

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 3 (77)

Issue 3 (77)

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 06 вересня 2024 року № 10).*

*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 • 2024

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., Чернігів, 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

НАВІГАЦІЯ ТА ГЕОІНФОРМАЦІЙНІ СИСТЕМИ

<i>Пащенко Р. Е., Марюшко М. В.</i> Оцінка впливу яскравості космічних знімків сільськогосподарських культур на величини фрактальних розмірностей та індексів NDVI	5
--	---

АВТОМОБІЛЬНИЙ, РІЧКОВИЙ, МОРСЬКИЙ ТА АВІАЦІЙНИЙ ТРАНСПОРТ

<i>Любий Є. В., Дорофєєв К. І.</i> Практика використання передсигналів	11
<i>Неронов С. М., Плєхова Г. А., Очеретенко С. В.</i> Синергія автомобільного трансферу та утримання автомобільних доріг	16

УПРАВЛІННЯ В СКЛАДНИХ СИСТЕМАХ

<i>Андрусевич А. О., Фролов А. В., Стародубцев М. Г., Демська Н. П., Макаренко Г. С.</i> Оптимізація стратегії управління технологічними процесами в умовах невизначеності	20
<i>Boriak B., Yanko A., Ovsii D., Prorok M.</i> Improvement of the control system model of a mobile platform under the influence of electromagnetic spectrum threats in the information environment	25
<i>Заковоротний О. Ю., Орлова Т. О.</i> Використання нечітких множин другого типу для проектування регуляторів швидкості	29
<i>Заповольський М. Й., Мезенцев М. В., Оліфір М. В.</i> Адаптація алгоритму векторного керування для розробки системи управління електроприводом змінного струму	34
<i>Качанов П. О., Тарасенко М. В.</i> Система автоматичного регулювання рівня води в барабані котла енергоустановки	41
<i>Олімпієва Ю. І.</i> Забезпечення функціональної стійкості виробничих процесів промислових підприємств на основі нейроадаптивної системи	46
<i>Radovanović M., Petrovski A., Behlić A., Zied Chaari M., Hashimov E. G., Fellner R., Agbeyangi A. O.</i> Unleashing Autonomous Forces: Integrating AI-Driven Drones in Modern Military Strategy	55
<i>Худов Г. В., Хиженяк І. А., Грідасов І. Ю., Збежховська У. Р., Юзова І. Ю., Соломоненко Ю. С., Калімулін Т. М.</i> Аналіз тактико-технічних характеристик та тактики застосування існуючих ударних FPV-дронів	70

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

<i>Бешта В. С., Комаричев А. В., Філімончук Т. В., Бараней Д. І.</i> Модель мобільного додатку, яка орієнтована на обробку даних	80
<i>Білоконь А. С., Борисов С. О., Сальников С. С., Федорченко В. М.</i> Аналіз функціонування розподілених систем обробки та зберігання даних	84
<i>Бовчалюк С. Я., Гайдай Я. А.</i> Аналіз методу опорних векторів у порівнянні з традиційними методами передбачення ринкових рухів	89
<i>Бурда А. С., Прудіус М. А., Стефанюк Я. Г., Фомічов О. О.</i> Методи обробки та інтелектуального аналізу даних з використанням штучних імунних систем	93
<i>Волк М. О., Поповкін М. М.</i> Методи моделювання масштабованих хмарних ресурсів	97
<i>Галицька О. О., Бологова Н. М., Кібірев Д. О., Скиба О. В.</i> Огляд підходів до захисту тривимірних моделей від несанкціонованого розповсюдження	101
<i>Гороховатський В. О., Стяглик Н. І., Мазур Є. В., Вечірська А. Д.</i> Способи метричної грануляції для опису зображень у задачі класифікації	106
<i>Datsenko S., Tarasenko D.</i> Research of industrial internet of things networks and systems	113
<i>Дяченко В. О., Колісник Є. Б., Ляшова А. О., Можжаєв О. О.</i> Методи маршрутизації для забезпечення якості мережного сервісу	116
<i>Зубко І. С., Мартовицький В. О., Пунченко А. В., Карачевцев Д. Д.</i> Огляд методів нанесення цифрових водяних знаків для захисту зображень	121
<i>Kariton A., Dziuban O., Baranenko R., Sokol H.</i> Security of information technologies in a hybrid working environment	126
<i>Кочина А. А., Азімов К. Н.</i> Розробка системи автоматизованого парсингу разових замовлень на вантажні перевезення у міжміському сполученні	130
<i>Лактіонов О. І., Педченко Н. М., Янко А. С.</i> Практичні кейси створення згорткових моделей штучного інтелекту для задач розпізнавання образів	136
<i>Малєєва О. В., Набокін Д. А., Малєєва Ю. А.</i> Вибір найбільш ефективного маршруту перевезень з урахуванням індексів небезпечності територій в умовах воєнного часу	141
<i>Radchenko V., Kuchuk N.</i> Cloud infrastructure service delivery specifics	146
<i>Tyshchenko D., Franchuk T., Zakharov R., Moskalenko V.</i> Supporting dynamic security needs with VPN tools	149
<i>Ткаченко О. С., Мелешко Є. В., Міхав В. В.</i> Комп'ютерна модель поширення інформаційних вірусів у соціальній мережі при різній поведінці користувачів	153
<i>Фастовський Е. Г., Роговий А. І., Ахієзер О. Б., Фролов А. В., Артюх Р. В.</i> Інформаційна технологія аналізу та синтезу пояснених моделей штучного інтелекту на основі вербальних методів	158

<i>Zhou M., Kuchuk N.</i> A robust method for embedding and extracting watermarks in video using discrete cosine transform	166
<i>Шевчук Є. В., Федорченко В. М.</i> Аналіз основних вразливостей і способів захисту механізму консенсусу в децентралізованих блокчейн системах	170
<i>Шматко О. В., Кулініч Д. В., Горбач Т. В.</i> Розробка та дослідження архітектурної моделі системи обміну персональними даними на основі блокчейн	175
<i>Шматко О. В., Сафанков Д. В.</i> Розробка та впровадження хмарної платформи для токенизації персональних даних студентів	185
<i>Шостак А. В.</i> Про особливості формування вхідних даних у сіамській нейронній мережі	193
<i>Andrusenko Yu.</i> IAAS performance assessment with service levels	196

ЦИВІЛЬНА БЕЗПЕКА

<i>Бірук Я. І., Климчук А. В.</i> Засади проектування неоднорідних електромагнітних екранів	199
<i>Глива В. А., Краснянський Г. Ю., Довгановський М. О., Краснянський Т. М.</i> Розроблення і дослідження захисних властивостей матеріалів для блокування електромагнітних випромінювань інфрачервоного діапазону	203

ЗВ'ЯЗОК, ТЕЛЕКОМУНІКАЦІЇ ТА РАДІОТЕХНІКА

<i>Serkov A., Dzheniuk N., Kasilov O., Sokol G., Tolkachov M., Arutiunian D.</i> Intelligent wireless communication system	206
<i>Почерняєв В. М., Магомедова М. С., Сивкова Н. М.</i> Мобільні цифрові тропосферні станції з комбінуванням просторово-рознесених сигналів	211

АЛФАВІТНИЙ ПОКАЖЧИК	216
----------------------------------	-----

Організації авторів

Азербайджанський технічний університет, Баку, Азербайджан
 Національний університет оборони, Баку, Азербайджан
 Університет оборони, Військова академія, Белград, Сербія
 «Університет Гоце Делчев», Штіп, Скоп'є, Північна Македонія
 Військова академія «Генерал Михайло Апостольський», Скоп'є, Північна Македонія
 Управління досліджень, експериментів і розвитку Міністерства оборони Катару, Катар
 Пожежний університет, Варшава, Польща
 Кріслендський університет, Абеокута, штат Огун, Нігерія
 Компанія «Shanghai Polyether Rui Information Technology Co., Ltd», Шанхай, Китай

ДП "Південний державний ПК і НД інститут авіаційної промисловості", Харків, Україна
 Державний торговельно-економічний університет, Київ, Україна
 Державний університет інформаційно-комунікаційних технологій, Київ, Україна
 Інститут радіофізики та електроніки імені О. Я. Усикова НАН України, Харків, Україна
 Київський національний університет будівництва і архітектури, Київ, Україна
 Національна академія Служби безпеки України, Київ, Україна
 Київський фаховий коледж зв'язку, Київ, Україна
 Національний аерокосмічний університет імені М. Є. Жуковського «ХАІ», Харків, Україна
 Національний технічний університет «Харківський політехнічний інститут», Харків, Україна
 Національний університет «Полтавська політехніка імені Юрія Кондратюка», Полтава, Україна
 Харківський національний автомобільно-дорожній університет, Харків, Україна
 Харківський національний університет імені В.Н. Каразіна, Харків, Україна
 Харківський національний університет радіоелектроніки, Харків, Україна
 Харківський національний університет Повітряних Сил імені Івана Кожедуба, Харків, Україна
 Український католицький університет, Львів, Україна
 Уманський національний університет садівництва, Україна
 Університет науки, підприємництва та технологій, Україна
 Центральноукраїнський національний технічний університет, Кропивницький, Україна
 Черкаський державний технологічний університет, Україна

Authors affiliation

Azerbaijan Technical University, Baku, Azerbaijan
 National Defense University, Baku, Azerbaijan
 University of Defence, Military Academy, Serbia
 "Goce Delcev University" Stip, North Macedonia
 Military Academy "General Mihailo Apostolski", Skopje, North Macedonia
 Office of Research, Experiments, and Development, Ministry of Defense, Qatar
 Fire University, Warsaw, Poland
 Chrisland University, Abeokuta, Ogun State, Nigeria
 Shanghai Polyether Rui Information Technology Co., Ltd., Shanghai, China

SE "Southern State Institute of Aviation Industry", Kharkiv, Ukraine
 State University of Trade and Economics, Kiev, Ukraine
 State University of Information and Communication Technologies, Kyiv, Ukraine
 O. Ya. Usikov Institute for Radiophysics and Electronics, Kharkiv, Ukraine
 Kyiv National University of Construction and Architecture, Kyiv, Ukraine
 National Academy of Security Services of Ukraine, Kyiv
 Kyiv Professional College of Communication, Ukraine
 National Aerospace University named after M.E. Zhukovsky "KHAI", Kharkiv, Ukraine
 National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine
 National University "Yuri Kondratyuk Poltava Polytechnic", Poltava, Ukraine
 Kharkiv National Automobile and Highway University, Kharkiv, Ukraine
 V.N. Kharkiv National University Karazina, Kharkiv, Ukraine
 Kharkiv National University of Radio Electronics, Kharkiv, Ukraine
 Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine
 Ukrainian Catholic University, Lviv, Ukraine
 Uman National University of Horticulture, Ukraine
 Science Entrepreneurship Technology University, Ukraine
 Central Ukrainian National Technical University, Kropyvnytskyi, Ukraine
 Cherkasy State Technological University", Ukraine

Навігація та геоінформаційні системи

УДК 528.88 + 515.127.1

doi: 10.26906/SUNZ.2024.3.005

Р. Е. Пащенко¹, М. В. Марюшко²

¹ Інститут радіофізики та електроніки імені О. Я. Усикова НАН України, Харків, Україна

² Національний аерокосмічний університет імені М. Є. Жуковського «ХАІ», Харків, Україна

ОЦІНКА ВПЛИВУ ЯСКРАВОСТІ КОСМІЧНИХ ЗНІМКІВ СІЛЬСЬКОГОСПОДАРСЬКИХ КУЛЬТУР НА ВЕЛИЧИНІ ФРАКТАЛЬНИХ РОЗМІРНОСТЕЙ ТА ІНДЕКСІВ NDVI

Анотація. На якість проведення оцінки стану сільськогосподарських культур може впливати яскравість космічних знімків, що використовуються для моніторингу сільськогосподарських земель. **Предметом** дослідження є оцінка впливу яскравості космічних знімків на величини фрактальних розмірностей та індексів NDVI. **Об'єктом** дослідження є космічні знімки супутника Sentinel-2 сільськогосподарських культур з різною яскравістю. **Метою** є оцінка впливу яскравості космічних знімків полів, засіяних сільськогосподарськими культурами, на величини фрактальних розмірностей та індексів NDVI. **Отримані наступні результати.** Проведена оцінка максимальних і мінімальних значень яскравості при зміні яскравості вихідних знімків різних каналів (каналів b4 і b8) супутника Sentinel-2 та визначено, що для різних каналів супутника Sentinel-2 збільшення і зменшення яскравості по різному впливає на їх поведінку. Зміна діапазону яскравості на знімку при зміні яскравості також залежить від каналу супутника Sentinel-2. Досліджено вплив яскравості космічних знімків на величини фрактальних розмірностей та індексів NDVI. Визначено, що під час збільшення яскравості знімка середні значення індексів NDVI зменшуються. При цьому різниця у величині індексу NDVI відбувається у другому знаку після коми, що не дозволяє оцінювати стан посівів при великих значеннях яскравості за допомогою індексу NDVI. Під час зменшення яскравості середні значення індексів NDVI спочатку різко збільшуються, а потім досягають приблизно початкових значень, що дозволяє оцінювати стан посівів. Показано, що для знімка каналу b4 середні фрактальні розмірності під час збільшення яскравості спочатку збільшуються, а потім зменшуються, але значення змінюються не значно. При цьому різниця у величині фрактальної розмірності відбувається у третьому знаку після коми. Характер змін середніх фрактальних розмірностей під час зменшення яскравості знімка каналу b4 є не прогнозованим. Визначено, що середні значення фрактальних розмірностей знімка каналу b8 під час збільшення яскравості збільшуються монотонно і прогнозовано, а під час зменшення яскравості середні значення фрактальних розмірностей спочатку незначно зменшуються, а потім досягають початкових значень, що також можна прогнозувати. **Висновки.** Проведені дослідження показали, що застосування середніх фрактальних розмірностей для оцінки стану посівів дозволяє проводити аналіз космічних знімків при більшій зміні їх яскравості, порівняно з використанням середніх індексів NDVI.

Ключові слова: оцінка стану сільськогосподарських культур, яскравість космічного знімка, фрактальна розмірність, індекс NDVI.

Вступ

Проводити аналіз впливу природних, техногенних і антропогенних факторів на посіви сільськогосподарських культур у теперішній час дозволяє агроекологічний моніторинг [1]. Такий моніторинг сільськогосподарських земель спрямований на отримання інформації про стан і структуру земельних ділянок та посівних площ. Крім того, необхідна оперативна оцінка їх стану для дослідження динаміки проведення сільськогосподарських робіт.

Останнім часом для моніторингу земель сільськогосподарського призначення дуже широко застосовуються космічні засоби дистанційного зондування Землі (ДЗЗ). Такі дані ДЗЗ з'явилися у вільному доступі у мережі Інтернет. Крім того, такі дані ДЗЗ (космічні знімки) мають високу роздільну здатність. Так, у мережі Інтернет існує актуальна та постійно поповнювальна база космічних знімків супутника Sentinel-2 [2, 3]. Ширококутові мультиспектральні зображення (13 спектральних діапазонів) супутників Sentinel-2 мають просторову роздільну здатність 10 м в одному пікселі [3], які можна отримувати 2 рази на тиждень

на задану територію. Таким чином, космічні знімки супутника Sentinel-2 можна застосовувати для вирішення різноманітних задач моніторингу земель сільськогосподарського призначення.

Найбільш часто для моніторингу сільськогосподарських земель використовуються методи обробки даних ДЗЗ (космічних знімків) з застосуванням різноманітних вегетаційних індексів (існує близько 160 варіантів таких індексів) [4]. При цьому вегетаційні індекси визначаються емпірично і враховують фізичні властивості світла, відбиваючу здатність рослин та інші важливі характеристики. Найбільше розповсюдження отримав нормалізований різницевий вегетаційний індекс (NDVI).

Для розрахунку індексу NDVI необхідно дані з ближнього інфрачервоного та червоного каналів, що отримуються з багатоспектральної апаратури супутників ДЗЗ. В роботі [5] приведені результати досліджень змін індексу NDVI різночасових космічних знімків сільськогосподарських земель і вказується, що використання даних з двох спектральних каналів для їх розрахунку, приводить до ускладнення отримання результатів. В роботі також зазначається, що

індекс NDVI у значній мірі залежить від кількості активної біомаси та стає неефективним в періоди, коли рослинність є послабленою або спостерігається низький рівень вегетації.

Для підвищення інформативності космічних знімків під час моніторингу різноманітних поверхонь, зокрема земної поверхні, застосовуються методи фрактального аналізу зображень [6, 7]. Розрахунок фрактальних розмірностей цифрових зображень з різною складністю поверхонь та рослинності приводить до різних їх величин, що дозволяє оцінити складність поверхонь і рослинності. При цьому використовуються космічні знімки в одному діапазоні хвиль. Можливість застосування фрактального аналізу космічних знімків супутника Sentinel-2 для оцінки стану сільськогосподарських культур на різних фазах вегетації показана у роботах [8, 9]. У роботі [8] проаналізовані поля, засіяні кукурудзою, а у роботі [9] – полів, засіяних іншими культурами (соняшником, пшеницею, ячменем і гречкою). Але у цих роботах не розглядається питання оцінки впливу яскравості космічних знімків на величини фрактальних розмірностей.

Мета статті: оцінити вплив яскравості космічних знімків полів, засіяних сільськогосподарськими культурами, на величини фрактальних розмірностей та індексів NDVI.

Формування космічних знімків сільськогосподарських земель з різною яскравістю

Щоб оцінити вплив яскравості космічних знімків сільськогосподарських культур на величини фрактальних розмірностей та індексів NDVI необхідно створити набір космічних знімків одного й того ж поля з різною яскравістю. З використанням комп'ютерної програми Microsoft Word 2013 було штучно створено набір елементів космічного знімка поля, засіяного кукурудзою, станом на 31.07.2016. Космічні знімки, які застосовувалися для дослідження були у градаціях сірого у форматі *.bmp і мали розміри 56×56 пікселів для отримання однорідної структури поля на знімках.

Яскравість є світловою характеристикою тіл, відношення сили світла, яка випромінюється поверхнею, до площі її проєкції на площу, яка перпендикулярна віссі спостереження, іншими словами – це кількість білого кольору на зображенні. Чим вище яскравість, тим світлішим воно стає. Значення яскравості знаходяться у діапазоні від 0 до 255. Для того щоб яскравість можна було зменшувати та збільшувати, значення яскравості беруть у діапазоні від -255 до 255, а потім за формулою розраховують колір і переводять до діапазону від 0 до 255. Під час використання засобів настройки, програма Microsoft Word 2013 автоматично здійснює зміни в діапазоні яскравості.

Коротко розглянемо створення бази космічних знімків поля з різною яскравістю. Під час зміни яскравості космічного знімка здійснювалося його завантаження на сторінку програми Microsoft Word 2013. Після цього знімок виділявся і здійснювався перехід до вкладки «Робота з рисунками Формат». У групі «Змінити» (ліворуч на верхній панелі) вибиралась вкладка

«Яскравість». Після її відкриття вибиралося потрібне значення контрастності (у відсотках). Позитивні значення відсотків збільшували яскравість знімка, а негативні – зменшували. Після проведення цих операцій, змінений знімок (з іншою яскравістю) виділявся і він зберігався з іншою назвою у форматі .bmp.

На рис. 1, а - м наведено елементи космічних знімків каналу b4 супутника Sentinel-2 поля з кукурудзою станом на 31.07.2016 при зміні яскравості від 0% до +50% і від 0% до -50%. Як видно з аналізу рис. 1, а - є, при збільшенні яскравості від 0% до +50% елементи космічних знімків каналу b4, як і повинно бути, стають більш світлими, але градієнт збільшення світлого на знімках не великий (відтінки сірого змінюються не значно). При зменшенні контрастності від 0% до -50% (див. рис. 1, ж - м) зміна яскравості на космічних знімках практично не спостерігається візуально. Скоріше за все, це обумовлено структурою космічних знімків каналу b4 супутника Sentinel-2. Таким чином, збільшення світлості знімка під час збільшення яскравості є не великою, а під час зменшення яскравості зміна світлості знімка візуально не спостерігається. Така поведінка зміни світлості знімка пов'язана з однорідністю космічного знімка каналу b4 і наявністю переважно темних пікселів на знімку (діапазон яскравості на знімку від 22 до 28 градацій).

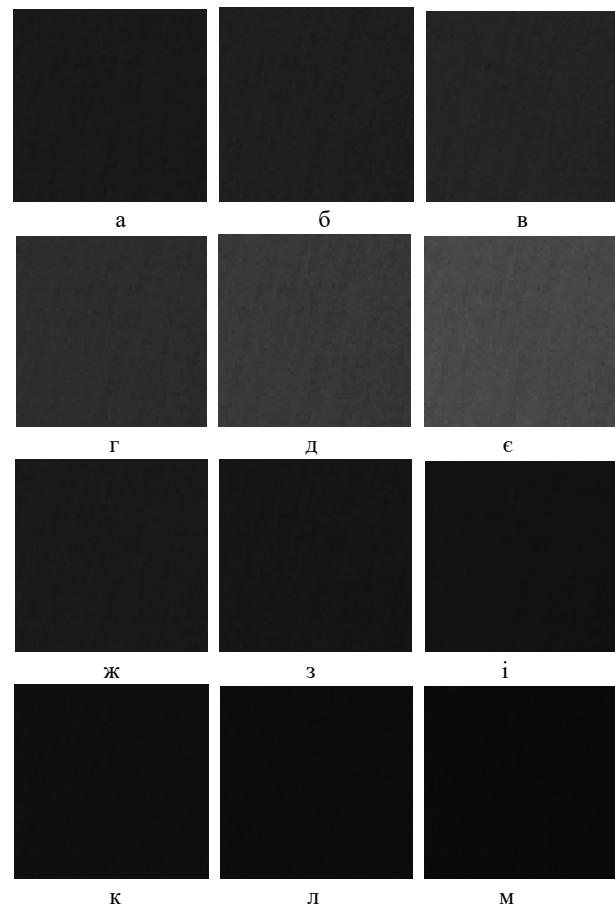


Рис. 1. Елементи космічних знімків (канал b4) поля з кукурудзою при зміні яскравості у Microsoft Word: 0% (а), +10% (б), +20% (в), +30% (г), +40% (д), +50% (є), 0% (ж), -10% (з), -20% (і), -30% (к), -40% (л), -50% (м)

На рис. 2, а - м наведено елементи космічних знімків каналу b8 супутника Sentinel-2 поля з кукурудзою станом на 31.07.2016 при зміні яскравості від 0% до +50% і від 0% до -50%. Аналіз зображень на рис. 2, а - є показує, що для знімка каналу b8 при збільшенні яскравості від 0% до +50% спостерігається значне збільшення світлості знімка, а для значення +50% знімок стає повністю білим (255 градацій), тобто його можна вважати аномальним. Таке збільшення яскравості приводить до того, що знімок не має корисної інформації про стан рослинності. Зменшення контрастності знімка каналу b8 від 0% до -50% (рис. 2, ж - м) спостерігається очікуване затемнення знімку, що приводить до його згладжування. Така поведінка знімку приводить до втрати деяких особливостей поля.

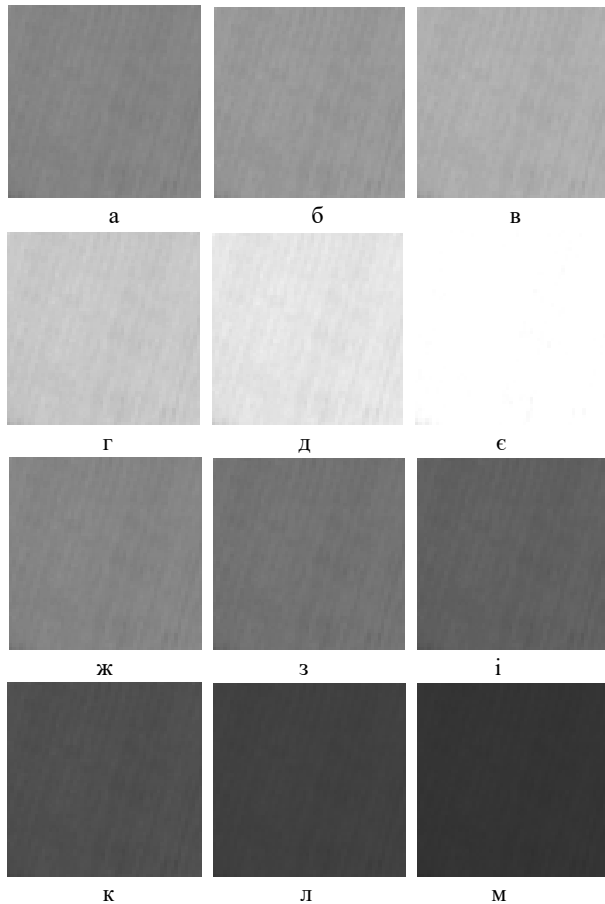


Рис. 2. Елементи космічних знімків (канал b8) поля з кукурудзою при зміні яскравості у Microsoft Word: 0% (а), +10% (б), +20% (в), +30% (г), +40% (д), +50% (е), 0% (ж), -10% (з), -20% (і), -30% (к), -40% (л), -50% (м)

Розглянемо як змінюються максимальні і мінімальні яскравості на знімку під час збільшення і зменшення його яскравості. У табл. 1 наведено максимальні, мінімальні яскравості та їх різниця (Δ) на знімках двох каналів супутника Sentinel-2 (b4 і b8) поля з кукурудзою станом на 31.07.2016 при збільшенні яскравості від 0% до +50% у програмі Microsoft Word. Як видно з аналізу даних табл. 1, при збільшенні яскравості знімків обох каналів (каналів b4 і b8) максимальні і мінімальні яскравості та їх різниця збільшуються, але для каналу b4 збільшення не таке велике (в 3,5 рази менше), як для каналу b8. Також, при збільшенні яскравості знімка до +50% у каналі b8 спостерігаються максимально можливі значення яскравості – 255 градацій яскравості, а різниця максимальних і мінімальних значень яскравості зменшується, тобто, як зазначалося вище, знімок є аномальним і має білий колір. З аналізу даних у табл. 1 також видно, що під час збільшення яскравості знімків обох каналів одночасно збільшуються, як максимальні, так і мінімальні яскравості. Але значення максимальних яскравостей збільшуються більше у порівнянні з мінімальними. Це характерно для знімків обох каналів (знімок каналу b8 з яскравістю +50% є аномальним). Такі зміни яскравостей на вихідних знімках може привести до змін величин індексів NDVI та фрактальних розмірностей.

У табл. 2 наведено максимальні, мінімальні яскравості та їх різниця (Δ) на знімках двох каналів супутника Sentinel-2 (b4 і b8) поля з кукурудзою станом на 31.07.2016 при зменшенні яскравості від 0% до -50% у програмі Microsoft Word.

У табл. 2 наведено максимальні, мінімальні яскравості та їх різниця (Δ) на знімках двох каналів супутника Sentinel-2 (b4 і b8) поля з кукурудзою станом на 31.07.2016 при зменшенні яскравості від 0% до -50% у програмі Microsoft Word.

Таблиця 1 – Максимальна і мінімальна яскравості та їх різниця на знімках поля з кукурудзою станом на 31.07.2016 при збільшенні яскравості

Яскравість знімку	Яскравість					
	канал b4			канал b8		
	Мін.	Макс.	Δ	Мін.	Макс.	Δ
0%	22	28	6	112	142	30
+10%	26	33	7	130	162	32
+20%	31	40	9	151	185	34
+30%	38	49	11	174	211	37
+40%	48	61	13	201	241	40
+50%	62	79	17	233	255	22

Таблиця 2 – Максимальна і мінімальна яскравості та їх різниця на знімках поля з кукурудзою станом на 31.07.2016 при зменшенні яскравості

Яскравість знімку	Яскравість					
	канал b4			канал b8		
	Мін.	Макс.	Δ	Мін.	Макс.	Δ
0%	22	28	6	112	142	30
-10%	19	24	5	95	123	28
-20%	16	20	4	80	105	25
-30%	13	17	4	66	87	21
-40%	11	14	3	55	69	14
50%	9	11	3	44	56	12

З даних табл. 2 видно, що під час зменшення яскравості знімків каналів b4 і b8 максимальні і мінімальні яскравості та їх різниця зменшуються, але характер змін для обох каналів має деякі особливості порівняно із збільшенням яскравості. Так, максимальні і мінімальні яскравості каналу b4 одночасно практично не змінюються, як це було під час збільшення яскравості знімків. Для каналу b8 також є особливості у поведінці максимальних і мінімальних яскравостей. При цьому мінімальні яскравості зменшуються з різною дискретністю, а максимальні – з однаковою (18 градацій яскравості). У табл. 2 також видно, що при зменшенні яскравості знімка каналу b4 різниця Δ практично однакова, діапазон яскравості змінюється у межах 3 – 5 градацій яскравості. Для каналу b8 різниця Δ зменшується моно-

тонно. При цьому знімок каналу b8 з яскравістю - 50% не є аномальним. Як зазначалося вище, це пов'язано, скоріше за все, різною структурою і яскравістю вихідних знімків каналів b4 і b8.

Таким чином, проведений аналіз показує, що збільшення і зменшення яскравості впливає на поведінку їх максимальних і мінімальних значень на знімках для різних каналів (каналів b4 і b8) супутника Sentinel-2. Зміна діапазону яскравості на знімках також залежить від каналу супутника Sentinel-2. Проведемо аналіз впливу цих змін на величини індексів NDVI і фрактальних розмірностей.

Вплив яскравості космічних знімків на величини фрактальних розмірностей та індексів NDVI

У теперішній час найчастіше для моніторингу сільськогосподарських земель використовуються вегетаційні індекси [4]. Більша частина вегетаційних індексів розраховується за допомогою двох діапазонів спектральної відбивної здатності рослин, які є найбільш стабільними і не залежать від інших чинників. Перший діапазон 0,62-0,75 мкм відповідає червоній зоні спектру, на якій спостерігається максимум поглинання сонячної радіації хлорофілом. Другий діапазон 0,75-1,3 мкм є ближньою інфрачервоною зоною спектру, де відбувається максимальне відбиття енергії клітинною структурою листа. Низькі значення коефіцієнтів відбиття в червоній зоні спектру і великі значення в ближній інфрачервоній зоні обумовлені, як правило, великою фітомасою рослинності та її високою фотосинтетичною активністю. Нормалізований різницевий вегетаційний індекс (NDVI) у теперішній час використовується у традиційних методах розрізнення рослинності від інших природних об'єктів. Нормалізований різницевий вегетаційний індекс розраховується як [4]:

$$NDVI = \frac{NIR - RED}{NIR + RED}, \quad (1)$$

де NIR – 760-900 нм (ближня інфрачервона зона спектру, канал b8 супутника Sentinel-2 (832 нм); RED – 630-690 нм (видима червона зона спектру), канал b4 супутника Sentinel-2 (665 нм).

Під час проведення фрактального аналізу космічних (цифрових) знімків спочатку проводиться розрахунок фрактальних розмірностей і потім будується поле фрактальних розмірностей. Існує велика кількість методів розрахунку фрактальних розмірностей, але на практиці для аналізу цифрових знімків найчастіше використовується метод покриття [10]. Цей метод забезпечує розрахунок розмірності Гаусдорфа – Безиковича, яка описується виразом:

$$D = \lim_{\varepsilon \rightarrow 0} \frac{\log N(\varepsilon)}{\log(1/\varepsilon)}, \quad (2)$$

де ε – довжина сторони куба (максимальна довжина дорівнює розміру «вікна»), яким покривають тривимірне представлення цифрового знімка; $N(\varepsilon)$ – кількість кубів, що накривають знімок.

Порядок визначення фрактальної розмірності за методом покриття докладно описано у [10], а у [7] – процес побудови поля фрактальних розмірностей. Під

час побудови поля фрактальних розмірностей проводиться сканування космічного знімка «вікном», розміром $n \times m$ пікселів, з кроком переміщення s (при $s = 1$ «вікно» є «ковзаючим», а при $s > 1$ – «стрибаючим»). Для кожного положення «вікна» розраховується числове значення фрактальної розмірності у «вікні» і записується в матрицю D , яка називається «полем фрактальних розмірностей». У роботі під час побудови поля фрактальних розмірностей розміри елемента космічного знімка вибрані 56×56 пікселів, а розміри «ковзаючого вікна» ($s = 1$) дорівнюють $n \times m = 4 \times 4$ пікселі. У роботах [5, 8] зазначалося, що найбільш інформативними під час оцінки стану сільськогосподарських культур є мінімальні, максимальні і середні значення індексів NDVI і фрактальних розмірностей.

Розглянемо вплив яскравості космічного знімка на мінімальні, максимальні і середні значення індексів NDVI і фрактальних розмірностей поля, засіяного кукурудзою. У табл. 3 наведено мінімальні, максимальні і середні значення індексів NDVI знімка поля з кукурудзою станом на 31.07.16 при різній яскравості. Як видно з аналізу даних табл. 3, під час збільшення яскравості знімка середні значення індексів NDVI зменшуються. При цьому різниця у величині індексу NDVI відбувається у другому знаку після коми (за виключенням збільшенні яскравості до +50%). Тобто зміни індексу NDVI при великих значеннях яскравості не дозволяють оцінювати стан посівів. Зміни мінімальних і максимальних значень індексів NDVI мають такий же характер, але мінімальні значення зменшуються не так швидко, як максимальні значення індексів NDVI.

Під час зменшення яскравості середні значення індексів NDVI спочатку різко збільшуються, а потім досягають приблизно початкових значень (при яскравостях +40% і +50%). Такі зміни індексу NDVI дозволяють оцінювати стан посівів під час зменшення яскравості знімка. Мінімальні значення індексів NDVI під час зменшення яскравості поведуться не регулярно, спочатку зменшуються, потім збільшуються, далі зменшуються і знову збільшуються. Максимальні значення індексів NDVI поведуться регулярно, спочатку збільшуються, а потім монотонно зменшуються.

У табл. 4 наведено мінімальні, максимальні і середні значення фрактальних розмірностей (каналів b4 і b8) знімка поля з кукурудзою станом на 31.07.16 при різній яскравості. З даних табл. 4 видно, що характер змін мінімальних, максимальних і середніх значень фрактальних розмірностей для каналів b4 і b8 знімка поля з кукурудзою відрізняються між собою. Для знімка каналу b4 середні фрактальні розмірності під час збільшення яскравості спочатку збільшуються, а потім зменшуються, але значення змінюються не значно. При цьому різниця у величині фрактальної розмірності відбувається у третьому знаку після коми. Мінімальні і максимальні фрактальні розмірності знімка каналу b4 при збільшенні яскравості практично не змінюються. Під час зменшення яскравості знімка каналу b4 середні фрактальні розмірності спочатку зменшуються, а потім збільшуються і знову зменшуються.

Таблиця 3 – Значення індексу NDVI знімка поля з кукурудзою станом на 31.07.16 при різній яскравості

Яскравість знімку	Індекс NDVI		
	NDVI _{мін}	NDVI _{макс}	NDVI _{сеп}
0%	0,612	0,701	0,679
+10%	0,605	0,688	0,669
+20%	0,598	0,674	0,656
+30%	0,575	0,655	0,637
+40%	0,546	0,628	0,609
+50%	0,508	0,578	0,563
0%	0,612	0,701	0,679
-10%	0,610	0,708	0,686
-20%	0,616	0,707	0,686
-30%	0,610	0,706	0,682
-40%	0,618	0,704	0,677
-50%	0,600	0,697	0,680

Таблиця 4 – Значення фрактальної розмірності знімка поля з кукурудзою станом на 31.07.16 при різній яскравості

Яскравість знімку	Фрактальна розмірність					
	канал b4			канал b8		
	D _{мін}	D _{макс}	D _{сеп}	D _{мін}	D _{макс}	D _{сеп}
0%	2,924	2,998	2,963	2,958	2,996	2,982
+10%	2,921	2,997	2,964	2,960	2,997	2,983
+20%	2,923	2,997	2,969	2,964	2,997	2,985
+30%	2,923	2,998	2,966	2,965	2,998	2,986
+40%	2,923	2,998	2,963	2,967	2,998	2,986
+50%	2,926	2,998	2,966	2,991	3,000	3,000
0%	2,944	2,994	2,973	2,958	2,996	2,982
-10%	2,910	2,998	2,956	2,957	2,996	2,981
-20%	2,919	3,000	2,970	2,949	2,998	2,979
-30%	2,904	3,000	2,965	2,943	2,998	2,977
-40%	2,907	3,000	2,963	2,953	2,998	2,982
-50%	2,885	3,000	2,952	2,957	3,000	2,982

Різниця у величині фрактальної розмірності відбувається у другому знаку після коми. Тобто характер змін середніх фрактальних розмірностей не прогнозований. Мінімальні фрактальні розмірності знімка каналу b4 змінюються, як і середні фрактальні розмірності. Максимальні фрактальні розмірності під час зменшення яскравості спочатку збільшуються, а потім не змінюються.

Таким чином, розрахунок фрактальних розмірностей знімка каналу b4 під час збільшення яскравості знімка дає прогнозовані результати, але під час зменшення яскравості фрактальні розмірності поведуться не прогнозовано.

Середні значення фрактальних розмірностей знімка каналу b8 під час збільшення яскравості збільшуються монотонно і прогнозовано. При цьому різниця у величині фрактальної розмірності відбувається у третьому знаку після коми (за виключенням збільшенні яскравості до +50%). Мінімальні фрактальні розмірності знімка каналу b8 збільшуються прогнозовано (за виключенням знімка з яскравістю +50%), а максимальні – практично не змінюються..

Під час зменшення яскравості знімка каналу b8 середні значення фрактальних розмірностей спочатку незначно зменшуються, а потім досягають початкових значень. Різниця у величині фрактальної розмірності відбувається у третьому знаку після коми.

Мінімальні фрактальні розмірності знімка каналу b8 змінюються, як і середні фрактальні розмірності. Максимальні значення фрактальних розмірностей практично не відрізняються при різних значеннях яскравості знімка. Таким чином, під час збільшення або зменшення яскравості знімка каналу b8 значення фрактальних розмірностей знімка поведуться прогнозовано, а різниця у величині фрактальної розмірності відбувається у третьому знаку після коми.

Можна зробити загальний висновок, що при великих значеннях яскравості зміни індексу NDVI не дозволяють оцінювати стан посівів (різниця у величині індексу NDVI відбувається у другому знаку після коми). Але під час зменшення яскравості знімка зміни індексу NDVI дозволяють оцінювати стан посівів. Застосування середніх фрактальних розмірностей для оцінки стану посівів дозволяє проводити аналіз космічних знімків при більшій зміні їх яскравості, порівняно з використанням середніх індексів NDVI. Для оцінки стану посівів під час зміни яскравості знімків доцільно застосовувати космічні знімки каналу b8 супутника Sentinel-2 для яких розраховуються середні фрактальні розмірності.

Висновки

Для моніторингу земель сільськогосподарського призначення дуже широко застосовуються космічні засоби ДЗЗ. У теперішній час дані ДЗЗ з'явилися у вільному доступі у мережі Інтернет і мають високу роздільну здатність. Різна яскравість космічних знімків може впливати на якість проведення оцінки стану сільськогосподарських культур.

Під час оцінки стану сільськогосподарських земель найбільш часто використовуються методи обробки космічних знімків з застосуванням різноманітних вегетаційних індексів, зокрема нормалізованого різницевого вегетаційного індексу NDVI. Але для розрахунку індексу NDVI необхідно дані з ближнього інфрачервоного та червоного каналів, що отримуються з багатоспектральної апаратури супутників ДЗЗ. Для підвищення інформативності космічних знімків під час моніторингу сільськогосподарських культур доцільно застосувати фрактальний аналіз, який базується на побудові поля фрактальних розмірностей. При цьому можуть використовуватися данні тільки з одного каналу супутника Sentinel-2 (канал b8). Різна яскравість космічних знімків полів, засіяних сільськогосподарськими культурами, приводить до необхідності проведення дослідження її впливу на величини фрактальних розмірностей та індексів NDVI.

Проведений аналіз показав, що збільшення і зменшення яскравості космічних знімків впливає на поведінку їх максимальних і мінімальних значень на знімках для різних каналів (каналів b4 і b8) супутника Sentinel-2. Зміна діапазону яскравості на знімках також залежить від каналу супутника Sentinel-2.

Під час збільшення яскравості знімка середні значення індексів NDVI зменшуються. При цьому різниця у величині індексу NDVI відбувається у другому знаку після коми (за виключенням збільшенні яскравості до +50%). Тобто зміни індексу NDVI при великих значеннях яскравості не дозво-

лять оцінювати стан посівів. Під час зменшення яскравості середні значення індексів NDVI спочатку різко збільшуються, а потім досягають приблизно початкових значень (при яскравостях +40% і +50%). Такі зміни індексу NDVI дозволяють оцінювати стан посівів під час зменшення яскравості знімка.

Для знімка каналу b4 середні фрактальні розмірності під час збільшення яскравості спочатку збільшуються, а потім зменшуються, але значення змінюються не значно. При цьому різниця у величині фрактальної розмірності відбувається у третьому знаку після коми. Середні фрактальні розмірності під час зменшення яскравості знімка каналу b4 спочатку зменшуються, а потім збільшуються і знову зменшуються, тобто характер змін середніх фрактальних розмірностей не прогнозований.

Середні значення фрактальних розмірностей знімка каналу b8 під час збільшення яскравості збільшуються монотонно і прогнозовано. При цьому

різниця у величині фрактальної розмірності відбувається у третьому знаку після коми (за виключенням збільшення яскравості до +50%). Під час зменшення яскравості знімка каналу b8 середні значення фрактальних розмірностей спочатку незначно зменшуються, а потім досягають початкових значень, що також можна прогнозувати. Застосування середніх фрактальних розмірностей для оцінки стану посівів дозволяє проводити аналіз космічних знімків при більшій зміні їх яскравості, порівняно з використанням середніх індексів NDVI. Для оцінки стану посівів під час змін яскравості знімків доцільно застосовувати космічні знімки каналу b8 супутника Sentinel-2 для яких розраховуються середні фрактальні розмірності. Під час подальших досліджень, доцільно здійснити оцінку впливу контрастності космічних знімків полів, засіяних сільськогосподарськими культурами, на величини фрактальних розмірностей та індексів NDVI.

СПИСОК ЛІТЕРАТУРИ

1. Тараріко О.Г., Сиротенко О.В., Льєнко Т.В., Кучма Т.Л. Агроєкологічний супутниковий моніторинг. К.: Аграр. наука, 2019. 204 с.
2. Copernicus Europe's eyes on Earth, Sentinel-2. Copernicus Europe's eyes on Earth [Електронний ресурс]. – Режим доступу: <https://www.copernicus.eu/en/about-copernicus/infrastructure/discover-our-satellites> – 09.05.2024 р.
3. ESA Sentinel Online, Spatial Resolution. ESA Sentinel Online [Електронний ресурс]. – Режим доступу: <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial> – 09.05.2024 р.
4. Черепанов А.С. Вегетационные индексы. Геоматика. 2011. № 2(11). С. 98 – 102.
5. Марюшко М.В., Пащенко Р.Е., Коблюк Н.С. Моніторинг сільськогосподарських культур із застосуванням космічних знімків SENTINEL-2. Радіоелектронні і комп'ютерні системи. 2019. №1(89). С. 99 – 108. DOI: 10.32620/reks.2019.1.11.
6. Feder J. Fractals. New York: Springer US, 1988. 263 p. DOI: <https://doi.org/10.1007/978-1-4899-2124-6>.
7. Доля Г.Н., Иванов В.К., Кучук Г.А., Пащенко Р.Э. и др. Фрактальный анализ процессов, структур и сигналов / Под. ред. Р.Э. Пащенко. Х.: “НЭО Экоперспектива”, 2006. 348 с.
8. Марюшко М.В., Пащенко Р.Е. Фрактальный анализ космических снимков SENTINEL-2 для мониторинга сельскохозяйственных культур. Радіоелектронні і комп'ютерні системи. 2020. №4(96). С. 34 – 47. DOI: <https://doi.org/10.32620/reks.2020.4.03>.
9. Пащенко Р.Е., Марюшко М.В. Оцінка стану різних сільськогосподарських культур з використанням фрактального аналізу. Сучасні інформаційні системи. 2023. Том. 7, № 3. С. 81 – 88. DOI: <https://doi.org/10.20998/2522-9052.2023.3.12>
10. Кроновер Р.М. Фракталы и хаос в динамических системах. М.: Постмаркет, 2000. 352 с.

Received (Надійшла) 20.03.2024

Accepted for publication (Прийнята до друку) 27.06.2024

Estimation influencing brightness of agricultural cultures spaces pictures on sizes fractals dimensions and indexes NDVI

Ruslan Pashchenko, Maksim Mariushko

Abstract. Spaces pictures in-use for monitoring agricultural earths can have a different contrast that can influence on quality through estimation of the state agricultural cultures. The subject of the study in the article is estimation of influencing brightness spaces pictures on the sizes fractals dimensions and indexes NDVI. The object of the study is the spaces pictures of agricultural cultures satellite Sentinel-2 with a different brightness. The goal is estimation influencing brightness spaces pictures of the fields sown by agricultural cultures, on the sizes fractals dimensions and indexes NDVI. The following results were obtained. Conducted estimation maximal and minimum values of brightness at the change brightness initial pictures of the different channel (channel b4 and b8) satellite Sentinel-2 and it is certain that for the different ducting of satellite Sentinel-2 the increase and diminishing of brightness on different influences on their conduct. Change range of brightness on a picture at the change brightness also depends on the channel of satellite Sentinel-2. Influence brightness spaces pictures is investigational on the sizes fractals dimensions and indexes NDVI. It is certain that during the increase brightness picture the mean values indexes of NDVI diminish. Thus a difference in the size index of NDVI takes place in the second sign after a comma that does not allow to estimate the state of sowing at the large values brightness by the index NDVI. During diminishing brightness the mean values of indexes NDVI are at first sharply multiplied, and then achieve the initial values approximately, that allows to estimate the state of sowing. Show that for the picture channel b4 middle fractals dimension during the increase of brightness at first is increased, and then diminish, but the values change not considerably. Thus a difference in the size fractal dimension takes place in the third sign after a comma. Character changes of middle fractals dimensions during diminishing brightness of picture channel b4 it is not forecast. It is certain that mean values of fractals dimensions picture channel b8 during the increase brightness is increased droningly and forecast, and during diminishing brightness mean values of fractals dimensions at first insignificantly, and then achieve the initial values, that it is also possible to forecast. **Conclusions.** The conducted researches showed that application average fractals dimensions for estimation of the state sowing allows to conduct the analysis spaces pictures at the greater change their brightness as compared to the use average indexes NDVI.

Keywords: estimation the state of agricultural cultures, brightness of space picture, fractal dimension, index NDVI.

Автомобільний, річковий, морський та авіаційний транспорт

УДК 656.051

doi: 10.26906/SUNZ.2024.3.011

Є. В. Любий, К. І. Дорофєєв

Харківський національний автомобільно-дорожній університет, Харків, Україна

ПРАКТИКА ВИКОРИСТАННЯ ПЕРЕДСИГНАЛІВ

Анотація. У представленій статті проаналізовано досвід використання передсигналів для вирішення проблемного питання щодо ефективного управління транспортними потоками на регульованих перехрестях. За результатами проведеного аналізу встановлено, що основними напрямками використання передсигналів як засобів управління транспортними потоками на міських магістралях є пріоритизація руху громадського транспорту на підході до перехрестя та вже безпосередньо на ньому, а також оптимізація світлофорного регулювання локальних перехресть або групи перехресть. Основними критеріями щодо оцінювання заходів щодо підвищення ефективності управління транспортними потоками на регульованих перехрестях з використанням передсигналів можна вважати транспортні затримки, витрати часу користувачів громадського транспорту, рівень обслуговування перехрестя та довжина черги транспортних засобів. Основними засобами, що використовуються для оцінювання ефективності роботи світлофорних систем з використанням передсигналів є імітаційне моделювання, для чого застосовуються найпопулярніші серед фахівців галузі програмні продукти – VISSIM, SUMO, AIMSUN. Потенційним напрямом подальших досліджень є розробка принципів вибору і обґрунтування місць розташування передсигналів на міських магістралях.

Ключові слова: дорожній рух, світлофорне регулювання, регульовані перехрестя, передсигнал, транспортні затримки, пропускна здатність.

Вступ

Постановка завдання. Підвищення рівня автомобілекористування міського населення впливає на рівень і якість транспортного обслуговування учасників дорожнього руху, особливо в міських умовах. Міські перехрестя є одним із основних місць концентрації конфліктних ситуацій для учасників дорожнього руху і головним джерелом виникнення транспортних затримок. На даний час вже відомо багато практик щодо скорочення затримок транспортних потоків як на регульованих, так і на нерегульованих перехрестях, впровадження яких дозволяє покращити якість транспортного обслуговування різних учасників дорожнього руху.

Відомо, що левова частка затримок транспортних засобів за умов руху в міських умовах відбувається саме на регульованих перехрестях, тому і увага фахівців (вчених і практиків) більше прикута саме до вирішення проблеми взаємодії транспортних потоків на цих елементах міської вулично-дорожньої мережі, як в локальному вигляді (окремі перехрестя), так і системному (група перехресть). Основними засобами мінімізації транспортних затримок на регульованих перехрестях є: визначення оптимальних світлофорних циклів; впровадження координації роботи світлофорних об'єктів; використання елементів адаптивних і інтелектуальних систем управління транспортними потоками; застосування передсигналів. Використання останніх, на даний час, не є розповсюдженою практикою, що пов'язано, в більшому ступені, з їх непопулярністю і необхідністю додаткового обґрунтування, а також специфікою застосування – в основному, використовуються в системах або на елементах вулично-дорожньої мережі, де існує потреба в наданні пріоритетного руху, наприклад, надання

пріоритету в русі міському громадському транспорту або спецтранспорту. Але не дивлячись на це доцільно більш глибоко розглянути практику використання передсигналів з точки зору покращення якості рівня транспортного обслуговування учасників дорожнього руху на регульованих перехрестях або підходах до них, що і є **основною метою** даної статті.

Виклад основного матеріалу

Загальна концепція використання передсигналів була запроваджена для скорочення втрат часу на регульованих перехрестях, а перші передсигнали були апробовані та встановлені в 1954 році в Дюссельдорфі [1]. Слід розуміти, що на практиці реалізація передсигналів є обмеженою. Це можна пов'язати із необхідністю виділення додаткової території на вулично-дорожній мережі під їх облаштування, що є надскладною задачею для сучасних міст з їх усталеним плануванням і забудовою. Але в останні часи почало з'являтися все більше прикладів впровадження систем з передсигналами, особливо це стосується міст країн азійського регіону, де вирішуються проблеми пріоритизації руху громадського транспорту на магістралях.

В сучасній практиці ефективної організації дорожнього руху найчастіше під передсигналами розуміють додаткові світлофори, які розміщуються перед регульованими перехрестями та забезпечують певний рівень пріоритету для руху окремих систем транспорту (громадських транспорт, транзитні транспортні потоки, спецтранспорт), дозволяючи їм оминати черги з транспортних засобів, що виникають перед цими перехрестями [2]. Але це не єдиний можливий варіант використання передсигналів у галузі управління транспортними потоками.

Передсигнали можуть використовуватись для підвищення ефективності координованого управ-

ліній транспортними потоками. У такому разі передсигнали встановлюються з метою підтримки параметрів групи транспортних засобів, що рухаються в скоординованих системах на магістралях, для зберігання структури і динамічних властивостей пачки зеленої хвилі. А також можуть застосовуватись для коригування подачі транспортних засобів, що прибувають до скоординованої магістралі або її ділянки з другорядних напрямків або проїздів (виїздів з прилеглих територій). В такому разі, в основному, передбачається встановлення передсигналів з організацією стоп-ліній поза перехрестя, використовуючи надлишкову тривалість зеленого сигналу.

Поширене розповсюдження отримали передсигнали для регулювання дорожнього руху на підходах і в межах залізничних переїздів. Метою встановлення передсигналу в цьому випадку є зменшити, а в оптимальному варіанті – ліквідувати ймовірність зупинки транспортних засобів на залізничній колії під час червоної фази циклу світлофору [3]. До того ж можна знайти приклади використання передсигналів для управління рухом велосипедистів [4, 5], хоча попит на велосипеди часто є занадто низьким, щоб виправдати такий управління дорожнім рухом [6].

Отже, з представленого вище, можна констатувати, що основними напрямками удосконалення організації дорожнього руху із використанням передсигналів є:

- пріоритезація руху громадського транспорту в межах регульованих перехресть на міських магістралях;
- підвищення ефективності світлофорного регулювання локальних перехресть або групи перехресть (координації роботи світлофорних об'єктів) на міських магістралях;
- та інші випадки, що зустрічаються значно рідше, до яких можна віднести – регулювання руху транспортних засобів і забезпечення його безпеки в межах залізничних переїздів; організація ефективної взаємодії вело- і транспортних потоків; надання пріоритету в русі спецтранспорту.

В подальшому матеріалі статті будуть представлені та описані приклади використання передсигналів для перших двох напрямків, оскільки вони характеризуються найбільшим поширенням і проробкою їх теоретичної та практичної складової.

Пріоритезація руху громадського транспорту. Надання громадському транспорту пріоритету є важливим і необхідним заходом для підвищення його привабливості та скорочення транспортних затримок у містах. Слід розуміти, що досягнення максимального скорочення затримок автобусного сполучення позитивно вплине й на загальний міський трафік [7].

Надання пріоритету громадському транспорту у поєднанні з використанням передсигналів можна вважати типовою формою управління пріоритетом автобусів, яка вивчалася і вивчається багатьма дослідниками всьому світу [8]. Вперше концепція використання передсигналів у цьому напрямку була висунута Міністерством транспорту Британії в 1991 р. [7].

У свою чергу, перший ґрунтовний теоретичний аналіз використання передсигналів для вирішення

проблем з організацією пріоритетного руху громадського транспорту був представлений авторами роботи [9], але запропонована ними реалізація передбачала постійну роботу передсигналу незалежно від часу прибуття автобусу. Типова схема передсигналу, що використовується з метою надання пріоритету громадському транспорту представлена на рис. 1.

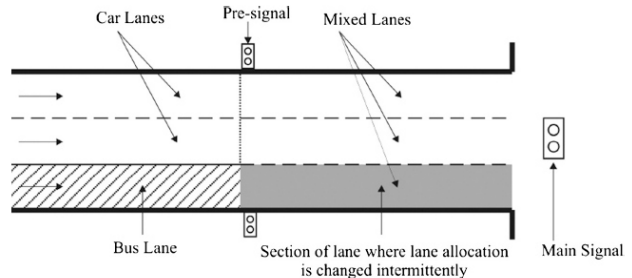


Рис. 1. Типова схема встановлення передсигналу для надання пріоритету руху громадському транспорту [10]

Автори роботи [10] проводили оцінювання переваг використання послідовних (групи) перехресть рпи вирішенні задачі надання пріоритету в русі громадському транспорту. За цільову функцію використано затримки і запропоновано метод оптимізації синхронізації сигналів і розподілу смуг руху. Результати дослідження засвідчують, що реалізація запропонованого методу може скоротити затримку транспортних засобів і значно покращити пропускну здатність перехресть. Для реалізації пріоритету громадського транспорту в роботі [11] запропоновано алгоритм передсигнального адаптивного управління трафіком для задоволення потреб приватного та громадського транспорту, що працює в режимі реального часу.

Автори іншої роботи [12] також досліджувалось питання ефективності надання пріоритету руху автобусам з використанням передсигналу. В наступній роботі [13] авторами було удосконалено координаційні зв'язки основного і передсигналу, а також склали криві прибуття та відправлення транспортного засобу відповідно до правил прибуття та відправлення на основний сигнал. До того ж у роботі запропоновано метод розрахунку зміни затримки громадського транспорту та інших транспортних засобів до і після встановлення передсигналу.

Оптимізації стратегії скоординованого управління основного сигналу і передсигналу для пріоритизації руху громадського транспорту присвячена робота [14] Для отримання результатів внесених в координацію змін авторами змодельовані різні ситуації за допомогою програмного забезпечення VISSIM.

В роботі [8] авторами запропоновано використовувати передсигнали для надання пріоритету громадському транспорту на підході до регульованого перехрестя, за рахунок чого у цих транспортних засобів виникає можливість обирати будь-яку смугу з переліку наявних для продовження руху. Встановлено, що після впровадження tandem design на перехресті затримка автобусів зменшується на 27,9%, а затримка транспортних засобів, що повертають ліворуч, на 5,9%.

Нову схему динамічного розподілу смуги руху автобусів запропонували автори [15], що дозволяє оптимізувати обрані параметри, мінімізуючи загальну затримку (включаючи затримку автобуса і затримку транспортних засобів), схема представлена на рис. 2. Недоліком статті є відсутність практичної реалізації запропонованої схеми.

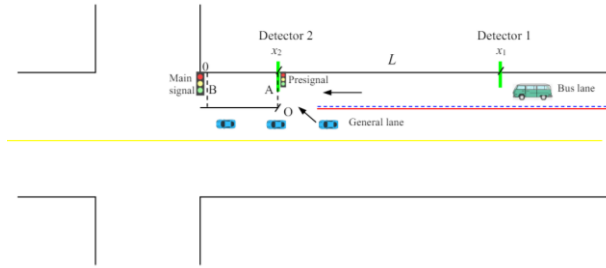


Рис. 2. Схема проектування динамічного розподілу для смуги громадського транспорту [15]

В статті [16] представлено принципи оптимізації управління передсигналами, щодо налаштування сигналів на основі мінімізації загальних витрат часу користувачів транспорту. Розроблено модель для опису динаміки черги для змішаних смуг руху після світлофора, а також для смуг для автобусів та автомобілів перед світлофором. З використанням AIMSUN виконано тестування, аналіз і оцінку ефективності запропонованих принципів оптимального управління, де враховано стохастичні прибуття транспортних засобів до перехрестя.

Автори [17] запропонували метод оцінювання довжини черги транспортних засобів, перед перехрестями, де використовуються передсигнали для надання пріоритету руху автобусам. Цей метод базується на використанні теорії ударних хвиль. Результати імітаційних експериментів показують, що запропонований метод на основі ударних хвиль є ефективним для оцінки довжини черги, пов'язаної з передсигналами для надання пріоритету автобусам.

В роботі [18] представлено теоретичні дослідження щодо доцільності використання передсигналів для збільшення пропускної здатності перехресть з автомобільним і автобусним рухом, які ґрунтуються на припущенні про те, що час, необхідний для роз'їзду сукупності класів транспортних засобів, дорівнює сумі часу, який би знадобився для роз'їзду кожного класу транспортних засобів окремо. Стаття не містить прикладу практичної реалізації запропонованої теорії. Результати імітаційного експерименту в роботі [19] також підтверджують ефективність використання передсигналів для пріоритизації руху автобусів на перехресті. Додатково авторами були змодельовані ситуації з середнім і високим рівнем інтенсивності трафіку, за яких використання розробленого алгоритму управління світлофорними об'єктами дозволяє скоротити загальні транспортні затримки на 30,1%.

Управління дорожнім рухом на локальних перехрестях і в скоординованих системах з використанням передсигналів. Іншим потужним напрямом застосування передсигналів у галузі організації

дорожнього руху є забезпечення ефективного руху транспортних засобів на елементах міської вулично-дорожньої мережі, з метою скорочення транспортних затримок. Авторами роботи [20] досліджувалась фактична дорожньо-транспортна ситуація на складному перехресті, де в'їзд на перехрестя з'єднаний з двома частинами, головною дорогою і дорогою, яка є вїздом з прилеглої території. Для підвищення якості транспортного обслуговування на цьому об'єкті авторами було використано передсигнали, ефективність використання яких підтверджена результатами мікромоделювання руху транспортних засобів на цьому перехресті з використанням VISSIM.

В роботі [21] авторами було проаналізовано вплив використання передсигналів на пропускну здатність регульованих перехресть. В результаті чого встановлено, що використання передсигналів дозволяє збільшити пропускну здатність регульованих перехресть.

В статті [22] запропоновано оптимізаційну модель розподілу смуг руху та синхронізації основного сигналу і передсигналу на перехресті, обладнаному двома світлофорами: основним і передсигналом (tandem intersection або tandem design), рис. 3. Крім того, в роботі доведено, що, за однакової кількості смуг на всіх підходах до перехрестя, середня транспортна затримка зменшиться, за умов якщо розподіл транспорту смугами буде пропорційним до частки поворотних транспортних засобів.

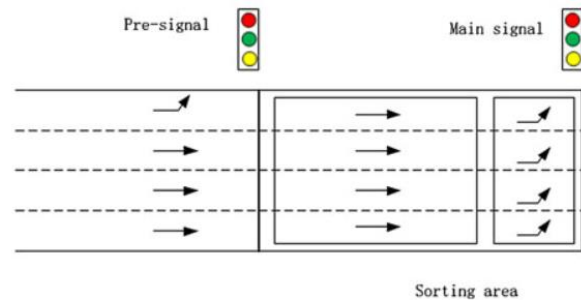


Рис. 3. Приклад використання передсигналу на підході до tandem intersection [22]

Рішення схожих задач до тих, що розглядалися в [22] з використанням передсигналів, представлено в роботі [23], рис. 4. Для оцінки ефективності запропонованих заходів із координації роботи основного світлофору і передсигналу використані засоби мікромоделювання VISSIM. За основні індикатори покращення транспортної ситуації на перехресті обрано три параметри: максимальну довжину черги, час зупинки та затримки транспортних засобів.

Авторами роботи [24] досліджено потенційну пропускну здатність перехресть за умов використання передсигналів. До того ж в статті представлено розрахунок довжини зони очікування основного попереднього сигналу у відповідності до взаємозв'язку між пропускну здатністю та швидкістю прибуття транспортних засобів. Результати показують, що використання передсигналів може покращити пропускну здатність перехресть і зменшити затримки транспортних засобів.

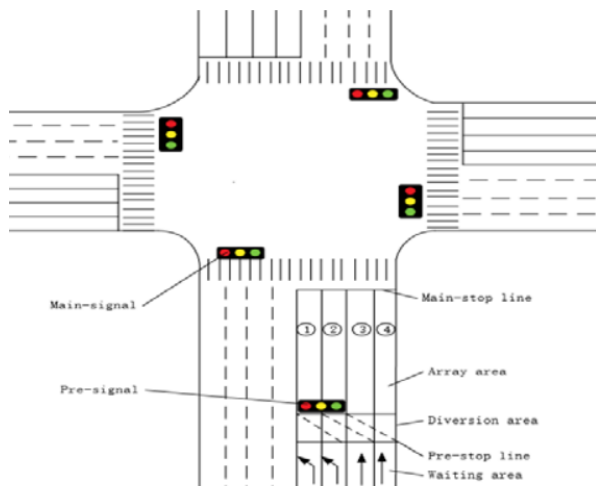


Рис. 4. Принципова схема перехрестя з обладнанням передсигналу лише на одному підході (single tandem intersection) [23]

Робота [25] присвячена розробці оптимальної стратегії управління дорожнім рухом на tandem intersection з передсигналами. Для отримання результатів авторами запропоновано вдосконалений метод розподілу смуг руху, який дозволяє уникати встановлення смуг руху з однаковим напрямком руху впритул одна до одної, наскільки це можливе, запропоновано оптимізаційну модель для мінімізації середньої затримки транспортного засобу та проведено експериментальне дослідження на перевантаженому перехресті в Чанчуні, Китай. Результати показують, що використання запропонованого методу розподілу смуг руху дозволяє зменшити середню затримку транспортних засобів на 22,9% і збільшити пропускну здатність перехрестя на 18,6%.

Оптимізації роботи основного і передсигнального світлофорів присвячена робота [26] на перехрестях з контрпотоковою смугою для лівоповоротних потоків (contraflow left-turn lane). Авторами запропоновано модифікований метод на основі ударних хвиль для оцінки максимальної довжини черги для лівого повороту на перехрестях, обладнаних світлофорними об'єктами, враховуючи унікальну поведінку черги перед передсигналом. Розроблена модель дає обґрунтовану оцінку максимальної довжини черги для лівоповоротного руху на перехрестях, обладнаних світлофорними об'єктами, що мають схему contraflow left-turn lane. Також було запропоновано процедуру, яка допоможе визначити мінімальну відстань між перехрестями, що регулюються світлофорними об'єктами, для впровадження схеми з contraflow left-turn lane.

В статті [27] розглядається питання використання принципу попереднього сортування і управління рухом за участі передсигналу (presorting and presignaling), який адаптовано під існуючий попит на рух автономних транспортних засобів. У цьому випадку передсигнал використовується як засіб для надання пріоритету в русі автономних автомобілів (рис. 5), ефективність використання яких доведена із застосуванням засобів мікросимуляції SUMO, на прикладі невеликих перехресть Німеччини.

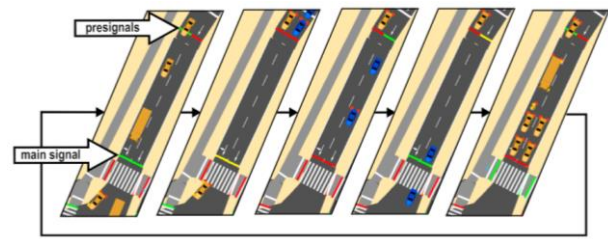


Рис. 5. Приклад роботи «попередньо відсортованого та попередньо сигналізованого» північного під'їзду до перехрестя Гостманнплац (автономні автомобілі позначені синім кольором) [27]

Підсумовуючи проведений аналіз практики використання передсигналів, можна сказати, що на сьогоднішній день науковці провели велику кількість досліджень щодо обладнання під'їзних доріг і перехресть передсигналами, їх налаштування та оцінки переваг їх впровадження як для пріоритизації руху громадського транспорту, так і для оптимізації роботи локальних і скоординованих систем світлофорних об'єктів. При цьому основною метою встановлення передсигналів є скорочення затримок транспорту і, як наслідок збільшення пропускну здатності перехресть. Основним засобом оцінки ефективності застосування передсигналів є проведення імітаційного експерименту з використанням програмних продуктів: VISSIN, SUMO, AIMSUN.

Висновки

Авторами чисельних досліджень доведено ефективність використання передсигналів для підвищення пропускну здатності регульованих перехресть, що дозволяє покращити транспортну ситуацію на них, особливо це стосується останнім за часом публікацій, де розглядаються питання дослідження транспортних систем з технологіями підключених та підключених автоматизованих транспортних засобів (Connected and Connected Automated Vehicle Technology).

Основними критеріями щодо оцінювання ефективності управління світлофорними об'єктами, в складі яких наявні передсигнали, є: транспортні затримки (громадського та (або) індивідуального транспорту), рівень обслуговування перехрестя, довжина черги та втрати часу учасниками руху (найчастіше користувачів громадського транспорту).

Результати аналізу практики використання передсигналів свідчать про відсутність цікавості та досвіду їх використання для оптимізації дорожнього руху на регульованих перехрестях серед вітчизняних науковців і практиків. На даний момент складно пояснити в чому саме полягають головні причини такої незацікавленості українських фахівців у використанні передсигналів. З позиції практиків - можливо в необхідності витрачання додаткових коштів на їх придбання і обладнання і, як наслідок, необхідності додаткового налаштування вже мінімум двох світлофорних об'єктів на синхронну роботу.

Викликає цікавість такий напрям досліджень як визначення та обґрунтування умов для вибору місць розташування передсигналів на міських магістралях.

СПИСОК ЛІТЕРАТУРИ

1. Von Stein, W., 1961. Traffic flow with pre-signals and the signal funnel. *Theory of Traffic Flow*, Elsevier, Amsterdam.
2. Pre-signals for Transit Priority [Електронний ресурс] One center research. URL: <https://umtcresearch.wordpress.com/2016/05/17/>
3. Ogden, B. D., 2001. Presignals: Current Applications and Issues. *Transportation Research Record*, 1762(1), 25-31. DOI: <https://doi.org/10.3141/1762-04>
4. Kuijper, D.H., 1982. The OFOS: A Description of the "Expanded Waiting Lane for Cyclists" (De OFOS: Een Beschouwing over de Opgeblazen Fietsopstelstrook). *Verkeerskunde*, 33(9), 472-476.
5. Wheeler, A.H., M.A.A., Leicester, G., Underwood, 1993. Advanced Stop-Lines for Cyclists. *Traffic Engineering & Control*, 34(2), 54-60.
6. Wheeler, A. 1995. Advanced Stop-Lines for Cyclists: A Simplified Layout. *Traffic Engineering & Control*, 36(5), 283-289.
7. Transit Signal Priority: A Planning and Implementation Handbook. URL: <https://trid.trb.org/View/772546>.
8. Sun, Y, Li, J, Wei, X, Jiao, Y., 2021. Tandem Design of Bus Priority Based on a Pre-Signal System. *Sustainability*, 13(18):10109. URL: <https://doi.org/10.3390/su131810109>.
9. Wu, J., Hounsell, N., 1998. Bus priority using pre-signals. *Transportation Research Part A: Policy and Practice*, 32(8), 563-583.
10. Guler, S.I., Menendez, M., 2014. Analytical formulation and empirical evaluation of pre-signals for bus priority. *Transportation Research Part B: Methodological*, 64, 41-53.
11. He, H., Guler, S.I., Menendez, M., 2016. Adaptive control algorithm to provide bus priority with a pre-signal. *Transp. Res. Part C Emerg. Technol.*, 64, 28-44.
12. Zhang, W., Wang, W., 2004. Study on the pre-signal setting method of intersection based on bus priority traffic. *Road Traffic Technol.*, 21, 101-104.
13. Zhang, W., Lu, H., 2005. Bus priority pre-signal control intersection vehicle delay analysis. *China Highw. J.*, 18, 78-82.
14. Zhang, B., Cheng, W., 2016. Study on the coordination control strategy of pre-signal bus priority intersection. *Roads Mot. Transp.*, 5, 28-32.
15. Chen, Qun, Tang, Yi, Wang, Yan, 2022. Dynamic Sharing Scheme and Design Optimization of Bus-Only Lanes Based on Presignal and Detector Settings Before Intersections. DOI: <http://dx.doi.org/10.2139/ssrn.4168506>
16. M., Khwais, J., Haddad, 2022. Optimal Presignal Control for Two-Mode Traffic at Isolated Signalized Intersections. *Transportation Science*, 57(2), 376-398. DOI: <https://doi.org/10.1287/trsc.2022.1172>.
17. Y., Liang, Z., Wu, J., Li, F., Li, Y., Wang, 2018. Shockwave-Based Queue Length Estimation Method for Presignals for Bus Priority. *Journal of Transportation Engineering, Part A: Systems*, 144(9). DOI: <https://doi.org/10.1061/jtepbs.0000175>.
18. Xuan, Y., Gayard, V.V., Cassidy, M.J., Daganzo, C.F., 2012. Presignal Used to Increase Bus- and Car-Carrying Capacity at Intersections, Theory and Experiment. *Transportation Research Record*, 2315, 191-196. DOI: 10.3141/2315-20
19. Y., Bie, Zh., Liu, H., Wang, 2020. Integrating Bus Priority and Presignal Method at Signalized Intersection: Algorithm Development and Evaluation. *J. of Transportation Eng, Part A: Systems*, 146(6). DOI: <https://doi.org/10.1061/JTEPBS.0000360>.
20. Zhao, J.; Wu, J., 2010. Application of Microscopic Simulation Technology in a Pre-Signal Control Method at a Complex Intersection; Asme Press: New York, NY, USA.
21. Xuan, Y., Daganzo, C.F., Cassidy, M.J., 2011. Increasing the capacity of signalized intersections with separate left turn phases. *Transp. Res. Part B Methodol.*, 45, 769-781.
22. Zhou, Y. Zhuang, H., 2014. The optimization of lane assignment and signal timing at the tandem intersection with pre-signal. *J. Adv. Transp.*, 48, 362-376.
23. Y, Liu, 2020. The Modeling and Simulation of the Tandem Intersection Considering the Vehicle Operation Law. *Academic Journal of Engineering and Technology Science*, 3(4), 145-159, DOI: 10.25236/AJETS.2020.030414.
24. Zhou, Y.P., Zhuang, H.B., 2013. Traffic Performance of Signalized Intersections with Protected Pre-Signal Established Upstream. *Appl. Mech. Mater.*, 253-255, 1299-1306.
25. J., Wan, C., Wang, Y., Bie, 2023. Optimal Traffic Control for a Tandem Intersection With Improved Lane Assignments at Presignals, in *IEEE Intelligent Transportation Systems Magazine*, doi: 10.1109/MITS.2023.3269670.
26. P., Liu, J., Wu, H., Zhou, J., Bao, Zh., Yang, 2019. Estimating Queue Length for Contraflow Left-Turn Lane Design at Signalized Intersections. *J. of Transportation Engineering, Part A: Systems*, 145(6). DOI: <https://doi.org/10.1061/JTEPBS.0000240>.
27. M., Barthauer, B., Friedrich, 2020. Evaluation of presorted and presignaled intersections with respect to traffic efficiency and traffic safety. *Transportation Research Procedia*, 47, 307-314. DOI: <https://doi.org/10.1016/j.trpro.2020.03.103>.

Received (Надійшла) 10.04.2024

Accepted for publication (Прийнята до друку) 03.07.2024

The practice of using pre-signals

Yevhen Liubiyi, Kostiantyn Dorofieiev

Abstract. The article analyzes the experience of using pre-signals to solve the problematic issue of effective traffic management at signalized intersections. Based on the results of the analysis, it has been found that the main directions of using pre-signals as a means of traffic control on cities highways are the prioritization of public transport on the approach to the intersection and directly at it, as well as the optimization of traffic signal control of local intersections or a group of intersections (coordination). The key criteria for evaluating measures to improve the efficiency of traffic flow management at controlled intersections with the use of pre-signals are traffic delays, time spent by public transport users, the level of service and the queue of vehicles. The main means used to assess the efficiency of traffic signal systems with the use of pre-signals is simulation modeling, for which the most popular software products among industry experts are used - VISSIM, SUMO, AIMSUN. A potential area for further research is the development of principles for selecting and justifying the location of pre-signals on cities highways.

Keywords: traffic, traffic signal control, signalized intersections, pre-signal, traffic delays, capacity.

С. М. Неронов, Г. А. Плехова, С. В. Очеретенко

Харківський національний автомобільно-дорожній університет, Харків, Україна

СИНЕРГІЯ АВТОМОБІЛЬНОГО ТРАНСФЕРУ ТА УТРИМАННЯ АВТОМОБІЛЬНИХ ДОРІГ

Анотація. В наш час в ІТ індустрії активно працюють над інтелектуалізацією електронних систем автомобілів та моніторингом стану та умов руху транспортних засобів з метою підвищення якості та оптимізації технічних характеристик вітчизняних наземних транспортних систем., поліпшення умов надання транспортних послуг мешканцям та підприємствам міст та регіонів України. Головне у цьому є стандартизація, керування стандартними правилами ISO/IEC 15288:2002 «Systems Engineering – systems life cycle processes» - системна інженерія – життєвий цикл складних систем. Існуючі окремі рішення ІТ індустрії повинні бути пов'язані із створенням нових систем та мереж саме автомобільного трансферу для переміщення пасажирів або вантажу у просторі та часі та умовами спеціальної просторово часової орієнтації, якісного та своєчасного виконання усього комплексу ІТ індустрії як автомобільного трансферу не тільки технічного, але і саме інтелектуального прийняття інженерних рішень згідно логіки професійної людини. Це віртуальна людська логіка водія- професіонала -перевізника. Це і є професіональний технічний, інструмент трансферної системи.

Ключові слова: комп'ютерні науки, автомобільний трансфер, утримання доріг, синергія, принципи та парадигми використання автомобільних комп'ютерних систем -АКС.

Вступ

Сьогодні розвиток транспортних систем, удосконалення відповідних транспортних засобів та транспортних комунікацій визначається рівнем інтелектуалізації транспорту у цілому. Оцінка рівня досконалості транспортних засобів, середовища руху та транспортної інфраструктури може бути порівняна з визначенням якостей, таких як автономність, ступінь уніфікації та спеціалізації мехатронних модулів та систем.

Це складова частина для проведення тренінга з комп'ютерних наук (КН) для пошукачів знань, навичок та вмінь, базових компетенцій ІТ індустрії. Повинно тільки чітко вважати, що сьогодні є наукова, теоретична та практична необхідність застосування КН.

Потрібні фахівцівнє тільки широкого профілю, але і саме з компетенціями нових рівнів кваліфікації.

Важливою умовою кар'єрного зростання є перехід на такий рівень фахівця з уміннями та знаннями КН в будь якій професійній галузі. Це є не тільки умовою кар'єрного зростання, але і становлення саме фахівців трансферних організацій (пересування пасажирів та вантажу).

Мається на увазі автомобільний транспорт АТ та дорожню галузь.

Усе так або інакше підлягає керуванню стандартними правилами ISO/IEC 15288:2002 «Systems Engineering – Systems life cycle processes» – системна інженерія – життєвий цикл складних систем.

Виведення наукових та практичних результатів свідчить про те, що використання хмарної інфраструктури, замість власної інфраструктури ІТ, приносить найбільші вигоди не тільки з технологічного, але й з фінансового погляду.

Оплата за фактично використані послуги, що є стандартом в хмарній інфраструктурі, виявляється значно економічнішою, ніж "передоплата за все", що є характерним для внутрішньої інфраструктури ІТ.

Таким чином, можна констатувати отримання зиску до 30 [% від витрат на особисту внутрішню інфраструктуру та 15 % від витрат порівняно з аутсорсером (керованим сервісом).

Програмно-апаратне забезпечення інформаційно-комунікаційної технології управління передбачає створення розподіленої внутрішньої та зовнішньої телематики транспортних систем, нових правил та концепції віртуального управління транспортними послугами.

Новим є створення єдиного інформаційного простору ринку ІТ індустрії на засадах Cloud Computing (хмарних обчислень). Запропоновано надання практично не обмежених додаткових комп'ютерних ресурсів усім учасникам автомобільного трансферу вантажу та/або пасажирів. Особливість – не звичайне віртуальне логістичне уявлення їх об'єднання, а синергетика, своєрідне узагальнення автотрансферу, його віртуальне управління.

Далі розглянемо це запитання на прикладі автотранспорту.

Як будь-яку складну систему автотранспорт можна розглядати подвійно: як сукупність автономних систем або як розподілену систему. Взагалі автономна система (autonomous systems) – це така система, яка усі функції з реєстрації, відбиття й обробки інформації виконує безпосередньо в автомобілі, що рухається.

Прикладом автономної системи є серія мобільних інформаційно-обчислювальних комплексів (ІОК-ХАДІ), розроблених у ХНАДУ для обстеження автомобільних доріг.

Аналогом автономної системи є телематична система легкового автомобіля з показчиком технічного обслуговування, бортовою системою контролю, маршрутним комп'ютером і навігаційною системою для забезпечення роботи спеціальна автомобільна технологія керування – RTI (Road and Traffic Information) – система інформації про дорогу і дорожнє середовище).

Аналіз публікацій

Основним джерелом цієї розробки є постановка задачі на застосування комп'ютерних наук в умовах розвитку ІТ індустрії віртуального управління перевізними процесами.

Основним джерелом цієї роботи були потреби інтелектуалізації інструментальних засобів керування процесами прийняття рішень з синергетики, телематики та мехатроніки трансферних систем, їх відповідність логіки своєчасного пересування вантажу або пасажирів із застосуванням комп'ютерних наук (КН) згідно особистих стартапів авторів, саме транспортного порталу (ikt.khadi.kharkov.ua) за змістом фундаментального посібника [1] та монографії [2, 3].

Когнітивом є формування понять:

– прийняття рішень з логістики уявлення теоретичних знань та практичних дій,

– ситуаційних задач конкретного сприйняття дійсності переміщення у просторі та часі пасажирів та вантажу,

– просторово-часової орієнтацію транспортних процесів, відповідного, перш за все, автомобільного трансферу у дослідженні [4, 5].

Когнітивом є формування понять прийняття рішень з логістики уявлення теоретичних знань. Це і є сприйняття дійсності переміщення у просторі та часі пасажирів та вантажу, перш за все, відповідних електронних ресурсів IBM Corp. та Український Академічний Грід [6-10].

Мета та постановка задачі

Складну транспортну систему можна розглядати двома способами: як сукупність автономних систем або як розподілену систему.

Автономна система (autonomous system) – це система, яка виконує всі функції з реєстрації, відбиття та обробки інформації безпосередньо в рухомому транспортному засобі.

Прикладом автономної системи є серія мобільних інформаційно-обчислювальних комплексів (ІОК-ХАДІ), які розроблені у ХНАДУ для обстеження автомобільних доріг.

Аналогом автономної системи є телематична система легкового автомобіля з показником технічного обслуговування, бортовою системою контролю, маршрутним комп'ютером та навігаційною системою, а для забезпечення роботи спеціальної технології керування автомобілем використовується система інформації про дорогу і дорожнє середовище (RTI – Road and Traffic Information). Ця інформація складається з легко розпізнаваних символів на карті, що відображається на моніторі. На монітор також виводиться інформація про стан транспортного засобу, якість роботи двигуна та системи електроживлення.

