставимо у (3) значення матриці P з (2), отримуємо наступне співвідношення:

$$\begin{aligned}\lambda_A &= 5 + 0.4\lambda_B + 0.3\lambda_C, \\ \lambda_B &= 3 + 0.3\lambda_A + 0.5\lambda_C, \\ \lambda_C &= 2 + 0.2\lambda_A + 0.3\lambda_B.\end{aligned}\quad (4)$$

Для відшукування $\lambda_A, \lambda_B, \lambda_C$ система рівнянь розв'язується за допомогою чисельних методів або підставлення.

В результаті отримано інтенсивності оброблюваного потоку у вузлах:

$$\text{вузол } A: \lambda_A = 11.14;$$

$$\text{вузол } B: \lambda_B = 9.95;$$

$$\text{вузол } C: \lambda_C = 7.21.$$

Розрахуємо коефіцієнти завантаження вузлів за формулою

$$\rho_i = \frac{\lambda_i}{\mu_i}, \quad (5)$$

де маємо такі характеристики вузлів:

$$\text{вузол } A: \rho_A = 0.45;$$

$$\text{вузол } B: \rho_B = 0.50;$$

$$\text{вузол } C: \rho_C = 0.48.$$

Отриманий кластер MESH-мережі було представлено у вигляді спрощеної мережі Джексона (рис. 2), що дозволило здійснити аналіз завантаження вузлів та оцінити стабільність системи.

Використання математичного апарату мереж Джексона забезпечило можливість розрахунку інтенсивності потоків у кожному вузлі, коефіцієнтів їх завантаження, а також середньої кількості заявок та часу їх обслуговування.

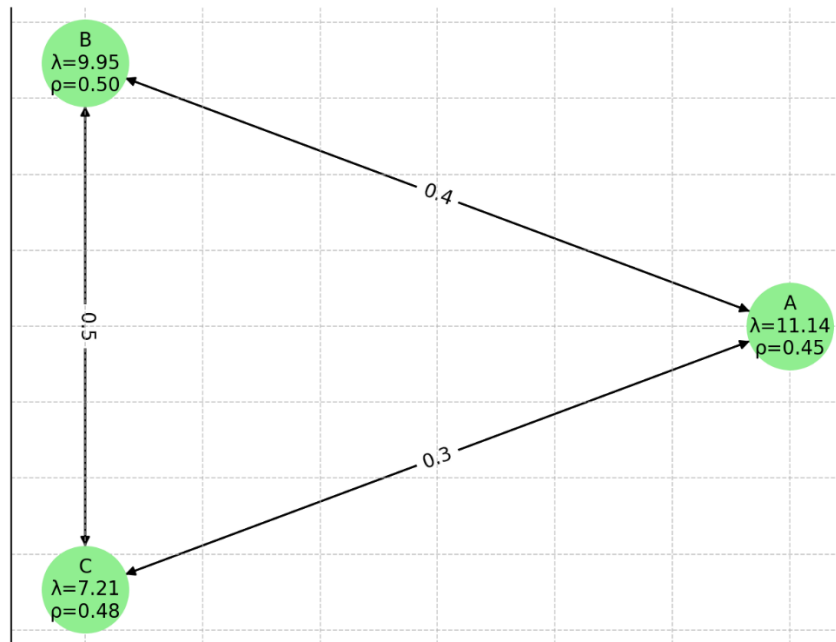


Рис. 2. Кластер MESH-мережі, представлений у вигляді спрощеної мережі Джексона

Це дозволило виявити вузли з критичним навантаженням, що є ключовим для оптимізації ресурсів мережі та підвищення її продуктивності.

Результати підтвердили ефективність запропонованого підходу для аналізу розподілених систем із різномірним трафіком.

3. Аналіз загальних показників якості обслуговування у кластері. Отримані коефіцієнти завантаження ρ_i для всіх вузлів менше за одиницю, що означає, що система працює у стабільному режимі.

Це забезпечує стаціонарність системи, а ймовірність того, що вузли вільні є більшою за нуль. Для кожного вузла визначається ймовірність $P_0(i)$, того, що у вузлі немає заявок:

$$P_0(i) = (1 - \rho_i). \quad (6)$$

Середня кількість заявок у кожному вузлі L_i

розраховується за формулою:

$$L_i = \frac{\rho_i}{1 - \rho_i}. \quad (7)$$

Середній час перебування заявки у черзі та обслуговуванні у вузлі (ping) W_i обчислюється за співвідношенням:

$$W_i = \frac{1}{\mu_i(1 - \rho_i)}. \quad (8)$$

Середній час перебування заявки у всій мережі Джексона розраховується за формулою Літтла:

$$W_{system} = \frac{\sum_1^3 L_i}{\sum_1^3 \lambda_i}. \quad (9)$$

Результати розрахунків параметрів для кожного вузла A, B, C за формулами (6) – (9) наведено у табл. 1, а інтегральний показник $W_{system} = 0.096$.

Таблиця 1 – Розрахунок показників системи

Вузол	λ_i	ρ_i	P_0	L_i	W_i
A	11.1442	0.4457	0.5542	0.8043	0.0721
B	9.9502	0.4975	0.5024	0.9901	0.0995
C	7.2139	0.4809	0.5190	0.9265	0.1284

Висновки

У цій роботі було запропоновано метод аналізу завантаження кластеру вузлів у MESH-мережах на основі математичного апарату мереж Джексона. Побудована модель дозволила оцінити ключові параметри роботи мережі, включаючи інтенсивність трафіку, коефіцієнти завантаження вузлів, середню кількість заявок і час їх перебування у вузлах. Це дало можливість визначити умови стабільної роботи системи та виявити вузли з критичним навантаженням.

Результати дослідження підтвердили ефективність запропонованого підходу для аналізу MESH-мереж, особливо в умовах динамічних змін трафіку та неоднорідного навантаження. Використання мереж Джексона забезпечує гнучкість у моделюванні складних розподілених систем, а також дозволяє здійснювати оптимізацію ресурсів і перерозподіл трафіку для досягнення максимальної продуктивності мережі.

Отримані результати можуть бути використані для вдосконалення існуючих MESH-мереж, плану-

вання нових систем та розробки алгоритмів управління трафіком у реальному часі. Подальші дослідження можуть бути спрямовані на розширення моделі з урахуванням більш складних типів трафіку, відмов вузлів та інших факторів, які впливають на роботу мережі в реальних умовах.

Запропонований метод аналізу завантаження кластеру вузлів у MESH-мережах може бути практично впроваджений у процесах проектування і оптимізації телекомунікаційних систем. Модель на основі мереж Джексона дозволяє ідентифікувати вузли з критичним навантаженням та прогнозувати їхній стан за умов змінного трафіку. Це може бути використано для побудови автоматизованих систем моніторингу та управління мережами, які забезпечуватимуть балансування навантаження у реальному часі.

Окрім того, результати дослідження можуть знайти застосування у сфері IoT-мереж, де необхідно забезпечити стабільну передачу даних між великою кількістю пристроїв. Інтеграція моделі в системи управління дозволить покращити ефективність використання ресурсів мережі, мінімізувати затримки передачі інформації та уникати перевантаження вузлів.

Це особливо актуально для критично важливих застосувань, таких як системи розумного міста, інфраструктура безпеки та сенсорні мережі для моніторингу.

СПИСОК ЛІТЕРАТУРИ

1. Doncel, J., & Assaad, M. (2021). Age of Information of Jackson Networks with Finite Buffer Size. *IEEE Wireless Communications Letters*, 10(4), 902-906. DOI: 10.1109/LWC.2021.3052811.
2. Sommer, J., Berkhout, J., Daduna, H., & Heidergott, B. (2017). Analysis of Jackson Networks with Infinite Supply and Unreliable Nodes. *Queueing Systems*, 87(1-2), 181-207. DOI: 10.1007/s11134-017-9542-1.
3. Sun, Y., Uysal, E., & Dhillon, H. S. (2023). *Age of Information*. Cambridge University Press. DOI: 10.1017/9781108943321.
4. Pustovoitov, P. Assessment of QoS indicators of a network with UDP and TCP traffic under a node peak load mode. / Pustovoitov, P., Voronets, V., Voronets, O., Sokol, H., Okhrymenko, M. // *Eastern-European Journal of Enterprise Technologies*, 2024. V 1. № 4(127). P. 23-31.
5. Пустовойтов П.С. Модель вузла електронної комунікації, що обслуговує tcp-трафік / Воронець В.М., Пустовойтов П.С. // *Системи управління, навігації та зв'язку*. Полтава, 2023. № 4(74). С. 152-155. (Б).

Received (Надійшла) 18.11.2024

Accepted for publication (Прийнята до друку) 22.01.2025

Method for analyzing the load of MESH-network cluster nodes based on the mathematical model of Jackson networks

Mykola Savchenko, Maksym Shyman

Abstract. The article presents a method for analyzing the load of a cluster of nodes in MESH-networks based on the mathematical apparatus of Jackson networks. A mathematical model is considered that describes the interaction of nodes in the cluster. In particular, the model describes their ability to process incoming traffic and redistribute flows among themselves. The model takes into account the probabilities of application transition between nodes, the service speed of each node, and external incoming traffic. The main attention is paid to the analysis of the node load and the conditions of system stability. The calculation of the node load coefficients and the probability of the absence of applications is carried out. The average number of applications in each node is also calculated. It is shown that to ensure stable operation of the system, it is necessary that the load coefficient of each node be less than one. The influence of system parameters, such as the intensity of incoming traffic and the speed of service, on the performance of the cluster is determined. The proposed approach is an effective tool for optimizing MESH networks, as it allows you to identify nodes with critical loads and assess the stability of the system as a whole. The model can be used to design new networks and improve existing ones. The model will be especially useful in conditions of heterogeneous traffic and dynamic changes in system parameters.

Keywords: Jackson network, cluster, mathematical model, MESH network, traffic.