Розглянута проблема інтеграції трьох компонентів – механічного, електронного та керуючого – у єдиний скомпонований блок, є однією з ключових проблем в мехатронній техніці. Хоча автономність є характерною рисою мехатронних об'єктів, наприклад, інтелектуальна керуюча частина

транспортної роботи не може бути поміщена на його рухомі частини.

Проте, як вже було доведено раніше, всі "автономні мехатронні системи" включають неавтономні модулі.

Це є додатковим підтвердженням того, що використання властивості "автономності" транспортних засобів для створення транспортних мехатронних систем є справедливим.

Таким чином, сучасні транспортні засоби можна розглядати як мехатронні системи. Вони представляють собою своєрідний обчислювальний комплекс, для розв'язання системних задач, що використовують загальні принципи інтелектуалізації обчислювальних приладів та пристроїв.

Виклад основного матеріалу

У відмінності від поточного стану логістики, розвиток ІТ-індустрії вимагає інтерактивного моніторингу як автомобілів, так і учасників перевезення на дорозі. Це передбачає розподілення комп'ютерних ресурсів між користувачами доріг та учасниками дорожнього руху.

Дослідження закономірностей розвитку телематики на автомобільному транспорті та використання новітньої мережної технології Cloud Computing у транспортних організаціях та фізичних осіб можуть допомогти досягти досяжності, спостережуваності та створення клієнтської частини телематики транспортної організації та автомобілів.

Фізичне, імітаційне моделювання, тестування та верифікація комп'ютеризованих інформаційних процесів можуть допомогти у підтвердженні інтерактивності транспортних процесів та управлінні наземним транспортом з використанням інформаційно-комунікаційних технологій.

Розробка внутрішньої автомобільної телематики та інтерактивної системи реєстрації, оцінки та накопичення даних можуть забезпечити оперативний моніторинг дорожньої ситуації та середовища дорожнього руху.

Наслідком є створення пропозицій щодо досяжності, спостережуваності та реалізації клієнтської частини телематики транспортної (дорожньої) організації, яка забезпечує інформаційну взаємодію між учасниками дорожнього руху та автомобілями.

Фізичне та імітаційне моделювання, тестування та верифікація комп'ютеризації процесів оцінки дорожніх ситуацій підтверджують експериментальні можливості інтерактивності транспортних процесів та шляхів втілення інформаційно-комунікаційної технології управління наземним транспортом.

Ці розробки повинні передувати створенню внутрішньої автомобільної телематики та інтерактивної системи реєстрації, оцінки та збору даних про оперативну ситуацію та середовище дорожнього руху.

Необхідно зменшити кількість управлінського та виробничого персоналу, який займається автотранспортом, скоротити час на прийняття рішень та усунути негативний вплив дорожнього руху. Таке

поліпшення підтверджено актами впровадження Служби автомобільних доріг та державного підприємства "Автодор" Харківської області.

Першим кроком впровадження веб-рішень на транспорті є використання інструментальних засобів, таких як дорожні сканери та інтерактивні дорожні тестери, які можна використовувати на смартфонах і планшетах.

Для інформування учасників дорожнього руху, користувачів доріг, робітників дорожньо-експлуатаційних організацій та будь-яких фізичних осіб про необхідність усунення негативних погіршень стану доріг протягом будь-якого періоду року, потрібно:

- розгорнути транспортний дорожній портал,
- автоматизувати та вести електронний журнальний облік робіт з утримання доріг,
- систематично спостерігати за транспортно-експлуатаційним станом доріг.

Сьогодні однією з найважливіших переваг використання WEB-рішень (тобто комп'ютерних програм для вирішення проблем) є простота їх реалізації та розгортання.

Фактично для цього достатньо мати такі ресурси:

- автомобіль,
- інформаційно-комунікаційний центр (ІКЦ) – вартість якого не перевищує вартості смартфона,
- інформаційний Інтернет-сайт – транспортний дорожній портал (вартість першого розгортання становить близько 20 000 грн.).

Усі витрати на створення такого ланцюга інформаційного комунікаційного центру (ІКЦ) та інформаційної комунікаційної технології (ІКТ) для дорожньої організації не перевищують звичайних витрат на утримання комп'ютерного ресурсу організації державної або комерційної компанії та фізичної особи (рис. 1).

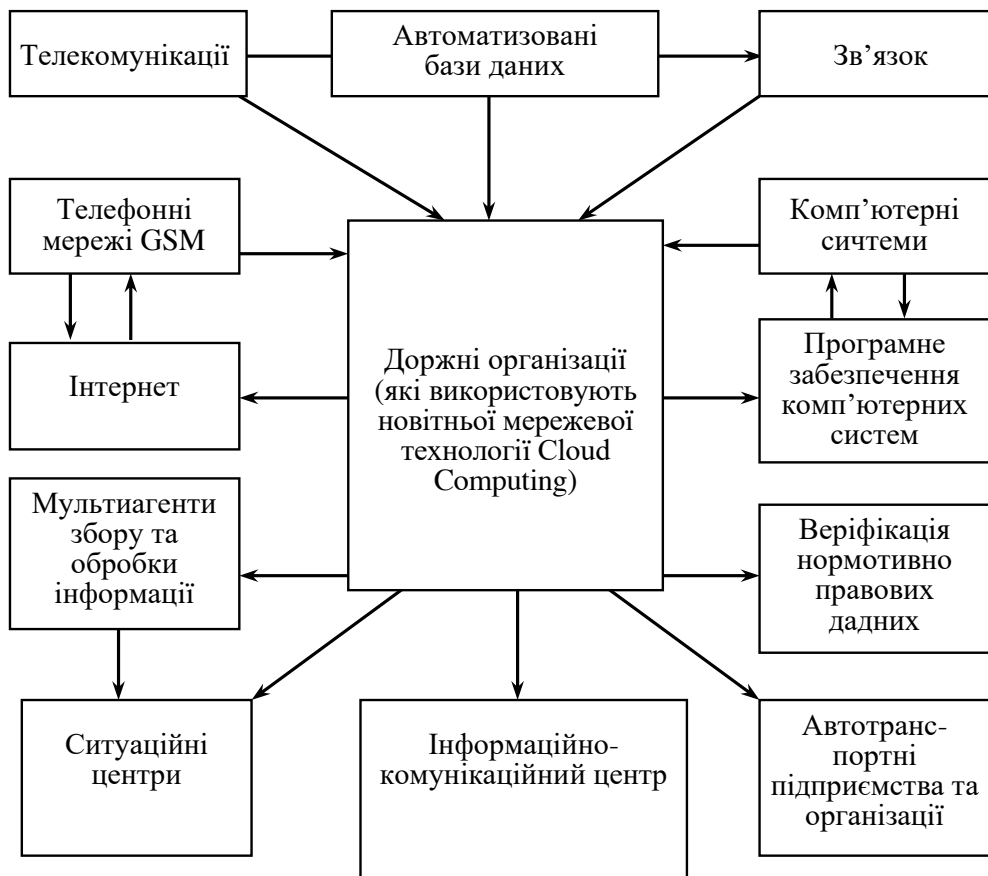


Рис. 1. Інформаційно- комунікаційна структура

Практичною підтверджено, що один водій зі смартфоном та клієнтським додатком транспортного дорожнього порталу ІКЦ-ІКТ може замінити гromізdky систему осіб, що приймають рішення з організації перевізних процесів.

Немає потреби залучати власників, перевізників та отримувачів вантажів (достатньо зусиль тільки саме учасників автотранспорту). Це підтверджено науковим обґрунтуванням, розрахунками та економією витрат на впровадження нового віртуального управління автомобільним транспортом:

установку обладнання, щомісячну плату за обслуговування, щомісячну зарплату персоналу звичайної віртуальної логістики.

Висновки

Утворення єдиного інформаційного простору у хмарних обчисленнях (Cloud Computing) передбачає не тільки технічну організацію розподілених обчислень у гетерогенному середовищі, але й створення певної соціальної структури, яка включає користувачів системи – учасників дорожнього руху.

Використання Інтернету забезпечує необхідну інформаційну базу для керування будь-яким потоком (матеріальним, інформаційним, трудовим, інвестиційним, фінансовим тощо) та оптимізації його руху.

Використання динамічного сайту-агрегатора WEB може бути найбільш корисним при синергетичному підході, який полягає в комбінації інформаційних можливостей сайту та картографічного сервісу для визначення розташування транспортних засобів на дорогах та візуалізації транспортної ситуації. Це дозволяє своєчасно приймати рішення щодо оцінки та усунення можливих негативних впливів на

транспортні процеси.

При цьому використання хмарної інфраструктури замість власної інфраструктури ІТ є набагато більш вигідним з фінансової точки зору, оскільки дозволяє платити лише за фактично спожиті послуги, не потребуючи значних капітальних витрат.

Крім того, імплементація цих наукових та практичних результатів свідчить про те, що витрати, пов'язані лише з залученням користувачів WEB рішень до цієї нової роботи, а не зі створенням власної інфраструктури ІТ.

СПИСОК ЛІТЕРАТУРИ

1. Alekseyev O. Development of automotive computer systems based on the virtualization of transportation processes management/ O. Alekseyev, V. Alekseyev D. Klets., V. Khabarov, et al. // Eastern-European Journal of Enterprise Technologies. – 2017. – Vol.6, N 3 (90). – P. 14-25.
2. Алексієв В.О. Інформаційний розвиток порталу віртуального управління процесами транспортного обслуговування / О. П. Алексієв, В. О. Алексієв, // Інформаційні технології: проблеми та перспективи : монографія[Текст] . – Х.: Видво: Рожко С. Г., 2017. – Розд. 2. – С. 32 – 47.
3. Introduction to Grid Computing. SG24-6778-00 [Electronic Resource] // IBM Corp. – 2005. – 268 p. – Access mode to mag.: <http://ibm.com/redbooks>.
4. Історія Грід в Україні [Електронний ресурс] // Український Академічний Грід. – 2007. – Режим доступу до журн.: <http://uag.bitp.kiev.ua/index.php/uk/grid-history.html>.
5. Varaiya Smart cars on smart roads/ Varaiya, Pravin. // problems of control: IEEE Transactions on Automatic Control, 1993, AC-38 (2). – P. 195–207.
6. Sviridov, A., Kovalenko, A. and Kuchuk, H. (2018), “The pass-through capacity redevelopment method of net critical section based on improvement ON/OFF models of traffic”, *Advanced Information Systems*, Vol. 2, No. 2, pp. 139–144, DOI: <https://doi.org/10.20998/2522-9052.2018.2.24>
7. Bogomolov V.O. Conceptually, a bathtub and a synergetic pid before the development of transport systems / V.O. Bogomolov, V.O. Alekseyev // Informatsionno-keruyuchi systems on zaliznichny transport : a science and technology magazine. - 2009. No. 5 (78). - S. 59–63.
8. Kovalenko, A. and Kuchuk H. (2018), “Methods for synthesis of informational and technical structures of critical application object’s control system”, *Advanced Information Systems*, 2018, Vol.2, No.1, pp. 22–27, DOI: <https://doi.org/10.20998/2522-9052.2018.1.04>
9. Pat. 97432 U Ukraine, IPC (2015.01) G01C 23/00. Interactive Road Tester / Ne-Ronov S.M., Alekseyev O.P., Alekseyev V.O.; the patent applicant is the Kharkiv National Auto-Road Universitet. Declared 10/27/2014; publ. 03/10/2015. Bull. No.
10. Костікова М. В., Неронов С. М., Плехова Г. А. Особливості математичного моделювання явищ економіки. Modern aspects of science. International collective monograph. International Economic Institute s.r.o.. Czech Republic: International Economic Institute s.r.o., 2024. Vol. 40. P. 500 – 512.

Received (Надійшла) 24.04.2024

Accepted for publication (Прийнята до друку) 03.07.2024

Cloud computing automobile transfer and road maintenance

S. Neronov, G. Pliekhova, S. Ocheretenko

Abstract. The goal is to develop a new quality level for domestic land transport systems and optimize their technical characteristics, improve the conditions for providing transport services to residents and enterprises of cities and regions of Ukraine. **Research objectives.** To study the existing separate solutions in the IT industry for automobile transfer. To connect these separate solutions for the creation of new systems and networks. To implement the entire complex of the IT industry as automobile transfer, not only technical, but also intellectually making engineering decisions according to the logic of a professional person. To achieve a virtual human logic of a professional transport driver. **The following results are received.** A comprehensive understanding of the existing IT solutions in the automobile transfer industry. A proposed framework for connecting these solutions to create new systems and networks. A system that can make intellectual engineering decisions according to the logic of a professional person. The achievement of a virtual human logic of a professional transport driver. **Conclusions.** The intellectualization of the car's electrical systems and monitoring the condition and driving conditions of transport vehicles is a crucial step towards optimizing domestic land transport systems. By following the standard rules of ISO/IEC 15288:2002 and connecting existing IT solutions, a new level of quality and efficiency can be achieved. This includes the development of a system that can make intellectual decisions, mimicking the logic of a professional transport driver. This research provides a significant contribution to the IT industry and the future of land transport systems in Ukraine.

Keywords: computer science, car transfer, road maintenance, synergy, principles and paradigms of using car computer systems - ACS.

А. О. Андрусевич, А. В. Фролов, М. Г. Стародубцев, Н. П. Демська, Г. С. Макаренко

Харківський національний університет радіоелектроніки, Харків, Україна

ОПТИМІЗАЦІЯ СТРАТЕГІЇ УПРАВЛІННЯ ТЕХНОЛОГІЧНИМИ ПРОЦЕСАМИ В УМОВАХ НЕВИЗНАЧЕНОСТІ

Анотація. В результаті аналізу задач оптимізації управління технологічними процесами виробництва виробів нової техніки в нечітких умовах виявлено, що при дослідженні технологічного об'єкта управління невисокої розмірності може бути використаний підхід, заснований на заміні нечітких множин множинами α -рівня і зведенні нечіткої оптимізаційної задачі до класичної задачі математичного програмування. Показано що для оптимізації управлінням багатоопераційним технологічними процесами більше раціональним є підхід Беллмана-Заде на основі принципу злиття цілей і обмежень, який вимагає подальшого розвитку на випадок нечітких станів технологічного об'єкта управління. Крім того, не вирішена проблема об'єднання нечітких цілей і обмежень у разі їх неоднакової важливості або взаємозалежності. Для вирішення цих проблем запропоновано метод оптимізації стратегії управління технологічними процесами виробництва за багатьма критеріями. В якості узагальненого критерію використовується згортка критеріїв у вигляді нечіткого інтеграла Сугено.

Ключові слова: оптимізація, управління, технологічний процес, технологічний об'єкт, невизначеність умов.

Вступ

Традиційно під терміном "керований технологічний процес" розуміється такий технологічний процес (ТП), для якого встановлені вхідні керуючі впливи та вихідні параметри та залежності між ними, визначено методи вимірювання вхідних впливів та вихідних параметрів та методи управління процесом [1].

Керованими змінними є:

- параметри ТП;
- стан засобів технологічного оснащення;
- техніко-економічні показники виробництва виробів нової техніки, що формуються в ході ТП (продуктивність, собівартість, точність, вихід придатної продукції та ін.).

На практиці управління ТП, як правило, відбувається в умовах невизначеності, пов'язаної з відсутністю достатньої інформації про поведінку керованих об'єктів, неоднорідністю вихідних матеріалів, високим рівнем шумів, дискретністю за часом вимірювання станів процесу, зміною контрольованих параметрів на кожному кроці процесу, неможливістю точного визначення фазових координат. Більш того, не всі координати вектора стану ТП можуть бути проконтрольовані. Частина факторів, що характеризують роботу промислового об'єкта, або взагалі невідома, або процес виміру пов'язаний із значними труднощами.

Ускладнення ТП призводить до збільшення невизначеності при їхньому функціонуванні. У виробництві виробів нової техніки посилюються тенденції до переходу на багатонаменклатурне дрібносерійне виробництво, яке характеризується малими партіями виробів та великим діапазоном їх розмірів та властивостей. При керуванні такими процесами зростає частота заміни настроювальних величин і підвищуються вимоги до точності та гнучкості настроювальної моделі. Гостро постає проблема адекватності

використовуваних моделей, що визначається, насамперед, реалізацією принципу інформаційної достатності [2].

Недолік кількісної інформації, одержуваної в результаті застосування об'єктивних, але обмежених методів дослідження, певною мірою може бути компенсований інформацією якісного характеру у вигляді інтуїтивних знань і досвіду технолога. Ця інформація носить суб'єктивний характер, і її подання природною мовою, як правило, містить велику кількість невизначеностей типу "багато", "мало", "сильно збільшити", "високий", "дуже ефективний" і т.п., які не мають аналогів мовою традиційної математики. Тому й опис подібної інформації мовою традиційної математики збіднює математичну модель досліджуваного реального процесу і робить його дуже грубим. Разом з тим наявність математичних засобів відображення невизначеності вихідної інформації дозволяє побудувати модель, більш адекватну реальності.

Найбільш поширеним видом невизначеності, що зустрічається при дослідженні ТП, є невизначеність, пов'язана з випадковістю. Теоретико-імовірнісне поняття випадковості відноситься до категорії об'єктивних понять та розглядається як додаткове до поняття причинності; таке уявлення підкріплюється концепцією відтворюваності експериментів, яка узгоджується зі спостереженнями в галузі природничих наук та в техніці. Для роботи з таким видом невизначеності використовуються теорія ймовірностей, математична статистика, теорія випадкових процесів.

Однак розгляд невизначених даних і зв'язків при аналізі ТП показує, що існують різні класи невизначеності, які не завжди пов'язані з випадковістю.

У світлі панівної думки, породженої так званою декартовою раціоналістською методологією, традиційно існує тенденція відкидати такі види невизначеності, як неясність, нечіткість і неточність через їхню

ненаукову або ірраціональну концепцію. У той же час, при управлінні ТП ми постійно стикаємося з ситуацією, коли неможливо уникнути проблеми обліку неясної, нечіткої або неточної інформації.

Одна з основних проблем при цьому – знайти спосіб обробки наявної інформації для вибору раціональних варіантів управління ТП.

Огляд методів обробки невизначеної інформації показав, що найбільш перспективним та систематичним є підхід, заснований на використанні поняття нечіткої множини Заде [3]. Теорія нечітких множин являє собою узагальнення і переосмислення найважливіших напрямків класичної математики. У її витоках лежать ідеї і досягнення багатозначної логіки, яка вказала на можливості переходу від двох до довільного числа значень істинності і поставила задачу оперування поняттями з мінливим змістом; теорії ймовірностей, яка, породивши велику кількість різних способів статистичної обробки експериментальних даних (наприклад, гістограми, функції розподілу), відкрила шляхи визначення та інтерпретації функції приналежності; дискретної математики (теорії матриць, теорії графів, теорії автоматів і т.д.), що запропонувала інструмент для побудови моделей багатовимірних і багаторівневих систем, зручний при вирішенні практичних задач.

Основною прагматичною метою Заде було створення апарату, здатного моделювати людські міркування та пояснювати людські прийоми прийняття рішень під час різних досліджень. Нечіткість розумілася як наслідок суб'єктивності чи індивідуальності людини, яка приймає рішення. Проте під час прикладних досліджень поняття нечіткості розширилося.

У реальних ситуаціях прийняття рішень – цілі, обмеження, критерій вибору може бути неточно визначеними. Джерелом неточності може бути принципова неможливість оперувати точними даними внаслідок складності системи. Тому при побудові моделей прийняття рішень виникає необхідність використання нечіткої логіки, нечітких множин і нечітких відносин. Нечіткі відносини дозволяють моделювати плавну, поступову зміну властивостей, а також невідомі функціональні залежності, частково або повністю виражені у вигляді якісних зв'язків.

Виклад основного матеріалу

Невизначеність, в тій чи іншій мірі властива всім процедурам прийняття рішень в управлінні ТП, має дві сторони. Одна сторона носить імовірнісний характер, друга пов'язана з неточністю і приблизністю цілей, альтернатив, ресурсів. У задачах управління ТП, як правило, присутні обидві сторони невизначеності. Так, у випадку багатоцільової задачі прийняття рішень на кожному етапі ТП ми маємо справу з невизначеністю стохастичної природи, пов'язаною з випадковими впливами зовнішнього середовища, і з невизначеністю, одержуваною в результаті формалізації якісної інформації про порівняльну перевагу можливих варіантів рішення.

Найбільш адекватною стратегією оптимізації ТП є стратегія послідовної оптимізації. В умовах невизначеності процедура прийняття рішень на кожному етапі

ТП орієнтована в загальному випадку не на вирішення чітко поставленої математичної задачі оптимізації, а на послідовне виявлення переваг особою, яка приймає рішення (на основі одержуваної від нього інформації) разом з визначенням допустимої множини дій.

При розробці і використанні математичних моделей прийняття рішень, наприклад, моделей математичного програмування, коли альтернатива являє собою множину рішень потужності континуум, і оптимальне рішення відповідає екстремуму цільової функції, питання про аналіз наслідків як таких в загальному випадку не ставиться. А він особливо важливий при дослідженні слабо структурованих об'єктів.

У ситуації вибору найкращої дії з множини можливих, коли результати цих дій носять імовірнісний характер, кожна дія характеризується своїм вектором ймовірностей всіх можливих альтернативних результатів. Одна з найбільш розроблених теорій вибору найкращої дії в цих умовах – теорія ймовірнісної корисності фон Неймана і Моргенштерна, згідно з якою для оцінки відносної переваги кожного альтернативного результату в умовах ризику вводиться спеціальна функція корисності (несуперечливий кількісний опис переваг за якісними критеріями), значення якої для кожної альтернативи встановлюються з врахуванням переваг особою, яка приймає рішення. На основі функції корисності кожної альтернативи та ймовірностей їх реалізації формується функція корисності кожної дії в цілому. Далі вибирається дія, що забезпечує максимум знайденої таким чином функції корисності. Однак альтернативи можуть відрізнятися між собою за цілою низкою факторів або критеріїв. Багатовимірна альтернатива оцінюється відповідно за допомогою багатовимірної функції корисності. Побудова багатовимірних функцій корисності є надзвичайно складною проблемою [4].

В [5] розглянуті методи звуження множини допустимих альтернатив. До них належать методи, що ґрунтуються на ідеях ефективних компромісів Парето, гарантовані оцінки Гермейера та вибір рішень на основі нечіткого опису Заде.

Відповідно до принципу Парето можливі рішення слід шукати лише серед невідомінованих (недомінованих) альтернатив, тобто альтернатив, поліпшення яких за одними критеріями призводить до їх погіршення за іншими критеріями. Оскільки ніякий математичний аналіз не може дати точного результату вибору альтернатив в умовах невизначеності, побудова множини невідомінованих альтернатив є нетривіальною задачею.

В основі методу гарантованих оцінок Гермейера лежить принцип максимуму. При вирішенні проблем прийняття рішень в умовах невизначеності може бути лише один строгий математичний результат – це оцінка, отримана на основі принципу максимуму. Гарантований результат є єдиною опорною точкою. Далі лежать гіпотези і ризик. Однак це твердження не означає, що потрібно вибирати саме ту альтернативу, яка реалізує цей гарантований результат. Він може бути і дуже хорошим і абсолютно неприйнятним.

Стиснути множину можливих альтернатив дозволяє також техніка Заде. У схемах аналізу будується

деяка система гіпотез. Вони формулюються в термінах «суб'єктивної» приналежності. Далі, в результаті аналізу отримуємо рішення в нечіткій формі – у формі приналежності деякій множині, а потім його дефазифікуємо. Використання інформації, заданої у формі нечітких множин, призводить до нових формулювань оптимізаційних задач і методів їх вирішення.

У загальному випадку задачу оптимізації можна сформулювати як задачу максимізації (мінімізації) заданої функції на заданій множині допустимих альтернатив, яка описується системою рівностей та/або нерівностей.

$$f(x) \rightarrow \max, \varphi_i(x) \leq 0, i = 1, \dots, m, x \in X,$$

де X – задана множина альтернатив; $f: X \rightarrow R^1$ і $\varphi_i: X \rightarrow R^1$ – задані функції.

При розгляді реальних задач прийняття рішень в управлінні ТП в розпорядженні дослідника можуть виявитися лише нечіткі описи функцій f і φ_i або параметрів, від яких залежать ці функції. Це може бути через недостатність інформації про цю ситуацію або служити формою наближеного опису ситуації, достатнього для вирішення поставленої задачі. Крім того, в реальній ситуації альтернативи поза множиною обмежень можуть бути не неприпустимими, а лише в тій чи іншій мірі менш бажаними для особи, яка приймає рішення, ніж альтернативи всередині цієї множини. Форми нечіткого опису вихідної інформації в задачах прийняття рішень можуть бути різними. Звідси і відмінності в математичних формулюваннях відповідних оптимізаційних задач. Залежно від виду функцій f і φ_i розрізняють наступні задачі [5]:

- оптимізація з нечіткими відносинами;
- оптимізація з нечіткими обмеженнями;
- оптимізація з нечіткою метою;
- оптимізація з нечіткою метою та нечіткими обмеженнями.

Аналіз методів вирішення різних типів нечітких оптимізаційних задач [5, 6, 7] показав, що в основі їх лежать два підходи:

- заміна нечітких множин множинами α -рівня і зведення нечіткої оптимізаційної задачі до класичної задачі математичного програмування;
- представлення цільової функції і обмежень нечіткими множинами і використання принципу злиття цілей і обмежень Беллмана-Заде [8].

При використанні першого підходу вихідна задача нечіткого математичного програмування зводиться до вирішення деякої множини класичних задач оптимізації. Розглянемо задачу оптимізації з нечіткими обмеженнями.

Нехай задані:

$f: X \rightarrow R^1$ – цільова функція;

X – універсальна множина альтернатив;

$\mu_c: X \rightarrow [0,1]$ – функція приналежності (ФП) нечіткої множини допустимих альтернатив.

Нехай задача оптимізації сформульована так: $\max f$ при μ_c , причому під максимізацією розуміється раціональний вибір конкретної альтернативи чи множини альтернатив.

У цьому випадку вихідна задача нечіткого математичного програмування представляється у вигляді

сукупності звичайних задач максимізації функції f на всіляких множинах рівня множини допустимих альтернатив.

$$c_\alpha = \{x | x \in X, \mu_c(x) \geq \alpha\}.$$

Якщо альтернатива $x_0 \in X$ є рішенням задачі $f(x) \rightarrow \max$ на множині рівня α , то α трактується як ступінь належності альтернативи x_0 нечіткій множині рішень вихідної задачі нечіткого математичного програмування. Перебравши всілякі α , отримаємо функцію приналежності нечіткого рішення. Для всякого $\alpha \geq 0$, такого, що $c_\alpha \neq \emptyset$, побудуємо множину

$$N(\alpha) = \{x | x \in X, f(x) = \sup_{x' \in c_\alpha} f(x')\} -$$

множину рішень звичайної задачі максимізації на множині допустимих альтернатив зі ступенем приналежності не менше α . Тоді рішення вихідної задачі визначається зі співвідношення

$$\sup p \mu_f(x) = \bigcup_{\alpha \geq 0} N(\alpha).$$

Очевидно, що стосовно до вирішення задач управління ТП даний підхід може бути використаний при дослідженні технологічного об'єкта невисокої розмірності, наприклад, при оптимізації окремих технологічних операцій, зважаючи на складність вирішення багатомірних задач математичного програмування.

Для оптимізації управління багатостадійним ТП більш раціональним є підхід Беллмана-Заде. Введення понять про нечіткі цілі і обмеження дозволяє застосувати іншу, відмінну від стандартної, логічну схему прийняття рішень. У стандартній схемі фіксується множина альтернатив, множина обмежень і критерій (критерії) або цільова функція (функції) вибору оптимальної альтернативи. При нечітких цілях і обмеженнях, заданих у просторі альтернатив або в просторі станів системи, логічна схема прийняття рішень виглядає по-іншому. Основним у даному підході є те, що цілі прийняття рішень і множина допустимих альтернатив (обмежень) розглядаються як рівноправні нечіткі множини деякої універсальної множини альтернатив. Рівноправність розуміється в тому сенсі, що рішення повинно задовольняти і цілям, і обмеженням. Результуюче рішення формується як перетин множин нечітких цілей і нечітких обмежень. Це означає, що між цілями і обмеженнями або ресурсами, по суті, втрачається відмінність. Серед множини можливих результуючих рішень вибирається так зване максимізує рішення, на якому ФП перетину множини цілей і множини обмежень приймає максимальне значення. Максимізує рішення є в деякому сенсі оптимальним.

Зважаючи на складність сучасних ТП виробництва виробів нової техніки, їх багатопараметричність, багатокритеріальності рішення задачі оптимізації не є єдиним.

При цьому неоднозначність рішення не може бути усунена шляхом внутрішнього, більш детального аналізу ТП. Необхідний зовнішній аналіз ТП, тобто процес повинен розглядатися як підсистема більш складної системи – виробничої, а критерії

оптимізації вибиратися і ранжуватися залежно від критеріїв оптимальності останньої. Це дозволить побудувати деякий результируючий показник якості ТП, який в принципі визначить єдине рішення задачі оптимізації ТП. Оскільки можливості об'єктивного вибору результируючого показника обмежені нашими знаннями властивостей ТП і систем більш високого ієрархічного рівня, то такий вибір неминуче стає суб'єктивним, і саме в цьому сенсі ми використовуємо термін субоптимальне, говорячи про єдине рішення.

Якщо розглядати багатокрокові процеси з детермінованою функцією переходів з нечіткими обмеженнями на управління та з нечіткою метою в просторі станів (спочатку з фіксованим числом кроків), то задача зводиться до знаходження послідовності управлінь, що максимізують результируючу ФП. Для знаходження максимізуючого рішення використовується стандартна процедура динамічного програмування з побудовою рекурентних співвідношень.

Ця ж задача з фіксованим часом розглянута в стохастичній постановці, коли максимізується ймовірність досягнення нечіткої мети. При цьому задача зводиться до детермінованої шляхом застосування оператора математичного очікування до ФП.

В задачі з довільним числом кроків, коли нечітка мета визначена у вигляді деякої нечіткої підмножини всієї множини можливих станів, процес закінчується, коли характеристики системи вперше виявляються в цій підмножині, і задача полягає у виборі такої послідовності управлінь з урахуванням нечітких обмежень, при якій максимізується ступінь наближення кінцевого стану до заданої підмножини.

Розглянуті варіанти задач прийняття рішень не вичерпують всього різноманіття можливих ситуацій. Підхід Беллмана-Заде розроблений тільки для детермінованих і стохастичних систем в умовах, коли нечіткими є цілі і обмеження на управління. Однак сам стан технологічного об'єкта (ТО) також може бути

$$\left. \begin{aligned} M_{N-v}(X_{N-v}) &= \max_{u_{N-v}} (\mu_{N-v}(u_{N-v}) \wedge M_{N-v+1}(X_{N-v+1})), \\ X_{N-v+1} &= f(X_{N-v}, u_{N-v}), \quad v = 1, \dots, N, \end{aligned} \right\} \quad (2)$$

де $M_{N-v}(X_{N-v})$ – ФП нечіткої мети на етапі ТП з номером $N - v$, індукованої заданою метою G_N на кінцевому етапі.

У нашому випадку стан X_{N-v} нечіткий, тому в (2) треба обчислити $M_{N-v+1}(X_{N-v+1})$, тобто визначити ступінь приналежності нечіткого стану X_{N-v+1} індукованої нечіткої цілі G_{N-v+1} .

Спочатку визначимо ступінь приналежності кожного елемента x_N нечіткого стану X_N нечіткої мети

$$\left. \begin{aligned} M_N(X_N) &= \max_{x_N} (m_N(x_N) \wedge m_N^G(x_N)), \\ M_{N-v}(X_{N-v}) &= \max_{u_{N-v}} (\mu_{N-v}(u_{N-v}) \wedge M_{N-v+1}(X_{N-v+1})), \\ X_{N-v+1} &= \tilde{f}(X_{N-v}, u_{N-v}), \quad v = 1, \dots, N. \end{aligned} \right\} \quad (3)$$

У реальних умовах при управлінні ТП виробництва виробів нової техніки можлива наявність декількох показників якості. Нехай G_j – нечітка множина, що характеризує j -й показник ($j = 1, \dots, n$).

нечітким. Крім того, не вирішена проблема об'єднання декількох цілей у разі їх неоднакової важливості або взаємозалежності.

Розглянемо ТП як N -стадійну нечітку систему, що описується нечітким рівнянням

$$X_{i+1} = \tilde{f}(X_i, u_i), \quad i = 0, 1, \dots, N - 1,$$

де « $\tilde{\cdot}$ » – оператор нечіткості;

$$X_i = \{(x_i, m_i(x_i))\}, \quad x_i \in X, u_i \in U;$$

X, U – області зміни стану та управління відповідно; $m_i(x_i)$ – ФП нечіткої множини X_i .

Нехай на управління u_i накладені нечіткі обмеження

$$C_i = \{(u_i, \mu_i(u_i))\},$$

де $\mu_i(u_i)$ – ФП нечіткої множини C_i .

Нехай задано початковий стан ТП – X_0 . Потрібно знайти послідовність чітких управлінь u_i ($i = 0, 1, \dots, N - 1$), що задовольняють нечітким обмеженням C_i , і послідовність нечітких станів X_1, \dots, X_N для якої показник якості G приймає максимальне значення.

Трансформуємо дану задачу до схеми багатоетапного прийняття рішень. Експертним шляхом визначимо передбачуване максимальне значення критерію G і представимо мету відповідною нечіткою множиною G_N в X з ФП $M_N(X_N)$. Тоді відповідно до принципу Беллмана-Заде рішення задачі можна записати у вигляді:

$$D = C_0 \cap C_1 \cap \dots \cap C_{N-1} \cap G_N.$$

Для відповідних ФП маємо:

$$\begin{aligned} \mu_D(u_0, \dots, u_{N-1}) &= \mu_0(u_0) \wedge \mu_1(u_1) \wedge \dots \\ &\dots \wedge \mu_{N-1}(u_{N-1}) \wedge M_N(X_N), \end{aligned} \quad (1)$$

де \wedge – знак операції \min .

У разі коли стани X_{N-v} – чіткі можна отримати систему рекурентних рівнянь, що дає рішення (1):

G_N , тобто сформуємо нову нечітку множину X_N^G з ФП $m_N^G(x_N)$. Як ступінь приналежності кожного нечіткого стану X_N нечіткої мети G_N вибирається максимальна ступінь приналежності елемента, що належить перетину нечітких множин X_N і X_N^G . Решта значень $M_{N-v+1}(X_{N-v+1})$, $v = 2, \dots, N$, виходять з рекурентних співвідношень.

В результаті отримуємо систему рекурентних рівнянь, що дають рішення поставленої задачі:

Тоді нечітка множина D , що є рішенням задачі, буде визначатися формулою

$$D = C_0 \cap C_1 \cap \dots \cap C_{N-1} \cap (\bigcap_{j=1}^n G_j).$$

Для отримання рішення можна використовувати систему рекурентних рівнянь, аналогічну (3), з урахуванням наявності декількох показників якості. Однак навіть у разі незалежності та однакової важливості показників рішення такої системи є дуже трудомістким.

У цьому випадку введемо узагальнений показник якості.

В якості узагальненого показника якості використовуємо згортку n приватних нечітких показників у вигляді нечіткого інтеграла Сугено [9]. Лінійний узагальнений показник використовується зазвичай в тому випадку, коли окремі показники взаємно незалежні. Згортка у вигляді нечіткого інтеграла може використовуватися, коли існує взаємозалежність показників, що характерно для більшості задач управління в нечіткій ситуації.

Позначимо через K сукупність показників G_j :

$$K = \{G_1, G_2, \dots, G_n\}.$$

Для кожного X_N з кінцевої множини можливих фінальних станів ТП визначимо функцію

$$h: K \rightarrow [0,1],$$

тобто $h(G_j)$ – нормалізована оцінка показника G_j .

Експертним шляхом визначимо g_j – ступінь важливості показника G_j , $0 \leq g_j \leq 1$.

Визначимо на $(K, 2^K)$ нечітку міру g так:

якщо $K' \subset K$, то $g(K') = \frac{1}{\lambda} \left[\prod_{G_j \in K'} (1 + \lambda g_j) - 1 \right]$, де параметр λ визначається за умови нормування:

$$\frac{1}{\lambda} \left[\prod_{j=1}^n (1 + \lambda g_j) - 1 \right] = 1.$$

Тоді як $M_N(X_N)$ візьмемо нечіткий інтеграл від функції h на множині K по нечіткій мірі g [9]:

$$M_N(X_N) = \oint_K h(x) \circ g = \sup_{\alpha \in [0,1]} \left(\alpha \wedge g \left(K \cap H_\alpha \right) \right),$$

де $H_\alpha = \{x | h(x) \geq \alpha\}$.

Далі стратегія управління ТП визначається із системи (3).

Висновки

Запропонований метод оптимізації стратегії управління ТП за багатьма критеріями з урахуванням їх неоднакової важливості і взаємозалежності узагальнює і розвиває підхід Беллмана-Заде на випадок нечіткої системи в нечітких умовах. Метод дозволяє об'єднати навіть суперечливі цілі управління, поставивши задачу досягнення максимального прибутку на єдиній основі, що поєднує управління кількістю (максимальна продуктивність) з управлінням якістю (максимальний вихід придатних) як складових елементів системи управління ТП.

СПИСОК ЛІТЕРАТУРИ

1. Ладанюк А. П., Архангельська К. С., Власенко Л. О. Теорія автоматичного керування технологічними об'єктами. К.: НУХТ, 2014. 274 с.
2. Nevliudov I.Sh., Chala O.O., Botsman I.V. Determination of technological process modes for surface formation of substrates for functional components of microoptoelectromechanical systems. *Functional Materials*. 2021. 28(2). P. 381-385.
3. Zadeh L. Fuzzy Sets // *Information and Control*. 1965. Vol. 8. P. 338–353.
4. Зайченко Ю.П. Дослідження операцій. Підручник. Сьоме видання, перероблене та доповнене. К.: Видавничий Дім «Слово», 2006. 816 с.
5. Желдак Т.А., Коряшкіна Л.С., Ус С.А. Нечіткі множини в системах управління та прийняття рішень: за редакцією С.А. Ус. Дніпро: НТУ «ДП», 2020. 387 с.
6. Дубовой В.М., Кветний Р.Н., Михальов О.І., Усов А.В. Моделювання та оптимізація систем: підручник. Вінниця: ПП «ТД«Едельвейс», 2017. 804 с.
7. Nevliudov I., Omarov M., Romashov Y. Numerical methods to solve optimal control problems for technical applications under novel global challenges. *ICONAT 2023 International Conference on Natural Science and Technologies*, 2023.
8. Раскін Л.Г., Сіра О.В., Кожевніков Г.К. Методи аналізу систем і прийняття рішень в умовах невизначеності: підручник. Харків: Факт, 2023. 256 с.
9. Ротштейн А.П. Интеллектуальные технологии идентификации: нечеткие множества, генетические алгоритмы, нейронные сети. Винниця: УНІВЕРСУМ-Вінниця, 1999. 320 с.

Received (Надійшла) 03.05.2024

Accepted for publication (Прийнята до друку) 24.07.2024

Optimization of the strategy of process management in conditions of uncertainty

Anatoliy Andrusevich, Andrei Frolov, Mykola Starodubcev, Nataliia Demaska, Hennadii Makarenko

Abstract. As a result of the analysis of tasks of optimization of technological processes of production of products of new technology in fuzzy conditions, it was found that in the study of a technological control object of low dimension, an approach can be used based on replacing fuzzy sets with sets of α -level and reducing the fuzzy optimization problem to the classical problem of mathematical programming. It is shown that the Bellman-Zade approach based on the principle of merging goals and constraints is more rational for optimizing the management of multi-operational technological processes, which requires further development in case of fuzzy states of the technological control object. In addition, the problem of combining fuzzy goals and restrictions in case of their unequal importance or interdependence is not solved. To solve these problems, a method of optimizing the strategy for managing production processes according to many criteria has been proposed. The convolution of criteria in the form of a fuzzy Sugeno integral is used as a generalized criterion.

Keywords: optimization, control, technological process, technological object, uncertainty of conditions.

Bohdan Boriak, Alina Yanko, Dmytro Ovsii, Maksym Prorok

National University "Yuri Kondratyuk Poltava Polytechnic", Poltava, Ukraine

IMPROVEMENT OF THE CONTROL SYSTEM MODEL OF A MOBILE PLATFORM UNDER THE INFLUENCE OF ELECTROMAGNETIC SPECTRUM THREATS IN THE INFORMATION ENVIRONMENT

Abstract. The article presents a control system model for a mobile robotic platform (unmanned vehicle) and proposes methods for improving the control system elements under the influence of electromagnetic spectrum threats in the information environment. The control model for the robotic platform includes the ability to operate the platform mechanisms in two modes: manual and automated. The use of the mobile platform in manual mode is necessary for operational control in the absence of electromagnetic interference aimed at disrupting the data transmission devices of the robotic platform. To ensure the operation of the mobile platform under the presence of electromagnetic interference, the control system must be capable of managing the platform devices in an automated or automatic mode. The aim of the article is to formalize the control system model for the mobile platform, identify the shortcomings of existing control systems, and explore ways to eliminate these shortcomings. **The results obtained:** quantitative indicators from the conducted research indicate the potential use of video information processing models, used by the unmanned vehicle operator for spatial orientation, for the autonomous operation of the unmanned vehicle. **Conclusions:** the implementation of data processing models in the mobile platform's control system will enable the automation of control processes for the unmanned vehicle under conditions of electromagnetic interference affecting the data transmission devices of the robotic platform.

Keywords: control system, unmanned vehicles, electromagnetic spectrum threats, video processing, mobile platform.

Introduction

Ensuring the safe and correct functioning of systems that utilize information computer technologies presents a range of challenges from the macro level [1] to the micro level.

The aim of this research was to formalize the control system model for a mobile platform, identify the shortcomings of existing control systems, and explore ways to eliminate these shortcomings. With the rapid development of mobile platforms in Ukraine and worldwide, a number of models are available on the market [2–4], whose configurations are characterized by a set of common features from an engineering perspective.

An analysis of the use of existing unmanned vehicles (UVs) indicates that controlling the platform can be complicated due to the limitations of data transmission systems, which, in turn, can lead to damage or loss of the mobile platform. The risks of damage or loss of unmanned vehicles significantly increase when UVs are used under conditions of electromagnetic interference.

Active use of the electromagnetic spectrum leads to the creation of interference that can affect both the control channel and the video data transmission channel, which are used by the mobile platform operator for spatial orientation.

One possible way to mitigate the risks of using UVs under conditions of electronic warfare is to utilize data processing models within the control system itself, which is integrated into the mobile platform.

Control System Model

The control system of a mobile platform can be conditionally divided into two subsystems.

The first subsystem, the manual control system, consists of the following main components:

- Control panel with a radio transmitter.
- Radio signal receiver.
- Antennas for receiving and transmitting control and video signals.
- Video camera and video signal transmitter (VTX).
- Controller (motor controller).
- Electric motors.
- Batteries, voltage converters, switching devices.
- Video output device.
- Actuators (determined based on the specific application of the UV).

In this configuration, direct control of the mobile platform's movement and its actuators is achieved by sending control signals via the control panel. The radio signal receiver ensures the transmission of the signal to the controller, which, in turn, generates the appropriate control signals for the electric motors and actuators of the robotic platform.

The second subsystem is the automated or automatic control system, which is designed to ensure the operation of the mobile platform under conditions of electronic warfare.

The tasks of this system include autonomous video signal processing, decision-making, and the generation of control signals, which the system sends to the controller (motor controller).

As hardware for the automated or automatic control system, single-board computers can be used. These computers are characterized by relatively low power consumption and have sufficient operational functionality for processing video data, generating, and sending control signals through data input/output ports. Single-board computers have the capability to

run software that utilizes algorithms incorporating elements of artificial intelligence [5] and modern data processing algorithms [6–8].

Analysis of Object Tracking Models Usage

To ensure the autonomy of the automated or automatic control system, typical tasks of robotic platforms need to be defined. One such task is the capture and tracking of objects in the surrounding environment.

This task can be addressed using modern video data processing models. One such technology is OpenCV [9], implemented as an open-source computer vision library used for real-time video data processing. This library implements a variety of models for object capture and tracking.

One of the tasks of this research is to analyze models implemented as video data processing algorithms.

Conducting this analysis is necessary to determine the models that can be implemented in the automated or automatic control system.

This research involves several steps:

1. **Implementation of Software:** Develop software to conduct experiments aimed at obtaining quantitative metrics of the performance of different data processing models.

2. **Model Selection:** Choose the data processing model(s) for experimentation using the technical equipment that can be used to implement the automated or automatic control system for the mobile platform.

3. **Implementation of Data Processing Model:** Implement the selected data processing model, conduct testing, and debug the control system.

In this article, the results of the first stage of the research are presented. To conduct the experiment and determine quantitative metrics, software was developed using the C++ programming language and the OpenCV 4.8 library [10].

The following data processing models were used for comparative analysis:

- MIL (Multiple Instance Learning),
- CSRT (Discriminative Correlation Filter with Channel and Spatial Reliability),
- KCF (Kernelized Correlation Filters), BOOSTING, MedianFlow,
- MOSSE (Minimum Output Sum of Squared Error),
- TLD (Tracking, Learning, and Detection).

To determine the quantitative performance metrics of each model, the following criteria were defined:

1. **Average Deviation between Object Center and Calculated Region Center:** This criterion measures the average difference between the center of the tracked object and the center of the region calculated and used by the model for displaying the captured object – the average tracking error.

2. **Average Frame Processing Time:** This criterion measures the average processing time of one video frame.

These criteria were chosen to assess the accuracy and efficiency of each model in tracking objects in real-time video streams.

The experiment involved running the software with the collection of quantitative metrics for 14 videos selected based on the following criteria:

- The video stream depicted changes in the environment with high dynamics in the object's position.
- The videos were obtained using technical equipment (camera) commonly used in modern robotic mobile platform models.

These criteria ensured that the selected videos provided a challenging and realistic scenario for testing the performance of the object tracking models.

Quantitative metrics were obtained for processing each individual frame of the video stream and recorded in a log file, structured and provided in Tables 1 and 2.

The notation "∞" in the tables indicates that the model lost track of the captured object.

The results of the OpenCV models investigation demonstrate that the effectiveness of tracking algorithms can significantly vary depending on the type of video fragment. Some algorithms perform well on videos where objects gradually change their position within the frame, but they may lose tracking on video fragments with unpredictable movements.

Thus, the choice of tracking model should take into account the specific requirements of the task and the characteristics of the video stream.

Analytically, the following characteristics of the models can be determined:

- KCF and MOSSE quickly process frames and perform well in tracking objects with uniform and smooth motion without abrupt movements. However, these algorithms are not suitable for tracking objects with unpredictable trajectories, as the object loss rate and tracking error are quite high.

• TLD can handle videos where objects enter obstacles or go beyond frame boundaries. It restores tracking but is characterized by longer frame processing time and higher tracking error compared to other models.

• MedianFlow demonstrates moderate quality for objects with uneven motion, but its accuracy heavily depends on the initially selected region of interest (ROI). It's important to precisely define object boundaries to avoid background capture.

• CSRT and BOOSTING, according to the obtained quantitative metrics, are characterized by relatively high performance in the conducted experiment, but BOOSTING has a sufficiently high frame processing time, and CSRT more frequently loses track of the object.

• MIL algorithm, based on the experiment results, is characterized by high quality in tracking objects. According to observations, this model enables object tracking in dynamic scenarios, with moderate frame processing time and the lowest average tracking error and object loss percentage.

Table 1 – Average Tracking Error (px)

No. of video	Model implemented in OpenCV						
	MIL	CSRT	KCF	BOOSTING	MedianFlow	MOSSE	TLD
1.	35.9	∞	∞	113.2	31.3	∞	∞
2.	45.4	∞	147.4	37.5	125.3	97.1	∞
3.	18.6	18.3	∞	16.9	∞	∞	65.3
4.	21.3	28.4	∞	39.9	107.6	∞	∞
5.	29.3	∞	∞	57.6	58.9	35.4	53.5
6.	38.0	117.7	∞	50.8	124.6	∞	52.9
7.	60.2	14.0	∞	21.2	43.1	∞	∞
8.	12.7	15.0	∞	15.2	118.4	∞	26.3
9.	12.5	14.2	∞	13.3	∞	∞	∞
10.	∞	∞	∞	∞	72.5	∞	∞
11.	37.7	7.5	∞	11.0	∞	∞	∞
12.	9.4	9.9	∞	42.2	110.9	∞	150.4
13.	8.1	∞	∞	∞	29.3	∞	∞
14.	∞	∞	∞	6.1	30.0	∞	64.6
Average tracking error, px	27,43	28,13	147,40	35,41	77,45	66,25	68,88
Object loss percentage, %	14,29	42,86	92,86	14,29	21,43	85,71	57,14

Table 2 – Average frame processing time (ms)

No. of video	Model implemented in OpenCV						
	MIL	CSRT	KCF	BOOSTING	MedianFlow	MOSSE	TLD
1.	243.9	∞	∞	415.0	48.2	∞	∞
2.	262.4	∞	158.6	698.0	153.4	73.9	∞
3.	237.0	166.6	∞	432.6	∞	∞	949.6
4.	215.0	122.7	∞	372.5	130.9	∞	∞
5.	236.4	∞	∞	533.9	72.7	13.8	598.1
6.	255.9	117.7	∞	475.5	144.1	∞	853.0
7.	260.8	149.4	∞	395.1	53.7	∞	∞
8.	206.5	103.4	∞	338.1	89.8	∞	966.0
9.	193.3	119.2	∞	341.7	∞	∞	∞
10.	∞	∞	∞	∞	44.0	∞	∞
11.	220.7	123.6	∞	393.8	∞	∞	∞
12.	218.9	110.7	∞	372.0	81.8	∞	1276.3
13.	195.3	∞	∞	∞	36.7	∞	∞
14.	∞	∞	∞	308.1	41.8	∞	961.2
Average processing time (ms)	228,84	126,66	158,60	423,03	69,01	43,85	934,03

Conclusions

This article presents a formalized model of ground mobile platform control system, identifies the main technical components, and principles of their interaction.

An analysis of the drawbacks of existing platforms under active electromagnetic spectrum utilization is conducted. A combined control system model utilizing manual and automated or automatic control subsystems is proposed.

An overview of typical tasks for unmanned vehicles operation is provided, and avenues for

improving the control system model of UVs to eliminate existing drawbacks are identified.

An experiment is conducted to investigate the potential use of modern object tracking models, obtaining quantitative metrics.

Based on these metrics, a data processing model is selected for further implementation into the mobile platform control system.

The next steps of this research involve conducting an experiment to obtain quantitative indicators of the data processing model performance based on the technical equipment used for implementing the mobile platform control system.

REFERENCES

1. Victor Krasnobayev, Alina Yanko and Alina Hlushko. Information Security of the National Economy Based on an Effective Data Control Method. Journal of International Commerce, Economics and Policy, 14(3), 2023, article no. 2350021. DOI: 10.1142/S1793993323500217.
2. До експлуатації в ЗСУ допустили понад 40 БПЛА та понад 20 наземних роботів [Електронний ресурс] // ГО «Український мілітарний центр». – 2024. – Режим доступу до ресурсу: <https://mil.in.ua/uk/news/do-ekspluatatsiyi-v-zsu-dopustyly-ponad-40-bpla-ta-ponad-20-nazemnyh-robotiv/>.
3. Milrem Robotics відкрила новий завод наземних дронів [Електронний ресурс] // ГО «Український мілітарний центр». – 2024. – Режим доступу до ресурсу: <https://mil.in.ua/uk/news/milrem-robotics-vidkryla-novyj-zavod-nazemnyh-droniv/>.
4. В Україні протестували наземні безпілотники [Електронний ресурс] // ГО «Український мілітарний центр». – 2023. – Режим доступу до ресурсу: <https://mil.in.ua/uk/news/v-ukrayini-protestuvaly-nazemni-bezpylotnyky/>.
5. Shefer, O., Laktionov, O., Pents, V., Hlushko, A., & Kuchuk, N. (2024). Practical principles of integrating artificial intelligence into the technology of regional security predicting. Advanced Information Systems, 8(1), 86–93. <https://doi.org/10.20998/2522-9052.2024.1.11>
6. Yanko A. Method for the control verification of digital information, represented in a residue number system / Yanko A., Krasnobayev V., Tur S. // Advanced Information Systems. 2020. Vol. 4, No. 2. ISSN2522-9052. – P. 149-156.
7. Boriak B. R. Filtering and forecasting signals algorithm based on exponential brown's filter / B. R. Boriak, A. M. Silvestrov // Control, Navigation and Communication Systems. Academic Journal. – Poltava: PNTU, 2017. – VOL. 4 (44). – PP. 150–152.
8. Боряк Б. Р. Аналіз ефективності застосування адаптивного двоконтурного фільтра-екстраполятора залежно від форми сигналу / Б. Р. Боряк, А. М. Сільвестров // Системи управління, навігації та зв'язку. Збірник наукових праць. – Полтава: ПНТУ, 2017. – Т. 6 (46). – С. 255–259.
9. OpenCV Tracking API Documentation [Електронний ресурс]. – 2023. – Режим доступу до ресурсу: https://docs.opencv.org/4.x/d9/df8/group_tracking.html
10. Object Tracking using OpenCV (C++/Python) Tutorial [Електронний ресурс] – 2017. – Режим доступу до ресурсу: <https://learnopencv.com/object-tracking-using-opencv-cpp-python/>

Received (Надійшла) 03.07.2024

Accepted for publication (Прийнята до друку) 21.08.2024

Удосконалення моделі системи керування мобільної платформи в умовах дії загроз електромагнітного спектру інформаційного середовища

Б. Р. Боряк, А. С. Янко, Д. Овсій, М. Ю. Пророк

Анотація. В статті наведено модель системи керування мобільною роботизованою платформою (безпілотного апарату), а також запропоновано методи удосконалення елементів системи керування, в умовах дії загроз електромагнітного спектру інформаційного середовища. Модель керування роботизованою платформою передбачає можливість управління механізмами роботизованої платформи у двох режимах: ручному та автоматизованому. Використання мобільної платформи в ручному режимі необхідне для оперативного керування за умов відсутності дії електромагнітного впливу з метою створення завад на пристрої передачі даних роботизованої платформи. Для забезпечення оперування мобільної платформи в умовах наявності дії електромагнітного впливу система керування повинна мати можливість керувати пристроями мобільної платформи в автоматизованому або автоматичному режимі. **Метою статті** є формалізація моделі системи керування мобільною платформою, виявлення недоліків існуючих систем керування та пошук шляхів усунення цих недоліків. **Отримані результати:** кількісні показники проведеного дослідження вказують на можливість використання моделей обробки відеоінформації, яка використовується оператором безпілотного апарату для орієнтації в просторі, для автономної роботи безпілотного апарату. **Висновки:** імплементація моделей обробки даних в систему керування мобільною платформою дозволить автоматизувати процеси керування безпілотного апарату в умовах дії електромагнітного впливу на пристрої передачі даних роботизованої платформи.

Ключові слова: система керування, безпілотні апарати, загрози електромагнітного спектру, обробка відео, мобільна платформа.

О. Ю. Заковортний, Т. О. Орлова

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

ВИКОРИСТАННЯ НЕЧІТКИХ МНОЖИН ДРУГОГО ТИПУ ДЛЯ ПРОЕКТУВАННЯ РЕГУЛЯТОРІВ ШВИДКОСТІ

Анотація. Інтервальні нечіткі множини другого типу використовуються при проектуванні нечітких систем, пов'язаних з керуванням швидкістю в умовах невизначеності. У статті розглядається можливість використання інтервальних нечітких множин другого типу для опису змінних, що можуть бути різними значеннями однієї і тієї ж динамічної системи або змінами технічних параметрів, що відбуваються під впливом зовнішніх факторів. Також представлено складні інтервальні нечіткі множини другого типу та наведено приклади їх використання для зменшення кількості нечітких множин та правил у нечіткій базі знань. Також наведено необхідні умови для побудови складних інтервальних нечітких множин другого типу та методи визначення нижньої та верхньої функцій належності в них. Слід невизначеності для складної інтервальної нечіткої множини другого типу представлено верхніми та нижніми функціями належності різних нечітких множин. Всі методи, описані в роботі, можуть бути використані при проектуванні систем підтримки прийняття рішень, для вибору оптимальних параметрів нелінійної системи або при проектуванні нечітких регуляторів швидкості.

Ключові слова: інтервальна нечітка множина другого типу; нечіткі системи; керування швидкістю; система підтримки прийняття рішень; нечіткі регулятори.

Вступ

Нечітка логіка широко застосовується для розв'язання задач, з якими стикаються дослідники в різних галузях, таких як прогнозування часових рядів [1], прийняття рішень [2], автоматизоване управління [3, 4], розпізнавання об'єктів [5] та інших задач [6]. Розглянемо застосування інтервальних нечітких множин другого типу в проектуванні регулятора швидкості та дамо визначення предмету дослідження. Регулятори швидкості використовуються для контролю або регулювання швидкості руху системи або процесу. Існують різні типи регуляторів, включаючи пропорційно-інтегрально-диференціальні (PID) регулятори, адаптивні регулятори, нечіткі регулятори та інші. Регулятори швидкості працюють на основі зворотного зв'язку, де вимірювальні сигнали порівнюються зі заданими значеннями і видаються керуючі сигнали для підтримання встановленої швидкості. Регулятори швидкості використовуються в широкому спектрі технічних систем, включаючи робототехніку, автомобільну промисловість, електроенергетику, авіаційну та космічну техніку, виробництво і багато інших галузей. Деякі з викликів, пов'язаних з регуляторами швидкості, включають точність регулювання, стабільність системи, компенсацію збурень та невизначеностей, а також врахування динамічних властивостей об'єктів керування.

Метою роботи є розгляд властивостей інтервальних нечітких множин другого типу для опису невизначеностей з якими працюють системи підтримки прийняття рішень під час пошуку оптимальних параметрів керування. Оцінка площі невизначеності в інтервальних нечітких множинах другого типу через визначення нижньої та верхньої функцій належності різних видів. Проектування інтервальних нечітких множин другого типу, що враховують різні форми невизначеності.

Огляд існуючих методів проектування регуляторів швидкості

При проектуванні регуляторів можуть бути прийняті до уваги такі наступні класичні види регу-

ляторів, як: PID-регулятори, адаптивні алгоритми, системи на основі штучних нейронних мереж та інші. Розглянемо деякі з них.

PID-регулятори є одними з найпоширеніших та найбільш використовуваних методів в системах автоматичного керування. Основною ідеєю PID-регулятора є керування величиною виходу системи, так щоб вона досягла заданого значення. Вони використовуються для стабілізації, регулювання та керування системами зворотного зв'язку, але мають деякі недоліки: надчутливість до параметрів (неадекватні значення параметрів можуть призвести до нестабільності або недосяжності цільового значення); затримка реакції (можуть бути повільними у реакції на зміни в системі, особливо коли інтегральний (I) та диференціальний (D) компоненти недостатньо налаштовані); чутливість до збурень та шуму в системі; непрактичність у складних системах або велику кількість невідомих параметрів.

Адаптивні алгоритми - це методи управління, які можуть автоматично змінювати свої параметри та структуру з метою адаптації до змін у середовищі або об'єкті керування. До переваг адаптивних алгоритмів відноситься: автоматична адаптація (гнучкість до умов), стійкість до змін, широкий спектр застосувань; можливість врахування невизначеності та забезпечення стійкості до неї; швидка реакція на зміни. Ці алгоритми попри досить високу ефективність також мають свої недоліки: складність налаштування через складну систему або велику кількість параметрів; деякі алгоритми вимагають значних обчислювальних ресурсів; потреба у великій кількості даних для ефективного навчання; нестійкість до шуму; складність інтерпретування результатів, які базуються на складних математичних моделях.

Системи на основі штучних нейронних мереж (ШНМ) моделюють структуру та функції нервової системи людини за допомогою комп'ютерного програмування [7]. До переваг цих систем відноситься: ШНМ моделюють біологічні нейронні мережі із взаємопов'язаних штучних нейронів, які обробляють інформацію за аналогією з реальними нейронами мережі

людини; ШНМ можуть навчатися з вчителем (коли є наявні вхідні та вихідні дані для навчання) або без вчителя (коли навчання відбувається без прямого використання вхідно-вихідних пар даних), ШНМ можуть бути використані для різних завдань, включаючи розпізнавання образів, прогнозування, класифікацію, оптимізацію та управління системами; ШНМ виявляють високу стійкість до шуму в даних та здатність адаптуватися до змін у вхідних параметрах. Однак є і недоліки: іноді складно інтерпретувати рішення, прийняті нейронною мережею, через її складну структуру та процеси внутрішньої обробки інформації; деякі типи нейронних мереж можуть вимагати значних обчислювальних ресурсів та часу для навчання; ШНМ можуть вимагати великої кількості даних для навчання, особливо для складних завдань, а недостатність даних може призвести до недооцінки або неправильного навчання моделі; мережі можуть страждати від перенавчання, коли модель надмірно адаптується до конкретних тренувальних даних і втрачає здатність узагальнення до нових даних.

Нечіткі множини другого типу (НМТ2) (які є вдосконаленням традиційних нечітких множин типу 1) – дозволяють моделювати та обробляти невизначеності більш ефективно завдяки двошаровій структурі, яка включає як первинні, так і вторинні функції належності. Це розширення дозволяє враховувати не лише основні невизначеності, але й варіативність цих невизначеностей. Нечіткі множини другого типу є розширенням звичайних нечітких множин, які дозволяють враховувати більшу невизначеність та неоднорідність у даних. Вони включають в себе не тільки нечіткість в межах значень, але й нечіткість в самій формі функцій належності. Нечіткі множини другого типу можуть бути корисним інструментом у проектуванні регуляторів швидкості, оскільки вони дозволяють враховувати та управляти складністю та невизначеністю в системах керування. Вони можуть забезпечити більш точне та стабільне керування, особливо в умовах зміни середовища або об'єктів керування. Використання нечітких множин другого типу може покращити якість регулювання, зменшити вплив збурень та невизначеностей, а також забезпечити більш гнучке та адаптивне керування в різних умовах роботи систем. Хоча нечіткі множини використовуються для опису та обробки нечітких даних, нечіткі множини типу 1 (НМТ1) мають функції належності (ФН), які є чіткими числами. Недоцільно використовувати чіткі функції для опису невизначеності та ненадійності даних [8]. У 1975 році Заде представив нечіткі множини другого типу як розширення звичайних ФН, які зараз відомі як функції належності нечітких множин типу 1 [9].

Невизначеності в нечітких системах другого типу

Інтраневизначеність та інтерневизначеність, як поняття невизначеностей різного типу при побудові функцій належності в нечітких системах розглядаються у роботі [10]. Інтраневизначеність виникає у суб'єкта або експерта через брак інформації про об'єкт дослідження чи неоднозначну інтерпретацію певного слова (лінгвістичної змінної). Виникає інтерневизна-

ченість в результаті різних оцінок від різних експертів термів лінгвістичної змінної. Описана інтраневизначеність може бути як інтервальні нечіткі множини другого типу (ІНМТ2), тоді як інтерневизначеність – через об'єднання різних ІНМТ2. При розробці систем нечіткого регулювання невизначеності можуть виникати не лише під час опису лінгвістичних змінних, але й у процесі функціонування комп'ютерної системи та об'єкта керування. Підстави, при яких можуть виникнути невизначеності різного типу надано на рис. 1.

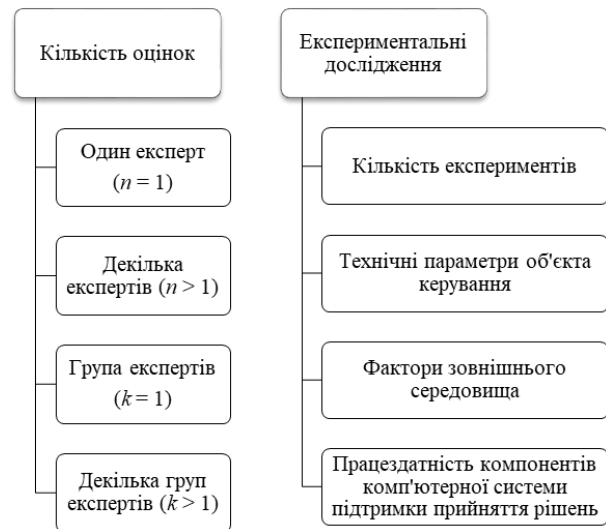


Рис. 1. Невизначеності при створенні систем нечіткого логічного виведення другого типу

Перш ніж розробляти системи нечіткого виводу другого типу, рекомендується перевірити на наявність статистичних невизначеностей вхідні дані. Це має значення для врахування нестандартних ситуацій, що стосуються ефективності компонентів комп'ютерної системи. При наявності набору даних, без статистичних невизначеностей або якщо відомі причини їх виникнення, можна створювати нечіткі множини другого типу. На рис. 1 показано як, при оцінці термів лінгвістичної змінної експертами, виникають невизначеності. Коли визначена модель однією оцінкою ($n = 1$), це спричиняє виникнення інтраневизначеності, і для її вирішення ми користуємось ІНМТ2. У цьому дослідженні оцінка однієї групи експертів ($k = 1$) також вважається інтраневизначеністю, оскільки рішення їхні повинні узгоджуватись. Крім того, ця оцінка може бути представлена як інтерневизначеність завдяки об'єднанню кожної окремої оцінки. Виконується, як правило, при $k = n = 1$ умова:

$$FOU(k) \geq FOU(n), \quad (1)$$

де $FOU(k)$ (footprint of uncertainty) – площа невизначеності для оцінки групою експертів, $FOU(n)$ – площа невизначеності для оцінки одного експерта.

Оцінки експертів щодо конкретного терміну, незалежно від причин виникнення невизначеностей, можуть змінюватися після ряду досліджень. Ці зміни можуть охоплювати збільшення кількості експериментів, коригування технічних параметрів об'єкта керування, уточнення впливу зовнішнього середовища і перевірку надійності компонентів комп'ютерної системи. Після збору оновлених даних внаслідок експериментів, варто

перевірити параметри та види функції належності другого типу, а також кількість термів для кожної лінгвістичної змінної. Для поліпшення системи нечіткого виводу корисним може бути моделювання нових лінгвістичних змінних, особливо при визначенні нових факторів зовнішнього середовища. Не слід проектувати нечіткі системи другого типу, якщо вхідні дані містять взаємозалежні параметри різних компонентів комп'ютерної системи, особливо коли цей зв'язок не досліджений і залежить від зовнішніх факторів. Це може збільшити похибку при перетворенні типів функцій належності на виході нечіткої системи.

Інтервальні нечіткі множини другого типу

У роботі [10] визначенні нечіткі множини другого типу, основні операції над ними та зазначено, що при проектуванні нечітких систем другого типу може бути використано математичний апарат нечітких множин типу-1. Нечітка множина другого типу (\tilde{A}) на універсальній множині X визначена функцією належності другого типу – $\mu_{\tilde{A}}(x, u)$, де $0 \leq \mu_{\tilde{A}}(x, u) \leq 1$ [11]:

$$\tilde{A} = \{(x, u), \mu_{\tilde{A}}(x, u) \mid \forall x \in X, \forall u \in J_x \subseteq [0,1]\}, \quad (2)$$

де J_x – первинна належність x . Об'єднання значень належностей для всіх допустимих значень x і u може бути іншим способом до визначення нечітких множин другого типу:

$$\tilde{A} = \int_{x \in X} \int_{u \in J_x} \mu_{\tilde{A}}(x, u) / (x, u), \quad J_x \subseteq [0,1]. \quad (3)$$

Через високу обчислювальну складність роботи з нечіткими множинами, при розробці систем нечіткого регулювання застосовуються інтервальні нечіткі множини другого типу.

ІНМТ2 є нечіткою множиною, де виконується умова – для всіх $\mu_{\tilde{A}}(x, u) = 1$. Отже ІНМТ2 є підмножиною нечітких множин другого типу та визначається наступним чином [12]:

$$\tilde{A} = \int_{x \in X} \int_{u \in J_x} 1 / (x, u), \quad J_x \subseteq [0,1]. \quad (4)$$

Для визначення інтервальних нечітких множин другого типу достатньо визначити область невизначеності (FOU), що описується через дві функції належності типу-1: нижньої ($\underline{\mu}_{\tilde{A}}(x)$) та верхньої ($\bar{\mu}_{\tilde{A}}(x)$), між якими виконується нерівність:

$$\underline{\mu}_{\tilde{A}}(x) \leq \bar{\mu}_{\tilde{A}}(x), \quad \forall x \in X. \quad (5)$$

Однією з основних задач при проектуванні систем нечіткого логічного виведення є визначення розміру області невизначеності (FOU), адже її збільшення впливає на точність моделі та збільшує час обчислень у комп'ютерній системі. Іншою важливою задачею є вибір типів функцій належності. На сьогодні, у розробці систем нечіткого логічного виведення найбільш застосовуваними функціями належності другого типу є гаусівська (gaussmf), трикутна (trimf), трапецієподібна (trapmf) та дзвоноподібна (gbellmf). На рис. 2 наведені приклади цих функцій належності другого типу. У [13] використані для моделювання невизначеності нестандартні функції належності другого типу.

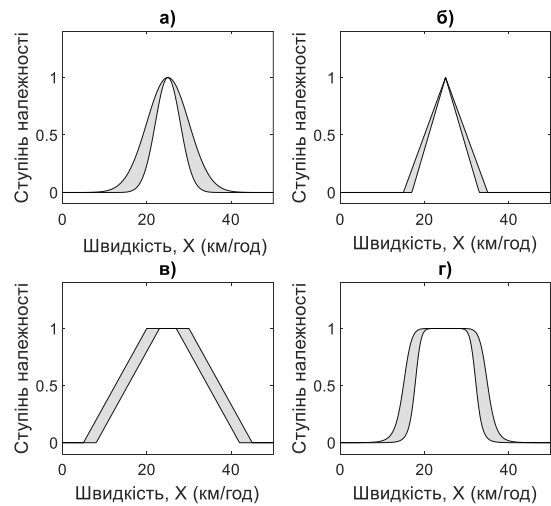


Рис. 2. Функції належності типу-2: а – гаусова (gaussmf); б – трикутна (trimf); в – трапецієподібна (trapmf); г – дзвоноподібна (gbellmf)

Рекомендується при розробці систем нечіткого регулювання використовувати функції належності з мінімальною кількістю параметрів і тому, для побудови функцій належності другого типу використовується гаусова функція належності типу-1:

$$\mu_A(x_{m,\sigma}) = \exp\left[-0.5 \cdot \left(\frac{(x-m)}{\sigma}\right)^2\right], \quad (6)$$

де $\mu_A(x_{m,\sigma})$ – гаусова функція належності типу-1, σ – параметр відхилення (крутизна функції), m – центр нечіткої множини.

Проаналізуємо методи визначення нечітких множин другого типу для моделі інтраневизначеності, де маємо одного експерта ($n = 1$).

Як вже зазначалося, для побудови нечітких множин другого типу достатньо визначити дві функції належності типу-1. У випадку гаусової функції належності існують три варіанти визначення нечітких множин другого типу з використанням таких функцій належності типу-1 [13]:

$$\mu_{\tilde{A}}(x) = f(m_1, m_2, \sigma); \quad (7)$$

$$\mu_{\tilde{A}}(x) = f(m, \sigma_1, \sigma_2); \quad (8)$$

$$\mu_{\tilde{A}}(x) = f(m_1, m_2, \sigma_1, \sigma_2). \quad (9)$$

Хоча практично варіант (9) майже не використовують. Надалі, задається умова: $\sigma_1 < \sigma_2$ для визначення нижньої та верхньої функції належності. З урахуванням нерівності (5) значення нижньої функції належності $\underline{\mu}_{\tilde{A}}(x)$ для (8) описується як [13]:

$$\underline{\mu}_{\tilde{A}}(x) = f(x, m, \sigma_1) = \exp\left[-0.5 \cdot \left(\frac{(x-m)}{\sigma_1}\right)^2\right]. \quad (10)$$

Матимемо для верхньої функції належності:

$$\bar{\mu}_{\tilde{A}}(x) = f(x, m, \sigma_2) = \exp\left[-0.5 \cdot \left(\frac{(x-m)}{\sigma_2}\right)^2\right]. \quad (11)$$

Ще можуть бути визначені нижня та верхня функції належності за допомогою коефіцієнта відхилення для значень $\sigma = k_\sigma$. Для коефіцієнта необхідна така умова: $0 \leq k_\sigma < 1$. Якщо $k_\sigma = 0$, то $\underline{\mu}_{\tilde{A}}(x) = \bar{\mu}_{\tilde{A}}(x)$.

Отже ми отримаємо функцію належності типу-1. У системі нечіткого регулювання коефіцієнт k_σ може використовуватися у трьох різних формах: для визначення нижньої функції належності; для визначення верхньої функції належності; для визначення нижньої та верхньої функцій належності. До того ж його значення напряму визначає розмір площі невизначеності. На рис. 3 показані площі невизначеності при різних значеннях коефіцієнта k_σ , що визначають значення для нижньої функції належності.

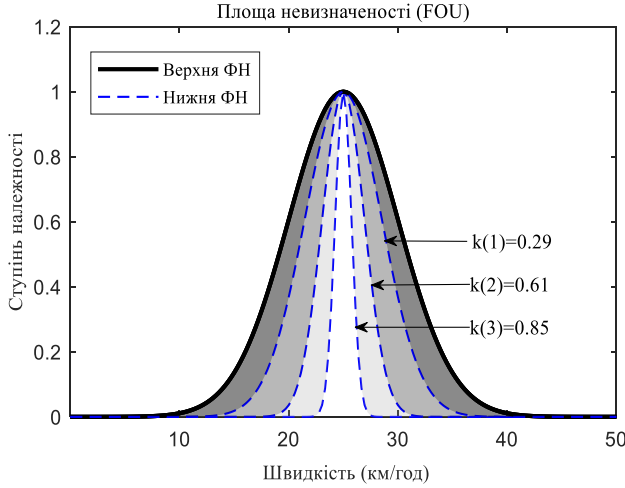


Рис. 3. Площі невизначеності для різних значень коефіцієнту відхилення k_σ ($\sigma = 5$, $m = 25$)

Аналізуючи рис. 3, бачимо, що зростання значення k_σ спричиняє зростання розміру FOU . Враховуючи нерівність (1) для оцінки розміру FOU , матимемо таку нерівність для значень коефіцієнта k_σ для одного експерта ($n = 1$) та 1 групи експертів ($k = 1$):

$$k_\sigma(k) \geq k_\sigma(n). \quad (12)$$

Використовуючи коефіцієнт k_σ для визначення двох функцій належності отримаємо наступний вираз для нижньої функції належності $\underline{\mu}_{\tilde{A}}(x)$:

$$\underline{\mu}_{\tilde{A}}(x) = f(x, m, \sigma, k) = \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma - k_\sigma \cdot \sigma}\right)^2\right]. \quad (13)$$

Верхня функції належності $\bar{\mu}_{\tilde{A}}(x)$ описується як:

$$\bar{\mu}_{\tilde{A}}(x) = f(x, m, \sigma, k) = \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma + k_\sigma \cdot \sigma}\right)^2\right]. \quad (14)$$

При проектуванні системи нечіткого регулювання зміна коефіцієнту k_σ ймовірна після проведення експериментальних випробувань. Тобто, після повторних експериментів значення k_σ може зменшитися, а при врахуванні факторів зовнішнього середовища або збоїв роботи комп'ютерної системи підтримки прийняття рішень, може збільшитися. Для визначення ІНМТ2 використовують декілька коефіцієнтів k_σ з окремою вагою з урахуванням оцінок між експертами в межах 1 групи.

Інтерневизначеність в системах нечіткого логічного виведення

Методи, що застосовуються для визначення інтраневизначеностей при проектуванні ІНМТ2 в системах нечіткого регулювання, можуть також бути застосовані для оцінки інтерневизначеностей (кількість експертів $n > 1$, кількість груп експертів $k > 1$). Однак існують особливості в визначенні параметрів ІНМТ2, у т.ч. функції (7-9) визначаються кількістю параметрів відповідної кількості експертів та груп. Отже отримаємо функції ($n > 1$), що мають декілька параметрів для центру нечіткої множини та відхилення:

$$\mu_{\tilde{A}(n)}(x) = f(m_1, m_2, \dots, m_n; \sigma); \quad (15)$$

$$\mu_{\tilde{A}(n)}(x) = f(m; \sigma_1, \sigma_2, \dots, \sigma_n); \quad (16)$$

$$\mu_{\tilde{A}(n)}(x) = f(m_1, m_2, \dots, m_n; \sigma_1, \sigma_2, \dots, \sigma_n). \quad (17)$$

Для декількох груп експертів ($k > 1$) матимемо:

$$\mu_{\tilde{A}(k)}(x) = f(m_1, m_2, \dots, m_k; \sigma); \quad (18)$$

$$\mu_{\tilde{A}(k)}(x) = f(m; \sigma_1, \sigma_2, \dots, \sigma_k); \quad (19)$$

$$\mu_{\tilde{A}(k)}(x) = f(m_1, m_2, \dots, m_k; \sigma_1, \sigma_2, \dots, \sigma_k). \quad (20)$$

Для опису нижньої $\underline{\mu}_{\tilde{A}(n,k)}(x)$ та верхньої $\bar{\mu}_{\tilde{A}(n,k)}(x)$ функцій належності, потрібно визначити σ_{\min} та σ_{\max} , при фіксованому значенні m (19). У такому разі нижня функція належності виглядатиме так:

$$\underline{\mu}_{\tilde{A}(n,k)}(x) = f(x, m, \sigma_{\min}) = \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma_{\min}}\right)^2\right]. \quad (21)$$

Верхня функція належності описується з використанням максимального значення відхилення у гаусовій функції належності для випадку інтерневизначеності:

$$\bar{\mu}_{\tilde{A}(n,k)}(x) = f(x, m, \sigma_{\max}) = \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma_{\max}}\right)^2\right]. \quad (22)$$

Отримуємо множину $\tilde{A}(n,k)$, яка є складеною, тобто визначеною декількома функціями належності другого типу [13, 14]. У цьому випадку площа невизначеності описується двома функціями належності типу-1, які можуть підмножинами різних множин другого типу. Водночас FOU визначається незалежно від кількості експертних оцінок, які використовуються у задачі системи нечіткого логічного виведення.

FOU для $\tilde{A}(n,k)$ може виглядати таким чином:

$$FOU(\tilde{A}(n,k)) = \bigcup_{x \in X} \left[\underline{\mu}_{\tilde{A}(n,k)}(x), \bar{\mu}_{\tilde{A}(n,k)}(x) \right]. \quad (23)$$

Проте ІНМТ2 може визначатись різними типами функцій належності, що ускладнює визначення нижньої, верхньої функцій належності, а також опис площі невизначеності. Коли немає функцій належності, де виконується умова:

$$\underline{\mu}_{\tilde{A}(n,k)}(x) \leq \mu_{\tilde{A}(n,k)}(x') \leq \bar{\mu}_{\tilde{A}(n,k)}(x), \quad \forall x \in X, \quad (24)$$

де $\mu_{\tilde{A}(n,k)}(x')$ – будь-яка функція належності, яка є підмножиною FOU , то площу невизначеності слід

описувати як об'єднання декількох *FOU*, у яких площі менші та виконується умова (24).

Висновки

Невизначеності різних типів, які виникають при оцінюванні експертами параметрів регуляторів швидкості, мають вплив на типи і кількість функцій належності, а також на розміри *FOU*. Розглянуті характеристики функцій належності і методи визначення *FOU* для різних типів невизначеностей не є остаточними, оскільки є можливість зміни розмірів *FOU* (збільшення або зменшення) після проведення експериментів.

Необхідно використовувати коефіцієнти відхилення (регулювання) параметрів під час розробки комп'ютерних систем підтримки прийняття рішень. Для інтраневизначеності коефіцієнт застосовується до параметрів функцій належності і необхідно досліджувати розміри *FOU*, оскільки збільшення площі

невизначеності погіршує точність роботи системи нечіткого регулювання. При інтерневизначеності *FOU* визначається декількома функціями належності другого типу і необхідно використовувати коефіцієнт ваги, який залежить від знань та досвіду експертів для кожної з них. Варто також дослідити побудову ІНМТ2 з різними функціями належності, оскільки це ускладнює визначення *FOU*.

ІНМТ2 дозволяють моделювати складніші невизначеності, ніж ІНМТ1. Це особливо корисно для систем, де дані можуть бути неповними або неоднозначними. Завдяки своїй структурі ІНМТ2 можуть краще справлятися з шумами та помилками в даних, що робить їх більш стійкими у порівнянні з традиційними методами. ІНМТ2 надають можливість більш гнучкого налаштування параметрів системи, що дозволяє досягати оптимальних характеристик керування в різних умовах експлуатації.