A. Serkov, V. Breslavets, J. Breslavets, I. Yakovenko

Національний технічний університет «Харківський політехнічний інститут», Харків, Україна

GENERATION OF SURFACE OSCILLATIONS OF SEMICONDUCTOR STRUCTURES BY CHARGED PARTICLE FLOWS

Abstract. The results of the work determine the degree of influence of flows of charged particles induced by external electromagnetic radiation on the performance of communication equipment. The **aim** is to determine the conditions for the development of hydrodynamic instabilities of electrostatic oscillations in communication system devices containing semiconductor layers surrounded by media with different electromagnetic properties. The following **results** are obtained: A mechanism for the occurrence and development of surface electrostatic oscillation (plasmon) instabilities is proposed under conditions where the interaction of electromagnetic oscillations and a flow of charged particles generated by external electromagnetic radiation is ensured by the presence of a boundary. Specific features of the transformation of the spectral characteristics of the energy of transition radiation associated with the presence of an interface between media with different electromagnetic properties in open-type radiophysical systems are determined. A new physical mechanism for the generation and amplification of surface electromagnetic oscillations by flows of charged particles, the characteristics of which are determined by the properties of the interface between conductive solids, is developed. **Conclusion.** The criteria for the occurrence and development of surface plasmon instabilities associated with the nonequilibrium of electronic systems obtained in the work are realized in conductive solids. Therefore, they can be used in the development of active devices of semiconductor electronics designed to amplify, generate and convert electromagnetic oscillations in the millimeter and submillimeter ranges. The comparative analysis of the increments of beam-plasma instabilities of various branches of electrostatic oscillations during the motion of a particle flow along the normal or along the interface of media, carried out in the work, allows us to solve the problems of optimizing existing mechanisms for amplifying oscillations in structures used in modern radiophysics (MDS, MOS, various types of p-n junctions).

Keywords: beam instability of electrostatic oscillations, semiconductor components, induced current, electromagnetic radiation, surface oscillations.

Introduction

One of the urgent tasks of modern radiophysics is the need to master the submillimeter and short-wave part of the millimeter ranges of electromagnetic oscillations. These ranges are important not only for many technical applications: radar, radio navigation, communication technology, computing technology, but also in biology, medicine, as well as in research in various areas of theoretical physics.

The possibility of mastering this range is impossible without creating appropriate sources of electromagnetic radiation. In solving it, two approaches were determined.

On the one hand, attempts are being made to improve devices that operate in a lower-frequency part of the spectrum, namely, avalanche-transit diodes, microwave transistors, Gunn diodes. These include studies of plasma-wave effects, unstable states and resonances in conducting solids. The results of these studies open up new possibilities for generating oscillations in this range, and also solve problems of radio spectroscopy of solids. On the other hand, research is being conducted aimed at using the laser principle of amplification and generation of electromagnetic waves in the submillimeter range, which is implemented in optics (in semiconductor lasers). A necessary condition for the successful solution of the tasks set is the availability of an appropriate element base created on the basis of materials with predictable parameters.

Modern technology allows creating conductive solid-state structures: semiconductors with a two-dimensional (2D) electron gas, a superlattice, as well as films and structures of the metal-insulator-semiconductor (MIS) type, etc. The study of the electronic properties of ultra-thin layers and plasma oscillations caused by the

collective behavior of charges determines the mechanisms of their formation.

Solutions to the issues of interaction of plasma oscillations with flows of charged particles are also important for diagnostics and practical use of conductive solid-state structures (for example, in micro- and nanoelectronics). The fact is that in structures that have submicron dimensions, the ballistic mechanism of charge transfer is implemented. Therefore, they may exhibit instabilities based on the effects of Cherenkov, transition and bremsstrahlung radiation of particles.

In addition, the results of studies of wave processes in a limited semiconductor plasma can be used for contactless methods of diagnostics of electron spectra of charge carriers and surface properties. Finally, new branches of electromagnetic oscillations arise in limited media, as well as a connection between different types of oscillations due to the presence of external boundaries.

This work to a certain extent compensates for the existing gaps in studies of wave processes in semiconductor plasma. It examines the mechanisms of generation of electromagnetic oscillations of the submillimeter and short-wave part of the millimeter ranges, caused by their interaction with flows of charged particles induced by pulsed electromagnetic radiation in semiconductor structures used in modern radio electronics and communication technology.

Task solution

Let us consider the dispersion characteristics of electrostatic oscillations in a system that is a semiconductor layer (or dielectric) surrounded by semi-infinite media with different electromagnetic properties [5]. To find the spectrum of electrostatic oscillations of a given structure, we will use the following electrostatic equations:

$$\operatorname{rot} \bar{E} = 0; \quad (1) \quad \operatorname{div} \bar{D} = 0. \quad (2)$$

The electric induction vector $\bar{D}(\bar{r}, t)$ is related to the electric field $\bar{E}(\bar{r}, t)$ by the material equation:

$$\bar{D}(\bar{r}, t) = \int_{-\infty}^t \hat{\varepsilon}(t-t') \bar{E}(\bar{r}, t') dt'. \quad (3)$$

We select the reference system in such a way that the y-axis is directed perpendicular to the interface boundaries, and the x,z-axes are parallel to them. The system is assumed to be infinite along the x,z-axis.

Let the plate with $\hat{\varepsilon} = \hat{\varepsilon}_1$ occupy the region $-d \leq y \leq d$; the half-space $-\infty < y < -d$ - is the medium "2" with $\hat{\varepsilon} = \hat{\varepsilon}_2$, , the half-space $d < y < \infty$ - is the medium "3" with $\hat{\varepsilon} = \hat{\varepsilon}_3$.

At the interface boundaries $y = \pm d$, the conditions of continuity of the tangential components of the electric field \bar{E} and continuity of the normal components of the induction vector are satisfied. At $y = \pm \infty$, the variables included in equations (1)-(2) vanish. We represent the field $\bar{E}(\bar{r}, t)$ in the form:

$$\bar{E}(\bar{r}, t) = \bar{E}(y) e^{i(\bar{q}\bar{\rho} - \omega t)} \quad (4)$$

Where \bar{q} is the wave vector, ω - is the oscillation frequency, $\bar{\rho} = (x, z)$. Since the medium is assumed to be isotropic, the x-axis can be directed along the wave vector \bar{q} . In this case $\bar{E}(\bar{r}, t) = \bar{E}(y, x, t)$; $\bar{E} = (E_x, E_y)$. The solution to system (1) - (2) takes the following form:

$$E_x(y) = \begin{cases} a_2 e^{qy}, & -\infty < y < -d; \\ a_1 e^{qy} + a_0 e^{-qy}, & -d \leq y \leq d; \\ a_3 e^{-qy}, & d < y < \infty; \end{cases} \quad (5)$$

$$E_y(y) = \frac{1}{iq} \frac{\partial E_x}{\partial y}, \rightarrow q > 0.$$

The boundary conditions at $y = \pm d$ give the following dispersion law for the natural oscillations of the system:

$$(\varepsilon_1 + \varepsilon_2)(\varepsilon_1 + \varepsilon_3) = (\varepsilon_1 - \varepsilon_2)(\varepsilon_1 - \varepsilon_3) \exp(-4qd), \quad (6)$$

where $\varepsilon_i(\omega) = \int_0^\infty \hat{\varepsilon}_i = \hat{\varepsilon}_i(\tau) e^{i\omega\tau} d\tau$ - the permittivity of the i-th medium. For the semiconductor layer it is assumed that: $\varepsilon_i(\omega) = \varepsilon_{0i} - \frac{\omega_{0i}^2}{\omega^2}$, $\omega_{0i}^2 = \frac{4\pi e^2 n_{0i}}{m_i}$, ε_{0i} - dielectric constant of the crystal lattice. n_{0i} , m_i - concentration, effective mass of conduction electrons of the medium. Expressions for ε_i are obtained from the equation of motion of conduction electrons.

The constants a_i are related to each other by the following relationships:

$$a_2 = \frac{2\varepsilon_1 a_1}{\varepsilon_1 + \varepsilon_2}; \quad a_0 = \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1 + \varepsilon_2} a_1 e^{-2qd};$$

$$a_3 = \frac{2\varepsilon_1}{\varepsilon_1 + \varepsilon_3} a_1 e^{-2qd}. \quad (7)$$

For large wave numbers ($qd \gg 1$) we obtain two independent solutions: $\varepsilon_1 + \varepsilon_2 = 0$ и $\varepsilon_1 + \varepsilon_3 = 0$ and describing surface plasma oscillations on isolated boundaries $y = \pm d$ of media "1-2" and "1-3". In the opposite limiting case ($d \rightarrow 0$) we have plasma surface oscillations on the boundary of media "2" and "3" ($\varepsilon_2 + \varepsilon_3 = 0$). For small but finite qd , plasma oscillations also arise in the layer $\varepsilon_1(\omega) = 0$.

From equation (6) it follows:

$$\omega_{1,2}^2 = \frac{\Omega_1^2}{2} \left\{ 1 + \frac{\alpha + \beta}{2} \pm \sqrt{1 + \frac{(\alpha + \beta)/2)^2 - 2 \times \frac{\alpha + \beta + (1 + \alpha\beta)th2qd}{1 + th2qd}} \right\}. \quad (8)$$

Here $\varepsilon_{01} = \varepsilon_{02} = \varepsilon_{03} = \varepsilon_0$; $\Omega_i^2 = \omega_{0i}^2 / \varepsilon_0$; $\Omega_2^2 = \alpha \Omega_1^2$; $\Omega_3^2 = \beta \Omega_1^2$; α and β ; are real numbers expressing the relationship between the concentrations of charge carriers in different media. An example of such a system is, for example, "R - P"- transitions at $y = \pm d$.

It is interesting to note one circumstance related to the symmetry of the system. If we put in expression (6), then it splits $\varepsilon_3 = \varepsilon_2$ into two independent equations:

$$(\varepsilon_1 + \varepsilon_2) = \pm (\varepsilon_1 - \varepsilon_2) \exp(-2qd) \quad (9)$$

The equation with the "+" sign describes oscillations with a symmetric distribution of the tangential component of the field in the layer $E_x(-d) = E_x(d)$, the second - with an antisymmetric one. If medium "1" is a dielectric with $\varepsilon_1 + \varepsilon_d$, and where second is a semiconductor $\varepsilon_2 = \varepsilon(\omega)$, where $\varepsilon(\omega) = \varepsilon_0 - \omega_{00}^2 / \omega^2$, then the spectra of symmetric and antisymmetric oscillations have the following form:

$$\omega_1(q) = \frac{\omega_0}{\sqrt{\varepsilon_0 + \varepsilon_d thqd}} \quad (10)$$

$$\omega_2(q) = \frac{\omega_0}{\sqrt{\varepsilon_0 + \varepsilon_d cthqd}} \quad (11)$$

On the contrary, in the case of $\varepsilon_1 = \varepsilon(\omega)$ and $\varepsilon_2 = \varepsilon_d$ the spectrum of symmetric oscillations is described by formula (11), and antisymmetric ones by formula (10). For the metal-insulator-semiconductor structure $\varepsilon_2 \rightarrow \infty$, $\varepsilon_1 = \varepsilon_d$, $\varepsilon_3 = \varepsilon(\omega)$, there is only one branch with the dispersion law:

$$\omega(q) = \frac{\omega_0}{\sqrt{\varepsilon_0 + \varepsilon_d cth2qd}} \quad (12)$$

Let us now consider the interaction of the flow of charged particles with the oscillations of the electric field in these structures. Let the flow of electrons with density n_0 and velocity v_0 move in the medium "1" ($\varepsilon_1 = \varepsilon_d$) along the X-axis (in the final formulas we can always put

($\varepsilon_d = 1$). At the boundaries $y = \pm d$ the normal component of the induction vector undergoes a discontinuity caused by the occurrence of surface charges. Thus, from the Poisson equation, at $y = d$ it follows that the boundary condition takes and is written as follows:

$$\varepsilon_d E_y(d+0) - D_y(d-0) = 4\pi\varepsilon \int_{d-0}^{d+0} ndy. \quad (13)$$

The disturbed concentration n of beam electrons is related to the velocity \vec{v} by the equations of hydrodynamics:

$$i(\varpi - qv_0)n = iqn_0v_x + \frac{\partial}{\partial y}[n_0(y)v_y]; \quad (14)$$

$$(\varpi - qv_0)\vec{v} = i\frac{e}{m}\vec{E}. \quad (15)$$

Using (14), (15), after integrating over the transition layer (13) we find:

$$\varepsilon_d E_y(d+0) - D_y(d-0) = \frac{\omega_b^2 \varepsilon_d}{(\varpi - qv_0)^2} E_y(d+0). \quad (16),$$

Where $\omega_b = (4\pi e^2 n_0 / m)^{1/2}$ is the plasma frequency of the beam electrons. Let us further assume that $\varepsilon_2 = \varepsilon_3 = \varepsilon(\omega)$. In other words, there is a gap in the semiconductor plasma through which a flow of charged particles passes. In this case, in the dispersion relations (9), it is necessary to replace ε_1 with $\varepsilon_d - \omega_b^2 / (\omega - qv_0)^2$. As a result, we obtain:

$$\left[\omega^2 - \omega_1^2(q) \right] (\omega - qv_0)^2 = \frac{\omega_b^2 \omega^2}{\varepsilon_0 \text{cthd} + \varepsilon_d}; \quad (17)$$

$$\left[\omega^2 - \omega_2^2(q) \right] (\omega - qv_0)^2 = \frac{\omega_b^2 \omega^2}{\varepsilon_0 \text{thqd} + \varepsilon_d}. \quad (18)$$

We obtain the solution of equation (17) using the approximation of low beam density $\omega_b^2 / \omega_0^2 \ll 1$ [7]. Assuming $\omega_b = 0$, we define two independent solutions:

$$\omega = \omega_1(q). \quad (19)$$

$$\omega = qv_0. \quad (20)$$

The first characterizes the already known natural oscillations of the system in the absence of a beam, the second – oscillations in a beam with low density. Taking into account the finite density of beam electrons leads to a change in frequencies and the emergence of instability (the Akhiezer and Feinberg mechanism). Under the conditions of Cherenkov resonance, when the frequencies and wave numbers of both branches coincide, the oscillation increment γ has a maximum value:

$$\frac{\gamma}{\omega_1} = \frac{\sqrt{3}}{2} \left(\frac{\omega_b^2 \text{thqd}}{2\omega_0^2} \right)^{1/3}. \quad (21)$$

The magnitude of the instability γ increment characterizes the oscillation generation mode since it

determines the exponential growth of the amplitude E of the electromagnetic fields of radiation of semiconductor devices with time:

$$E \approx \exp(+\gamma t); \quad \gamma = 1/\Delta t_{ei},$$

where Δt_{ei} – the time of effective interaction of induced currents and natural oscillations of the semiconductor structure (the time it takes a particle to pass through the semiconductor layer).

The relative increment for antisymmetric oscillations with a frequency $\omega = \omega_2(q)$ is found in a similar way. It differs from formula (21) by replacing $\text{th} qd$ with $\text{cth} qd$. For small qd , the relative increment for a branch $\omega_1(q)$ is proportional to $(qd)^{1/3}$, and for a branch $\omega_2(q)$ it turns out to be significantly larger and proportional to $(qd)^{-1/3}$. This is due to the fact that the perturbed electron density in the flow is localized at the boundaries and therefore the interaction of space charge waves is more effective with antisymmetric plasmons.

The interaction of an electron beam with plasma oscillations was experimentally studied in indium antimonide and germanium [9, 10] in the millimeter and submillimeter ranges. Due to the high frequency of electron collisions ν , the weak coupling condition is realized in the experiment, where $\Delta\omega < \nu$, where $\Delta\omega$ is the change in the frequency of natural oscillations in the presence of a beam. Therefore, compensation of the wave amplitude attenuation is observed due to the energy of the particle flow.

Analysis

Let us present quantitative estimates of the conditions of resonant (Cherenkov) interaction of surface plasmons with flows of charged particles induced by external electromagnetic radiation, i.e. the possibilities of their generation (amplification) in currently used microwave semiconductor devices. The frequency of surface plasmons for typical values of semiconductor structures used in modern radio electronics is $\omega_s \approx 10^9 - 10^{11} \text{ s}^{-1}$. The drift velocity of carriers for fields in the range of amplitudes of electric E and magnetic H field strengths acting on the semiconductor structure of radiation is $E < 100 \text{ kV/m}$; $H < 600 \text{ A/m}$ consists $v_{dp} \approx 10^{-7} - 10^{-9} \text{ sm/s}$. Therefore, the conditions of resonant interaction of waves and particles (equality of the phase velocity of the wave v_ϕ and the drift velocity of $v_\phi = \omega_s / q \approx v_{dp}$ the particles of the induced current are fulfilled for millimeter (submillimeter) wavelengths $\lambda = 2\pi/q \approx 10^{-1} - 10^{-3} \text{ sm}$. Thus, the proposed physical model of the emergence of conditions for the generation (amplification) of oscillations is realized for most semiconductor devices used in modern radio electronics since their dimensions are within $10^{-1} - 10^{-3} \text{ sm}$.

Conclusion

The conditions for the development of hydrodynamic instabilities of electrostatic oscillations in

communication system devices containing semiconductor layers surrounded by media with different electromagnetic properties are determined.

The problem of the development of the initial perturbation of the charged particle flow, the appearance of which is caused by the action of external electromagnetic radiation, when it passes through the interface of media with different electromagnetic properties, is solved.

Determination of the mechanisms of collisionless attenuation of surface electrostatic oscillations and the conditions of its reversal associated with the interaction of waves and flows of charged particles at the boundary of conducting solids.

Expressions are obtained for the increments of hydrodynamic beam instabilities in semiconductor

structures and structures with ballistic charge transfer.

Ratios are obtained that connect the characteristics of semiconductor structures and the parameters of charged particle flows, ensuring the occurrence of unstable states similar to the Akhiezer-Fainberg instability.

A new physical mechanism for generating and amplifying surface electromagnetic oscillations by flows of charged particles, the characteristics of which are determined by the properties of the interface between conducting solids, has been developed. Specific features of the transformation of the spectral characteristics of the energy of transition radiation associated with the presence of an interface between media with different electromagnetic properties in open-type radiophysical systems have been determined.

REFERENCE

1. Potylitsyn A.P. Transition radiation and diffraction radiation. Seminars and differences // Nucl. Instrum. Methods Phys. Res. – 2019. - v. 145, - P. 67.
2. Rule D.W., Fiorito R.B., Kimura W.D. Noninterceptive beam diagnostics based on diffraction radiation // AIP Conf. Proc. – 2020. – v. 590. – P. 510.
3. Fiorito R.B., Rule D.W. Diffraction radiation diagnostics for moderate to high energy beam // Proc. of the 4. Int. Symp. On Radiation From Relativistic Electrons. – 2020. – v. 155. – P. 67.
4. Mkrthyan A.R. Coherent diffraction radiation from an electron bunch. Nucl. Instrum. Methods Phys. Res. B. 2021. v. 56., P. 69.
5. Aronov I.E., Beletskii N.N. Fundamental steps of group velocity of 2D surface polaritons in high magnetic field // Czechoslovak Journal of Physics. – 2020. - Vol. 46(S5), - P. 2473-2474.
6. Perez-Rodriguez F. and Yampolskii V.A. Hysteresis del campo acustico excitado electromagneticamente en una pelicula metalica // XI Congreso Nacional de la SMCSV. Programa. Cancun, Mexico. – 2020
7. Krowne C.M., Blakey P.A. On the existence of submillimeter wave negative conductance in n – gallium arsenide diodes // J. Appl. Phys. – 2021. - t. 62 №6 - P. 2257 - 2266.
8. Wunsch D.C. and Bell R.R. Determination Of Threshold Failure Of Semiconductor Diodes And Transistors Due To Pullse Voltages // IEEE Trans. – 2020. - Vol. NS-15, No 6. - P. 244-259.
9. Schroen W. and Hooper W.W. Failure Mechanisms in Silicon Semiconductors. // Rome Air Development Center Report No. RADS-TR-64-524. - 2020. Also AD 615312.
10. Shilliday T.S. and Vaccaro J. (Editors). // Physics of Failure in Electronics. Vol. 5, RADS Series in Reliability, Rome Air Development Center. - June 2021. Also AD. 655397.
11. Queisser H.J. Failure Mechanisms in Silicon Semiconductors. // Final Report Contract AF 30 (602)-2556. Rome Air Development Center, Report No. RADC- TDR-62-533. - 2020.
12. Antinone J. Electrical Overstress Protection for Electronic Devices. - New York. – 2021. – 387p.

Received (Надійшла) 25.11.2024

Accepted for publication (Прийнята до друку) 26.02.2025

Генерація поверхневих коливань напівпровідникових структур потоками заряджених часток

О. А. Серков, В. С. Бреславец, Ю. В. Бреславец, І. В. Яковенко

Анотація. Результати роботи визначають рівень впливу потоків заряджених частинок, наведених зовнішнім електромагнітним випромінюванням, на робочі характеристики засобів зв'язку. **Мета статті** – визначення умов розвитку гідродинамічних нестійкостей електростатичних коливань у пристроях систем зв'язку, що містять напівпровідникові шари, оточені середовищами з різними електромагнітними властивостями. The following **results** are obtained: Запропоновано механізм виникнення та розвитку нестійкостей поверхневих електростатичних коливань (плазмонів), в умовах, коли взаємодія електромагнітних коливань та потоку заряджених частинок, створених зовнішнім електромагнітним випромінюванням, забезпечується наявністю кордону. Визначено специфічні особливості трансформації спектральних характеристик енергії перехідного випромінювання, пов'язані з наявністю межі поділу середовищ з різними електромагнітними властивостями у радіофізичних системах відкритого типу. Розроблено новий фізичний механізм генерації та посилення поверхневих електромагнітних коливань потоками заряджених частинок, характеристики яких визначаються властивостями межі розділу твердих тіл, що проводять. **Conclusion.** Отримані в роботі критерії виникнення та розвитку нестійкостей поверхневих плазмонів, пов'язані з нерівноважністю електронних систем, реалізуються у твердих тілах, що проводять. Тому вони можуть бути використані при розробці активних пристроїв напівпровідникової електроніки, призначених для посилення, генерування та перетворення електромагнітних коливань міліметрового та субміліметрового діапазонів. Проведений у роботі порівняльний аналіз інкрементів пучково-плазмових нестійкостей різних гілок електростатичних коливань при русі потоку частинок за нормаллю або вздовж межі розділу середовищ, дозволяє вирішувати завдання оптимізації існуючих механізмів посилення коливань у структурах, що використовуються в сучасній радіофізиці (МДП, МОП, різних р-п переходів).