СПИСОК ЛІТЕРАТУРИ

1. Reddy, P. V. S. Fuzzy logic based on Belief and Disbelief membership functions. *Fuzzy Information and Engineering*, 2017, 9(4), pp. 405-422, doi: 10.1016/j.fiae.2017.12.001.
2. Babanezhad, M., Masoumian, A., Nakhjiri, A. T. Influence of number of membership functions on prediction of membrane systems using adaptive network based fuzzy inference system. *Scientific Reports*, 2020, vol. 10(1), pp. 1-20, doi: 10.1038/s41598-020-73175-0.
3. Pelalak, R., Nakhjiri, A. T., Marjani, A. Influence of machine learning membership functions and degree of membership function on each input parameter for simulation of reactors. *Scientific Reports*, 2021, 11(1), pp.1-11, doi:10.1038/s41598-021-81514-y.
4. Razak, T.R., Garibaldi, J.M., Wagner, C., Pourabdollah, A., Soria, D., Toward a Framework for Capturing Interpretability of Hierarchical Fuzzy Systems—A Participatory Design Approach. *IEEE Transactions on Fuzzy Systems*, 2020, 29(5), pp.1160-1172, doi: 10.1109/TFUZZ.2020.2969901.
5. Zadeh, L. A. The concept of a linguistic variable and its application to approximate reasoning-I. *Information sciences*, 1975, vol. 8(3), pp.199-249. doi: 10.1016/0020-0255(75)90036-5.
6. Liang Q., Mendel J. M. Interval type-2 fuzzy logic systems: theory and design. *IEEE Transactions on Fuzzy Systems*, 2000, 8(5), pp. 535-550, doi: 10.1109/91.873577.
7. Zakovorotniy A.Y. Neural Networks Art: Solving Problems with Multiple Solutions and New Teaching Algorithm / A.Y. Zakovorotniy, V.D. Dmitrienko, S.Yu. Leonov, I.P. Khavina // The Open Neurology Journal. – 2014. – № 8. – P. 15 – 21.
8. Mendel, J. M. Type-2 fuzzy sets: some questions and answers. *IEEE Connections, Newsletter of the IEEE Neural Networks Society*, 2003, 1, pp. 10-13.
9. Devenci, M., Cali, U., Kucuksari, S. and Erdogan, N. Interval type-2 fuzzy sets based multi-criteria decision-making model for offshore wind farm development in Ireland. *Energy*, 2020, p.117317, doi:10.1016/j.energy.2020.117317.
10. Biswas, R., Sil, J. An improved canny edge detection algorithm based on type-2 fuzzy sets. *Procedia Technology*, 2012, 4, pp. 820-824, doi: 10.1016/j.protcy.2012.05.134.
11. Zhao, F., Chen, Y., Liu, H., Fan, J. Alternate PSO-based adaptive interval type-2 intuitionistic fuzzy C-means clustering algorithm for color image segmentation. *IEEE Access*, 2019, 7, pp. 64028-64039, doi: 10.1109/ACCESS.2019.2916894.
12. Yatak, M. Ö., Şahin, F. Ride Comfort-Road Holding Trade-off Improvement of Full Vehicle Active Suspension System by Interval Type-2 Fuzzy Control. *Eng. Science and Techn.*, 2021, 24(1), pp. 259-270, doi: doi.org/10.1016/j.jestech.2020.10.006.
13. Zakovorotniy, A., Kharchenko, A. Optimal Speed Controller Design with Interval Type-2 Fuzzy Sets. In 2021 *IEEE 2nd KhPI Week on Advanced Technology (KhPIWeek)*, 2021 pp. 363-366, doi:10.1109/KhPIWeek53812.2021.9570045.
14. Заковоротний О.Ю. Синтез оптимальних законів управління рухом дизель-поїзда за допомогою математичної моделі у формі Бруновського / О.Ю. Заковоротний, В.Д. Дмитрієнко, Н.В. Мезенцев // Інформаційно-керуючі системи на залізничному транспорті. – Харків, 2010. – № 5-6 (84-85). – С. 7 – 13.

Received (Надійшла) 17.05.2024

Accepted for publication (Прийнята до друку) 31.07.2024

The use of fuzzy sets of the second type for the design of speed controllers

O. Zakovorotnyi, T. Orlova

Abstract. Interval fuzzy sets of the second type are used in the design of fuzzy systems related to speed control under conditions of uncertainty. The article considers the possibilities of using interval fuzzy sets of the second type to describe variables that can be different values of the same dynamic system or changes in technical parameters that occur under the influence of external factors. Complex interval fuzzy sets of the second type are also presented and examples of their use for reducing the number of fuzzy sets and rules in the fuzzy knowledge base are given. The necessary conditions for constructing complex interval fuzzy sets of the second type and methods for determining the lower and upper membership functions in them are also given. The uncertainty trace for a complex interval fuzzy set of the second type is represented by the upper and lower membership functions of different fuzzy sets. All the methods described in the work can be used in the design of decision support systems, to select the optimal parameters of a nonlinear system, or in the design of fuzzy speed controllers.

Keywords: interval fuzzy set of the second type; fuzzy systems; speed control; decision support system; unclear regulators.

М. Й. Заповловський, М. В. Мезенцев, М. В. Оліфір

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

АДАПТАЦІЯ АЛГОРИТМУ ВЕКТОРНОГО КЕРУВАННЯ ДЛЯ РОЗРОБКИ СИСТЕМИ УПРАВЛІННЯ ЕЛЕКТРОПРИВОДОМ ЗМІННОГО СТРУМУ

Анотація. Розглянуті питання розробки моделей для синтезу систем керування тяговим електроприводом змінного струму дизель-поїзда та їх дослідження за допомогою пакету MATLAB. На основі аналітичного огляду існуючих алгоритмів, зокрема систем керування з використанням алгоритмів векторного управління, запропонована структурна схема регулятора контуру САР електроприводу. Для стабілізації потокозчеплення в структурі САР запропоновано використання ПД-регулятора та визначені за допомогою моделювання коефіцієнти підсилення із забезпеченням вимог до якості перехідних процесів та установленні можливі діапазони зміни коефіцієнтів підсилення. Отримані аналітичні співвідношення та математичні моделі можуть бути використані для розробки САР електроприводу дизель-поїзда і розрахунку її параметрів при заданому критерію якості.

Ключові слова: системи керування, тяговий електропривод змінного струму, дизель-поїзд, алгоритми векторного управління, ПД-регулятор, математична модель, перехідний процес, критерій якості.

Вступ

Завдання створення системи управління електроприводом змінного струму є складовою частиною загального завдання створення оптимальної системи керування транспортними засобами, в тому разі і тяговим електроприводом дизель-поїздів на базі асинхронних тягових електродвигунів (ТАД). Вирішення цих завдань пропонується виконувати з використанням відомих досягнень, як теоретичних так і практичних, застосування алгоритмів та методів керування, використання сучасних комп'ютерних технологій, в основу яких покладено методи математичного моделювання, аналізу і синтезу складних технічних систем. Тому адаптація алгоритму векторного керування для розроблення системи управління електроприводом змінного струму та моделей для її дослідження є актуальною задачею.

Постановка задачі і аналіз відомих публікацій. Ряд транспортних засобів, в тому числі електропоїзди і дизель-поїзди, в якості тягових електроприводів використовують електроприводи змінного струму з частотним способом керування. Однак, як слідує із джерел інформації, все більшої уваги приділяється розробкам систем керування електроприводами змінного струму на основі алгоритмів векторного управління [1–6]. В даних алгоритмах закладена ідея орієнтування потоку асинхронного двигуна [1, 2, 5, 6]. Напруги статора трифазного асинхронного двигуна (живлення ТАД) в системі управління визначаються по рівняннях векторного керування, які забезпечують відповідну якість функціонування системи в цілому в процесі експлуатації.

Рішення задачі розробки ефективної системи управління електроприводом змінного струму в першу чергу пов'язане зі створенням нелінійних математичних моделей з врахуванням особливостей використання того чи іншого алгоритму керування та його модифікації відповідно до умов експлуатації певного транспортного засобу, критеріїв оцінки якості функціонування системи керування, проведенням комплексних досліджень динаміки електромеханічної системи електроприводу в процесі проектування.

Для рішення завдань розробки подібних систем керування присвячено значне число публікацій [3–9], де використовується математичне моделювання, сучасні методи теорії автоматичного керування та оптимізації, комп'ютерні технології.

Метою даної роботи є адаптація алгоритму векторного керування для розробки системи управління електроприводом змінного струму та моделей для проведення досліджень з метою отримання її якісних характеристик в процесі моделювання з використанням пакету MATLAB.

Основна частина

Математична модель ТАД електроприводу в синхронній системі координат представлена рівняннями електричних кіл статора і ротора в системі координат, що обертається з швидкістю ω_s і приведена у вигляді системи диференціальних рівнянь [3, 4]:

$$\frac{d\Psi_{s1}}{dt} = U_{s1} - a_s\Psi_{s1} + a_s k_R \Psi_{r1} + \Psi_{s2}\omega_s, \quad (1)$$

$$\frac{d\Psi_{s2}}{dt} = U_{s2} - a_s\Psi_{s2} + a_s k_R \Psi_{r2} + \Psi_{s1}\omega_s, \quad (2)$$

$$\frac{d\Psi_{r1}}{dt} = -a_r\Psi_{r1} + a_r k_s \Psi_{s1} + (\omega_s - \omega)\Psi_{r2}, \quad (3)$$

$$\frac{d\Psi_{r2}}{dt} = -a_r\Psi_{r2} + a_r k_s \Psi_{s2} - (\omega_s - \omega)\Psi_{r1}, \quad (4)$$

$$M = \frac{3}{2} p \frac{k_r}{\sigma L_S} (\Psi_{r1}\Psi_{s2} - \Psi_{s1}\Psi_{r2}), \quad (5)$$

$$\frac{d\omega}{dt} = \frac{p}{J} (M - M_c), \quad (6)$$

де ω_s – кутова частота напруги живлення статора ТАД при $p = 1$; ω – кутова швидкість ротора; Ψ_{s1} , Ψ_{s2} , Ψ_{r1} , Ψ_{r2} , U_{s1} , U_{s2} – відповідно проекції векторів потоків статора і ротора та напруги живлення на осі координат; k_s і k_r – коефіцієнти електромагнітного зв'язку відповідно статора та ротора;

$$L_s = L_m + L_{\sigma s}, \quad L_r = L_m + L_{\sigma r}, \quad a_r = \frac{L_m}{L_s}, \quad k_s = \frac{L_m}{L_s},$$

$$k_r = \frac{L_m}{L_r}, \quad \sigma = 1 - \frac{L_m^2}{L_s L_r}, \quad T_s = \frac{L_s}{R_s}, \quad T_r = \frac{L_r}{R_r}, \quad a_s = \frac{1}{\sigma T_s},$$

$$a_r = \frac{1}{\sigma T_r}; \quad L_r, L_m, L_s, R_s, R_r - \text{параметри схеми}$$

заміщення ТАД; T_s і T_r – відповідно електромагнітна постійна часу статора та ротора; σ – повний коефіцієнт розсіювання; p – постійний коефіцієнт ТАД (число пар полюсів); J – момент інерції.

Чисельні значення параметрів схеми заміщення ТАД:

$$L_s = 0,093331 \text{ (мГн)}; \quad L_r = 0,092819 \text{ (мГн)};$$

$$R_s = 0,0833 \text{ (Ом)}; \quad R_r = 0,068 \text{ (Ом)};$$

$$k_r = 0,98816; \quad k_s = 0,98274; \quad T_s = 1,123117 \text{ (с)};$$

$$T_r = 1,3730621 \text{ (с)}; \quad \sigma = 0,028896;$$

$$a_s = 30,813233; \quad a_r = 25,204152.$$

У випадку синхронного обертання осей $\omega_s = \omega_k$ і початкової фази $\phi_K = 0$, функції задання напруги U_{s1} , U_{s2} – постійне значення напруги живлення величиною U_M по осі s_1 і нульове по осі s_2 .

Оскільки в процесі проектування систем керування для аналізу протікання перехідних процесів, а також можливих датчиків вимірювання та фіксації результатів, в якості змінних використовуються струми в обмотках ТАД, то систему рівнянь (1) – (6) можна записати відносно проекцій векторів струмів, використовуючи відомі співвідношення:

$$I_{s1} = (a_s \Psi_{s1} - a_s k_R \Psi_{r1}) / R_s;$$

$$I_{s2} = (a_s \Psi_{s2} - a_s k_R \Psi_{r2}) / R_s;$$

$$I_{r1} = (a_r \Psi_{r1} - a_r k_s \Psi_{s1}) / R_r;$$

$$I_{r2} = (a_r \Psi_{r2} - a_r k_s \Psi_{s2}) / R_r.$$

В результаті перетворення систему рівнянь (1) – (6) в синхронній системі координат представимо в наступному вигляді:

$$\frac{d\Psi_{s1}}{dt} = \Psi_{s2} \omega_k - R_s I_{s1} + U_{s1}; \quad (7)$$

$$\frac{d\Psi_{s2}}{dt} = \Psi_{s1} \omega_k - R_s I_{s2} + U_{s2}; \quad (8)$$

$$\frac{d\Psi_{r1}}{dt} = (\omega_k - p\omega) \Psi_{r2} - R_r I_{r1}; \quad (9)$$

$$\frac{d\Psi_{r2}}{dt} = -(\omega_k - p\omega) \Psi_{r1} - R_r I_{r2}, \quad (10)$$

де 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 електромагнітного моменту:

$$M = \frac{3}{2} p \frac{k_R}{\sigma L_s} (\Psi_{r1} \Psi_{s2} - \Psi_{s1} \Psi_{r2}). \quad (11)$$

Алгоритм векторного керування полягає в наступному [3, 5].

Якщо допустити, що швидкість обертання координат співпадає із швидкістю вектору потоку ротора

$$\omega_k = \omega_\Psi,$$

то вектор потоку Ψ_r буде на осі (9) своїм модулем Ψ_{rm} , а його проекція на вісь (10) дорівнюватиме нулю, тобто вектор потокозчеплення ротора буде мати тільки дійсну складову.

Рівняння за умови, що $\omega_k = \omega_\Psi$, приймуть наступний вигляд:

$$\frac{d\Psi_{s1}}{dt} = \Psi_{s2} \omega_\Psi - R_s I_{s1} + U_{s1}; \quad (12)$$

$$\frac{d\Psi_{s2}}{dt} = \Psi_{s1} \omega_\Psi - R_s I_{s2} + U_{s2}; \quad (13)$$

$$\frac{d\Psi_{rm}}{dt} = -R_r I_{r1}; \quad (14)$$

$$(\omega_\Psi - p\omega) \Psi_{rm} + R_r I_{r2} = 0. \quad (15)$$

Виразимо проекції струмів ротора I_{r1} та I_{r2} через проекції струмів статора I_{s1} та I_{s2} .

Для цього використаємо рівняння зв'язку:

$$\Psi_r = I_r L_r + I_s L_m; \quad (16)$$

$$\Psi_s = I_s L_s + I_r L_m. \quad (17)$$

Із рівняння (16) отримаємо:

$$I_r = \frac{1}{L_r} \Psi_r - I_s \frac{L_m}{L_r}.$$

Із рівняння (17) отримаємо:

$$\begin{aligned} \Psi_s &= I_s L_s + I_r L_m = I_s L_s + \frac{L_m}{L_r} \Psi_r - I_s \frac{L_m^2}{L_r} = \\ &= I_s \left(L_s - \frac{L_m^2}{L_r} \right) + \frac{L_m}{L_r} \Psi_r. \end{aligned}$$

Тоді:

$$\Psi_s = I_s L_s' + K_r \Psi_r, \quad (18)$$

де $L_s' = L_s - \frac{L_m^2}{L_r}$; $K_r = \frac{L_m}{L_r}$.

Виконаємо перетворення системи рівнянь (12) – (15).

$$\frac{d\Psi_{s1}}{dt} = \frac{dI_{s1}}{dt} L_s' + K_r \frac{d\Psi_{rm}}{dt};$$

$$\frac{dI_{s1}}{dt} L_s' + K_r \frac{d\Psi_{rm}}{dt} = \Psi_{s2} \omega_\Psi - R_s I_{s1} + U_{s1};$$

$$\Psi_{s2} \omega_\Psi = I_{s2} L_s' \omega_\Psi + K_r \Psi_{r2} \omega_\Psi.$$

Із врахуванням того, що $\Psi_{r2} = 0$,

$$\Psi_{s2}\omega_{\Psi} = I_{s2}L_s' \omega_{\Psi}.$$

З урахуванням перетворень рівняння (12) прийме вигляд:

$$\frac{dI_{s1}}{dt} L_s' + K_r \frac{d\Psi_{rm}}{dt} - I_{s2}L_s' \omega_{\Psi} + R_s I_{s1} + U_{s1} = 0. \quad (19)$$

Перетворення рівняння (13).

$$\begin{aligned} \frac{dI_{s2}}{dt} L_s' + \frac{d\Psi_{r2}}{dt} K_r &= \\ = \omega_{\Psi} (I_{s1}L_s' + K_r \Psi_{rm}) - R_s I_{s2} + U_{s2}. \end{aligned}$$

Оскільки $\Psi_{r2} = 0$, то рівняння (13) набуде вигляду:

$$\begin{aligned} \frac{dI_{s2}}{dt} L_s' &= I_{s1}L_s' \omega_{\Psi} + \\ + K_r \Psi_{rm} \omega_{\Psi} - R_s I_{s2} + U_{s2}. \end{aligned} \quad (20)$$

Перетворення рівняння (14).

$$\frac{d\Psi_{rm}}{dt} = -R_r I_{r1}.$$

Виконаємо операцію знаходження струму I_{r1} , в залежності від струму I_{s1} .

Для цього скористаємося рівнянням зв'язку:

$$I_{r1} = \frac{1}{L_r} \Psi_{r1} - I_{s1} \frac{L_m}{L_r}.$$

В результаті рівняння (14) набуде вигляду:

$$\frac{d\Psi_{rm}}{dt} + \frac{R_r}{L_r} \Psi_{rm} - I_{s1} K_r \frac{L_m}{L_r} = 0; \quad (21)$$

За умови завдання модуля потокозчеплення $\Psi_{rm} = const$ із рівняння (21) знайдемо струм I_{s1} :

$$I_{s1} = \frac{\Psi_{rm}}{L_m}. \quad (22)$$

Перетворення рівняння (15).

$$(\omega_{\Psi} - p\omega) \Psi_{rm} + R_r I_{r2} = 0.$$

Відповідно до рівняння зв'язку

$$I_{r2} = \frac{1}{L_r} \Psi_{r2} - I_{s2} \frac{L_m}{L_r}.$$

Враховуючи, що $\Psi_{r2} = 0$, рівняння (15) набуде вигляду:

$$\beta \Psi_{rm} - K_r R_r I_{s2} = 0, \quad (23)$$

де $\beta = \omega_{\Psi} - p\omega$.

В результаті струм I_{s2} можна визначити як:

$$I_{s2} = \frac{\beta \Psi_{rm}}{K_r R_r}. \quad (24)$$

Рівняння для електромагнітного моменту знаходиться за умови, що швидкість обертання координат співпадає із швидкістю вектору потоку ротора, тобто

$$\omega_k = \omega_{\Psi}.$$

Відповідно до рівняння (11) знаходження електромагнітного моменту ТАД визначається як:

$$M = \frac{3}{2} p \frac{K_r}{\sigma L_s} (\Psi_{r1} \Psi_{s2} - \Psi_{s1} \Psi_{r2}).$$

Враховуючи, що $\Psi_{r2} = 0$, і з урахуванням рівняння зв'язку

$$\Psi_{s2} = I_{s2}L_s' + K_r \Psi_{r2},$$

отримаємо:

$$M_d = \frac{mpK_r}{2} \Psi_{rm} I_{s2}. \quad (25)$$

В результаті перетворень отримали систему рівнянь, що дозволяє формування електромагнітного моменту ТАД, який є визначальним в системі керування для приведення дизель-поїзда до його руху.

Це сукупність рівнянь (19) – (25), яка в кінцевому разі визначає електромагнітний момент через модуль потокозчеплення і величину сковзання β , що залежать від кутової частоти напруги живлення обмотки статора і кутової швидкості ротора (за певних умов швидкості руху дизель-поїзда).

Крім цього струми I_{s1} і I_{s2} є визначальними для взаємного зв'язку з фазними напругами U_{s1} і U_{s2} . Таким чином отримали математичну модель електроприводу, яка дозволяє виділити управління – це значення амплітуди і частоти напруги живлення ТАД. Частота напруги живлення визначається на основі рівняння (23).

Функції впливу (управління) U_{s1} і U_{s2} знаходяться з рівнянь (19) і (20). За умови стабілізації току I_{s2} функції впливу будуть знаходитись відповідно до рівнянь:

$$U_{s1} = -I_{s2}L_s' \omega_{\Psi} + R_s I_{s1}; \quad (26)$$

$$U_{s2} = -\omega_{\Psi} (I_{s1}L_s' + K_r \Psi_{rm}) + R_s I_{s2}. \quad (27)$$

За умови стабілізації потокозчеплення Ψ_{rm} ($\Psi_{rm} = const$) функції впливу визначаються як:

$$U_{s1} = -I_{s2}L_s' \omega_{\Psi} + R_s \frac{\Psi_{rm}}{L_m}; \quad (28)$$

$$U_{s2} = -\omega_{\Psi} \Psi_{rm} \left(\frac{L_s'}{L_m} + K_r \right) + R_s \frac{\beta \Psi_{rm}}{K_r R_r}. \quad (29)$$

Швидкість вектору потоку ротора ω_{Ψ} визначається як $\omega_{\Psi} = \omega_s$.

Таке допущення цілком можливе, постільки при знаходженні оптимального закону керування, а це функції впливу U_{s1} і U_{s2} , формуємо їх відповідно

до заданого критерія якості (функціоналу) з подальшим уточненням за допомогою моделювання.

Рівняння (28) і (29) при заданому значенні поточкозчеплення Ψ_{rm} і рівняння (26), що визначає темп розгону (величина β) дизель – поїзда за умови його завантаження та рівняння, які описують математичну модель його руху (наприклад, наведені в [3, 4]) можуть бути використані при розробці оптимальної системи керування згідно заданого критерія якості.

Для практичної реалізації даного алгоритму необхідно забезпечити стабілізацію поточкозчеплення. Це можливо здійснити шляхом введення в канал

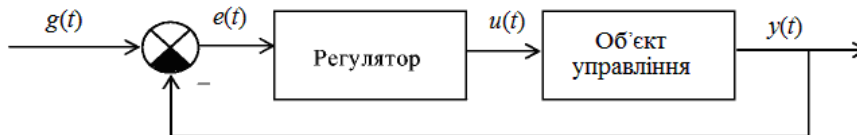


Рис. 1. Управління з негативним зворотним зв'язком

Призначення регулятора системи полягає в корекції динамічних властивостей об'єкта управління за допомогою керуючого сигналу $u(t)$ таким чином, щоб реальний вихідний сигнал $y(t)$ якнайменше відрізнявся від бажаного вихідного сигналу $g(t)$. Регулятор виробляє управління, використовуючи помилку регулювання $e(t) = g(t) - y(t)$.

Для оцінки динамічних властивостей системи використовувалась її реакція на одиничний вхідний вплив згідно заданим показникам якості: величина статичної помилки $\varepsilon(t)$, час перехідного процесу (t_p), величина перерегулювання $\Delta\%$ й коливальність (N).

В якості регулятора використано традиційну схему ПІД – регулятора (ПІД – пропорційно-інтегродиференціальний), що широко застосовується в системах автоматичного керування. Основне рівняння ПІД-регулятора має вигляд:

$$u(t) = K(k_p e(t) + k_i \int_{\tau=0}^t e(\tau) d\tau + k_d \frac{de(t)}{dt}), \quad (30)$$

де K , k_p , k_i , k_d – константи, обираєні в процесі проектування (в нашому разі в процесі досліджень за допомогою моделювання).

Пропорційна складова виробляє вихідний сигнал, що протидіє відхиленню регульованої величини від заданого значення, яке спостерігається в поточний момент часу. Вона тим більша, чим більше це відхилення.

Однак, при використанні тільки пропорційного регулятора значення регульованої величини ніколи не стабілізується. Існує так звана статична помилка, що дорівнює такому відхиленню регульованої величини, що забезпечує вихідний сигнал, який стабілізує вихідну величину саме на цьому значенні. Чим більший коефіцієнт пропорційності між вхідним і вихідним сигналом ($K \cdot k_p$ – коефіцієнт підсилення), тим

формування поточкозчеплення відповідного регулятора, який забезпечує якість перехідних процесів.

В результаті схему управління стабілізації поточкозчеплення, а відповідно і струмів I_{s1} та I_{s2} , згідно алгоритму векторного керування для підтримки постійного значення тягового моменту ТАД в процесі розгону дизель-поїзда можливо представити у вигляді класичної схеми управління з одиничним негативним зворотним зв'язком, яка приведена на рис. 1.

Модель об'єкта управління представлена передавальною функцією ТАД на основі Т-образної схеми заміщення та передавальними функціями системи зворотного зв'язку (датчиків сигналів).

менша статична помилка, однак при занадто великому коефіцієнті підсилення за наявності затримок (запізнювання) у системі можуть початися автоколивання, а при подальшому збільшенні коефіцієнта система може втратити стійкість.

Інтегровальна складова пропорційна інтегралу за часом від відхилення регульованої величини. Її використовують для усунення статичної помилки.

Інтегровальна складова також може призводити до автоколивань при неправильному виборі її коефіцієнта (K , k_i).

Складова, що диференціює, пропорційна темпу зміни відхилення регульованої величини й призначена для протидії відхиленням від цільового значення, яке прогнозується в майбутньому. Відхилення можуть бути викликані зовнішніми збурюваннями або запізнюванням впливу регулятора на систему.

Відповідно до Т-образної схеми заміщення і за умови, що $L_{\sigma s} \ll L_m$ передавальна функція ТАД може бути представлена у вигляді:

$$W(p) = \frac{1}{(a_1 p T_1 + 1)(p T_2 + 1)}, \quad (31)$$

де $a_1 = \frac{L_s}{L_m}$, $T_1 = \frac{L_m}{r_1}$, $T_2 = \frac{L_{\sigma r} S}{r_2}$, S – величина ско-
взання.

За умови реальних значень електричних параметрів схеми заміщення ТАД

$$L_{\sigma s} = 1.6 \cdot 10^{-3} \text{ (мГн); } L_{\sigma r} = 1.099 \cdot 10^{-3} \text{ (мГн);}$$

$$L_m = 92 \cdot 10^{-3} \text{ (мГн); } r_1 = 0.0831 \text{ (Ом);}$$

$$r_2 = 0.0676 \text{ (Ом)}$$

передавальною функцією об'єкта управління має вид:

$$W(p) = \frac{1}{(1.1p + 1)(0.016p + 1)}. \quad (32)$$

Структурна схема моделі в середовищі MATLAB для визначення параметрів ПІД – регулятора (коефіцієнтів підсилення) наведена на рис. 2.

В ній можливо виділити чотири складові:

- блок формування сигналу завдання,
- безпосередньо ПІД-регулятор,
- об’єкт керування,
- блок формування сигналу зворотного зв’язку.

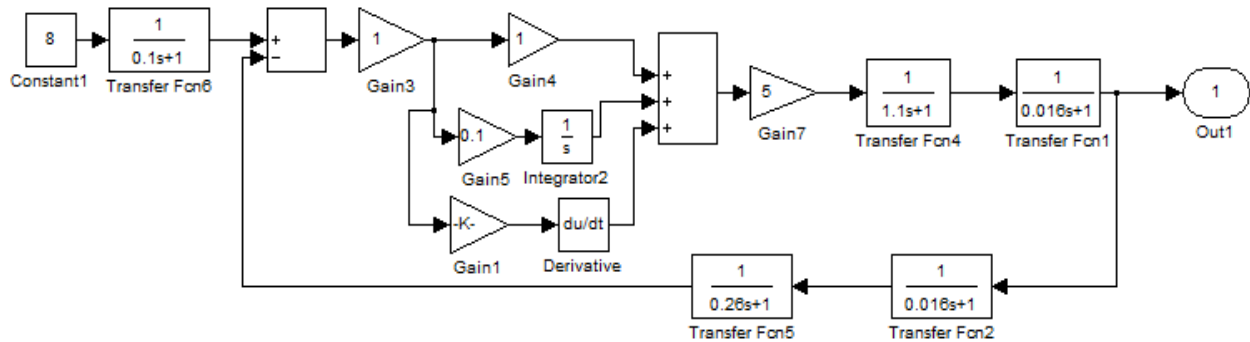


Рис. 2. Структурна схема моделі

На рис. 3–6 наведені графіки перехідних процесів в процесі дослідження моделі з метою визначення параметрів (коефіцієнтів підсилення) K , k_p , k_i , k_d ПІД-регулятора, які забезпечують необхідні значення показників якості.

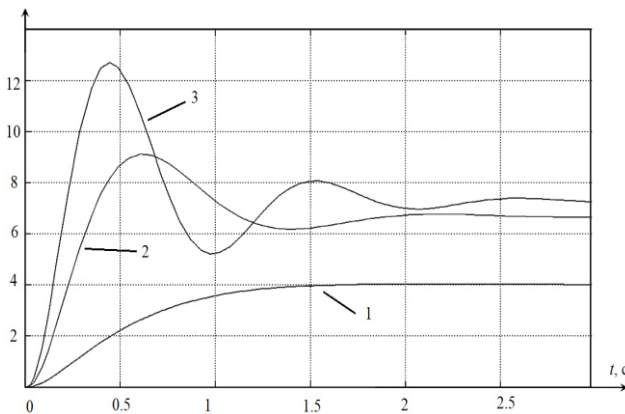


Рис. 3. Графіки перехідних процесів (пропорційна складова) при коефіцієнтах підсилення K : $K = 1$ (1); $K = 5$ (2); $K = 10$ (3)

В даному разі (при використанні тільки пропорційної складової) статичні помилки $\varepsilon(t)$ в кінцевому терміні часу процесу регулювання t_p відповідно становлять величину: 25%, 16.25%, 8.4%.

Крім того, що при загальному коефіцієнті підсилення $K=10$ і $k_p=1$ статична помилка становить величину $\varepsilon(t)=8.4\%$, а також присутні автоколивання. Це вказує на необхідність використання і інших складових ПІД-регулятора.

На рис. 4 наведені графіки перехідних процесів при різних значеннях коефіцієнтів інтегральної складової k_i регулятора за умови постійних значень пропорційної і диференційної складових:

$$K=5; k_p=1; k_d=0.05.$$

При цьому статичні помилки $\varepsilon(t)$ за час регулювання t_p відповідно становлять величину: 8.75%, 3.75%, 0.05%.

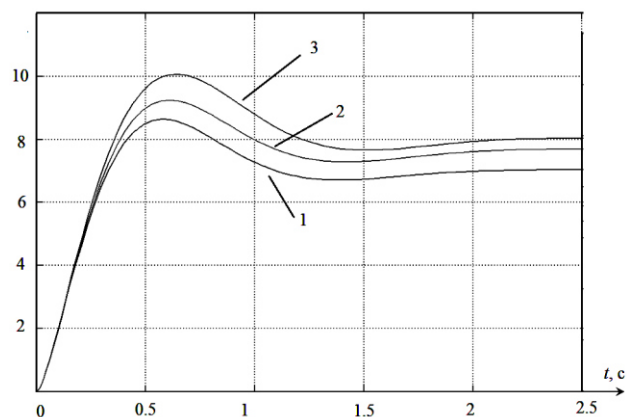


Рис. 4. Графіки перехідних процесів $k_i=0.1$ (1); $k_i=0.4$ (2); $k_i=0.8$ (3)

Час регулювання, що визначає швидкодію системи, становить $t_p = 2$ с. При даних значеннях коефіцієнтів ПІД-регулятора величина пере регулювання Δ відповідно становить: 7.80%, 15.25%, 25,0%. На підставі значень статичних помилок і величин пере регулювання при відповідних значеннях коефіцієнтів найбільш оптимальними є значення коефіцієнтів: $K=5$; $k_d=0.05$; $k_p = 1$; $k_i = 0.4$.

Однак величина перерегулювання Δ не задовольняє вимогам щодо якості перехідного процесу. Та і значення статичні помилки $\varepsilon(t)$ не близьке до нульового.

Покращення цих показників можливо за рахунок каналу диференційної складової – значення коефіцієнта k_d .

На рис. 5 наведені графіки перехідних процесів при різних значеннях коефіцієнтів диференційної складової k_d ($k_d=0.05$; 0.1; 0.15) регулятора за умови постійних значень пропорційної і інтегральної складових ($K=5$; $k_p=1$; $k_i=0.4$).

При цьому статичні помилки $\varepsilon(t)$ за час регулювання t_p відповідно становлять величину: 3.75%, 3.70%, 3.65%.

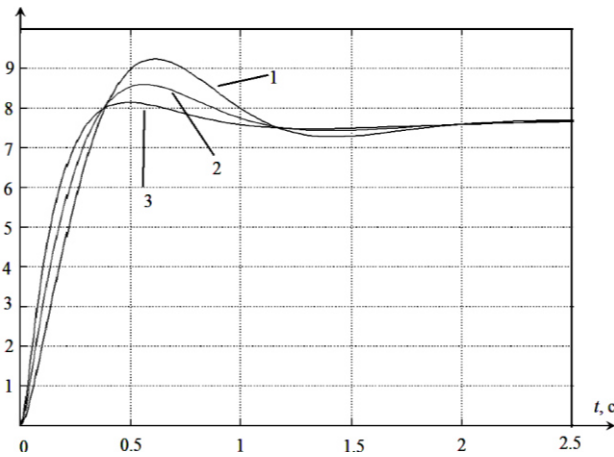


Рис. 5. Графіки перехідних процесів
 $k_d=0.15$ (1); $k_d=0.1$ (2) $k_d=0.05$ (3)

Час регулювання t_p , що визначає швидкодію системи, становить відповідно (2; 1.8; 1.6) с. При даних значеннях коефіцієнтів ПІД-регулятора величина перерегулювання Δ відповідно становить: 15.5%; 7.5%; 1.8%. Звідси слідує, що значення коефіцієнта диференційної складової рівному $k_d=0.15$, покращує показники якості – як значення статичної помилки, так і величину перерегулювання.

На рис. 6 наведені графіки перехідних процесів при різних значеннях коефіцієнтів диференційної складової k_d ($k_d=0.05$; 0.1; 0.15) регулятора за умови постійних значень пропорційної і інтегральної складових ($k_p=1$; $k_i=0.4$) та загального коефіцієнта підсилення ПІД-регулятора K (при ($K=5$; $K=6$)).

Графіки перехідних процесів демонструють вплив на якість показників системи регулювання за рахунок загального коефіцієнта підсилення K для всіх каналів ПІД-регулятора. Такий підхід побудови системи керування спрощує процес її налагодження в процесі проектування.

Перехідні процеси, що наведені на графіках рис.6, відповідають значенням коефіцієнтів: $K=5$ ($k_d=0.05$; 0.1); $K=6$ ($k_d=0.15$). При цьому статичні помилки $\varepsilon(t)$ за час регулювання t_p відповідно становлять величину: 3.75%, 3.70%, 3.70%.

Таблиця 1 – Узагальнені результати досліджень

Номер експерименту	K	k_p	k_i	k_d	$\varepsilon, \%$	$\Delta, \%$	$t_p, \text{с}$	Рисунок
1	1	1	0	0	25.0	-	2.0	3
	5	0.9	0	0	16.25	-	2.0	
	10	0.8	0	0	8.4	2.0	2.0	
2	5	1	0.1	0.05	8.75	7.8	2.0	4
	5	1	0.4	0.05	3.75	15.25	2.0	
	5	1	0.8	0.05	0.05	25.0	2.0	
3	5	1	0.4	0.05	3.75	15.50	2.0	5
	5	1	0.4	0.1	3.70	7.50	1.8	
	5	1	0.4	0.15	3.65	1.80	1.6	
4	5	1	0.4	0.05	3.75	15.50	2.0	6
	5	1	0.4	0.1	3.70	7.50	1.8	
	6	1	0.4	0.15	3.70	8.25	1.2	

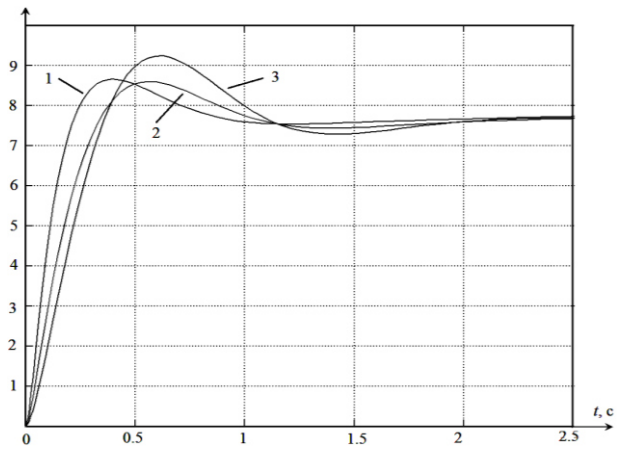


Рис. 6. Графіки перехідних процесів ($K=6$)
 $k_d=0.15$ (1); ($K=5$) $k_d=0.1$ (2); $k_d=0.05$ (3)

Час регулювання t_p , що визначає швидкодію системи, становить відповідно (2; 1.8; 1.2) с.

При даних значеннях коефіцієнтів ПІД-регулятора величина перерегулювання Δ відповідно становить: 15.5%; 7.5%; 8.25%.

Звідси слідує, що збільшення загального коефіцієнта підсилення K при значенні коефіцієнта диференційної складової рівному $k_d=0.15$, покращує показники якості – збільшення швидкодії системи керування при забезпеченні задовільних показників якості.

Згідно з результатами досліджень для забезпечення необхідної якості протікання перехідних процесів коефіцієнти підсилення каналів ПІД-регулятора мають наступні значення: $K=6$; $k_p=1$; $k_i=0.4$; $k_d=0.15$. При цьому статична помилка $\varepsilon=3.7\%$, величина перерегулювання $\Delta=8.25\%$, час регулювання $t_p=1.2$ с.

Крім того результати досліджень дають можливість установлення можливих діапазонів зміни коефіцієнтів підсилення: $4 \leq K \leq 6$; $0.8 \leq k_p \leq 1.2$; $0.3 \leq k_i \leq 0.4$; $0.1 \leq k_d \leq 0.15$.

В табл. 1 наведені узагальнені результати досліджень.

Висновки

1. На основі аналітичного огляду існуючих алгоритмів та систем автоматичного керування електроприводу з тяговим електродвигуном дизель-поїзда запропонована структурна схема регулятора контуру САР електропередачі для стабілізації поточкозчеплення та математична модель для дослідження системи управління з метою визначення параметрів регулятора за умови забезпечення певних критеріїв якості.

2. З метою можливості реалізації алгоритму векторного керування в САР електроприводу дизель-поїзда для цілей оптимізації проведена адаптація алгоритму в плані розроблення математичної моделі системи управління у вигляді передавальної функції на основі Т- схеми заміщення тягового електродвигуна.

3. Для стабілізації поточкозчеплення в структурі САР електропривода запропоновано використання ПІД – регулятора та визначені за допомогою моделювання параметри (коефіцієнтів підсилення) його каналів із забезпеченням вимог до якості перехідних процесів. Крім того, в процесі досліджень установлені можливі діапазони зміни коефіцієнтів підсилення, що важливо при технічній реалізації системи управління в цілому.

4. Отримані моделі алгоритмів управління (регулятора поточкозчеплення для формування напруги живлення інвертора ТАД можуть бути використані при побудові системи управління за допомогою засобів мікропроцесорної техніки, що дозволить створення цифрової системи управління електропередачі дизель-поїзда згідно сучасних світових тенденцій.

5. Запропоновані математичні моделі, які реалізовані у вигляді машинної моделі і проведені за її допомогою дослідження, розроблені алгоритми САР показали перспективність такого підходу при проектуванні сучасних систем управління електроприводами локомотивів та дизель-поїздів з тяговими двигунами змінного струму.

6. На підставі досліджень, які проведені шляхом моделювання показано, що структури регуляторів із значеннями параметрів, що пропонуються, задовольняють вимогам щодо працездатності в робочому проміжку швидкостей та можливих навантажень.

7. Запропоновані алгоритми у виді математичних моделей можуть бути покладені в основу розроблення системи керування в цілому рухом дизель-поїзда, де одним з основних принципів керування – підтримка заданої величини тягового моменту.

СПИСОК ЛІТЕРАТУРИ

1. Blaschke F. Das Prinzip der Feldorientierung. Die Grundlage für die Transvektor-Reglung von Drehfeldmaschine / F. Blaschke // Siemens Zeitschrift. – 1971. – № 45. – Н. 10.
2. Bose Bimal K. Modern Power Electronics and AC drives / Bimal K. Bose. – Prentice Hall PTR: Prentice-Hall Inc. – 2002. – 712 p.
3. М.І. Заполовський, М.В. Мезенцев. Оптимізаційна модель для синтезу системи керування електроприводу змінного струму на основі використання алгоритму векторного управління // Системи управління, навігації та зв'язку. – №4. Полтава. – 2022.
4. М.І. Заполовський, М.В. Мезенцев, О.І. Баленко, М.В. Оліфір Оптимізаційна модель тягового асинхронного електроприводу дизель-поїзда та її дослідження // Системи управління, навігації та зв'язку. – №3. Полтава. – 2023. С. 45-53.
5. O.K. Ablesimov, E.E. Alexandrov, and I.E. Alexandrova, Automatic control of moving objects and technological processes. Kharkov: NTU "KhPI", – 2008. – 443 p. (in Ukrainian).
6. Volkov A.V. Asynchronous motor drive based on self-excited current inverter with switched-off thyristors and provided with redicting relay and vector regulation of stator current / A.V. Volkov, I.A. Kosenko // Published in Elektrotehnika. – 2008. – No 10. – P. 6-17.
7. Bose, Bimal K. (2006). Power Electronics and Motor Drives: Advances and Trends. Amsterdam: Academic. p. 22.
8. Rafiq, Md Abdur (2006). "Fast Speed Response Field-Orientation Control of Induction Motor Drive with Adaptive Neural Integator". Istanbul University Journal of Electrical & Electronics Engineering. 6 (2): P. 229.
9. Dong, Gan (Dec 2007). "Sensorless and Efficiency Optimized Induction Machine Control with Associated Converter PWM Modulation Schemes". Tennessee Technological University. P. 10.

Received (Надійшла) 18.05.2024

Accepted for publication (Прийнята до друку) 17.07.2024

Adaptation of the vector control algorithm for the development of an AC drive control system

M. Zapolovskyi, M. Mezentsev, M. Olifir

Abstract. The issues of developing models for the synthesis of AC traction electric drive control systems of a diesel train and their research using the MATLAB package are considered. On the basis of an analytical review of existing algorithms, in particular, control systems using vector control algorithms, a structural diagram of the controller of the Automated Control System (ACS) circuit of the electric drive is proposed. To stabilize the flow coupling in the ACS structure, the use of a PID controller is proposed, and the gain coefficients are determined by modeling to ensure the requirements for the quality of transient processes, and the possible ranges of gain coefficient changes are established. The obtained analytical relations and mathematical models can be used for the development of the ACS of the electric drive of a diesel train and the calculation of its parameters for a given quality criterion.

Keywords: control system, alternating current traction electric drive, diesel train, vector control algorithms, PID controller, mathematical model, transition process, quality criteria.

П. О. Качанов, М. В. Тарасенко

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

СИСТЕМА АВТОМАТИЧНОГО РЕГУЛЮВАННЯ РІВНЯ ВОДИ В БАРАБАНИ КИТЛА ЕНЕРГОУСТАНОВКИ

Вступ. Розглянуто систему автоматичного регулювання рівня води в барабані парового котла. В контексті сучасних вимог до енергетичних установок, важливість точного та стабільного контролю рівня води в барабані стає ключовою для забезпечення безпеки та ефективності їх експлуатації. **Метою** статті є розробка та оцінка ефективної системи регулювання, яка усуває коливання рівня води в барабані, викликані змінами витрати живильної води або перегрітої пари. Проведено аналіз існуючих технічних рішень реалізації систем регулювання рівня в барабані. Запропоновано модель трьохімпульсної САР, яка набула найбільшого поширення в даному напрямку через свої характеристики. У **результаті** аналізу та моделювання системи в середовищі Simulink було підтверджено, що розроблена система відповідає всім поставленим вимогам та дозволяє прогнозувати її функціонування в реальних умовах. На основі проведених досліджень можна зробити висновок про значний потенціал запропонованої системи для підвищення надійності та ефективності енергетичних установок.

Ключові слова: система автоматичного регулювання, барабанний паровий котел, рівень живильної води.

Вступ

Ефективна експлуатація енергоустановок безпосередньо залежить від точної підтримки технологічних параметрів. Це можливо завдяки якісній роботі автоматичних регуляторів, які застосовуються в системах автоматичного регулювання (САР). Одним із ключових моментів при проектуванні таких систем є вибір та налаштування автоматичних регуляторів.

Регулювання рівня води в барабані є одним з ключових контурів регулювання системи керування енергетичними котлоагрегатами. Високі вимоги до якості регулювання обумовлені кількома чинниками. Найсуттєвішим із них є те, що навіть незначні відхилення рівня можуть активувати захисні механізми, що призводять до аварійного відключення котла. Тому регулятор потребує точного налаштування для утримання рівня в невеликому діапазоні допустимих значень. Важливість точного та стабільного регулювання рівня в барабані незмінно підкреслюється у контексті забезпечення безпечної та ефективної роботи котельних установок.

Ще одним важливим моментом є відсутність самовирівнювання каналів регулювання та обурення, що є причиною відсутності рівноваги в системі. Це призводить до постійних коливань рівня барабана під впливом зовнішніх факторів. Оптимальна структура та налаштування системи регулювання допомагають зменшити ці коливання, що продовжує термін служби ключових елементів котла.

Таким чином, ключовим напрямком стає не лише вибір та налаштування автоматичних регуляторів, але й створення надійних математичних моделей, що сприяє більш точному та стабільному функціонуванню систем автоматичного регулювання в енергетичних установках.

Огляд існуючих рішень

На сьогоднішній день актуальною темою є використання нейронних мереж у системах автоматичного регулювання.

У статті [1] описується використання штучних нейронних мереж (ANN) для моделювання термодинамічних процесів вугільних теплових електростанцій. Зокрема, вона порівнює ефективність двох видів ANN - Back Propagation Neural Network (BPNN) та Radial Basis Neural Network (RBNN) - у прогнозуванні показників ефективності та тепловіддачі котлів. Стаття підкреслює значення використання цих моделей для аналізу впливу різних параметрів на ефективність та визначення напрямку змін, які можуть покращити роботу станції.

Автори у [2] досліджують використання різних нейронних мереж, таких як Multilayer Perceptron (MLP), Nonlinear Autoregressive Exogenous (NARX) і Long Short Term (LSTM), для створення моделі. Вони роблять висновок, що LSTM має найкращу продуктивність. Також пропонується нова архітектура, яка інтегрує модель процесу у наявний контролер з використанням адаптивних ваг для підвищення ефективності контролю.

У роботі [3] запропоновано метод, що використовує сигнал парового потоку для контролю, забезпечуючи компенсацію впливу змін навантаження та уникнення феномену "фальшивого рівня води". Комбінація нейронних мереж із робастним контролером забезпечує стабільність і ефективність регулювання навіть при зміні умов експлуатації.

Дослідники у [4] використовують нейронні мережі для оцінки параметрів контролера та забезпечення оптимального налаштування зворотнього зв'язку в системах автоматичного керування. Вона порівнює підхід з нейронними мережами із традиційними методами, такими як метод Циглера-Нікольса, та демонструє переваги швидкості, ефективності та безпеки, які може забезпечити використання нейронних мереж у процесі налаштування контролерів.

Нейронні мережі передбачають рівень води в барабані на основі поточних та попередніх вимірювань рівня води, витрати пари та витрати подачі води. Запропонована ідея може підвищити точність САР, оскільки нейронні мережі можуть вловлювати тонкі

закономірності в них, що дозволяє їм передбачати майбутню поведінку системи з високою точністю.

Однак, слід зауважити, що внутрішні процеси у нейронних мережах які приводять до висновків, не завжди легко інтерпретувати. Це створює проблеми при спробі зрозуміти, як нейронна мережа приймає рішення, що в свою чергу ускладнює процес обслуговування системи та підвищує вимоги до кваліфікації персоналу.

У статті [5] розглянуто налаштування різних регуляторів, включаючи PID, PD-PI, PI-PD та 2DOF регулятори для контролю рівня води в барабані котла. Чотири контролери були налаштовані за допомогою інструментів оптимізації MATLAB з індексами продуктивності ITAE/ISTSE з метою забезпечення стабільної системи керування та хороших динамічних характеристик. В цілому кожен регулятор проявив різні сильні та слабкі сторони. Виходячи з наведених даних, вибір "найкращого" регулятора залежить від конкретних критеріїв продуктивності та компромісів між перерегулюванням, часом встановлення та відкиданням дії.

У роботі [6] проведено аналіз схем регулювання рівня в барабані парового котла. Було описано принцип роботи одноімпульсної, двоімпульсної та трьохімпульсної схем, а також схеми з диференціатором. Однак недоліком трьохімпульсної схеми наведено те, що при зникненні будь-якого з трьох сигналів схема припиняє роботу. На практиці ж для таких важливих сигналів завжди використовують принцип "2оо3", коли для кожного параметра встановлюють по 3 давачі і при відмові одного з давачів система продовжує роботу спираючись на два, що залишилися.

Метою даної статті є розробка системи автоматичного регулювання рівня води у барабані котлоагрегата. Автоматичний регулятор повинен забезпечувати сталість середнього рівня незалежно від навантаження котла та інших збуджуючих впливів.

Властивості барабанного котла як об'єкта регулювання рівня

Регулювання живлення котельних агрегатів зводиться до підтримки матеріального балансу між відведенням пари та подачі води. Параметром, що характеризує матеріальний баланс є рівень води в барабані котла.

Надійність роботи котельного агрегату багато в чому визначається якістю регулювання рівня. Зниження рівня нижче допустимих меж може призвести до порушення циркуляції в екранних трубах, в результаті нього відбудеться підвищення температури стінок труб, що обігриваються, і їх перепал. Значне підвищення рівня води в барабані також може призвести до аварійних наслідків, оскільки при підвищеному рівні можливе закидання води в пароперегрівач і турбіну, що викличе занесення пароперегрівача солями або поломку турбіни.

Зі сказаного ясно, що навіть короткочасне зниження або підвищення рівня понад задані межі неприпустимо. У зв'язку з цим до точності підтримки рівня висуваються дуже високі вимоги.

Якість регулювання живлення визначається як точністю підтримки рівня на заданому значенні, а й рівномірністю подачі поживної води. Необхідно забезпечити рівномірне живлення котла водою, оскільки часті та глибокі зміни витрати поживної води можуть спричинити значні температурні зміни у металі економайзера.

Барабаним котлам з природною циркуляцією властива значна акумулююча здатність, яка проявляється у перехідних режимах. Якщо в стаціонарному режимі положення рівня води в барабані котла визначається станом матеріального балансу, то в перехідних режимах положення рівня впливає велика кількість збурень. Основними з них є:

- зміна витрат поживної води;
- зміна парозйому котла за зміни навантаження споживача;
- зміна паропровідності котла під час зміни навантаження топки.

У барабанних котлів відхилення рівня перехідних режимах може відповідати знаку математичного небалансу. У стаціонарних режимах барабан та екранні труби котла заповнені водою при температурі кипіння, в якій містяться бульбашки пари. Питома вага пари в барабанних казанах значно відрізняється від питомої ваги води. Різниця питомої ваги пари і води, а, отже, і обсягів при зміні режиму, призводить до відхилення рівня в той чи інший бік. Відхилення рівня води в барабані котла при порушенні стаціонарного режиму у бік, протилежний знаку матеріального балансу, отримала назву "набухання" рівня. Так, наприклад, збільшення пароутворення в екранних трубах при матеріальному балансі призведе до витіснення деякої кількості води з екранних труб і підвищення рівня води в барабані. І, навпаки, зниження паропроductивності котла спричинить зменшення парового обміну в екранних трубах і, відповідно, зниження рівня.

Відмінною рисою регулювання рівня в барабані котла є те, що в цьому випадку регулюється рівень двофазного середовища. У зв'язку з цим вплив кожного фактора призводить до зміни стану двофазного середовища і як наслідок – зміни рівня.

Прийнято, що максимально допустимі відхилення води в барабані складають 100 мм від середнього значення, встановленого заводом-виробником. Середнє значення може не збігатися з геометричною віссю барабана. Зниження рівня за межі 100 мм вважається "опуском" води, а перевищення його - "переживленням".

Зниження рівня до місця приєднання опускних труб циркуляційного контуру може призвести до порушення живлення та охолодження водою підйомних труб. Наслідком цього може бути порушення міцності труб в місцях стикування з корпусом барабана, а в найбільш важких випадку - перепал. Надмірне підвищення рівня може призвести до зниження ефективності внутрішньо-барабанних сепараційних пристроїв та передчасного заносу солями пароперегрівача. Переживлення барабана та закидання частинок води в турбіну може спричинити тяжкі механічні пошкодження її ротора та лопаток. Поста-

чання барабана водою здійснюється за двома нитками трубопроводів живильної води, одна з яких служить резервною.

Відхилення рівня води в барабані від середнього значення пов'язано з наявністю небалансу між припливом живильної води та витратою пари; воно відбувається в наслідок зміни вмісту пари в пароводяної суміші підйомних труб за рахунок коливань тиску пари в барабані або зміни теплосприйняття випарних поверхонь нагріву. Зміна рівня під дією небалансу між витратою пари та води описується рівнянням:

$$F(\rho_B - \rho_{\Pi}) \frac{d\Delta H_6}{dt} = G_{\Pi B} - G_{\Pi \Pi}, \quad (1)$$

де F – площа дзеркала випаровування, m^2 ; ρ_B, ρ_{Π} – щільність води та насиченої пари, kg/m^3 , $G_{\Pi \Pi}$ та $G_{\Pi B}$ – витрати пари та води, kg/c . Приймаючи, що

$$F(\rho_B - \rho_{\Pi}) \frac{H_6^0}{G_{\Pi B}^0} = \frac{1}{k_{\Pi}}; \quad (2)$$

$$(G_{\Pi B} - G_{\Pi \Pi}) \frac{1}{G_{\Pi B}^0} = x; \frac{\Delta H_6}{H_6^0} = y$$

та перейшовши до безрозмірних величин після інтегрування одержимо:

$$y(t) = k_{\Pi} \int_0^t x(t) dt. \quad (3)$$

З цього рівняння можна визначити час проходження рівня від мінімального H_6^{min} до максимального H_6^{max} допустимого значення при ступінчастому збудженні витратою води:

$$\Delta t = \frac{F(\rho_B - \rho_{\Pi})(H_6^{max} - H_6^{min})}{G_{\Pi B} - G_{\Pi \Pi}}. \quad (4)$$

Крива перехідного процесу за рівнем води в барабані та збудженні витратою живильної води приведена на рис. 1, а. Динаміка цієї ж ділянки при збудженні витратою пари, рис. 1, б [7].

Опукла форма кривої зміни рівня при збудженні збільшенням витрати пари пояснюється тим, що в перший момент після збудження, рівень води в барабані зростає в результаті різкого зменшення тиску пари. Це в свою чергу призводить до збільшення вмісту пари в підйомних трубах циркуляційного контуру та зростання рівня.

Після того, як тиск пари в барабані прийме нове стале значення, відповідне новому значенню витрати пари, зміна рівня буде протікати у відповідності з рівнянням матеріального балансу. Дане явище носить назву "збудження" рівня.

При математичному моделюванні динаміки ділянки за рівнем при збудженні витратою пари, описується різницею аперіодичної та інтегруючої ланок.

Синтез САР

Виходячи з вимог до регулювання рівня води у барабані, автоматичний регулятор повинен забезпечувати сталість середнього рівня незалежно від навантаження котла та інших збуджуючих впливів.

У перехідних режимах зміна рівня може відбуватися досить швидко, тому регулятор живлення для забезпечення малих відхилень рівня повинен підтримувати сталість витрат живильної води та пари.

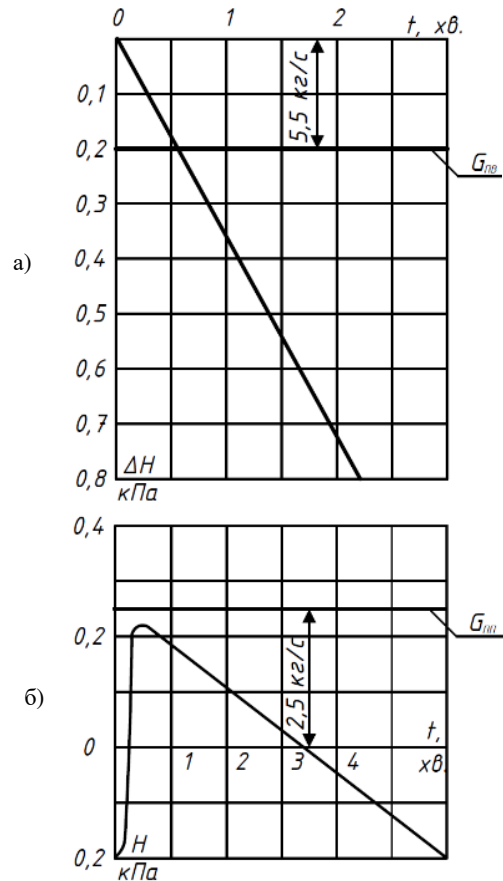


Рис. 1. Динамічні характеристики котла за рівнем води в барабані при збудженнях

Це завдання виконує трьохімпульсний ПІД-регулятор, принципова схема якого зображена на рис. 2. Регулятор переміщує клапан при появі сигналу небалансу між витратою живильної води $G_{\Pi B}$ та пари $G_{\Pi \Pi}$.

Модель парового котла як об'єкту регулювання має наступну передавальну функцію (далі ПФ) з входом - витратою живильної води і виходом - рівнем води в барабані

$$W_B(s) = \frac{1}{T_1 s} - \frac{K}{T_2 s + 1}, \quad (5)$$

У процесі синтезу САР прийемо, що передавальні функції об'єкту регулювання по витраті води та пари протилежні за знаком, а рівні за величиною, тобто $W_{\Pi}(s) = -W_B(s)$.

За експериментальними даними отримана ПФ об'єкта регулювання по каналу витрати води:

$$W_B(s) = \frac{1}{60s} - \frac{2}{22.4s + 1}, \quad (6)$$

По каналу керуючого впливу є два контури регулювання. Перший контур складається з регулятора $W_P(s)$, трубопроводу $W_{TP}(s)$, датчика витрати води та підсилювальної ланки K_2 . Цей контур призначений для стабілізації витрати живильної води при збудженні v , що характеризує зміну тиску в трубопроводі. Еквівалентна передавальна функція першого контуру має наступний вигляд:

$$W_{P*}(s) = \frac{W_K(s) * W_{TP}(s)}{1 + K_2 * W_P(s) * W_{TP}(s)} \quad (7)$$

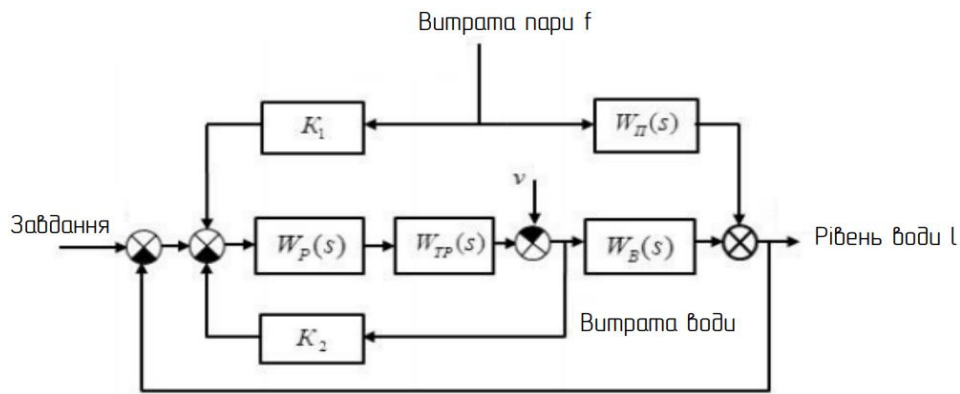


Рис. 2. Структурна схема трьохімпульсної САР:

$W_{\Pi}(s), W_B(s)$ - ПФ об'єкта регулювання по каналах витрати води та витрати пари;
 $W_{TP}(s)$ - ПФ трубопроводу; $W_P(s)$ - ПФ регуляторів; $W_K(s)$ - ПФ коригувального пристрою

Другий контур містить об'єкт регулювання $W_{P*}(s)$ та датчик рівня $W_B(s)$. Оскільки інерційність елементів в першому контурі незначна порівняно з інерційністю об'єкта регулювання, то перехідні процеси в першому контурі завершаться раніше, ніж у другому контурі. Перший контур стабілізує витрату води, де застосовуються ПІ або ПІД регулятори. Після завершення перехідного процесу в першому контурі, його передавальну функцію можна розглядати як передавальну функцію деякого умовного регулятора для другого контуру.

Маємо:

$$W_{P*}(s) = \frac{W_P(s) \cdot W_{TP}(s)}{1 + K_2 \cdot W_P(s) \cdot W_{TP}(s)} = \frac{1}{\frac{1}{W_P(s) \cdot W_{TP}(s)} + K_2}, \quad (8)$$

Оскільки:

$$K_2 = \frac{1}{W_P(s) \cdot W_{TP}(s)}, \quad (9)$$

то

$$W_{P*}(s) \approx \frac{1}{K_2} = K_{P*}, \quad (10)$$

Таким чином при охопленні ПІ або ПІД регулятора глибоким зворотнім зв'язком у вигляді підсилювальної ланки, це з'єднання можна розглядати як деякий наведений П-регулятор.

Розрахунок зовнішнього контуру полягає в виборі параметру цього умовного П-регулятора. Користуючись методом Циглера - Нікольса для зовнішнього контуру САР, отримуємо:

$$K_P = 0,307 = \frac{1}{K_2}, \quad (11)$$

Відповідно:

$$K_2 \frac{1}{0,307} \approx 3,257 \frac{1}{K_2}, \quad (12)$$

Використовуємо ПІ-регулятор з параметрами настройки $K_{\Pi}=0,1$; $T_I=4$ сек.

Проведемо розрахунок коефіцієнта K_1 коригувального пристрою по каналу витрати води. Передаточна функція системи по каналу витрати пара має вигляд:

$$W_{lf}(s) = \left(K_1 + W_{\Pi}(s) * \frac{1}{W_P(s)} * \frac{1}{W_B(s)} \right) * W_{lg}(s), \quad (13)$$

де

$$W_{lg}(s) = \frac{W_P(s) \cdot W_B(s)}{1 + W_{P*}(s) \cdot W_B(s)} \quad (14)$$

передавальна функція системи по заданому впливу. Беручи $W_{\Pi}(s) \approx -W_B(s)$ та $W_{P*}(s) = K_{P*}(s)$, отримуємо $W_{lf}(s) \approx \left(K_1 * -\frac{1}{K_{P*}} \right) * W_{lg}(s)$. Для забезпечення умови інваріантності беремо $W_{lf}(s) = 0$, тобто $K_1 = \frac{1}{K_{P*}} = K_2 = 3,257$.

Моделювання САР

Моделювання САР проводилось в середовищі Simulink. Схема САР представлена на рис. 3. Результати моделювання наведені у табл. 1 та 2.

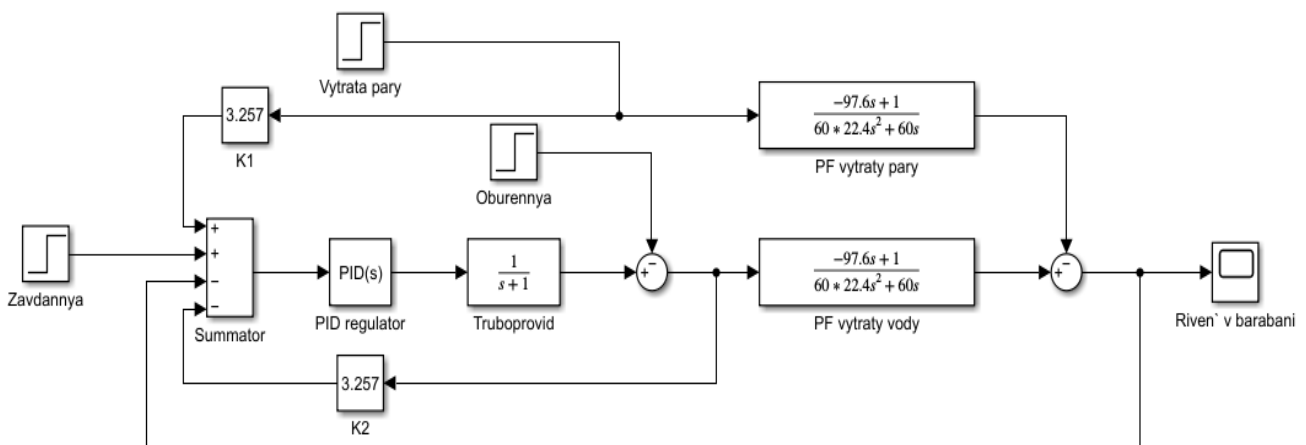


Рис. 3. Схема САР у середовищі Simulink

Таблиця 1 – Показники якості перехідного процесу САР

Час регулювання (сек)	Перерегулювання (%)
438	5,09

Таблиця 2 – Зміни рівня при ступінчастому збільшенні витрати пари на 1 тон/годину та при ступінчастому зменшенні витрати води через зміну тиску в трубопроводі на 1 тон/годину

Вид збуджування	Максимальне відхилення рівня (мм)
Збільшення витрати пари на 1 тон/год	0,065
Зменшення витрати пари на 1 тон/год	0,07

Висновки

Отримана САР з показником перерегулювання у 5,09% та часом регулювання 438 секунд. При ступінчастому впливі відхилення рівня становить

0,065мм при збільшенні витрати пари на 1 тон/год та 0,07мм при зменшенні на 1 тон/год. Одержані результати задовольняють вимогам до САР рівня води у барабані.

Результати моделювання показали, що САР з синтезованим трьохімпульсним регулятором забезпечує необхідну якість регулювання. САР забезпечує швидкодію перехідного процесу та точність регулювання, крім цього вона ефективно усуває впливи зміни витрати живильної води та пари, які часто відбуваються в процесі роботи котельного агрегату.

Серед майбутніх напрямків досліджень у сфері САР слід виділити технології машинного навчання. Сучасні котли оснащені датчиками, які записують інформацію про численні робочі параметри, включаючи температуру, тиск і витрати. Після цього ці дані можуть бути відправлені на платформу обробки та аналізу в реальному часі. Оператори можуть передбачати такі проблеми та вживати запобіжних заходів, використовуючи алгоритми машинного навчання для виявлення тенденцій і відхилень у даних.

СПИСОК ЛІТЕРАТУРИ

- Deshpande Purva, Nilima Warke, Prakash Khandare, Vijay Deshpande. "Thermal power plant analysis using artificial neural network." In 2012 Nirma University International Conference on Engineering (NUiCONE), pp. 1-6. IEEE, 2012, DOI:10.1109/NUiCONE.2012.6493290
- Douglas T. Mugweni, Hadi Harb, "Neural Networks-based Process Model and its Integration with Conventional Drum Level PID Control in a Steam Boiler Plant ", International Journal of Engineering and Manufacturing (IJEM), Vol.11, No.5, pp. 1-13, 2021, DOI:10.5815/ijem.2021.05.01
- Yang P., Peng D.G., Yang Y.H., Wang Z.P. "Neural networks internal model control for water level of boiler drum in power station" – Proceedings of 2004 International Conference 302 on Machine Learning and Cybernetics, vol. 5, 26-29 Aug. 2004, p.3300 – 3303, DOI:10.1109/ICMLC.2004.1378607
- S. Surendran and V. Kumar, "Neural network based pi controller parameter calculation on a boiler drum level system" Procedia Technology, ICETES, vol. 24, pp. 1616–1622, 2016 DOI:10.1016/j.protcy.2016.05.164
- G. A. Hassaan, "Control of Boiler-drum water level using PID, PD-PI, PI-PD and 2DOF controllers", International Journal of Engineering and Techniques, vol.10, issue 1, 10 pages, 2024
- Пура А.С., Тарахтій О.С. "Аналіз існуючих схем регулювання рівня в барабані парового котла." Молодий вчений. 2019. No 11(75). С. 163–166.
- Попович М.Г., Ковальчук О.В. Теорія автоматичного керування – К.: Либідь. 2007. – 655 с.
- Zak S.H. Systems and Control. Oxford University Press, 2003. – 770 p
- Александров Е.Е., Голуб О.П., Костенко Ю.Т. Теорія автоматичного керування. Т1.- Харків: ХДПУ, 2000. -154с.
- Александров С.Є., Козлов С.П., Кузнецов Б.І. Автоматичне керування рухомими об'єктами і технологічними процесами: Підручник у 3-х томах. Т.1 Теорія автоматичного керування. – Харків: НТУ „ХПІ”, 2002. - 490 с.

Received (Надійшла) 27.05.2024

Accepted for publication (Прийнята до друку) 24.07.2024

System of automatic regulation of the water level in the boiler drum of the power plant

P. Kachanov, M. Tarasenko

Abstract. Introduction. The article discusses the automatic water level control system in the drum of a steam boiler. In the context of modern requirements for energy installations, the importance of precise and stable control of water level in the drum becomes crucial for ensuring safety and efficiency of their operation. The aim of the article is to develop and evaluate an effective regulation system that eliminates water level fluctuations in the drum caused by changes in feed water flow or superheated steam. An analysis of existing technical solutions for implementing water level control systems in the drum has been conducted. A model of a three-element control system has been proposed, which has gained the most popularity in this direction due to its characteristics. Through analysis and simulation of the system in the Simulink environment, it has been confirmed that the developed system meets all specified requirements and allows for predicting its performance under real conditions. Based on the conducted research, it can be concluded that the proposed system has significant potential for increasing the reliability and efficiency of thermal power plants.

Keywords: Automatic Regulation System, Drum Steam Boiler, Feed Water Level.

Ю. І. Олімпієва

Державний університет інформаційно-комунікаційних технологій, Київ, Україна

ЗАБЕЗПЕЧЕННЯ ФУНКЦІОНАЛЬНОЇ СТІЙКОСТІ ВИРОБНИЧИХ ПРОЦЕСІВ ПРОМИСЛОВИХ ПІДПРИЄМСТВ НА ОСНОВІ НЕЙРОАДАПТИВНОЇ СИСТЕМИ

Анотація. У статті розглядається проблема забезпечення функціональної стійкості виробничих процесів промислових підприємств на основі нейроадаптивних систем. В умовах індустрії, де виробничі процеси стають дедалі складнішими та автоматизованими, важливість надійних і стійких систем управління зростає. Проводиться аналіз сучасних підходів до використання нейронних мереж для адаптивного управління виробничими процесами, що дозволяє підвищити їхню ефективність та надійність. Аналізуються різні методи підвищення функціональної стійкості, включаючи адаптивні алгоритми навчання, регуляризацію та техніки виявлення і корекції помилок. Особлива увага приділяється інтеграції нейроадаптивних систем з існуючими технологіями та виробничими лініями, а також їх здатності до швидкого відновлення після збоїв і адаптації до змінних умов експлуатації. На основі проведеного аналізу запропоновано новий алгоритм діагностування на основі нейроадаптивної системи, що сприятиме підвищенню стійкості та надійності виробничих процесів. При проектуванні апаратної реалізації нейромереж враховуються такі характеристики нейрочипів, як масштабованість, вартість розробки, сумісність з минулими та майбутніми версіями. Тому серед існуючих нейрочипів було обрано Google TPU v4, оскільки у нього високий коефіцієнт продуктивності, а фреймворки машинного навчання, такі як TensorFlow та PyTorch, надають набір операцій для управління мережевим взаємодією вузлів системи. Важливо зазначити, що нейронні мережі можуть використовуватися в комплексі з іншими методами та системами діагностування, такими як експертні системи, фізичні моделі, статистичні методи.

Ключові слова: функціональна стійкість, діагностування, нейронні мережі, нейроадаптивні системи, алгоритм, нейрочипи, виробничий процес, промислові підприємства.

Вступ

Більшість інформаційних систем, які використовуються сьогодні в різноманітних сферах життя, є складними. Вони утворюють багаторівневі структури, які не можна описати звичайною сумою взаємодії її елементів. Такі складні технічні системи створюються для розв'язання спеціальних задач. Стрімкий розвиток таких систем приводять до ускладнення їх функціонування та виникнення непередбачуваних збоїв. Тому підтримка безперервної роботи складних технічних систем, виявлення та локалізація різних несправностей набуває все більшого значення.

Для забезпечення автономного функціонування системи можна досягти за допомогою властивості функціональної стійкості [1]. Дана властивість забезпечує надійну роботу системи, незважаючи на різноманітні потоки відмов і несправностей протягом часу її роботи.

Постановка завдання. Поняття функціональної стійкості вперше було введено науковцем Машковим О.А. в 1990-х роках, які на той час розв'язував конкретні прикладні задачі. Інші науковці, такі як Барабаш О.В., Собчук В.В. та ін., продовжують вести дослідження в даному напрямі [2-3]. У своїх роботах вони описують методи забезпечення функціональної стійкості для вдосконалення технічних характеристик складних технічних систем, що працюють в екстремальних умовах. Також приділяється увага зменшенню кількості порушень при роботі систем завдяки вчасному діагностуванню таких збоїв та перерозподілу задач між обчислювальними вузлами [4-6]. Однак розвиток елементної бази обчислювальних систем, складності сучасних і, особливо, перспективних автономних динамічних систем, дозволяє розширити область застосування методів забезпечення

функціональної стійкості. У сучасних умовах дуже активно розвивається область нейронауки, де вже представлено багато різних методів для вирішення різних інформаційно-технологічних проблем, серед яких одним з найбільш ефективних є штучні нейронні мережі. Тому для забезпечення високого рівня функціональної стійкості потрібно створювати нові методи або вдосконалювати існуючі. Нейронні мережі можуть бути ефективним інструментом, який дозволить створити глибоку ієрархію прийняття рішень з врахуванням рекомендацій, отриманих з моніторингу роботи виробничих процесів у системі. Буде розглянуто особливості основних положень теорії штучного інтелекту, а саме нейронних мереж, з порівняльними характеристиками апаратної реалізації нейромереж.

Аналіз останніх досліджень і публікацій. Забезпеченню функціональної стійкості присвячено низка робіт. В роботі [7] автори описують початковий етап забезпечення функціональної стійкості, а саме контроль стану інформаційної системи, і якщо є відмова, то включається самодіагностування системи. У [8] науковці розробили методологію побудови ефективної системи самодіагностики інформаційних систем на прикладі підприємств металургійної та енергетичної промисловості, за результатами сформульовано критерії достатності діагностичної інформації за відсутності обмежень щодо виконання елементарних перевірок. У статті [9] розглядається методологія забезпечення функціональної стійкості об'єктів критичної інфраструктури через представлення функціонування системи у вигляді формалізованого процесу. У [10] розкрито механізми самоорганізації інформаційних неоднорідних мереж, проаналізовано нові показники та критерії визначення функціонально стійких систем, запропоновано мережі відповідно до концепції SON. Розглянуто

особливості застосування методу обернених задач динаміки для керування відновленням, отримано вираз для керуючої сили та проведено моделювання для етапу визначення керуючої сили [11]. Для об'єктів критичної інфраструктури запропоновано метод побудови закону управління безпекою критичних об'єктів інфраструктури в умовах зовнішніх неконтрольованих впливів, а саме метод побудови надійної системи управління, що дозволяє компенсувати неконтрольовані зовнішні впливи [12]. У розглянутих роботах для забезпечення функціональної стійкості не використовувалися нейронні мережі

У статті [13] вже розглянуто особливості основних положень теорії штучного інтелекту, а саме нейронних мереж для забезпечення функціональної стійкості виробничих процесів промислових підприємств, досліджено можливість застосування нейронних мереж для діагностики стану систем та практичного застосування інструментарію нейронних мереж для виявлення та локалізації дефектів в роботі систем, що є запорукою забезпечення функціональної стійкості виробничих процесів підприємства. Також удосконалено методіку забезпечення властивості функціональної стійкості інформаційної системи підприємства. У статті [14] запропоновано новий підхід до визначення технічного стану мікропроцесорних систем в базисі програмно-реконфігурованої логіки, які є основою термінальних компонентів технологічної системи. З метою автоматизації системи контролю та діагностування мікропроцесорних систем запропоновано реалізувати принцип самодіагностування, в основу якого покла-

дено ідеї штучного інтелекту. У [15] досліджуються процеси глобальної трансформації інфраструктури інформаційних технологій на фоні масового впровадження кіберфізичних систем та проривних здобутків в галузях штучного інтелекту і робототехніки у виробництво та технологічні процеси. Описано способи застосування нейромереж в моделюванні процесів механічної обробки металів різанням. Дано універсальну методіку побудови нейромережевих моделей процесу механічної обробки на базі штучної нейронної мережі зустрічного поширення. У [16] досліджується стійкість класу нейронних мереж із затримкою, що змінюється в часі. Для GSC мережевих мереж, коли кількість змінних рішення залишається незмінною, інформація про затримку часу та її похідні додатково вихопується за допомогою підходу розподілу затримки. На основі цих методів наведено покращені критерії стабільності.

Детальний аналіз останніх досліджень і результатів досліджень вказує на те, що задача забезпечення функціональної стійкості за допомогою нейронних мереж є новою та перспективною в сучасному світі.

Метою дослідження є забезпечення функціональної стійкості виробничих процесів промислових підприємств на основі нейроадаптивної системи.

Виклад основного матеріалу

Автоматизація процесу контролю параметрів виробничих процесів на сучасних підприємствах є важливим фактором для організації випуску якісної продукції (рис. 1).

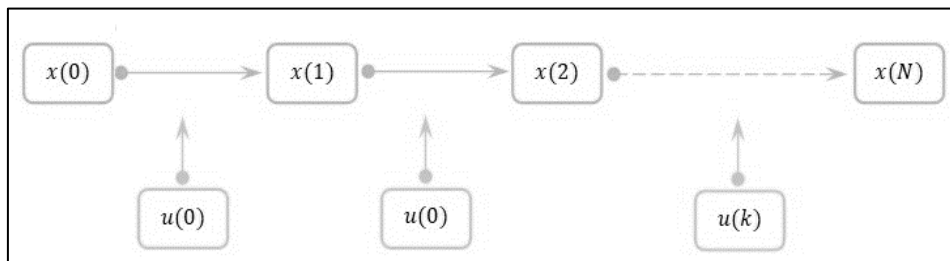


Рис. 1. Топологія лінійного технологічного виробничого процесу

Випуск продукції складається з декількох етапів, на кожному з яких висуваються вимоги до параметрів та характеристик сировини та готової продукції [15]. Нехай $x(i)$ – це набори параметрів на кожному i -му етапі, $i=1, 2, \dots, N$. На кожному етапі для гарантування досягнення параметрів $x(i)$ впливають зовнішні параметри $u(i)$, наприклад, ефект від роботи, енергетичний ефект, хімічні та ін. Вважаємо, що якість продукції на кожному етапі залежить від дотримання технології та забезпечення контролю за необхідними параметрами на кожному попередньому кроці.

Ще введемо додаткові позначення:

- $A(i)$ — матриця залежності показників якості продукції на $(i+1)$ -му етапі від показників на i -му етапі (матриця виробничого процесу);

- $C(i)$ — матрицю, яка визначає структуру впливу на виробничий процес $u(i)$.

Математична модель технологічного процесу, що забезпечується інформаційними системами виробничого підприємства, представляється так:

$$x(i+1) = A(i)x(i) + C(i)u(i), \quad i = 1, 2, \dots, N, \quad (1)$$

$$x(i) \in R^n, \quad A(i) \in R^{n \times n}, \quad C(i) \in R^{n \times m}, \quad u(i) \in R^m,$$

Відтак забезпечення функціональної стійкості виробничого процесу залежить від здатності забезпечувати керованість процесом в кожному виробничому центрі та контролювати виникаючі несправності. Тому для забезпечення цієї властивості важливо застосовувати діагностику для виявлення несправних елементів у системі та вчасно їх локалізувати. Отже, функціональна стійкість виробничих процесів надає такі переваги:

- безперебійність роботи: стійке функціонування обладнання та технологічних систем гарантує

безперебійний випуск продукції, мінімізуючи простоту та втрати;

- якість продукції: стійкість параметрів технологічних процесів забезпечує стабільну якість виробленої продукції, що відповідає стандартам та очікуванням споживачів;

- ефективність виробництва: безвідмовна робота обладнання та мінімізація браку сприяють підвищенню ефективності виробництва, знижуючи витрати та збільшуючи рентабельність;

- безпека праці: стійкість до відмов та аварій забезпечує безпечне робоче середовище для персоналу, мінімізуючи ризики травм та виробничих аварій.

Наразі для діагностування складних технічних систем пропонується застосовувати штучні нейронні мережі, які дозволяють оптимізувати виробничі процеси, підвищувати їх надійність та стійкість до відмов. Зазначимо, що нейронні мережі є інструментом, який дозволяє створювати глибоку ієрархію прийняття рішень з врахуванням місця, виду та рівня дефекту, який виник в системі управління і, як наслідок, може використовуватись в діагностуванні. Нейронні мережі можуть аналізувати дані з сенсорів та інших джерел, щоб передбачити можливі несправності обладнання. Це дозволяє проводити обслуговування на основі реальних потреб, а не за фіксованим графіком, що підвищує надійність та зменшує простоту.