Ключові слова: пучкова нестійкість електростатичних коливань напівпровідникові комплектуючі, наведений струм, електромагнітне випромінювання, поверхневі коливання.

М. А. Штомпель¹, І. В. Ковтун¹, І. І. Гусєва²

¹ Український державний університет залізничного транспорту, Харків, Україна

² Національний технічний університет України «КПІ імені Ігоря Сікорського», Київ, Україна

ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ ДЕКОДУВАННЯ ЗА ВПОРЯДКОВАНИМИ СТАТИСТИКАМИ АЛГЕБРАЇЧНИХ ЗГОРТКОВИХ КОДІВ

Анотація. У роботі розглянуто застосування декодування за впорядкованими статистиками до коротких алгебраїчних згорткових кодів. Показано, що даний підхід доцільно застосовувати для надійного передавання службових повідомлень у каналах керування новітніх систем радіозв'язку. У роботі запропоновано алгоритм для програмної реалізації процесу завадостійкого передавання інформації для даних умов. Розглянуто загальні принципи та особливості реалізації основних етапів моделювання. Дослідження проводились для алгебраїчних згорткових кодів з максимальною кодовою відстанню для різних довжин кодового обмеження. Декодування даних кодів реалізовувалось для двох обраних степенів впорядкованих статистик у каналі з адитивним білим гаусовим шумом. У розробленій програмній реалізації декодування здійснювалось за критерієм мінімізації зваженої ваги Хемінга між згенерованими тестовими кодовими словами та прийнятим словом. Моделювання завершувалось при досягненні цільового значення коефіцієнту помилок. За результатами проведених досліджень визначено, що декодування за впорядкованими статистиками є ефективним для коротких алгебраїчних згорткових кодів, особливо у діапазоні високого відношення сигнал/шум. Представлений метод декодування доцільно використовувати для підвищення ефективності передавання службової інформації на базі коротких алгебраїчних згорткових кодів у системах радіозв'язку нового покоління.

Ключові слова: алгебраїчні згорткові коди, впорядковані статистики, декодування, ефективність, моделювання, системи радіозв'язку.

Вступ

Постановка проблеми. Новітні системи радіозв'язку передбачають надання різноманітних інформаційних та комунікаційних послуг з підвищеними вимогами до швидкості, надійності та затримки обробки інформації. Для забезпечення заданої якості надання послуг на фізичному рівні даних систем реалізуються удосконалені методи модуляції, цифрової обробки сигналів та завадостійкого кодування. Зокрема, завадостійке кодування відіграє важливу роль при досягненні таких ключових показників ефективності систем радіозв'язку як надійність передавання інформації, загальна продуктивність системи, затримка обробки даних у вузлах мережі, розмір зони покриття мережі, спектральна ефективність та енергетична ефективність мережевого обладнання [1 – 3].

Важливу роль у сучасних технологіях радіозв'язку відіграють згорткові коди. На даний момент існує значна кількість різних типів даних кодів, що використовуються окремо або в складі каскадних кодових конструкцій. Наприклад, нерекурсивні згорткові коди знаходять застосування у системах радіозв'язку другого покоління. В подальшому у системах радіозв'язку третього та четвертого покоління було здійснено перехід до турбокодів, побудованих на основі рекурсивних згорткових кодів. Стандартним методом декодування згорткових кодів є алгоритм Вітербі, обчислювальна складність якого значно зростає для великої довжини кодового обмеження. З іншого боку, застосування спеціального класу алгебраїчних згорткових кодів дозволяє гарантувати високі корегувальні характеристики та використовувати достатньо прості алгебраїчні методи декодування. Традиційною областю згорткових кодів є захист від помилок довгих інформаційних

послідовностей. Проте необхідність передавання коротких повідомлень у каналах керування систем радіозв'язку п'ятого та шостого поколінь вимагає вирішення актуальної задачі пошуку більш ефективних методів декодування [4, 5].

Аналіз останніх досліджень і публікацій. У [6] представлено списковий метод декодування згорткових кодів, що мають перевірочну матрицю з низькою щільністю перевірок на парність. Перевагою даного методу є забезпечення досить малої затримки декодування за рахунок використання фіксованого вікна декодування малого розміру. У [7] розглянуто декодування на основі ковзного вікна локально відновлювальних алгебраїчних згорткових кодів. Показано, що ці коди характеризуються високою корегувальною здатністю та практично досягають границі Сінглтона. У [8] для декодування диференціальних згорткових кодів використано алгебраїчний метод Пітерсона-Горенштейна-Цирлера. З отриманих у роботі результатів випливає, що через обробку лише жорстких рішень ефективність декодування є досить невисокою. Для підвищення ефективності алгебраїчних згорткових кодів у [9] представлено біоінспірований метод декодування, що заснований на процедурі диференціальної еволюції. У роботі розглянуто особливості основних етапів декодування та підтверджено досягнення меншої імовірності помилки декодування порівняно з алгебраїчним методом декодування. У [10] запропоновано алгоритми для програмної реалізації даного методу декодування алгебраїчних згорткових кодів.

У [11] показано, що декодування за впорядкованими статистиками можна розглядати у якості універсального методу декодування коротких завадостійких кодів. Зокрема, у [12] проведено дослідження особливостей цього методу декодування, проаналізовано статистичні властивості та вплив

зваженої ваги Хемінга між оцінками кодового слова та прийнятим з каналу зв'язку словом на реалізацію декодера. Для зменшення обчислювальної складності декодування у [13] застосовується локальне обмеження впорядкованих статистик для пошуку тестових помилок з використанням модифікованого декодування Вітербі. Введення специфічних параметрів та імовірнісних оцінок для зменшення області пошуку тестових помилок дозволяють додатково покращити ефективність декодування за впорядкованими статистиками [14].

Таким чином, проведений аналіз показав, що доцільно розглянути застосування процедури впорядкованих статистик для підвищення ефективності декодування коротких алгебраїчних згорткових кодів у системах радіозв'язку нового покоління.

Мета статті: оцінювання ефективності декодування за впорядкованими статистиками коротких алгебраїчних згорткових кодів у системах радіозв'язку нового покоління.

Виклад основного матеріалу

Деякий (n, k, m) алгебраїчний згортковий код повністю задається узагальненою породжувальною матрицею, побудованою з використанням обраного недвійкового блокового коду, та характеризується наступними параметрами: довжина інформаційного кадру k , довжина кодового кадру n , довжина кодового обмеження m , швидкість кодування R . Теоретично дані коди можуть застосовуватись для кодування нескінченної інформаційної послідовності, але на практиці інформаційне повідомлення завжди обмежується та застосовуються різні механізми усичення згорткових кодів. Це дозволяє розглянути можливість м'якого декодування алгебраїчних згорткових кодів на основі стандартних методів, наявних у теорії блокових кодів. Відомо, що декодування за впорядкованими статистиками є універсальним та ефективним методом для блокових кодів з довільною структурою, тому у роботі проведено дослідження ефективності застосування даного підходу до алгебраїчних згорткових кодів.

Оцінювання ефективності методу декодування алгебраїчних згорткових кодів за впорядкованими статистиками засновано на розробленій програмній реалізації за використанням мови програмування Python. Блок-схема алгоритму проведення моделювання представлена на рис. 1.

Вихідними даними для моделювання є параметри алгебраїчного згорткового коду, довжина інформаційного повідомлення, степінь декодування за впорядкованими статистиками, діапазон відношення сигнал/шум та цільовий коефіцієнт помилок. Моделювання передбачає генерування кодових слів для кожного випадкового інформаційного повідомлення на основі узагальненої породжувальної матриці заданого алгебраїчного згорткового коду. При моделюванні було використано алгебраїчні згорткові коди з максимальною кодовою відстанню, побудовані на базі породжувальної матриці коду Ріда-Соломона [9].