Виділимо деякі можливості нейронних мереж в контексті забезпечення функціональної стійкості виробничих процесів:

1. Прогнозування відмов обладнання та обслуговування: нейронні мережі можуть аналізувати історичні дані про виробничі процеси та прогнозувати майбутні відмови обладнання або неполадки. Наприклад, рекурентні нейронні мережі (RNN) можуть враховувати часові залежності між даними і передбачати моменти, коли обладнання може вийти з ладу. Це дозволяє планувати профілактичне обслуговування та уникнути непередбачених зупинок у виробництві.

2. Виявлення аномалій у виробничому процесі: нейронні мережі засновані на зворотному розповсюдженні помилок (backpropagation), можуть навчитися виявляти аномалії у виробничих процесах. Вони можуть виявляти відхилення від нормальної роботи обладнання або процесів, що може свідчити про потенційні проблеми, такі як пошкодження або несправності.

3. Оптимізація параметрів технології: нейронні мережі можуть бути використані для оптимізації параметрів виробничих процесів з метою підвищення продуктивності та зниження витрат. Наприклад, з використанням глибоких нейронних мереж (DNN) можна розробити моделі, які адаптуються до різноманітних умов виробництва та швидко знаходять оптимальні режими роботи.

4. Автоматизоване управління виробничим процесом: НМ можуть використовуватися для автоматизованого управління виробничим процесом, реагуючи на зміни в режимі реального часу та підтримуючи його стійкість. Один з ефективних прикладів

нейронної мережі для автоматизованого управління виробничим процесом – це нейронні контролери або адаптивні контролери, які використовуються для автоматичного керування об'єктами зі складною динамікою або змінюваними параметрами. Одним із найпоширеніших прикладів таких мереж є нейронні мережі зі зворотнім поширенням помилок (Backpropagation Neural Networks – BPNN).

5. Прогнозування попиту та управління запасами сировини та готової продукції: нейронні мережі можуть бути використані для прогнозування попиту на продукцію та управління запасами сировини та готової продукції. Це допомагає уникнути надлишкових запасів або нестачі, що може призвести до перебоїв у виробництві. Наприклад, може бути ефективним використання глибокої нейронної мережі, такої як Long Short-Term Memory (LSTM) або Gated Recurrent Unit (GRU). Вони часто використовуються для аналізу часових рядів та здатні ефективно моделювати довгострокові залежності у даних.

Кожен конкретний випадок вимагає вибору підходящої архітектури нейронної мережі з урахуванням конкретних характеристик виробничого процесу. Наприклад, для аналізу часових рядів можуть бути ефективними рекурентні нейронні мережі, а для класифікації даних – згорткові нейронні мережі. Вибір конкретної архітектури також може залежати від обсягу даних, складності моделі та доступних обчислювальних ресурсів. Отже, для забезпечення функціональної стійкості виробничих процесів можна широко застосовувати різні класи нейронних мереж для забезпечення діагностики стану обладнання на кожному виробничому центрі. Навчаючи нейронні мережі з урахуванням умов функціональної стійкості технологічного процесу, буде забезпечена ефективна робота як виробничого обладнання, так і поточний контроль дотримання якості продукції згідно визначеної системи толерансів.

Для досягнення функціональної стійкості системи важливо проводити її діагностування, що дозволить враховувати всі особливості системи та визначити, чи працюють модулі системи належним чином, оперативно виявляти несправності та вживати відповідні заходи для забезпечення надійності та продуктивності системи. Система діагностування повинна складатися з об'єкту управління, області діагностичних ознак та алгоритму прийняття рішень. Від об'єкту до області надходять сигнали, які аналізуються алгоритмом та видається висновок про правильність функціонування досліджуваного об'єкту. В залежності від місця та способу використання нейронної мережі можна отримати різні структури систем діагностування (рис. 2). На рис. 2, а представлена структура, де в якості області ознак використовуються змінні параметри об'єкту (наприклад, вхідні сигнали). Нейронна мережа тут використовується в якості пристрою прийняття рішень. На рис. 2, б представлена структура, де нейронна мережа використовується як модель динамічної системи, а в якості алгоритму прийняття рішень – будь-який алгоритм класифікації, наприклад, алгоритм виявлення розлагодження. Структура, яка представлена

на рис. 2, в, використовує стандартний підхід з використанням моделі для генерації нев'язок, в той час як нейронний класифікатор розв'язує задачу виявлення та локалізації дефектів. У структур, які зображені на рис. 2, б та 2, в, відбувається суміщення аналітичних

методів з нейромережним підходом до розв'язання пов'язаних між собою задач генерації ознак та прийняття рішень. На рис. 2, г представлена структура для випадку, коли система діагностування виявляє дефекти, але не буде їх локалізувати.

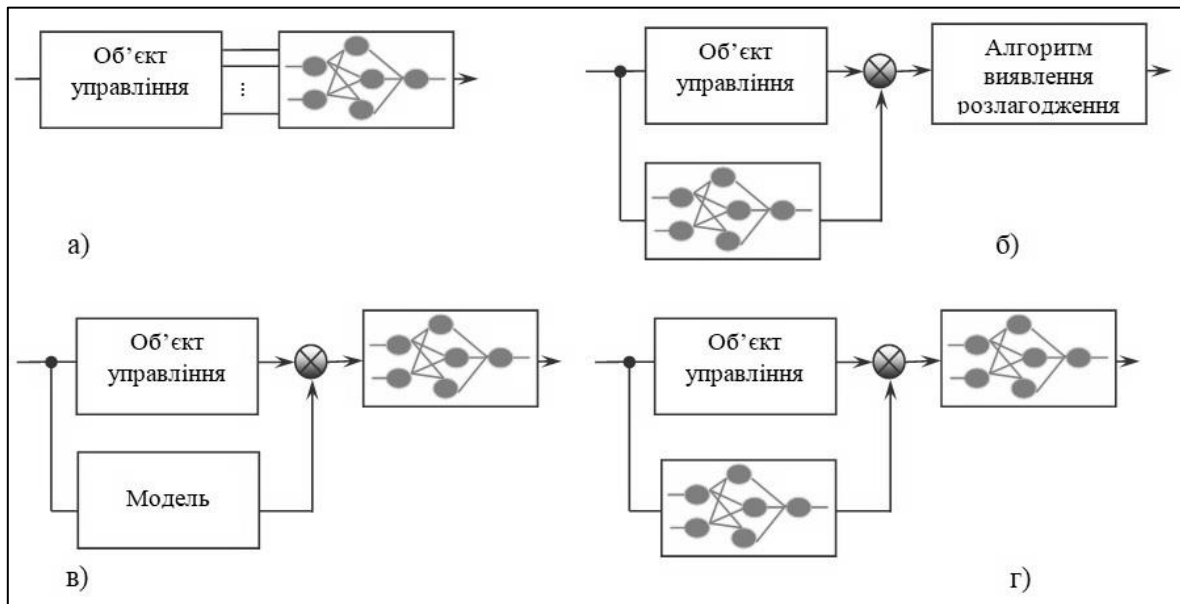


Рис. 2. Структури діагностування на основі нейронних мереж

У випадку використання нейронної мережі в якості алгоритму класифікації незалежно від структури і якщо нейронна мережа має два виходи, то розв'язується задача виявлення. Коли ж входів більше двох, розв'язується задача виявлення та локалізації дефектів і, можливо, визначення величини дефекту.

Отже, при використанні нейромережевого підходу можливо проводити обчислення паралельно, а це в свою чергу дає можливість реалізувати системи управління вищого порядку за прийнятних показників збіжності і, отже, досягти вищої якості управління. Нейромережний підхід до реалізації багатовимірних просторових систем управління дозволяє вирішувати проблеми, що стояли перед розробниками щодо необхідності виконання векторно-матричних операцій високої розмірності в реальному часі.

Проектування нейронної системи представляє собою складний і трудомісткий процес, у якому вибір алгоритму діагностування системи є дуже важливим кроком. Розглянемо алгоритм діагностування системи за допомогою нейроадаптивної системи.

Алгоритм діагностування

Крок 1. Дослідження та аналіз процесу з точки зору нейромережних технологій.

Крок 2. Визначення типу нейронів в мережі.

Крок 3. Визначення кількості шарів нейронної мережі і число нейронів в шарах в залежності від заданої кількості областей розбиття області ознак при умові мінімізації числа нейронів в мережі.

Крок 4. Синтез нейромережі зі змінною структурою:

1) так – вибір алгоритму побудови структури багатоваріантної нейромережі з метою досягнення найкращої відповідності складності розв'язуваної задачі;

2) ні – синтезувати структуру нейромережі з визначенням способу створення зв'язків між шарами: прямими, перехресними, оберненими, комбінованими.

Крок 5. Вибір алгоритму адаптації багатоваріантної нейромережі.

Крок 6. Діагностика синтезованої нейромережі.

Крок 7. Оцінка функціональної стійкості системи.

Крок 8. Вибір варіанту перерозподілу задач між працюючими модулями.

Крок 9. Оцінка результатів діагностування.

Промисловий попит на комп'ютери для машинного навчання зріс у мільйон разів за останні шість років і щороку продовжує збільшитись удесятеро. Для апаратної реалізації нейромережі потрібно обрати нейрочіп із вбудованими нелінійними перетвореннями. Їх виробляють в багатьох країнах світу й вони створюються для конкретних систем. Розглянемо класифікацію чипів [17]:

- за способом подання інформації – цифрові, аналогові та гібридні;

- за типом реалізації нейроалгоритмів – з апаратною та програмно-апаратною реалізацією;

- за характером нелінійних перетворень – апаратно реалізовані та перепрограмуючі;

- за побудовою мереж – з жорсткою та змінною нейромережевою структурою;

- систолічні;

- нейросигнальні.

Орієнтація у виконанні нейромережових операцій зумовлює з одного боку підвищення швидкостей обміну між пам'яттю та паралельними арифметичними пристроями, а з іншого боку – зменшення часу вагового підсумовування (множення та накопичення) за рахунок застосування фіксованого набору команд типу реєстр-реєстр.

Відмітимо, що при проектуванні апаратної реалізації нейромереж враховуються такі характеристики, як масштабованість, вартість розробки, сумісність з минулими та майбутніми версіями. І якщо для розробника критичні терміни розробки і вартість, то слід звернути увагу на сигнальні чіпи або чіпи для каскадних архітектур. Отже, загальноприйнятих рекомендацій по вибору конкретної бази в наш час не існує. Все залежить від вимог розробленої системи.

Для оцінки продуктивності нейрочіпів будемо використовувати такі показники.

- Теоретична максимальна продуктивність нейрочіпа в операціях з рухомою комою з одинарною точністю (FP32) за секунду. Високий показник TFLOPS вказує на те, що нейрочіп може обробляти великі обсяги даних з високою швидкістю, що робить його придатним для складних завдань машинного навчання. Однак важливо зазначити, що TFLOPS – це лише теоретичний показник, і реальна продуктивність може відрізнятись залежно від конкретного завдання та програмного забезпечення.

- Максимальна тактова частота – це максимальна швидкість, на якій може працювати нейрочіп.

- Розмір кристала – це загальна площа кремнієвої пластини, на якій розміщені транзистори нейрочіпа. Більший розмір кристала може дозволити розмістити більше транзисторів, що може призвести до кращої продуктивності.

- Кількість транзисторів – це загальна кількість транзисторів на нейрочіпі. Більша кількість транзисторів зазвичай вказує на те, що нейрочіп є більш складним і може виконувати більше операцій.

- Чіпів на CPU-хост- це кількість нейрочіпів, які можна підключити до одного CPU. Більша кількість нейрочіпів може призвести до кращої масштабованості та загальної продуктивності системи.

- Теплова розсіююча потужність – це максимальна кількість тепла (теоретичний максимум), яку може генерувати нейрочіп. Низький показник вказує на те, що нейрочіп є більш енергоефективним.

- Найбільша масштабована конфігурація – це максимальна кількість нейрочіпів, які можна використовувати разом у одній системі. Більша масштабована конфігурація може призвести до кращої продуктивності для дуже складних завдань.

В [17] виділяють дві базові лінії розвитку обчислювальних високопродуктивних систем з масовим паралелізмом (ВСМП): з модифікованими послідовними алгоритмами та надпаралельних нейромережових алгоритмів вирішення різних завдань. До цієї інформації слід додати, що сучасні тенденції розвитку ВСМП характеризуються наявністю більшої кількості базових ліній розвитку, які неможна чітко окреслити, бо вони часто перетинаються та поєднуються одна з одною.

Умовно ці лінії розвитку можна подати таким чином:

1) підхід використання модифікованих послідовних алгоритмів як і раніше, використовується для адаптації класичних алгоритмів до багатопроекторних систем, проте з'являються нові методи оптимізації та розробки алгоритмів, які враховують специфіку сучасних архітектур, таких як GPU, TPU та нейроморфні процесори;

2) підхід використання нейромережових алгоритмів продовжує розвиватися, з'являються нові архітектури та алгоритми навчання, що значно розширюють можливості нейромереж. Нейромережі використовуються для нових задач, таких як генерація тексту, переклад мов та управління складними системами;

3) підхід використання нейроморфних обчислень прагне створити комп'ютерні системи, які імітують роботу людського мозку. Нейроморфні системи мають потенціал значно перевершити традиційні комп'ютери за енергоефективністю та продуктивністю при вирішенні певних задач;

4) підхід використання квантових обчислень. Квантові комп'ютери мають потенціал вирішувати деякі задачі, які неможливо вирішити на традиційних комп'ютерах, наприклад, факторизацію великих чисел;

5) підхід використання гетерогенних обчислювальних систем, які поєднують в собі різні типи процесорів, такі як CPU, GPU, TPU та нейроморфні процесори.

Наведемо декілько найбільш популярних нейрочіпів у країнах Європи та США:

NVIDIA A100 (США, 2020) використовується у широкому спектрі застосунків, включаючи штучний інтелект, машинне навчання, обробку природної мови та наукові обчислення;

Intel Habana Labs Gaudi (США, 2020) спеціально розроблений для Intel AI Framework, він пропонує високу продуктивність та енергоефективність, що робить його популярним вибором для застосунків штучного інтелекту, які потребують низької затримки;

Graphcore Colossus GC30 (Великобританія, 2020) пропонує високу продуктивність та масштабованість, що робить його популярним вибором для великих проектів машинного навчання; Google TPU v4 (США, 2021) та наступні версії v5e, v5p спеціально розроблені для TensorFlow, платформи машинного навчання Google;

AMD MI25X (США, 2021) є конкурентом NVIDIA A100 і пропонує схожу продуктивність за нижчою ціною;

Imveo NeuroScale 2 (Франція, 2021) спеціально розроблений для штучного інтелекту в галузі охорони здоров'я, популярний вибір для медичної діагностики та обробки зображень;

NVIDIA H200 (США, 2023) – це перший графічний процесор із пам'яттю HBM3E, яка відрізняється від звичайної HBM3 (high bandwidth memory, HBM) вищою швидкістю, нова пам'ять дозволить прискорювачу швидше працювати з величезними обсягами даних для генеративного ШІ та високопродуктивних обчислювальних навантажень;

Cerebras Wafer-Scale Engine 3 (WSE-3) (США, Cerebras Technologies, 2024) – це третє покоління суперкомп'ютерної платформи Cerebras, є одним із найпотужніших нейрочіпів у світі, пропонує безпрецедентну продуктивність, що робить його популярним вибором для дослідницьких проектів штучного інтелекту, які потребують обчислення великих обсягів даних;

Google TPU Trillium (США, очікується в кінці 2024) – це шосте покоління фірмового тензорного процесора, новий чіп майже вп'ятеро продуктивніший за попередника TPU v5e, при цьому Trillium на 67% енергоефективніший, ніж TPU v5e [18-21].

Найбільш перспективними є нейрочіпи розробників NVIDIA та Google.

В табл. 1 наведено їх характеристики [21-24].

Таблиця 1 – Порівняльна характеристика деяких нейрочіпів

Характеристики	Nvidia A100	Graphcore MK2 IPU	Google TPU v4	Google TPU v3
Вихід на ринок	2020	2021	2020	2018
Пікова продуктивність, TFLOPS	312 (bf16), 624 (i8)	250 (bf16)	275 (bf16 або int8)	123 (bf16)
Тактова частота, МГц	1095/1410	1850	1050	940
Тех. процес, нм Розмір кристала, мм ²	7 826	7 832	7 <600	16 <700
Кількість транзисторів, мільярдів	54	59	22	10
Мікросхем на хост ЦП	4	4	4	8
Номинальна теплова потужність, Вт	400	300	інформація не доступна	
Енергоспоживання нейрочіпа, коли він не використовується, мін/сер/макс, Вт	інформація не доступна		90 121/170/192	123 175/220/262
Міжчіпове з'єднання: кількість з'єднань та швидкість передачі даних, ГБ/с	12 25	3 64	6 50	4 70
Конфігурація найбільшого масштабу, к-ть чипів	4216	256	4096	1024
Архітектура процесора	SIMT	MIMD	SIMD 2D	
Процесори / Чип	108	1472	2	2
Потоки / Ядра	32	6	1	1
Розріджені ядра/ Чип			4	2
Оперативна пам'ять, МБ	38	865	123 (CMEM)+ 30 (VMEM) + 10 (spMEM)	30 (VMEM) + 5 (spMEM)
Розмір файлу реєстрів	26	1,3	0,24	0,24
Ємність НВМ2, ГБ пропускна здатність, ГБ/с	74 2039	0	30 1200	30 900

В графі «Архітектура процесора» (табл. 1) розшифруємо аббревіатури:

Single Instruction Multiple Threads (SIMT) – це архітектура процесора, яка дозволяє йому виконувати одну інструкцію над декількома потоками даних одночасно. Це може значно прискорити обробку даних, адже процесор може виконувати багато операцій одночасно.

Принцип роботи SIMT.

Крок 1. Процесор отримує одну інструкцію.

Крок 2. Процесор розбиває інструкцію на декілька менших частин.

Крок 3. Процесор виконує кожну частину інструкції над декількома потоками даних одночасно.

Multiple Instruction Multiple Data (MIMD) – це архітектура процесора, яка дозволяє йому виконувати декілька інструкцій над декількома потоками даних одночасно, що робить MIMD-процесори дуже універсальними, адже вони можуть виконувати широкий спектр задач.

Вони більш гнучкі, адже можуть виконувати декілька різних інструкцій одночасно для неоднорідного набору даних.

Принцип роботи MIMD.

Крок 1. Процесор отримує декілька інструкцій.

Крок 2. Процесор розбиває кожну інструкцію на декілька менших частин.

Крок 3. Процесор виконує кожну частину інструкції над декількома потоками даних одночасно.

Його можна використовувати в ЦП, ГП, нейрочіпах та інших комп'ютерних системах, може бути менш ефективним для задач машинного навчання,

адже йому може бути складно оптимізувати виконання декількох різних інструкцій.

Single Instruction 2D Data (SIMD 2D) – це архітектура процесора, яка є розширенням класичної SIMT архітектури. Вона дозволяє процесору виконувати одну інструкцію над двовимірними матрицями даних одночасно.

Принцип роботи SIMD 2D.

Крок. Процесор отримує одну інструкцію.

Крок 2. Процесор розбиває інструкцію на декілька менших частин.

Крок 3. Процесор виконує кожну частину інструкції над двовимірними матрицями даних одночасно.

SIMD 2D поєднує в собі переваги SIMT та MIMD, пропонуючи високу продуктивність для задач з двовимірними даними.

В графі «Оперативна пам'ять» (табл. 1) є три типи ОП.

СМЕМ – це кеш-пам'ять першого рівня (L1), яка використовується для зберігання інструкцій та даних, які використовуються найчастіше.

VMEM – це кеш-пам'ять другого рівня (L2), яка використовується для зберігання даних, які використовуються трохи рідше, ніж дані в СМЕМ.

spMEM – це спеціальна пам'ять, яка використовується для зберігання даних, які потрібні для роботи певних блоків нейрончипа.

Отже, отримано наступні результати порівняння характеристик нейрончипів:

- технологічний вузол (розмір транзистора, що використовується в мікросхемі) у перших трьох вдвічі менший за TPU v3, що дозволяє розмістити більше транзисторів на одній мікросхемі, а це може призводити до кращої продуктивності та енергоефективності;

- TPU v4 має вдвічі більше матричних множників, ніж TPU v3. Матричні множники використовуються для виконання основних операцій машинного навчання, таких як множення матриць та згортки, тому вони є важливими для виконання завдань машинного навчання, а отже їх збільшення може суттєво покращити продуктивність.

- Graphcore MK2 IPU має найвищу тактову частоту: на 68%/31% вищу за Nvidia A100, на 76% вищу, ніж TPU v4 та на 97% вищу, ніж TPU v3. Тактова частота стосується кількості циклів, які процесор може виконати за одну секунду. Більша тактова частота може призвести до кращої продуктивності.

- Graphcore MK2 IPU має 12 міжз'єднань (порівняно з 6 та 4 у Google TPU) на чіп дозволяють створювати мережеві топології з меншим діаметром мережі.

- частина покращення продуктивності на ват (ефективності) походить від самого 7-нм техпроцесу (близько 40%). Решта покращення походить від змін у дизайні TPU v4, таких як балансування конвеєра та реалізація тактового гетування (для динамічного відключення блоків, які не використовуються в даний момент, що допомагає зменшити споживання енергії). Продуктивність на ват – це міра енергоефективності процесора. Вона вимірюється в GFLOPS/W

(гігафлопсах на ват) і показує, скільки GFLOPS процесор може виконувати на один ват споживаної потужності.

- HBM (High Bandwidth Memory) - це тип пам'яті, який може дуже швидко передавати дані. TPU v4 має в 1,4 та 1,3 рази більшу пропускну здатність пам'яті HBM порівняно з TPU v3 та MK2 IPU відповідно, в свою чергу Nvidia A100 в 2 рази більшу порівняно з TPU v4, що означає, що він може швидше отримувати доступ до даних з пам'яті.

- TPU v4 має 128 МБ вбудованої пам'яті scratchpad (СМЕМ), якої немає в TPU v3. Scratchpad-пам'ять – це тип дуже швидкої пам'яті, до якої процесор може отримувати доступ дуже швидко. Це може бути корисно для зберігання даних, які часто використовуються під час обчислень.

На рис. 3 представлені результати експерименту застосування TPU v4 та TPU v3 однакової конфігурації (кількості TPU чипів) для різних застосунків (наскільки суперкомп'ютери TPU v4 перевершують TPU v3).

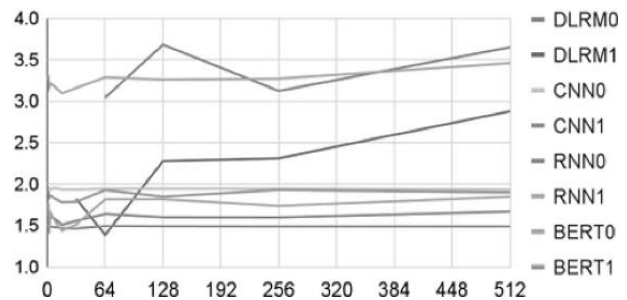


Рис. 3. Прискорення TPU v4 порівняно з v3 за однакової кількості чипів

Виходячи з порівнянь у таблиці, не дивно, що для більшості застосунків при однаковій конфігурації, TPU v4 працює в 1.5-2.0 рази швидше, ніж TPU v3. Застосунок DLRM0 працює в 3.0-3.5 рази швидше, а DLRM1 - в 2.8 рази швидше на конфігурації з 512 чипами. Це пояснюється тим, що TPU v4 має вдвічі більше обчислювальних блоків (SC) та їхня тактова частота є вищою. Але найбільш неочікуваний результат отримано для застосунку RNN1 - він працює в 3.3 рази швидше на TPU v4. Це зумовлено тим, що малі ваги та малий розмір пакетів даних RNN1 отримують значну перевагу від пропускну здатності внутрішньої пам'яті СМЕМ порівняно з зовнішньою пам'яттю HBM [21].

TPUv2/v3 мають менші кристали у більш старому напівпровідниковому процесі та нижчі ціни на хмару, незважаючи на те, що вони менш зрілі на багатьох рівнях стеку апаратного/програмного забезпечення, ніж CPU та GPU. Ці добрі результати, незважаючи на технологічні недоліки, свідчать, що підхід TPU є економічно ефективним і може забезпечити високу архітектурну ефективність у майбутньому [22].

В цілому, Google TPU v4 – це достатньо потужний нейрончип, який розроблений спеціально для задач машинного навчання. Він має значно більшу продуктивність, ніж A100, але дорожчий. TPU v4 добре підходить для вирішення складних задач машинного

навчання, таких як навчання великих мовних моделей та розробка нових алгоритмів штучного інтелекту. TPU v4 підтримує TensorFlow та PyTorch, що робить його зручним для використання з популярними фреймворками машинного навчання. TPU v4 доступний через Google Cloud TPU, що робить його доступним для розробників у всьому світі. Зазначається, що платформа Cloud TPU v4 в 1,2-1,7 рази продуктивніша і витрачає в 1,3-1,9 рази менше енергії, ніж платформи на базі NVIDIA A100 у системах аналогічного розміру. Хоча, поки компанія не порівнювала TPU v4 з більш новими прискорювачами NVIDIA H100 через їхню обмежену доступність і 4-нм архітектури (порівняно з 7-нм у TPU v4) [21-22].

Отже, нейронні мережі володіють широким спектром можливостей для застосування в задачах діагностування стану виробничих процесів та складних технічних систем. Їхня здатність до навчання, швидкість обробки інформації, гнучкість, автоматизація, можливість виявлення неявних несправностей та прогнозування несправностей роблять їх цінним інструментом для підвищення надійності, безпеки та ефективності експлуатації систем. Важливо зазначити, що нейронні мережі можуть використовуватися в комплексі з іншими методами та системами діагностування, такими як експертні системи, фізичні моделі, статистичні методи. Їх застосування для діагностування системи промислового підприємства може призвести до значної економічної вигоди, таких як зниження витрат на ремонт, простої та брак, а також до підвищення якості продукції та конкурентоспроможності підприємства.

Висновки

Функціональна стійкість є критичним фактором для успішного впровадження технологій у промислових підприємствах. У контексті швидкого розвитку індустрії, підприємства дедалі частіше використовують передові технології, такі як нейронні мережі та інші методи штучного інтелекту, для оптимізації виробничих процесів, контролю якості та прогнозування технічного обслуговування. Однак, висока залежність від таких технологій робить

питання їхньої надійності і стійкості до збоїв надзвичайно актуальними.

Функціональна стійкість у промислових умовах означає здатність систем підтримувати свою продуктивність і точність при змінних робочих умовах, впливі зовнішніх збурень і можливих апаратних відмовах. Це включає стійкість до непередбачуваних змін вхідних даних, адаптивність до нових умов експлуатації, а також здатність до швидкого відновлення після збоїв. Недоліки в стійкості систем можуть призвести до зупинки виробничих ліній, значних фінансових втрат та зниження якості продукції.

У сучасних умовах дуже активно розвивається область нейронауки, де вже представлено багато різних методів для вирішення різних інформаційно-технологічних проблем, серед яких одним з найбільш ефективних є штучні нейронні мережі. Нейронні мережі можуть бути ефективним інструментом, який дозволить створити глибоку ієрархію прийняття рішень з врахуванням рекомендацій, отриманих з моніторингу роботи виробничих процесів у системі. У статті розглянуто особливості основних положень теорії штучного інтелекту, а саме нейронних мереж, з порівняльними характеристиками апаратної реалізації нейромереж. Визначено структури діагностування виробничих процесів, описано алгоритм діагностування системи за допомогою нейроадаптивної системи. Для апаратної реалізації нейромережі обрано нейрочіп із вбудованими нелінійними перетвореннями, а саме Google TPU v4, оскільки у нього високий коефіцієнт продуктивності, системне програмне забезпечення TPU v4 відповідає за розподіл обчислень, оптимізацію передачі даних та моніторинг системи та фреймворки машинного навчання, такі як TensorFlow та PyTorch, надають набір операцій для управління мережевим взаємодією вузлів системи.

Подальшим перспективним шляхом розвитку роботи є застосування експертних систем, які дозволять вдосконалити процес діагностування виробничих процесів, надаючи їм більшої гнучкості та універсальності.

СПИСОК ЛІТЕРАТУРИ

1. Собчук В.В. Основи забезпечення функціональної стійкості інформаційних систем підприємств в умовах впливу дестабілізуючих факторів: монографія / В.В. Собчук, О.В. Барабаш, А.П. Мусієнко. К: Міленіум. – 2022. – 272 с. https://www.researchgate.net/publication/363474851_Basis_for_functional_stability_of_information_systems_businesses_under_the_influence_of_destabilizing_factors
2. Sobchuk A.V. Assessment methods of functional stability of wireless sensor networks / A.V. Sobchuk, O.V. Barabash, A.P. Musienko // Телекомунікаційні та інформаційні технології. – 2019. – № 3 (64). – С. 46–54. <https://doi.org/10.31673/2412-4338.2019.034654>
3. Собчук В.В. Методи контролю і діагностування інформаційної системи підприємства за принципом адаптивного накопичення діагностичної інформації / В.В. Собчук, О.В. Барабаш, А.П. Мусієнко, О.А. Капустян // Вісник Київського національного університету ім. Т. Шевченка. – 2020. – Вип. 4. – С. 69–78. <https://doi.org/10.17721/1812-5409.2020/4.10>
4. Barabash O. Ensuring the functional stability of the information system of the power plant on the basis of monitoring the parameters of the working condition of computer devices / O. Barabash, O. Svynchuk, I. Salanda, V. Mashkov, M. Myroniuk // Advanced Information Systems. – 2024. – Vol. 8, no. 2. – P. 107–117. <https://doi.org/10.20998/2522-9052.2024.2.12>
5. Барабаш О. В. Програмне забезпечення контролю справного стану інформаційних систем в енергетичній галузі для забезпечення функціональної стійкості / О. В. Барабаш, О. В. Свинчук, Бандурка О. І. // Сучасний захист інформації. – 2024. – № 2 (58). – С. 41–49. <https://doi.org/10.31673/2409-7292.2024.020005>
6. Svynchuk O. Development of the information system for monitoring time changes in forest plantations based on the analysis of space images / O. Svynchuk, O. Bandurka, O. Barabash, O. Ilin, A. Lapin // Eastern-European Journal of Enterprise Technologies. – 2022. – Vol. 5, no. 2 (119). – P. 31–41. <https://doi.org/10.15587/1729-4061.2022.265039>

7. Mashkov V.A. Self-Checking of modular systems under random performance of elementary checks. *Engineering Simulation / V.A. Mashkov, O.V. Barabash // Amsterdam: OPA. – 1995. – Vol. 12. P. 433–445.*
8. Sobchuk V. Adaptive accumulation and diagnostic information systems of enterprises in energy and industry sectors / V. Sobchuk, O. Barabash, A. Musienko, O. Svynchuk // *E3S Web of Conferences. – 2021. – Vol. 250. – P. 82–87. <https://doi.org/10.1051/e3sconf/202125008002>*
9. Barabash O. System analysis and method of ensuring functional sustainability of the information system of a critical infrastructure object / O. Barabash, V. Sobchuk, A. Musienko, O. Laptiev, V. Bohomia, S. Kopytko // In: Zgurovsky, M., Pankratova, N. (eds) *System Analysis and Artificial Intelligence. Studies in Computational Intelligence. – 2023. – Vol 1107. – Springer, Cham. – P. 117–192. https://doi.org/10.1007/978-3-031-37450-0_11*
10. Maksymuk O.V. A system of indicators and criteria for evaluation of the level of functional stability of information heterogenic networks / O.V. Maksymuk, V.V. Sobchuk, I.P. Salanda, Yu.V. Sachuk // *Mathematical modeling and computing. – 2020. – Vol. 7, no. 2. – P. 285–292. <http://doi.org/10.23939/mmc2020.02.285>*
11. Mashkov O. Features of determining controlling effects in functionally-stablesystems with the recovery of a control / O. Mashkov, V. Chumakevych, O. Sokulsky, L. Chyrun // *Mathematical Modeling And Computing. – 2019. – Vol. 6, no. 1. – P. 85–91. <http://doi.org/10.23939/mmc2019.01.085>*
12. Laptiev O. The method of construction of the law of safety management of critical infrastructure objects under the conditions of external uncontrolled influences / O. Laptiev, O. Barabash, I. Tsyganivska, D. Obidin, A. Sobchuk // *CEUR Workshop Proceedings. – 2023. – Vol. 3624. – P. 291–300. https://ceur-ws.org/Vol-3624/Paper_24.pdf*
13. Собчук А.В. Застосування нейромереж для забезпечення функціональної стійкості виробничих процесів / А.В. Собчук, Ю.І. Олімпієва // *Телекомунікаційні та інформаційні технології. – 2020. – № 2 (67). – С. 13–28. <http://doi.org/10.31673/2412-4338.2020.021328>*
14. Тюлюпа С.В. Самодіагностування як спосіб підвищення кіберстійкості термінальних компонентів технологічної системи / С.В. Тюлюпа, Ю.Я. Самохвалов, П.В. Хусаїнов, С.С. Штатенко // *Кібербезпека: освіта, наука і техніка. – 2023. – № 2 (22). С. 134–147. <https://doi.org/10.28925/2663-4023.2023.22.134147>*
15. Собчук В.В. Функціональна стійкість технологічних процесів на основі нелінійної динаміки із застосуванням нейромереж / В.В. Собчук, І.В. Замрій, Ю.І. Олімпієва, С.О. Лаптев // *Сучасні інформаційні системи. – 2021. – Т.5 (2). – С.49–57. <https://doi.org/10.20998/2522-9052.2021.2.08>*
16. Guo-Qiang Kong Stability analysis of delayed neural networks based on improved quadratic function condition / Guo-Qiang Kong, Liang-Dong Guo // *Neurocomputing. – 2023. – Vol. 524. – P. 158–166. <https://doi.org/10.1016/j.neucom.2022.12.012>*
17. Кожем'яко В.П. Сучасний стан, елементна база та порівняльний аналіз характеристик нейрообчислювачів / В.П. Кожем'яко, А.В. Кожем'яко, О.С. Васильківа // *Оптико-електронні інформаційно-енергетичні технології. – 2017. – Т. 32 (2). – С. 29–38.*
18. The future of AI is Wafer-Scale [Електронний ресурс]. – Режим доступу: <https://www.cerebras.net/product-chip/>
19. NVIDIA H200 Tensor Core GPU [Електронний ресурс]. – Режим доступу: <https://www.nvidia.com/en-us/data-center/h200/>
20. NVIDIA DGX H100 [Електронний ресурс]. – Режим доступу: <https://nvdam.widen.net/s/kpwzdrwrwbv/ai-for-enterprise-dgx-h100-datasheet-nvidia-a4-2146027-r3-web>
21. Introducing TPU v4: Googles Cutting Edge Supercomputer for Large Language Models [Електр. ресурс]. – Режим доступу: <https://www.kdnuggets.com/2023/04/introducing-tpu-v4-googles-cutting-edge-supercomputer-large-language-models.html>
22. Jouppe N.P. A domain-specific supercomputer for training deep neural networks / N.P. Jouppe, D.H. Yoon, G. Kurian, S. Li, N. Patil, J. Laudon, C. Young, D. Patterson // *Communications of the ACM. – 2020. – Vol. 63, no. 7. – P. 67–78. <https://doi.org/10.1145/3360307>*
23. Graphcore/ IPU-POD64 Reference Design Datasheet [Електронний ресурс]. – Режим доступу: <https://docs.graphcore.ai/projects/ipu-pod64-datasheet/en/latest/overview.html>
24. NVIDIA A100 Tensor Core GPU Architecture [Електронний ресурс]. – Режим доступу: <https://images.nvidia.com/aem-dam/en-zz/Solutions/data-center/nvidia-ampere-architecture-whitepaper.pdf>

Received (Надійшла) 06.05.2024

Accepted for publication (Прийнята до друку) 10.07.2024

Ensuring functional sustainability of production processes of industrial enterprises based on the neuroadaptive system

Yu. Olimpiyeva

Abstract. The article considers the problem of ensuring the functional stability of production processes of industrial enterprises based on neuroadaptive systems. In an industry where production processes are becoming increasingly complex and automated, the importance of reliable and sustainable control systems is growing. An analysis of modern approaches to the use of neural networks for adaptive management of production processes is carried out, which allows to increase their efficiency and reliability. Various methods of increasing functional stability are analyzed, including adaptive learning algorithms, regularization, and error detection and correction techniques. Special attention is paid to the integration of neuroadaptive systems with existing technologies and production lines, as well as their ability to quickly recover from failures and adapt to changing operating conditions. Based on the analysis, a new algorithm for diagnosing the main neuroadaptive systems is proposed, which will contribute to increasing the stability and reliability of production processes. When designing the hardware implementation of neural networks, such characteristics as scalability, development cost, compatibility with past and future versions are taken into account. Therefore, among existing neurochips, Google TPU v4 was chosen because it has a high performance factor and the system part of the software implements set of operations for managing the network interaction of system nodes. It is important to note that neural networks can be used in combination with other diagnostic methods and systems, such as expert systems, physical models, statistical methods.

Keywords: functional stability, diagnostics, neural networks, neuroadaptive systems, algorithm, neurochips, production process, industrial enterprises.

Marko Radovanović¹, Aleksandar Petrovski^{2,3}, Aner Behlić^{2,3}, Mohamed Zied Chaari⁴, Elshan Giyas Hashimov^{5,6}, Radosław Fellner⁷, Abayomi O. Agbeyangi⁸

¹ University of Defence, Military Academy, Belgrade, Serbia

² “Goce Delcev University” Stip, Skopje, North Macedonia

³ Military Academy “General Mihailo Apostolski”, Skopje, North Macedonia

⁴ Office of Research, Experiments, and Development, Ministry of Defense, Qatar

⁵ Azerbaijan Technical University, Baku, Azerbaijan

⁶ National Defense University, Baku, Azerbaijan

⁷ Fire University, Warsaw, Poland

⁸ Chrisland University, Abeokuta, Ogun State, Nigeria

UNLEASHING AUTONOMOUS FORCES: INTEGRATING AI-DRIVEN DRONES IN MODERN MILITARY STRATEGY

Abstract. The impact of Artificial Intelligence (AI) on international security is unmistakable, as machines are now capable of undertaking tasks traditionally reserved for human intelligence. This shift brings forth a myriad of challenges in international security, affecting both conventional military capabilities and hybrid threats. Simultaneously, AI presents new opportunities to address these challenges, influencing key aspects of collective defense, cooperative security systems and crisis management. Given its profound implications for prosperity and security, effective management of AI requires collaborative efforts. The scope of promise and peril associated with AI is vast, necessitating collective action to mitigate security risks and leverage its potential to restructure operational processes, support missions, and streamline operations. This paper primarily focuses on introducing drones equipped with artificial intelligence and autonomous learning capabilities, exploring their applications in military contexts. It examines the potential for the independent use of AI-driven drones in both combat and non-combat army operations. By leveraging GIS, C5IRS (Command, Control, Computers, Communications, Cyber-Defense, Intelligence, Surveillance, and Reconnaissance), and AI, these drones provide a significant battlefield advantage by operating autonomously and adapting to dynamic ground situations. In a battlespace where the amalgamation of human, information, and physical components is crucial for strategic advantage, interoperability emerges as a vital factor.

Keywords: UAV, YOLOv8, Drone, GIS, Artificial Intelligence (AI), Security, C5IRS.

Introduction

Contemporary combat operations require the deployment of sophisticated combat technologies to effectively carry out assigned missions. Over the years, there has been a substantial enhancement in combat activities, with the application of scientific and technological advancements significantly improving the efficiency and effectiveness of contemporary military units. The use of the most sophisticated combat means is essential for the successful execution of modern combat missions. To carry out these operations effectively, state-of-the-art command and information systems are required. The primary objective is to collect, analyse, and visualize real-time battlefield data, facilitating and expediting the decision-making process for commanders (decision-makers) [1].

The utilization of unmanned aerial vehicles (UAVs) or Unmanned Aircraft Systems (UAS) - with their supporting equipment (for instance drop-off mechanism, reflectors, speakers, thermal cameras, streaming devices) is an integral component of contemporary combat operations. Engagement in military operations demands the adoption of cutting-edge combat technologies to efficiently accomplish designated objectives. Given their adaptable applications and varied capabilities, such vehicles provide a wide array of benefits to units equipped with such combat gear. The integration of UAV into the C5IRS framework enables the acquisition of real-time imagery from the battlefield. This capability empowers

decision-makers to command forces efficiently and promptly during operations. Various types of modern unmanned aerial vehicles, ranging from commercial to armed variants, are employed in present-day combat operations.

The utilization of drones equipped with artificial intelligence and streaming devices in military combat operations represents an advanced technological dimension that is transforming the way battles are led and tactical strategies are executed. The integration of artificial intelligence into drones enables them to operate autonomously and swiftly analyze large volumes of real-time data. This creates the capability to make precise decisions on the battlefield, often within seconds, significantly enhancing the reactivity and efficiency of military forces.

The advantages of this technology range from improved situational awareness and more precise reconnaissance to the rapid identification of potential threats and targets. The use of artificial intelligence enables drones to recognize behavioral patterns, differentiate between friendly and hostile units, and even predict the moves of adversaries. This paper will explore key aspects of integrating artificial intelligence into drones for military purposes, analyze current research, and consider the implications of this technology on the future of warfare.

The purpose of the work is to present drones with artificial intelligence, operating autonomously, used for military and civilian purposes. The following tasks are solved in the article:

1) definition and classification of unmanned aircraft and combat operations;

2) The YOLOv8 approach for enhancing military drone surveillance through AI software capabilities.

The following research methods are used: comparison, abstraction, axiomatic, analysis, synthesis, formalization and induction, modeling. The following results were obtained: This paper introduces a real-time vital detection system for Military Objects utilizing a Convolutional Neural Network within the YOLOv8 (You Only Look Once) framework.

Conclusions: This obtained value and prediction results there is compelling evidence to support the effective utilization of the YOLOv8 algorithm for real-time detection of Military Targets during drone surveillance. The model demonstrates consistently high confidence levels in recognizing Military Targets. Moreover, its lightweight design and swift processing capabilities make it highly deployable. Integrated seamlessly with onboard surveillance cameras on drones, this model ensures near-instantaneous recognition with minimal latency, enhancing the efficiency of military surveillance operations. Furthermore, its adaptability allows for potential integration with other defense systems, promising even greater operational effectiveness.

Literature review

In the paper Bayramov et al. consider the issues of detecting invisible enemy objects using GIS technologies [2]. Bares conducts interoperability modeling for the C4IRS system within a framework of collective security [3]. Petrovski and Toshevski demonstrate the utilization of GIS in geographic reconnaissance and its integration with C5IRS for military objectives [4]. Jović investigates the tactical deployment of drones in counter-terrorism operations [5]. Petrovski et al. investigate the application of GIS in conjunction with the C5IRS system for military geography [6]. Milić et al. examine the potential utilization of drones in urban operational environments [7]. Adamski assesses the efficacy of UCAVs in contemporary armed conflicts [8]. Žnidaršič et al. present various types of drones and counter-drone measures for integration into Serbian Army units [9]. Petrovski and Radovanović scrutinize the integration of drones with the C5IRS system for military purposes [10]. Hashimov and Huseynov provide a brief analysis of the combat application and capabilities of some unmanned aerial vehicles in the armies of developed countries and the Azerbaijani Army [11]. Radovanovic et al. examines the process of selecting UAVs for military and police tactical units through the fuzzy AHP - VIKOR model of multicriteria decision making [12]. Ilić and Tomašević analyze the ramifications of the Nagorno-Karabakh conflict on perceptions of combat drones [13]. Radovanovic et al. explore the feasibility of incorporating drones into mortar units to enhance fire support efficiency through collaborative integration with the C4IRS system [14]. Ciolponea analyzes the integration of unmanned aerial vehicle systems into current combat operations [15]. Szulc analyzes the

possibility of using unmanned combat vehicles in tactical actions in mountainous terrain [16]. Zaher analyses drones and their role in the evolution of warfare generation [17]. Hashimov et al. devoted his research to the use of GIS and photogrammetric technologies in military affairs, determining the coordinates of the desired target, drawing up an orthophoto map of the area for operational decision-making and other actions, constructing a detailed 3-dimensional model of the area for organizing future operational actions [18]. Li et al. presents an algorithm for glove detection based on improved YOLOv8 [19]. Roy and Bahaduri shows DenseSPH-YOLOv5: An automated damage detection model based on DenseNet and Swin-Transformer prediction head-enabled YOLOv5 with attention mechanism [20]. In their study, Duan et al. aims to enhance the speed and accuracy of YOLOv8 algorithm detection by combining its lightweight features with ShuffleNetV2, proposing a simple closure detection method based on ShuffleNetV2 and YOLOv8 [21]. Wang et al. explore automatic crack identification on roads using an improved YOLOv8 approach [22]. Sun et al. apply an enhanced YOLOv8 model for pest detection in agriculture within a complex environment [23]. Wu et al. utilize an enhanced YOLOv8 model for overhead power line damage detection [24]. Wen et al. shows the improved YOLOv8 algorithm based on EMSPConv and SPE-head modules [25]. Niu et al. present an improved YOLOv8 model for application in agriculture [26]. Fazekas explores the potential use of artificial intelligence tools for application in military operations [27]. Kania explores Chinese military innovations in the artificial intelligence revolution [28]. Morgan et al. analyze the military application of artificial intelligence [29]. Walsh et al. showcase geospatial-temporal visualizations for military operations [30]. Bhagat examines the use of artificial intelligence in offensive and defensive military operations [31]. Castro et al. illustrate the use of artificial intelligence in military logistics support operations and its application in Logistics 4.0 [32]. Lee et al. investigates the shift towards intelligent drones with minimal human dependence, showcasing the necessary technological advancements [33]. Shah et al. examines the utilization of artificially intelligent drones in smart city environments [34]. Pomortseva et al. illustrate the specifics of attracting additional resources during the conduct of military operations in modern conditions using the latest technologies and geographic information systems [35].

1 Definition and classification of unmanned aerial vehicles

Until now, there hasn't been a universally accepted definition of unmanned aerial vehicles (UAVs), nor a standardized classification. The European Association of Unmanned Vehicle Systems (EUROUVS) has endeavored to tackle this issue by establishing classifications rooted in factors such as purpose, duration, flight altitude, dimensions, speed, signal range, Maximum Takeoff Weight (MTOW), and more [36].

In terms of control and management models, UAVs are categorized into autonomous systems, self-control systems, radar or radio beam control systems, telecommand control systems, and combined systems (autonomous and non-autonomous). The US Department of Defense has classified UAVs into five distinct categories, as illustrated in Table 1 [36].

Another classification system, originating from Europe, is presented in Table 2. This system appears to

alternate its criteria between size categories (Nano, Micro, Mini), range (Close, Short, Medium), and altitude and endurance. A significant differentiation in the European system is observed at the threshold of 150kg (approximately 330 lbs.). Remarkably, the European agency overseeing aircraft certification exempts unmanned aircraft systems below this weight from their standard certification requirements, leaving airworthiness standards to be managed by individual nations [37-39].

Table 1 – Categorization of unmanned aerial vehicles as outlined by the US Department of Defense [37]


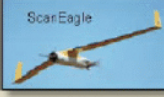


UAS Groups	Maximum Weight (lbs) (MGTOW)	Normal Operating Altitude (ft)	Speed (kts)	Representative UAS	
Group 1	0 – 20	<1200 AGL	100	Raven (RQ-11), WASP	
Group 2	21 – 55	<3500 AGL	< 250	ScanEagle	
Group 3	< 1320	< FL 180		Shadow (RQ-7B), Tier II / STUAS	
Group 4	>1320		> FL 180	Any Airspeed	Fire Scout (MQ-8B, RQ-8B), Predator (MQ-1A/B), Sky Warrior ERMP (MQ-1C)
Group 5		Reaper (MQ-9A), Global Hawk (RQ-4), BAMS (RQ-4N)			

Table 2 – European UAS Classification System

UAS Categories	Acronym	Range (km)	Flight Altitude (m)	Endurance (hours)	MTOW (kg)	Currently Flying
Tactical						
Nano	η	< 1	100	< 1	< 0,025	yes
Micro	μ (Micro)	< 10	250	1	< 5	yes
Mini	Mini	< 10	150 ^a to 300 ^a	< 2	< 30 (150 ^b)	yes
Close Range	CR	10 to 30	3.000	2 to 4	150	yes
Short Range	SR	30 to 70	3.000	3 to 6	200	yes
Medium Range	MR	70 to 200	5.000	6 to 10	1.250	yes
Medium Range Endurance	MRE	> 500	8.000	10 to 18	1.250	yes
Low Altitude Deep Penetration	LADP	> 250	50 to 9.000	0,5 to 1	350	yes
Low Altitude Long Endurance	LALE	> 500	3.000	> 24	< 30	yes
Medium Altitude Long Endurance	MALE	> 500	14.000	24 to 48	1.500	yes
Strategic						
High Altitude Long Endurance	HALE	> 2000	20.000	24 to 48	(4.500 ^c)12.000	yes
Special Purpose						
Unmanned Combat Aerial Vehicle	UCAV	approx. 1500	10.000	approx. 2	10.000	yes
Lethal	LETH	300	4.000	3 to 4	250	yes
Decoy	DEC	0 to 500	5.000	< 4	250	yes
Stratospheric	STRATO	> 2000	>20.000 & <30.000	> 48	TBD	no
Exo-stratospheric	EXO	TBD	> 30.000	TBD	TBD	no
Space	SPACE	TBD	TBD	TBD	TBD	no

TBD = To Be Defined ^a = according to national legislation ^b = in Japan ^c = Predator B

According to the maximum take-off mass, ISO standard characterized even six classes [40]:

- I (0 < mass ≤ 0,25),
- II (0,25 < mass ≤ 0,9),
- III (0,9 < mass ≤ 4),
- IV (4 < mass ≤ 25),

- V (25 < mass ≤ 150),
- VI (150 < mass).

It is also worth mentioning that NATO adopted their own classification in which distinguishes 3 classes (Fig. 1) [41, 42]:

- I (< 150 kg), II (150 kg - 600 kg), III (> 600 kg).

Class	Category	Normal Employment	Normal Operating Altitude	Normal Mission Radius	Primary Supported Commander	Example Platform
Class III (> 600 kg)	Strike/Combat *	Strategic/National	Up to 65,000 ft MSL	Unlimited (BLOS)	Theatre	Reaper
	HALE	Strategic/National	Up to 65,000 ft MSL	Unlimited (BLOS)	Theatre	Global Hawk
	MALE	Operational/Theatre	Up to 45,000 ft MSL	Unlimited (BLOS)	JTF	Heron
Class II (150 kg - 600 kg)	Tactical	Tactical Formation	Up to 18,000 ft AGL	200 km (LOS)	Division, Brigade	Watchkeeper
Class I (< 150 kg)	Small (>15 kg)	Tactical Unit	Up to 5,000 ft AGL	50 km (LOS)	Battalion, Regiment	Scan Eagle
	Mini (<15 kg)	Tactical Sub-unit (manual or hand launch)	Up to 3,000 ft AGL	Up to 25 km (LOS)	Company, Platoon, Squad	Skyjark
	Micro ** (<66 J)	Tactical Sub-unit (manual or hand launch)	Up to 200 ft AGL	Up to 5 km (LOS)	Platoon, Squad	Black Widow

Fig. 1. NATO UAS classification

Petrovski and Radovanović offer a detailed exploration of the definitions and classification of "drone" and "UAV" (depicted in Fig. 2) [10].

UAV encompasses motorized devices controlled remotely or possessing varying degrees of autonomy. Control typically involves communication software, often integrating artificial intelligence and diverse sensors. UAVs serve diverse purposes, from carrying cargo and transmitting real-time data to functioning as WiFi stations. They vary in purpose, construction, environment of use, and energy source. UAV applications span numerous fields including defense, security, agriculture, construction, communication, science, and more.

The term "drone" is broader than "unmanned aerial vehicle," encompassing all UAVs but not vice versa [10].

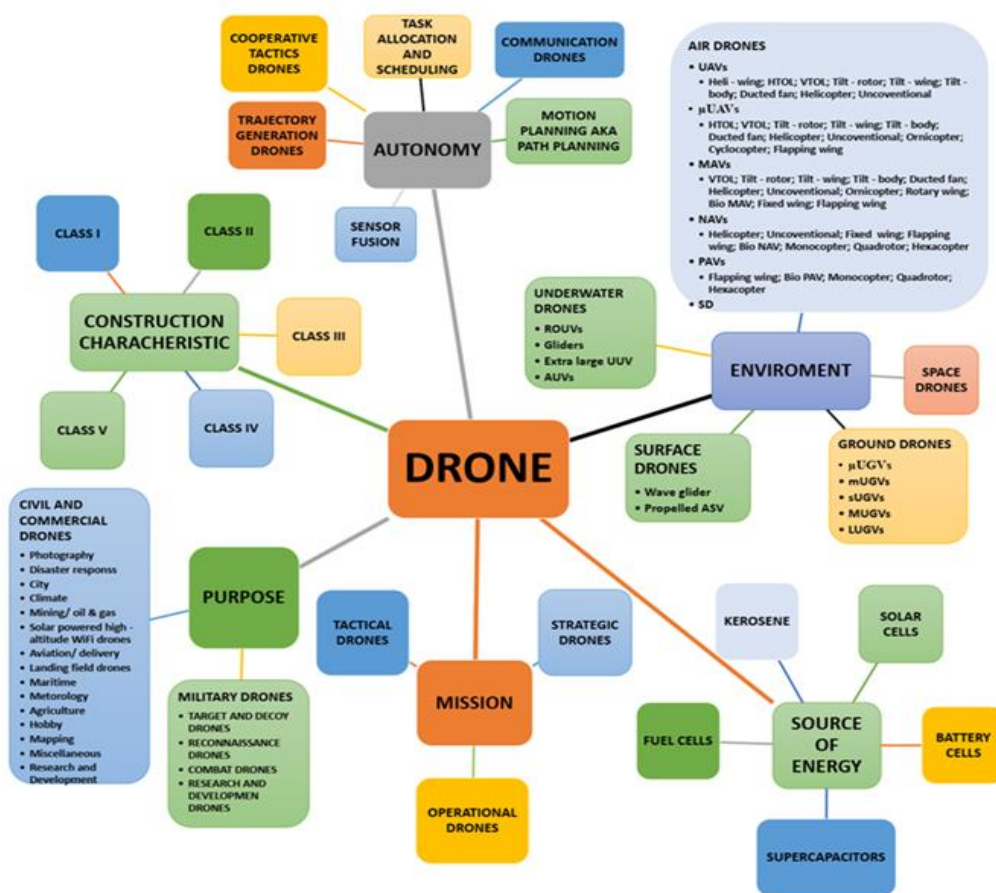


Fig. 2. NATO UAS classification

The categorization of unmanned aerial vehicles (UAVs) implies that their attributes are typically determined by their intended purpose. Analyzing the tactical and technical features of UAVs is essential to assess their suitability for deployment in modern combat operations.

Military utilization of UAVs can be classified into three primary categories: naval, land, and aerial deployment. On the civilian front, UAVs have diverse applications spanning numerous domains, such as geodesy (photogrammetry), agriculture, industrial production, civil protection, disaster management, surveillance of critical infrastructure, environmental

protection, police surveillance, search and rescue operations, intelligence and security services, journalism, commercial activities, and recreational pursuits. Additionally, their versatility and adaptability continue to drive innovation and expand their utility in various fields [43].

2 Combat operations

Land-based military operations form is the bedrock of overall military endeavor, as almost all capabilities, regardless of their deployment in other domains, are fundamentally rooted in or overseen from land. While specific domains may take precedence in certain

situations, resolutions to conflicts typically unfold on land because of its importance as a place where people live, where political decisions are made, and where national power is located. Military forces achieve their objectives through the execution of operations, defined as a series of tactical actions with a common purpose or central theme. Operations vary in several dimensions and encompass different physical environments such as urban, subterranean, desert, jungle, mountain, maritime, and arctic environments. There is also variability in the scale of forces deployed and the duration of operations. Operations bring about changes in the physical, informational, and human dimensions of the operational environment.

Also important are technologies and devices that increase situational awareness and support decision making by providing live images from the battlefield, such as the Mobile Command Centre (Fig. 3) or Tactical Command Case that enable video transmission from several drones to any location, two-way radio communication between commander and pilot, and the creation of a secure wireless WiFi network [43, 44]. Although designed for civilian use and emergency services, they can be successfully used for military purposes, just as civilian commercial-off-the-shelf drones (COTS) were used in the Ukrainian-Russian war [45].



Fig. 3. Example of Mobile Command Centre

The complexity of the operational environment requires leaders who understand both the scientific and artistic aspects of operations. A comprehension of the scientific facets, including combat power ratios, weapons ranges, and movement tables, aids leaders in enhancing synchronization and mitigate risk. However, uncertainty persists, and leaders must employ operational art to make decisions and embrace risk. Intangible elements like the influence of leadership on morale, leveraging shock effect against enemy forces, and gaining the support from populations are inherently human factors that can overcome physical challenges and frequently determine the outcomes of an operation. Army forces face a wide range of challenges, contributing to national objectives across various operational categories, such as large-scale combat operations, limited contingency operations, crisis response, and support to security cooperation.

Armed conflicts are typically characterized by major combat operations that demand intense combat activities and significant logistical resources. These major operations often involve large-scale maneuvers conducted by complex joint forces organized and commanded as functional components. The primary

focus is on preserving freedom of action for one's own forces while denying the same to the adversary. Combat operations involve conventional force-on-force engagements of varying scale, frequency, and intensity between opposing armed forces. The armed forces of a state act to implement the state's national policy and assert dominance over other instruments of power. Combat operations comprise a series of battles and major engagements, leading to intense activity and high logistic consumption. The tempo of activities is usually rapid, requiring the prioritization of resources and the generation of additional combat power. These operations often involve large-scale manoeuvres by complex and multifaceted Joint Task Forces, organized and commanded as functional components [41].

Combat operations refer to military actions that involve direct engagement with enemy forces in a physical confrontation. These operations are a subset of military activities and focus on achieving specific tactical the use of force. Here are key elements related to combat operations:

Direct Engagement: Combat operations involve direct and deliberate engagement with enemy forces. This may include infantry clashes, armored warfare, artillery bombardment, and other forms of direct confrontation.

Tactical Objectives: The primary purpose of combat operations is to achieve immediate and specific tactical objectives on the battlefield. These objectives could include seizing key terrain, neutralizing enemy positions, or disrupting enemy movements.

Fire and Maneuver: Fire and maneuver tactics are fundamental to combat operations. This involves using firepower to suppress or destroy enemy positions while maneuvering troops to gain a positional advantage.

Close Combat: Combat operations often involve close combat, where opposing forces engage each other at close range. This may occur in urban environments, open terrain, or other battlefield conditions.

Strategic Context: While combat operations are tactical in nature, they are conducted within the broader context of a strategic plan. Tactical victories contribute to achieving overall strategic goals in a conflict.

Dynamic and Fluid: Combat situations are dynamic and fluid, requiring commanders to make rapid decisions based on real-time information. The ability to adapt to changing circumstances is critical for success.

Combined Arms: Successful combat operations often involve the coordinated use of combined arms, including infantry, armor, artillery, and air support. The synergy of these elements enhances the effectiveness of military forces.

Risk and Uncertainty: Combat operations inherently involve risk and uncertainty. Commanders must make decisions with incomplete information and navigate through fog of war.

Casualties and Medical Support: Combat operations can result in casualties, and medical support is a critical component. The evacuation and treatment of casualties contribute to the overall effectiveness of the force.

Conclusion and Follow-Up:

Combat operations conclude with an assessment of achieved objectives and may be followed by the consolidation of gains, preparation for subsequent operations, or other strategic actions.

In summary, combat operations are intense, direct engagements with enemy forces, focused on achieving specific tactical goals within the broader context of a military campaign.

3 Utilizing AI-Enhanced Drones in Military Combat Operations

Unmanned aerial vehicles (UAVs) have significantly bolstered global military capacities, catalysing revolutionary shifts in military tactics. The utilization of military drones has profound implications for land forces and legal considerations, in the area of command and control. Drones play a pivotal role in relaying vital information about by providing intelligence on enemy movements, positions, and key targets, commanders are enabled to operate with greater efficiency and make well-informed decisions in the field. While drone technology is relatively new in the military domain, rapid advancements have been made by integrating drones with artificial intelligence. Companies like Shield.AI, AeroVironment, and Lockheed Martin exemplify how defense contractors utilize computer vision technology and image recognition to tackle military challenges while maintaining... jeopardizing human lives. Shield AI's drone, for instance, boasts the capability to navigate unfamiliar terrain without relying on GNSS tracking. This class of UAV empowers military units with rapid data collection capabilities, elevating their agility and situational awareness in tactical reconnaissance, tracking, combat assessment, and cartographic missions. Furthermore, drones afford operators the liberty to make decisions autonomously, alleviating concerns regarding potential rear ambushes. With their versatile functionalities, UAVs play a pivotal role in modern warfare, ensuring swift and effective responses to dynamic battlefield scenarios.

The integration of AI into drones involves a combination of physical components, machine vision, and navigational instruments. Training the artificial intelligence behind the drone involves a supervised learning process, a task considered more challenging than it may seem. From my perspective, combat drone technology emerges as among the most potent and effective tools ever created for intermediate and close combat situations. Its precision strikes and real-time reconnaissance capabilities redefine battlefield strategies, granting military forces unparalleled tactical advantages. Additionally, the continuous advancements in drone technology promise further enhancements in combat effectiveness and mission success rates, cementing their position as indispensable assets in modern warfare.

Recognizing the strategic importance of countering small drones, the Army emphasizes machine learning and AI as pivotal technologies in neutralizing opposing systems. Swarms of drones, capable of simultaneous multi-directional attacks, pose the potential to overwhelm human defenders. Initially designed for anti-

insurgency and defense, drones have demonstrated immense value in such contexts. The aerospace and defense industry is now actively integrating drone technology into various global military programs, offering diverse benefits and advantages across different roles.

Military drones have revolutionized warfare by operating on land, soaring through the skies, and navigating underwater. Evolving over more than fifty years, drones have become key artificial intelligence weapons seamlessly integrated into military forces worldwide. Consequently, armed forces increasingly turn to drones to enhance their capabilities in combat and surveillance.

The benefits of autonomous weapons systems serve as force multipliers, reducing the personnel required for a mission while amplifying the effectiveness of each individual. Advocates also attribute autonomous weapons systems with expanding the operational scope of the battlefield, facilitating engagement in remote or otherwise inaccessible regions. Moreover, these systems offer the potential to minimize casualties by eliminating the need for human intervention in hazardous missions. Additionally, the integration of artificial intelligence enables swift adaptation to dynamic combat scenarios, further enhancing operational efficiency and mission success rates.

4 Transforming Military Drone Surveillance: Harnessing YOLOv8 and AI Software Capabilities

This paper looks at the integration of real-world models into either drones or the AI software powering them (Fig. 4). The focus is on various military drones and UAVs equipped with AI capabilities. AI plays a crucial role in multiple use cases within drone technology, especially in the military context where it is often used to enable autonomous flight through machine vision. It should be noted that although COTS drones have object recognition capabilities (e.g. Active Track in Autels and in DJIs), the algorithms do not identify whether the visible objects are vehicles or people, nor whether they are military vehicles or soldiers.

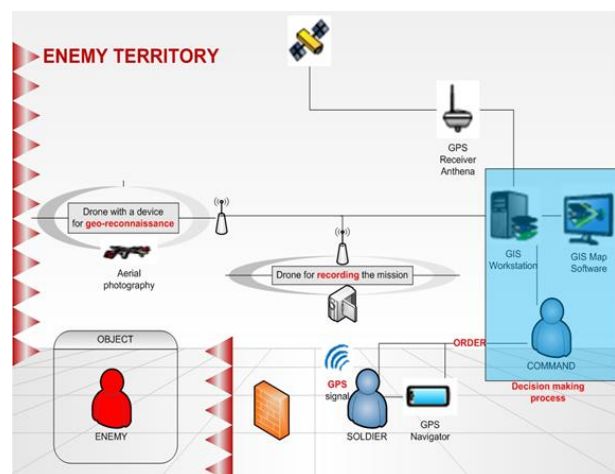


Fig. 4. Model of Geo-Reconnaissance and Command Information System (GRCIS) in UAVs Enhanced with C4IRS Systems and AI

4.1 Advanced Military Drone Surveillance Utilizing the YOLOv8 Method

The outlined framework is designed to identify objects of significant military relevance using images extracted from the live feed from military surveillance drones. Subsequent actions will be contingent upon the acquired outcome. Fig. 5 provides a concise overview of the proposed methodology in this article.

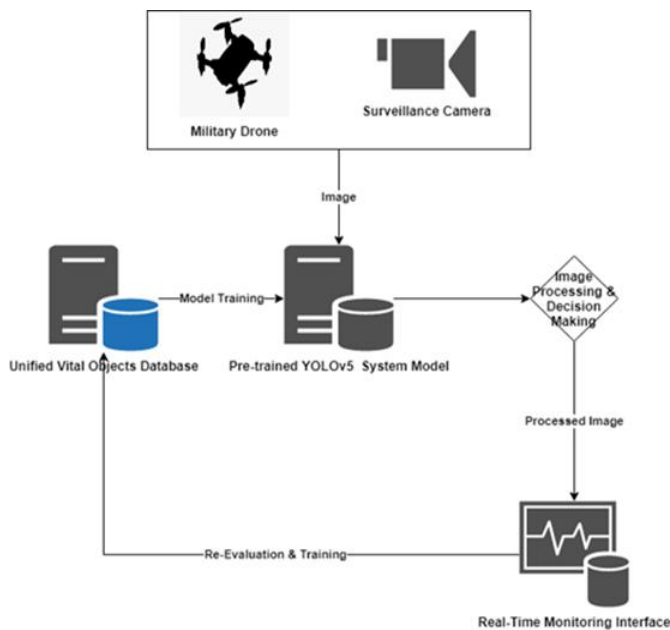


Fig. 5. Methodology Overview

Initially, a comprehensive Database of Vital Military Objects is established by aggregating verified images sourced from diverse online and local repositories. These images encompass crucial surveillance items such as weapons, bunkers, military structures, vehicles, tanks, and artillery pieces. Subsequently, the YOLOv8 model undergoes rigorous training and evaluation using this extensive dataset, ensuring robust performance in real-world scenarios.

To facilitate real-time surveillance, a Surveillance Camera is seamlessly integrated onto a military drone, forming a crucial component of the System Model. This camera transmits live video feeds through the system model, enabling instantaneous data processing. Moreover, sophisticated image filtering algorithms are applied to enhance image clarity and stability, ensuring reliable performance even in adverse weather conditions.

The processed video feed is then presented on the Real-Time Monitoring Interface, providing instant insights into detected vital military objects. Alarms are triggered to alert operators upon the identification of such objects, enhancing situational awareness and enabling prompt decision-making. Instances of false positives and undetected objects identified by human operators are meticulously logged and transmitted to the consolidated Vital Objects Database for further analysis and system refinement.

In essence, this paper introduces a cutting-edge real-time system for the detection of vital military objects, leveraging state-of-the-art Convolutional

Neural Network technology within the YOLOv8 framework. Key contributions of this innovative approach include its seamless integration onto existing surveillance drones, high-precision detection capabilities over considerable distances, and utilization of open-source development software. Additionally, the establishment of a comprehensive dataset for vital military objects, along with its support for military decision-makers in tracking and identifying such objects, underscores its significance in modern warfare.

The development of a real-time system for detecting vital military objects represents a significant advancement in military surveillance technology. By leveraging the power of Convolutional Neural Networks within the YOLOv8 framework, this approach enables seamless integration onto existing surveillance drones and ensures high-precision detection capabilities over vast distances. The establishment of a comprehensive database of vital military objects, coupled with its support for decision-makers in tracking and identifying such objects, underscores its pivotal role in enhancing situational awareness and enabling prompt response in dynamic battlefield scenarios. Moreover, the utilization of open-source development software underscores its accessibility and potential for widespread adoption across military contexts. Moving forward, continued refinement and enhancement of this system hold the promise of further strengthening military capabilities and safeguarding national security interests.

4.2 Advancing Military Surveillance: YOLOv8-Based Detection of Vital Objects

YOLO is built on Convolutional Neural Network (CNN) and serves as a highly efficient end-to-end object detection algorithm [46]. Continuous enhancements have positioned it as a top performer on two official object detection datasets: Pascal VOC (Visual Object Classes) [47] and Microsoft COCO (Common Objects in Context) [48].

The network architecture of Yolov8 is depicted in Fig. 6. Several factors contribute to the selection of Yolov8 as the primary method for training the system model. First and foremost, Yolov8 integrates Cross Stage Partial Network (CSPNet) [49] into Darknet, forming CSPDarknet as its backbone. CSPNet addresses issues related to repeated gradient information in large-scale backbones by incorporating gradient changes into the feature map. This leads to a reduction in model parameters and FLOPS (Floating-Point Operations Per Second), balancing inference speed and accuracy while also minimizing the model size. In military surveillance operations, the speed and accuracy of detection are paramount, and a compact model size significantly enhances inference efficiency on edge devices with limited resources. Additionally, the streamlined model architecture contributes to improved real-time processing capabilities, enabling swift and precise identification of vital military objects in dynamic environments.

Secondly, Yolov8 implements the Path Aggregation Network (PANet) [50, 51] as its neck to enhance information flow. PANet adopts a novel Feature Pyramid Network (FPN) structure with an

improved bottom-up path, enhancing the propagation of low-level features. Simultaneously, adaptive feature pooling, linking the feature grid and all feature levels, facilitates the direct propagation of useful information

in each feature level to the subsequent subnetwork. PANet improves the utilization of accurate localization signals in lower layers, significantly enhancing the location accuracy of detected objects.

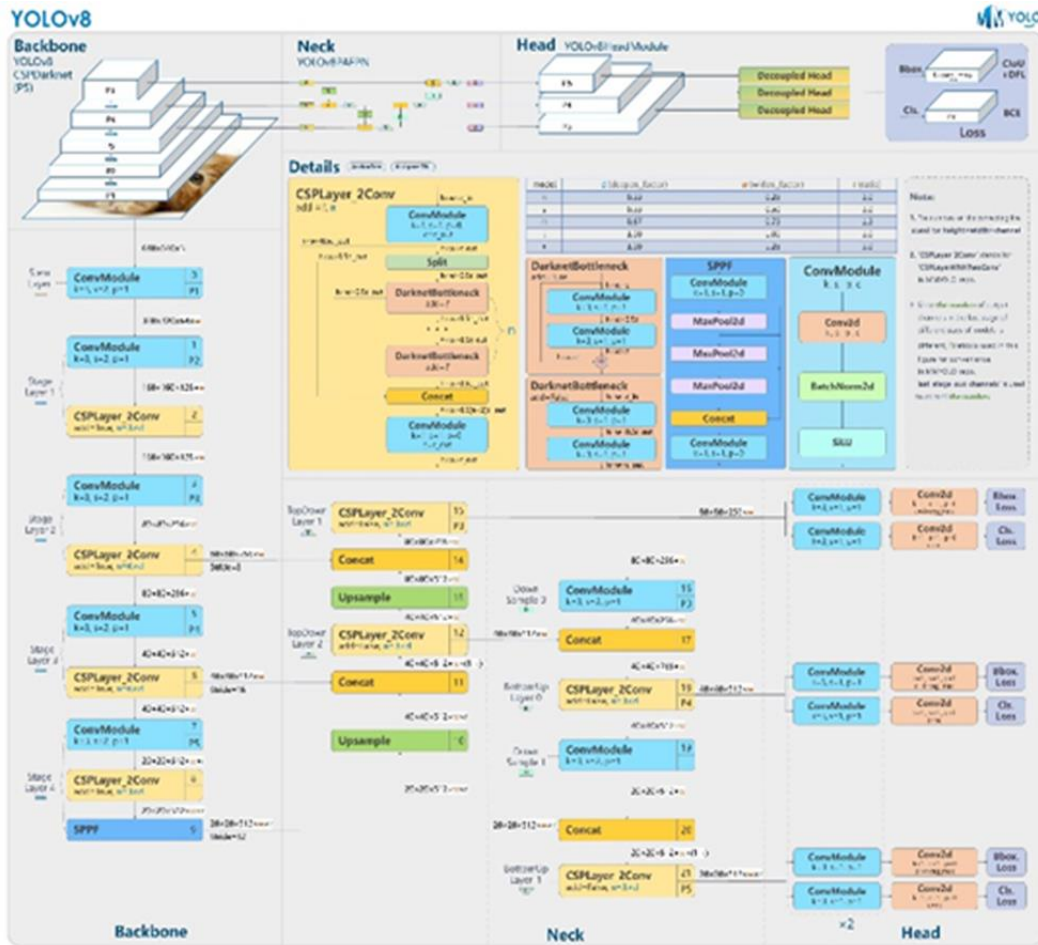


Fig. 6. Network Architecture of Yolov8: Backbone (CSPDarknet), Neck (PANet), and Head (Yolo Layer) [52]

Thirdly, the head of Yolov8, namely the Yolo layer, generates feature maps of three different sizes (18×18 , 36×36 , 72×72) to achieve multi-scale prediction [51]. This capability allows the model to efficiently detect objects of various sizes, ranging from small to large. Moreover, its adaptive nature ensures robust performance across diverse environmental conditions and operational scenarios, further enhancing its utility in military surveillance missions.

In the YOLO method, the input images undergo a single pass through the neural network, predicting the detected objects in the image (Fig. 7). The process involves dividing the input image into various grids based on a predefined grid size. Subsequently, the algorithm predicts the probability of the desired object in each grid. This approach enables the prediction of all classes and the bounding boxes of objects in the image in a single run of the algorithm [53].

The proposed solution, utilizing pre-trained YOLOv8, boasts exceptional lightweight design and rapidity in identifying vital military objects. Maintaining both high accuracy and a swift frame rate (fps) is imperative for effective real-time surveillance and data collection in military operations. Moreover, the

adaptability of the YOLOv8 model ensures robust performance across diverse environmental conditions, further enhancing its suitability for critical military applications.



Fig. 7. Overview of Military Vehicles Recognizable with Pre-Input Images in Drone Databases

To better understand the process, we divide and analyses it step by step [54]:

Grid-Based Approach:

— YOLOv8 divides the input image into a grid of cells.

— Each cell is responsible for making predictions within its boundaries.

— This grid-based approach ensures efficient processing and simplifies object localization.

Predictions in Each Cell:

— For every cell, YOLOv8 predicts:

— **Bounding Boxes:** These represent the coordinates (x, y, width, height) of detected objects.

— **Class Probabilities:** The likelihood that an object belongs to a specific class (e.g., car, person, dog).

— These predictions are made independently for each cell, allowing parallel computation.

Anchor Boxes:

— YOLOv8 introduces anchor boxes to refine localization.

— Each anchor box corresponds to a specific aspect ratio and scale.

— The predicted bounding boxes are adjusted based on anchor box properties.

Feature Extraction Backbone:

— YOLOv8 employs a neural network backbone (often based on Darknet or CSPDarknet).

— This backbone extracts high-level features from the input image.

— These features are crucial for accurate predictions.

Loss Function: Guiding Training:

— During training, YOLOv8 optimizes a loss function.

— The loss considers both localization (bounding box accuracy) and classification (class probabilities).

— It guides the model to learn from mistakes and improve predictions.

Non-Maximum Suppression (NMS):

— After predictions, YOLOv8 applies NMS.

— NMS removes duplicate or overlapping bounding boxes.

— Only the most confident bounding box for each object remains.

Inference: Real-Time Detection (Fig. 8):

— During inference, YOLOv8 processes an image.

— The final output includes bounding boxes and associated class labels.

In summary, YOLOv8 is a highly efficient algorithm that incorporates image classification, Anchor-Free object detection, and instance segmentation. Its detection component incorporates numerous state-of-the-art YOLO algorithms to achieve new levels of performance.

4.3 Training Process and Dataset Overview for Model Development

For model training, validation, and evaluation purposes, a tailored dataset was meticulously curated. This finalized dataset includes around 10,000 images depicting military trucks and tanks, sourced from diverse

repositories such as publicly available datasets, Google Images, and local archives, as detailed in Table 3.

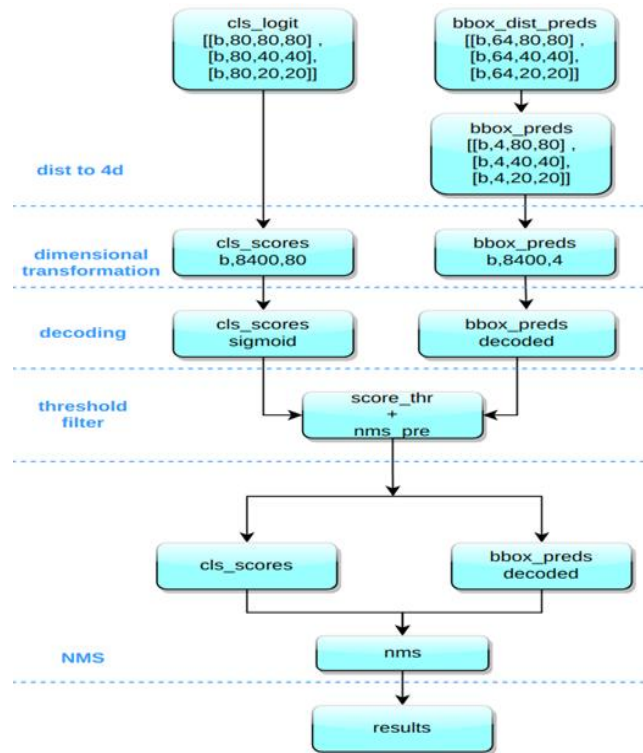


Fig. 8. Inference process implemented in MMYOLO

4.4 Training Process and Dataset Overview for Model Development

For model training, validation, and evaluation purposes, a tailored dataset was meticulously curated. This finalized dataset includes around 10,000 images depicting military trucks and tanks, sourced from diverse repositories such as publicly available datasets, Google Images, and local archives, as detailed in Table 3.

Tabl. 3 – Image Distribution in the Dataset

Dataset type	Size & Type	Source
APC, Tanks, Military Trucks	6772 images	Military Vehicles - Mendeley Data Collection [55]
Military Tanks	1078 images	Military Tanks – Kaggle [56]
Military Artillery, Tanks, MRL Systems, Trucks	1000 images	Moving and Stationary Target Acquisition and Recognition (MSTAR) Dataset [57]
Command posts & fire stations, Artillery, Stationary Military Equipment	20 videos & 2000 images	Indigenous Sources

The images encompass a wide range of scenarios involving military targets across diverse terrains, including day and night settings, forests, deserts, and urban environments, as depicted in Fig. 9. Despite this diversity, the dataset exhibits limitations. To overcome these constraints, several videos featuring various military targets in different conditions, such as size, shape, and color, were incorporated.



Fig. 9. Unified Dataset Training Process: Annotating Military Targets by Class (Tanks, APCs, Artillery, MRL Systems)

To prepare the dataset for training with YOLOv8, labels and bounding boxes were meticulously assigned to each image, with the coordinates of annotating boxes normalized between 0 and 1. This annotation process was facilitated by Roboflow [58], ensuring efficient data labeling in the desired format. Additionally, the dataset was thoughtfully divided into training, validation, and testing sets, maintaining balance across categories to prevent model bias. During the model training phase, fine-tuning of the YOLOv8 model, initially pre-trained on the COCO dataset, was conducted using transfer learning on the unified dataset. To clarify further, the training process involved utilizing pre-trained weights from a model that was trained on a large and diverse dataset. Transfer learning helps leverage the knowledge gained from the initial training, which is beneficial when working with limited labeled data.

Various augmentation processes, such as HSV (Hue, Saturation, Value) augmentation adjusting the color space parameters to make the model invariant to changes in lighting conditions., color spacing including transforming images between different color spaces, which helps the model handle variations in how colors are represented., mosaic combining multiple images into a single mosaic image, providing the model with diverse contexts and

increasing its ability to handle complex scenes, and image scaling including resizing images to different dimensions, allowing the model to learn from variations in object sizes., were applied. Hyperparameters that were fine-tuned, including SGD optimizer used for updating model parameters during training, 0.01 learning rate which is the step size which the model's weights are updated during training., 0.0005 weight decay as a regularization term that helps prevent overfitting by penalizing large weights, and finally, 600 epochs as the number of times the entire dataset was processed by the model during training, on batch size 32 as the number of training examples used in each iteration of a training epoch.