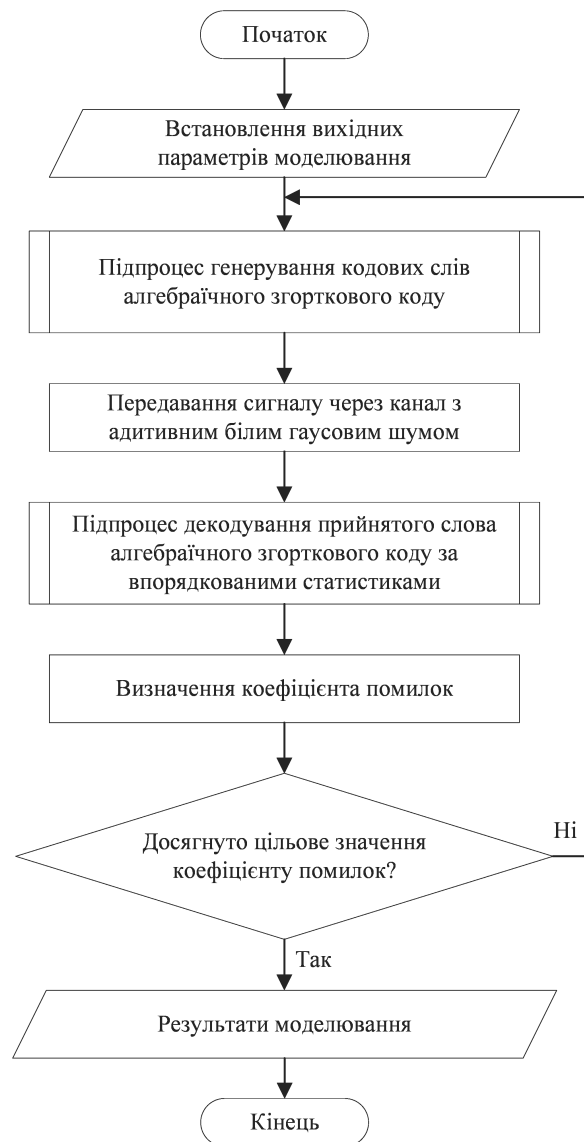


Рис. 1. Блок-схема алгоритму проведення моделювання

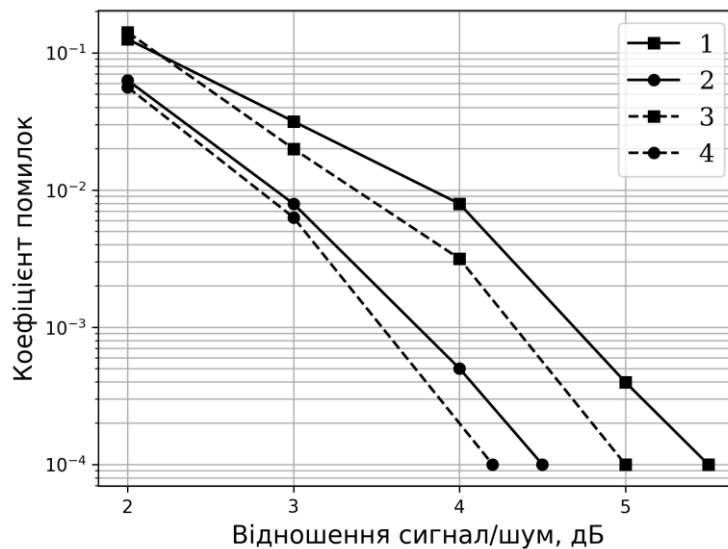
Передавання інформації здійснюється фазомодульованим сигналом, що відповідає кодовому слову, через канал зв'язку з адитивним білим гаусовим шумом (АБГШ). Декодування прийнятого слова реалізується для обраної степені впорядкованих статистик D із визначенням відповідного коефіцієнту помилок. У розробленій програмній реалізації декодування здійснюється за критерієм мінімізації зваженої ваги Хемінга між згенерованими тестовими кодовими словами та прийнятим словом [11, 12]. Процес моделювання продовжується до досягнення цільового значення коефіцієнту помилок p . За результатами моделювання формується набір знайдених оцінок кодових слів алгебраїчного згорткового коду та зберігаються коефіцієнти помилок для кожного значення сигнал/шум.

Параметри моделювання наведено у табл. 1.

Результати моделювання методу декодування за впорядкованими статистиками для обраних алгебраїчних згорткових кодів у каналі з АБГШ наведено на рис. 2.

Таблиця 1 – Параметри моделювання

Назва параметру	Значення			
Модель каналу	АБГШ			
Метод модуляції	Фазова модуляція			
Тип завадостійкого коду	Алгебраїчний згортковий код			
Довжина інформаційного повідомлення	64			
Швидкість кодування	1/2			
Довжина кодового обмеження	6	8		
Метод декодування	Декодування за впорядкованими статистиками			
Степінь впорядкованих статистик	1	2	1	2
Діапазон відношення сигнал/шум, дБ	від 2 до 6			
Цільовий коефіцієнт помилок	10^{-4}			

Рис. 2. Залежність коефіцієнту помилок від відношення сигнал/шум для алгебраїчних згорткових кодів з $R = 1/2$

На рис. 2 позначено:

1, 2 – результати декодування за впорядкованими статистиками (2, 1, 6) алгебраїчного згорткового коду при $D=1$ та $D=2$ відповідно;

3, 4 – результати декодування за впорядкованими статистиками (2, 1, 8) алгебраїчного згорткового коду при $D=1$ та $D=2$ відповідно.

З аналізу графіків на рис. 2 слідує, що у випадку (2, 1, 6) алгебраїчного згорткового коду збільшення степеня впорядкованих статистик D дозволяє отримати додатковий енергетичний вигоду близько 1,3 дБ для цільового коефіцієнту помилок $p = 10^{-4}$. Одним із шляхів підвищення ефективності декодування є збільшення довжини кодового обмеження m , хоча це, в свою чергу, призводить до зростання затримки обробки інформації та обчислювальної складності декодування [6, 9]. З проведених експериментів випливає, що застосування (2, 1, 8) алгебраїчного згорткового коду та декодування за впорядкованими статистиками при $D=1$, підвищує енергетичну ефективність на 0,8 дБ для заданого значення коефіцієнту помилок $p = 10^{-4}$. При цьому

збільшення степеня впорядкованих статистик до $D=2$ дозволяє отримати лише близько 0,3 дБ енергетичного вигоду порівняно з (2, 1, 6) алгебраїчним згортковим кодом.

Таким чином, вигоду від збільшення степеня впорядкованих статистик є не досить значимим для алгебраїчних згорткових кодів зі швидкістю кодування $R = 1/2$ та довжиною кодового обмеження $m = 8$.

Висновки

Для забезпечення високої якості надання послуг у системах радіозв'язку різних поколінь широке застосування знаходять згорткові коди.

Передавання коротких службових повідомлень у каналах керування новітніх систем радіозв'язку доцільно реалізувати з використанням алгебраїчних згорткових кодів.

У роботі розглянуто застосування декодування за впорядкованими статистиками до коротких алгебраїчних згорткових кодів.

Для оцінювання ефективності даного підходу було проведено моделювання на основі розробленої

програмної реалізації процесу завадостійкого передавання інформації через канал з АБГШ.

Визначено, що декодування за впорядкованими статистиками дозволяє забезпечити ефективне декодування алгебраїчних згорткових кодів для цільового коефіцієнту помилок, особливо у діапазоні високого

відношення сигнал/шум (у області понад 4 дБ).

Представлений метод декодування доцільно використати для підвищення ефективності передавання службової інформації на базі коротких алгебраїчних згорткових кодів у системах радіозв'язку нового покоління.

СПИСОК ЛІТЕРАТУРИ

1. Aslam A. M., Chaudhary R., Bhardwaj A., Budhiraja I., Kumar N., Zeadally S. Metaverse for 6G and beyond: the next revolution and deployment challenges. *IEEE Internet of Things Magazine*. 2023. Vol. 6. No. 1. P. 32-39. <https://doi.org/10.1109/IOTM.001.2200248>.
2. Masaracchia A., Li Y., Nguyen K. K., Yin C., Khosravirad S. R., Benevides Da Costa D. UAV-enabled ultra-reliable low-latency communications for 6G: a comprehensive survey. *IEEE Access*. 2021. Vol. 9. P. 137338–137352. <https://doi.org/10.1109/ACCESS.2021.3117902>.
3. Pocovi G., Kolding T., Pedersen K. I. On the cost of achieving downlink ultra-reliable low-latency communications in 5G networks. *IEEE Access*. 2022. Vol. 10. P. 29506–29513. <https://doi.org/10.1109/ACCESS.2022.3158361>.
4. Zhang H., Tong W. Channel coding for 6G extreme connectivity – requirements, capabilities, and fundamental tradeoffs. *IEEE BITS the Information Theory Magazine*. 2023. Vol. 3. No. 1. P. 54-66. <https://doi.org/10.1109/MBITS.2023.3322978>.
5. Geiselhart M., Krieg F., Clausius J., Tandler D., Ten Brink S. 6G: a welcome chance to unify channel coding? *IEEE BITS the Information Theory Magazine*. 2023. Vol. 3. No. 1. P. 67-80. <https://doi.org/10.1109/MBITS.2023.3322974>.
6. Wang Q., Cai S., Chen L., Ma X. Semi-LDPC convolutional codes: construction and low-latency windowed list decoding. *Journal of Communications and Information Networks*. 2021. Vol. 6. No. 4. P. 411-419. <https://doi.org/10.23919/JCIN.2021.9663105>.
7. Martínez-Peñas U., Napp D. Locally repairable convolutional codes with sliding window repair. *IEEE Transactions on Information Theory*. 2020. Vol. 66. No. 8. P. 4935-4947. <https://doi.org/10.1109/TIT.2020.2977638>.
8. Gómez-Torrecillas J., Lobillo F.J., Navarro, G., Sánchez-Hernández J. P. Peterson–Gorenstein–Zierler algorithm for differential convolutional codes. *Applicable Algebra in Engineering, Communication and Computing*. 2021. 32. P. 321-344. <https://doi.org/10.1007/s00200-020-00464-6>.
9. Panchenko, S., Prykhodko, S., Kozelkov, S., Shtompel, M., Kosenko, V., Shefer, O., Dunaievska, O. Analysis of efficiency of the bioinspired method for decoding algebraic convolutional codes. *Eastern-European Journal of Enterprise Technologies*. 2019. 2(4 (98)). P. 22–30. <https://doi.org/10.15587/1729-4061.2019.160753>.
10. Приходько С.І., Штомпель М. А., Власов А. В. Принципи програмної реалізації біоінспірованого методу декодування алгебраїчних згорткових кодів. *Інформаційно-керуючі системи на залізничному транспорті*. 2019. № 2. С. 18-24. <https://doi.org/10.18664/iksz.v0i2.164877>.
11. Shirvanimoghaddam M. et al. Short block-length codes for ultra-reliable low latency communications. *IEEE Communications Magazine*. 2019. Vol. 57. No. 2. P. 130–137. <https://doi.org/10.1109/MCOM.2018.1800181>.
12. Yue C., Shirvanimoghaddam M., Vucetic B., Li Y. A revisit to ordered statistics decoding: distance distribution and decoding rules. *IEEE Transactions on Information Theory*. 2021. Vol. 67. No. 7. P. 4288–4337. <https://doi.org/10.1109/TIT.2021.3078575>.
13. Liang J., Wang Y., Cai S., Ma X. A low-complexity ordered statistic decoding of short block codes. *IEEE Communications Letters*. 2023. Vol. 27. No. 2. P. 400–403. <https://doi.org/10.1109/LCOMM.2022.3222819>.
14. Yue C., Shirvanimoghaddam M., Park G., Park O.-S., Vucetic B., Li Y. Probability-based ordered-statistics decoding for short block codes. *IEEE Communications Letters*. 2021. Vol. 25. No. 6. P. 1791–1795. <https://doi.org/10.1109/LCOMM.2021.3058978>.

Received (Надійшла) 14.01.2025

Accepted for publication (Прийнята до друку) 05.03.2025

Evaluation of efficiency of ordered statistics decoding for algebraic convolutional codes

M. Shtompel, I. Kovtun, I. Husyeva

Abstract. The paper considers the application of ordered statistics decoding to short algebraic convolutional codes. It is shown that this approach is expedient to use for reliable transmission of service messages in control channels of modern radio communication systems. The paper proposes an algorithm for software implementation of the process of noise-resistant information transmission for given conditions. The general principles and features of the implementation of the main stages of modeling are considered. The research was conducted for algebraic convolutional codes with maximum code distance for different code constraint lengths. Decoding of these codes was implemented for two selected values of ordered statistics in additive white Gaussian noise channel. In the developed software implementation, decoding was carried out according to the criterion of minimizing the weighted Hamming weight between the generated test codewords and the received word. The modeling was completed when the target value of the error rate was reached. According to the results of the conducted research, it was determined that ordered statistics decoding is effective for short algebraic convolutional codes, especially in the range of high signal-to-noise ratio. The presented decoding method is advisable to use to increase the efficiency of service information transmission based on short algebraic convolutional codes in new generation radio communication systems.