5 Results and Analysis

To evaluate the effectiveness of the proposed system, the trained model was tested across diverse environments, such as forests, mountains, and open fields. Results were obtained using a confidence threshold of 0.4, indicating that predictions with a confidence score below this threshold were excluded from consideration during model evaluation were likely filtered out. A confidence threshold acts as a cutoff, ensuring that only predictions with a certain level of confidence are considered. Initially, when applied to the handpicked test data from the unified dataset, we achieved an average of 0.922 mAP@0.5. This metric, ranging from 0 to 1, indicates the model's accuracy in correctly identifying and localizing military targets in images. An mAP@0.5 score of 0.922 suggests a high level of precision in the predictions, considering a 50% IoU overlap threshold. In essence, the model exhibits robustness and effectiveness in its ability to recognize military targets with a notable degree of accuracy.

Furthermore, the displayed confusion matrix in Fig. 10 indicates that the model can label most of the data quite accurately and with high precision when it comes to depicting classes of Military Targets, it does not overlap them with each other which ensures proper and accurate target recognition.

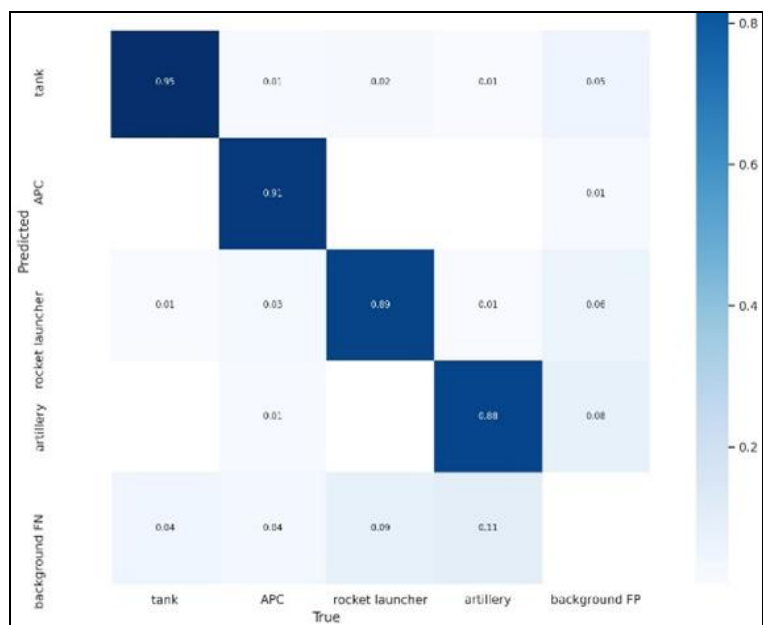


Fig. 10. Confusion Matrix of the model

Additionally, Fig. 11 shows that the model can accurately confirm the presence of a Military Target in the image.

The obtained value and prediction results depicted in the performance graphs in Fig. 12 demonstrate a significant potential to use the YOLOv8 algorithm for

real-time detection of Military Targets during drone surveillance.

The confidence value of the Military Targets recognized by the model is notably high. Furthermore, the performance of the proposed model applied to validation data can be observed in Fig. 8.



Fig. 11. Predicted Output Examples

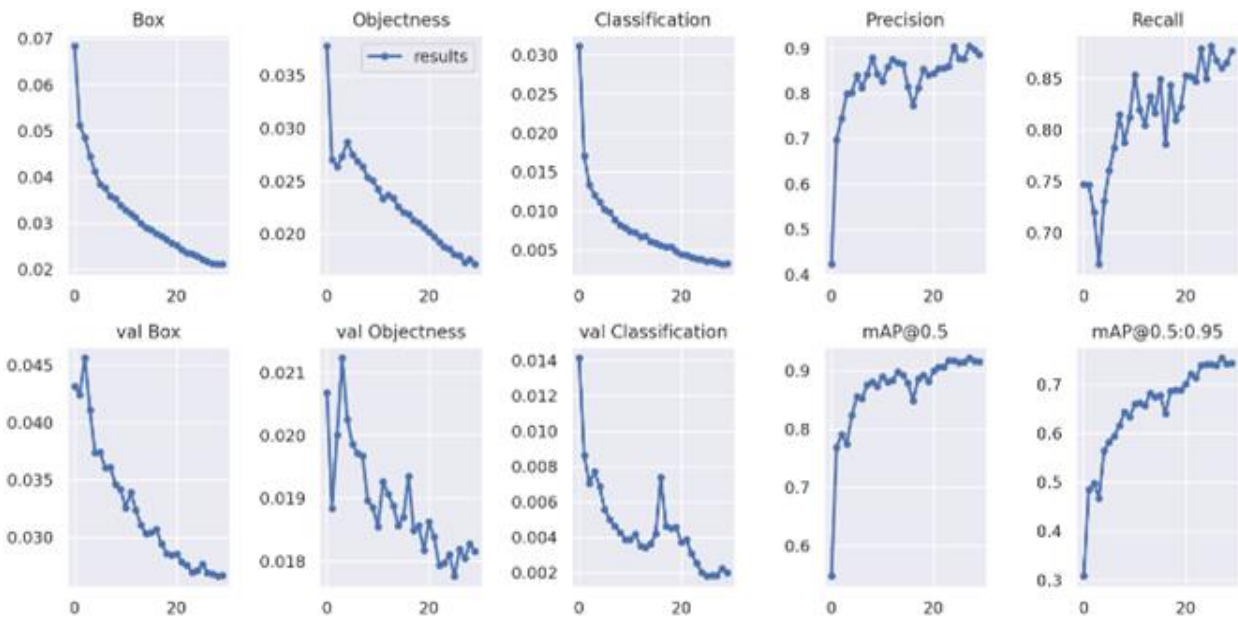


Fig. 12. Graphs illustrating Model Training and Validation Results

Box Accuracy refers to how accurately the model is able to predict the bounding boxes around objects in the images. High box accuracy shown in Fig. 12 means that the predicted bounding boxes closely match the actual positions of the objects in the test data. Objectness measures how well the model distinguishes between object and non-object regions in an image. A high objectness score as shown in Fig. 12 indicates that the model is effective at identifying regions containing relevant objects. Classification Accuracy assesses how accurately the model classifies the detected objects. As shown in Fig. 12, high classification accuracy means that the model correctly assigns the appropriate class labels to the recognized objects, such as different types of military targets. Precision as the ratio of true positive predictions to the total number of positive predictions made by the model. A high precision as shown in

Fig. 12 indicates that the model is making accurate positive predictions, minimizing false positives. Recall or sensitivity, is the ratio of true positive predictions to the total number of actual positive instances in the dataset. High recall as shown in Fig. 12 means that the model is adept at capturing a significant portion of the positive instances in the data, minimizing false negatives.

Having all these metrics at very high levels collectively suggests that the model is well-optimized, demonstrating strong performance in accurately detecting, localizing, and classifying military targets.

The model is lightweight and sufficiently fast for deployment. It can be integrated with built-in capturing tools, such as Surveillance Cameras mounted on COTS drones, to predict recognition in almost real-time with minimal delays.

In contrast to this, the results may vary depending on the image quality, therefore, another challenge to get the desired results from this model is to secure a constant good image quality feed. In the operational environment this can be a challenging task, and depends on multiple factors, but mostly weather conditions. The first step in mitigating those issues would be manually fine-tuning YOLOv8 on datasets specifically captured in rainy or foggy scenarios to improve its robustness. However, that focuses on the model itself rather than image feed quality, and it overall may bring better detection results, it will not eliminate the problem completely. Furthermore, image feed quality needs to be filtered properly before getting to the model, so the model can perform at its highest capacity and achieve the results which are presented in this study. To effectively mitigate weather conditions impact on the image feed several algorithms can be employed, depending on the severity and the impact of negative factors. First and foremost, is image dehazing. According to recent studies, Dehazing algorithms can be utilized to mitigate the impacts of fog and haze on the image feed up to certain heights and those algorithms include; Dark Channel Prior (DCP), Color Attenuation Prior (CAP), and Fast Guided Filter Dehazing [59].

Alongside fog and haze, rain streaks can significantly degrade image quality, therefore, motion compensation, temporal filtering, and deep learning-based rain removal networks can be utilized to mitigate its effects to a certain extent

Additionally, other image processing algorithms and tools may be utilized to improve the overall image quality. Contrast enhancement techniques like histogram equalization or adaptive contrast stretching can make objects more distinguishable for the pre-trained YOLOv8 detection model. Gamma Correction can also be utilized to improve brightness and overall image quality, furthermore it can compensate for variations caused by weather conditions. Alongside with that, denoising filters (e.g., Gaussian, median) can be used to reduce noise caused by raindrops or fog. White Balance Correction can be used to adjust color balance to compensate for color shifts due to weather conditions, while proper white balance ensures accurate color representation [60].

With successful employment of the beforementioned algorithms and solutions, camera feed from the drone to the YOLOv8 model can be brought to a decent enough level so that even in negative operating conditions the model can perform on its highest levels.

Conclusion

The evolution of military capabilities has been closely linked to the relentless integration of advanced technologies into modern combat operations. Advances in science and technology have ushered in a new era, significantly enhancing the efficiency and effectiveness of military units. This transformation is particularly evident in the utilization of unmanned aerial vehicles (UAVs) equipped with artificial intelligence (AI) and advanced technical instruments within the C5IRS

system. The integration of UAVs into the C5IRS system represents a quantum leap in increasing the operational efficiency of military units engaged in combat. The versatility of these vehicles, equipped with high-resolution cameras, infrared and thermal cameras, microphones, various sensors, guided and non-guided missiles, and other accessories, provides a comprehensive solution for real-time battlefield information. This capability not only facilitates the destruction of targets without risking human lives but also ensures the swift transmission of critical data to central command centers.

The wide range of applications for UAVs extends beyond combat scenarios, encompassing humanitarian efforts such as the delivery of medical supplies and essential equipment to operational areas. Despite the undeniable advantages, it is crucial to acknowledge the potential shortcomings of the proposed model. The risk of misidentifying non-military objects as actual targets, particularly at long range and with poor imagery is a significant challenge. Addressing these issues requires sustained efforts, including longer training times, expansive datasets, and improvements in accuracy to minimize false positives.

Looking ahead, the integration of modern C6ISR systems with existing combat frameworks emerges as a focal point for future research. This seamless integration holds the promise of optimizing military operations across diverse scenarios, aligning with the dynamic nature of modern warfare. The imperative for ongoing refinement and adaptation of these systems underscores the commitment to maintaining strategic advantages on the contemporary battlefield. The comprehensive literature review provides insights into the diverse applications of UAVs in military operations. Researchers have explored the possibilities of using drones in urban environments, protecting and monitoring of land security zones, effectiveness in modern armed conflicts, combat use in counter-terrorism operations, and integration with C5IRS systems for military needs. These studies collectively contribute to the growing body of knowledge on the subject, addressing challenges and proposing innovative solutions.

The following sections discuss the definition and classification of unmanned aircraft, emphasizing the need for a standardized classification system. The EUROVS classification based on various factors such as purpose, flight altitude, duration, speed, and dimensions provide a foundation for understanding the categorization of UAVs. Control and management models further delineate categories based on take-off weight, maximum range and take-off weight.

The discussion focuses on combat operations, underscoring the foundational role of land-based military operations in overall military endeavors. Operations, characterized by tactical actions with a common purpose, unfold across diverse physical environments and vary in scale and duration. The complex operational environment necessitates leaders with a nuanced understanding of both the science and art of operations. While the science includes factors like

combat power ratios and weapons ranges, the art encompasses intangible human factors that often determine the outcomes of operations.

The focus then shifts to combat operations, which are defined as intense, direct engagements with enemy forces aimed at achieving specific tactical goals. The elements of direct engagement, tactical objectives, fire and maneuver tactics, close combat, and the strategic context define the nature of combat operations. The discussion acknowledges the inherent risk and uncertainty in combat situations, emphasizing the importance of adaptability and the combined use of arms for success. The subsequent section explores the application of drones with AI in military combat operations. Drones, with their diverse capabilities, significantly enhance military capabilities by providing crucial information about enemy movements and targets. The integration of artificial intelligence further enhances the capabilities of drones, enabling autonomous operation and rapid analysis of real-time data. The advantages range from improved situational awareness to the rapid identification of threats and targets. The discussion highlights the transformative impact of drone technology on modern warfare.

The literature review covers a wide range of studies analyzing the use of drones in various military contexts, including urban operations, land security, counter-terrorism, reconnaissance, and cooperation with C5IRS systems. These studies contribute valuable insights into the practical applications and challenges associated with drone technology in military settings. The subsequent sections delve into the YOLOv8 method for enhanced military drone surveillance with AI software capabilities. The methodology involves training a YOLOv8 model on a customized dataset of military targets, including trucks, tanks, artillery, and other vital objects. The results indicate a high potential for real-time detection of military targets during drone surveillance, with a focus on lightweight and efficient deployment. The developed and validated algorithm may also be successfully used in COTS drones, increasing their functionality.

In conclusion, the integration of advanced technologies, particularly UAVs with AI capabilities, represents a paradigm shift in modern military operations. The versatility and efficiency of these systems offer strategic advantages while addressing complex challenges. The ongoing research and development efforts, coupled with a nuanced understanding of operational dynamics, enable the military to navigate the evolving landscape of contemporary warfare. As technology continues to advance, the commitment to innovation and adaptation remains paramount in ensuring military readiness and effectiveness on the global stage.

In conclusion, despite the achievements in integrating drones with artificial intelligence into military operations, future research can focus on optimizing performance, minimizing target identification errors, and enhancing drone autonomy. Continued research should aim to further integrate modern C6ISR systems with existing military frameworks, exploring how this integration can further optimize operations in diverse scenarios. Keeping abreast of technological advancements and adapting systems to new challenges on the battlefield remains crucial for maintaining military effectiveness and strategic advantage in the future.

As we look ahead, ongoing investigations may delve into refining the YOLOv8 algorithm for even more accurate real-time detection of military targets during drone surveillance. Addressing any remaining limitations, such as potential misidentifications and errors, particularly over longer distances, will be crucial for advancing the capabilities of the proposed model. Additionally, future studies could explore the deployment of advanced AI models in conjunction with drones to enhance decision-making processes and situational awareness in dynamic military environments.

The evolution of drone technology and artificial intelligence continues to reshape modern warfare, and further research will play a pivotal role in harnessing these advancements for more effective and precise military operations.

REFERENCES

1. Petrovski, A., Radovanović, M., Behlić, A. and Ackovska, S. Advantages of Implementation of C6ISR in Low Budget Armies. In: *Int. Scientific Conference GEOBALCANICA 2023*, 2023, pp. 47-60. <https://doi.org/10.18509/GBP23047p>
2. Bayramov, A.A., Hashimov, E.G., and Amanov, R.R. Identification of invisible objects using GIS technology. In: *Proceedings of the Azerbaijan Geographical Society Geography and Natural Resources*, 2016, pp. 124-126.
3. Bares, M. Interoperability Modeling of the C4ISR Systems. In: *RTO SCI Symposium on "System Concepts for Integrated Air Defense of Multinational Mobile Crisis Reaction Forces"*, Valencia, Spain, 2021, pp. 16-1 – 16-16.
4. Petrovski, A. & Toshevski, M. GIS in Army: Application of GIS in Geo-Reconnaissance and C4IS in Army Purposes. In: *2nd International Scientific Conference GEOBALCANICA*, 2016, pp. 153-160.
5. Jović, Ž. Combat engagement of drones in USA counterterrorist operations. *Bezbednost*, 2016, 58(3), pp. 171-190.
6. Petrovski, A., Taneski, N. and Bogatinov, D. Geography in Geospatial Intelligence - C4IRS and Cyber Security. In: *5th Int. Scientific conference Safety and crisis management – Theory and practise Safety for the future – SecMan 2019*, pp. 64-72.
7. Milić, A., Randelović, A. and Radovanović, M. Use of drones in operations in the urban environment. *5th International Scientific conference Safety and crisis management – Theory and practise Safety for the future – SecMan 2019*, 2019, pp. 124-130.
8. Adamski, M. Effectiveness analysis of UCAV used in modern military conflicts. *Aviation*, 2020, 24(2), 66-71. <https://doi.org/10.3846/aviation.2020.12144>
9. Žnidaršič, V., Radovanović, M. and Stevanović, D. Modeling the organisational implementation of a drone and counter-drone operator into the Serbian Armed Forces rifle section. *Vojno delo*, 2020, 72(3), pp. 84-109. <https://doi.org/10.5937/vojdela2003084Z>

10. Petrovski, A. And Radovanović, M. (2021). Application of detection reconnaissance technologies use by drones in collaboration with C4IRS for military interested. *Contemporary Macedonian Defence*, 2021, 21(40), pp. 117-126.
11. Hashimov, E.G. and Huseynov, B.S. Some aspects of the combat capabilities and application of modern UAVs. *National security and military sciences*, 2021, 3(7), pp. 14-24.
12. Radovanović, M., Petrovski, A., Žindrašić, V., and Randelović, A. Application of the fuzzy AHP -VIKOR hybrid model in the selection of an unmanned aircraft for the needs of tactical units of the armed forces. *Scientific Technical Review*, 2021, 71(2), 26-35. <https://doi.org/10.5937/str2102026R>
13. Ilić, D. and Tomašević, V. The impact of the Nagorno-Karabakh conflict in 2020 on the perception of combat drones. *Serbian Journal of Engineering Management*, 2021, 6(1), pp. 9-21. <https://doi.org/10.5937/SJEM21010091>
14. Radovanović, M., Samopjan, M. and Petrovski, A. Possibility of Implementation of Drons in Mortar Units in Order to Increase the Efficiency of Fire Support Units. In: *24. Međunarodna DQM konferencija Upravljanje kvalitetom i pouzdanošću ICDQM -2021*, 2021, pp. 307–315.
15. Ciolponea, C.A. The Integration of Unmanned Aircraft System (UAS) in Current Combat Operations. *Land Forces Academy Review*, 2022, 27(4), pp. 333–47.
16. Szulc, T. Possibilities of using unmanned combat assets in tactical operations in the mountains. *Scientific Journal of the Military University of Land Forces*, 2023, 208(2), pp. 112-127. <https://doi.org/10.5604/01.3001.0053.7270>
17. Zaher, R. A. H. Drones and their Role in the Evolution of Generations of War. *The International and Political Journal*, 2023, 56(Sep. 2023), pp. 69–86. <https://doi.org/10.31272/ipj.i56.246>
18. Hashimov, E.G., Bayramov, A.A., and Khalilov, B.M. Orthophotomap making of terrain for detection military targets. *National security and military sciences*, 2016, 4(2), pp. 14-20.
19. Li, S., Huang, H., Meng, X., Wang, M., Li, Y. and Xie L. A Glove-Wearing Detection Algorithm Based on Improved YOLOv8. *Sensors*, 2023, 23(24):9906. <https://doi.org/10.3390/s23249906>
20. Roy A. M., Bhaduri, J. DenseSPH-YOLOv5: An automated damage detection model based on DenseNet and Swin-Transformer prediction head-enabled YOLOv5 with attention mechanism. *Advanced Engineering Informatics*, 2023, 56(12), 102007, <https://doi.org/10.1016/j.aei.2023.102007>
21. Duan, J., Bai, T., Lv, B., Zong, H. and Fu, H. (2024). Lightweight Detection of Fasteners with YOLOv8 Combined with ShuffleNetV2. In: *Proceedings of the 6th Int. Conf. on Electrical Eng. and Information Techn. for Rail Transp. (EITRT 2023). Lecture Notes in Electrical Engineering*, 2024, 1136, pp. 480-489, https://doi.org/10.1007/978-981-99-9315-4_46
22. Wang, H., Han, X., Song, X., Su J., Li Y., Zheng W. and Wu X. Research on automatic pavement crack identification Based on improved YOLOv8. *Int. J. on Int. Design and Manufacturing*, 2024, 18(2), <https://doi.org/10.1007/s12008-024-01769-3>
23. Sun, D., Zhang, K., Zhong, H., Xie J., Xue X., Yan, M., Wu, W. and Li, J. Efficient Tobacco Pest Detection in Complex Environments Using an Enhanced YOLOv8 Model. *Agriculture*, 2024, 14(3):353. <https://doi.org/10.3390/agriculture14030353>
24. Wu, Y., Liao, T., Chen, F., Zeng, H., Ouyang, S. and Guan, J. Overhead Power Line Damage Detection: An Innovative Approach Using Enhanced YOLOv8. *Electronics*, 2024, 13(4):739. <https://doi.org/10.3390/electronics13040739>
25. Wen, G., Li, M., Luo, Y., Shi, C. & Tan, Y. *The improved YOLOv8 algorithm based on EMSPConv and SPE-head modules*, 2024.
26. Niu, S., Xu, X., Liang, A., Yun, Y., Li, L., Hao, F., Bai, J. and Ma, D. Research on a Lightweight Method for Maize Seed Quality Detection Based on Improved YOLOv8. *IEEE Access*, 2024, 12, pp. 32927-32937, <https://doi.org/10.1109/ACCESS.2024.3365559>
27. Fazekas, F. Application of Artificial Intelligence in Military Operations Planning. *AARMS – Academic and Applied Research in Military and Public Management Science*, 2022, 21(2), pp. 41–54. <https://doi.org/10.32565/aarms.2022.2.3>
28. Kania, E. B. (2019). Chinese Military Innovation in the AI Revolution. *The RUSI Journal*, 2019, 164(5–6), pp. 26–34. <https://doi.org/10.1080/03071847.2019.1693803>
29. Forrest E.M, Boudreaux, B., Lohn, J.A., Ashby, M., Curriden, C., Klima, K. and Grossman, D. *Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World*. Santa Monica, CA: RAND Corporation, 2020. https://www.rand.org/pubs/research_reports/RR3139-1.html
30. Walsh, G., Andersen, N., Stoianov, N. and Jänicke, S. A Survey of Geospatial-Temporal Visualizations for Military Operations. In: *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023) – IVAPP*, 2023, pp. 115-129. <https://doi.org/10.5220/0011902500003417>
31. Bhagat R. S. Artificial Intelligence and Data Applications In Military Operations. *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)*, 2023, 3(1), pp. 264-271, <https://doi.org/10.48175/IJARSCT-12040>
32. De Castro, B.A., Pochmann, P.G.C. and Neves, E.B. (2024). Artificial Intelligence Applications in Military Logistics Operations. In: *Developments and Advances in Defense and Security. MICRADS 2023. Smart Innovation, Systems and Technologies*, 2024, 380, pp. 89-100 https://doi.org/10.1007/978-981-99-8894-5_8
33. Lee, M., Choi, M., Yang, T., Kim, J., Kim, J., Kwon, O. and Cho, N. A Study on the Advancement of Intelligent Military Drones: Focusing on Reconnaissance Operations. *IEEE Access*, 2024, 12, pp. 55964-55975, <https://doi.org/10.1109/ACCESS.2024.3390035>
34. Shah, I. A., Jhanjhi, N. Z. and Brohi, S. N. Use of AI-Based Drones in Smart Cities. In: *Cybersecurity Issues and Challenges in the Drone Industry*, IGI Global, 2024, pp. 362-380, <https://doi.org/10.4018/979-8-3693-0774-8.ch015>
35. POMORTSEVA O., KOBZAN, S., and SHTERNDOK E. Use of geo-information technologies when conducting combat operations in modern conditions. *Municipal Economy of Cities*, 2023, 1(175), pp. 69–73. <https://doi.org/10.33042/2522-1809-2023-1-175-69-73>
36. Agbeyangi, A., Odiete, J. and Olorunlomeye, A. Review on UAVs used for Aerial Surveillance. *Journal of Multidisciplinary Engineering Science and Technology*, 2016, 3(10), pp. 5713-5719.
37. *Unmanned Aircraft System Airspace Integration Plan*, Department of Defense, 2011.
38. *Regulation No 1592/2002 of the European Parliament and of the Council on Common Rules in the Field of Civil Aviation and Establishing a European Aviation Safety Agency*, 2002.

39. Radovanović, M., Alexandar Petrovski, Vinko Žnidaršič and Aner Behlić. The C5ISR System Integrated With Unmanned Aircraft in the Large-Scale Combat Operations. *Vojenské rozhledy*. 2023, 32 (2), 098- 118. [10.3849/2336-2995.32.2023.02.098-118](https://doi.org/10.3849/2336-2995.32.2023.02.098-118)
40. *ALLIED JOINT DOCTRINE*, NATO Standardization Office (NSO), NATO Standard AJP-01 2017, point 2.20, 2-12
41. *Categorization and classification of civil unmanned aircraft systems*, ISO 21895:2020, 2020.
42. *Minimum Training Requirements for Unmanned Aircraft Systems (UAS) Operators and Pilots*, NATO STANDARD ATP-3.3.8.1, (2019), Edition B Version 1, NATO, 2-2 (16),
43. Fellner, A. and Fellner, R. Navigation preparation of unmanned aircraft systems for the operational activities of Volunteer Fire Services – selected examples [Nawigatorskie przygotowanie systemów bezzałogowych statków powietrznych do operacyjnych działań jednostek ochotniczych straży pożarnych–wybrane przykłady]. *Acta Iuridica Resoviensia*, 2022, 38(120), pp. 83 - 85. <https://dx.doi.org/10.15584/actaires.2022.3.5>
44. Fellner, R. and Burnos, A. Unmanned aerial vehicles in operational activities—example of selected volunteer fire brigade units. *Scientific Reports of Fire University*, 2023, 86(null), pp. 175-190. <https://doi.org/10.5604/01.3001.0053.7153>
45. Thompson, K. *How the Drone War in Ukraine Is Transforming Conflict*, 2024, Retrieved January 30, 2024, from <https://www.cfr.org/article/how-drone-war-ukraine-transforming-conflict>.
46. REDMON, J., DIVVALA, S., GIRSHICK, R. and FARHADI, A. You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779-788
47. Everingham, M., Eslami, S.A., Van Gool, L., Williams, C.K., Winn, J. and Zisserman, A. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 2015, 111(1), pp. 98–136.
48. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C.L. Microsoft coco: Common objects in context. In: *Proceedings of the 13th European Conference on Computer Cision (ECCV 2014)*, 2014, pp. 740–755.
49. Wang, C.Y., Mark Liao, H.Y., Wu, Y.H., Chen, P.Y., Hsieh, J.W. and Yeh, I.H. CSPNet: A new backbone that can enhance learning capability of CNN. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2020)*, Washington, DC, USA, 2020, pp. 390–391.
50. Wang, K., Liew, J.H., Zou, Y., Zhou, D. and Feng, J. (2019). Panet: Few-shot image semantic segmentation with prototype alignment. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV 2019)*, Seoul, Korea, 2019, pp. 9197–9206.
51. Redmon, J. and Farhadi, A. *Yolov3: An incremental improvement*, 2018, <https://doi.org/10.48550/arXiv.1804.02767>
52. Xu, R., Lin, H., Lu, K., Cao, L. and Liu, Y. A Forest Fire Detection System Based on Ensemble Learning. *Forests*, 2021, 12(2), 217:1-17, <https://doi.org/10.3390/f12020217>
53. Redmon, J. and Farhadi, A. Yolo9000: better, faster, stronger. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 6517-6525. <https://doi.org/10.1109/CVPR.2017.690>
54. Jocher, G., Munawar, R. and Exel, A. *YOLOv8 Official Documentation*, Retrieved Mart 25, 2024, from <https://docs.ultralytics.com>
55. Gupta, P., Pareek, B., Singal, G. and Rao, D. V. Military and Civilian Vehicles Classification. *Mendeley Data*, 2021, V1, <https://doi.org/10.17632/njdjkbxdpn.1>
56. Jana, A. *Military Tanks Dataset (Images)*, 2021, Retrieved Mart 22, 2024, from <https://www.kaggle.com/datasets/antorecepjana/military-tanks-dataset-images>
57. MSTAR overview, 2023, Retrieved Mart 25, 2024, from <https://www.sdms.afrl.af.mil/index.php?collection=mstar>
58. Roboflow, 2024, Retrieved Mart 15, 2024, <https://universe.roboflow.com/>
59. Wang, W. & Yuan, X. Recent advances in image dehazing. *IEEE/CAA Journal of Automatica Sinica*, 2017, 4(3), pp. 410-436. <https://doi.org/10.1109/JAS.2017.7510532>.
60. Yang, W., Tan, R., Wang, S., Kot, A. and Liu, J. Learning to Remove Rain in Video with Self-Supervision, *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2024, 46(03), pp. 1378-1396, <https://doi.org/1109/TPAMI.2022.3186629>

Received (Надійшла) 05.06.2024

Accepted for publication (Прийнята до друку) 31.07.2024

Вивільнення автономних сил:

інтеграція безпілотних літальних апаратів із штучним інтелектом у сучасну військову стратегію

М. Радовановіч, А. Петровські, А. Вехлік, М. Зід Хаарі, Е. Г. Гашимов, Р. Фельнер, А. О. Агбєянгі

Анотація. Вплив штучного інтелекту (ШІ) на міжнародну безпеку на сьогодні є безсумнівним, оскільки тепер машини здатні виконувати завдання, які традиційно покладені на людський інтелект. Ця зміна породжує безліч викликів у міжнародній безпеці, впливаючи як на звичайні військові можливості, так і на гібридні загрози. Водночас ШІ відкриває нові можливості для вирішення цих викликів, впливаючи на ключові аспекти колективної оборони, кооперативних систем безпеки та управління кризами. Враховуючи його глибокі наслідки для процвітання та безпеки, ефективне управління ШІ вимагає спільних зусиль. Обсяг перспектив і небезпек, пов'язаних зі штучним інтелектом, величезний, що вимагає колективних дій для пом'якшення ризиків безпеці та використання його потенціалу для реструктуризації операційних процесів, підтримки місій і оптимізації операцій. Ця стаття в основному зосереджена на представленні дронів, оснащених штучним інтелектом і можливостями автономного навчання, досліджуючи їх застосування у військових умовах. У статті розглядається потенціал незалежного використання безпілотних літальних апаратів із штучним інтелектом як у бойових, так і в небойових армійських операціях. Завдяки використанню ГІС, C5IRS (командування, управління, комп'ютери, зв'язок, кіберзахист, розвідка, спостереження та розвідка) і штучного інтелекту ці дрони забезпечують значну перевагу на полі бою, працюючи автономно та адаптуючись до динамічних наземних ситуацій. У бойовому просторі, де поєднання людських, інформаційних і фізичних компонентів має вирішальне значення для стратегічної переваги, оперативна сумісність стає життєво важливим фактором.

Ключові слова: БПЛА, YOLOv8, дрон, ГІС, штучний інтелект, безпека, C5IRS.

Г. В. Худов, І. А. Хижняк, І. Ю. Грідасов, У. Р. Збежховська, І. Ю. Юзова,
Ю. С. Соломоненко, Т. М. Калімулін

Харківський національний університет Повітряних Сил імені Івана Кожедуба, Харків, Україна

АНАЛІЗ ТАКТИКО-ТЕХНІЧНИХ ХАРАКТЕРИСТИК ТА ТАКТИКИ ЗАСТОСУВАННЯ ІСНУЮЧИХ УДАРНИХ FPV-ДРОНІВ

Вступ. Предметом вивчення в статті є тактико-технічні характеристики та тактика застосування ударних FPV-дронів. **Метою** є аналіз тактико-технічних характеристик існуючих ударних FPV-дронів та їх тактики застосування. **Завдання:** ідентифікація та класифікація безпілотних літальних апаратів, аналіз основних технічних параметрів ударних FPV-дронів, дослідження методів та стратегій застосування ударних FPV-дронів у бойових діях, дослідження вразливостей ударних FPV-дронів до засобів радіоелектронної боротьби та методів захисту від них, розробка рекомендацій для покращення технічних характеристик та бойових можливостей ударних FPV-дронів. Використовуваними **методами** є: порівняльний аналіз, методи кластеризації даних, аналітичні та емпіричні методи аналізу використання ударних FPV-дронів у військових конфліктах, зокрема у російсько-українській війні. Отримані такі **результати**. Проведений структурований аналіз тактико-технічних характеристик та тактики застосування ударних FPV-дронів, наведені їх переваги та недоліки, сформувані пропозиції щодо шляхів їх удосконалення. Практична значимість отриманих результатів в майбутньому може стати основою для удосконалення існуючих ударних FPV-дронів, розробки нових моделей, а також для підвищення ефективності їх використання. **Висновки.** Аналіз тактико-технічних характеристик та тактики застосування ударних FPV-дронів показав, що такі безпілотні літальні апарати суттєво впливають на хід сучасних військових операцій, надаючи нові можливості для нанесення ефективних точкових ударів по техніці та особовому складу противника, проведення розвідки, коригування вогню та виконання функцій транспортування. Ударні FPV-дрони демонструють високу ефективність завдяки маневреності, відносно низькій вартості та можливості модифікації під конкретні бойові задачі. Водночас, їх вразливість до засобів радіоелектронної боротьби вимагає подальшого удосконалення систем зв'язку та управління. Дослідження були проведені за рахунок грантової підтримки Національного фонду досліджень України в рамках конкурсу “Наука для зміцнення обороноздатності України”, назва проєкту “Інформаційна технологія автоматизованого сегментування зображень об'єктів в системах націлювання ударних FPV-дронів на основі алгоритмів ройового інтелекту”, реєстраційний номер 2023.04/0153.

Ключові слова: безпілотний літальний апарат, FPV-дрон, оптико-електронне обладнання, тактико-технічні характеристики, тактика застосування, сегментування зображень.

Вступ

Постановка проблеми у загальному вигляді.

Досвід ведення бойових дій у світовій практиці останніх років показав, що використання безпілотних літальних апаратів (БпЛА) або дронів відіграє важливу роль, змінюючи традиційні підходи до ведення війн. БпЛА надають можливість збройним угрупованням військ (сил) проводити розвідку в режимі реального часу, коригувати артилерійський вогонь, наносити вогневі удари по об'єктах противника, виконувати завдання транспортування тощо. Також використання БпЛА дозволяє значно зменшити ризики для особового складу своїх сил, підвищуючи ефективність військових операцій і знижуючи бойові втрати [1].

Перше масове застосування БпЛА у світі для завоювання переваги на полі бою відбулось у 2020 році під час військового конфлікту між Азербайджаном і Вірменією в Нагірному Карабасі. Збройні Сили (ЗС) Азербайджану в результаті проведення операції “Залізний кулак” встановили повний контроль над Нагірним Карабахом. Азербайджан мав на озброєнні сучасні БпЛА турецького та ізраїльського виробництва, такі як Bayraktar TB2, Heron TP, Hermes 4507, Sky Striker, і Harop, які використовувались для проведення ударів по вірменським системам протиповітряної оборони (ППО), зокрема ЗРК “Оса” та “Стріла-10”. Дані засоби ППО, хоча і були призначені для боротьби з літаками та вертольота-

ми, проте, виявились недостатньо ефективними проти сучасних БпЛА. Масоване застосування Азербайджаном дронів призвело до значних втрат вірменських засобів ППО вже в перші дні конфлікту, що суттєво послабило оборонні можливості Вірменії [2].

Саме знищення вірменських систем ППО забезпечило Азербайджану перевагу в повітрі, що дозволило безперешкодно атакувати вірменські наземні військові цілі, завдаючи їм значних втрат ще до початку наземної операції. Зазначене сприяло успішному наступу азербайджанських військ (сил), що призвело до значних оперативних успіхів. Збройні Сили Вірменії виявились неготовими до масового використання БпЛА противником, і після втрати основних сил ППО не змогли швидко поповнити втрати новими системами, що стало однією з ключових причин поразки у війні.

Таким чином, війна за Нагірний Карабах наочно продемонструвала ефективність застосування БпЛА, визначивши нову стратегію і тактику збройного протистояння на театрах бойових дій. Вона сформувала концепцію “війни майбутнього” – коли в бій ідуть машини, якими здійснюють управління на відстані [3].

Повномасштабне вторгнення ЗС російської федерації на територію України також підтвердило, що сучасні БпЛА відіграють важливу роль у завоюванні переваги на полі бою. З їх допомогою ЗС України знищили велику кількість озброєння і військової

техніки та особового складу противника, як на окупованих територіях, так й на території російської федерації. Використання БпЛА здатне забезпечити наступальний потенціал проти більшого, технологічно спроможного супротивника [4].

Характерною особливістю, що значно відрізняє війну в Україні від війни в Нагірному Карабасі є те, що ЗС України почали активно використовувати FPV (first person view) - дрони. На відміну від великих БпЛА, які використовувалися Азербайджаном, таких як Bayraktar TB2, FPV-дрони здатні забезпечити більшу маневреність і точність при значно менших витратах. Відповідно до звіту американського Інституту вивчення війни [5] саме FPV-дронами українські сили частково компенсують недостатність артилерійських боєприпасів на бої бою, використовуючи їх для відбиття атак російської бронетехніки і піхоти.

FPV-дрони надають можливість оператору керувати літальним апаратом у реальному часі в режимі "від першої особи", завдяки встановленим на дроні камері та передавачу відеосигналу. Це дозволяє оператору бачити саме те, що бачить дрон, і здійснювати точне управління, навіть у складних умовах міського бою або пересіченої місцевості. Одна із основних переваг FPV-дронів полягає у їхній маневреності та швидкості, що робить їх ідеальними для виконання різноманітних тактичних завдань. FPV-дрони можуть виконувати як розвідувальні місії, так і удари по наземним цілям. Завдяки їх невеликим розмірам і здатності швидко змінювати висоту і напрямок польоту, такі дрони важко виявити та збити засобами ППО. Збройні Сили України активно використовують FPV-дрони для проведення розвідки та коригування артилерійського вогню. З їх допомогою українські військові можуть отримувати детальну інформацію про розташування ворожих сил і техніки, що дозволяє точніше планувати атаки та зменшувати кількість помилкових ударів. Окрім того, FPV-дрони можуть бути оснащені невеликими вибуховими пристроями для виконання точкових ударів по ворожих об'єктах. Ефективність застосування FPV-дронів також полягає в їхній відносно невеликій вартості. Вони значно дешевші у виробництві та обслуговуванні порівняно з традиційними БпЛА. Це дозволяє ЗС України швидко нарощувати кількість дронів у своєму арсеналі, що в свою чергу забезпечує безперервний моніторинг і атакуючий потенціал на полі бою.

Масовим використанням дронів разом із ракетами характеризувалась атака Ірану на Ізраїль 13 квітня 2024 року [6]. Тактика Ірану, спираючись на досвід, отриманий під час війни в Україні, полягала в тому, щоб подолати передову оборону Ізраїлю дешевими дронами та надати більш складним балістичним ракетами більш шанси вразити свої цілі. Хоча ця операція не була успішною, вона вказує на велику актуальність застосування БпЛА у стратегіях ведення бойових дій для досягнення своїх військових цілей та суттєво трансформує традиційні концепції ведення війни.

Таким чином, сучасні конфлікти показують, що використання FPV-дронів є важливим елементом

військової стратегії. Вони забезпечують значну перевагу на полі бою, підвищуючи ефективність військових операцій і мінімізуючи ризики для особового складу. FPV-дрони стають невід'ємною частиною концепції "війни майбутнього", де високі технології та дистанційне керування визначають хід бойових дій.

Аналіз останніх досліджень і публікацій. Враховуючи масове застосування FPV-дронів з метою завоювання переваги на полі бою, постає задача проведення аналізу їх тактико-технічних характеристик (ТТХ) та тактики застосування. Зазначене обумовлене необхідністю постійного удосконалення можливостей застосування FPV-дронів в умовах ведення бойових дій та дії засобів радіоелектронної боротьби (РЕБ) противника. Тому, автори у своїх дослідженнях зосереджуються на різних аспектах застосування та удосконалення FPV-дронів для застосування у військових операціях. На теперішній час, тематиці досліджень присвячені роботи [7]–[15].

Так, у роботі [7] автори розглядають розробку військового FPV-дрона з використанням Raspberry Pi та алгоритмів машинного навчання для розпізнавання об'єктів противника і забезпечення їх відстеження в режимі реального часу. Дрон оснащений камерою та мережею Jio-Fi для передачі відео в реальному часі, а також сучасним мережевим глушиком, що запобігає перехопленню сигналів каналів управління протидіючою стороною. Використання згорткових нейронних мереж та інших алгоритмів машинного навчання дозволяє дроні розпізнавати об'єкти та обличчя людей, що значно підвищує його ефективність у бойових умовах.

У [8] автори здійснюють огляд різних застосувань БпЛА, підкреслюючи їх зростаюче стратегічне значення та технологічні виклики. В дослідженні обговорюються різноманітні сфери застосування дронів, такі як військові операції, картографування, пошуково-рятувальні місії та розвідка. Автори наголошують на важливості модульної гнучкості та програмованості дронів, що дозволяє здійснювати їх швидку адаптацію до різноманітних завдань. Основна увага приділяється викликам автономії польоту, включаючи обмежену тривалість роботи батареї, маневреність, інтерфейс користувача та економічну ефективність.

Автори роботи [9] досліджують гіростабілізоване управління камерами FPV-дронів для підвищення їх ефективності в умовах ведення бойових дій. Використання гіроскопа для керування кутами нахилу камери дозволяє оператору отримувати більш стабільне зображення навіть під час руху дрона. Це покращує якість відео та збільшує ймовірність успішності виконання бойового завдання.

У [10] було розглянуто можливості автоматизації БпЛА з фіксованим крилом для їх автономного та точного виконання розвідувальних операцій. Завдяки введенню плану польоту в наземну систему управління (НСУ), БпЛА виконає поставлене завдання автономно для чого керуватиметься оперативним планом польоту, де активні маршрутні точки нанесені на НСУ. НСУ може бути прив'язана до

телеметрії та системи глобального позиціонування, щоб отримувати повну інформацію про місцезнаходження БпЛА.

Комплексний огляд сфер застосування дронів, включаючи військову, телемедицину, сільське господарство, будівництво та пошуково-рятувальні операції, наведений у роботі [11]. Автори підкреслюють важливість використання дронів для забезпечення безпеки та ефективності виконання завдань в різних сферах діяльності людини. Дослідження також акцентує увагу на технологічних досягненнях, що сприяють зростанню популярності дронів.

Систематизований огляд літератури щодо сонячних БпЛА, їх переваги та недоліки проведений авторами у [12]. Сонячні БпЛА використовують сонячну енергію для продовження тривалості польоту, що робить їх екологічно чистими та економічно вигідними. Дослідження також показує останні досягнення в галузі сонячних елементів, акумуляторів та електродвигунів, які є перспективними для використання під час розробки нових видів БпЛА.

У [13] було досліджено використання БпЛА у військових операціях. Автори показують, що застосування БпЛА дозволяє зберегти життя пілота, при цьому виконавши бойове завдання з високою ефективністю. Однак, дослідження наголошує на вразливості систем управління БпЛА та показує можливі шляхи їх подальшого удосконалення. Іншим можливим застосуванням БпЛА для військових цілей є виявлення за допомогою них мін та особового складу противника [14]. Автори пропонують розробити дрон з металодетектором для виявлення мін та камерою для моніторингу підозрілої присутності людей. Дослідження також підкреслює важливість використання GPS для відстеження місця розташування мін та передачі їх оператору.

У роботі [15] показані порівняльні характеристики ефективності роботи кодеків H.264 та H.265 для 4K FPV-дронів, акцентуючи увагу на компресії відео, енергоспоживанні та якості зображення. Автори для визначення якості стисненого відео використовують об'єктивні методи оцінки якості відео, такі як PSNR, SSIM, VMAF та BRISQUE. Результати дослідження показують, що H.265 забезпечує кращу якість відео при меншому споживанні енергії у порівнянні з H.264.

Однак, враховуючи стрімкий розвиток сучасних технологій машинного навчання та алгоритмів обробки зображень, якість відео, що передається з дронів, стає критично важливою для успішного виконання бойового завдання. Висока роздільна здатність і чіткість зображення дозволяють точно ідентифікувати об'єкти та оцінити їх стан для прийняття ефективних рішень у бойових умовах.

Інтеграція таких систем у FPV-дрони забезпечить здатність оперативного реагувати на зміни у визначеному районі, сприяючи підвищенню загальної ефективності операції. Слід також зазначити, що із збільшенням застосування на полі бою БпЛА, противник постійно розробляє нові шляхи протидії, використовуючи засоби РЕБ та радіоелектронної розвідки (РЕР). Тому, постає задача постійного удо-

сконалення ТТХ і тактики застосування БпЛА. Однак, проведений аналіз досліджень і публікацій показав, що на теперішній час є недостатня кількість комплексних досліджень ТТХ і тактики застосування FPV-дронів, в тому числі тих, де аналізується їх оптико-електронне обладнання.

Мета статті – аналіз тактико-технічних характеристик існуючих ударних FPV-дронів та їх тактики застосування.

Основна частина

1. Класифікація безпілотних літальних апаратів. Масове використання БпЛА у бойових діях між українськими та російськими силами кардинально змінило концепцію сучасної війни. БпЛА – це безпілотне повітряне судно, керування польотом якого і контроль за яким здійснюються дистанційно за допомогою пункту дистанційного пілотування, що розташований поза повітряним судном, або повітряне судно, що здійснює політ автономно за відповідною програмою [16].

На першому етапі проведення аналізу доцільно провести класифікацію БпЛА. Класифікація БпЛА дозволяє систематизувати різноманітні типи БпЛА за певними ознаками, що, у свою чергу, сприятиме кращому розумінню їхніх можливостей та обмежень. Зазначене важливо для розробки стратегії застосування БпЛА, вибору оптимальних моделей для конкретних завдань, а також для удосконалення існуючих і розробки нових технологій. Крім того, класифікація допомагає визначити стандарти та вимоги до технічних характеристик БпЛА, що є необхідним для забезпечення їхньої надійності та ефективності у бойових умовах.

Таким чином, необхідність класифікації БпЛА обумовлена потребою у систематизації знань про різноманітні їх типи, що забезпечує більш раціональне їх використання та сприятиме подальшому розвитку безпілотних технологій у військовій сфері.

Класифікація БпЛА за призначенням [16]:

- ударні;
- винищувальні;
- розвідувальні;
- корегування вогню та артилерії;
- ретранслятори радіосигналів;
- транспортні.

Класифікація БпЛА за типом ЛА [16]:

- літаковий тип;
- вертолітний тип;
- мультироторний.

Класифікація БпЛА за максимальною висотою польоту [1]:

- гранично маловисотні – до 200 метрів;
- маловисотні – від 200 метрів до 1000 метрів включно;
- середньовисотні – понад 1000 метрів до 4000 метрів включно;
- великовисотні – понад 4000 метрів до 12000 метрів включно;
- стратосферні – понад 12000 метрів.

Класифікація БпЛА за максимальною швидкістю польоту [16]:

- замалошвидкісні – до 100 км/год;
- малошвидкісні – від 100 км/год до 300 км/год включно;
- середньошвидкісні – понад 300 км/год до 600 км/год включно;
- високошвидкісні – понад 600 км/год до швидкості звуку;
- надзвукові – максимальна швидкість польоту перевищує швидкість звуку до 5 разів;
- гіперзвукові – максимальна швидкість польоту перевищує швидкість звуку вп'ятеро і більше.

Зведена класифікація БпЛА за основними ознаками наведена у табл. 1 [16].

Таблиця 1 – Зведена класифікація БпЛА

Клас	Рівень застосування	Бойовий радіус
I клас < 150 кг	мікро (тактичні) злітна маса < 2 кг	до 5 км
	міні (тактичні) злітна маса 2-15 кг	5-25 км
	малі (тактичні) злітна маса >15 кг	понад 25 км
II клас 150-600 кг	оперативно-тактичні	50-200 км
III клас > 600кг	оперативні	понад 200 км
	стратегічні	понад 200 км

Наведена класифікація створює основу для подальшого розвитку технологій БпЛА, що є важливим для забезпечення переваги у військових конфліктах. При наявності ТТХ будь-якого типу БпЛА у оператора не буде виникати труднощів щодо його класифікації за різними показниками та вибору тактики застосування в різних бойових ситуаціях. Далі у роботі наведений аналіз ТТХ існуючих ударних FPV-дронів.

2. Аналіз тактико-технічних характеристик існуючих ударних FPV-дронів. Перш ніж перейти до аналізу технічних характеристик ударних FPV-дронів необхідно зрозуміти, що це за тип БпЛА та з яких компонентів зазвичай він складається.

FPV-дрон – це БпЛА, який передає відео в реальному часі з камери, встановленої на борту в напрямку польоту. Оператор керує дроном за допомогою спеціальних окулярів, спостерігаючи за польотом так, ніби знаходиться всередині повітряного судна. Сама аббревіатура FPV походить від англійського терміну “First Person View”, що означає “вид від першої особи”.

Важливо мати повноцінне уявлення та розуміння з яких компонентів складаються ударні FPV-дрони. На практиці, для забезпечення повноцінного функціонування ударних FPV-дронів необхідна наявність наступних обов'язкових компонентів [17]:

- рами, яка є конструкційною основою дрону, що забезпечує монтаж інших компонентів;
- польотного контролера, що включає в себе гіроскоп, акселерометр, компас, барометр і сонар, які контролюють стабільність та навігацію дрону;
- апаратури радіокерування, яка здійснює формування і перетворення сигналів та включає в се-

бе вбудовану телеметрію для передачі команд від пілота до дрону;

- антени для забезпечення зв'язку між дроном та оператором;
- двигуна, що забезпечує рух дрону;
- повітряних гвинтів, які генерують підйомну силу для польоту;
- регуляторів обертів, які контролюють швидкість обертання двигунів;
- батареї та зарядного пристрою (джерело живлення дрону);
- джерела бортового живлення з індикатором заряду, яке контролює та відображає рівень заряду батареї.

Додаткові компоненти включають [17]:

- обладнання для відеоканалу, яке складається із відеокамери, відеокомутатора, підвісу для стабілізації та приймача відеосигналу, які забезпечують відеозйомку та передачу відеосигналу оператору на вивідний пристрій;
- приймачі GPS, які забезпечують визначення точних координат дрону, використовуючи принципи визначення координат, режими старту та технології позиціонування;
- адаптер Bluetooth, який забезпечує бездротовий зв'язок з іншими пристроями;
- засоби пошуку, які допомагають знаходити дрон у випадку втрати зв'язку або аварії;
- дисплеї та індикатори, необхідні для відображення інформації про стан дрону та його параметри польоту.

Слід зазначити, що повна комплектація FPV-дронів з усіма обов'язковими та додатковими компонентами значно збільшує їхню комерційну вартість та підвищує експлуатаційні витрати.

Розглянувши з яких компонентів складаються ударні FPV-дрони, можемо зробити висновки, що це є каркасні коптери (тобто конструктори), які складаються з різних елементів та оснащені закріпленою камерою. Цей тип дронів може бути зібраний будь-якою особою з кількох основних компонентів, що робить їх значно дешевшими порівняно з заводськими моделями, наприклад, такими як DJI Mavic 3. Завдяки своїй модульній конструкції, FPV-дрони можуть бути адаптовані під різні завдання, включаючи виконання функцій камікадзе, бомбардувальників, розвідувальних апаратів тощо.

Необхідно чітко розрізнити основні відмінності за призначенням між розвідувальними дронами, наприклад, такими як DJI Mavic 3, які використовуються зазвичай для проведення розвідки ближнього сектору на відстані до 10 км, пошуку цілей та корегування вогню, та ударними FPV-дронами. На відміну від розвідувальних БпЛА, ударні FPV-дрони здебільшого функціонують як камікадзе, несучи на собі корисне навантаження у вигляді боєприпасу. Хоча можливо оснастити такі дрони високоякісними камерами та використовувати їх для розвідувальних цілей, проте, така практика на теперішній час не є широко поширеною.

Оскільки ударний FPV-дрон є конструктором, то при достатніх зусиллях можливо укомплектувати

модель, яка буде дорожчою за умовний Mavic. Вартість FPV-дрона визначають внутрішні компоненти, процес його складання та кінцеві технічні характеристики. DJI Mavic є дроном заводської збірки, виготовленим з високоякісних матеріалів та укомплектованим пультом керування, різними сенсорами, датчиками, і найголовніше – високоякісною камерою, яка складає значну частину вартості дрона. На противагу цьому, ударний FPV-дрон може бути зібраний самостійно, придбавши основні компоненти через Інтернет. Залежно від завдань, які буде виконувати дрон, підбираються його складові частини. Наприклад, для дрона типу камікадзе (одноразовий дрон до рами якого закріплений боєприпас), зазвичай, не потрібна високоякісна камера, достатньо невеликої камери, яка дозволяє орієнтуватися на місцевості під час виконання польоту.

Проведений аналіз ТТХ ударних FPV-дронів показав, що переважна більшість дронів функціонують на цивільних частотах, а саме: 900 МГц, 1,2 ГГц, 1,3 ГГц, 2,4 ГГц, 5,8 ГГц, а також використовують технологію Wi-Fi. Слід зазначити, що сучасний ринок FPV-дронів переважно пропонує апаратуру, що функціонує на частотах 2,4 ГГц або 5,8 ГГц. Оскільки зазначені частоти призначені для цивільного використання, а не для вирішення військових задач, то вони є відомими та постійно піддаються радіоелектронному подавленню. Тому, оператор дрона повинен бути обережним, знати місця розташування засобів РЕБ противника, а також повітряні коридори, якими можна безпечно здійснювати політ. БпЛА військового призначення, наприклад, літакового типу, є більш стійкими до РЕБ завдяки використанню різних частот. У разі подавлення однієї частоти, оператор може продовжити роботу на іншій. Водночас, щодо цивільних FPV-дронів, постійно триває робота в напрямку підвищення стійкості до засобів РЕБ, шляхом удосконалення технічних характеристик. Під час виконання бойових завдань застосовуються активні антени, які підсилюють сигнал, а також робляться деякі прошивки самих дронів, які дозволяють працювати в різних діапазонах частот. Дані технологічні різноманітності дозволяють оператору обирати оптимальні параметри для конкретних умов польоту.

Далі у статті представлені структуровані ТТХ п'яти ударних FPV-дронів, які поєднують у собі передові технології, високу маневреність та значну бойову потужність, необхідну для виконання військових завдань у сучасних конфліктах.

Першим авторами роботи запропоновано розглянути дрон DJI FPV, який є технологічною розробкою китайської приватної компанії, провідного виробника квадрокоптерів, мікроконтролерів та відеообладнання. DJI є одним із лідерів ринку БпЛА та інноватором у сфері літальних дронів, контролерів для БпЛА і обладнання для стабілізації відеозйомки. Дрон DJI FPV зображено на рис. 1 [18].

Основні ТТХ DJI FPV наведено в табл. 2 [18]. Аналіз ТТХ DJI FPV показує, що цей дрон є високотехнологічним та багатофункціональним інструментом для виконання різноманітних завдань. Завдяки

своїм компактним розмірам і відносно малій вазі, DJI FPV забезпечує високу маневреність. Максимальна швидкість у 140 км/год та максимальна висота польоту до 6000 метрів роблять його важкою ціллю для виявлення противником.



Рис. 1. Модель FPV дрону – DJI FPV

Таблиця 2 – ТТХ DJI FPV

Характеристика	Значення
Розмір	255 x 312 x 127 мм
Вага	795 г (з акумулятором)
Максимальна швидкість	140 км/год (в режимі Manual)
Максимальна висота польоту	6000 метрів
Час польоту	до 20 хвилин
Діапазон робочих частот	2,4-2,48 ГГц 5,725-5,850 ГГц
Максимальна дальність польоту	до 16.8 км (залежить від умов)
Сенсор камери	1/2.3" CMOS, 12 МП
Об'єктив камери	FOV: 150°, діафрагма: f/2.8
Відео	4K/60fps, 1080p/120fps
Стабілізація камери	Електронна стабілізація зображення (EIS)
Система передачі відео	DJI O3 (OcuSync 3.0)
Дальність передачі відео	До 10 км (FCC), до 6 км (CE/SRRC/MIC)
Тип акумулятора	LiPo 6S
Ємність акумулятора	2000 мА·годин
Напруга акумулятора	22.2 В
Режими польоту	N Mode (Normal), S Mode (Sport), M Mode (Manual)
Режими повернення	Автоматичне повернення при низькому заряді акумулятора або втраті зв'язку
Наявність автосупроводу цілі	не передбачено
Ціна	~30000 гривень

Відносно великий час польоту для такого типу БпЛА (до 20 хвилин), а також велика дальність польоту (до 16.8 км), дозволяють виконувати доволі тривалі завдання. Камера з сенсором 1/2.3" CMOS і можливістю знімання відео в 4K/60fps забезпечує високоякісне зображення, що є критично важливим при наведенні на ціль ураження.

На рис. 2 наведений FPV-дрон – ProDrone 10inch VTx1.2 (2w) / TxES720, який є FPV-дроном української збірки (компоненти дрону виготовляються в Китайській Народній Республіці) [19].

У порівнянні з іншими дронами, розглянутими у статті, це найдорожча модель. Ціна обумовлена

повною комплектацією (акумулятор, окуляри, зарядний пристрій, пульт керування) та встановленою високоякісною камерою. Тому, враховуючи його характеристики та можливості, цей дрон є ефективним засобом для виконання спеціалізованих військових завдань. У табл. 3 наведені основні ТТХ ProDrone 10inch VTx1.2 (2w) / TxES720 [19].



Рис. 2. Модель FPV-дрону – ProDrone 10inch VTx1.2 (2w) / TxES720

Таблиця 3 – ТТХ ProDrone 10inch VTx1.2 (2w) / TxES720

Характеристика	Значення
Розмір	297×297×80 мм
Вага	918 грам (без батареї)
Максимальна висота польоту	5000 метрів
Максимальна швидкість	72 км/годину
Камера	Thermal cam 640x512, 9.1mm
Тип камери	з тепловізором
Діапазон робочих температур	від -10° до 40° C (14° до 104° F)
Протокол зв'язку	ELRS
Робоча частота	720 MHz
Рекомендовані акумулятори	6S2P 10.000Mah
Вантажопідйомність	ефективна до 3 кілограм
Навігація	без GPS
Анени	1.2g antenna
Максимальний час в повітрі	залежить від АКБ та корисного навантаження (БЧ)
Пропелери	HQProp 1550
Двигун	3115 motor kv900 6S
Рама	кастомне карбонове шасі 427 мм, 2/3 мм товщина, товщина променів 7,5 мм
Контролер польоту	F405 FPV drone flight controller
ESC	6s 60A 4-in-1 ESC/10 inch version
Передача відео	2W, 1.2G transmitter
Наявність автоматичного супроводу цілі	не передбачено
Ціна	~ 63000 гривень

Робоча частота 720 МГц та протокол зв'язку ELRS забезпечують надійну комунікацію між дронами і оператором та є суттєвою відмінністю між іншими представленими у роботі FPV-дронами.

Достатньо велика вантажопідйомність бойової частини (до 3 кілограм) робить дану модель досить

пріоритетною серед інших моделей дронів.

Ядро інфрачервоної камери Thermal cam 640x512, 9.1mm об'єднує інфрачервоний тепловий детектор 640x512/12 мкм на рівні пластини (WLP). Вона відрізняється різким і чітким зображенням, компактними розмірами та низькою вартістю.

На рис. 3 зображений FPV-дрон AvengeAngel Hercules-13, виробництва країни Китайської Народної Республіки [20]. У табл. 4 наведені основні ТТХ AvengeAngel Hercules-13 [20].



Рис. 3. Модель FPV дрону – AvengeAngel Hercules-13

Таблиця 4 – ТТХ AvengeAngel Hercules-13

Характеристика	Значення
Розмір	294,3×200×600
Максимальна дальність польоту	до 15 кілометрів
Максимальна швидкість	74 км/годину
Камера	AvengeAngel A10
Діапазон робочих температур	-10°C ~ 40°C
Передавач	AvengeAngel 5.8G 2.5W
Приймач	Elsr750/Elsr915/TBS RX
Акумулятор	Lipo 12000mAh 8s
Вантажопідйомність	до 7 кілограм
Навігація	без GPS
Антенна	AvengeAngel 5.8G
Максимальний час в повітрі	до 15 хвилин
Пропелери	13 дюймів
Двигун	AvengeAngel 4214-380KV
Наявність автоматичного супроводу цілі	не передбачається
Ціна	~ 60000 гривень

Велика потужність акумулятора FPV-дрону AvengeAngel Hercules-13 дає можливість застосовувати дрони даного типу на великі відстані (до 15 кілометрів) та нести відносно велику бойову частину в польоті (до 7 кілограмів). Управління дроном здійснюється на стандартній для такого типу БПЛА частоті 5.8 МГц. Слід зауважити, що до цього типу дрону виробником передбачено велика кількість ретрансляторів для посилення сигналу.

Далі проаналізуємо FPV-дрон – BetaFPV Pavo25 Whoop зображений на рис. 4 [21].

Даний FPV-дрон з діаметром гвинтів 2,5 дюйма, розроблений компанією BETA FPV (Китайська Народна Республіка), характеризується новітньою конструкцією рами, що оптимізує її структуру та зменшує кількість складних компонентів. Для скла-

дання повітропроводу і дрона необхідно всього шість гвинтів, що значно спрощує процес монтажу. Рама виготовлена із високоякісного матеріалу PA12, що забезпечує її міцність і довговічність. Цей дрон підтримує використання широкого спектра камер FPV, доступних на ринку, що робить його універсальним у застосуванні. Для цифрової HD-версії використовується комплект Caddx Nebula Pro Nano Vista, тоді як для аналогової версії з передавачем VTX – Caddx Baby Ratel2+A03 з потужністю 400 мВт на частоті 5,8 ГГц. У табл. 5 наведені основні ТТХ Pavo25 Whoop [21].



Рис. 4. Модель FPV дрону – BetaFPV Pavo25 Whoop

Таблиця 5 – ТТХ BetaFPV Pavo25 Whoop

Характеристика	Значення
Розмір	108×108 мм
Вага	153,3 г (без акумулятора)
Рамка	Комплект рами Pavo25
Двигун	1404 безщітковий двигун 4500KV
Камера	SMO 4K
Час польоту	до 7,5 хвилин
Відеопередавач	Caddx Nebula Pro Nano Vista Kit (HD Digital)/Caddx Baby Ratel2+A03 400 мВт 5,8G (аналоговий), Walksnail Avatar HD Digital Kit (HD Digital)
Регулює кут камери	0°-50°
Приймач	PNP/ELRS 2,4G/TBS
Рекомендований акумулятор	4S 750 мАч/4S 850 мАч
Наявність автосупроводу цілі	не передбачається
Ціна	~ 10000 гривень

Версія Analog VTX обладнана камерою Caddx Baby Ratel2, що має 1/18-дюймовий HDR-датчик зоряного світла з роздільною здатністю 1200 ТВЛ. Відеопередавач A03 400 мВт VTX, який входить у комплект, характеризується відкритим вихідним кодом і забезпечує регульовану вихідну потужність у режимах PIT, RCE, 100 мВт та 400 мВт.

Версія Walksnail HD VTX укомплектована набором Walksnail Avatar HD Kit, який забезпечує запис відео у форматі 1080P з наднизькою

затримкою 22 мс. Конструкція з двома антенами дозволяє здійснювати передачу сигналу на відстань до 4 км. Зазначена вартість FPV-дрону BetaFPV Pavo25 Whoop не враховує повний комплект (окуляри, акумулятор, пульт керування).

Іншим ударним FPV-дроном, який широко використовується в бойових діях на території України є SPIDER-2 TBS, розроблений компанією DRONE UKRAINE, і є інноваційним продуктом українських фахівців. Даний дрон характеризується відмінними технічними параметрами та здатністю переносити вантажі до 4 кілограмів. Також дрон відрізняється високою швидкістю для даного типу БПЛА (до 140 км/год). FPV-дрон SPIDER-2 TBS зображений на рис. 5 [22]. У табл. 6 наведені основні ТТХ SPIDER-2 TBS [22].



Рис. 5. Модель FPV-дрону – SPIDER-2 TBS

Таблиця 6 – ТТХ SPIDER-2 TBS

Характеристика	Значення
Розмір	350×350×93
Вага	1525 грам
Максимальна вага навантаження	до 4000 грам
Максимальна швидкість польоту	140 км/годину
Максимальна висота польоту	7000 метрів
Максимальна дальність польоту	до 20 кілометрів
Максимальний час польоту	до 30 хвилин
Діапазон робочих температур	-20°C до +50°C
Камера	RunCam Phoenix 2
Двигун	Mamba KN 3115 730KV
Антенa	Mamba KN 150 mm
Відеопередавач	RUSH MAX Solo Tank 5.8G 2.5W
Приймач	TBS Crossfire Nano RX (SE)
Дистанційне керування	2,4 ГГц або 915 МГц/868 МГц
Пропелери	Gemfan 7040-3
Наявність автосупроводу цілі	не передбачається
Ціна	~ 25000

Особливою відмінністю FPV-дрону SPIDER-2 TBS є можливість здійснювати управління дроном на різних частотах та максимальна висота польоту (до 7 000 метрів).

Таким чином, проведений аналіз ТТХ FPV-дронів дозволив зробити наступні висновки:

- на практиці у більшості випадків FPV-дрони застосовуються як ударні дрони камікадзе;

- FPV-дрони є малошвидкісними БпЛА, за висотою польоту належать до великовисотних та є БпЛА I класу (міні(тактичні));

- управління розглянутими FPV-дронами здійснюється на різних частотах, враховуючи постійне застосування та удосконалення противником засобів РЕБ. Отже постає актуальне завдання удосконалення даного параметру (зміна діапазону робочих частот);

- FPV-дрони можуть нести в польоті боєприпаси до 15 кілограмів, що робить їх доволі рушійною силою для знищення військової техніки;

- адаптивність дронів до компонентів дозволяє використовувати різні типи камер, акумуляторів, пультів тощо;

- в більшості моделей FPV-дронів відсутній GPS;

- розглянуті моделі FPV-дронів своєю конструкцією не передбачають автоматичний супровід цілі у разі подавлення каналу засобами РЕБ.

Ударним FPV-дронам властиві численні переваги, які обумовлюють їх високу затребуваність в ЗС України. Основними перевагами є:

- відносно невелика ціна. Вартість таких типів БпЛА рідко перевищує 30 000 тисяч гривень (неповного комплекту), а це в рази менше ніж коштує артилерійський снаряд;

- висока точність при ураженні цілі. Відеошолом забезпечує оператора різними можливостями для роботи з ціллю. На відміну від звичайних моніторів або екранів смартфона, FPV-окуляри не відблискують, що дозволяє комфортно керувати БпЛА навіть у яскравий сонячний день. За умови належної вправності оператора FPV-дрон може вражати цілі мінімального розміру, такі як люки танків або бліндажі;

- велика швидкість польоту. FPV-дрони здатні літати дуже швидко. Деякі FPV-дрони, можуть розвивати швидкість до 150 кілометрів на годину. Зазначений фактор значно ускладнює рятування від таких типів ударних БпЛА, зазвичай особовий склад противника не встигає навіть ідентифікувати БпЛА;

- невеликі розміри FPV-дронів, що ускладнює можливість їх виявлення для противника;

- достатньо велика вантажопідйомність бойової частини (снаряду). Це дозволяє використовувати підвіси достатньої ваги (до 7 кілограм). Для порівняння ручна граната типу Ф-1 важить близько 690 грамів;

- відносна стійкість до засобів РЕБ. FPV-дрони управляються вручну оператором за допомогою радіоканалу, що робить їх нечутливими до засобів РЕБ, спрямованих на пригнічення супутникової навігації. Крім цього, на відміну від цивільних БпЛА, у випадках перетворення їх на бойові апарати, ці

дрони можуть бути оснащені спеціальною прошивкою, яка ускладнює декодування сигналу противником. Проте, слід враховувати, що будь-які БпЛА можуть бути подавлені засобами РЕБ шляхом виявлення місця розташування операторів за допомогою засобів РЕБ. Тому важливо застосовувати FPV-дрони у поєднанні з ретрансляторами та виносними антенами для захисту оператора від нанесення удару противником;

- широкий діапазон застосування. FPV-дрони можуть застосовуватися не тільки для нанесення ударів по противнику, а й для доставки вантажів на небезпечних ділянках в районах ведення бойових дій, ведення розвідки, корегування вогню артилерії тощо;

- висока адаптивність. Це дозволяє модифікувати FPV-дрони до конкретних завдань, що робить їх універсальними та ефективними в різних умовах;

- використання ударних FPV-дронів створює психологічний ефект на противника, що призводить до зниження його морально-психологічного стану, створюючи відчуття постійної загрози.

Проте, як і будь-яка зброя, FPV-дрони не позбавлені певних недоліків. Основними недоліками ударних FPV-дронів є:

- обмежений час польоту. Акумулятори FPV-дронів мають обмежений час роботи, що суттєво зменшує тривалість проведення завдання та дальність польоту;

- високі вимоги до професійних навичок оператора. Ефективне управління FPV-дроном потребує високих навичок оператора, що вимагає тривалого навчання та досвіду. Недосвідчені оператори можуть знизити ефективність використання дронів і підвищити ризик їх втрати;

- уразливість до погодних умов. FPV-дрони можуть бути обмежені у використанні за несприятливих погодних умов, таких як сильний вітер, дощ або сніг, що може впливати на їхню стабільність та ефективність.

3. Тактика застосування ударних FPV-дронів. Тактика застосування ударних FPV-дронів під час ведення бойових дій включає кілька ключових етапів, які забезпечують ефективність та максимальну результативність їх використання.

По-перше, перед застосуванням ударного FPV-дрону доцільно провести попередню розвідку місцевості за допомогою інших БпЛА (коптерного типу) або розвідувальних підрозділів. Зазначене необхідно для виявлення точного місцезнаходження особового складу або військової техніки противника. Далі слід провести аналіз потенційних цілей ураження для визначення їх захищеності та вразливості.

По-друге, необхідно налагодити координацію та взаємодію з іншими суміжними підрозділами в районі виконання бойового завдання. Даний етап є невід'ємною складовою та проводиться для того, щоб свої підрозділи не вразили FPV-дрон та при необхідності надали підтримку з повітря наявними засобами.

По-третє, здійснюється атака на ціль (противника) із використанням боєприпасів, під'єднаних до FPV-дронів. Далі відбувається оцінка успішності

застосування та збір інформації про нанесенні ураження противнику.

Під час застосування ударних FPV-дронів на кожному етапі слід забезпечити безпеку оператора і збереження отриманих даних. Це досягається шляхом застосування ретрансляторів та виносних антен.

У бойових умовах типовий екіпаж ударного FPV-дрону складається із оператора та помічника. Оператор, керуючи дроном через окуляри віртуальної реальності, концентрується на точному наведенні на ціль для її ураження. Водночас, помічник через спеціальний монітор відстежує рухи БпЛА, контролює важливі параметри якості польотного процесу. Він забезпечує необхідне наведення на ціль у разі дезорієнтації пілота, моніторить вплив погодних умов, напрям вітру, рівень заряду батарей і якість сигналу. Додатково він спостерігає та аналізує навколишнє середовище щодо змін в тактичній обстановці з метою своєчасного виявлення противника в районі виконання завдання та проведення заходів щодо збереження життя екіпажу.

Висновки

Таким чином, аналіз ГТХ та тактики застосування ударних FPV-дронів показав, що такі БпЛА суттєво впливають на хід сучасних військових операцій, надаючи нові можливості для нанесення ефективних точкових ударів по техніці та особовому складу противника, проведення розвідки, коригування вогню та функцій транспортування. Ударні FPV-дрони демонструють високу ефективність завдяки маневреності, відносно низькій вартості та можливості модифікації під конкретні бойові задачі. Водночас, їх вразливість до засобів РЕБ вимагає подальшого удосконалення систем зв'язку та управління.

Напрямок подальших досліджень. Аналіз досвіду також дозволяє запропонувати можливі напрямки подальших досліджень для підвищення ефективності застосування БпЛА під час ведення бойових дій. Засоби РЕБ можуть суттєво вплинути на

канали управління та навігації FPV-дронів, проте вони не мають впливу на відеоканали отримання зображення в системах націлювання ударних FPV-дронів. Тому з метою підвищення ймовірності ураження техніки противника в умовах дії РЕБ, перспективним є використання ударних FPV-дронів з самонаведенням. Однією з умов успішної імплементації такої системи є розробка алгоритмів сегментування отриманих з дрону зображень в режимі реального часу. Для цього можна використовувати методи сучасного комп'ютерного зору, в тому числі з елементами штучного інтелекту. Результатом сегментування таких зображень має бути розподіл зображення на штучні об'єкти (цілі ураження) та природні об'єкти (фон). За наявності такої системи, оператору ударного FPV-дрона необхідно лише захопити ціль, виділивши об'єкт на екрані, після чого автоматика візьме на себе подальші дії.

Також, одним з найскладніших та найважливіших етапів обробки зображень в системах націлювання ударних FPV-дронів є дешифрування зображень. Від якості виконання операції сегментування оптико-електронного зображення залежить результат обробки та дешифрування зображень. Тому, в подальших дослідженнях доцільно приділити увагу сегментуванню зображень в системах націлювання ударних FPV-дронів.

Результати досліджень можуть стати підґрунтям для розробки рекомендації щодо оптимізації конструкції, вибору компонентів та навчання операторів, що сприятиме підвищенню ефективності ударних FPV-дронів.

Дослідження були проведені за рахунок грантової підтримки Національного фонду досліджень України в рамках конкурсу "Наука для зміцнення обороноздатності України", назва проекту "Інформаційна технологія автоматизованого сегментування зображень об'єктів в системах націлювання ударних FPV-дронів на основі алгоритмів ройового інтелекту", реєстраційний номер 2023.04/0153.

СПИСОК ЛІТЕРАТУРИ

1. Коршець О., Горбенко В. Уроки застосування безпілотних літальних апаратів у російсько-українській війні. Повітряна міць України. 2023. № 1(4). С. 9–17. <https://doi.org/10.33099/2786-7714-2023-1-4-9-17>.
2. Титаренко О., Горобець Ю. Аналіз застосування безпілотних літальних апаратів в збройних конфліктах в Сирії, Лівії та Нагірному Карабасі в період 2017-2020 рр. для удосконалення форм і способів застосування Повітряних Сил. Повітряна міць України. Том.1 № 2(3). С. 5–10. [https://doi.org/10.33099/2786-7714-2022-1-2\(3\)-5-10](https://doi.org/10.33099/2786-7714-2022-1-2(3)-5-10).
3. Волошин В. Чи можуть змінити дрони перебіг війни? Армія inform: веб-сайт. URL: <https://armyinform.com.ua/2020/10/24/chy-mozhut-zminyty-drony-perebig-vijny/>
4. Байрактари та квадрокоптери зі скидами. PS Magazine українською: веб-сайт. URL: <https://drukarnia.com.ua/articles/bairaktari-ta-kvadrokopteri-zi-skidami-g28dH>
5. Volkov N., Evans A., Mappes G., Bailey R., Kagan F. W. Russian Offensive Campaign Assessment, April 10, 2024. Institute for the study of war. URL: <https://www.understandingwar.org/backgrounder/russian-offensive-campaign-assessment-april-10-2024>
6. Іран атакував Ізраїль дронами та ракетами. BBC News Україна: веб-сайт. URL: <https://www.bbc.com/ukrainian/articles/c6pyjevnuq93o>
7. Kumari J., Ravishankar M., Jatkar M., Kumar P., Arya P., Garg, A. Military Grade FPV Drone for Enemy Recognition. Journal of Cyber Security, Privacy Issues and Challenges. 2023. № 2. P. 7–13. <https://doi.org/10.46610/JCSPIC.2023.v02i01.002>.
8. Alotaibi A., Chatwin C., Birch P. Ubiquitous Unmanned Aerial Vehicles (UAVs): A Comprehensive Review. Shanlax International Journal of Arts, Science and Humanities. 2023. № 11. P. 62–90. <https://doi.org/10.34293/sijash.v11i2.6650>.
9. Singh B., Bennet J. Gyro-Stabilized Camera Control in drones for Military Applications. IOP Conference Series: Materials Science and Engineering. 2021. Vol. 1012. P. 012017. <https://doi.org/10.1088/1757-899X/1012/1/012017>.
10. Anand T., Kashyappan A., Kailash K., Nithiyantham K. Development and Automation of Fixed Wing UAV for Reconnaissance Mission with FPV Capability. INCAS BULLETIN. 2022. № 14. P. 111–118. <https://doi.org/10.13111/2066-8201.2022.14.4.9>.

11. Dolgin D. L., Van Der Like D., London J., Holdman C. From a Drones Point of View. *HCI International 2021 - Late Breaking Posters*. 2021. P. 517–520. https://doi.org/10.1007/978-3-030-90176-9_66.
12. Dhafari L., Afzal A., Al-Abri O., Khan A. Solar-Powered UAVs: A systematic Literature Review. 2024 2nd International Conference on Unmanned Vehicle Systems-Oman. 2024. P. 1–8. <https://doi.org/10.1109/UVS59630.2024.10467158>.
13. Wang H., Cheng H., Hao H. The Use of Unmanned Aerial Vehicle in Military Operations. *Man-Machine-Environment System Engineering*. 2020. P. 939–945. https://doi.org/10.1007/978-981-15-6978-4_108.
14. Reddy C., Venkatesh B. Unmanned Aerial Vehicle for Land Mine Detection and Illegal Migration Surveillance Support in Military Applications. *Drone Technology: Future Trends and Practical Applications*. 2023. <https://doi.org/10.1002/9781394168002.ch13>.
15. Benjak J., Hofman D., Knezović J., Žagar M. Performance Comparison of H.264 and H.265 Encoders in a 4K FPV Drone Piloting System. *Applied Sciences*. 2022. № 12(13). P. 6386. <https://doi.org/10.3390/app12136386>.
16. Про затвердження Правил виконання польотів безпілотними авіаційним комплексами державної авіації України : Наказ Міністра оборони України від 08.12.2016 р. №661. URL: <https://zakon.rada.gov.ua/laws/show/z0031-17#Text>.
17. Толмачов В.С., Маринченко Є.О. Особливості вивчення принципів роботи та використання FPV-дронів під час підготовки майбутніх спеціалістів професійної освіти. *Аграрні інновації*. №22. 2023. С.97-100. <https://doi.org/10.32848/agrar.innov.2023.22.16>.
18. DJI FPV Redefine Flying. [Dji.com/global. веб-сайт](https://www.dji.com/global/dji-fpv). URL : <https://www.dji.com/global/dji-fpv>.
19. Квадрокоптер ProDrone 10inch VTx1.2 (2w) \ TxE5720 (THERMAL ver.) without battery. PRODRONE професійно про дрони. веб-сайт. URL: <https://prodrone.com.ua/ru/5300000886/>.
20. Hercules-13. AvengeAngel. веб-сайт. URL : <https://avengeangel.com/products/avengeangel-hercules-13>, 13-inch-heavy-load-professional-military-kamikaze-fpv,extreme-large-and-powerful-fpv-with-motor,-can-load-7-kg,15-minutes-flying-time-with-7kg-payload,first-person-view-achieve-precise-strike-target
21. Квадрокоптер BetaFPV Pavo25 Whoop. RC Store. веб-сайт. URL : <https://rc-store.com.ua/kvadrokopter-betafpv-pavo25-whoop>
22. FPV-дрон камікадзе DRONE UKRAINE SPIDER-2 TBS вантажопідйомність 4,0 кг. SAFETY. веб-сайт. URL : <https://safety.kiev.ua/fpv-dron-kam-kadze-drone-ukraine-spider-2-tbs-vantazhop-dyomn-st-4-0-kg/>

Received (Надійшла) 21.05.2024

Accepted for publication (Прийнята до друку) 31.07.2024

Analysis of the tactical and technical characteristics and tactics of using existing strike FPV-drones

H. Khudov, I. Khizhnyak, I. Hridasov, U. Zbezhkhovska, I. Yuzova, Y. Solomonenko, T. Kalimulin

Abstract. The subject of the study in the article is the tactical and technical characteristics and tactics of strike FPV drones. The aim is to analyze the tactical and technical characteristics of existing strike FPV drones and how they are commonly used in combat operations. **Tasks:** identification and classification of unmanned aerial vehicles, analysis of the main technical parameters of strike FPV drones, research on the methods and strategies for the use of strike FPV drones in combat, investigation of the vulnerabilities of strike FPV drones to electronic warfare means and methods of protection against them, and development of recommendations for improving the technical characteristics and combat capabilities of strike FPV drones. **The methods** used are comparative analysis, data clustering, analytical and empirical methods of analyzing the use of strike FPV drones in military conflicts, particularly in the Russian-Ukrainian war. **The following results were obtained.** A structured analysis of the tactical and technical characteristics and tactics of strike FPV drones has been conducted, highlighting their advantages and disadvantages and forming proposals for improvement. The practical significance of the obtained results can serve as a basis for enhancing existing strike FPV drones, the development of new models, and increasing their effectiveness in combat conditions. **Conclusions:** The analysis of the tactical and technical characteristics and the tactics of use of strike FPV drones has shown that such unmanned aerial vehicles significantly influence the course of modern military operations, providing new opportunities for effective precision strikes on enemy equipment and personnel, conducting reconnaissance, adjusting fire, and transportation functions. Strike FPV drones demonstrate high efficiency due to their maneuverability, relatively low cost, and the ability to be modified for specific combat tasks. At the same time, their vulnerability to electronic warfare necessitates further improvement of communication and control systems. The analysis of experience also allowed the proposal of possible directions for further research to increase the effectiveness of unmanned aerial vehicles during combat operations. Electronic warfare can significantly impact the control and navigation channels of FPV drones, but it does not affect the image acquisition channels in the targeting systems of strike FPV drones. Therefore, to increase the probability of hitting enemy equipment under electronic warfare conditions, the use of self-guiding strike FPV drones is promising. One condition for successfully implementing such a system is the development of real-time image segmentation algorithms received from the drone. Modern computer vision methods, including those with artificial intelligence elements, can be used for this. The result of such image segmentation should be the division of the image into artificial objects (targets) and natural objects (backgrounds). With such a system, the operator of a strike FPV drone only needs to capture the target by selecting the object on the screen, after which the automation will take over further actions. Also, image decryption is one of the most complex and important stages of image processing in the targeting systems of strike FPV drones. The quality of the segmentation operation of the optoelectronic image affects the result of image processing. Therefore, in further research, it is advisable to focus on image segmentation in the targeting systems of strike FPV drones. The research results can form the basis for developing recommendations for optimizing the design, selecting components, and training operators, which will increase the combat effectiveness of strike FPV drones in future wars. The research was conducted with grant support from the National Research Foundation of Ukraine within the framework of the "Science for Strengthening Ukraine's Defense Capability" competition, project title "Information Technology for Automated Image Segmentation of Objects in Strike FPV Drone Targeting Systems Based on Swarm Intelligence Algorithms," registration number 2023.04/0153.