Keywords: algebraic convolutional codes, ordered statistics, decoding, efficiency, modeling, radio communication systems.

V. Yareschenko¹, V. Kosenko^{1,2}

¹ National University «Yuri Kondratyuk Poltava Polytechnic», Poltava, Ukraine

² Kharkiv National University of Radio Electronics, Kharkiv

METHOD OF GENERATING ENERGY EFFICIENT CODES

Abstract. The article considers the current problem of reducing the power that is dissipated in global communication lines while maintaining high performance. High switching activity leads to significant losses through coupling capacities between long lines. They are one of the most energy-intensive components of the embedded system. The use of Gray codes not only reduces dynamic energy losses in the address bus, but also minimizes communication losses between closely spaced lines. However, the Gray code has low balance and a large number of bit switches. The methods for constructing alternative Gray codes are investigated, which makes it possible to reduce the number of switches on the buses. The purpose of the article is to develop a method for constructing unit distance codes, determining the type of code transformations and forming a system of various typical representatives. Estimates of their number are obtained, characteristics are determined, catalogs of typical representatives are formed. The application of the developed method will make it possible to analyze and select codes with the best properties and, as a result, obtain the best results in terms of network delays, energy costs and other design constraints for computer systems.

Keywords: coding, unit distance codes, energy efficient coding, switching activity, equivalence.

Introduction

Statement of the problem. Computer systems are widely used in various fields of science and technology to build systems for controlling, regulating, transmitting, and processing discrete information.

Modern processors are becoming increasingly bottlenecked by the energy required to move data across different levels of the memory hierarchy and between different input and output devices.

High switching activity leads to significant losses due to capacitance between long lines. They are one of the most power-consuming components of an embedded system [1, 2].

Various methods are used to encode the address bus with the least switching activity. Gray codes are used to reduce dynamic energy. Since access to the instruction memory is often continuous, the use of Gray codes not only reduces dynamic energy losses in the address bus, but also minimizes communication losses between closely spaced lines. However, the Gray code has a low balance and a large number of bit changes [3, 4]. In this regard, the problem of finding low-power coding methods that allow efficient transmission or storage of information is relevant.

Analysis of recent studies and publications. Gray codes are widely used in practice, but they have low balance and, in some cases, cannot be used. In this regard, alternative coding options are being sought [5, 6].

The conceptual foundations of Gray transforms are considered in the works of A. Beletsky [7, 8]. A generalized structure of Gray codes is proposed, which includes the classical “left-handed” and the proposed “right-handed” Gray codes. In the new class of right-handed transformations, the value of the lowest (right) bit of the transformed number remains unchanged during forward and reverse transformations. The introduction of the left and right Gray transforms (both forward and backward) together with the reverse permutation of the codes led to the possibility of constructing combined or composite Gray codes. The

use of composite Gray codes has proven to be quite successful in cryptography, in solving problems of synthesis and analysis of discrete systems. The author considered the properties of transformations for only three bits.

Paper [9] presents algorithms for generating reflected and modular (shifted) m-ary Gray codes. For both variants, the ranking and de-ranking functions are presented, as well as algorithms for generating codewords between two given vectors. A. Phillips and M. Wick [10] developed a recursive method for creating n-bit binary Gray codes. The paper [11] proposes innovative approaches to generating classical and quantum reflected binary Gray codes. A technique is described that generates both long and short Gray codes with the desired properties of the same number of column changes. M. Ali [12] developed a method for constructing an n-bit reflected binary Gray code sequence in space and time.

T. Mutze and J. Nummenpalo studied the Gray code at medium levels, which is a cyclic enumeration of all $2n+1$ bit strings with n or $n+1$ entries equal to 1, such that any two consecutive bit strings in the list differ by exactly one bit for any integer n . They also provided an efficient algorithm for computing the Gray code at medium levels [13].

For image denoising, Y. Zhou, K. Panetta and C. Chen proposed a new parametric n-dimensional Gray code, the so-called (n, k, p) -Gray code, which includes well-known codes such as the binary-reflected code and the (n, k) -Gray code [14]. The (n, k, p) -Gray code varies depending on the values of the base n and the distance parameter p . It is a new type of non-Boolean Gray code when its base is greater than two. X. Wang, Y. Su and H. Zhang proposed to use the (n, k, p) -Gray code for encrypting color images [15]. Experimental results have shown that the encryption algorithm shows high performance in image encryption. It can be used to protect privacy in biometrics, medical imaging systems, and video surveillance systems. M. Tahiri, H. Karmouni, and A. Bencherqui used the (n, k, p) -Gray

code to develop a hybrid optimization and fractional transform algorithm [16].

The article [17] considers the problem of constructing Gray codes using mathematical methods. The authors analyze methods for constructing Gray codes using Hamiltonian graphs and the iterative method. The problem of generating Gray codes of any given length is considered. G. Meenakshi and S. Gupta in [18] proposed an algorithm for generating (n, r) Gray codes. As an example, they consider the design details of the generalized binary Gray code for 5 and 6 bits.

An analysis of known studies has shown that there are other codes that have the same characteristics as Gray codes. This class of codes is called generalized (alternative, extended) Gray codes in the literature. The name "unit-distance codes" reflects the properties of these codes more accurately. Unfortunately, most of the researchers study only particular cases of codes, and there is no general approach to their study.

Aim of the article: development of a method for constructing unit distance codes and forming systems of typical representatives.

Presentation of the main study material

In general, a code is a bijective mapping of a finite ordered set of symbols belonging to some finite alphabet Y to another, not necessarily ordered set of symbols X to encode the transmission, storage, or transformation of information [20, 21]. For codes, the bijective function has the form $f: Y \rightarrow X$, where X is a set of codewords, $X = \{X^1, X^2, \dots, X^k\}$; Y is a finite ordered set of symbols, $Y = \{y_1, y_2, \dots, y_k\}$, k is the number of codewords, n is the number of bits of a binary code.

The code word X^i consists of n characters (the number of bits):

$$X^i = \{x^i_1 \dots x^i_n\}; x^i_j \in \{0,1\}, i=1, \dots, k; j=1, \dots, n,$$

where k is the number of codewords, n is the number of bits of the binary code.

Two adjacent words differ only in one bit, i.e.

$$\rho(X^i, X^{i+1}) = 1, i = 1, \dots, k-1, \quad (1)$$

where ρ is the Hamming distance between codewords X^i and X^{i+1} .

The code in which $\rho(X^1, X^k) = 1$ is called cyclic. The canonical form of a code $W(X)$ is its representation in the form:

$$W(X) = x^1_1 \dots x^1_n \dots x^i_1 \dots x^i_n \dots x^{k-1}_1 \dots x^{k-1}_n \dots x^k_1 \dots x^k_n. \quad (2)$$

For compactness, the binary form of the representation $W(X)$ can be converted to hexadecimal.

On the set of unit distance codes, the following types of transformations are considered: column permutation (P transformation) and column inversion (N transformation). These transformations preserve the property of unit distance between adjacent binary words in the code. The set of transformations is denoted by $\Pi = \{\pi_1, \dots, \pi_{L_{PN}}\}$, where L_{PN} is the number of PN transformations.

The X^i code corresponds to the canonical form $W(X^i)$. As a result of the transformation π_j , code X^i is transformed into code X^j with the canonical form $W(X^j)$. A typical representative of a code X^i is a code (denoted by $T(X^i)$) that has the smallest canonical form in the lexicographic sense among the codes obtained as a result of a set of P transformations, i.e.

$$T(X^i) = \min\{W(X^j)\}; j=1, \dots, L_{PN}. \quad (3)$$

Based on the set of typical representatives, a set of different typical representatives is formed $MT = \{T^1, \dots, T^{KT}\}$, where KT is the number of different typical representatives. For each typical representative included in this set, its structure is determined $S(X) = \langle H(X), \leq \rangle$, where

$$h_i = \sum_{j=2}^k (x_{i,j-1} \oplus x_{i,j}), i = 1, \dots, n. \quad (4)$$

The balance of C code is defined as follows:

$$C = \sum_{i=1}^n \left| h_i \cdot \sum_{j=1}^n h_j / n \right|. \quad (5)$$

Typical representatives $T^1 - T^{KT}$ are divided into non-intersecting equivalence classes S . Typical representatives T^i and T^j are called S-equivalent if $S(T^i) = S(T^j)$. As a result, a set of typical S-equivalent representatives is formed, denoted by $MS = \{S^1, \dots, S^{NS}\}$, where NS is the number of typical S-equivalent representatives.

To build a system of typical representatives for a group of Π transformations, a procedure has been developed, which is described using the following notation:

- i - the current number of the code under review,
- X^i - the code under consideration
- $T(X^i)$ - a typical representative of the code X^i ,
- S^i - the structure of code X^i ,
- KT - the number of different typical representatives,
- MT - the set of different typical representatives,
- KS - the number of code structures,
- MS - the set of code structures,
- $KTC(S_j)$ - the number of different typical representatives that have the structure S_j , $j = 1, \dots, KS$.
- $MTC(S^i, KTC(S^j))$ - the set of different typical representatives that have the structure S^i .

The procedure for constructing a system of typical representatives for a group of Π transformations consists of the following steps.

1. $i = 0$.
2. $KT = 0, KS = 0$.
3. $i = i + 1$.
4. Let's form the step X^i .
5. We define a typical code representative $X^i - T(X^i)$.
6. We define the code $X^i - S^i$ structure.
7. If $T(X^i) \in MT$, then proceed to the stage 19.
8. $KT = KT + 1$.
9. $MT(KT) = T(X^i)$.
10. If $S^i \in MS$, then proceed to the stage 12.
11. Proceed to the stage 15.
12. $KTC(S^i) = KTC(S^i) + 1$.

13. $MTC(S^i, KTC(S^i)) = T(X^i)$.
14. Proceed to the stage 19.
15. $KS = KS + 1$.
16. $MS(KS) = S^i$.
17. $KTC(S^i) = 1$.
18. $MTC(S^i, KTC(S^i)) = T(X^i)$.

19. If all code options are considered, then proceed to the stage 21.
20. Proceed to the stage 3.
21. End.

Table 1 shows an example of building a system of typical representatives for a group of Π transformations.