Keywords: application tactics, FPV-drone, image segmentation, optical and electronic equipment, tactical and technical characteristics, unmanned aerial vehicle.

В. С. Бешта, А. В. Комаричев, Т. В. Філімончук, Д. І. Бараней

Харківський національний університет радіоелектроніки, Харків, Україна

МОДЕЛЬ МОБІЛЬНОГО ДОДАТКУ, ЯКА ОРІЄНТОВАНА НА ОБРОБКУ ДАНИХ

Анотація. Актуальність цього дослідження полягає у тому, що розвиток сучасних технологій та швидкий темп життя суспільства вимагають від мобільних додатків не лише високої продуктивності, але й забезпечення високого рівня зручності для користувачів. **Об'єктом** дослідження в даному контексті є модель, яка спроектована для мобільного додатку. Модель складається з декількох взаємопов'язаних елементів, включаючи інтерфейс користувача, клієнтську та серверну частини, API, заходи безпеки, аналітику та моніторинг, хмарні сервіси та модулі тестування. Кожен компонент відповідає за конкретний аспект функціоналу, сприяє стабільності та ефективності мобільного додатку. Робота також враховує можливість розширення та адаптації моделі в майбутньому для відповіді на зростаючі вимоги користувачів та технологічні тенденції. **Предметом** дослідження виступає взаємодія компонентів моделі мобільного додатку. В роботі аналізується, як кожен компонент моделі впливає на функціональність та продуктивність додатку в цілому, спрямовуючи увагу на оптимізацію його роботи для досягнення кращих результатів. **Результати.** Кожен компонент запропонованої моделі відіграє важливу роль у стабільності та ефективності мобільного додатку, а також враховує розширення моделі за рахунок додавання необхідних складових. **Висновки.** Вивчення моделі мобільного додатку є важливим кроком у забезпеченні його успішного функціонування. Розуміння взаємодії компонентів та оптимальних шляхів їхньої конфігурації дозволяє досягти високої ефективності та конкурентоспроможності на ринку мобільних додатків.

Ключові слова: мобільний додаток, модель додатку, клієнт-серверна архітектура.

Вступ

Сучасний світ невинно рухається вперед із стрімкими темпами технологічного розвитку. Однією з ключових областей, яка зазнала значущих трансформацій, є розробка мобільних додатків. Мобільні застосунки визначають нову реальність для широкого спектру послуг та взаємодії з користувачами у реальному часі. Однак успіх будь-якого мобільного додатку визначається не лише його інноваційними можливостями, але й ефективністю його архітектурного рішення. Робота присвячена науковому дослідженню моделі мобільного додатку, яка орієнтована на обробку даних. Метою роботи є визначення ключових компонентів та їх взаємодії, які спрямовано на забезпечення стабільності, ефективності та розширюваності мобільного додатку. Дослідження включає в себе аналіз інтерфейсу користувача, клієнтської та серверної частин, архітектурних принципів безпеки, аналітики та моніторингу, а також впровадження хмарних сервісів та модулів тестування.

В результаті проведених авторами досліджень буде розкрито сутність та ключові особливості архітектурної моделі мобільного додатку, що може слугувати підґрунтям для подальших інженерних рішень у сфері їх впровадження. Розробка якісного мобільного додатку починається з визначення його мети та напряму за яким слід рухатись, щоб досягти потрібного результату. Кожен додаток, що розробляється, має низку етапів, якими проходить вся розробка. Як правило, розробка починається з ідеї. Далі виконується підготовка та аналіз, здійснюється пошук концепції та шляхи створення відповідного додатку. На етапі аналітики ідея перетворюється на чіткий план дій, де здійснюється підбір стеку технологій та методології розробки програмного забезпечення. Це і є

підхід, за яким команда розробників буде створювати та наповнювати додаток. Тому вибір моделі розробки додатку визначить порядок реалізації задач, які заплановано, сформує терміни створення додатку та визначить його вартість [1].

Модель – це представлення об'єкта в будь-якій формі (математичній, фізичній, символічній), яка призначена для дослідження окремо взятих аспектів об'єкту для отримання відповідей на поставлені запитання [2]. Модель мобільного застосунку (додатку) – це структурований план або опис, який визначає архітектурні, функціональні та технічні елементи розробки мобільного додатка. Така модель, як правило, може оперувати різними методологіями, які орієнтовано на створення різноманітних додатків [3].

Методологія розробки програмного забезпечення – це сукупність методів, які застосовуються на різних стадіях життєвого циклу розробки додатку та мають спільний підхід, що дозволяє забезпечити найкращу ефективність процесів розробки [4]. В ІТ-інфраструктурі існує декілька фундаментальних методик для розробки мобільних додатків, кожна з яких має свою спрямованість.

На даний час актуальна модель мобільного застосунку має такі складові: UI (User Interface) – інтерфейс користувача; BL (Business Logic) – бізнес-логіка, яка реалізує основні функції додатку; DB (Database) – база даних, яка відповідає за зберігання та організацію даних; API (Application Programming Interface) – інтерфейс програмування застосунків, який забезпечує взаємодію між клієнтською та серверною частинами:

$$M = \{UI, BL, DB, API\}. \quad (1)$$

Наведена модель служить основою для розробки та впровадження мобільного додатку з метою забезпечення його ефективності та задоволення вимог

користувачів. Проте існуюча модель має ряд недоліків, таких як:

– проста модель може бути менш точною або недостатньо розгорнутою, оскільки вона не враховує всі можливі аспекти та взаємодії компонентів, що може впливати на результати та ефективність застосування [5];

– прості моделі легше розробляти та підтримувати, але це може обмежувати їхню гнучкість та можливість удосконалення з часом [6];

– для конкретних задач, де потрібна докладна або спеціалізована обробка, проста модель може бути недостатньою [7, 8].

Метою роботи є аналіз існуючої моделі мобільних додатків, її вдосконалення та розширення, яке забезпечує більшу модульність та її гнучкість.

Основна частина

На основі проведеного аналізу [9,10] було виявлено ряд недоліків, які впливають не лише на вартість розробки та підтримки мобільного додатку, а також на неможливість залучити та зберегти цільову аудиторію, на яку він був орієнтований. У зв'язку з цим було прийнято рішення сформулювати модель, яка була б орієнтована на розробку та подальше впровадження мобільного додатку:

$$M = \{UI, CF, IL, BL, DB, CS, API, SEC, AM, MT\}. \quad (2)$$

В порівнянні з існуючою моделлю (1) модель, що запропонована (2), включає додаткові складові, а саме: CF (Client Frontend) – клієнтська частина, яка реалізована на мобільній платформі (iOS та Android); IL (Interaction Logic) – логіка взаємодії, що містить компоненти для обробки даних, що були внесені, та взаємодії з сервером через API; CS (Cloud Services) – хмарні сервіси, які надають можливість здійснювати зберігання файлів користувачів, управління мультимедійним контентом та іншими ресурсами, що потребують значної потужності для зберігання; SEC (Security) – механізм безпеки, який включає аутентифікацію, авторизацію та інші заходи безпеки; AM (Analytics and Monitoring) – механізм аналітики та моніторингу, який забезпечує збір та подальший аналіз даних про використання додатку; MT (Mobile Testing) – модуль тестування, який відповідає за створення та проведення низки тестів для гарантії вірності роботи окремих компонентів додатку та їх взаємодії.

Проаналізуємо доцільність використання складових моделі мобільного додатку, що запропонована з точки зору розробника [11]. У світі, насиченому технологіями, мобільні додатки стали невід'ємною частиною нашого повсякдення. Однак, щоб ці додатки стали дійсно вдалими та користувалися популярністю, важливою є не лише їхня функціональність, але й інтерфейс користувача (UI). Аналіз та оптимізація цього аспекту є важливою задачею для забезпечення ефективності та задоволення від використання мобільного додатку. Одним із ключових складників інтерфейсу користувача є ретельно підібрані елементи, такі як кнопки, текстові поля, графічні об'єкти тощо. Їхня структура повинна не лише відповідати функціоналу, але й забезпечувати легкість сприйняття та взаємодії з додатком.

Високий рівень зручності взаємодії (UX) є також необхідним елементом успішного інтерфейсу користувача. Естетичний дизайн сприяє не лише вигляду додатку, але й враженню від взаємодії з ним. Оптимізація UX включає в себе аналіз потреб користувача та створення інтуїтивного інтерфейсу. З огляду на різні типи пристроїв та їх розміри екранів, UI повинен бути адаптивним. Це включає в себе відзначення та відповідність до різних роздільних здатностей, щоб користувачі могли отримати оптимальний досвід навіть на різних пристроях.

Інтерфейс користувача повинен чітко відображати можливість додатку та його функціонал. Інтуїтивність та логічність розміщення елементів дозволяють користувачеві ефективно взаємодіяти з додатком та використовувати його функціонал без зайвих труднощів. Механізм, який реалізує мобільний клієнт (CF) є ключовим елементом моделі мобільного додатку, відповідаючи за відображення інтерфейсу та взаємодію з користувачем на мобільних платформах. Елементи інтерфейсу та його дизайн створюють зручний та привабливий для користувача досвід.

При розробці мобільного клієнта важливо враховувати адаптацію до платформ, забезпечуючи сумісність з графічними стандартами та елементами управління на iOS та Android. Оптимізація продуктивності призводить до енергоефективності, яка в свою чергу оптимізує використання ресурсів пристроїв [12]. В цілому, мобільний клієнт виступає як інтерфейс між користувачем та системою, забезпечуючи ефективну та приємну взаємодію на платформі мобільних пристроїв. Серверна частина, відома як Back-End, є основою складовою мобільного додатку, тому що відповідає за обробку даних та забезпечує коректне функціонування додатку. Вона включає в себе низку складових (бізнес-логіку (BL), базу даних (DB) та API), які забезпечують взаємодію з клієнтською частиною. Готовність серверної частини адаптуватися до змін у навантаженні та здатність ефективно працювати при різних обсягах запитів користувачів є ключовою вимогою при розробці мобільного додатку.

Архітектура серверної частини повинна бути готова до легкого розширення та інтеграції нового функціоналу, забезпечуючи ефективну роботу додатку в цілому. Залучення серверної частини до існуючої інфраструктури та її інтеграція з іншими системами є важливим етапом для створення комплексного та згуртованого середовища додатку [13]. Така інтеграція дозволяє оптимізувати взаємодію з іншими компонентами системи та покращити її узгодженість.

За допомогою механізму, який реалізує логіку взаємодії мобільного клієнта (IL) здійснюється обробка дій користувача та спілкування з сервером через API. Даний механізм забезпечує введення та обробку даних, а також управління локальними операціями, відправляючи необхідні запити на сервер. Це забезпечує коректну та швидку реакцію на дії користувача, покращуючи загальний досвід використання додатку. Окрім того, механізм взаємодії відповідає за ефективне управління локальними операціями та передачу відповідних запитів на сервер. Це важливий елемент для забезпечення синхронності та актуаль-

ності інформації, що зберігається на клієнтському пристрої та сервері.

Бізнес-логіка мобільного додатку (BL) визначає його основні функції, включаючи обробку запитів користувача, валідацію введених даних та виконання різних бізнес-процесів. Важливою частиною є валідація введених даних, що гарантує коректність та надійність інформації, яку вводить користувач. Бізнес-логіка також відповідає за виконання різноманітних бізнес-процесів в мобільному додатку. Це може включати в себе розрахунок результатів, формування звітів, роботу з базою даних та інші ключові операції, необхідні для ефективної роботи додатку. Бізнес-логіка створює основу для функціональності додатку, роблячи його адаптованим до конкретних потреб користувачів та бізнес-вимог.

За допомогою використання баз даних (DB) здійснюється зберігання та організації даних, які необхідні для роботи додатку. Це може бути SQL або NoSQL рішення в залежності від конкретних потреб.

Хмарні сервіси (CS), які також є складовою моделі мобільного додатку, надають додаткові можливості для зберігання даних та взаємодії з різноманітними хмаровими ресурсами. Одна із основних переваг використання хмарних сервісів – це зберігання великих обсягів даних, таких як файли користувачів, мультимедійний контент чи інші ресурси, які вимагають великого об'єму. Іншим аспектом є використання хмарних сервісів для реалізації функціоналу в режимі реального часу. Наприклад, сервіси для реального часу можуть бути використані для відправлення пуш-повідомлень, синхронізації даних між різними пристроями чи отримання оновлень з сервера.

Інтеграція з хмарними сервісами може також полегшити розширення функціоналу додатку та забезпечити гнучкість в роботі. Слід розуміти, що використання готових хмарних рішень дозволяє розробнику уникнути необхідності розробляти та підтримувати власні сервери для зберігання та обробки даних. При цьому важливо враховувати питання безпеки та конфіденційності даних при використанні хмарних сервісів та обирати сервіси, які відповідають стандартам безпеки та вимогам до захисту інформації. API відіграє ключову роль у взаємодії між клієнтською та серверною частинами мобільного додатку, визначаючи, як дані передаються та обробляються.

Слід розуміти, що фактор безпеки в моделі мобільного додатку є важливим елементом для забезпечення конфіденційності, цілісності та доступності даних. Вони забезпечують захист від несанкціонованого доступу та зберігають конфіденційність інформації окремих частин додатку.

Аутентифікація та авторизація є ключовими елементами безпеки, забезпечуючи визначення ідентичності користувачів та регулювання їх доступу до функціоналу. Аутентифікація визначає процес перевірки ідентичності користувача, забезпечуючи впізнання його облікових даних. Це може включати в себе введення логіну та пароля, використання біометричних даних чи інших методів ідентифікації. Авторизація, у свою чергу, визначає, які дії та ресурси користувач може використовувати після успішної аутентифікації.

Цей процес регулюється на основі ролей та прав доступу, що забезпечує обмеження користувачів в їх можливостях відповідно до їхньої ролі в системі.

Шифрування даних також відіграє критичну роль у забезпеченні конфіденційності. Не слід забувати, що всі дані, які передаються між клієнтом та сервером, повинні бути зашифровані для захисту від несанкціонованого доступу. Заходи безпеки повинні включати моніторинг та журналювання подій для виявлення та відстеження потенційних загроз. Система повинна відповідати стандартам безпеки та регулярно оновлюватися для виправлення виявлених вразливостей. Механізм, який реалізує моніторинг працездатності мобільного додатку (AM) відіграє ключову роль у зборі, аналізі та використанні даних для покращення його продуктивності та ефективності. Ці фактори забезпечують здатність відслідковувати та реагувати на різноманітні чинники функціонування додатку. Інструменти аналітики – це, як правило, набори програмних засобів та методик, які розроблено для збору, аналізу та візуалізації даних, пов'язаних з різними аспектами роботи мобільного додатку. Дані інструменти відіграють важливу роль у розумінні даних та перетворенні їх на цінні інсайти для прийняття обґрунтованих рішень у подальшому розвитку та просуванню мобільного додатку. В якості інформації, яку збирають інструменти аналітики може бути інформація про кількість завантажень, активність користувачів, використані функції та інші параметри, що допомагають зрозуміти, як ефективно взаємодіє аудиторія з додатком.

Моніторинг продуктивності також спрямований на виявлення та усунення проблем, які можуть виникнути в процесі розробки та використання додатку. Це включає в себе відстеження швидкодії додатку, виявлення помилок та збоїв, а також контроль за завантаженням серверів. Здійснення аналізу та моніторингу дозволяє вчасно реагувати на зміни та покращувати додаток. Важливим етапом є також аналіз поведінки користувачів для визначення популярних функцій та тих функцій, які можуть бути покращені.

Використання модулю тестування (MT) у складі моделі мобільного додатку відіграє важливу роль у забезпеченні надійності, стійкості та правильності його роботи. Тестування включає різні методики, спрямовані на перевірку якості програмного забезпечення та визначення його відповідності вимогам.

Юніт-тести та інтеграційне тестування спрямовані на перевірку роботи окремих компонентів та їх взаємодії в мобільному додатку. Як правило юніт-тести перевіряють правильність роботи конкретних функцій чи методів, тоді як інтеграційне тестування оцінює взаємодію різних частин додатку в цілому.

Тестування витривалості та безпеки визначається для перевірки стійкості та відповідності додатка вимогам щодо обробки великої кількості запитів, а також для виявлення потенційних проблем безпеки та їх подальшого усунення.

Модуль тестування також включає фактори, які пов'язані із забезпеченням високої стабільності додатку. Тестування в реальних умовах допомагає виявити помилки та аномалії, які можуть виникнути в різних сценаріях використання.

Тому при розробці мобільного додатку слід пам'ятати, що забезпечення повного та ефективного покриття тестами дозволяє впевнитися в надійності та стабільності роботи додатку, а регулярне тестування в процесі розробки дозволяє виявляти та вирішувати проблеми на ранніх етапах роботи над ним.

Висновки

В результаті проведених авторами досліджень виявлено, що модель мобільного додатку є визначальною для забезпечення його ефективної та стабільної роботи. Взаємодія різних компонентів, таких як інтерфейс користувача, клієнтська та серверна частини, API, системи безпеки, грає вирішальну роль в успіху додатку. Гнучка архітектура моделі дозволяє легко адаптуватися до змін в умовах ринку та вимог

користувачів. Система безпеки, включаючи аутентифікацію, авторизацію та шифрування даних, є необхідною для забезпечення конфіденційності та цілісності інформації. Застосування інструментів аналітики та моніторингу дозволяє постійно вдосконалювати продуктивність додатку та вчасно виявляти можливі проблеми. Використання хмарних сервісів допоможе оптимізувати функціонал та забезпечує додаткові переваги при розміщенні інформації, а використання модуля тестування гарантує стабільність та надійність кожного компонента мобільного додатку.

Загальний висновок полягає в тому, що ретельно пророблена модель мобільного додатку є ключем до успішної реалізації та його функціонування, що може відповідати високим стандартам якості та вимогам користувачів.

СПИСОК ЛІТЕРАТУРИ

1. Ключові методології розробки програмного забезпечення: робота команди зсередини, [Електронний ресурс] – Режим доступу до ресурсу: <https://wezom.com.ua/ua/blog/metodologija-razrabotki-programnogo-obespechenija>
2. Сеспедес Гарсія Н.В., Сеспедес Гарсія П.Д. (2023), “Моделі життєвого циклу розробки програмного забезпечення”, *Молодий вчений*, №2 (114), С. 17-20, doi: <https://doi.org/10.32839/2304-5809/2023-2-114-4>.
3. Розробка мобільних додатків від А до Я: повний гайд, [Електронний ресурс] – Режим доступу до ресурсу: <https://dan-it.com.ua/uk/blog/rozrobka-mobilnih-dodatkov-vid-a-do-ya-povnij-ghajd>.
4. Ічанська Н.В., Улько С.І. (2020), “Основні аспекти створення мобільних додатків та вибір інструментів їх розробки”, *Системи управління, навігації та зв'язку*, Полтава: ПНТУ, 1 (59). С. 74-78. doi: <https://doi.org/10.26906/SUNZ.2020.1.074>.
5. Паламарчук К.А. (2020), “Тестування. Моделі розробки програмного забезпечення”, Матеріали конференції “XLIX Науково-технічна конференція підрозділів Вінницького національного технічного університету”. Вінниця: ВНТУ, URL: <https://conferences.vntu.edu.ua/index.php/mn/mn2020/paper/viewFile/8611/7186>.
6. Вакалюк Т.А., Лук'янчук Я.В. (2020), “Теоретичні аспекти розробки мобільного додатку для прогнозування ймовірної вартості земельної ділянки”, *Актуальні проблеми автоматизації та інформаційних технологій*: зб. наук. пр. / наук. ред. О. Г. Байбуз. Дніпро: ДНУ, Том 24, С. 16-31.
7. Лудан Д.В. (2017), “Розробка мобільного додатку для організації інтерактивної квестової системи”, Матеріали конференції “XLVI НТК підрозділів ВНТУ. Вінниця: ВНТУ, Том 2, С. 1113-1114.
8. Ляшенко О., Пославський С., Киричук Д., Прачик В., Бунккус С. (2023). “Розроблення мобільного додатку для відстеження стану здоров'я та планування харчування маломобільних груп населення в умовах надзвичайних ситуацій”. *Інформаційні технології та суспільство*, №2 (8), С. 42-50, doi: <https://doi.org/10.32689/maup.it.2023.2.5>.
9. Ткачов, В., Коваленко, А., Кучук, Г., Ні, Я. (2021). Метод забезпечення живучості високомобільної комп'ютерної мережі. *Сучасні інформаційні системи*, том 5, № 2. С. 159–165. Doi: <https://doi.org/10.20998/2522-9052.2021.2.24>
10. R. Whitaker (2020), “Developing Inclusive Mobile Apps: Building Accessible Apps for iOS and Android”. USA: Apress, 359 с.
11. Козолуп П., Любчак В. (2023). “Огляд методів та інструментів для розробки інформаційного сервісу обліку особистих активів”. *Інформаційні технології та суспільство*, №3 (9), 47-53, doi: <https://doi.org/10.32689/maup.it.2023.3.6>.
12. Carmine Zaccagnino (2020), “Programming Flutter: Native, Cross-Platform Apps the Easy Way (The Pragmatic Programmers)”. United States of America: Pragmatic Bookshelf, 370 с.
13. Процес розробки мобільних додатків [Електронний ресурс] – Режим доступу до ресурсу: <https://horosho.com.ua/proces-rozrobki-mobilnih-dodatkov>.

Received (Надійшла) 22.04.2024

(Accepted for publication) Прийнята до друку 19.06.2024

A mobile application model that is focused on data processing

Viacheslav Beshta, Andrii Komarychev, Tetiana Filimonchuk, Dmytro Baraniec

Abstract. The relevance of this study lies in the fact that the development of modern technologies and the fast pace of society require mobile applications not only to be highly productive, but also to provide a high level of user convenience. The object of study in this context is a model designed for a mobile application. The model consists of several interconnected elements, including the user interface, client and server parts, APIs, security measures, analytics and monitoring, cloud services, and testing modules. Each component is responsible for a specific aspect of the functionality, contributes to the stability and efficiency of the mobile application. The work also takes into account the possibility of expanding and adapting the model in the future to respond to growing user requirements and technological trends. The subject of the study is the interaction of components of the mobile application model. The paper analyzes how each component of the model affects the functionality and performance of the application as a whole, focusing on optimizing its performance to achieve better results. The following results were obtained. Each component of the proposed model plays an important role in the stability and efficiency of the mobile application, and also takes into account the expansion of the model by adding the necessary components. **Conclusions.** Studying the mobile application model is an important step in ensuring its successful operation. Understanding the interaction of components and the best ways to configure them allows achieving high efficiency and competitiveness in the mobile application market.

Keywords: mobile application, application model, client-server architecture.

А. С. Білоконь, С. О. Борисов, М. В. Усатенко, В. М. Федорченко

Харківський національний технічний університет радіоелектроніки, Харків, Україна

АНАЛІЗ ФУНКЦІОНУВАННЯ РОЗПОДІЛЕНИХ СИСТЕМ ОБРОБКИ ТА ЗБЕРІГАННЯ ДАНИХ

Анотація. Актуальність. Зі зростанням обсягу даних, генерованих користувачами, IoT пристроями, соціальними медіа, бізнес-процесами тощо, потреба в масштабованих рішеннях для зберігання стає все більш очевидною. Розподілені системи дозволяють ефективно масштабуватися, забезпечуючи збільшення обсягів зберігання і обчислювальної потужності без значних затрат. Сучасний бізнес вимагає високої доступності та надійності систем, оскільки навіть мінімальний час простою може призвести до значних фінансових втрат та зниження довіри клієнтів. Розподілені системи забезпечують високу доступність і витривалість, автоматично відновлюючи роботу після збоїв і реплікуючи дані для забезпечення їх цілісності. Глобалізація бізнесу вимагає роботи з даними у різних географічних локаціях. Розподілені системи дозволяють локалізувати зберігання та обробку даних ближче до кінцевих користувачів, знижуючи затримки і підвищуючи загальну продуктивність систем. Збільшення загроз безпеці та посилення нормативних вимог до захисту даних змушують організації шукати більш надійні рішення для зберігання даних. Розподілені системи пропонують розширені можливості для шифрування даних, регулювання доступу, аудиту та відповідності нормативним актам. Для обробки великих обсягів даних часто потрібні великі обчислювальні потужності. Розподілені системи зберігання даних ідеально підходять для роботи в парі з розподіленими обчисленнями, такими як обробка потокових даних, машинне навчання, великі дані, дозволяючи ефективно розподіляти завдання і обробляти великі обсяги інформації. Серед викликів, з якими можуть стикатися розподілені системи зберігання даних, - забезпечення консистентності даних між вузлами, управління затримками мережі, захист даних і забезпечення безпеки. Для вирішення цих викликів застосовуються різні стратегії та технології, включаючи алгоритми консистентного хешування, реплікацію даних, транзакційні протоколи з гарантією атомарності, консистентності, ізоляції та довговічності та моделі послідовної консистентності. Таким чином, в умовах постійного зростання обсягів даних та збільшення вимог до їх обробки, розподілені системи зберігання даних є ключовим елементом інфраструктури будь-якої організації, що прагне до інновацій та ефективності. **Метою даної роботи** є аналіз функціонування розподілених систем обробки та зберігання даних. **Об'єктом дослідження** є розподілені системи обробки та зберігання даних. **Предметом дослідження** є архітектурні рішення розподілених систем зберігання та обробки даних. **Результати.** Проведено аналіз функціонування розподілених систем обробки та зберігання даних. Вибір архітектурного рішення для розподіленої системи залежить від специфіки задач, які необхідно вирішити, вимог до продуктивності, масштабованості, надійності та доступності. Зазвичай, ефективні розподілені системи використовують комбінацію зазначених підходів для досягнення оптимальних результатів. Вибір архітектурного рішення для розподілених систем - це комплексний процес, який вимагає балансування між технічними, бізнесовими та операційними вимогами. Врахування майбутнього зростання, потенційних викликів і гнучкості системи є ключовими факторами для забезпечення її довгострокового успіху.

Ключові слова: розподілена інформаційна система, обробка даних, SOA, P2P, Web-Scale, сховище даних, центр обробки даних, event-driven architectures.

Вступ

Розподілені системи зберігання даних – це комплексні інформаційні системи, які дозволяють зберігати дані в розподіленому вигляді по декількох фізично рознесених серверах чи вузлах [1]. Такі системи забезпечують високу доступність, масштабованість і надійність зберігання даних. Вони є основою для багатьох сучасних веб-сервісів, облачних платформ і корпоративних додатків. Система може бути легко розширена шляхом додавання нових вузлів, що дозволяє збільшувати обсяги зберігання та обчислювальні можливості без перерв у роботі. Дані реплікуються між декількома вузлами, що забезпечує їх збереження у разі виходу з ладу одного чи декількох вузлів. Система забезпечує доступ до даних навіть при збоях окремих компонентів, завдяки реплікації даних і резервуванню. Розподілені системи намагаються забезпечити прозорість розміщення, виконання та міграції, щоб кінцевий користувач або додаток не відчував розподіленості ресурсів.

В теперішній час існує багато архітектурних рішень розподілених систем, які відіграють критичну роль у забезпеченні їх ефективності, масштабованості,

надійності та доступності. Мікросервісна архітектура полягає у розбитті додатку на невеликі, незалежні сервіси, кожен з яких виконує певну функцію та спілкується з іншими через легковісні механізми, такі як HTTP/REST або черги повідомлень. Цей підхід дозволяє незалежно розгортати та масштабувати кожен сервіс, що підвищує гнучкість та відмовостійкість системи. SOA - це архітектурний стиль, який забезпечує взаємодію різних сервісів через веб-сервіси або API. Він спрямований на забезпечення високого рівня інтеграції та взаємодії між різноманітними застосунками і системами в розподіленому середовищі [2].

У архітектурі з шарами розподілена система організована у вигляді шарів, кожен з яких має свої відповідальності (наприклад, презентаційний шар, бізнес-логіка, доступ до даних). Це спрощує розробку та тестування, але може вплинути на продуктивність через додаткові затримки при переході між шарами. У peer-to-peer (P2P) архітектурі всі вузли є однорідними та можуть виконувати роль як клієнта, так і сервера [3]. Це дозволяє системі ефективно масштабуватися та витримувати високі навантаження, розподіляючи ресурси та завдання між вузлами. Архітектура, яка заснована на подіях, передбачає реагування компонентів системи на

події (наприклад, повідомлення про зміни даних), що дозволяє створювати високоефективні та гнучкі системи, здатні швидко адаптуватися до змін умов роботи. Слід відзначити, що розподілена система може одночасно забезпечити тільки дві з трьох характеристик: консистентність, доступність та стійкість до розподілу (переривань мережі). Вибір між цими характеристиками залежить від конкретних вимог до системи.

Отже, **метою цієї роботи** є аналіз функціонування розподілених систем обробки та зберігання даних, а також огляд існуючих архітектур.

Основна частина

Існуючі на сьогодні вимоги до проектування та розробки сховищ даних (СД), що забезпечують прийнятну надійність та продуктивність, зумовлюються ефективною тенденцією високого зростання обсягу даних, які потребують довготривалого зберігання. Фахівці багатьох організацій зазначають, що Web-Scale IT включає реалізацію різних підходів до побудови IT-інфраструктури. Основною відмінністю Web-Scale IT є реалізація стійкості до відмови на рівні програмного забезпечення і вимагає оптимізації коду під апаратне забезпечення. Слід відзначити, що відмова від будь-яких технологій віртуалізації вносить ще один шар взаємодії з апаратним забезпеченням (і, цілком імовірно, відповідні затримки). Апаратне забезпечення розробляється з урахуванням архітектури програмного забезпечення, а також може бути інтегрованим або розподіленим на компоненти архітектури, побудовані за Web-Scale IT, високо масштабуються, оскільки рівні продуктивності досягаються великою кількістю ефективних неспеціалізованих вузлів та використанням типових серверів. Сукупні ресурси багатьох цих вузлів об'єднуються в пул для створення обчислювальної матриці та розподілу відповідно до потреби. Обладнання Web-масштабу - відносно недороге та доступне. При відмові вузол просто замінюється. Керує цією мережею вузлів інтегроване програмне забезпечення, яке, на відміну від апаратних засобів управління, може перепрограмувати систему, щоб вона відповідала мінливій динаміці. Така система по суті своїй еластична, оскільки дані, метадані та операції розподіляються на велику кількість малих вузлів, що замінюються, і, можливо, навіть на кілька центрів обробки даних.

Ключовими елементами архітектурного рішення СД з Web-Scale IT є: орієнтація використання центрів обробки даних, використання веб-орієнтованої чи мікросервісної архітектури надання послуг; наявність програмованого управління СД; можливість налаштування будь-яких бізнес-процесів, орієнтованих високу продуктивність; застосування методології «корпоративна культура організації, що навчається», сфокусована на інноваціях і безперервному навчанні.

Також виділяються два підходи до забезпечення масштабованості: scale-up та scale-out. Вертикально масштабовані системи (scale-up), яким властиве збільшення кількості доступних ресурсів за рахунок збільшення потужності застосовуваних обчислювальних засобів, та горизонтально-масштабовані системи (scale-out), що характеризуються наявністю значного числа невеликих серверів, об'єднаних у кластер,

на яких розподіляється поставлена замовником прикладне завдання. Незважаючи на різницю, scale-up і scale-out зазначені підходи можуть комбінуватися в рамках однієї СД. Однак масштабування до великої кількості вузлів, у тому числі з можливим розподілом по географічно рознесеним ЦОД, може призвести до проблеми забезпечення доступності та консистентності даних.

В даний час найбільш характерними представниками горизонтально-масштабованих систем, спроектованих згідно з Web-Scale IT, є СД компанії EMC Isilon та IBM XIV. Кластер горизонтально масштабованої системи, зазначених рішень складається з серверів, що включають процесори, кеш-пам'ять, диски, Ethernet-порти засобів підключення. Крім даних пристроїв, вузол EMC Isilon містить також порти Infiniband RDMA, за допомогою яких сервери об'єднуються у внутрішню мережу з метою забезпечення стійкості до відмови кластера [4]. Загальною для цих горизонтально масштабованих систем є одиниця масштабування, що є апаратним модулем з фіксованою кількістю приєднаних до нього дисків. Основні відмінності цих систем одна від одної полягають в особливостях організації доступу до даних та способу їх зберігання. В одному випадку забезпечується доступ до даних через організацію файлової системи, в іншому випадку використовується блоковий пристрій. Всі дані розподілені між усіма вузлами даних СД, і будь-який вузол може на рівних обслуговувати запити до будь-якого файлу, незалежно від того, на дисках або кеші якого вузла знаходяться дані. Тим самим досягається висока продуктивність роботи з будь-яким файлом.

Класичні платформи зберігання SAN та NAS, хоч і критично важливі для корпоративних додатків, але не призначені для роботи з сучасними хмарними додатками та не відповідають вимогам масштабування у хмарі [5]. Крім того, постійне зростання обсягів неструктурованих даних вимагає створення більш простої архітектури зберігання, яка призначена для управління мільярдами та трильйонами файлів, прискорює розробку хмарних та мобільних додатків та додатків для великих даних і при цьому дозволяє знизити витрати, пов'язані із зберіганням, та витрати в цілому. Щоб вирішити ці завдання, IT-служби та постачальники послуг почали тестувати та впроваджувати економічні типові та відкриті інфраструктури. Типові компоненти та відкриті технології на базі стандартів знижують витрати на зберігання, але окремо забезпечують нижчі показники продуктивності та надійності. Також слід пам'ятати, що інфраструктура зручна в обслуговуванні, якщо IT-фахівці замовника мають достатній досвід її експлуатації.

Безпрецедентне зростання обсягів даних (структурованих та неструктурованих) призводить до того, що файли великих обсягів (наприклад, зображення, відео тощо) зберігаються у дорогих ізолюваних СД. У класичній інфраструктурі, зазвичай, часто дані ізолюються, ускладнюючи цим обмін і управління ними і збільшуючи витрати. Зазначений фактор не підтримує ефективне та економічне масштабування, що змушує замовників шукати рішення, яке поєднує в собі переваги публічної та приватної хмари. Рішення дозволяють будь-якій організації об'єднати кілька

систем зберігання та архівів змісту в єдиний глобально доступний та ефективний репозиторій змісту, який може обслуговувати велику кількість програм.

Покладена в основу рішення архітектура типу «активний-активний» з підтримкою кількох майданчиків, єдиний глобальний простір імен та універсальна доступність забезпечують доступ до змісту з будь-якої точки для будь-якої програми чи пристрою. Завдяки розподілу контейнерів даних між майданчиками операції запису та читання можуть виконуватися в будь-якій точці світу. Рішення забезпечують високий рівень семантичної узгодженості, що дозволяє спростити розробку додатків та отримувати доступ до даних із будь-якої точки. Також у рішенні є функція геокешування, яка визначає схеми доступу з декількох майданчиків та кешує дані на майданчику, де ці дані використовуються найчастіше.

У загальноприйнятих класичних інфраструктурах зберігання даних, як правило, запровадити системи бізнес-аналітики може бути непросто. Так, дані часто розподілені між кількома складними системами, і ускладнює та збільшує витрати на доступ та управління. Організації змушені отримувати дані від операційних систем і потім завантажувати їх у виділений кластер для аналізу. В основі ефективної аналітики лежать точність та своєчасність даних. Рішення надає замовникам можливість видобувати практично значущі дані бізнес-аналітики з великих обсягів розподіленого змісту, не використовуючи робочого процесу ETL (Extract, Transform, Load). Рішення можуть підтримувати озеро даних, надаючи можливості доступу та управління будь-яким змістом організації за допомогою кількох дистрибутивів Hadoop [6].

При використанні звичних СД зі зростанням обсягів даних неминує зростає кількість контролерів, дискових груп, томів, файлових систем, так чи інакше з'являються «острівці» різнорідних систем, стає все складніше та складніше підтримувати ефективність їх використання. Нарощування систем вимагає попереднього аналізу, складних робіт з конфігурації та міграції даних [6]. При проектуванні рішень СД, їх загальна продуктивність визначається продуктивністю модуля управління, вплинути на який з боку персоналу (адміністраторів) практично неможливо. А ось перспективнішим варіантом управління продуктивністю є виділення двох рівнів масштабування: масштабування за ресурсами, яке виконується управлінням кількості апаратно-керованих масивів жорстких дисків, і масштабування за продуктивністю - управлінням кількістю апаратних блоків підготовки та обробки інформації. Далі перейдемо до існуючих рішень. Аналіз існуючих мікросервісних архітектур розподілених систем виявляє цілий ряд ключових компонентів, практик та викликів, які є актуальними для більшості сучасних високонавантажених та високодоступних систем.

Netflix є одним з найвідоміших прикладів успішного застосування мікросервісної архітектури. Система розподілена на сотні мікросервісів, які відповідають за різні аспекти платформи, від потокового відтворення до рекомендацій контенту та обробки користувацьких даних. Платформа базується на використанні контейнеризації, оркестрації, внутрішньої

мережевої інфраструктури для забезпечення зв'язку між сервісами (Netflix Zuul як API Gateway).

Amazon перетворив свою монолітну архітектуру на мікросервісну, щоб забезпечити масштабування та гнучкість своєї величезної електронної комерції та хмарних сервісів. Даний сервіс використовує AWS для розгортання та управління мікросервісами, таких як Elastic Container Service (ECS), Lambda для безсерверних обчислень, та інших сервісів для моніторингу, логування та автоматизації.

Обробкою величезної кількості запитів з мінімальною затримкою, зберігання та аналіз великих обсягів даних займається Google, який використовує мікросервісну архітектуру для різних своїх продуктів, включаючи пошук, Gmail, Maps. Їхня система розроблена для високої продуктивності, масштабування та надійності. В роботі використовуються власні технологічні розробки для управління розподіленими системами, таких як Borg для оркестрації контейнерів, та інших інструментів для мікросервісів, моніторингу, логування.

Uber перейшов на мікросервісну архітектуру, щоб впоратися зі швидким зростанням та потребами в глобальному масштабуванні. Сервіс використовує георозподілених мікросервісів, розробка власних рішень для геопросторових обчислень, інтенсивне застосування контейнеризації та оркестрації.

Таким чином, кожен мікросервіс управляє своїми даними незалежно, що забезпечує гнучкість та високу доступність. Інтенсивне використання CI/CD для автоматизації тестування, розгортання та моніторингу. Використання контейнерів та оркестрації типу Docker, Kubernetes та інших інструментів значно спрощують розгортання, масштабування та управління мікросервісами. Забезпечення видимості та прозорості всієї системи через централізоване логування та моніторинг. Використання API шлюзів для спрощення зв'язку між мікросервісами та клієнтами, а також для забезпечення безпеки, маршрутизації запитів, обробки помилок. Мікросервісна архітектура дозволяє створювати високоадаптивні та масштабовані системи, здатні ефективно вирішувати завдання різної складності.

Сервісно-орієнтована архітектура (SOA) для розподілених систем орієнтована на використання веб-сервісів та інших технологій для забезпечення модульності, масштабованості, та гнучкості систем. У таких системах компоненти, або "сервіси", надають стандартизовані інтерфейси для взаємодії, що дозволяє легко інтегрувати та перевикористовувати їх у різних додатках та бізнес-процесах. AWS надає широкий спектр хмарних сервісів, кожен з яких може вважатися компонентом у SOA. Сервіси охоплюють обчислення, зберігання даних, бази даних, мережеві функції, і багато іншого, дозволяючи розробникам легко інтегрувати потрібні компоненти у свої додатки. Salesforce використовує SOA у своїй платформі для керування відносинами з клієнтами. Через API Salesforce надає доступ до своїх сервісів, таких як продажі, обслуговування клієнтів, маркетинг тощо, що дозволяє інтегрувати ці сервіси з іншими бізнес-додатками. SAP пропонує комплексні бізнес-рішення, які також базуються на SOA, дозволяючи компаніям оптимізувати свої бізнес-процеси. SAP

NetWeaver дозволяє інтегрувати дані та бізнес-процеси з SAP додатками та зовнішніми системами.

SOA спрощує інтеграцію різних сервісів і систем, незалежно від їхньої реалізації та місця розташування, завдяки використанню стандартизованих протоколів і форматів даних. Системи, побудовані за принципами SOA, складаються з незалежних сервісів, які можуть бути розроблені, розгорнуті, і вдосконалені окремо один від одного. SOA дозволяє легко масштабувати системи, додаванням або модифікацією сервісів, для задоволення зростаючих потреб бізнесу. Розподіленість сервісів у SOA сприяє відмовостійкості системи, дозволяючи їй продовжувати роботу навіть у разі збоїв окремих компонентів.

Таким чином, SOA залишається важливим підходом у розробці розподілених систем, оскільки він пропонує значну гнучкість та ефективність при інтеграції різноманітних сервісів та додатків. Водночас успіх впровадження SOA залежить від уважного планування, управління змінами, та заходів безпеки.

Архітектура з шарами, або багаторівнева архітектура, є традиційним підходом до проектування розподілених систем, який включає розділення функціоналу програми на окремі рівні, кожен з яких має свою специфічну відповідальність. Такий підхід дозволяє досягти модульності, спрощує розробку та тестування, а також полегшує масштабування та адміністрування системи. Презентаційний шар відповідає за взаємодію з користувачами системи, представлення даних та обробку введення користувача. Веб-інтерфейси, мобільні додатки чи настільні застосунки є прикладами компонентів цього шару. Бізнес-логіка (Business Logic Layer): Містить основну функціональність системи, виконує обробку даних згідно з правилами бізнесу. Цей шар взаємодіє з презентаційним шаром та шаром доступу до даних, обслуговуючи їх запити. Шар доступу до даних забезпечує абстракцію та управління доступом до даних, збережених у базі даних або інших сховищах. Це може включати виконання запитів до бази даних, обробку транзакцій та керування підключеннями. Шар даних зберігає дані, з якими працюють інші шари системи. Архітектура з шарами залишається популярним вибором для розробки як простих, так і складних розподілених систем, завдяки її гнучкості, модульності та здатності до масштабування.

Peer-to-Peer (P2P) архітектура розподілених систем відіграє ключову роль у створенні мереж, де кожен вузол (peer) функціонує як клієнт, так і сервер для інших вузлів. Це контрастує з традиційними клієнт-серверними архітектурами, де централізований сервер надає ресурси, а клієнти їх споживають. P2P архітектури забезпечують різноманітні переваги, але також стикаються з певними викликами. P2P мережі не мають єдиної точки відмови, на відміну від централізованих систем, що робить їх більш відмовостійкими та надійними. Додавання нових вузлів до P2P мережі покращує загальну продуктивність системи, оскільки кожен новий вузол збільшує доступну пропускну спроможність та обчислювальні ресурси. P2P мережі ефективно розподіляють ресурси серед великої кількості вузлів, зменшуючи навантаження на окремі вузли та забезпечуючи більш рівномірне

використання ресурсів. Вузли в P2P мережах можуть динамічно приєднуватися та від'єднуватися без значного впливу на роботу мережі в цілому. P2P мережі, такі як BitTorrent, дозволяють ефективно обмінюватися файлами між великою кількістю користувачів, зменшуючи навантаження на індивідуальні сервери. Біткойн та інші криптовалюти використовують P2P мережі для розподіленого ведення реєстру транзакцій (блокчейн), забезпечуючи високий рівень безпеки та незалежність від централізованих фінансових установ. Деякі децентралізовані соціальні мережі побудовані на P2P архітектурі для забезпечення більшої приватності та контролю користувачів над їхніми даними. Платформи на кшталт Golem, використовують P2P мережі для створення децентралізованих обчислювальних мереж, що дозволяє користувачам "орендувати" невикористані обчислювальні ресурси інших учасників. P2P архітектура продовжує розвиватися, надаючи нові можливості для створення масштабованих, відмовостійких та ефективних розподілених систем. Незважаючи на виклики, пов'язані з безпекою та управлінням, переваги, які надає P2P, роблять цей підхід привабливим для широкого спектру застосувань.

Архітектури, основані на подіях (event-driven architectures), є ключовим компонентом у проектуванні розподілених систем, забезпечуючи високу гнучкість, масштабованість та реактивність. У таких архітектурах компоненти системи взаємодіють один з одним за допомогою подій, що дозволяє їм залишатися слабо зв'язаними та асинхронно реагувати на зміни стану або зовнішні запити. Компоненти системи обробляють події асинхронно, що зменшує залежності між ними та покращує загальну продуктивність. Взаємодія між компонентами заснована на подіях, що зменшує прямі залежності та спрощує заміну або модифікацію окремих компонентів. Систему легко розширити, додавши нові обробники подій для реагування на існуючі або нові події. Асинхронна обробка подій може сприяти більшій відмовостійкості, оскільки система здатна продовжувати роботу навіть при виході з ладу окремих компонентів. Використовують таку архітектуру системи, такі як Apache Flink або Spark Streaming, що підтримують обробку поточкових даних в реальному часі, використовуючи події як основні одиниці даних. Архітектури, основані на подіях, надають потужний механізм для побудови розподілених систем, які вимагають високої гнучкості, масштабованості та реактивності. Водночас, ефективне використання таких архітектур вимагає уважного планування та розуміння потенційних викликів.

Консистентне хешування є фундаментальною технікою у проектуванні розподілених систем, особливо тих, що вимагають масштабованості, високої доступності та гнучкого управління ресурсами. Вперше запропоноване як спосіб розподілу запитів між кеш-серверами у веб-мережах, консистентне хешування широко використовується в системах кешування, базах даних NoSQL, розподілених файлових системах та інших додатках. Консистентне хешування використовує кільцевий простір хешу, де кожен вузол та кожен ключ даних отримують хеш-значення, яке визначає їх позицію на

хеш-кільці. Ключі даних призначаються вузлу на кільці, що знаходиться найближче за годинниковою стрілкою. Це забезпечує розподіл даних між вузлами. При додаванні або видаленні вузлів з системи, лише невелика частина ключів потребує перерозподілу. Це значно спрощує масштабування системи, мінімізуючи вплив на доступність даних та продуктивність. Для забезпечення більш рівномірного розподілу даних вузли можуть мати кілька віртуальних хеш-позицій на кільці. Це допомагає уникнути проблеми "гарячих точок", коли один вузол обслуговує непропорційно велику частину запитів. Консистентне хешування залишається фундаментальною технікою у розробці ефективних, масштабованих та відмовостійких розподілених систем, пропонуючи унікальні рішення для викликів, пов'язаних з розподілом та управлінням даними.

Висновки

Вибір архітектурного рішення для розподілених систем є критичним етапом, що вимагає ретельного аналізу вимог до системи, цілей бізнесу та очікуваних викликів. Вибір правильної архітектури може значно

покращити продуктивність, масштабованість, відмовостійкість та легкість управління системою. Проведено аналіз архітектурних рішень для розподілених систем зберігання та обробки даних, який показав, що мікросервіси ідеально підходять для комплексних застосунків, що потребують швидкого розвитку, масштабованості та легкої підтримки; сервісно-орієнтована архітектура підходить для інтеграції різноманітних бізнес-додатків та сервісів у єдину екосистему; архітектура з шарами підходить для традиційних веб-додатків із чітким розділенням логіки, даних і інтерфейсу користувача; P2P ідеально підходить для додатків, які потребують високої відмовостійкості, децентралізації та розподілених обчислень; архітектура, яка основана на подіях підходить для систем, що потребують високої реактивності та асинхронної обробки подій. Вибір архітектурного рішення для розподілених систем - це комплексний процес, який вимагає балансування між технічними, бізнесовими та операційними вимогами. Врахування майбутнього зростання, потенційних викликів і гнучкості системи є ключовими факторами для забезпечення її довгострокового успіху.

СПИСОК ЛІТЕРАТУРИ

1. Tanenbaum, Andrew S., and Maarten van Steen. *Distributed Systems: Principles and Paradigms*. 3rd ed., Pearson, 2017. p.17-30.
2. Thomas Erl, Benjamin Carlyle, Cesare Pautasso, Raj Balasubramanian. *SOA with REST*. – Prentice Hall, 2013.
3. Kovalenko, A., Kuchuk, H., Kuchuk, N. and Kostolny, J. (2021), "Horizontal scaling method for a hyperconverged network", *2021 Int. Conf. on Information and Digital Technologies (IDT)*, Zilina, doi: <https://doi.org/10.1109/IDT52577.2021.9497534>
4. Panda, Dhableswar K.; Sayantan Sur. *Network Speed Acceleration with IB and HSE. Designing Cloud and Grid Computing Systems with InfiniBand and High-Speed Ethernet*. Newport Beach, CA, USA: CCGrid – 2011. p. 23.
5. HWM Singapore. *An introduction to network attached storage*. SPH Magazines – 2003. pp. 90–92.
6. Wang, Yandong; Goldstone, Robin; Yu, Weikuan; Wang, Teng. *Characterization and Optimization of Memory-Resident MapReduce on HPC Systems*. IEEE 28th International Parallel and Distributed Processing Symposium. IEEE – 2014. pp. 799–808. doi:10.1109/IPDPS.2014.87

Received (Надійшла) 25.04.2024

Accepted for publication (Прийнята до друку) 17.07.2024

Analysis of the functioning of distributed data processing and storage systems

Anton Bilokon, Stanislav Borisov, Maxim Usatenko, Volodymyr Fedorchenko

Abstract. Relevance. As the amount of data generated by users, IoT devices, social media, business processes, etc. grows, the need for scalable storage solutions becomes more and more evident. Distributed systems allow for efficient scaling, providing an increase in storage volumes and computing power without significant costs. Modern business requires high availability and reliability of systems, because even minimal downtime can lead to significant financial losses and a decrease in customer confidence. Distributed systems provide high availability and resilience by automatically recovering from failures and replicating data to ensure data integrity. Globalization of business requires working with data in different geographical locations. Distributed systems allow localizing data storage and processing closer to end users, reducing delays and increasing the overall performance of systems. Increasing security threats and increasing regulatory requirements for data protection are forcing organizations to look for more secure data storage solutions. Distributed systems offer enhanced capabilities for data encryption, access control, auditing, and regulatory compliance. Processing large amounts of data often requires large computing power. Distributed data storage systems are ideal for working together with distributed computing, such as streaming data processing, machine learning, big data, allowing to efficiently distribute tasks and process large amounts of information. Among the challenges that distributed data storage systems may face are - ensuring data consistency between nodes, network delay management, data protection and security. Various strategies and technologies are used to address these challenges, including consistent hashing algorithms, data replication, transaction protocols with guaranteed atomicity, consistency, isolation, and durability, and sequential consistency models. Thus, in conditions of constant growth of data volumes and increasing requirements for their processing, distributed data storage systems are a key element of the infrastructure of any organization striving for innovation and efficiency. **The purpose** of this work is to analyze the functioning of distributed data processing and storage systems. **The object** of research is distributed data processing and storage systems. **The subject** of research is architectural solutions of distributed data storage and processing systems. **The results.** The analysis of the functioning of distributed data processing and storage systems was carried out. The choice of an architectural solution for a distributed system depends on the specifics of the tasks to be solved, requirements for performance, scalability, reliability and availability. Usually, effective distributed systems use a combination of these approaches to achieve optimal results. Choosing an architectural solution for distributed systems is a complex process that requires balancing technical, business, and operational requirements. Consideration of future growth, potential challenges and flexibility of the system are key factors to ensure its long-term success.

Keywords: distributed information system, data processing, Web-Scale, data storage, data processing center, event-driven architectures.

С. Я. Бовчалюк, Я. А. Гайдай

Харківський національний технічний університет радіоелектроніки, Харків, Україна

АНАЛІЗ МЕТОДУ ОПОРНИХ ВЕКТОРІВ У ПОРІВНЯННІ З ТРАДИЦІЙНИМИ МЕТОДАМИ ПЕРЕДБАЧЕННЯ РИНКОВИХ РУХІВ

Анотація. **Актуальність.** Розробка та налаштування універсального алгоритму прийняття ринкових рішень є першочерговою задачею експертів та інвесторів фондових ринків з усього світу. Сучасні технології пропонують актуальні рішення з використанням нейронних мереж та штучного інтелекту. Однак існуючі рішення підходять лише для обмеженого використання та потребують великих об'ємів навчальних даних. **Метою даної роботи** є аналіз можливості застосування методу опорних векторів для прийняття ринкових рішень, у порівнянні з традиційними стратегіями та методами аналізу ринкових рухів. **Об'єктом дослідження** є процес прийняття ринкових рішень та використання нейронних мереж. **Предметом дослідження** є метод опорних векторів. **Результати.** У роботі було проаналізовано метод опорних векторів на ринкових моделях за різних підходів і модифікацій та запропоновано шляхи підвищення ефективності прийняття рішень. **Висновок.** Метод опорних векторів продемонстрував більшу ефективність та надійність за класичні методи аналізу ринку, на моделях з великою волатильністю. Даний алгоритм показує позитивні результати на кризових та нестабільних моделях ринку, що робить його використання оптимальним для збереження капіталу.

Ключові слова: фондові ринки, нейронні мережі, генетичний алгоритм, економіка, штучний інтелект, система підтримки ринкових рішень.

Вступ

Антиципація рухів на фондовому ринку становить значний виклик для експертів через його критичне значення. Щодня тисячі осіб у всьому світі займаються інвестуванням на фондовому ринку. Прогнозування точних патернів і цін на акції стає складним завданням у сфері фондових спекуляцій. Ці патерни безпосередньо або опосередковано залежать від різних факторів, таких як новини, громадський настрій, політичні події тощо. Тому використання лише історичних даних для прогнозування цін на акції є вже недостатнім. Незважаючи на складність прогнозування цін на акції, вирішення цієї задачі залишається важливим для автоматизації прийняття обґрунтованих рішень, мінімізації спекулятивних ризиків і максимізації прибутків з мінімальним капіталовкладенням. Протягом років було розроблено безліч традиційних прогностичних моделей [1, 7, 8], таких як індикатор MACD (Moving Average Convergence Divergence), авторегресійне інтегроване ковзне середнє (ARIMA), індекс відносної сили (RSI), тощо. Однак у останні десятиліття дослідники все більше зосереджуються на підходах машинного навчання, таких як штучні нейронні мережі та мережі зворотного поширення. Прогнозування лінійних часових рядів досить просте порівняно з нелінійними часовими рядами.

Серед різних підходів машинного навчання метод опорних векторів [3, 9] (SVM) виділяється своєю здатністю роботи з обома видами часових рядів, а також до надання більш точних результатів, особливо у роботі з нелінійними часовими рядами. Хоча SVM використовується переважно для класифікації об'єктів, розпізнавання структур та регресійного аналізу, його основним принципом є мінімізація структурного ризику, на відміну від інших підходів машинного навчання, що базуються на принципі емпіричної мінімізації ризику. SVM зменшує помилки класифікації на основі раніше отриманих даних та оцінює функцію

зменшення верхньої межі узагальнення. Шляхом ідентифікації оптимальних розташувань з мінімальними тренувальними наборами на основі геометричних властивостей простору ознак алгоритм демонструє високий рівень узагальнення. Навпроти, традиційні методи та інші підходи машинного навчання часто потребують великої кількості тренувальних даних для оцінки вхідних патернів, що робить узагальнення результатів важким через їх схильність до перенавчання [10]. Дана стаття демонструє аналіз роботи та переваг методу опорних векторів стосовно інших підходів машинного навчання.

Метою цієї роботи є аналіз методу опорних векторів у порівнянні з традиційними стратегіями прийняття ринкових рішень.

Основна частина

У дослідженні, проведеному Шомом Прасад Дасом і Сударсаном Падхі [2], для побудови прогностичної моделі використовувалися два підходи машинного навчання: метод опорних векторів SVM (рис. 1) та нейронні мережі зі зворотним поширенням помилки BPNN (рис. 2).

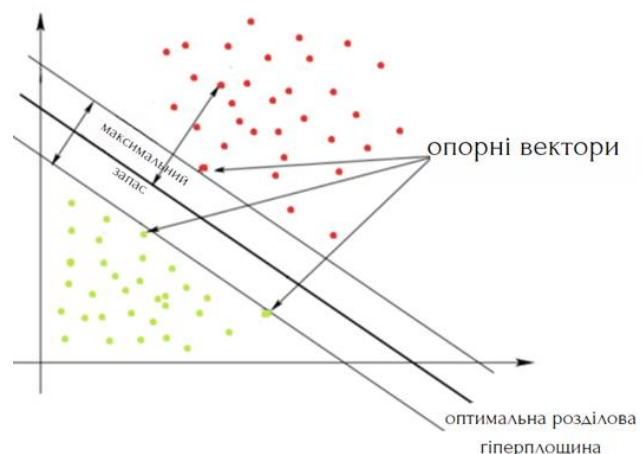


Рис. 1. Схема роботи методу опорних векторів

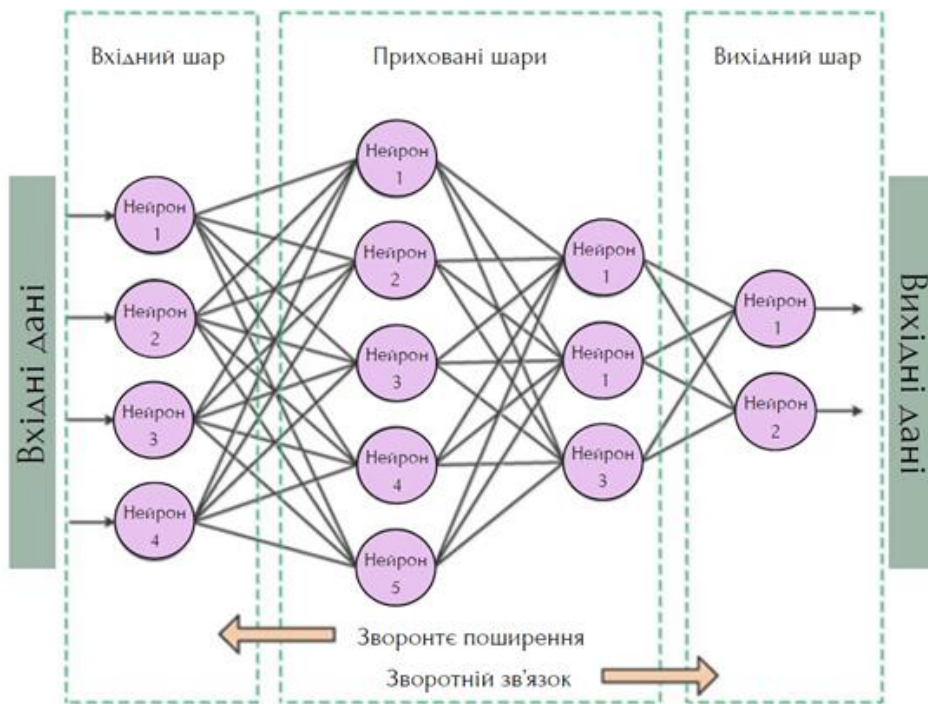


Рис. 2. Схема роботи нейронної мережі зі зворотним поширенням помилки

Набір даних складався з вибірки щоденних цін відкриття та закриття, максимальних та мінімальних цін, а також обсягу та вартості угод більш ніж за сім років.

Вхідні значення визначалися на основі чотирьох значень RDP (відносна відсоткова різниця) у п'ятиденному проміжку. Експоненційна середня рухома змінна використовувалася для збереження інформації про фактичну ціну закриття, оскільки застосування відносної відсоткової різниці може призвести до втрати цінної інформації. Було визначено, що прогностична ефективність нейронних мереж покращується, коли застосування згладжування переходить у залежну змінну. Тому за допомогою триденної експоненційної середньої рухоми змінної було надано вихідне значення відносній відсотковій різниці.

Для нормалізації значень даних також використовувалось масштабування, щоб розмістити їх у діапазоні від -0.9 до 0.9, охоплюючи як позитивні, так і негативні значення. Якість прогнозу оцінювалася з використанням статистичних показників, таких як нормалізована середньоквадратична помилка (NMSE), спрямована симетрія (DS) та середня абсолютна помилка (MAE). MAE та NMSE кількісно визначають відхилення між фактичними та прогнозованими значеннями. Після навчання було виявлено, що SVM перевершує BPNN за різними показниками.

У своєму дослідженні Каратанасопулос та співавтори [4] представили еволюційний метод опорних векторів ESVM [5], як модель для прогнозування поведінки акцій. ESVM поєднує в собі метод опорних векторів (SVM) та генетичні алгоритми (GA). В той час як SVM зазвичай використовуються для аналізу даних та розпізнавання патернів у завданнях класифікації, генетичний алгоритм викори-

стовується в цій моделі для оптимізації параметрів SVM та вибору найбільш значущих ознак. Модель пропонується з метою уникнення перенавчання, застосовуючи генетичний алгоритм для модифікації властивостей моделі SVM та вибору оптимальних вхідних даних для досягнення найвищої статистичної та торговельної ефективності. Ефективність такого підходу обумовлена здатністю працювати у високовимірних просторах, покращеними можливостями інтеграції та здатністю виявляти прибуткові торговельні стратегії. Порівняльний аналіз з п'ятьма традиційними методами продемонстрував високі результати як з точки зору торговельних, так і статистичних показників. Зокрема, ESVM проявив вражаючу торговельну ефективність у відсотковому вираженні річних доходів, завдяки здатності спростувати класифікатори SVM і при цьому забезпечувати високу торговельну ефективність як у тестовому середовищі, так і на реальному наборі даних.

Росілло Гінер та Де Ла Фуенте [6] досліджували використання відносного індексу сили (RSI) та значення збігу і розходження середнього руху (MACD) як входів для SVM, при цьому RSI був налаштований для акцій великого капіталу, а MACD - для акцій малого капіталу. Класифікатор SVM був обраний для кількісного прийняття рішень через його перевагу перед іншими підходами машинного навчання в допомозі інвесторам. Тестування проводилося протягом періодів у 200, 250, 300 та 500 днів, з яких період тривалістю 250 днів показав оптимальні результати. Тому дані за 250 днів були обрані як тренувальні дані, при цьому значення на 5 днів вперед враховувалися при прийнятті рішень щодо торгівлі акціями. Спочатку SVM категоризує вхідні дані, а потім ділить їх на два класи для купівлі та продажу. В якості ядра для SVM була використана функція важкої радіальної базисної (HTRBF).

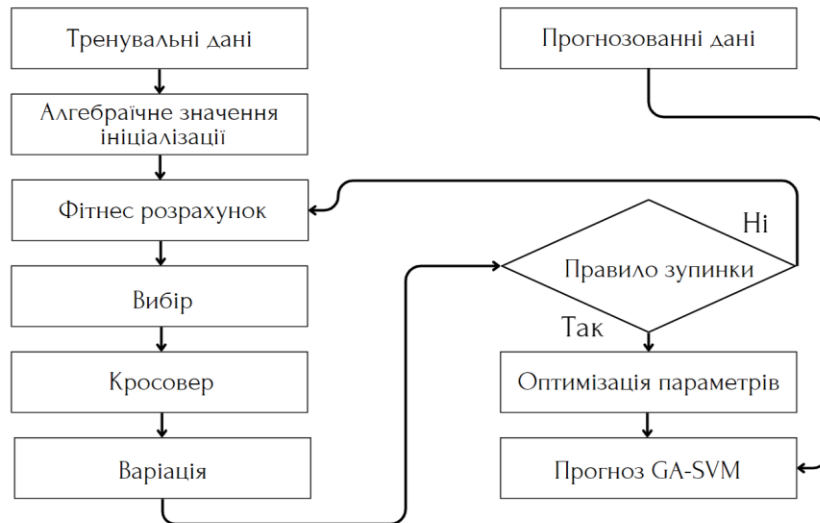


Рис. 3. Схема роботи еволюційного методу опорних векторів

Представлені результати у табл. 1 і 2 демонструють порівняльні значення отримані під час симуляцій на бичачому і ведмежому ринках, а результати табл. 3 і 4, показують значення отримані під час симуляцій на моделі з високою і низькою волатильністю відповідно.

У таблицях порівнюються метод опорних векторів (SVM) з методом купівлі і утримання (ВН) та зі наївною стратегією (N). Метод купівлі і утримання полягає у придбанні інвестицій і зберіганні їх протягом заданого часу, у цьому дослідженні мінімальний період утримання склав 5 днів.

Таблиця 1 – Результат симуляції на бичачому ринку

	SVM	ВН	N
Симуляція_1	337.0	800.1	-425.5
Симуляція_2	-303.3	1147.3	-592.5
Симуляція_3	148.8	377.1	-193.8
Середнє значення	60.8	774.8	-403.9

Таблиця 2 – Результат симуляції на ведмежому ринку

	SVM	ВН	N
Симуляція_1	153.1	-17.0	-425.5
Симуляція_2	104.8	-389.9	-592.5
Симуляція_3	-9.0	377.1	-193.8
Середнє значення	83.0	-216.1	-187.6

Таблиця 3 – Результат симуляції з високою волатильністю

	SVM	ВН	N
Симуляція_1	-284.8	655.8	297.0
Симуляція_2	2599.4	22.8	789.5
Симуляція_3	-288.9	-755.5	-458.1
Середнє значення	867.8	-25.6	209.5

Таблиця 4 – Результат симуляції з низькою волатильністю

	SVM	ВН	N
Симуляція_1	145.2	-43.9	-84.4
Симуляція_2	9.1	-181.8	-3.8
Симуляція_3	-111.8	143.1	33.8
Середнє значення	14.2	-27.5	-18.2

Наївна стратегія в свою чергу приймає зміни за останній ринковий період, як основу для подальшого прогнозування. Результати представлені у вигляді умовних одиниць, які демонструють продуктивність прийнятих ринкових рішень, та не мають валютного еквіваленту.

Порівняльний аналіз з базовими стратегіями торгівлі показав, що метод опорних векторів демонструє кращий результат на моделях ринку як з високою волатильністю, так і низькою. Також, застосування методу опорних векторів на моделі ведмежого ринку продемонструвало єдиний позитивний результат серед інших стратегій, що є безумовною перевагою даного алгоритму. Однак, на бичачому ринку стратегія купівлі і утримання показала найкращий результат, що є закономірною ситуацією, з урахуванням специфікації та призначення цього методу аналізу ринку. Під час дослідження метод опорних векторів продемонстрував загальну позитивну продуктивність на кожній ринковій моделі. Було доведено перевагу SVM та його здатність знизити максимальні збитки і річне стандартне відхилення.

Метод опорних векторів має високу ефективність у порівнянні зі стандартними стратегіями прийняття ринкових рішень. Використання цього методу рекомендовано на ринках з високою волатильністю та ринках зі спадаючим трендом. Використання SVM є найкращим вибором стратегії під час фінансової нестабільності та може зберегти інвестиції під час тривалих ведмежих трендів, або навіть зафіксувати прибутки під час кризових ситуацій. Однак метод не

є достатньо чутливим до маленьких цінових змін, що робить його не рекомендованим до використання на ринках з низькою волатильністю. Для покращення роботи алгоритму рекомендується використовувати індикатор тренду, як одне із значень вхідних даних, для вчасного визначення переходу ринку у стан високої або низької волатильності.

Об'єднання методу опорних векторів з експертною системою оцінки предметної області, гарантовано знизить ризики втрати капіталу під час ринкових операцій [11]. У ролі експертної системи можуть виступати уже існуючі рішення, такі як Neural Forex Network, Algo Trader тощо. Більшість таких систем також базується на роботі штучного інтелекту. Поєднання підвищить надійність алгоритму шляхом надання вхідних значень відповідних до специфічної предметної області. Серед цих значень можуть бути такі, що впливають на волатильність ринку, наприклад політичні події, економічні фактори, рішення центрального банку тощо. Алгоритм може використовувати ці дані для прийняття більш обґрунтованих рішень та оптимізації торговельних стратегій, що в свою чергу покращить ефективність прогнозування.

Для підвищення ефективності навчання нейронної моделі рекомендовано використовувати віднос-

ний індекс сили та значення збігу і розходження на входах для SVM. Також продуктивність нейронної мережі можливо підвищити використовуючи залежну змінну для згладжування, наприклад експоненційну рухому змінну.

Висновки

Валютні ринки відіграють важливу роль у сфері бізнесу, тому прогнозування рухів валютного ринку є складним завданням для розробників. Існує кілька традиційних моделей для передбачення, проте вони часто не досягають точності і вимагають значної кількості даних. У порівнянні з традиційними моделями та іншими підходами машинного навчання метод опорних векторів показує кращі результати при меншій кількості даних. SVM виявляється ефективним як у лінійних, так і у нелінійних сценаріях даних. Крім того, еволюційний метод опорних векторів, як гібридний підхід, проявляє найкращу ефективність з урахуванням річних доходів. У висновку це дослідження стверджує, що підхід SVM є оптимальною стратегією для прогнозування тенденцій на валютному ринку у порівнянні з традиційними моделями та іншими підходами машинного навчання.

СПИСОК ЛІТЕРАТУРИ

1. Iacomin, R. (2015), "Stock market prediction", 2015 19th International Conference on System Theory, Control, and Computing
2. Prasad Das (2012). Support Vector Machines at Prediction of Futures Costs in Indian Stock Market. International Journal of Computer Applications.
3. F.E.H. Tay *et al.* "Application of support vector machines in financial time series forecasting", Omega (2001)
4. Karathanasopoulos (2016). "Stock market anticipation utilizing evolutionary support vector machines: an application of the ASE20 index", The European Journal of Finance
5. Yaloveha V., Podorozhniak A., Kuchuk H. Convolutional neural network hyperparameter optimization applied to land cover classification. *Radioelectronic and computer systems*. 2022. No. 1. С. 115–128. DOI: <https://doi.org/10.32620/reks.2022.1.09>
6. Rosillo (2014), "Stock market simulation utilizing support vector machines. Journal of Forecasting"
7. M.T. Leung *et al.* "Forecasting stock indices a comparison of classification and level estimation models", International Journal of Forecasting (2000)
8. Y.S. Abu-Mostafa *et al.* "Introduction to financial forecasting", *Applied Intelligence* (1996)
9. Studies, E. (2000). A Stock Market Prediction Method Based on Support Vector Machines and Independent Component Analysis
10. Yetis, Y., Kaplan, H., & Jamshidi, M. (2014), "Stock market prediction by using the artificial neural network", 2014 World Automation Congress
11. Ichinose, K., & Shimada, K. (2016), "Stock market prediction from news on the web and a new evaluation approach in trading. Proceedings" 2016 5th IIAI International Congress on Advanced Applied Informatics

Received (Надійшла) 16.03.2024

Accepted for publication (Прийнята до друку) 05.06.2024

Analysis of the support vector machine algorithm in comparison to traditional market movements prediction methods

Stanislav Bovchaliuk, Yaroslav Haidai

Abstract. Topicality. The development and adjustment of a universal algorithm for making market decisions is the primary task of experts and investors of stock markets around the world. Modern technologies offer relevant solutions using neural networks and artificial intelligence. However, existing solutions are only suitable for limited use and require large amounts of training data. The goal of this work is analysis of the support vector machine performance on market decisions adjustment, and its comparison with traditional strategies and methods of analysis of market movements. The object of research is the process of making market decisions based on the neural networks. The subject of research is the support vector machine algorithm. Results. In this paper, was analyzed the support vector machine algorithm on market models for various approaches and modifications, also were offered ways to improve the efficiency of the decisions made. Conclusions. The support vector machine has demonstrated greater efficiency and reliability than classical methods of analysis, on models with high market volatility. This algorithm shows positive results in crisis and unstable market models, its use is optimal for capital preservation.

Keywords: stock markets, neural networks, genetic algorithm, economy, artificial intelligence, market decisions support system.

А. С. Бурда, М. А. Прудіус, Я. Г. Стефанюк, О. О. Фомічов

Харківський національний технічний університет радіоелектроніки, Харків, Україна

МЕТОДИ ОБРОБКИ ТА ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ З ВИКОРИСТАННЯМ ШТУЧНИХ ІМУННИХ СИСТЕМ

Анотація. Актуальність. Штучні імунні системи, завдяки своїй здатності працювати в розподіленому середовищі, можуть бути ефективно використані для виявлення та запобігання вторгненням у мережі. Штучні імунні системи здатні адаптуватися до змінюваних умов та нових загроз. Це особливо актуально в умовах швидких технологічних змін, де традиційні методи часто вимагають тривалого перенавчання. Вони активно використовуються для виявлення та запобігання кіберзагрозам. Вони можуть аналізувати мережевий трафік, виявляти аномальні дії та забезпечувати захист у режимі реального часу. Таким чином, актуальність використання штучних імунних систем для обробки та класифікації даних полягає у їх унікальних властивостях адаптивності, здатності до розпізнавання складних патернів та аномалій, а також ефективній роботі в розподілених системах. **Метою даної роботи** є дослідження існуючих методів обробки та штучного аналізу даних з використанням штучних імунних систем. **Об'єктом дослідження** є інтелектуальний аналіз даних штучними імунними системами. **Предметом дослідження** є методи інтелектуального аналізу даних штучними імунними системами. **Результати.** Проведено аналіз існуючих методів обробки та інтелектуального аналізу даних з використанням штучних імунних систем. Імунні алгоритми кластеризації є потужним інструментом для аналізу та обробки даних. Вони дозволяють ефективно групувати дані, виявляти аномалії та адаптуватися до змін у середовищі даних, що робить їх цінними для широкого спектра застосувань, від маркетингу та медицини до фінансів і промисловості. Артилерійні алгоритми є потужними інструментами для інтелектуального аналізу даних, які пропонують ефективні, точні та продуктивні методи оптимізації та аналізу. Вони знаходять застосування у багатьох галузях, від фінансів і логістики до медицини та кібербезпеки, забезпечуючи вирішення складних задач обробки даних і прийняття рішень. Дендритні клітинні алгоритми є потужним інструментом для інтелектуального аналізу даних, зокрема для виявлення аномалій і класифікації даних. Їх здатність агрегувати різні типи сигналів і приймати рішення на основі загального рівня небезпеки робить їх особливо ефективними в складних і динамічних середовищах, таких як кібербезпека та біоінформатика. Вони забезпечують високу чутливість і адаптивність, що дозволяє їх успішно застосовувати у різних сферах для вирішення задач обробки та аналізу даних.

Ключові слова: штучна імунна система, алгоритми негативного відбору, імунні алгоритми кластеризації, артилерійні алгоритми, дендритні клітинні алгоритми, кібербезпека, афінність, клонадний відбір, патерн.

Вступ

Штучні імунні системи (ШИС) – це обчислювальні моделі, натхнені біологічною імунною системою, які використовуються для вирішення різноманітних задач, включаючи обробку та класифікацію даних. Основними компонентами ШИС є імунні алгоритми [1], які імітують процеси розпізнавання і адаптації, що відбуваються в імунній системі живих організмів. Методами ШИС є імунні алгоритми негативного відбору (ІАНВ), імунні алгоритми кластеризації (ІАК), артилерійні алгоритми (АА), дендритні клітинні алгоритми (ДКА) [2].

ІАНВ імітують процес видалення незрілих Т-лімфоцитів, які розпізнають власні клітини організму (self-cells). У контексті обробки даних використовуються для виявлення аномалій та невідомих даних. ІАНВ є однією з ключових технік штучних імунних систем, що моделюють механізм негативного відбору в біологічній імунній системі. Вони призначені для виявлення аномалій або невідомих патернів у даних шляхом навчання системи розрізняти "свої" (нормальні) дані від "чужих" (аномальних) даних.

ІАК базуються на процесі клонального відбору та афінної зрілості В-лімфоцитів. Ці алгоритми створюють та зберігають популяцію антитіл, що розпізнають специфічні антигени (патогени). ІАК також є частиною штучних імунних систем, які моделюють процес клонального відбору та афінної зрілості в біологічній імунній системі. Ці алгоритми викорис-

товуються для кластеризації даних, оптимізації та інших задач, що вимагають розпізнавання патернів та адаптації.

АА моделюють взаємодію між антитілами у мережі [3]. АІН використовуються для кластеризації даних, де мережа антитіл саморегулюється та адаптується до вхідних даних. АА, або штучні імунні мережі (Artificial Immune Networks, АІН), моделюють взаємодію між антитілами в біологічній імунній системі, створюючи мережу антитіл, яка здатна до самоорганізації та адаптації. Ці алгоритми використовуються для кластеризації даних, виявлення схожих структур і патернів у даних, а також для оптимізації та прогнозування [4, 5].

ДКА моделюють роботу дендритних клітин [6], які відповідають за презентацію антигенів Т-лімфоцитам. ДКА використовуються для виявлення аномалій та класифікації даних. ДКА [7] – це інноваційний підхід у сфері штучних імунних систем, який моделює роботу дендритних клітин у біологічній імунній системі. Ці алгоритми розроблені для виявлення аномалій і класифікації даних, зокрема у сфері кібербезпеки та біоінформатики [8–10].

Отже, **метою цієї роботи** є дослідження існуючих методів обробки та штучного аналізу даних з використанням штучних імунних систем.

Основна частина

Штучні імунні системи – це інноваційна методологія, що базується на принципах біологічної

імунної системи для вирішення складних задач обробки даних. Вони використовуються для класифікації, кластеризації, виявлення аномалій, оптимізації та прогнозування. Основні типи алгоритмів ШІС включають негативний відбір, клональний відбір, штучні імунні мережі та дендритні клітинні алгоритми [11].

ІАНВ є важливою складовою ШІС і моделюють процеси, що відбуваються у біологічній імунній системі. Основна мета цих алгоритмів – виявлення аномалій шляхом навчання на нормальних даних і подальшої перевірки нових даних на предмет їх відповідності.

У біологічній імунній системі негативний відбір відбувається в тимусі, де Т-клітини, які реагують на власні антигени організму, видаляються. Аналогічно, в NSA створюється набір детекторів, які навчаються на нормальних даних і не реагують на них. Ці детектори потім використовуються для виявлення аномалій у нових даних.

На першому етапі створюється набір детекторів, які представляють випадкові можливі стани системи. Ці детектори проходять процес навчання, під час якого вони перевіряються на відповідність "своїм" даним (нормальним зразкам). Детектори, які реагують на нормальні дані, відсіюються (негативний відбір). Це означає, що лише ті детектори, які не реагують на нормальні дані, залишаються в системі.

У біологічній імунній системі негативний відбір відбувається в тимусі, де Т-клітини, які реагують на власні антигени організму, видаляються. Аналогічно, в ІАНВ створюється набір детекторів, які навчаються на нормальних даних і не реагують на них. Ці детектори потім використовуються для виявлення аномалій у нових даних. Після процесу навчання детектори використовуються для моніторингу нових даних. Якщо детектор реагує на нові дані, ці дані вважаються аномальними або підозрілими. Процес моніторингу може відбуватися в режимі реального часу, що дозволяє швидко виявляти аномалії.

Алгоритм ІАНВ передбачає ініціалізацію (створення початкового набору випадкових детекторів), негативний відбір (перевірка кожного детектора на відповідність нормальним даним та відсіювання детекторів, які розпізнають нормальні дані), а також моніторинг (застосування детекторів до нових даних та ідентифікація аномальних даних на основі реакції детекторів). Таким чином, ІАНВ є порівняно простими у реалізації, що робить їх доступними для широкого кола застосувань. Також вони можуть бути адаптовані для роботи з великими обсягами даних і в умовах розподілених систем. Алгоритм може навчатися новим патернам і адаптуватися до змін у даних.

Можливе застосування ІАНВ в кібербезпеці (виявлення вторгнень і аномалій у мережевому трафіку, моніторинг активності користувачів для виявлення підозрілих дій), моніторингу систем (виявлення несправностей у промисловому обладнанні, аналіз логів систем для виявлення аномалій у роботі програмного забезпечення) та медичній діагностиці (виявлення, наприклад, аномалій у кардіограмах або ракових клітин у зображеннях). Імунні алгоритми кластеризації використовують принципи біологічної імун-

ної системи для групування подібних даних. Ці алгоритми особливо корисні для обробки великих обсягів даних, виявлення патернів та аномалій, і є потужним інструментом для аналізу даних у різних сферах. ШІС для кластеризації використовують механізми самоорганізації та адаптації для створення мережі антитіл (або агентів), що представляють кластери даних. Основні принципи роботи включають ініціалізацію мережі антитіл, клонування та мутацію, взаємодію між антитілами, і самоорганізацію мережі.