Table 1 – An example of building a system of typical representatives for a group of Π transformations

i	X^i	$T(X^i)$	S^i	MT	MS
1	01326457fb98aec	01326457fb98aec	1356	MT(1) = T(X1)	MS(1) = S1
2	01326457fb9dc8ae	01326457fb9dc8ae	1455	MT(2) = T(X2)	MS(2) = S2
3	01326457fb9dcea8	01326457fb9dcea8	1446	MT(3) = T(X3)	MS(3) = S3
4	01326457fba89dce	01326457fba89dce	1356	MT(4) = T(X1)	
5	01326457fbaec89d	01326457fbaec89d	1455	MT(5) = T(X5)	
6	01326457fbaecd98	01326457fbaecd98	1446	MT(6) = T(X6)	
7	01326457fd98ceab	01326457fd98ceab	1455	MT(7) = T(X7)	
8	01326457fd9ba8ce	01326457fd9ba8ce	1347	MT(8) = T(X8)	MS(4) = S8
9	0132a89bfd5764ce	01326457fd9ba8ce	1347	$T(X9) \in MT$	
10	0132abfd98c4576e	013267fd54c89bae	3336	MT(9) = T(X10)	MS(5) = S10
11	0132abfd98ce6457	013267fd54cea89b	2355	MT(10) = T(X11)	MS(6) = S11
12	0137feab98cd5462	0137feab98cd5462	2346	MT(11) = T(X12)	MS(7) = S12
13	01546237fd9baec8	01326457fb9dcea8	1446	$T(X13) \in MT$	
14	015467fb32a89dce	013267fd54c89bae	3336	$T(X14) \in MT$	
15	0154cdfb98a2376e	013267fd54c89bae	3336	$T(X15) \in MT$	
16	0154cdfb98ae6237	013267fd54cea89b	2355	$T(X16) \in MT$	
17	015d9b32a8cef764	01375d98c46efba2	3444	MT(12) = T(X17)	MS(8) = S17
18	015d9b37fe2a8c4	01375d9bfea8c462	2256	MT(13) = T(X18)	MS(9) = S18
19	015dfb98aec46732ae	0137fd54ce62ab98	2445	MT(14) = T(X19)	MS(10) = S19
20	015dfb98c46732ae	0137fd5462ab98ce	3345	MT(15) = T(X20)	MS(11) = S20
21	015dfb98c46ea237	0137fd5462aec89b	3345	MT(16) = T(X21)	
22	015dfb98cea23764	0137fd546ec89ba2	2445	MT(17) = T(X22)	
23	015dfe64c89ba237	0137fea2645dc89b	3345	MT(18) = T(X23)	
24	015dfe6732ab98c4	0137feab98cd5462	2346	$T(X24) \in MT$	
25	015dfea23764c89b	0137fec89ba2645d	3345	MT(19) = T(X25)	
26	0198a2376ec45dfb	0132645dcea89bf7	2355	MT(20) = T(X26)	
27	0198a23bf7546ecd	01326457fd98ceab	1455	$T(X27) \in MT$	
28	0198a23bf75dc46e	01326457fd9ba8ce	1347	$T(X28) \in MT$	
29	0198a23bf75dce64	01326457fd9baec8	1446	MT(21) = T(X29)	
30	0198a23bf7645dce	01326457fdc89bae	1356	MT(22) = T(X30)	
31	0198a23bf76ec45d	01326457fdcea89b	1257	MT(23) = T(X31)	MS(12) = S31
32	0198c45df76ea23b	01326457fdcea89b	1257	$T(X32) \in MT$	
33	019d573bfea264c8	01375d9bfea8c462	2256	$T(X33) \in MT$	
34	023bae64c89df751	01375d98c46efba2	3444	$T(X34) \in MT$	
35	023bae64cdf75198	01375d98cefba264	2445	MT(24) = T(X35)	
36	023bae67fd5198c4	01375d9bfea264c8	2346	MT(25) = T(X36)	
37	023bae67fd54c891	01375d9bfea8c462	2256	$T(X9)$	
38	023bae67fd9154c8	01375d9bfe62a8c4	2445	MT(26) = T(X38)	
39	023bae67fd98c451	01375d9bfe64c8a2	2445	MT(27) = T(X39)	
40	023bae67fdc45198	01375d9bfec8a264	2256	MT(28) = T(X40)	
41	02a8c46efb319d57	01326457fb98aec	1356	$T(X41) \in MT$	
42	045dce62a8913bf7	01375d98c462aefb	3336	MT(29) = T(X42)	
43	045dce62a89bf731	01375d98c46efba2	3444	$T(X43) \in MT$	
44	0462a8cef7315d9b	01326457fba89dce	1356	$T(X44) \in MT$	
45	08a264cefb9d5137	01326457fb9dc8ae	1455	$T(X45) \in MT$	

Table 2 shows the formed system of different typical representatives. Table 3 shows the binary codes and their characteristics: binary positional code X^1 , Gray code X^2 , binary code X^3 ($i = 17$ in Table 1), corresponding to the structure $S(X^3) = \{3,4,4,4\}$ and its typical representative X^4 . Fig. 1 shows the assessment of the options' balance. The number of changes in the values of each bit for these codes is as follows:

$H^1 = \{1,3,7,15\}$, $H^2 = \{1,2,4,8\}$, $H^3 = \{4,3,4,4\}$, $H^4 = \{4,4,3,4\}$. The values of the balance of the codes: $C^1 = 4,5$; $C^2 = 9$; $C^3 = 1,5$; $C^4 = 1,5$. A comparative analysis of the results shows that the type of code significantly affects its characteristics.

Thus, in the above example, the balance of the obtained typical representative X^4 is 6 times better than this indicator in the Gray code.

Table 2 – System of different typical representatives

KS	S	KTC (KS)	MTC(S)	KS	S	KTC (KS)	MTC(S)
1	1257	1	01326457fdcea89b	8	2355	1	013267fd54cea89b
2	1347	1	01326457fd9ba8ce			2	0132645dcea89bf7
3	1356	1	01326457fb98aecd	9	2445	1	0137fd54ce62ab98
		2	01326457fba89dce			2	0137fd546ec89ba2
		3	01326457fdc89bae			3	01375d98cefba264
4	1446	1	01326457fb9dcea8			4	01375d9bfe62a8c4
		2	01326457fbaecd98			5	01375d9bfe64c8a2
5	1455	3	01326457fd9baec8	10	3336	1	013267fd54c89bae
		1	01326457fb9dc8ae			2	01375d98c462aefb
		2	01326457fbaec89d	11	3345	1	0137fd5462ab98ce
3	01326457fd98ceab	2	0137fd5462aec89b				
6	2256	1	01375d9bfea8c462			3	0137fea2645dc89b
		2	01375d9bfec8a264			4	0137fec89ba2645d
7	2346	1	0137feab98cd5462	12	3444	1	01375d98c46efba2
		2	01375d9bfea264c8				

Table 3 – Binary codes and their features

X ¹				X ²				X ³				X ⁴			
x ¹ ₁	x ¹ ₂	x ¹ ₃	x ¹ ₄	x ² ₁	x ² ₂	x ² ₃	x ² ₄	x ³ ₁	x ³ ₂	x ³ ₃	x ³ ₄	x ⁴ ₁	x ⁴ ₂	x ⁴ ₃	x ⁴ ₄
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
0	0	1	0	0	0	1	1	0	1	0	1	0	0	1	1
0	0	1	1	0	0	1	0	1	1	0	1	0	1	1	1
0	1	0	0	0	1	1	0	1	0	0	1	0	1	0	1
0	1	0	1	0	1	1	1	1	0	1	1	1	1	0	1
0	1	1	0	0	1	0	1	0	0	1	1	1	0	0	1
0	1	1	1	0	1	0	0	0	0	1	0	1	0	0	0
1	0	0	0	1	1	0	0	1	0	1	0	1	1	0	0
1	0	0	1	1	1	0	1	1	0	0	0	0	1	0	0
1	0	1	0	1	1	1	1	1	1	0	0	0	1	1	0
1	0	1	1	1	1	1	0	1	1	1	0	1	1	1	0
1	1	0	0	1	0	1	0	1	1	1	1	1	1	1	1
1	1	0	1	1	0	1	1	0	1	1	1	1	0	1	1
1	1	1	0	1	0	0	1	0	1	1	0	1	0	1	0
1	1	1	1	1	0	0	0	0	1	0	0	0	0	1	0

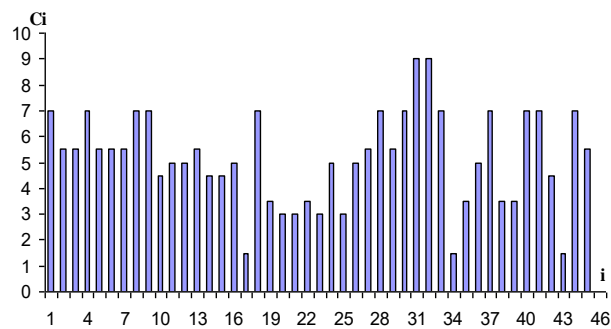


Fig. 1. Assessment of the options' balance

Conclusions from this study and prospects for further research in this area

The paper considers the actual problem of reducing power dissipation in global interconnection lines while maintaining high performance. It is shown that the switching activity of buses is the cause

of a significant share of the total power dissipation. One of the effective methods for reducing switching activity during device-to-device or system-on-chip communication is the use of low-power coding methods.

A method for generating energy-efficient codes and systems of typical representatives is proposed, which allows choosing the optimal coding without going through the options. Estimates of the number of typical structures are obtained and catalogs of typical representatives are formed.

The application of the developed method will allow analyzing and selecting codes with better properties and obtaining better results in terms of network delays, energy costs, and other design constraints for computer systems.

Further research in this area: development of a method for constructing codes with specified properties and a method for constructing code converters.