Алгоритм ІАК містить ініціалізацію мережі антитіл, коли ініціалізується початкова популяція антитіл (або агентів), кожна з яких представляє потенційний кластер, обчислення афінності (афінність антитіла до антигену (даних) визначається на основі відстані або подібності між ними), клонування та мутації (антитіла з високою афінністю клонуються і мутують для створення нових варіантів, що можуть краще відповідати даним), взаємодію та самоорганізацію мережі (антитіла взаємодіють між собою та з даними, утворюючи мережу, яка самоорганізується та адаптується до нових даних), а також оновлення мережі (мережа антитіл постійно оновлюється на основі нових даних та взаємодій у мережі).

Під час дії цього алгоритму відбувається генерація початкової популяції антитіл, кожне з яких представляє можливе рішення або центр кластера. Потім визначається ступінь відповідності (афінності) між антитілами та антигенами (даними). Висока афінність вказує на добру відповідність даним. Антитіла з високою афінністю проходять процес клонування, створюючи копії з невеликими мутаціями. Це дозволяє досліджувати простір можливих рішень навколо антитіла з високою афінністю. Також проходить відбір кращих клонів на основі їх афінності до антигенів. Цей процес повторюється до досягнення оптимального розподілу клонів, що представляють центри кластерів. Дані кластеризуються на основі близькості до обраних центрів кластерів (антитіл).

Перевагами використання ІАК є адаптивність, коли алгоритми можуть адаптуватися до нових даних та змін у даних; гнучкість, при якій ІАК можуть використовуватися для різних задач кластеризації та оптимізації та пошук глобальних оптимумів, що, в свою чергу, дозволяє ІАК ефективно знаходити глобальні оптимуми в просторі рішень. Також слід відзначити, що можливе застосування ІАК для кластеризації різних типів даних, включаючи текстові, числові та зображення; вирішення задач оптимізації яких як оптимізація розкладів, маршрутів та ресурсів; а також для виявлення схожих структур у великих обсягах даних, що корисно для аналізу трендів та патернів. ІАК є потужним інструментом для кластеризації та оптимізації даних завдяки своїй здатності до адаптації, гнучкості та ефективному пошуку глобальних оптимумів. Вони знаходять застосування у різних сферах, включаючи аналіз даних, медицину, промисловість та маркетинг, забезпечуючи ефективні рішення для задач кластеризації та розпізнавання патернів.

Артилерійні алгоритми – це специфічний підхід до інтелектуального аналізу даних, який може використовуватися для оптимізації та покращення продук-

тивності в завданнях обробки великих даних. Цей термін може відноситися до різних алгоритмічних стратегій, але в основному має на увазі точні, ефективні та добре структуровані методи, які застосовуються до розв'язання складних задач у галузі штучного інтелекту та машинного навчання. АА зазвичай включають методи оптимізації, такі як генетичні алгоритми, рій частинок, алгоритми мурашиних колоній, які дозволяють ефективно знаходити найкращі рішення серед великого простору можливих варіантів. Ці алгоритми мають високу продуктивність, що досягається через паралельні обчислення, ефективні структури даних і спеціалізовані методи обробки. Точність і надійність є ключовими характеристиками артилерійних алгоритмів, що дозволяє їм успішно застосовуватися в критичних додатках, таких як фінансове моделювання, кібербезпека та медичний аналіз.

Генетичні алгоритми (ГА) використовують принципи природного відбору та генетики для оптимізації і пошуку рішень. Процес роботи ГА наступний: ініціалізація популяції випадкових рішень, оцінка кожного рішення за допомогою функції придатності, відбір кращих рішень для створення нового покоління через кросовер і мутацію, повторення процесу до досягнення оптимального рішення. Застосовуються ГА для оптимізації торгових стратегій на фінансових ринках та розв'язання задач комівояжера.

Алгоритми рою частинок моделюють поведінку зграї птахів або риби для пошуку оптимальних рішень. Процес роботи включає ініціалізацію популяції частинок з випадковими положеннями та швидкостями, оцінку кожної частинки на основі її положення, оновлення швидкостей і положень частинок на основі їхньої історії і найкращих положень сусідів, повторення процесу до досягнення оптимального рішення. Застосовуються ці алгоритми для оптимізації параметрів машинного навчання, а також розв'язання задач оптимізації у виробництві.

Алгоритми мурашиних колоній використовують принципи поведінки мурах для знаходження найкоротших шляхів і оптимізації. Спочатку відбувається ініціалізація мурах у випадкових позиціях, потім моделюється рух мурах по графу на основі феромонових слідів і евристичної інформації, далі оновлення феромонових слідів на основі якості знайдених рішень. Застосовується даний тип алгоритмів для розв'язання задач маршрутизації, а також для оптимізації логістичних ланцюгів.

Артилерійні алгоритми базуються на ідеях створення та підтримки мережі антитіл, яка може змінювати свою структуру відповідно до вхідних даних. Основні процеси включають ініціалізацію мережі антитіл, тобто створення початкової популяції антитіл, кожне з яких представляє можливе рішення або кластер; афінну зрілість і вибір (антитіла взаємодіють з антигенами (даними) та з іншими антитілами в мережі, що сприяє їх афінній зрілості. Антитіла з високою афінністю до антигенів залишаються в мережі, а антитіла з низькою афінністю видаляються); клонування та мутацію (клонування антитіл з високою афінністю та їх мутація для створення нових антитіл, які можуть краще розпізнавати антигени); само-

організацію мережі. В даному випадку антитіла утворюють мережу, яка самоорганізується на основі взаємодій між антитілами та антигенами. Ця мережа адаптується до змін у вхідних даних.

Перевагами використання АА є: самоорганізація, масштабованість та гнучкість. АА є потужним інструментом для кластеризації, виявлення аномалій, оптимізації та прогнозування завдяки своїм властивостям самоорганізації та адаптивності. Вони знаходять застосування в різних сферах, включаючи аналіз даних, кібербезпеку, медицину та промисловість, забезпечуючи ефективні рішення для складних задач у сучасному інформаційному середовищі.

Дендритні клітини у біологічній імунній системі відіграють ключову роль у виявленні патогенів та ініціації імунної відповіді. Вони збирають інформацію з навколишнього середовища і, на основі отриманих сигналів, активуються або залишаються неактивними. ДКА моделюють цей процес, використовуючи три типи сигналів: сигнали, що свідчать про наявність патогенів (патоген-асоційовані молекулярні патерни), сигнали, що свідчать про нормальні, "безпечні" умови (безпечно-асоційовані молекулярні патерни, сигнали), що свідчать про стрес або пошкодження клітин (небезпечно-асоційовані молекулярні патерни). Алгоритм ДКА включає в себе ініціалізацію, коли створюється популяція дендритних клітин; збір сигналів, при якому кожна клітина збирає сигнали з навколишнього середовища даних; агрегацію сигналів, при якій дендритні клітини агрегують отримані сигнали, визначаючи рівень "небезпеки"; розподіл сигналів, при якому клітини розподіляють зібрані сигнали і приймають рішення про свою активацію на основі загального рівня безпеки; активацію, якщо рівень безпеки перевищує певний поріг, клітини активуються і сигналізують про наявність аномалій, та якщо рівень безпеки низький, клітини залишаються неактивними, що свідчить про нормальні умови; класифікацію активованих клітин, які використовуються для класифікації даних як аномальні або нормальні.

Перевагами ДКА є здатність до виявлення навіть незначних аномалій завдяки багаторівневій агрегації сигналів, можливість адаптуватися до змінюваних умов у середовищі даних, ефективна робота з великими обсягами даних та у розподілених системах. Застосовують ДКА в кібербезпеці, моніторингу систем та біоінформатиці.

Висновки

Проведено аналіз існуючих методів обробки та інтелектуального аналізу даних з використанням штучних імунних систем. Імунні алгоритми негативного відбору є потужним інструментом для виявлення аномалій та класифікації даних. Завдяки своїй здатності адаптуватися до змін та працювати у розподілених середовищах, вони залишаються актуальними в сучасних умовах, забезпечуючи ефективні рішення для моніторингу та аналізу даних у різних сферах, від кібербезпеки до медичної діагностики. Імунні алгоритми кластеризації є потужним інструментом для аналізу та обробки даних. Вони дозволяють ефективно групувати дані, виявляти аномалії та адап-

туватися до змін у середовищі даних, що робить їх цінними для широкого спектра застосувань, від маркетингу та медицини до фінансів і промисловості.

Артилерійні алгоритми є потужними інструментами для інтелектуального аналізу даних, які пропонують ефективні, точні та продуктивні методи оптимізації та аналізу. Вони знаходять застосування у багатьох галузях, від фінансів і логістики до медицини та кібербезпеки, забезпечуючи вирішення складних задач обробки даних і прийняття рішень.

Дендритні клітинні алгоритми є потужним інструментом для інтелектуального аналізу даних, зокрема для виявлення аномалій і класифікації даних. Їх здатність агрегувати різні типи сигналів і приймати рішення на основі загального рівня небезпеки робить їх особливо ефективними в складних і динамічних середовищах, таких як кібербезпека та біоінформатика. ДКА забезпечують високу чутливість і адаптивність, що дозволяє їх успішно застосовувати у різних сферах для вирішення задач обробки та аналізу даних.

СПИСОК ЛІТЕРАТУРИ

1. Lauren M. Sompayrac How the Immune System Works. 4th Edition // Wiley-blackwell. 2012. – 153 p.
2. Kephart, J. O. A biologically inspired immune system for computers". Proceedings of Artificial Life IV: The Fourth International Workshop on the Synthesis and Simulation of Living Systems. MIT Press. 1994. – pp. 130–139.
3. Andrews and Timmis "A Computational Model of Degeneracy in a Lymph Node". Artificial Immune Systems. Lecture Notes in Computer Science. Vol. 4163. 2006. – pp. 164–177. doi:10.1007/11823940_13.
4. Kuchuk, H., Kovalenko, A., Ibrahim, B.F. and Ruban, I. (2019), "Adaptive compression method for video information", International Journal of Advanced Trends in Computer Science and Engineering, vol. 8(1), pp. 66–69, doi: <http://dx.doi.org/10.30534/ijatcse/2019/1181.22019>
5. Худов В.Г. Аналіз відомих методів сегментування зображень, що отримані з бортових систем оптикоелектронного спостереження / В.Г. Худов, Г.А. Кучук, О.М. Маковейчук, А.В. Крижний // Системи обробки інформації, 2016. – Вип. 9 (146). – С. 77-80.
6. Mendao; et al. "The Immune System in Pieces: Computational Lessons from Degeneracy in the Immune System". 2007 IEEE Symposium on Foundations of Computational Intelligence. 2007. – pp. 394–400. doi:10.1109/FOCI.2007.371502.
7. Greensmith, J.; Aickelin, U. "Artificial Dendritic Cells: Multi-faceted Perspectives". Human-Centric Information Processing Through Granular Modelling (PDF). Studies in Computational Intelligence. Vol. 182.2009. – pp. 375–395. doi:10.1007/978-3-540-92916-1_16.
8. Kuchuk G.A., Mohammad A.S. and Kovalenko, A.A. (2011), "Method for reducing data transmission time in a wireless network", *Control, navigation and communication systems*, No. 3 (19), pp. 209-213.
9. Kovalenko, A., Kuchuk, H., Kuchuk, N. and Kostolny, J. (2021), "Horizontal scaling method for a hyperconverged network", *2021 International Conference on Information and Digital Technologies (IDT)*, Zilina, Slovakia, doi: <https://doi.org/10.1109/IDT52577.2021.9497534>
10. Semenov, S., Mozhaiev, O., Kuchuk, N., Mozhaiev, M., Tiulieniev, S., Gnusov, Yu., Yevstrat, D., Chyrva, Yu. & Kuchuk, H. (2022), "Devising a procedure for defining the general criteria of abnormal behavior of a computer system based on the improved criterion of uniformity of input data samples", *Eastern-European Journal of Enterprise Technologies*, 2022, 6 (4(120)), pp. 40–49, doi: <https://doi.org/10.15587/1729-4061.2022.269128>
11. Yaloveha V., Podorozhniak A., Kuchuk H. Convolutional neural network hyperparameter optimization applied to land cover classification. Radioelectronic and computer systems. 2022. No. 1. С. 115–128. DOI: <https://doi.org/10.32620/reks.2022.1.09>

Received (Надійшла) 22.05.2024

Accepted for publication (Прийнята до друку) 31.07.2024

Methods of data processing and intellectual analysis using artificial immune systems

Anatoly Burda, Nikita Prudius, Yaroslav Stefanyuk, Oleksandr Fomichev

Abstract. Relevance. Artificial immune systems, especially algorithms of negative selection and dendritic cells, demonstrate high efficiency in detecting rare and unknown anomalies. Distributed computing systems and networks require effective monitoring and protection methods. Their ability to recognize specific patterns and classify medical images makes them indispensable in these fields. They are also actively used to detect and prevent cyber threats. They can analyze network traffic, detect anomalous activity, and provide real-time protection. Thus, the relevance of using artificial immune systems for data processing and classification lies in their unique properties of adaptability, ability to recognize complex patterns and anomalies, as well as effective work in distributed systems. **The purpose** is to research existing methods of processing and artificial analysis of data using artificial immune systems. **The object** of the study is the intellectual analysis of data by artificial immune systems. **The subject** are the methods of intellectual analysis of data by artificial immune systems. **Results** an analysis of existing methods of data processing and intellectual analysis using artificial immune systems was carried out. Immune clustering algorithms are a powerful tool for data analysis and processing. They allow you to efficiently group data, detect anomalies, and adapt to changes in the data environment, making them valuable for a wide range of applications, from marketing and medicine to finance and industry. Artillery algorithms are powerful tools for intelligent data analysis, offering efficient, accurate and productive methods of optimization and analysis. They find applications in many industries, from finance and logistics to medicine and cyber security, providing solutions to complex data processing and decision-making tasks. Dendritic cell algorithms are a powerful tool for intelligent data analysis, particularly for anomaly detection and data classification. Their ability to aggregate different types of signals and make decisions based on the overall level of danger makes them particularly effective in complex and dynamic environments such as cyber security and bioinformatics. they provide high sensitivity and adaptability, which allows them to be successfully used in various fields to solve data processing and analysis problems.

Keywords: artificial immune system, negative selection algorithms, immune clustering algorithms, artillery algorithms, dendritic cell algorithms, cyber security, affinity, clonal selection, pattern.

М. О. Волк, М. М. Поповкін

Харківський національний університет радіоелектроніки, Харків, Україна

МЕТОДИ МОДЕЛЮВАННЯ МАСШТАБОВАНИХ ХМАРНИХ РЕСУРСІВ

Анотація. В епоху стрімкого розвитку хмарних технологій основною задачею є не лише дослідження існуючих можливостей, але й практичне впровадження рішень, спрямованих на оптимізацію використання хмарних ресурсів. Метою статті є виявлення найбільш ефективних підходів до моделювання хмарних ресурсів, що дозволить організації досягти значного зниження витрат на використання хмарних ресурсів при одночасному забезпеченні високого рівня продуктивності та надійності хмарних сервісів. В роботі пропонуються ефективні стратегії автоматизації розгортання та управління хмарною інфраструктурою на основі платформи Azure Cloud та інструменту Terraform. Сучасні дослідження, підкреслюють важливість інтеграції автоматизованих інструментів управління для підвищення ефективності використання хмарних ресурсів, що включає аналіз поточних викликів у масштабованих хмарних ресурсах, таких як балансування навантаження, забезпечення неперервної доступності сервісів та оптимізація використання ресурсів. Огляд методів оптимізації демонструє стратегії зниження витрат та покращення продуктивності в хмарних середовищах. Результати дослідження призначені для широкого кола фахівців у галузі.

Ключові слова: хмарні обчислення, масштабовані хмарні ресурси, методи моделювання, terraform, azurecloud, оптимізація інфраструктури, ефективне використання ресурсів, аналіз та оцінка, продуктивність, стабільність, ефективність витрат, інтеграція ресурсів, моделювання інфраструктури.

Вступ

У сучасному світі хмарні технології стають все більш важливими для управління та розгортання інфраструктури в інформаційних системах [1]. Хмарні обчислення відкривають нові можливості для динамічного аналізу та обробки великих обсягів даних та послуг в Інтернеті. Ці системи забезпечують доступ до розподілених ресурсів, таких як хости, пристрої зберігання даних, процесорні модулі та інші, за запитом користувачів.

Згідно з дослідженням [2], використання хмарних обчислень приносить передові технології в обробку та аналіз інформації, що охоплює широкий спектр від робочого столу користувача до великих систем обробки даних.

Зазвичай доступ до такої інформації реалізується через веб-додатки на основі хмарної моделі надання послуг.

Хмарні системи представляють собою глобальні мережі датацентрів, що об'єднуються в єдину інфраструктуру високошвидкісними комп'ютерними мережами. Вони можуть ефективно забезпечувати збір, обробку та зберігання інформації.

В [3] наголошують, що послуги хмарних систем розподіляються між користувачами з використанням віддалених ресурсів у конкретний момент часу через Інтернет. Для ефективного використання ресурсів хмарних систем необхідно вирішити проблему розподілу ресурсів, що полягає у динамічному перерозподілі програмних завдань та потоків даних в реальному часі.

Розподіл ресурсів у хмарних системах [4] - це процес розподілу віддалених ресурсів через комп'ютерну мережу для виконання відповідних хмарних програм. На високому рівні абстракції такий механізм розподілу отримав назву

Інфраструктура як послуга (IaaS), яка надає комп'ютерні ресурси за запитами користувачів завдяки попередньо визначеному або динамічному механізму розподілу ресурсів. **Аналіз публікацій та**

постановка проблеми. Останніми роками хмарні технології зазнали значних трансформацій, що спонукало наукову спільноту до інтенсивних досліджень у сфері ефективності використання хмарних ресурсів.

Широкий спектр публікацій в області інформаційних технологій [5] охоплює різні аспекти хмарних обчислень, включаючи архітектуру хмарних систем, управління даними, безпеку та приватність, а також оптимізацію ресурсів. Важливим напрямком досліджень є розробка алгоритмів і методів для забезпечення адаптивного масштабування хмарних ресурсів, що реагують на змінні потреби використання без зайвих затрат.

За даними [6], алгоритми масштабування мають критичне значення для підтримки високої доступності та ефективності хмарних сервісів.

Паралельно, робота [7] акцентує увагу на використанні машинного навчання для автоматизації рішень щодо розподілу ресурсів, вказуючи на можливість значного підвищення продуктивності хмарних систем.

Однак, існуючі дослідження часто обмежуються теоретичними моделями або специфічними випадками використання, не враховуючи складності реальних хмарних середовищ [8]. Проблема оптимізації розподілу ресурсів у хмарних системах залишається актуальною, особливо у контексті забезпечення високої продуктивності та економічної ефективності. Ключовим викликом є знайти баланс між потребами додатків у ресурсах та доступними обчислювальними потужностями [9], щоб мінімізувати витрати без шкоди для якості обслуговування.

В контексті цього дослідження, особлива увага приділяється інтеграції та використанню таких інструментів, як Azure Cloud та Terraform, які надають можливості для гнучкого управління хмарними ресурсами. Azure Cloud з його обширним набором сервісів та Terraform як потужний інструмент для "інфраструктури як коду" відкривають нові перспективи для оптимізації розподілу ресурсів.

Методи оптимізації розподілу ресурсів у масштабованих хмарних середовищах на основі Azure Cloud та Terraform

У даній статті осліджуємо методи моделювання та оптимізації розподілу ресурсів у масштабованих хмарних середовищах, використовуючи платформу Azure Cloud та інструмент Terraform. Основною метою є забезпечення ефективного використання ресурсів та підвищення продуктивності системи.

Першим кроком у нашому дослідженні є аналіз існуючих методів розподілу ресурсів у хмарних середовищах. Розглядаємо різноманітні підходи, включаючи алгоритми планування та оптимізації, що застосовуються у відкритій літературі.

Далі пропонуємо гібридний підхід, який поєднує можливості Azure Cloud та Terraform. На першому етапі використовується Terraform для моделювання та автоматизації інфраструктури хмарного середовища. Детально розглядаємо процес створення та конфігурування інфраструктури за допомогою Terraform, зосереджуючись на можливостях автоматизації та масштабованості.

На другому етапі розглядаємо методи оптимізації розподілу ресурсів у масштабованих хмарних середовищах. Вивчаємо різні підходи до розподілу віртуальних машин, контейнерів та мікросервісів, а також вплив різних критеріїв, таких як продуктивність, ефективність використання ресурсів та вартість.

У наступних розділах представляємо практичні експерименти, в яких застосовується розроблений підхід до оптимізації розподілу ресурсів у хмарних середовищах на базі Azure Cloud та Terraform. Оцінюємо ефективність розроблених методів за різними критеріями та проводимо порівняння з існуючими рішеннями.

Загалом, робота спрямована на розробку та впровадження ефективних методів розподілу ресурсів у масштабованих хмарних середовищах з використанням передових технологій, що дозволяють підвищити продуктивність та знизити витрати.

Подальше дослідження у цій області може включати розширення підходу на інші хмарні платформи та вдосконалення методів оптимізації з метою підвищення надійності та ефективності хмарних середовищ.

Оптимізація розподілу ресурсів у хмарних середовищах вимагає комплексного підходу [10], що поєднує в собі передові технології та інноваційні методики. Azure Cloud і Terraform стали ключовими інструментами у цьому процесі, дозволяючи ефективно управляти ресурсами та автоматизувати розгортання інфраструктури.

Використання Azure Cloud для оптимізації. Azure Cloud пропонує ряд послуг та інструментів [11], що допомагають оптимізувати розподіл ресурсів. Одним із таких інструментів є Azure Autoscale, який дозволяє автоматично масштабувати кількість віртуальних машин відповідно до поточного навантаження, забезпечуючи таким чином ефективне ви-

користання ресурсів та оптимальну продуктивність за мінімальних витрат.

Крім того, Azure Cost Management [12] допомагає відстежувати та аналізувати витрати на хмарні ресурси, дозволяючи ідентифікувати можливості для зниження вартості.

Azure Autoscale – це потужний інструмент для автоматичного масштабування кількості віртуальних машин або інших ресурсів на основі поточного навантаження або запланованих параметрів. Використовуючи Azure Autoscale, можна:

1. Адаптувати ресурси до поточного навантаження. Гарантувати, що додатки завжди мають достатньо ресурсів для оптимальної роботи, без надмірного резервування ресурсів.

2. Мінімізувати витрати. Автоматичне зменшення кількості ресурсів у періоди низького навантаження допомагає знизити витрати на хмарні сервіси.

3. Покращити доступність та продуктивність. Забезпечується швидке масштабування ресурсів у відповідь на сплески навантаження, підтримуючи високий рівень продуктивності.

Terraform як інструмент для управління інфраструктурою

Terraform використовується для опису інфраструктури за допомогою коду, що дозволяє автоматизувати розгортання, зміни та версіонування хмарних ресурсів. Це забезпечує високий рівень гнучкості та контролю над хмарною інфраструктурою. За допомогою Terraform можна легко інтегрувати різні хмарні платформи, включаючи Azure, та ефективно управляти розподілом ресурсів, знижуючи витрати та покращуючи доступність сервісів.

Terraform дозволяє описати інфраструктуру у вигляді коду (Infrastructure as Code, IaC), що значно спрощує процеси розгортання, управління та масштабування хмарних ресурсів. Використовуючи Terraform для оптимізації розподілу ресурсів, можна:

1. Декларативно моделювати інфраструктуру. Описуючи потрібні ресурси та їх конфігурації у вигляді коду, можна легко реплікувати та масштабувати інфраструктуру відповідно до потреб.

2. Використовувати модулі Terraform. Модулі дозволяють групувати ресурси та перевикористовувати конфігурації, забезпечуючи краще управління інфраструктурою та зниження можливості помилок.

3. Автоматизувати зміни інфраструктури. Terraform підтримує планування змін та безпечно застосування цих змін до хмарної інфраструктури, дозволяючи динамічно адаптувати ресурси під змінні потреби.

Для представлення процесу оптимізації розподілу ресурсів у хмарних середовищах, можна використати графік (рис. 1), який ілюструє принципи роботи Azure Autoscale та ефективність використання Terraform для управління інфраструктурою.

Для представлення впливу оптимізаційних стратегій на економічну ефективність та продуктивність хмарних ресурсів, ми зробили порівняльний аналіз.

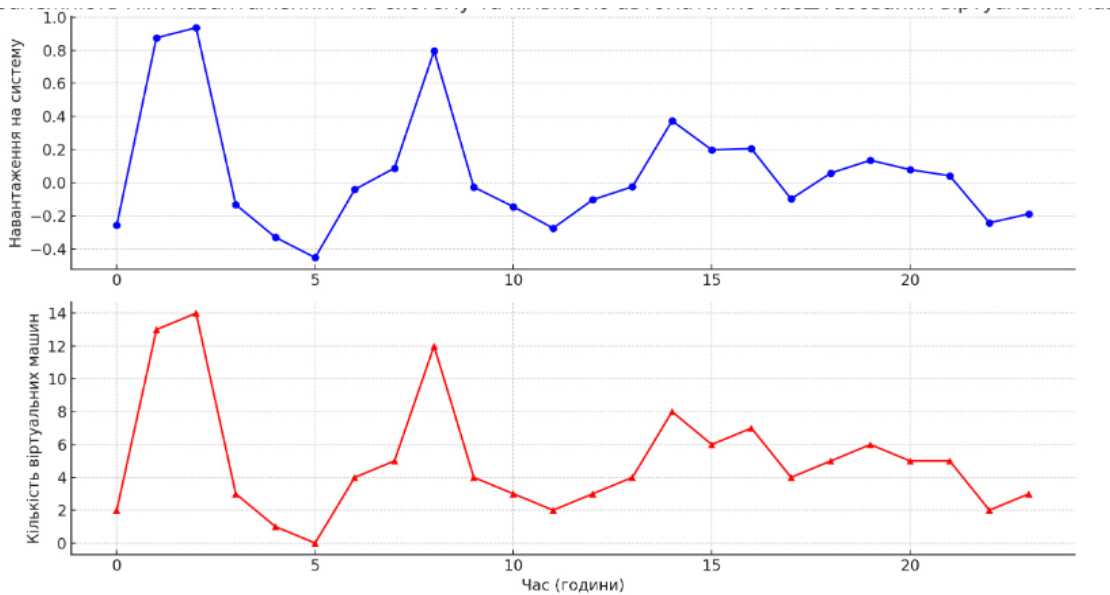


Рис. 1. Залежність між навантаженням на систему та кількістю автоматично масштабованих віртуальних машин в Azure Cloud

Цей аналіз візуалізується за допомогою графіку (рис. 2), який відображає зміни в загальних витратах на хмарні ресурси та продуктивності хмарних сервісів до та після впровадження інструментів автоматизації та оптимізації, таких як Terraform і

Azure Autoscale. Графік демонструє не лише фінансові вигоди, але й поліпшення у якості обслуговування, що є критично важливим для забезпечення високого рівня задоволеності користувачів та ефективності бізнес-операцій.

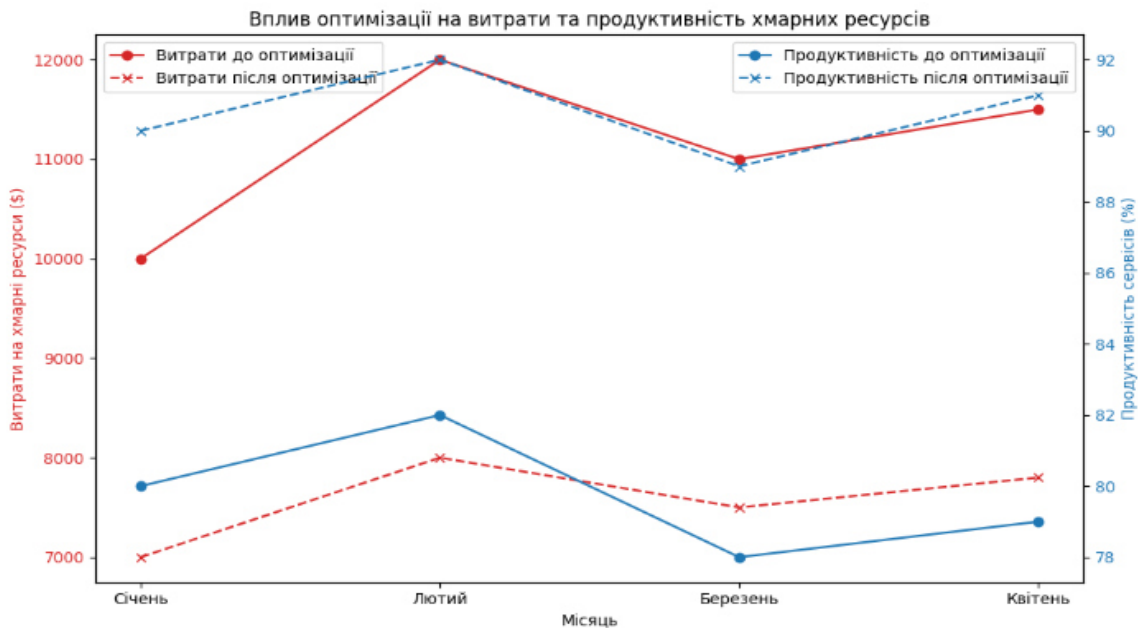


Рис. 2. Вплив оптимізації хмарних ресурсів на витрати та продуктивність сервісів

Цей графік ілюструє, як застосування вдосконалених методів управління інфраструктурою може призвести до суттєвого зниження операційних витрат на хмарні ресурси та одночасного підвищення продуктивності сервісів, гарантуючи таким чином більш ефективне та економічно вигідне використання хмарних технологій.

Висновки

Дослідження методів оптимізації розподілу ресурсів у масштабованих хмарних середовищах, з осо-

бливим акцентом на використанні Azure Cloud та Terraform, відкрило нові перспективи для покращення ефективності, надійності та економічності хмарних інфраструктур. Розроблені та проаналізовані методики демонструють важливість автоматизації, декларативного опису інфраструктури та впровадження інтелектуального масштабування для задоволення змінних потреб сучасних додатків та сервісів.

Основні висновки роботи:

1. Автоматизація інфраструктури за допомогою Terraform значно спрощує управління хмарними

ресурсами, забезпечує високу швидкість внесення змін та підтримку великої кількості середовищ і конфігурацій, водночас знижуючи ризик людських помилок.

2. Динамічне масштабування в Azure Cloud, зокрема використання Azure Autoscale, дозволяє ефективно реагувати на зміни навантаження, оптимізуючи використання ресурсів і витрати на хмарні послуги, при цьому підтримуючи необхідний рівень продуктивності та доступності сервісів.

3. Інтеграція інструментів і платформ, таких як Terraform і Azure, відкриває додаткові можливості для більш гнучкого управління хмарною інфраструктурою, дозволяючи компаніям швидко адаптуватися до змін на ринку та технологічних інновацій.

Вплив оптимізації на економічну ефективність та продуктивність. Аналіз витрат і продуктивності до та після оптимізації яскраво демонструє, що ефективне управління хмарними ресурсами призводить

до значного зниження витрат – приблизно на 30-35% від початкових значень, а також до підвищення продуктивності сервісів на 10-15%. Це підтверджує важливість впровадження автоматизованих інструментів управління та оптимізації, таких як Terraform і Azure Autoscale. Використання цих інструментів для досягнення більш ефективного розподілу ресурсів не тільки забезпечує суттєве зниження витрат, але й сприяє збільшенню рентабельності інвестицій у хмарну інфраструктуру, забезпечуючи водночас вищу якість наданих хмарних сервісів.

Загалом, застосування сучасних підходів до управління хмарними ресурсами, таких як Terraform для декларативного опису інфраструктури та інструментів автомасштабування в Azure, є ключем до побудови ефективних, гнучких та економічно вигідних хмарних рішень. Отже, компанії, які активно впроваджують та оптимізують такі рішення, отримують зниження витрат.

СПИСОК ЛІТЕРАТУРИ

1. Wang, D., Zhong, D. and Li, L. "A comprehensive study of the role of cloud computing on the information technology infrastructure library (ITIL) processes", *Library Hi Tech*, (2022), Vol. 40 No. 6, pp. 1954-1975. DOI: <https://doi.org/10.1108/LHT-01-2021-0031>
2. Omar Alzakholi, Lailan M. Haij, Hanan M. Shukur, Rizgar R. Zebari, Shakir M. Abas, Mohammad A. M. Sadeeq Comparison Among Cloud Technologies and Cloud Performance (2020) Vol. 1 No. 1 DOI: <https://doi.org/10.38094/jastt1219>
3. Рудь Л. І. Войцеховська О. В. Використання хмарних технологій в методології devops та CI/CD процесі. 2021 DOI: <https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/34093/89359.pdf?sequence=2&isAllowed=y>
4. Коломицев М. В. Підхід до побудови системи безпеки хмарних баз даних Johnson та Smith 2020 No УДК 004.7. 2-3 URL: <https://ela.kpi.ua/server/api/core/bitstreams/d740a6c4-936d-4d95-a2e5-b41c5686a75e/content>
5. Sururah A. Bello, Lukumon O. Oyedele, Olugbenga O. Akinade, Muhammad Bilal, Juan Manuel Davila Delgado, Lukman A. Akanbi, Anuoluwapo O. Ajayi, Hakeem A. Owolabi Review Cloud computing in construction industry: Use cases, benefits and challenges 2020 Vol. 6 No. 3, pp. 54-60 DOI: <https://doi.org/10.1016/j.autcon.2020.103441>
6. Mohammad S., Sukhpal S., Adel N. Performance evaluation metrics for cloud, fog and edge computing: A review, taxonomy, benchmarks and standards for future research, 2020 Vol. 40 No. 6, pp. 54-75. DOI: <https://doi.org/10.1016/j.iot.2020.100273>
7. Кулик В.В. Дослідження методів оптимізації обчислень у хмарних технологіях 2020 12 – 15 DOI: <https://openarchive.nure.ua/server/api/core/bitstreams/c0243f1e-34a0-42e0-b127-d87f104e6c43/content>
8. Kuchuk, H. and Malokhvii, E. (2024), "Integration of IOT with Cloud, Fog, and Edge Computing: A Review", *Advanced Information Systems*, vol. 8(2), pp. 65–78, doi: <https://doi.org/10.20998/2522-9052.2024.2.08>
9. Petrovska, I., Kuchuk, H., Kuchuk, N., Mozhaiev, O., Pochebut, M., Onishchenko, Yu. (2023), "Sequential Series-Based Prediction Model in Adaptive Cloud Resource Allocation for Data Processing and Security", *2023 13th International Conference on Dependable Systems, Services and Technologies, DESSERT 2023*, 13–15 October, Athens, Greece, code 197136, doi: <https://doi.org/10.1109/DESSERT61349.2023.10416496>
10. Гуржій В. В. ТЕОРЕТИЧНІ АСПЕКТИ ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ В УПРАВЛІННІ ПРОЕКТАМИ 2023 No УДК 005:004.8 4 - 5 DOI: <https://doi.org/10.32702/2307-2105.2023.12.73>
11. Огляд служб оптимізації виконання й повернення замовлень. 2024. URL: <https://learn.microsoft.com/uk-ua>
12. Microsoft Cost Management 2024. URL: <https://azure.microsoft.com/en-us/products/cost-management>

Received (Надійшла) 12.03.2024

Accepted for publication (Прийнята до друку) 29.05.2024

Methods of modelling scaled cloud resources

Maksym Volk, Maksym Popovkin

Abstract. In the era of rapid development of cloud technologies, the main task is not only the research of existing opportunities, but also the practical implementation of solutions aimed at optimizing the use of cloud resources. The purpose of the article is to identify the most effective approaches to modeling cloud resources, which will allow organizations to achieve a significant reduction in the costs of using cloud resources while simultaneously ensuring a high level of performance and reliability of cloud services. The work offers effective strategies for automating the deployment and management of cloud infrastructure based on the Azure Cloud platform and the Terraform tool. Modern research emphasizes the importance of integrating automated management tools to increase the efficiency of cloud resource use, which includes the analysis of current challenges in scalable cloud resources, such as load balancing, ensuring continuous availability of services and optimization of resource use. An overview of optimization techniques demonstrates strategies for reducing costs and improving performance in cloud environments. The results of the study are intended for a wide range of specialists in the field.

Keywords: cloud computing, scalable cloud resources, modeling methods, terraform, azurecloud, infrastructure optimization, efficient use of resources, analysis and evaluation, performance, stability, cost effectiveness, resource integration, infrastructure modeling.

О. О. Галицька, Н. М. Бологова, Д. О. Кібірев, О. В. Скиба

Харківський національний університет радіоелектроніки, Харків, Україна

ОГЛЯД ПІДХОДІВ ДО ЗАХИСТУ ТРИВИМІРНИХ МОДЕЛЕЙ ВІД НЕСАНКЦІОНОВАНОГО РОЗПОВСЮДЖЕННЯ

Анотація. Стаття присвячена огляду сучасних підходів до захисту тривимірних (3D) моделей від несанкціонованого розповсюдження. У зв'язку з розвитком технологій тривимірного моделювання та широким використанням 3D моделей у різних галузях, питання захисту інтелектуальної власності набуває особливої актуальності. Розглянуто основні методи захисту, включаючи криптографічні техніки, цифрові водяні знаки, стеганографію та методи машинного навчання. Проведено аналіз переваг та недоліків кожного підходу, а також їх ефективності в різних контекстах застосування. Особлива увага приділена новітнім розробкам у сфері захисту 3D моделей та перспективам їх подальшого розвитку. На основі проведеного аналізу запропоновано рекомендації щодо вибору найбільш оптимальних методів захисту в залежності від специфіки використання тривимірних моделей.

Ключові слова: тривимірні моделі, захист інтелектуальної власності, криптографія, цифрові водяні знаки, стеганографія, машинне навчання, несанкціоноване розповсюдження.

Вступ

Тривимірні графічні моделі знаходять широке застосування в кіно, архітектурі, іграх, віртуальній реальності, автоматизованому проєктуванні (САПР), симуляції органів у медицині, військовій справі та біоінформатиці - і це лише деякі з безлічі сфер застосування. З появою 3D-телевізорів та доступних високопродуктивних 3D-відеокарт для настільних комп'ютерів, 3D-моделі стають ще більш розповсюдженими. А з перспективою використання 3D-принтерів у побуті ринок створення цифрового 3D-контенту перетворився на багатомільярдну індустрію. Існують веб-сайти [1, 2], які дозволяють художникам замовляти свої роботи і продавати ці 3D-моделі. Існують пошукові системи [3] для пошуку 3D-моделей, а також низка інструментів для створення таких 3D-моделей, зокрема [4–6]. Однак проєктування та створення високоякісних 3D-графічних моделей вимагає значних навичок та використання спеціалізованого програмного забезпечення та/або обладнання, наприклад, лазерних сканерів. Зважаючи на високий попит і популярність 3D-моделей, а також враховуючи вартість, час і зусилля, необхідні для створення таких моделей, виникає загроза широкого розповсюдження незаконного копіювання 3D-моделей. Водяні знаки - метод, який запобігає нелегальному копіюванню шляхом вставки прихованого повідомлення в 3D-модель.

Методи управління цифровими правами (DRM), засновані на технологіях шифрування, в минулому використовувалися для запобігання копіюванню цифрових мультимедійних матеріалів, які потрібно було розшифрувати і розблокувати протягом декількох місяців, якщо не тижнів або днів. Алгоритми шифрування не залежать від формату мультимедіа, чи то аудіо, відео, пісні з альбомів, фільмів, електронних книг або 3D-моделей. Численні копії 3D-моделей Шрека (з фільму "Шрек"), 3D-модель Голлума з фільму "Володар перснів" можна знайти в Інтернеті. Голлівуд щорічно втрачає мільйони доларів через піратські фільми, які іноді розповсюджуються ще до того, як вони вийшли в кінотеатрах. Запобігти копіюванню мультимедійних матеріалів не лише важко, але й важко відстежити

піратів або походження порушення в ланцюжку дистрибуції. Незважаючи на суворе законодавство навіть в США, яке було прийнято Законом про захист авторських прав у цифрову епоху (DMCA), та судові позови Американської асоціації звукозаписної індустрії (RIAA), незаконне копіювання мультимедійних матеріалів продовжується й досі. Системи DRM намагаються забезпечити антипіратські рамки, які обмежують використання контенту його законним користувачем. Провал DRM, що базується на шифруванні, можна підкреслити тим, що Apple відмовилася від свого FairPlay DRM в iTunes у січні 2009 року [7].

Водяні знаки - це технологія, яка вставляє повідомлення або код у цифровий контент. Ініціатива із захисту цифрової музики (Secure Digital Music Initiative, SDMI), консорціум компаній музичної індустрії, у 2000 році провела змагання, щоб перевірити надійність їхніх технологій нанесення водяних знаків. Едвард Фелтон (Edward Felton) та його команда з Принстонського та Райського університетів [8] перемогли всі чотири алгоритми водяних знаків, тим самим довівши, що самі по собі водяні знаки є неефективними. Однак, з 2000 року було проведено значну кількість важливих досліджень. У 2008 році Fox Studios почала використовувати систему водяних знаків на вимогу для автоматичного і безперешкодного вбудовування непомітної інформації в кожен кадр відеоконтенту з метою захисту від піратства. Хоча методи шифрування і водяних знаків, коли вони використовуються окремо, виявилися неефективними, коли вони використовуються разом в рамках DRM, вони можуть стати потужним підходом для виявлення незаконного копіювання. Наприклад, фільм "Люди Ікс: Росомаха" у квітні 2009 року незаконно розповсюджувався з видимими водяними знаками "Rising Sun Pictures" у пірингових мережах (P2P), і врешті-решт було заарештовано особу, яка завантажила цей фільм на сайт megaupload.com [9]. Хоча наявність водяного знаку, очевидно, не сприяла арешту, тим не менш, він дав змогу локалізувати джерело витоку інформації. Для кіностудій стало звичайною практикою додавати водяні знаки до своїх попередніх версій фільмів, щоб відстежити первісного завантажувача.

Стан досліджень у сфері нанесення водяних знаків на 3D-моделі (3D стеганографія) все ще перебуває на початковому етапі порівняно з опублікованими роботами у сфері нанесення водяних знаків на зображення та відео. Однак досвід, отриманий в кіноіндустрії, свідчить про те, що водяні знаки є життєздатною технологією, яка залишиться в тренді і може бути поширена на 3D-моделі. Сьогодні для художників-аматорів не є звичайною практикою вставляти водяні знаки під час продажу свого оригінального 3D-контенту продавцю, і вони довіряють продавцю, що той не порушить права власності художника. Хоча продавці 3D-моделей виплачують художникам роялті за право власності на 3D-моделі, не існує бізнес-моделей для підтримки перерозподілу 3D-моделей, які б захищали інтелектуальні права художників. Небагато роботи було зроблено щодо нанесення декількох водяних знаків для підтримки такої бізнес-моделі розповсюдження через торгових посередників. Таким чином, існує бізнес-потреба в надійних алгоритмах нанесення 3D водяних знаків.

Загальний принцип нанесення цифрових водяних знаків на 3D-моделі

Полігональна сітка (англ. Polygon mesh) — це набір вершин, ребер, та граней, що описують форму багатогранного об'єкта в тривимірній графіці та твердотілому моделюванні. Грані зазвичай складаються з трикутників (сітка з трикутників), чотирикутників, чи інших опуклих багатокутників, що спрощує їх рендеринг, хоча можуть використовуватись і загальніші, неопуклі багатокутники, чи багатокутники з дірками. Список вершин містить координати у тривимірному просторі кожної вершини моделі, а список граней описує, як вершини з'єднані одна з одною. Список ребер може бути отриманий шляхом обходу списку граней і списку вершин. Рис. 1 є прикладом каркасної сітки.

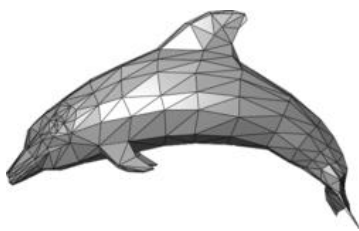


Рис. 1. Приклад трикутної сітки, яка використовується для зображення дельфіна

Технології водяних знаків вбудовують непомітні дані в мультимедійний контент. Приховані дані називаються цифровими водяними знаками і можуть складатися з унікального ідентифікатора користувача, криптографічних ключів, повідомлень про авторські права, умов доступу до контенту, логотипів, зображень, біометричних даних або інформації, що базується на контенті. Процес вбудовування та пошуку цифрових водяних знаків відбувається за допомогою секретного ключа, в якому міститься інформація про те, де і в якій мірі оригінальний контент був змінений для розміщення водяного знаку. Непомітність є важливою вимогою до кожної схеми нанесення цифрових водяних знаків, оскільки водяний знак не по-

винен спотворювати оригінальний зміст або заважати його використанню за призначенням чи виконанню функцій. Надійність необхідна для того, щоб гарантувати, що звичайна обробка сигналів, геометричні операції та зловмисні модифікації не вплинуть на виявлення або відновлення водяного знаку. Мета полягає в тому, щоб полегшити власникам контенту довести своє право власності шляхом вилучення водяного знаку з піратського носія, а потім подати позов проти порушника. На рис. 2 та 3 показано два компоненти системи захисту водяними знаками: вбудований пристрій та детектор.

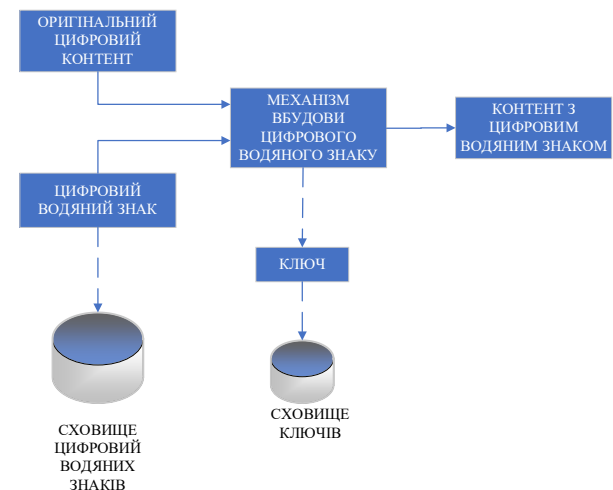


Рис. 2. Технологія нанесення цифрового водяного знаку



Рис. 3. Технологія виявлення цифрового водяного знаку

Процес виявлення водяних знаків може бути несліпим (процес пошуку вимагає доступу до повного оригінального вмісту), напівсліпим (детектору потрібен доступ до деякої побічної інформації та/або цифрового водяного знаку, але не до оригінального вмісту) або сліпим (виявлення виконується без доступу до оригінального вмісту). Несліпі методи виявлення є більш надійними, але непрактичними для використання в системах DRM. Оскільки несліпі методи вимагають, щоб оригінальний контент був доступний детектору, це призводить до необхідності доступу до оригінального контенту з боку програмного забезпечення системи DRM, що створює діру в безпеці системи. Напівсліпі методи є найбільш придатними для

використання в цьому контексті, оскільки сліпі методи не відповідають вимогам надійності.

Алгоритми водяних знаків також можна умовно поділити на перше, друге та третє покоління водяних знаків. Існуючі алгоритми водяного маркування в цій статті були класифіковані на два покоління алгоритмів водяного маркування. Покоління класифіковано на основі можливостей алгоритмів. З кожним наступним поколінням алгоритми водяних знаків мають більші можливості при вбудові, не спричиняючи при цьому помітних спотворень і залишаючись стійкими. На рис. 4 показано класифікацію алгоритмів за поколіннями на основі домену вставки.

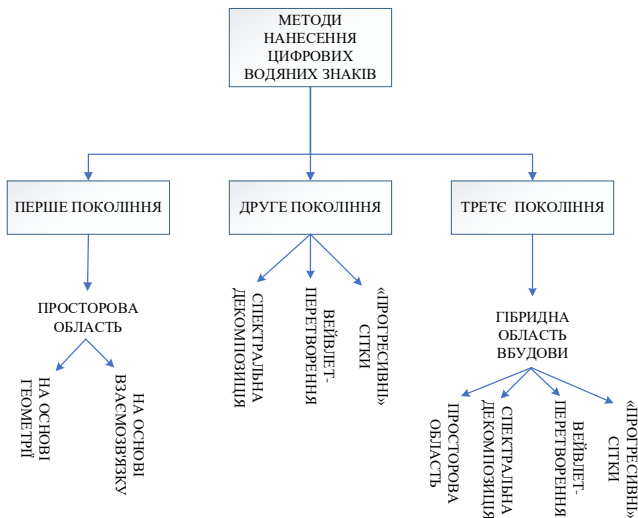


Рис.4. Класифікація алгоритмів за поколіннями на основі домену вставки

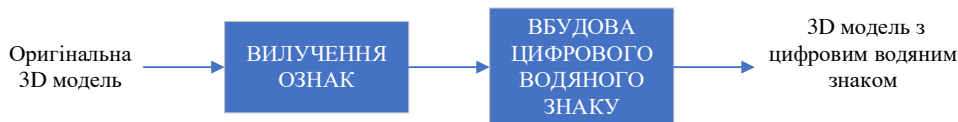


Рис. 5. Загальна схема алгоритмів нанесення водяних знаків першого покоління

Для кожного трикутника, що задовольняє функції допустимості, в локальні інваріанти вносяться невеликі модифікації шляхом зміни положення сусідніх точок. Як наслідок, вони є чутливими до модифікацій з додаванням шуму. Серед цього класу схем водяних знаків було запропонувати чотири різні алгоритми водяних знаків у першій опублікованій роботі, присвяченій 3D-водяним знакам. Ці схеми відповідно називаються Triangle Similarity Quadruple (TSQ), Tetrahedral Volume Ratio (TVR), Triangle Strip Peeling Sequence (TSPS) та Macro Density Pattern (MDP) [11, 12]. В статті [13] автори вставляють випадковий водяний знак на основі коефіцієнта маскувального у вершинах. Коефіцієнт маскувального базується на оцінці середньої різниці між позицією та з'єднаними вершинами. Розрядність вставленого водяного знаку становить 100 біт, алгоритм було протестовано на двох моделях з кількістю граней від 3 500 до 5 000 граней. Алгоритм несліпого нанесення водяних знаків виявився стійким до адитивного шуму, стиснення MPEG4 та атак спрощення сітки. Однак стійкість алгоритму була зумовлена

Методи третього покоління ґрунтуються на існуючих алгоритмах першого і другого поколінь, а також включають гібридні домени, що дозволяють об'єднувати інформацію з різних доменів. Запропоновані алгоритми третього покоління досліджують використання методів обчислювального інтелекту для вставки водяних знаків високої ємності як у просторову область, так і в область перетворення.

Більш детальний опис алгоритмів нанесення цифрових водяних знаків був висвітлений в огляді методів нанесення 3D водяних знаків [10], тому тут буде представлено лише короткий огляд.

Перше покоління методів нанесення цифрових водяних знаків

Алгоритми першого покоління вставляють водяний знак у просторову область, змінюючи положення вершин або змінюючи зв'язність вершин. На рис. 5 показано схему алгоритмів нанесення водяних знаків першого покоління. Ознаки витягуються з просторової області. Схеми 3D-водного маркування, які вбудовують дані в просторову область, можна розділити на дві основні категорії: схеми водяних знаків, що базуються на зв'язках, і схеми водяних знаків, що базуються на геометрії. Схеми просторових водяних знаків зазвичай менш стійкі до таких атак, як стиснення і додавання шуму. Однак вони витримують атаки обрізання і є менш складними. Алгоритми накладання водяних знаків на основі зв'язності - це алгоритми, які явно використовують зв'язність сітки (деякі автори також називають їх топологічними особливостями). Ці схеми, як правило, базуються на обході всіх трикутників сітки.

багаторазовим вставлянням водяного знаку та використанням кодів з надлишковою стійкістю до помилок, а отже, низькою здатністю до вбудовування.

Алгоритми накладання водяних знаків на основі зв'язності - це алгоритми, які явно використовують зв'язність сітки (деякі автори також називають їх топологічними особливостями). Ці схеми, як правило, базуються на обході всіх трикутників сітки. Для кожного трикутника, що задовольняє функції допустимості, в локальні інваріанти вносяться невеликі модифікації шляхом зміни положення сусідніх точок. Як наслідок, вони є чутливими до модифікацій з додаванням шуму. Серед цього класу схем водяних знаків було запропонувати чотири різні алгоритми водяних знаків у першій опублікованій роботі, присвяченій 3D-водяним знакам. Ці схеми відповідно називаються Triangle Similarity Quadruple (TSQ), Tetrahedral Volume Ratio (TVR), Triangle Strip Peeling Sequence (TSPS) та Macro Density Pattern (MDP) [11, 12]. В статті [13] автори вставляють випадковий водяний знак на основі коефіцієнта маскувального у вершинах. Коефіцієнт маскувального базується на оцінці середньої різниці між позицією та

з'єднаними вершинами. Розрядність вставленого водяного знаку становить 100 біт, алгоритм було протестовано на двох моделях з кількістю граней від 3 500 до 5 000 граней. Алгоритм несліпого нанесення водяних знаків виявився стійким до адитивного шуму, стиснення MPEG4 та атак спрощення сітки. Однак стійкість алгоритму була зумовлена багаторазовим вставлянням водяного знаку та використанням кодів з надлишковою стійкістю до помилок, а отже, низькою здатністю до вбудовування.

Друге покоління методів нанесення цифрових водяних знаків

Порівняно з першим поколінням було зроблено кілька удосконалень, які покращили характеристики з точки зору ємності, невидимості та стійкості водяного знаку. На рис. 6 показано блок-схему алгоритмів

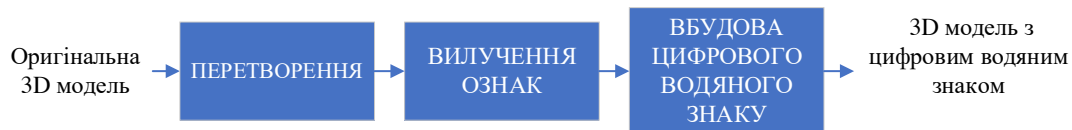


Рис. 6. Загальна схема алгоритмів нанесення водяних знаків другого покоління

Таким чином, водяний знак існує навіть при більш низькій роздільній здатності 3D-моделі. Це дає дві основні переваги над традиційними підходами: По-перше, це допомагає захиститися від атаки методом зниженої вибірки. При атаці з використанням низхідної дискретизації зломисник знижує роздільну здатність 3D-моделі і, таким чином, зменшує її роздільну здатність, щоб видалити водяний знак. Але оскільки водяний знак вставляється навіть у нижчу роздільну здатність 3D-моделі, він не знищується. Таким чином, підхід з використанням декількох роздільних здатностей робить водяний знак стійким до атак. По-друге, оскільки водяний знак додається навіть при низькій роздільній здатності, кількість доданого водяного знаку є більшою, ніж водяний знак, доданий без аналізу з декількома роздільними здатностями. Це збільшує ємність водяного знаку, що

підвищує стійкість до таких атак, як згладжування, обрізання та додавання шуму.

Алгоритми другого покоління використовують різні перетворення для вставки водяного знаку в коефіцієнти області перетворення для підвищення стійкості. У другому поколінні для декомпозиції 3D-моделі до більш низької роздільної здатності використовуються спектральна декомпозиція та методи з декількома роздільними здатностями, такі як вейвлет-перетворення та прогресивні сітки, а водяний знак вставляється в бітовий потік. Таким чином, алгоритми другого покоління дозволяють застосовувати підхід до нанесення водяних знаків до потокових сіток і підвищують надійність алгоритму, вставляючи водяний знак з різною роздільною здатністю. Вейвлет-перетворення дає багаторівневе представлення 3D-моделі. На кожному рівні вейвлет-перетворення в 3D-моделі вставляється водяний знак.

підвищує стійкість до таких атак, як згладжування, обрізання та додавання шуму. Обмеженням вейвлет-методів є те, що сітка повинна мати зв'язність від 1 до 4 підрозділів. Вейвлет-перетворення можна застосовувати лише до сіток з напіврегулярною зв'язністю через процес четвертинного поділу/спрощення.

Третє покоління методів нанесення цифрових водяних знаків

Методи нанесення водяних знаків третього покоління базуються на методах першого і другого поколінь, додаючи інтелектуальний рівень оптимізації для вставки водяних знаків високої щільності. Таким чином, ці алгоритми можуть бути розширені для використання на потокових сітках. На рис. 7 показано блок-схему алгоритмів накладання водяних знаків третього покоління.



Рис. 7. Загальна схема алгоритмів нанесення водяних знаків другого покоління

Нанесення водяних знаків можна розглядати як оптимізаційну задачу, де метою є максимізація кількості вершин для нанесення водяних знаків, а також максимізація кількості водяних знаків, які можна вставити, не спричиняючи помітних спотворень. Методи третього покоління також поширюються на алгоритми для крихких водяних знаків. Новизна цієї роботи полягає в оцінці методів обчислювального інтелекту для вирішення проблеми вставки водяних знаків високої щільності як оптимізаційної задачі. Еволюційні методи, такі як генетичні алгоритми, бу-

ли включені в це покоління. Алгоритми на основі нечіткої логіки та нейронних мереж також входять до цієї нової генерації алгоритмів. Як приклад в роботі [14], в якій автори використовували квадратичне програмування (QP) для обмеженої оптимізації 3D-сіток. Було запропоновано метод на основі гістограми для нанесення водяних знаків на 3D полігональній сітці за допомогою квадратичного програмування для мінімізації середньоквадратичної похибки між вихідною сіткою та сіткою з водяними знаками. Однак цей метод має труднощі в роботі з великими сітками через

обмеження складності існуючих QR-розв'язувачів. Не існує жодної опублікованої роботи, яка б досліджувала використання генетичних алгоритмів, нечіткої логіки або штучних нейронних мереж для нанесення 3D водяних знаків. Однак генетичні алгоритми (ГА), нечітка логіка (НЛ) і штучні нейронні мережі (ШНМ) з частковим успіхом використовувалися для нанесення водяних знаків на зображення і відео.

Висновки

3D стеганографія є відносно новою та перспективною галуззю досліджень, що об'єднує аспекти комп'ютерної графіки, криптографії та інформаційної безпеки. У ході огляду було виявлено кілька ключових напрямів та тенденцій розвитку цієї сфери:

– використання 3D моделей для приховування інформації забезпечує вищий рівень захисту порівняно з традиційними 2D методами. Це обумовлено складністю та багатовимірністю 3D об'єктів, що ускладнює виявлення і витяг прихованих даних;

– існує багато методів 3D стеганографії, таких як методи на основі геометричних властивостей, спектрального аналізу, топологічних змін та методи з використанням текстур і матеріалів. Кожен з мето-

дів має свої переваги та недоліки, що робить їх придатними для різних застосувань;

– 3D стеганографія знаходить застосування в багатьох галузях, включаючи захист інтелектуальної власності, безпечний обмін медичними даними, військові технології та розваги. Це свідчить про широкую корисність та потенціал для подальших досліджень;

– попри численні переваги, 3D стеганографія стикається з певними технічними викликами, такими як необхідність високих обчислювальних потужностей, складність у створенні стійких до атак методів, а також питання стандартизації та сумісності;

– подальші дослідження в цій сфері можуть зосередитися на розвитку більш стійких алгоритмів стеганографії, інтеграції з іншими технологіями захисту даних, а також на вирішенні проблем сумісності та стандартизації.

Загалом, 3D стеганографія є перспективною технологією, що має потенціал для значного впливу на галузь інформаційної безпеки. Подальші дослідження та розвиток у цій сфері можуть сприяти створенню більш ефективних та безпечних методів захисту даних, а також розширенню можливостей використання 3D технологій у різних галузях.

СПИСОК ЛІТЕРАТУРИ

- 3D Materials | 3D Textures | Photoreal Textures. URL: <https://www.a23d.co/textures>
- TurboSquid: 3D Models for Professionals [Електронний ресурс]. – Режим доступу: <https://www.turbosquid.com/>
- Thingiverse [Електронний ресурс]. – Режим доступу: <https://www.thingiverse.com/>
- Create massive worlds and high-quality designs. URL: <https://www.autodesk.com/products/3dmax/overview?term=1-YEAR>
- Kuchuk, H., Kovalenko, A., Ibrahim, B.F. and Ruban, I. (2019), "Adaptive compression method for video information", International Journal of Advanced Trends in Computer Science and Engineering, vol. 8(1), pp. 66–69, doi: <http://dx.doi.org/10.30534/ijatcse/2019/1181.22019>
- ZBrush [Електронний ресурс]. – Режим доступу: <https://www.maxon.net/en/zbrush>
- FairPlay DRM – 5 Things to Know About DRM Technology [Електронний ресурс]. – Режим доступу: <https://pallycon.com/blog/5-things-you-need-to-know-about-multi-drm-technology-part-3/>
- Reading Between the Lines: Lessons from the SDMI Challenge. S.A. Craver, M. Wu, B. Liu, A. Stubblefield, and E. W. Felten. Proc. of 10th USENIX Security Symposium, 2001. URL: <https://www.usenix.org/legacy/events/sec01/craver.pdf>
- FBI makes arrest in 'Wolverine' uploading case. URL: <https://www.cnet.com/culture/fbi-makes-arrest-in-wolverine-uploading-case/>
- Beugnon, S., Itier, V., & Puech, W. (2022). 3D Watermarking. Multimedia Security I: Authentication and Data Hiding, 219.
- Semenov, S., Mozhaiev, O., Kuchuk, N., Mozhaiev, M., Tiulieniev, S., Gnusov, Yu., Yevstrat, D., Chyrva, Yu. & Kuchuk, H. (2022), "Devising a procedure for defining the general criteria of abnormal behavior of a computer system based on the improved criterion of uniformity of input data samples", Eastern-European Journal of Enterprise Technologies, 2022, 6 (4(120)), pp. 40–49, doi: <https://doi.org/10.15587/1729-4061.2022.269128>
- A. G. Bors, "Watermarking mesh-based representations of 3-D objects using local moments," in IEEE Transactions on Image Processing, vol. 15, no. 3, pp. 687-701, March 2006, doi: 10.1109/TIP.2005.863116
- Wang, K., Lavoué, G., Denis, F., Baskurt, A. (2007). Three-Dimensional Meshes Watermarking: Review and Attack-Centric Investigation. Lecture Notes in Computer Science, vol 4567. Springer, Berlin, https://doi.org/10.1007/978-3-540-77370-2_4
- Narendra, M., Valarmathi, M.L. & Anbarasi, L.J. Watermarking techniques for three-dimensional (3D) mesh models: a survey. Multimedia Systems 28, 623–641 (2022). <https://doi.org/10.1007/s00530-021-00860-z>

Received (Надійшла) 10.04.2024

Accepted for publication (Прийнята до друку) 19.06.2024

Review of approaches to protecting 3D models from unauthorized distribution

O. Galitska, N. Bolohova, D. Kibirev, O. Skiba

Abstract. The article is devoted to a review of modern approaches to protecting three-dimensional (3D) models from unauthorized distribution. In connection with the development of three-dimensional modeling technologies and the widespread use of 3D models in various industries, the issue of intellectual property protection is of particular relevance. Basic security methods are covered, including cryptographic techniques, digital watermarks, steganography and machine learning methods. The advantages and disadvantages of each approach, as well as their effectiveness in different application contexts, are analyzed. Particular attention is paid to the latest developments in the field of protecting 3D models and prospects for further development. Based on the analysis, recommendations are proposed for choosing the most optimal protection methods depending on the specific use of three-dimensional models.

Keywords: 3D models, intellectual property protection, cryptography, digital watermarking, steganography, machine learning, unauthorized distribution.

В. О. Гороховатський¹, Н. І. Стяглик², Є. В. Мазур¹, А. Д. Вечірська¹

¹ Харківський національний університет радіоелектроніки, Харків, Україна

² ННІ Каразінський банківський інститут Харківського національного університету ім. В.Н. Каразіна, Харків, Україна

Способи метричної грануляції для опису зображень у задачі класифікації

Анотація. Предметом досліджень статті є методи класифікації зображень у системах комп'ютерного зору. Мета – удосконалення структурних методів класифікації задля підвищення їх швидкодії шляхом впровадження редукованої системи класифікаційних ознак на підрунті здійснення грануляції еталонних описів. Застосовані методи: детектори ключових точок *brisk* та *orb*, апарат теорії множин і векторних просторів, метричні моделі кластеризації і грануляції векторів, моделі «мішок слів» і голосування для класифікації, програмне моделювання. Отримані результати: розроблено моделі для трансформації опису зображень у вигляді набору кластерних центрів і побудови редукованого опису з використанням грануляції за подібністю векторів, досліджено швидкісні методи класифікації зображень на основі впровадження трансформованих описів. Результативність розроблених модифікацій класифікатора залежить від способу грануляції даних і від параметрів порогів на еквівалентність векторів, мінімум значення метрики і число голосів дескрипторів для класу – переможця у схемі голосування. Запропоновано побудову ієрархічної системи ознак з кількома рівнями грануляції. На підставі впровадження модифікацій вдалося в сотні разів скоротити обчислювальні витрати при забезпеченні результативності класифікації на навчальній вибірці даних. Практична значущість роботи полягає у побудові програмних моделей для здійснення швидкісної класифікації, підтвердженні працездатності і точності запропонованих модифікацій для прикладної бази зображень, створенні програмного застосунку для впровадження модифікацій класифікаторів у комп'ютерному баченні.

Ключові слова: комп'ютерний зір; структурні методи класифікації; множина дескрипторів; кластеризація, грануляція векторів, точність класифікації.

Вступ. Аналіз літературних джерел

На сьогодні однією із найбільш вагомих проблем у системах комп'ютерного зору є забезпечення необхідних показників їх прикладного функціонування, зокрема, точності та швидкодії оброблення даних [1-4]. Особливо актуальною є вирішення цієї проблеми для задач класифікації візуальних об'єктів, де використовується значна кількість прототипів класів, образи яких подаються у формі множини векторів [2, 5]. Традиційно такі задачі відносяться до сфери *big data* і потребують здійснення об'ємного лінійного пошуку у багатовимірному просторі даних.

Зараз інтенсивно розвиваються структурні методи класифікації, де класифікаційне рішення базується на поданні зображення у формі множини дескрипторів ключових точок [2, 6]. Дескриптори ключових точок (КТ) формуються спеціальними фільтрами – детекторами [7, 8]. У структурних методах процес класифікації здійснюється за моделлю «мішок слів», а клас об'єкту визначається за результатом голосування складових компонентів – дескрипторів розпізнаваного об'єкту [2]. Такі методи мають переваги перед сучасними нейромережами, так як базуються на узгодженні з описом чи параметрами еталонів і не потребують довготривалого навчання. Крім того, при їх застосуванні є можливість оперативної зміни складу бази еталонів і прийняття рішення за окремими деталями візуального об'єкту. Функціонал апарату дескрипторів має необхідну в реальних умовах властивість інваріантності до геометричних перетворень об'єктів у полі зору.

Одним із способів зниження ресурсоемності при впровадженні структурних методів є застосування засобів грануляції даних із формуванням кластерів, хеш-кошиків чи оцінювання центрів [9]. Такі

способи можна вважати метричною грануляцією, так як вони базуються на оцінюванні близькості даних з використанням метрики або параметрів даних [10, 11]. За рахунок впровадження таких підходів суттєво покращується швидкодія оброблення з допустимим зниженням показника точності класифікації. Однак досягнення необхідних значень критерію результативності потребує більш глибокого вивчення усіх можливостей і нюансів прикладного застосування апарату грануляції з необхідністю підтвердження об'єктивності і універсальності цього апарату у сфері інтелектуального аналізу багатовимірних даних. Зокрема, є потреба у впровадженні більш простих у порівнянні з кластеризацією засобів формування результативних чи найбільш інформативних підмножин даних структурного опису [2, 11], а також використання роздільних схем для кластерного подання еталонів у створених модифікаціях швидкісних методів класифікації.

У статтях [3, 5] розглянуто три основні способи отримання редукованої системи ознак на прикладі метричного критерію інформативності для кожного елемента множини дескрипторів опису. Це такі, як:

- 1) відбір фіксованої кількості інформативних ознак;
- 2) визначення числа ознак із інформативністю вище заданого порогу;
- 3) формування інформативної підмножини ознак, що забезпечує необхідну результативність.

Розробляються також більш ресурсозатратні способи, наприклад *Relief* [12], де редукована множина формується випадковою процедурою аналізу на усій наявній множині даних з визначенням оптимальних відстаней до свого і чужого класів. Перспективним вважається також застосування оцінок інформативності на базі центрів даних [5, 13].

Кластерування є одним із найбільш ефективних прикладних підходів, що здійснює апроксимацію даних шляхом побудови деякої системи центрів кластерів. Швидкодія оброблення підвищується пропорційно зменшенню числа центрів. При цьому число центрів є параметром, яким можна керувати задля забезпечення необхідної результативності класифікації [10, 14].

Ефективна в обчислювальному плані ідея щодо побудови інтегральних ознак зображення реалізується пірамідальним поданням [16, 17]. Пірамідальна структура даних дає можливість суттєво скоротити багатомасштабний обсяг зорової інформації шляхом інтегрування яскравості для просторово близьких координат пікселів зображення. Такий підхід гранулювання доцільно поширити і на опис у формі множини дескрипторів. Загалом ідея грануляції даних набула популярності у системах розпізнавання образів [20]. Різні аспекти використання новітніх способів формування ефективного простору ознак у класифікаторах зображень обговорюються у роботах [22-25].

Зосередимося на вивченні і експериментальному впровадженні способів аналізу даних, що базуються на відношенні грануляції на множині дескрипторів. Для конкретизації розглянемо бінарне відношення для пар метрично найближчих елементів опису, в результаті грануляції у редукованому списку залишається тільки один із них.

Задачами дослідження є опрацювання новітніх модифікацій способів грануляції описів для впровадження у класифікаторах зображень, аналіз їх ефективності у порівняльному аспекті з традиційним методом. Увага зосереджена на апроксимації опису за рахунок скорочення кількості його інгредієнтів за принципом подібності, а також на роздільному способі кластерування для описів еталонів. Практичними моментами є вивчення результативності (швидкодії, точності) впровадження пропонованих модифікацій простору даних у методах класифікації шляхом проведення експериментального оцінювання за наслідками програмного моделювання для прикладної бази зображень.

Формалізація структурного методу

Будемо здійснювати класифікацію у рамках фіксованої бази із N еталонів (представники чи прототики класів) у формі деякої множини E описів еталонних зображень: $E = \{E_1, E_2, \dots, E_N\}$. Множина E – це фактично навчальна вибірка, яка одночасно є підґрунтям для побудови класифікатора на базі зіставлення з еталоном [2, 5]. Кожний еталонний опис E_k у формалізмі класифікатора репрезентує окремих клас. Клас k з описом E_k формально визначаємо як нескінченну множину зображень, отриманих із еталону (прототип класу з номером k) шляхом застосування до нього багатопараметричної групи геометричних перетворень, яка найчастіше у прикладних застосуваннях включає зміщення, поворот, масштабування, дія яких не виводять об'єкт інтересу із поля зору [18]. Сучасні детектори ключових точок, наприклад ORB, BRISK, AKAZE забезпечують інваріантність отриманого опису до такої групи перетворень [6–8].

Опис еталону $E_k = \{e_v(k)\}_{v=1}^s$, $E_k \subseteq V^n$, подаємо як скінченну множину потужністю в s дескрипторів КТ у просторі V^n бінарних векторів, $e_v(k) \in V^n$, $s = \text{card}E_k$ – число дескрипторів у множині. Усякий дескриптор $e_v(k)$ у складі бази E характеризується параметром k номеру класу, а загалом число ознак – дескрипторів у базовій множині E складає $\text{card}E = sN$.

Процес R структурної класифікації здійснимо у два етапи як $R = R_2 R_1$, де етап R_1 реалізує локальне рішення стосовно класу окремого дескриптора об'єкту, а етап R_2 на підставі рішень (голосів) компонентів усього складу опису отримує значення параметра класу для аналізованого об'єкту.

Класифікатор R_1 першого етапу реалізує по-елементний аналіз вхідного опису $Z = \{z_v\}_{v=1}^s$ об'єкта і відносить кожний дескриптор $z_v \in Z$ до одного із класів за правилом

$$R_1: z_v \rightarrow \{1, \dots, N\}. \quad (1)$$

Клас k аналізованого дескриптора z_v об'єкту у моделі (1) визначимо як аргумент мінімуму відстані на множині E дескрипторів усіх класів

$$R_1: k = \arg \min_{i=1, \dots, N; d=1, \dots, s} \rho(z_v, e_d(i)). \quad (2)$$

Тут $\rho: V^n \times V^n \rightarrow [0, \infty]$ – відстань у векторному просторі V^n .

У конкурентній моделі (2) доцільно використовувати відстань Хеммінга, що визначає число відмінних бітів у двох бінарних векторах [3, 4]. У виразі (2) мінімізується значення відстані за змінною i (номер класу) на множині дескрипторів бази E . У формулі (2) може бути організовано визначення мінімуму як за повним списком $s \cdot N$ дескрипторів бази, так і роздільно для кожного із N еталонів (з встановленим лімітом на величину мінімуму).

У роздільному способі аналізу для кожного із класів підраховується число дескрипторів об'єкту, що знайшли відповідність до фіксованого еталону. У такій моделі маємо множинне врахування голосів, коли один і той же дескриптор об'єкту може знайти відповідність для кількох еталонів різних класів.

Таким чином, модель (2) втілює багатокласову класифікацію для окремого дескриптора у векторному просторі V^n на підставі його найкращого узгодження з множинами еталонних дескрипторів.

Для побудови R_2 введемо вектор $\{h_i\}_{i=1}^N$ з цілими значеннями для накопиченої кількості голосів класів, отриманих застосуванням моделі (2) до усієї множини дескрипторів об'єкту Z . На підставі впровадження R_1 для кожного $z_v \in Z$ відповідно до (2) визначимо номер класу k , а потім інкрементуємо акумулятор голосів $h_k = h_k + 1$ для відповідного номеру класу. Фактично отриманий вектор $\{h_i\}_{i=1}^N$ є розподілом (гістограмою) числа голосів для множини дескрипторів об'єкту у системі класів.

За результатом оброблення опису Z об'єкту обчислюємо вектор $\{h_i\}_{i=1}^N$. Клас об'єкту визначимо правилом R_2 як аргумент максимуму

$$R_2: Z \rightarrow E_k | (k = \arg \max_{i=1, \dots, N} h_i) \ \& \ (h_k \geq \delta_n), \quad (3)$$

де δ_h – деякий поріг для мінімально допустимого числа голосів. Якщо умова $h_k \geq \delta_h$ не виконується, клас об'єкту не встановлюється (відмова від класифікації, немає підстави віднести об'єкт до жодного із класів).

Значення δ_h визначають експериментально для заданої бази E еталонів. Як правило, його обирають як мінімальне число голосів (з допуском), необхідне для впевненої класифікації тестової вибірки на основі еталонів. Із загальних теоретичних позицій δ_h є компромісом у задачі розрізнення еталонів від решти зображень, що можуть поступити на вхід системи класифікації [5, 9].

Послідовність правил R_1, R_2 реалізує класифікатор на підґрунті рішень набору класифікаторів, отриманих для множини складових компонентів об'єкту. Він впроваджує принцип структурного аналізу і забезпечує стійкість до просторових викривлень окремих компонентів (дескрипторів) із-за можливого впливу завад.

Модифікації простору ознак впровадженням грануляції даних

Здійснимо кластеризацію множини даних у базі E еталонних описів. Процедура кластеризації відносять до апарату грануляції даних з визначенням системи центрів. Будемо розглядати два альтернативних підходи до кластеризації з однаковим числом M кластерів – сукупно для усієї множини E (система центрів W^1) і роздільно для кожного із еталонів E_i , $i = 1, \dots, N$ (система центрів W^2). У результаті маємо два варіанти центрів

$$W^1 = \{w_j^1\}_{j=1}^M, \quad W^2 = \{w_j^2\}_{j=1}^M. \quad (4)$$

Однакове число кластерів у (4) обрано нами виключно задля можливості порівняння результативності роздільного і спільного кластерного подання для еталонних описів. Фактично центри (4) є апроксимацією множини E , вони створюють нові простори систем класифікаційних ознак.

При цьому для отримання векторного подання кожного еталону за побудованою системою центрів у першому варіанті (спільна кластеризація для множини E) треба додатково на підготовчому етапі виконати конкурентний аналіз опису для кожного із еталонів E_k . Для цього для усіх елементів $z \in E_k$ визначимо номер v найближчого центру у множині W^1 , $w_j \in W^1$

$$v = \arg \min_{j=1, \dots, M} \rho(w_j, z). \quad (5)$$

Схема (5) є втіленням моделі (2) з пошуком найближчого кластерного центру. У результаті застосування (5) отримуємо подання для еталону E_k у формі цілочислового вектору $g(E_k) = [g_1, g_2, \dots, g_M]$ з кількістю компонентів, що дорівнює числу кластерів. Тут g_d – число дескрипторів опису, що потрапили у кластер з номером d . Вектор g є розподілом даних опису за набором кластерів на відміну від $\{h_i\}_{i=1}^N$ – розподілу за системою класів. У процесі класифікації для аналізованого зображення з описом Z аналогічно

також треба отримати вектор $g(Z)$ кластерного подання.

Для системи центрів W^2 вже маємо закріплення фіксованого рівноцінного числа m центрів за кожним із еталонів, причому $M = mN$, так як їх кластеризація виконувалась роздільно. Тому додаткових дій з аналізу чи трансформації еталонного представлення даних не потрібно.

Загалом число центрів m є параметром, що залежить від складу даних.

Наряду з отриманням центрів на підґрунті здійснення кластеризації розглянемо процедуру гранулювання для складу еталонних описів на підставі критерію подібності дескрипторів. Схожі трансформації застосовуються у пірамідальному поданні зображень, де впроваджено інтегроване подання у формі ієрархічної структури [16]. Близьке перетворення використано нами при трансформації дескрипторів опису шляхом складання значень бітів в рамках системи фрагментів [2].

Визначимо для опису еталону E_k потужністю в s дескрипторів трансформацію

$$T(E_k) \rightarrow E_k^*, \quad \text{card}(E_k^*) \ll s, \quad (6)$$

результатом якої є відфільтрований опис E_k^* , отриманий відбором підмножини $E_k^* \subset E_k$ компонентів із E_k з використанням деякого критерію грануляції. У прикладному аспекті пропонується орієнтовне зниження потужності опису у два рази на одному етапі оброблення.

Принцип аналізу метричної близькості дескрипторів може бути основою для побудови різноманіття варіантів фільтрації (редукції) даних. Розглянемо конкретні варіанти побудови T на основі здійснення грануляції Ω даних на підставі значення метрики ρ для дескрипторів.

1. Грануляція за порогом еквівалентності. Визначимо для кожного із еталонних дескрипторів $e_d(k) \in E_k$ найближчий до нього за метрикою ρ дескриптор $e_*(k)$ всередині множини E_k , що відповідає моделі лінійного метричного пошуку

$$e_*(k) = \arg \min_{v \neq d} \rho(e_v(k), e_d(k)). \quad (7)$$

Після визначення дескриптора з мінімальним значенням метрики перевіримо умову еквівалентності $\rho(e_d(k), e_*(k)) \leq \delta_\rho$, δ_ρ – заданий граничний параметр для значення метрики, який наближено встановлює еквівалентність двох дескрипторів. У просторі багатовимірних векторів величину δ_ρ часто визначають як 25% від максимуму метрики [4].

Фактично тут грануляція реалізована як функція Ω двох аргументів $\Omega[e_a(k), e_b(k)] \rightarrow e_*(k)$, $a \neq b$, де значення $e_*(k) \in E_k^*$ – це один із аргументів, які визначено метрично еквівалентними між собою. Функція Ω повертає значення одного із аргументів, якщо вони визнані рівнозначними на підставі істинності побудованого бінарного відношення еквівалентності. Якщо ж еквівалентність за порогом δ_ρ не встановлено, то перший аргумент заносимо до E_k^* , виключаємо цей елемент із оброблення, а другий продовжує брати участь у аналізі.

Результат такого способу грануляції залежить від параметра порогу δ_ρ , а також від порядку слідування дескрипторів у описі.

2. Грануляція з відсіюванням за найменшою відстанню. Ще один спосіб можна збудувати на підґрунті аналізу повної матриці відстаней між дескрипторами опису E_k [19]. Застосуємо відбір необхідної кількості дескрипторів шляхом відсіювання, починаючи із значення найменшої відстані. Список E_k^* поповнюємо черговим дескриптором, виключаємо його відповідні рядок і стовпець із матриці, і переходимо до аналізу наступного дескриптора з більшим значенням відстані. Таким чином послідовно набираємо потрібну фіксовану кількість дескрипторів у множину E_k^* . Ця процедура дещо складніша за попередню, може містити сортування за відстанню, зате вона не залежить від порядку слідування дескрипторів і не потребує адаптації порогу δ_ρ до даних.

Зрозуміло, що обидва розглянуті способи грануляції можна об'єднати у єдиному більш складному різновиді аналізу.

Наприклад, другий спосіб може бути доповнено використанням порогу еквівалентності.

Запропоновані способи аналізу і трансформації даних, крім результативного скорочення складу опису еталону, створюють основу для побудови ієрархії ознак, так як перетворення (6) можна застосувати повторно до уже трансформованого опису E_k^* .