REFERENCE

1. Taha T. B., Barzinjy A. A., Hussain F. H. Nanotechnology and computer science: Trends and advances. Memories-Materials, Devices, Circuits and Systems. 2022. Vol. 2. P. 100011. DOI: <https://doi.org/10.1016/j.memori.2022.100011>

2. Samanth R., Nayak S. G., Nempu P. B. A Novel Multiply-Accumulator Unit Bus Encoding Architecture for Image Processing Applications. Iranian Journal of Electrical and Electronic Engineering. 2023. Vol. 19. №. 1. P. 1 – 11. DOI: <https://doi.org/10.22068/IJEEE.19.1.2391>
3. Chennakesavulu M., Prasad T. J., Sumalatha V. Data encoding techniques to improve the performance of system on chip. Journal of King Saud University-Computer and Information Sciences. 2022. Vol. 34. №. 2. P. 492–503. DOI: <https://doi.org/10.1016/j.jksuci.2018.12.003>
4. Huffman W. C., Kim J. L., Solé P. Concise encyclopedia of coding theory. Chapman and Hall/CRC, 2021. 998p. <https://doi.org/10.1201/9781315147901>
5. Serkov A. Noise stability of mobile telecommunication systems / A. Serkov, K. Trubchaninova, B. Lazurenko // Control, Navigation and Communication Systems. Academic Journal. – Poltava: PNTU, 2020. – VOL. 2 (60). – P. 169-172. DOI: <https://doi.org/10.26906/SUNZ.2020.2.169>
6. Mittal S., Nag S. A survey of encoding techniques for reducing data-movement energy //Journal of Systems Architecture. – 2019. – Vol. 97. – P. 373-396. DOI: <https://doi.org/10.1016/j.sysarc.2018.11.001>
7. Beletsky A. Ya. Generalized Gray Transforms and Synthesis Symmetric Systems of Walsh functions, Telecommunications and Radio Engineering - Vol. 78. - № 4, 2019, p.p. 305-326. DOI:10.1615/TelecomRadEng.v78.i4.30
8. Beletsky A. Conceptual Bases of Gray Transformations and Discrete Vilenkin-Crestenson Functions Systems //WSEAS Transactions on Computers. – 2023. – Vol. 22. – C. 44-56. DOI: 10.37394/23205.2023.22.5
9. Bouyuklieva S., Bouyukliev I., Bakoev V., Pashinska-Gadzheva, M. Generating m-Ary Gray Codes and Related Algorithms //Algorithms. – 2024. – Vol. 17. – №. 7. – C. 311. 331. -<https://doi.org/10.3390/a17070311>
10. Phillips A. T., Wick M. R. A dynamic programming approach to generating a binary reflected gray code sequence //Proc. 38th Annual Midwest Instruction and Computing Symposium. – 2005. 11p [RS 95]. Rasmussen D. J., Savage C. D. Gray Code Enumeration of Families of Integer Partitions./Journal of Combinatorial Theory, Series A. / – 1995. P. 201-229. [https://doi.org/10.1016/0097-3165\(95\)90090-X](https://doi.org/10.1016/0097-3165(95)90090-X)
11. Nimbe P., Yeng P., Okyere E., Weyori B., Adekoya A. Innovative Approaches to Classical and Quantum Reflected Binary Code Generation using Pascal Triangle, Reversible N-Input C-Gate and Reversible N-Input Q-Gate //Journal of Algebraic Statistics. – 2022. – Vol. 13. – №. 3. – P. 4708-4734.
12. Islam, M. N., & Foysal, A. B. (2009). Algorithms for generating binary reflected gray code sequence: Time efficient approaches. Proceedings - 2009 International Conference on Future Computer and Communication, ICFCC . – 2009.- № 79. - P. 79–83. <https://doi.org/10.1109/ICFCC.2009.41>
13. Mutze, T., & Nummenpalo, J. (2018). Efficient computation of middle levels gray codes. ACM Transactions on Algorithms, 14(2), 915–927. https://doi.org/10.1007/978-3-662-48350-3_76
14. Zhou Y., Panetta K., Chen C. (n, k, p)-Gray code for image systems //IEEE transactions on cybernetics. – 2013. – Vol. 43. – №. 2. – C. 515-529. doi: <https://10.1109/TSMCB.2012.2210706>
15. Wang, X., Su, Y., Zhang, H., & Zou, C. A new hybrid image encryption algorithm based on Gray code transformation and snake-like diffusion //The Visual Computer. – 2022. – P. 1-22. <https://doi.org/10.1007/s00371-021-02224-0>
16. Tahiri M., Karmouni H., Bencherqui A., Daoui A., Sayyouri M., Qjidaa H., Hosny K. New color image encryption using hybrid optimization algorithm and Krawtchouk fractional transformations //The Visual Computer. – 2023. – Vol. 39. – №. 12. – P. 6395-6420. <http://dx.doi.org/10.1007/s00371-022-02736-3>
17. Arora U. A review of mathematical methods for construction of Gray codes./GANITA, Vol.71(2), 2021, P 153-164.
18. Meenakshi G., Gupta S. Advanced level cyclic gray codes with application //International Journal of Electronics Communication and Computer Technology (IJECCT). – 2014. – Vol. 4. – №. 3. – P. 619-622.
19. Jacques, M., Wong, D., & Woo, K. (2020). Generating Gray codes for weak orders in constant amortized time. Discrete Mathematics, 343(10), 111992. doi:10.1016/j.disc.2020.111992
20. Yareshchenko V. Coding to reduce the energy of data movement / V. Yareshchenko, V. Kosenko // Системи управління, навігації та зв'язку. Збірник наукових праць. – Полтава: ПНТУ. – 2023. – Vol. 1 (71). – P. 159-162. DOI: <https://doi.org/10.26906/SUNZ.2023.1.159>.
21. Yareshchenko, V. V., Kosenko, V. V. (2023), “Low energy coding.” Electronic and mechatron systems: theory, innovation, practice: Coll. Sciences. BC according to the materials of IX All -Ukrainian. scientific-practical. conf., 10 November. Poltava: Nat. Univ. Yuri Kondratyuk. P. 67–68.

Received (Надійшла) 02.12.2024

Accepted for publication (Прийнята до друку) 19.02.2025

Метод формування енергоефективних кодів

В. В. Ярещенко, В. В. Косенко

Анотація. У статті розглядається актуальна проблема зниження потужності, що розсіюється, в глобальних лініях зв'язку при збереженні високої продуктивності. Висока комутаційна активність призводить до значних втрат через ємності зв'язку між довгими лініями. Вони є одним із самих енергоємних компонентів убудованої системи. Використання кодів Грея не тільки зменшує динамічні втрати енергії в адресній шині, але також мінімізує втрати зв'язку між близько розташованими лініями. Однак, код Грея має низьку збалансованість і велику кількість перемикачів бітів. Досліджено методи побудови альтернативних кодів Грея, що дозволяють зменшити кількість перемикачів на шинах. Мета статті полягає у розробці методу побудови кодів одичної відстані, визначення видів кодових перетворень та формування системи різних типових представників. Отримано оцінки їх кількості, визначено характеристики, сформовано каталоги типових представників. Застосування розробленого методу дозволить аналізувати та вибирати коди з найкращими властивостями та в результаті отримувати найкращі результати з погляду мережних затримок, витрат на електроенергію та інших конструктивних обмежень для комп'ютерних систем.

Ключові слова: кодування, коди одичних відстаней, енергоефективне кодування, комутаційна активність, еквівалентність.

АЛФАВІТНИЙ ПОКАЖЧИК

Бараніченко О. М.	67	Карпішен Б. С.	5	Почерняєв В. М.	196
Барковська О. Ю.	143	Кічата Н. М.	178	Проценко А. С.	124
Бірук Я. І.	162	Климчук А. В.	162	Ремська А. В.	178
Бреславець В. С.	205	Клівець С. І.	107	Решетнікова П. Е.	49
Бреславець Ю. В.	205	Коваленко А. А.	156	Резнік Д. В.	172
Бульба С. С.	116	Ковтун І. В.	209	Романенко А. О.	143
Бурдейна Н. Б.	165	Коломійцев О. В.	116	Савченко М. В.	201
Буслов П. В.	188	Кононенко А. І.	91	Сапальський О. А.	83
Гавращенко А. О.	143	Косенко В. В.	213	Сас Н.	37
Гаврилець Д. С.	172	Костікова М. В.	5	Серков О. А.	205
Глива В. А.	169	Костін К. Д.	150	Сивкова Н. М.	196
Голуб А. Ю.	129	Крук О. О.	72	Ситник О. В.	156
Гончаренко С. О.	72	Кулешов О. В.	107	Скакаліна О. В.	129
Гороховатський О. В.	77	Куценко Т. Г.	156	Соболь О. Р.	31
Гриньов Д. В.	31	Кучук Н. Г.	188	Тимошенко Д. О.	91
Гук А. С.	188	Лашко Є. Є.	172	Тиртишний Д. А.	111
Гурін Д. В.	42	Леонов С. Ю.	111	Тищенко Д.	37
Гусєва І. І.	209	Леві Л. І.	193	Третяков О. В.	178
Гуць В. В.	77	Лисиця Д. О.	116	Удянський Р. О.	188
Десятко А.	37	Магомєдова М. С.	196	Усиченко В. Д.	188
Довгановський М. О.	169	Малєєв Л. В.	55	Федорович О. Є.	55
Євсєєв В. В.	42	Мартовицький В. О.	138	Федорченко В. М.	124
Єрмілов Р. О.	62	Мезенцев М. В.	23	Франчук Т.	37
Єрмілова Н. В.	184	Молчанов Г. І.	116	Халмурадов Б. Д.	178
Заковоротний О. Ю.	49, 83	Невлюдов І. Ш.	42	Царьков Р. М.	62
Заповольський М. Й.	23	Нєронов С. М.	5	Чєнчєва О. О.	172
Запорожець О. В.	31	Ні Я. С.	143, 156	Чумак О. О.	17
Зачепило М. О.	87	Носко С. В.	116	Шаповал А. С.	156
Зиков І. С.	31	Оліфір М. В.	23	Шаповалова С. І.	67
Зима О. Є.	193	Орехов О. О.	10	Шєфєр О. В.	62
Зозуля Л. А.	165	Очерєтенко С. В.	5	Шиман М. В.	201
Зоураб Ю. Р.	184	Перєтяка Є. О.	143	Шостак А. В.	134
Зубко Р. В.	107	Петренко І. С.	172	Штомпєль М. А.	209
Іващенко Г. С.	91	Петруньок Т. Б.	165	Шупилюк М. В.	138
Ігнатюк С. О.	96	Плєхова Г. А.	5	Яковєнко І. В.	205
Ільїна І. В.	150	Положий Д. С.	10	Янко А. С.	72
Капітон А.	37	Попов А. В.	96	Ярєщєнко В. В.	213

Наукове видання

СИСТЕМИ УПРАВЛІННЯ, НАВІГАЦІЇ ТА ЗВ'ЯЗКУ

Збірник наукових праць

Випуск 1 (79)

Відповідальна за випуск *К. С. Нєстерєнко*Тєхнічний рєдактор *Т. В. Уварова*Корєктор *О. В. Морозова*Комп'ютерна верстка *Н. Г. Кучук, І. Ю. Петровська*Оформлення обкладинки *І. В. Ільїна*Ідєнтифікатор мєдіа R30-04135 згідно з рїшеннєм Національної ради України
з питань тєлєбачєння і радіомовлєння від 25.04.2024 № 1416

Підписано до друку 15.03.2025. Формат 60×84/8. Ум.-друк. арк. 27,25. Тираж 120 прим. Зам. 315-25

Адреса рєдакції: Україна, 36011, м. Полтава, Першотравнєвий проспєкт, 24, тєл. (050) 302-20-71

Національний університєт «Полтавська політєхніка імені Юрія Кондратюка»

Віддруковано з готових оригінал-макетів у цифровій друкарні Impress

61002, м. Харків, вул. Пушкінська, 56, тєл. + 38 (057) 714-52-11

e-mail: irina@impress.biz.ua