За результатом здійснення етапу грануляції елементам сформованого редукованого опису E_k^* можна призначити деякі вагові коефіцієнти, які можуть бути враховані під час класифікації. Наприклад, новоствореному елементу, який отримано грануляцією значень двох близьких дескрипторів, можна присвоїти коефіцієнт 2.

А елементу, що не знайшов найближчого із заданим обмеженням – залишити вагу 1.

Класифікація у трансформованому просторі

Традиційний метод класифікації реалізує послідовність моделей (2), (3).

Класифікатор з використанням спільної кластеризації на основі системи центрів W^1 спочатку отримує за моделлю (5) векторне подання

$$g(Z) = [g_1, g_2, \dots, g_M]$$

для розпізнаваного об'єкту, а далі з використанням метрики μ для векторів цілих чисел визначає клас об'єкту як аргумент мінімуму метрики μ на множині кластерних подань для еталонів бази E

$$R_2: Z \rightarrow E_v \mid v = \arg \min_{k=1, \dots, N} \mu(g(Z), g(E_k)). \quad (8)$$

Варіантом метрики μ в (8) може бути манхетенська відстань для простору векторів цілих чисел. Використання правила (8) замість правил (2), (3) для традиційного методу дає вигреш у обсязі обчислень приблизно пропорційний величині $\gamma = s/M$.

Відмітимо, що для відсіювання зображень поза базою оптимальне значення метрики μ повинне задовольняти умов $\mu_v \geq \delta_\mu$.

Класифікатор на підставі роздільної кластеризації на етапі R_1 втілює модель (5) на множині центрів $w_j^2 \in W^2$. Клас дескриптора при цьому однозначно визначається отриманим центром кластеру, а клас усього об'єкту оцінюється за моделлю (3). Вигреш у обсязі обчислень теж пропорційний значенню γ .

На підставі отриманих редукованих множин $\{E_k^*\}_{k=1}^N$ впровадимо класифікатор з використанням грануляції, що реалізує традиційну модель (2), (3). У той же час і до множин $\{E_k^*\}_{k=1}^N$ теж можна застосувати обговорювані схеми кластеризації, що призводить до ієрархічної структури даних опису.

При дії локальних завод класифікація у розглянутих методах фактично здійснюється на підставі невикривленої частини опису аналізованого об'єкту. У цьому випадку число голосів природно знижується, тому параметри δ_ρ , δ_h треба адаптувати до допустимого рівня перешкод.

Зважаючи на те, що при способі роздільної кластеризації апроксимація здійснюється шляхом формування центрів для еталонних даних, критерієм, що впливає і на результативність класифікації, можна прийняти середню відстань між системами центрів для пар еталонів. Якщо позначити вектори для центрів пари еталонів $A = \{a_i\}_{i=1}^m$ та $B = \{b_j\}_{j=1}^m$, то такий критерій β має вигляд

$$\beta(A, B) = [\sum_{i=1}^m \sum_{j=1}^m \rho(a_i, b_j)]/m^2. \quad (9)$$

Значення (9) знаходиться у межах визначення метрики ρ . Для бінарних векторів розмірністю n і метрики Хемінга показник (9) належить діапазону $[0, n]$. Чим більше значення (9), тим маємо більший розкид між системами сформованих центрів і тим вище ступінь розрізнення класів A, B між собою.

Результативність класифікації будемо оцінювати критерієм точності pr , який відображає відношення кількості q даних з правильною оцінкою класу до загального числа Q експериментів

$$pr = q/Q. \quad (10)$$

Експериментальні результати і обговорення

Основною метою проведення експериментів було підтвердження працездатності і результативності запропонованих способів грануляції у задачі класифікації в порівняльному аспекті з традиційним методом.

При програмному моделюванні класифікатора з використанням двох способів кластеризації використано базу із п'яти зображень картин української художниці М. Приймаченко [21], а також три її інші картини поза базою класифікації.

На рис. 1 наведено дві картини із використаної бази і координати КТ дескрипторів, отриманих детектором brisk (вектор із 512 бітів) [8, 26].

У решті експериментів використано детектор orb (вектор із 256 бітів) [7].

Як бачимо, художниця використовує оригінальну техніку для своїх картин, що в той же час робить їх візуально схожими.

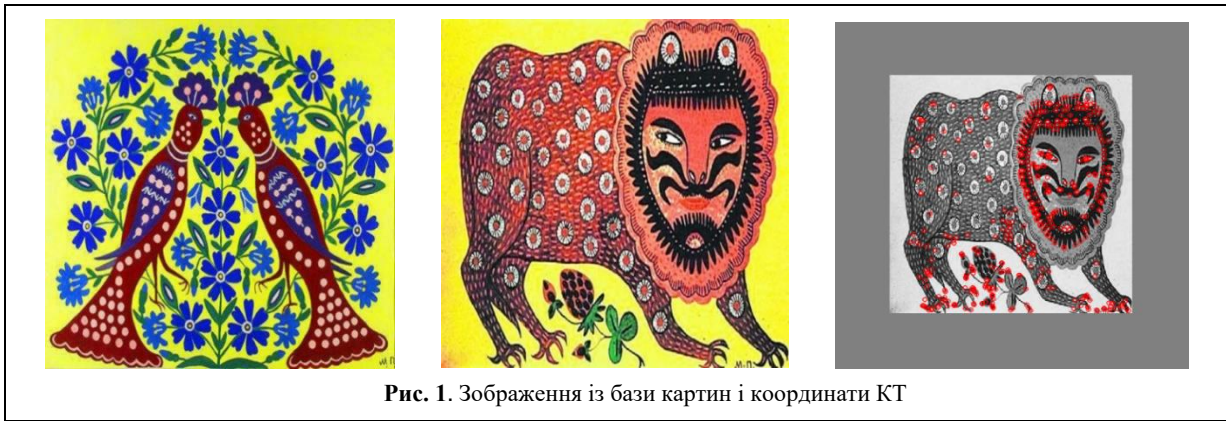


Рис. 1. Зображення із бази картин і координати КТ

Тестовий набір у експерименті склав 80 зображень і включав ці 8 картин та зображення, отримані із них геометричними перетвореннями зсуву (± 10 пікселів), повороту ($\pm 10\%$) та масштабування ($\pm 10\%$).

На рис. 2 показано тестове зображення і його координати КТ.

Для кластеризації використано $M = 15$ кластерів для бази, і $m = 3$ для кожного із еталонів.

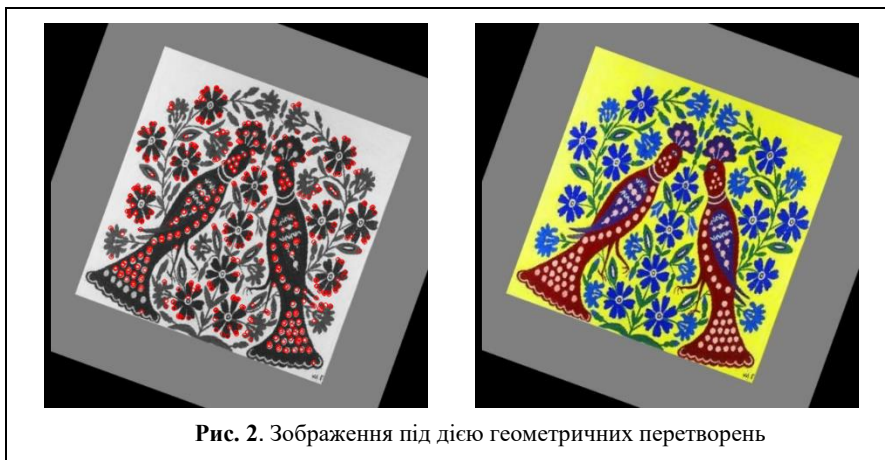


Рис. 2. Зображення під дією геометричних перетворень

Проведено попередні експерименти, щоб визначити для наявної бази зображень параметри порогів δ_p , δ_n для відстані Хемінга між дескрипторами та центрами і необхідного для прийняття рішення числа голосів дескрипторів.

Зважаючи на те, що результативність безпосередньо залежить від процедури кластерування, у обраній схемі кластеризації k -середніх початкові дані для центрів кластерів визначалися випадковим чином із складу опису. При цьому застосована модифікована процедура обчислення поточних центрів шляхом усереднення числа бітів з формуванням бінарного подання для новоствореного центру [10, 19].

У результаті проведених експериментів визначено параметри $\delta_p = 192$ для методів з кластеруванням, $\delta_n = 300$ для традиційного методу, $\delta_\mu = 180$ для методу з кластеруванням повної бази і $\delta_n = 140$ для методу з роздільним кластеруванням еталонів.

Результати експериментів засвідчили таке. Традиційний метод показав найвищу точність $pr = 1$, успішно класифікуючи усі зображення із бази і відділивши зображення не із бази. Модифікований метод з кластеризацією повної бази показав точність $pr = 0,82$, а з роздільною – $pr = 0,90$. Як бачимо, роздільна кластеризація дає дещо кращий рівень точності. При застосуванні числа кластерів $m = 5$

для кожного еталону точність методу з кластеризацією повної бази зросла до $pr = 0,88$, а з роздільною кластеризацією – до $pr = 0,98$.

Заміри часу класифікації одного зображення: традиційний метод – 6500 мс, для $m = 3$ із спільною кластеризацією – 41 мс, із роздільною – 42 мс.

Як бачимо, введення кластерування даних у процес класифікації суттєво знизило обчислювальні витрати (у нашому експерименті – у 154 рази!) за рахунок деякого зниження показника точності.

За моделлю (9) середні відстані між системами центрів першого і другого еталонів склали для трьох кластерів – 168,2, для п'яти – 182,8 (максимум відстані – 512). Як бачимо, відстань зростає зі збільшенням числа кластерів.

Зауважимо, що отримані показники точності безпосередньо залежать від обраної бази зображень, а також від набору альтернативних зображень не із бази, і можуть змінюватися у будь-яку сторону. Ключовим фактором є візуальна схожість аналізованих зображень між собою, що відображається і у близькості дескрипторів. У той же час значний вигравш у часі оброблення при застосуванні кластерування є досить об'єктивним фактором.

Моделювання за схемою грануляції (6) здійснювалося з використанням зображень птахів за анало-

гічною схемою із 80 зображень (5 із бази та 3 поза базою) і дією фіксованого набору геометричних перетворень (рис. 3). Експериментально вивчено обидва запропоновані способи грануляції – за порогом еквівалентності і з відсіюванням за найменшою відстанню. Значення порогу для визначення еквівалент-

ності для пари дескрипторів обрано як $\delta_p = 64$ як 25% від максимальної відстані 256.

Після здійснення грануляції на редукованому описі еталонів застосовано традиційний метод класифікації (2), (3) з підрахунком голосів для дескрипторів вхідного об'єкту.



Рис. 3. Зображення із бази птахів і координати ключових точок

Для підтвердження ефективності введення грануляції взято простий спосіб скорочення складу еталонних дескрипторів у два рази шляхом відсіювання через один без застосування апарату грануляції.

Встановлено, що точність класифікації для обраної бази зображень залишилася на найвищому рівні $rg = 1$ як для повного складу дескрипторів опису еталонів, так і для редукованого складу з грануляцією і без неї. При цьому число голосів для перетворених еталонів без застосування грануляції склало 340–440 із 500 і є приблизно на 5–10% нижче, ніж із використанням грануляції. Як бачимо, у нашому експерименті грануляція на базі зображень птахів дає дещо кращу результативність, ніж кластерне подання для бази картин. Проведено пробні експерименти із повторним впровадженням грануляції до уже гранульованого опису, що здійснює перехід на наступний рівень ієрархічного подання даних. Число дескрипторів із використанням адаптованого до бази порогу δ_p еквівалентності у 37% (значення 96) знизилось поетапно від 500 до 250...255, і далі до 140...160, що призвело до виграшу у швидкодії оброблення приблизно у три рази.

Усі еталонні зображення на двох рівнях гранульованого подання класифікуються правильно. У той же час максимальне число голосів для еталонних вхідних зображень з повним складом опису у 500 дескрипторів знижується на першому рівні до 410...450, а на другому рівні – до 300...400. Це вказує на перспективу впровадження наступних рівнів грануляції з ще більшим ступенем виграшу у швидкодії.

Висновки

Впровадження засобів грануляції даних у структурних методах класифікації зображень показало їх працездатність і достойний рівень результативності

із можливістю суттєвого прискорення часу оброблення. Обчислювальні затрати на класифікацію знижуються пропорційно ступеню скорочення опису еталонів і числа кластерів. Результативність класифікації за критерієм точності залежить від встановлених експериментально значень параметрів порогів та способу грануляції, обраного для конкретної бази розпізнаваних зображень.

Якщо порівнювати ефективність способів із сукупною і роздільною кластеризацією для описів бази еталонів, то із застосуванням кластерування к-середніх метод з роздільною кластеризацією показав вищий рівень точності класифікації. За обсягом часових затрат при однаковому числі кластерів ці способи приблизно рівноцінні. Обидва застосовують по-елементний метричний аналіз опису вхідного зображення, що забезпечує відсіювання завад. Рівень апроксимації для обох способів може бути покращено збільшенням числа кластерів, але за рахунок зростання обсягу обчислень.

Відпрацьована нами на практиці технологія визначення порогів, що використовуються у класифікаторі за множиною дескрипторів ключових точок, дає можливість універсально застосувати запропоновані методи для довільної бази зображень.

Перспективою дослідження може бути прикладне застосування розроблених способів для об'ємних колекцій зображень. Ще одним напрямком є вивчення ефективності щодо більш високих рівнів гранульованого подання описів.

Результати даного дослідження отримано за міжнародним науковим проектом INITIATE у рамках гранту № 101136775-HORIZON-WIDERA-2023-ACCESS-03.

СПИСОК ЛІТЕРАТУРИ

1. Tymchyshyn R., Volkov O., et al. (2018), Modern Approaches to Computer Vision, Control systems and computers, 6, 46–73.
2. Daradkeh Y. I., Gorokhovatskyi V., Tvoroshenko I., Gadetska S., and Al-Dhaifallah M. (2023), Statistical data analysis models for determining the relevance of structural image descriptions, IEEE Access, vol. 11, pp. 126938–126949, Nov. 2023, doi: 10.1109/ACCESS.2023.3332291.
3. Gorokhovatskyi V. Vlasenko N. (2021), The image description reduction in the set of descriptors on informativeness metric criteria base. Advanced Information Systems, 5 (4), 10–16. DOI: <https://doi.org/10.20998/2522-9052.2021.4.02>

4. P. Flach. (2012), Machine learning. The Art and Science of Algorithms that Make Sense of Data. New York, NY, USA: Cambridge University Press.
5. Gorokhovatskyi V. A. (2018), Image classification methods in the space of descriptions in the form of a set of the key point descriptors, Telecommunications and Radio Engineering, 77(9), 787-797.
6. Kuchuk H., Kovalenko A., Ibrahim B.F., Ruban I. (2019), Adaptive compression method for video information. International Journal of Advanced Trends in Computer Science and Engineering. Vol. 8 (1). P. 66–69. DOI: <http://dx.doi.org/10.30534/ijatcsc/2019/1181.22019>
7. Rublee E., Rabaud V., Konolige K., and Bradski G., ORB: An efficient alternative to SIFT or SURF, in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2564–2571, doi: 10.1109/ICCV.2011.6126544.
8. Leutenegger S., Chli M. and Siegwart R. Y. BRISK: Binary robust invariant scalable keypoints, in 2011 Int. Conf. on Computer Vision, Barcelona, Spain, pp. 2548–2555, 2011.
9. Gorokhovatskyi V., Tvoroshenko I. (2023), Identification of visual objects by the search request. International scientific symposium «Intelligent Solutions-S». Computational intelligence. Decision making theory: proc. of the international symposium, September 28, 2023, Kyiv-Uzhorod, Ukraine, pp. 25-27.
10. Gorokhovatskyi O., Gorokhovatskyi V. and Peredrii O. (2018), Analysis of application of cluster descriptions in space of characteristic image features, Data, vol. 3, no. 4, pp. 52.
11. Gorokhovatsky V., Putyatin Y. and Stolyarov V. (2017), Research of Effectiveness of Structural Image Classification Methods using Cluster Data Model, Radio Electronics, Computer Science, Control, vol. 3 (42), pp. 78–85.
12. Robnik-Sikonja M., Kononenko, I. (2003), Theoretical and empirical analysis of ReliefF and RReliefF. Machine Learning 53 (1-2): 23-69.
13. Stańczyk U. (2015), Feature Evaluation by Filter, Wrapper, and Embedded Approaches. In: Stańczyk U., Jain L. (eds) Feature Selection for Data and Pattern Recognition. Studies in Computational Intelligence, Springer, Berlin, Heidelberg, vol. 584, 568 p.
14. Xiong, H. and Z. Li. (2014), Data Clustering: Algorithms and Application. 1st ed. Boca Raton: CRC Press.
15. Zhang X., Yu F. X., Karaman S. and Chang S.-F. (2017), Learning discriminative and transformation covariant local feature detectors,” in IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 4923–4931.
16. Lowe D. G. (2004), Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision. 60 (2). [doi:10.1023/B:VISI.0000029664.99615.94](https://doi.org/10.1023/B:VISI.0000029664.99615.94)
17. Crowley, J., Riff O. (2003), Fast computation of scale normalized Gaussian receptive fields, Proc. Scale-Space'03, Isle of Skye, Scotland, Springer Lecture Notes in Computer Science, 2695.
18. Putyanin, E.P., Averin, S.I. (1990), Image processing in robotics. Moscow, Mashinostroeniye, 320 p.
19. Gorokhovatskyi V., Tvoroshenko I., and Yakovleva O. (2024), Transforming image descriptions as a set of descriptors to construct classification features, Indonesian Journal of Electrical Engineering and Computer Science, vol. 33, no. 1, pp. 113-125, doi: 10.11591/ijeecs.v33.i1.
20. Martino A., De Santis E., Rizzi A. (2023), On Information Granulation via Data Filtering for Granular Computing-Based Pattern Recognition: A Graph Embedding Case Study. SN Computer Science, 4:314, <https://doi.org/10.1007/s42979-023-01716-1>
21. WikiArt. Енциклопедія візуальних мистецтв. <https://www.wikiart.org/uk/mariya-primachenko> (01.04.2024)
22. Gorokhovatsky V.A. Putyatin Y. P. (2009), Image Likelihood Measures of the Basis of the Set of Conformities. Telecommunications and Radio Engineering, 68 (9), p. 763–778.
23. Tvoroshenko I., and Zarivchatskyi R. (2020), Analysis of existing methods for searching object in the video stream, Abstracts of VI International Scientific and Practical Conference «About the problems of science and practice, tasks and ways to solve them» (October 26-30, 2020). Milan, Italy, pp. 500–505.
24. Yakovleva, O., Kovtunencko, A., Liubchenko, V., Honcharenko, V., & Kobylin, O. (2023), Face Detection for Video Surveillance-based Security System (COLINS-2023). In CEUR Workshop Proceedings (Vol. 3403). pp. 69-86.
25. Kuchuk, H., Podorozhniak, A., Liubchenko, N., and Onishchenko, D. (2021), System of license plate recognition considering large camera shooting angles, Radioelectronic and Computer Systems, 4(100), 82–91.
26. OpenCV. URL: <https://docs.opencv.org/>

Received (Надійшла) 08.04.2024

(Accepted for publication) Прийнята до друку 26.06.2024

Metric granulation methods for image description in the image classification problem

Volodymyr Gorokhovatskyi, Natalia Stiahlyk, Yehor Mazur, Anna Vechirska

Abstract. The subject of the research of the article is the methods of image classification in computer vision systems. The aim is to improve the structural methods of classification to increase their speed by implementing a reduced classification features system based on the granulation of reference descriptions. Applied methods: BRISK and ORB key point detectors, set theory and vector spaces apparatus, metric models of vector clustering and granulation, “bad of words” and voting for classification models, software modeling. Results are obtained: models for the transformation of image description in the form of a set of cluster centers and the construction of a reduced description using granulation based on the similarity of vectors were developed, high-speed image classification methods based on the implementation of transformed descriptions were researched. The effectiveness of the developed modifications of the classifier depends on the data granulation method and the threshold parameters for the equivalence of vectors, the minimum metric value and the number of votes of the descriptors for the winner class in the voting scheme. The construction of a hierarchical system of features with several granulation levels was proposed. Based on the implementation of modifications, it was possible to reduce the computational costs by hundreds of times while ensuring the effectiveness of classification on the training sample of data. The practical significance of the work includes the software models for high-speed classification, confirming the functionality and accuracy of the proposed modifications for the applied image dataset, creating software application for implementing modifications of classifiers in computer vision.

Keywords: computer vision; structural methods of image classification; set of descriptors; clustering; vector granulation; classification accuracy.

Sergii Datsenko¹, Denys Tarasenko²¹ National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine² Cherkasy State Technological University, Cherkasy, Ukraine

RESEARCH OF INDUSTRIAL INTERNET OF THINGS NETWORKS AND SYSTEMS

Abstract. The article discusses the concept of the Industrial Internet of Things. The main protocols for data transmission of the Industrial Internet of Things were identified. IIoT systems have high requirements for the reliability of message delivery, which is why special industry data transmission protocols used in industrial automation are often used to develop IIoT systems. IIoT technologies are an integral part of the Internet of Things concept. The paper presents the architecture of an automated enterprise based on IIoT, highlighting its advantages and disadvantages.

Keywords: IoT IIoT, automation, integration.

Introduction

Currently, the concept of the Internet of Things is actively developing, which implies the active implementation of embedded technologies in modern infocommunication networks [1–4]. This concept covers an increasing number of areas of human activity every day, such as unmanned transport control, urban infrastructure management, control of medical operations, automation of residential premises, offices, etc. [5–8]. Recently, a new direction has been developing within the concept of the Internet of Things (IoT) - the industrial Internet of Things (IIoT), which affects the issues of creating a heterogeneous intelligent system for automating industrial enterprises. IIoT systems are used in a variety of production areas, such as agriculture, electronic equipment manufacturing, mechanical engineering, machine tool manufacturing, automation of data collection and accounting from energy networks, production of aviation, space and military equipment, etc. Currently, the implementation of technologies related to the industrial Internet of Things is associated with problems both with the integration of these technologies with outdated solutions for industrial automation, and with the absence of any automation system at the enterprise [9–14]. The solution to these problems lies in the development of unified technological standards for the integration of both industrial automation systems and

equipment that does not have digital control interfaces with IIoT systems [15, 16].

Basic material

There are a number of areas in industrial enterprises whose work is subject to automation. In addition to the obvious automation of production equipment, there are a number of areas subject to automation within the enterprise and related to human safety and social activities. The following areas should be highlighted as the main areas of automation [17]:

- automation of production equipment;
- automation of monitoring and management of the state of the enterprise's products;
- automation of business applications, such as enterprise resource planning (ERP), customer relationship management (CRM), product lifecycle management (PLM), manufacturing execution systems (MES), human resource management (HRM), etc.;
- automation of multimedia systems for monitoring industrial enterprise safety;
- automation of local and global positioning systems;
- automation of collection and analysis of openly accessible data from the Internet.

IIoT technologies are an integral part of the Internet of Things concept. Fig. 1 shows an architecture that consists of four functional and two non-functional layers.

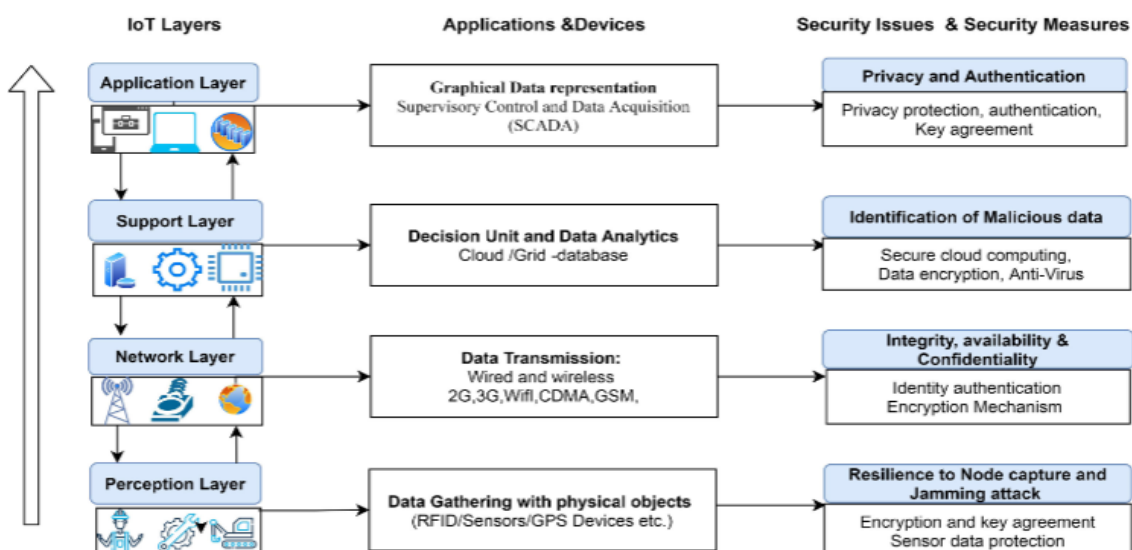


Fig. 1. IIoT layered architecture

The functional layers include the following [18]:

1. Application layer.
2. Service support and application support layer.
3. Network layer.

4. Device layer.

Fig. 2 shows the functional architecture of smart manufacturing systems based on Industrial Internet of Things solutions.

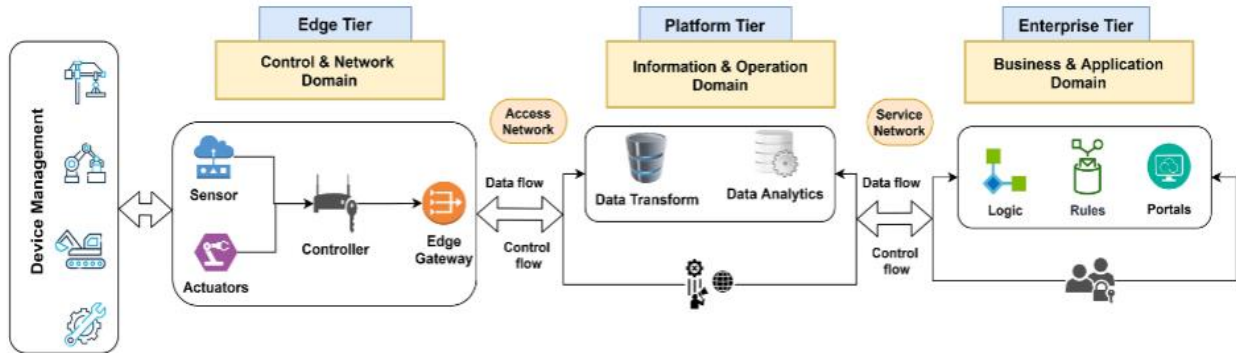


Fig. 2. Three T IIoT System architecture

It is divided into the following levels [19]:

1. **Life Cycle Management Level.** This level is responsible for managing the various life cycles of the manufacturing enterprise and covers a wide range of systems in the manufacturing enterprise, such as systems for design, production, management and maintenance of manufactured products.

The following can be distinguished as the main life cycles of an enterprise:

- The product cycle, which includes the design and development of the corresponding production system, design, development, production, testing, maintenance of the manufactured product, use of the product by the user, recycling or destruction of the used product.
- The production system cycle, which includes the design, assembly, operation, maintenance and decommissioning of the entire production system.
- The implementation cycle, which includes functions related to the interaction of the supplier and the customer.

Other life cycles that can also be included in the system under consideration, if there are special requirements for a “smart enterprise”.

2. **The integration level,** which ensures the integration of all resources, systems and processes involved in the various life cycles associated with the creation of products, through all levels of the production system to create an environment for smart manufacturing applications. This level includes the following types of integration:

- Vertical integration, which is responsible for ensuring that new technological solutions are included in an existing system through higher levels of the production system.
- Horizontal integration, which is responsible for ensuring that new technological solutions are integrated into an existing system at one system level.
- End-to-end integration, which allows connecting each phase of the life cycle of the manufactured product through the entire chain of product implementation, including through various enterprises.

3. **The application layer,** which is responsible for implementing end applications used to solve various

problems that arise in smart manufacturing. Examples include the following applications:

- Innovative manufacturing is a type of “smart manufacturing” and includes applications such as virtual manufacturing, flexible manufacturing, and customized manufacturing. Virtual manufacturing allows for the simulation of production processes at the enterprise and computer modeling. Flexible manufacturing allows for a quick response to predictable and unpredictable changes. Customized manufacturing allows for the sale of goods in a way that meets the needs of each individual customer.

- Preventive maintenance is responsible for analyzing data coming from production tools equipped with various types of sensors that collect information about the current state of the equipment, and making decisions based on the analysis results, which can potentially lead to the prevention of equipment failures.

To highlight the task that will be revealed in this paper, we will highlight the main data transfer protocols of the Industrial Internet of Things. IIoT systems have high requirements for the reliability of message delivery, as a result of which IIoT system developers very often use special industry data transfer protocols used in the field of industrial automation. The following technologies were selected for the study of IIoT data transfer protocols: CoAP (Constrained Application Protocol)

- MQTT (Message Queuing Telemetry Transport);
- XMPP (Controller Area Network);
- HTTP (HyperText Transfer Protocol);
- ModBus;
- OPC UA (Open Platform Communications Unified Architecture).

Currently, there is an active development and implementation of such a class of devices in the networks of the Internet of Things and the Industrial Internet of Things, such as gateways. These devices are responsible for ensuring the interaction of specific communication technologies and data transfer of IoT and IIoT, both among themselves and with public communication networks.

Conclusions

The article discusses the concept of the Industrial Internet of Things. The main protocols for data transmission of the Industrial Internet of Things were identified. IIoT systems have high requirements for the reliability of message delivery, which is why special

industry data transmission protocols used in industrial automation are often used to develop IIoT systems. IIoT technologies are an integral part of the Internet of Things concept.

The article presented the architecture of an automated enterprise based on IIoT, highlighting its advantages and disadvantages.

REFERENCES

1. Rafique, H., Shah, M. A., Islam, S. U., Maqsood, T., Khan, S. and Maple, C. (2019), "A novel Bio-Inspired Hybrid Algorithm (NBIHA) for efficient resource management in Fog computing", *IEEE Access*, vol. 7, pp. 115.760–115.773, doi: <https://doi.org/10.1109/access.2019.2924958>
2. Wang, T., Zhou, J., Chen, X., Wang, G., Liu, A. and Liu, Y. (2018), "A Three-Layer privacy preserving cloud storage scheme based on computational intelligence in FOG computing", *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2(1), pp. 3–12, doi: <https://doi.org/10.1109/tetci.2017.2764109>
3. Petrovska, I. and Kuchuk, H. (2023), "Adaptive resource allocation method for data processing and security in cloud environment", *Advanced Information Systems*, vol. 7, is. 3, pp. 67–73, doi: <https://doi.org/10.20998/2522-9052.2023.3.10>
4. Dang, L. M., Piran, M. J., Han, D., Min, K. and Moon, H. (2019), "A survey on internet of things and cloud computing for healthcare", *Electronics*, vol. 8(7), 768, doi: <https://doi.org/10.3390/electronics8070768>
5. Botta, A., De Donato, W., Persico, V. and Pescapè, A. (2016), "Integration of Cloud computing and Internet of Things: A survey", *Future Generation Computer Systems*, vol. 56, pp. 684–700, doi: <https://doi.org/10.1016/j.future.2015.09.021>
6. Díaz, M., Martín, C. L. and Rubio, B. (2016), "State-of-the-art, challenges, and open issues in the integration of Internet of things and cloud computing", *Journal of Network and Computer Applications*, vol. 67, pp. 99–117, doi: <https://doi.org/10.1016/j.jnca.2016.01.010>
7. Liu, Y., Fieldsend, J. E. and Min, G. (2017), "A framework of FOG Computing: architecture, challenges, and optimization", *IEEE Access*, vol. 5, pp. 254.45–254.54, doi: <https://doi.org/10.1109/access.2017.2766923>
8. Kuchuk, N., Mozhaiev, O., Haichenko, A., Semenov, S., Kuchuk, H., Tiulieniev, S., Mozhaiev, M., Davydov, V., Brusakova, O. and Gnusov, Y. (2023), "Devising a method for balancing the load on a territorially distributed foggy environment", *Eastern-European Journal of Enterprise Technologies*, vol. 1(4 (121)), pp. 48–55, doi: <https://doi.org/10.15587/1729-4061.2023.274177>
9. Chiang, M. and Zhang, T. (2016), "Fog and IoT: An Overview of Research Opportunities", *IEEE Internet of Things Journal*, vol. 3(6), pp. 854–864, doi: <https://doi.org/10.1109/jiot.2016.2584538>
10. Puliapito, C., Mingozzi, E. and Anastasi, G. (2017), "Fog Computing for the Internet of Mobile Things: Issues and Challenges", *2017 IEEE Int. Conf. Smart Comput*, doi: <https://doi.org/10.1109/smartcomp.2017.7947010>
11. Khan, S., Parkinson, S. and Qin, Y. (2017), "Fog computing security: a review of current applications and security solutions", *Journal of Cloud Computing*, vol. 6(1), doi: <https://doi.org/10.1186/s13677-017-0090-3>
12. Hamdan, S., Ayyash, M. and Almajali, S. (2020), "Edge-Computing Architectures for Internet of Things Applications: A survey", *Sensors*, vol. 20(22), 6441, doi: <https://doi.org/10.3390/s20226441>
13. Li, Y., Qi, F., Wang, Z., Yu, X. and Shao, S. (2020), "Distributed edge Computing offloading algorithm based on deep reinforcement learning", *IEEE Access*, vol. 8, pp. 85.204–85.215, doi: <https://doi.org/10.1109/access.2020.2991773>
14. Premsankar, G., Di Francesco, M. and Taleb, T. (2018), "Edge Computing for the Internet of Things: a case study", *IEEE Internet of Things Journal*, vol. 5(2), pp. 1275–1284, doi: <https://doi.org/10.1109/jiot.2018.2805263>
15. Xue, H., Huang, B., Qin, M., Zou, H. and Yang, H. (2020), "Edge Computing for Internet of Things: A Survey", *2020 Int. Conf. Internet Things IEEE Green Comput. Commun. IEEE Cyber. Phys. Soc. Comput. IEEE Smart Data IEEE Congr. Cybermatics*, IEEE, doi: <https://doi.org/10.1109/ithings-greencom-cpscom-smartdata-cybermatics50389.2020.00130>
16. Liu, J., Mao, Y., Zhang, J. and Letaief, K. B. (2016), "Delay-optimal computation task scheduling for mobile-edge computing systems", *2016 IEEE Int. Symp. Inf. Theory*, IEEE, doi: <https://doi.org/10.1109/isit.2016.7541539>
17. Kuchuk, N., Ruban, I., Zakovorotnyi, O., Kovalenko, A., Shyshatskyi, A. 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, 194480, doi: <https://doi.org/10.1109/khpiweek61412.2023.10312856>
18. Παλαεωργίου, Α., Poormohammady, E. and Cheng, B. (2016), "Edge-Computing-Aware Deployment of Stream Processing Tasks Based on Topology-External Information: Model, Algorithms, and a Storm-Based Prototype", *2016 IEEE Int. Congr. Big Data*, doi: <https://doi.org/10.1109/bigdatacongress.2016.40>
19. Zhou, J., Leppänen, T., Harjula, E., Ylianttila, M., Ojala, T., Chen, Y., Jin, H. and Yang, L. T. (2013), "CloudThings: A common architecture for integrating the Internet of Things with Cloud Computing", *2013 IEEE 17th Int. Conf. CSCWD 2013*, doi: <https://doi.org/10.1109/cscwd.2013.6581037>

Received (Надійшла) 12.05.2024

Accepted for publication (Прийнята до друку) 14.07.2024

Дослідження мережі і систем промислового Інтернету Речей

С. С. Даценко, Д. А. Тарасенко

Анотація. У статті розглянуто концепцію промислового Інтернету речей. Було виділено основні протоколи передачі даних промислового Інтернету речей. Системи IIoT мають високі вимоги до показників надійності доставки повідомлень, внаслідок чого часто-густо для розробки систем IIoT використовують спеціальні галузеві протоколи передачі даних, що використовуються у сфері промислової автоматизації. Технології IIoT є складовою концепції Інтернету речей. У роботі представлено архітектуру автоматизованого підприємства на основі IIoT, виділено її переваги та недоліки.

Ключові слова: IIoT, IIoT, автоматизація, інтеграція.

В. О. Дяченко, Є. Б. Колісник, А. О. Ляшова, О. О. Можасв

Харківський національний технічний університет радіоелектроніки, Харків, Україна

МЕТОДИ МАРШРУТИЗАЦІЇ ДЛЯ ЗАБЕЗПЕЧЕННЯ ЯКОСТІ МЕРЕЖНОГО СЕРВІСУ

Анотація. Актуальність. Актуальність методів маршрутизації для забезпечення якості мережного сервісу залишається високою в сучасному цифровому світі. Суттєве збільшення обсягів передачі даних у мережах зумовлене зростанням кількості підключених пристроїв і розвитком інтернету речей. Це вимагає більш ефективного управління пропускнуою спроможністю та затримкою для забезпечення відповідної якості обслуговування. Сучасні мережі використовуються для широкого спектра застосунків: від веб-перегляду та електронної пошти до відеоконференцій, онлайн-ігор, і потокового відео високої чіткості. Кожен застосунок має свої унікальні вимоги до якості сервісу, що зумовлює потребу в різноманітних методах маршрутизації для забезпечення оптимальної продуктивності. Слід також зазначити, що збільшення популярності потокових мультимедійних сервісів, таких як Netflix, YouTube, і Spotify, вимагає високої пропускнуої спроможності та низької затримки для забезпечення задовільного рівня сервісу. З огляду на ці фактори, методи маршрутизації для забезпечення якості мережного сервісу залишаються ключовим елементом в проектуванні та управлінні мережевою інфраструктурою, а їх розвиток і оптимізація є невід'ємною частиною забезпечення високої продуктивності та надійності мережевих сервісів. **Метою даної роботи** є дослідження існуючих методів маршрутизації для забезпечення якості мережного сервісу. **Об'єктом дослідження** є методи багатопоточної маршрутизації комп'ютерних мереж. **Предметом дослідження** є метод керування контролером програмно-конфігурованої мережі для реалізації побудованих маршрутів багатопоточного з'єднання. **Результати.** Проведено аналіз існуючих методів багатопоточної маршрутизації. Проаналізовано існуючі роботи з аналізу ефективності багатопоточної маршрутизації. Для проведення порівняльного аналізу обрані протоколи Multipath TCP та Flow Distribution Management Protocol як представники статичного та динамічного методів. Основною методологією дослідження є поєднання імітаційного моделювання та статистичних тестів. Основна гіпотеза полягає в тому, що динамічні методи будуть перевершувати статичні методи багатопоточної маршрутизації. Ця гіпотеза перевіряється на основі моделювання поведінки багатопоточних методів маршрутизації з розглянутих стратегій роботи контролера SDN. Наведено опис методу управління контролерами в програмно-конфігурованих мережах для реалізації побудованих маршрутів багатопоточного з'єднання.

Ключові слова: багатопоточна маршрутизація, метод, програмно-конфігурована мережа, QoS, мережний сервіс, методика, мультимедійний сервіс, Multipath TCP, Flow Distribution Management Protocol.

Вступ

Методи забезпечення якості обслуговування (Quality of Service, QoS) є критично важливими для багатопоточної маршрутизації, оскільки вони дозволяють керувати пріоритетами трафіку, забезпечуючи відповідні рівні пропускнуої спроможності, затримки та надійності для різних типів мережевого трафіку. Актуальність QoS у сучасних мережах не може бути переоцінена, особливо з огляду на зростаючі вимоги до мережевих ресурсів та диверсифікацію застосунків, які використовуються як у бізнес-сфері, так і в повсякденному житті. QoS є критично важливою концепцією та набором технологій, що забезпечують пріоритизацію трафіку, гарантію пропускнуої спроможності, мінімізацію затримок та забезпечення надійності для різних видів мережевих застосунків. Потокове відео, відеоконференції, VoIP (Voice over Internet Protocol) та інші мультимедійні застосунки мають строгі вимоги до затримок, джитеру (варіації затримки) та втрат пакетів. QoS дозволяє забезпечити стабільність та якість таких сервісів, навіть у переповнених або нестабільних мережах, а також підвищити енергозбереження [1].

З розвитком IoT, кількість підключених пристроїв та генерований ними трафік продовжують зростати. Мережі повинні ефективно обробляти величезні обсяги даних від сенсорів та інших IoT пристроїв, деякі з яких вимагають мінімальної затримки та високої надійності для критичних застосунків, наприклад, у сферах охорони здоров'я та автономного транспорту. Існують також і інші питання, які роблять забезпе-

чення QoS актуальним: додатки критичного значення, такі як телемедицина, системи управління транспортом та промисловий контроль, потребують гарантованих параметрів мережевого сервісу; QoS забезпечує необхідний рівень відповідальності та ефективності для підтримки цих застосунків; хмарні обчислення та сервіси вимагають ефективного управління мережевим трафіком для забезпечення доступності, продуктивності та безпеки. QoS є ключовим елементом у забезпеченні надійності та продуктивності хмарних сервісів, особливо при розгортанні масштабованих застосунків.

Методи маршрутизації для забезпечення QoS є важливим аспектом у проектуванні та управлінні мережами. Вони дозволяють визначати пріоритети для різних типів трафіку, забезпечуючи, щоб критично важливі дані були передані з належною швидкістю, мінімальною затримкою та надійністю. Використання методів багатопоточної маршрутизації для забезпечення QoS пропонує низку переваг, які покращують продуктивність та надійність мережі. Багатопоточна маршрутизація дозволяє використовувати декілька маршрутів для передачі даних між вузлами, оптимізуючи використання мережевих ресурсів і забезпечуючи більш високий рівень якості обслуговування. Використання кількох маршрутів для передачі даних знижує ризик втрати з'єднання у випадку виходу з ладу одного або декількох маршрутів. У випадку виникнення збою на одному шляху, трафік може бути автоматично перенаправлений через інший доступний маршрут, забезпечуючи безперебійність сервісу. Багатопоточна маршрутизація дозволяє балансувати навантаження

між різними маршрутами, ефективно розподіляючи трафік для уникнення перевантажень на певних ділянках мережі. Це забезпечує оптимальне використання доступної пропускної спроможності та зменшує ймовірність утворення "вузьких горловин". За допомогою багатопоточної маршрутизації можливо віддавати пріоритет трафіку від критично важливих застосунків, таких як VoIP або відеоконференції, забезпечуючи їм необхідну пропускну спроможність і мінімальну затримку. Таким чином, користувачі отримують високий рівень якості обслуговування навіть під час пікових навантажень на мережу. Розподіл трафіку через кілька маршрутів може зменшити загальну затримку, оскільки дані можуть бути передані через найменш завантажені або найшвидші шляхи. Це особливо важливо для застосунків, чутливих до затримок. Методи багатопоточної маршрутизації надають адміністраторам мережі розширені можливості для управління трафіком, включаючи моніторинг, аналіз і оптимізацію потоків даних у реальному часі, забезпечуючи високий рівень ефективності мережних операцій. Загалом, використання багатопоточної маршрутизації для забезпечення якості мережного сервісу дозволяє створити більш надійну, ефективну та високопродуктивну мережну інфраструктуру, здатну задовольнити сучасні вимоги до передачі даних. **Метою цієї роботи** є дослідження існуючих методів маршрутизації для забезпечення якості мережного сервісу

Основна частина

При розробці багатопоточних протоколів виникають завдання розподілу транспортних сегментів даних між відповідними підпотоками, розробки алгоритмів управління навантаженням для відповідних підпотоків, узгодження кількості підпотоків і створення маршрутів для кожного відповідного підпотoku.

Існує два класи методів для вирішення цієї проблеми: статичні та динамічні. У статичному методі кількість пов'язаних підпотоків транспортного з'єднання заздалегідь визначена в кожному конкретному випадку і не змінюється з часом. Кількість пов'язаних гілок може залежати від конфігурації користувача та кількості мережних інтерфейсів пристроїв-відправників і пристроїв-одержувачів (у цьому випадку гілки відкриваються на кожну пару інтерфейсів).

У динамічних методах кількість пов'язаних підпотоків у транспортному з'єднанні залежить від вимог до QoS застосунку та стану мережевого середовища. Якщо ресурсів на маршруті існуючого набору пов'язаних підпотоків недостатньо для забезпечення необхідної якості обслуговування, відкриваються додаткові підпотоки. У дослідженнях, спрямованих на аналіз ефективності багатопоточної маршрутизації, основним критерієм в більшості випадків є збільшення швидкості транспортного зв'язку при використанні статичних багатопотокових схем в заданому середовищі передачі даних або при заданому навантаженні.

В роботі [2] розглядається збільшення продуктивності статичного методу під час використання різних алгоритмів планування сегментів між підпотоками. Також є робота [3], де досліджують ефективність статичного методу для автомобілів, підключених до мережі. Ці автомобілі можуть координувати свою роботу між собою з міськими службами дорожнього руху, а

також виконувати допоміжні обчислення на інших мережних пристроях. Робота [4] присвячена дослідженню продуктивності статичного методу багатопоточної маршрутизації в бездротових децентралізованих мережах, що самоорганізуються. У статті [5] розглядається вплив різних алгоритмів управління навантаженням у разі статичного багатопотокового методу. В рамках цієї роботи було проведено дослідження масового використання статичного методу багатопоточної маршрутизації в рамках однієї автономної системи. Дослідження показало, що застосування статичного багатопотокового транспортного сполучення має перевагу перед однопотоковим з'єднанням протоколу TCP тільки за умови завантаження мережі, меншою 37%. Під завантаженням мережі розуміється середнє значення завантаження ліній зв'язку, де завантаження лінії зв'язку розраховувалася як відношення сумарної швидкості TCP потоків, що проходять цю лінію зв'язку, до її пропускної спроможності.

В роботі [6] проведено аналіз ефективності динамічного методу багатопоточної маршрутизації в контексті безпроводних мереж. Для проведення порівняльного аналізу статичного та динамічного методів багатопоточної маршрутизації необхідно вибрати протоколи, на яких буде досліджено ефективність роботи методів. Вибір проводився за такими критеріями як: клас методу, підтримка протоколу, опорний транспортний протокол, можливість використання довільного алгоритму управління навантаженням. Багатопотоковий транспортний протокол може відповідати статичному або динамічному методу. Для отримання актуальних достовірних результатів необхідно, щоб реалізація багатопотокового протоколу була відкритою та доступною, а також мала діючу підтримку, щоб відповідати поточним вимогам мереж. Багато протоколів використовують такі поширені транспортні протоколи, як TCP, UDP, для того щоб спростити впровадження цих протоколів в існуючу інфраструктуру мережі. Так як розвиток інших методів забезпечення якості сервісу не стоїть на місці, то важливо, щоб багатопотоковий протокол міг використовувати сучасні алгоритми управління навантаженням у своїй роботі.

Згідно проведеного аналізу робіт за тематикою досліджень, найбільш популярним багатопоточним протоколом, що підтримує статичний спосіб, є Multipath TCP (MPTCP) протокол [7]. Незважаючи на велику кількість різних багатопоточних протоколів багато хто з них є різними модифікаціями протоколу MPTCP. Інші багатопоточні протоколи або засновані на Stream Control Transmission Protocol (SCTP-CMT), або репрезентують новий транспортний протокол, який не має стандарту. В обох випадках багатопоточний протокол буде складніше інтегрувати в існуючу мережу інфраструктури. Також необхідно відзначити протокол Quick User Datagram Protocol Internet Connections (MP-QUIC) заснований на тих же принципах роботи, що і протокол MPTCP, тому протокол MPTCP був обраний як основний представник статичного протоколу багатопотокового для транспортних з'єднань для проведення порівняльного аналізу ефективності статичного і динамічного методів багатопоточної маршрутизації. Крім цього, протокол MPTCP має доступну відкриту реалізацію, що підтримується, на відміну від протоколу MP-QUIC. Побудована модель

роботи методу багатопоточної маршрутизації для Bandwidth-Need Driven Energy Efficiency Improvement of MPLS (BNDEEI-MPLS) [7] як один з критеріїв відкриття підпотіку використовує енергоефективність каналу, що мало застосовується для провідних мереж Інтернет-провайдерів. Крім того, BNDEEI-MPLS спирається на припущенні, що використовується алгоритм управління навантаженням, який визначає перевантаження по втраті пакета (loss-based алгоритми).

MPLS – це модифікація протоколу TCP, що підтримує статичний метод багатопоточної маршрутизації. MPLS є розширенням традиційного протоколу передачі даних TCP (Transmission Control Protocol), яке дозволяє встановлення одного TCP-з'єднання через кілька фізичних шляхів. Це значно покращує використання ресурсів мережі, надійність з'єднання та пропускну спроможність. MPLS дозволяє передавати дані між двома кінцевими точками через кілька мережевих інтерфейсів та шляхів одночасно, забезпечуючи ефективніше використання мережевих ресурсів. У випадку відмови одного з каналів зв'язку, MPLS автоматично перерозподіляє трафік між доступними шляхами, зменшуючи перерви в обслуговуванні та підвищуючи загальну надійність з'єднання. Додатки, які використовують TCP, не потребують змін для роботи з MPLS, оскільки весь процес управління кількома шляхами виконується на транспортному рівні без залучення додаткового програмного забезпечення. Використання декількох маршрутів дозволяє агрегувати пропускну спроможність різних мережевих з'єднань, підвищуючи загальну швидкість передачі даних.

MPLS ідеально підходить для мобільних пристроїв та багатомісних середовищ, де доступно кілька мережевих інтерфейсів (наприклад, Wi-Fi та мобільний зв'язок), забезпечуючи краще використання доступних мережевих ресурсів. Розподіл трафіку через кілька шляхів може зменшити загальну затримку зв'язку, особливо для додатків, чутливих до затримок, таких як голосовий зв'язок та відеоконференції. Використання MPLS підвищує стійкість мережі до відмов, оскільки автоматичне переключення між маршрутами дозволяє уникнути перерв у з'єднанні через проблеми з одним або декількома каналами.

Протокол Flow Distribution Management Protocol (FDMP) не є широко відомим або стандартизованим протоколом у сфері мережевих технологій чи мережевої інженерії, ані не згадується як загальноприйнятий стандарт або протокол. В контексті мережевих комунікацій, ідея "управління розподілом потоків" може бути асоційована з різними механізмами та стратегіями для оптимізації проходження даних через мережу, включаючи маршрутизацію, балансування навантаження, та управління QoS. Такі механізми спрямовані на підвищення ефективності мережі, забезпечення надійності з'єднань, та гарантію, що критично важливі потоки даних обробляються з належним пріоритетом. Завдання вибору та встановлення маршруту в комп'ютерній мережі відіграє критично важливу роль у забезпеченні ефективності та надійності передачі даних між вузлами мережі. Цей процес включає в себе ряд кроків та механізмів, які дозволяють визначити найоптимальніший шлях для передачі пакетів даних від джерела до призначення. При статичній

маршрутизації маршрути встановлюються вручну адміністратором мережі. Цей метод є простим та передбачуваним, але не має гнучкості для автоматичного адаптування до змін у мережі, таких як відмови обладнання чи зміни в топології мережі. Використовуючи протоколи динамічної маршрутизації (наприклад, OSPF, EIGRP, BGP), маршрутизатори автоматично обмінюються інформацією про доступні шляхи та визначають оптимальні маршрути в реальному часі. Це забезпечує вищу стійкість та адаптивність мережі.

Маршрутизатори використовують таблиці маршрутизації для зберігання інформації про відомі маршрути. Коли пакет даних надходить до маршрутизатора, він використовує цю таблицю для визначення найкращого шляху до призначення. Для визначення найкращого маршруту маршрутизатори використовують різні метрики, такі як пропускну спроможність, затримка, кількість хопів (переходів через маршрутизатори), навантаження та надійність. Вибір маршруту залежить від алгоритму маршрутизації та налаштувань протоколу. Деякі протоколи маршрутизації підтримують балансування навантаження, дозволяючи використовувати кілька маршрутів для розподілу трафіку і таким чином покращуючи загальну пропускну спроможність та надійність мережі. Динамічна маршрутизація дозволяє мережі швидко адаптуватися до змін, автоматично переконфігуруючи маршрути в разі відмови обладнання чи інших змін у топології мережі.

Для забезпечення необхідного рівня якості для різних типів трафіку, таких як голосовий зв'язок або відео, можуть застосовуватися спеціальні правила QoS, які дозволяють віддавати пріоритет або резервувати пропускну спроможність для критично важливих даних. Завдання вибору та встановлення маршруту в комп'ютерній мережі є фундаментальним елементом мережевої інфраструктури, який впливає на продуктивність, надійність та загальну ефективність мережевих комунікацій. У програмно-конфігурованих мережах використовується централізоване управління, яке вносився в окремий логічний пристрій. Контролер програмно-конфігурованої мережі знає поточний стан мережі та може керувати установкою побудовою маршрутів. Централізоване управління програмно-конфігурованих мереж відноситься до парадигми, в якій управління мережею та її конфігурація відбуваються через централізований програмний інтерфейс. Цей підхід є ключовим елементом технології Software-Defined Networking (SDN), яка дозволяє адміністраторам мережі динамічно управляти мережевими ресурсами, оптимізувати трафік та автоматизувати процеси настройки мережі через програмне забезпечення. В традиційних мережах план управління (логіка маршрутизації, правила доступу тощо) інтегрований безпосередньо в обладнання. У програмно-конфігурованих мережах ці площини розділені. Логіка управління вносився на централізований контролер, що дозволяє більш гнучке та ефективне управління мережею. Ядром SDN є централізований програмний контролер, який комунікує з мережевими пристроями через стандартизований інтерфейс (наприклад, OpenFlow). Контролер має глобальний огляд усієї мережі, що дозволяє ефективно управляти потоками даних, маршрутизацією та політиками безпеки. Централізоване управління дозволяє автоматизувати багато аспектів конфігурації мережі, зменшуючи

витрати на її обслуговування та впроваджуючи зміни швидше та з меншим ризиком помилок. Також це забезпечує можливість швидко масштабувати мережеві ресурси відповідно до змінних потреб. SDN дозволяє мережі бути гнучкою та адаптивною до змін у вимогах до трафіку та політик управління. Це включає можливість швидко реагувати на зміни в мережевому навантаженні, автоматично перенаправляючи трафік для оптимізації продуктивності та надійності. Централізоване управління дозволяє ефективніше імплементувати політики безпеки на всій мережі, забезпечуючи консистентність та оновлення в реальному часі. Це також полегшує виявлення та реагування на мережеві загрози.

Важливим питанням є встановлення правил SDN для реалізації побудованих маршрутів багатопоточного з'єднання. Існують дві стратегії роботи контролера SDN: реактивна та проактивна. У реактивній стратегії контролер будує маршрут і встановлює правила щодо настання деякої події, наприклад, появи першого пакета нового багатопоточного з'єднання в мережі. Для цього контролер SDN налаштовує підконтрольні йому комутатори для перехоплення всіх пакетів TCP і відбирає пакети, які відносяться до багатопоточного з'єднання. У термінах протоколу OpenFlow [8], що розглядається в роботі як протокол взаємодії між контролером SDN і підконтрольними комутаторами, контролер встановлює на комутатори низькопріоритетні правила для відправлення PacketIn повідомлень на контролер для кожного TCP пакета. При надходженні PacketIn повідомлення контролер аналізує опціональну частину транспортного заголовка та шукає опції, що стосуються багатопоточного протоколу MPTCP або FDMP. Якщо відповідні опції присутні, то контролер сприйматиме цей TCP потік як підпотік багатопоточного з'єднання.

Щоб визначати, до якого багатопоточного з'єднання відноситься підпотік, що розглядається, контролер зберігає інформацію про кожен підпотік в Subflow Information Base (SIB). При надходженні першого потоку багатопоточного з'єднання в мережу контролер витягує ключ, що передається в опції MP_CAPABLE. Якщо пакет відповідає ініціації з'єднання, тобто пакет з прапором SYN, то створюється новий запис в SIB і зберігається ключ відправника. Інакше запис вже створено у SIB і зберігається ключ одержувача. На основі двох ключів контролер може обчислити хеш, який передаватиметься в наступних споріднених потоках. Виявлення розрахованого хешу в пакетах інших потоків означатиме належність невідомого до того моменту потоку до багатопоточного з'єднання, інформацію про яке тільки збережено в SIB.

За допомогою SIB контролер може визначати споріднені потоки та прокладати для них маршрути з найменшою кількістю перетинів. Наступні пакети після процедури повинні йти згідно з правилами, встановленими для відповідних маршрутів. Кожен підпотік визначається правилом OpenFlow за парою IP адрес та парою транспортних портів. Щоб підтримувати передачу пакетів в обидві сторони, контролеру необхідно встановити правил удвічі більше довжини маршруту. На противагу реактивній стратегії проактивна стратегія передбачає попереднє завантаження правил передсилання пакетів на етапі ініціалізації контролера

SDN. Отже, комутатори SDN мають обробляти пакети повністю самостійно, без втручання контролера SDN. Зокрема, комутатори SDN мають розподіляти споріднені підпотоки різними маршрутами, що мають мінімальну кількість перетинів. Самі маршрути можуть бути побудовані заздалегідь, і відповідні правила можуть бути встановлені заздалегідь, однак пакети MPTCP або FDMP не несуть жодних додаткових ознак, які можуть допомогти у визначенні приналежності тому чи іншому багатопоточному з'єднанню. Можна використовувати хеш функції для балансування за побудованими маршрутами, проте, як було зазначено раніше, відсутня гарантія того, що родинні підпотоки вирушають різними маршрутами. Одним із можливих підходів для запобігання збігу маршрутів споріднених під потоків може стати відстеження багатопоточних з'єднань у комутаторах SDN так само, як це зроблено у реактивній стратегії в контролері SDN. Однак переважним буде підхід, коли транспортні агенти допомагають комутаторам визначати споріднені підпотоки.

Пропонується використовувати змішану стратегію для реалізації багатопоточної маршрутизації у SDN. Контролер SDN заздалегідь буде маршрути з мінімальною кількістю перетинів між кожною парою точок у мережі. Кількість маршрутів визначається зі ступеня демультимплексування мережі. Побудувавши маршрути заздалегідь, ми можемо зменшити затримку на встановлення з'єднання, тому що не потрібно перераховувати їх щоразу при надходженні нового підтоку. Крім того, ми можемо заздалегідь встановити правила для кожного з маршрутів за винятком точки входу та точки виходу, тим самим скоротивши час, що витрачається на встановлення маршруту, залишивши встановлення лише двох правил. У динаміці контролер SDN має підтримувати стан цих маршрутів, де під станом розуміється залишкова пропускна спроможність маршруту. Контролер SDN може отримати стан маршруту, збираючи статистику про обсяг передається трафіку в одиницю часу (назвемо цю величину завантаженням лінії зв'язку) на кожній лінії зв'язку на маршруті і беручи мінімум з різниць пропускної спроможності та поточного завантаження. Тоді при появі нового багатопоточного з'єднання в мережі можна оцінити, чи вистачає ресурсів побудованих маршрутів для забезпечення потрібної якості сервісу для цього потоку класу чи ні. Якщо ресурсів маршрутів вистачає для задоволення вимог багатопоточного транспортного сполучення, що розглядається, то за цим транспортним з'єднанням резервуються необхідні ресурси.

Ресурси, що займаються багатопоточним транспортним з'єднанням, повинні звільнитися після завершення з'єднання. Понад те, ресурси, займані одним підпотіком, мають звільнитися під час завершення цього підпотіку, оскільки динамічний метод багатопоточної маршрутизації дозволяє закривати зайві підпотоки. Контролер SDN може дізнатися про завершення підпотіку за допомогою idle-timeout - таймера, який прив'язаний до певного правила в таблиці OpenFlow і веде свій відлік за відсутності пакетів, які відповідають цьому правилу. Якщо з'являється новий пакет мережі, який відповідає цьому правилу, то відлік починається спочатку. Після закінчення відведеного часу в idle-timeout комутатор SDN надішле

службове повідомлення на контролер SDN із зазначенням, за яким правилом спрацював idle-timeout. Використовуючи цю інформацію, контролер SDN зможе зрозуміти, який із підпотоків завершив свою роботу. Так як для кожного багатопотокового транспортного з'єднання відбувається резервація ресурсів і за умови відсутності інших потоків в SDN, можна використовувати спрощену реалізацію з наступною моделлю. Зберігатимемо поточний стан кожної лінії зв'язку. У разі появи нового з'єднання MPTCP визначаємо резервовану пропускну спроможність, так як всі підпотоків MPTCP відкриваються відразу. У разі з'єднання FDMP визначаємо резерв пропускну спроможності, послідовно розглядаючи кожен маршрут і забираючи всю залишкову пропускну спроможність маршруту, поки не наберемо необхідну для виконання вимоги потоку, що розглядається. Запропонований метод дозволяє не витратити ресурси контуру управління на моніторинг завантаження ліній зв'язку.

Висновки

При впровадженні SDN виникають виклики, такі як забезпечення масштабованості та високої доступності централізованих контролерів, захист від єдиної точки відмови та забезпечення безпеки комунікацій між контролерами та мережевими елементами. Централізоване управління програмно-конфігурованих мереж представляє собою потужний підхід до управління мережею, який пропонує значні переваги у гнучкості, автоматизації та ефективності.

Для проведення порівняльного аналізу обрані протоколи Multipath TCP та Flow Distribution Management Protocol як представники статичного та динамічного методів. Основна гіпотеза перевіряється на основі моделювання поведінки багатопоточних методів маршрутизації з розглянутих стратегій роботи контролера SDN. Наведено опис методу управління контролерами в програмно-конфігурованих мережах.

СПИСОК ЛІТЕРАТУРИ

1. Diachenko V. Інтелектуальні підходи енергозбереження у безпроводних сенсорних комп'ютерних мережах. Системи управління, навігації та зв'язку. Полтава: ПНТУ, 2020. Т. 4 (62). С. 114-118. doi: <https://doi.org/10.26906/SUNZ.2020.4.114>
2. Kovalenko, A., Kuchuk, H., Kuchuk, N. and Kostolny, J. (2021), "Horizontal scaling method for a hyperconverged network", *2021 International Conference on Information and Digital Technologies (IDT)*, Zilina, Slovakia, doi: <https://doi.org/10.1109/IDT52577.2021.9497534>
3. Petrovska, I., Kuchuk, H., Mozhaiev, M. (2022), Features of the distribution of computing resources in cloud systems, 2022 IEEE 4th KhPI Week on Advanced Technology, KhPI Week 2022 - Conference Proceedings, 03-07 October 2022, Code 183771, doi: <https://doi.org/10.1109/KhPIWeek57572.2022.9916459>
4. Kuchuk, H., Kovalenko, A., Ibrahim, B.F. and Ruban, I. (2019), "Adaptive compression method for video information", *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8(1), pp. 66–69, doi: <http://dx.doi.org/10.30534/ijatcse/2019/1181.22019>
5. Petrovska, I. and Kuchuk, H. (2023), "Adaptive resource allocation method for data processing and security in cloud environment", *Advanced Information Systems*, Vol. 7, No. 3, pp. 67–73, doi: <https://doi.org/10.20998/2522-9052.2023.3.10>
6. Semenov, S., Mozhaiev, O., Kuchuk, N., Mozhaiev, M., Tiulieniev, S., Gnusov, Yu., Yevstrat, D., Chyrva, Yu. & Kuchuk, H. (2022), "Devising a procedure for defining the general criteria of abnormal behavior of a computer system based on the improved criterion of uniformity of input data samples", *Eastern-European Journal of Enterprise Technologies*, 2022, 6 (4(120)), pp. 40–49, doi: <https://doi.org/10.15587/1729-4061.2022.269128>
7. Datsenko, S. and Kuchuk, H. (2023), "Biometric authentication utilizing convolutional neural networks", *Advanced Information Systems*, Vol. 7, no. 2, pp. 87–91, doi: <https://doi.org/10.20998/2522-9052.2023.2.12>
8. Pfaff B. et al. OpenFlow Switch Specification—1.3 Version //Open Networking Foundation: Menlo Park, CA, USA. – 2012. NoviFlow Inc. 2021. NoviSwitch 2122 Datasheet (2019). Retrieved "May 17, 2021" from https://noviflow.com/wp-content/uploads/2019/11/NoviSwitch-2122-Datasheet-400_V5.pdf.

Received (Надійшла) 15.05.2024

Accepted for publication (Прийнята до друку) 17.07.2024

Routing methods for ensure quality of network service

Vladyslav Diachenko, Evgeny Kolisnyk, Anastasia Lyashova, Oleksandr Mozhaev

Abstract. Relevance. The relevance of routing methods to ensure the quality of network service remains high in today's digital world. A significant increase in the volume of data transmission in networks is due to the growth of the number of connected devices and the development of the Internet of Things. This requires more efficient management of bandwidth and latency to ensure adequate quality of service. Today's networks are used for a wide range of applications, from web browsing and e-mail to video conferencing, online gaming, and high-definition video streaming. Each application has its own unique requirements for quality of service, which leads to the need for various routing methods to ensure optimal performance. It should also be noted that the increasing popularity of streaming multimedia services, such as Netflix, YouTube, and Spotify, requires high bandwidth. and low latency to ensure a satisfactory level of service. Given these factors, routing methods to ensure the quality of network service remain a key element in the design and management of network infrastructure, and their development and optimization are an integral part of ensuring high performance and reliability of network services. **The purpose** is a study of existing routing methods to ensure the quality of network service. **The object** of the study is the evolution of FPGA accelerators for data center (DC) and cloud computing. **The subject** are methods of multithreaded routing of computer networks. **Results.** An analysis of the existing methods of multithreaded routing was carried out. The existing works on the analysis of the effectiveness of multithreaded routing are analyzed. Multipath TCP and Flow Distribution Management Protocol protocols were selected for comparative analysis as representatives of static and dynamic methods. The main research methodology is a combination of simulation modeling and statistical tests. The main hypothesis is that dynamic methods will outperform static methods in multithreaded routing. This hypothesis is tested on the basis of modeling the behavior of multithreaded routing methods from the considered SDN controller operation strategies. A description of the method of controlling controllers in software-configured networks for the implementation of built routes of a multi-flow connection is given.

Keywords: multiflow routing, method, software-configurable network, QoS, network service, method, multimedia service, Multipath TCP, Flow Distribution Management Protocol.

І. С. Зубко, В. О. Мартовицький, А. В. Пунченко, Д. Д. Карачевцев

Харківський національний університет радіоелектроніки, Харків, Україна

ОГЛЯД МЕТОДІВ НАНЕСЕННЯ ЦИФРОВИХ ВОДЯНИХ ЗНАКІВ ДЛЯ ЗАХИСТУ ЗОБРАЖЕНЬ

Анотація. Стаття присвячена огляду методів нанесення цифрових водяних знаків з метою захисту зображень від несанкціонованого використання та копіювання. Об'єктом дослідження є самі методи нанесення цифрових водяних знаків, їхні технічні особливості, ефективність у захисті зображень та вартість впровадження. Автори досліджують різні підходи до вбудовування водяних знаків у зображення та проводять їх порівняльний аналіз з точки зору ефективності, стійкості до атак та вартості впровадження. Стаття детально розглядає різні підходи до вбудовування цифрових водяних знаків у зображення, включаючи методи на основі текстури, частотного домену, та алгоритмічні підходи. Кожен метод аналізується з точки зору його ефективності, стійкості до різних видів атак, таких як ретушування або обрізання, а також вартості впровадження та обслуговування. Наслідком дослідження є комплексне порівняльне обґрунтування методів нанесення водяних знаків, яке допомагає вибрати оптимальний підхід для конкретних потреб захисту зображень у цифровому середовищі. Висновки статті можуть бути корисними для фахівців у сфері цифрових медіа, авторів контенту та розробників програмного забезпечення, які цікавляться збереженням інтелектуальної власності та захистом авторських прав в онлайн-середовищі. Результати огляду допомагають визначити найбільш оптимальні методи для конкретних потреб захисту зображень у цифровому середовищі.

Ключові слова: стеганографія, цифровий водяний знак, захист інформації, зображення.

Вступ

Метод вбудовування цифрового водяного знаку в зображення дає змогу розв'язати проблему права власності, тому дослідження в області стеганографії має ширший спектр застосування, наприклад, захист авторських прав, автентифікація контенту, ідентифікація власника тощо. Яремчук та інші [1], Рубан та інші [2], A Ray та інші [3] представили огляд сучасних методів вбудовування цифрових водяних знаків в зображення, які застосовують у різних галузях. У цій статті представлено мотивацію подальших досліджень, області застосування, вимоги та питання проектування, класифікацію атак і методів вбудовування цифрового водяного знаку. Крім того, визначено деякі проблеми в галузі цифрових водяних знаків для зображень.

Постановка проблеми та її зв'язок із важливими науковими чи практичними завданнями

У наш час спостерігається зростання кількості різноманітних електронних пристроїв, призначених для зберігання та обробки мультимедійних даних, такі як:

- смартфони;
- планшетні комп'ютери;
- вбудовані мультимедійні програвачі;
- фотоапарати з підтримкою Wi-Fi або Bluetooth;
- портативні жорсткі диски та флеш-накопичувачі;
- електронні книги з можливістю відтворення аудіо та відеофайлів;
- портативні геймінгові консолі;
- умовно-портативні компактні комп'ютери, такі як ноутбуки та нетбуки.

Разом з появою портативних пристроїв також з'являються ефективні методи стиснення мультиме-

дійних даних. Це в сукупності з високошвидкісним інтернет-з'єднанням, сприяло поширенню різноманітних програм та сервісів, що базуються на використанні цифрового контенту. Незважаючи на те, що цифрові дані мають багато переваг порівняно з аналоговою версією, їх справжність або право власності на них є найбільшою проблемою. Цифрові мультимедійні дані можна легко дублювати та/або маніпулювати ними, що створює реальну загрозу неправомірного використання мультимедійних даних для власника контенту. Щоб вирішити проблему неправомірного використання мультимедійних даних в мережі Інтернет треба забезпечити надійність і оригінальність переданих мультимедійних даних. У світлі цих негативних факторів, в сучасній ері цифрових технологій стає важливим забезпечити захист мультимедійних даних від незаконного використання.

Сьогодні власники мультимедійного контенту шукають технології, здатні захистити їхні права та убезпечити контент від піратства, несанкціонованого використання, а також ті технології, що дають змогу відстежувати й засуджувати медіапиратів. За останні десятиліття дослідники запропонували різні рішення для захисту мультимедійних даних від несанкціонованого використання. Одним із рішень цієї проблеми є вбудовування невидимих ідентифікаторів у вихідні мультимедійні дані для доказу їхньої приналежності. Цей тип методів називається приховуванням інформації, яке можна розділити на різні підкласи, такі як криптографія, стеганографія та водяні знаки [4–6]. Криптографія – найпоширеніший метод захисту цифрових мультимедійних даних, де мультимедійний контент шифрують перед виданням, а ключ для розшифрування надають тим, хто придбав справжні або легальні копії [7, 8]. Однак криптографія не може допомогти контент-провайдерам контролювати вміст після процесу розшифрування; зловмисник може легко викрасти

справжню або легальну копію, а потім перепродати її або розповсюдити безплатно в загальнодоступній мережі. Стеганографія - це запобігання виявленню зашифрованих даних, які були захищені криптографічними алгоритмами. Однак повідомлення, приховане за допомогою стеганографії, не є надійним. До водяних знаків порівняно з алгоритмами стеганографії висуваються додаткові вимоги щодо стійкості до різних атак, пов'язаних з обробкою сигналу та геометричними перетвореннями. Тому вкрай важливо знайти спосіб захисту цифрового мультимедійного контенту за допомогою більш точного методу, який дав би змогу власникам контенту бути впевненими в розміщенні та поширенні своїх матеріалів в Інтернеті. Таким засобом може стати водяний знак.

Тому виникає завдання з дослідження застосування різних методів водяних знаків.

Результати досліджень

Сфери застосування водяних знаків. Методи вбудовування цифрових водяних знаків мають багато застосувань, а саме:

- захист авторських прав: одним з причин розробки методів нанесення водяних знаків є захист авторських прав. У цьому випадку дані/інформація про авторське право вбудовуються в основний об'єкт без втрати якості [9]. Вбудовані дані перешкоджають іншим сторонам претендувати на право власності на ці дані. Крім того, водяний знак повинен бути відомий лише автору і повинен бути стійким до різних атак;

- прихована комунікація: методи нанесення водяних знаків також можуть використовуватися для прихованої передачі інформації, оскільки різні відомства або уряди встановлюють обмеження на використання шифрування. У цьому випадку люди можуть надсилати свої секретні повідомлення, використовуючи методи вбудови цифрових водяних знаків [10];

- контроль копіювання: ця функція обмежує незаконне копіювання матеріалів, захищених авторським правом, шляхом вбудови цифрового водяного знаку, який не можна копіювати, або обмеження кількості разів копіювання [11]. Наприклад, сьогодні в Інтернеті доступно багато документів, які не можна зберегти та роздрукувати, щоб контролювати незаконне копіювання;

- автентифікація вмісту: крихкий водяний знак може бути вбудований у зображення хоста для перевірки автентичності даних. Крихкий водяний знак вказує на те, чи були дані змінені, а також надає інформацію про те, де ці дані були змінені [12]. Тому ця задача не вимагає надійного водяного знаку, оскільки нам потрібно лише виявити зміни;

- зчитування цифрового відбитка: метод зчитування цифрового відбитка, застосовується власником для того, щоб відстежити джерело нелегальних копій. Для цього власник може вбудовувати різні водяні знаки в кожен копію, яка розповсюджується серед різних клієнтів [13]. Наприклад, унікальні серійні номери присвоюються покупцям і використовуються для ідентифікації покупця;

- моніторинг трансляції: власники захищених авторським правом телепрограм повинні знати про нелегальну трансляцію або рекламу, що транслюється телеканалами в певний час і в певному місці відповідно до умов контракту. Водяні знаки можуть бути вбудовані в будь-який тип даних для трансляції в мережі автоматизованими системами, які здатні контролювати канали розповсюдження, щоб відстежувати контент у потрібний час і в потрібному місці [14];

- медична безпека: останнім часом телемедицина полегшує медичну діагностику, надсилаючи медичні дані/звіт пацієнта через загальнодоступну мережу для подальшого аналізу там, де доступне сучасне медичне обладнання. Це обладнання виробляє велику кількість даних щодня. Отже, необхідно захистити ці важливі дані. Нанесення водяних знаків на медичні зображення є підходящим методом для підвищення безпеки та автентифікації медичних даних, які використовуються для подальшої діагностики та довідок [15]. Вбудовані дата та ім'я пацієнта в медичні зображення можуть бути корисними заходами безпеки;

- індексування: Одним з відомих застосувань цифрових водяних знаків є індексування мультимедійного контенту, такого як фільми, новини, відеопшта, зображення тощо [16]. При цьому коментарі або будь-який тег/жанр вбудовуються у вміст таким чином, щоб ці коментарі або теги використовувалися будь-якою пошуковою системою для пошуку цього вмісту в Інтернеті.

Вимоги та особливості проектування методів вбудови цифрових водяних знаків. Існують різні аспекти проектування та вимоги, пов'язані з будь-яким методом нанесення водяних знаків, такі як прозорість, надійність, пропускну здатність, безпека тощо. Завданням дослідників у галузі нанесення водяних знаків є максимізація всіх цих параметрів для конкретного методу.

Крім того, ці параметри взаємозалежні один від одного, як показано на рис. 1.

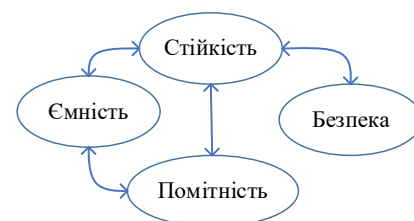


Рис. 1. Взаємна залежність між параметрами проектування

Три параметри, а саме: прозорість, стійкість та сміність обернено пропорційно пов'язані між собою, тобто, якщо прозорість методу водяних знаків зростає, то його стійкість погіршується, і навпаки. Цей взаємозв'язок зображено на рис. 2.

Отже, відносна важливість цих параметрів залежить від конкретного застосування, як зазначено в попередньому розділі. Крім того, деякі програми вимагають більшої надійності у порівнянні з непрозорістю.

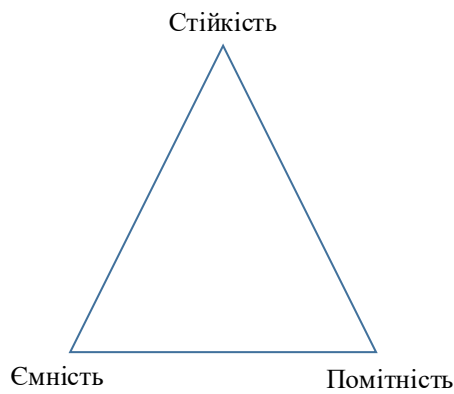


Рис. 2. Три основні суперечливі аспекти використання водяних знаків

Таким чином, процес розробки методу нанесення водяних знаків включає в себе компроміс між суперечливими параметрами.

Систематизація атак на цифрові водяні знаки. Будь-яку процедуру, яка може знизити результативність методу нанесення водяних знаків, можна назвати атакою. Тестування стійкості та безпеки методу нанесення водяних знаків до атак є настільки ж важливим, як і процес розробки. Атаки не завжди видаляють або знищують водяний знак, але також унеможливають його виявлення.

Спотворення, спричинені будь-якими атаками, погіршують функціонування методу нанесення водяних знаків.

Загалом, атаки на цифрові водяні знаки можна розділити на два класи, а саме: ненавмисні та навмисні атаки. Щоб досягти високої надійності виявлення цифрових водяних знаків, процес виявлення повинен бути стійким до змін у даних, спричинених як ненавмисними, так і навмисними атаками. Ненавмисні атаки відбуваються за допомогою операцій обробки сигналів над даними з цифровими водяними знаками, а саме: стиснення, друк, сканування, фільтрація, зашумлення, геометричні перетворення, обрізання тощо. Наприклад, мультимедійні дані зазвичай зберігаються у стислому форматі з втратами для того, щоб використовувати менше пам'яті. Ці алгоритми стиснення відкидають неважливі частини даних. Таке спотворення може призвести до пошкодження даних із вставленими водяними знаками. Це означає, що проста атака полягає в стисненні мультимедійних даних з втратами. Крім того, обертання або масштабування може змінити значення пікселів і знищити дані водяного знаку. Операції обробки сигналу, такі як квантування, декомпресія, повторна вибірка і зменшення кольору, можуть зіпсувати водяний знак. У випадку навмисних атак, людина може цілеспрямовано атакувати вставлені дані цифрового водяного знаку, щоб скопіювати мультимедійні дані. В обох випадках будь-який метод нанесення цифрових водяних знаків повинен бути здатним виявити і витягти водяний знак після атаки. Таксономія різних навмисних і ненавмисних атак на методи нанесення водяних знаків наведена в табл. 1.

Таблиця 1 – Систематизація атак на водяні знаки

Атака	Короткий опис
Шум	Будь-який випадковий небажаний сигнал із заданим розподілом, а саме: гауссіан, «сіль і перець», Пуассона.
Фільтрація	Атаки типу фільтрації - це лінійна фільтрація, а саме: фільтрація нижніх/середніх частот, гауссова фільтрація, фільтрація з підвищенням різкості тощо.
Стиснення	Якщо цифровий водяний знак повинен протистояти різним рівням стиснення, зазвичай рекомендується виконувати вбудовування водяного знаку в той самий домен, де відбувається стиснення.
Множинні водяні знаки	Одним з рішень такого типу проблем є включення інформації про час нанесення ЦВЗ сертифікаційним центром
Геометричні атаки	Геометричні атаки спотворюють цифровий водяний знак шляхом просторових змін зображення. Найпоширенішими геометричними атаками є обертання, масштабування тощо.
Обрізання	Це дуже поширена атака, яка обрізає потрібну область від зображення з цифровим водяним знаком.
Атаки на видалення водяних знаків і перешкоди	Мета таких атак - визначити або підмінити водяний знак.
Статистичне узагальнення	Метою таких атак є відновлення основного зображення та/або даних водяного знаку шляхом статистичного дослідження декількох наборів даних з водяними знаками.

Зазначимо, що з розвитком технологій захисту й аналізу цифрових даних, з'являються нові методи захисту від цих атак, а також нові способи їх виявлення та протидії.

Огляд підходів нанесення водяних знаків на зображення. Цифрове зображення може бути представлено/зберігатися або в просторово-часовій області, або в області перетворень. Зображення в просторово-часовій області характеризується пікселями,

тоді як зображення в області перетворення описується в термінах його коефіцієнтів перетворення. Іншими словами, представлення зображення в області перетворення розділяє коефіцієнти перетворення на декілька частотних діапазонів. Для перетворення зображення в область перетворення можна використовувати різні доступні методи зворотнього перетворення, а саме: дискретне перетворення Фур'є (DFT), дискретне косинусне перетворення (DCT),

дискретне вейвлет-перетворення (DWT), кероване пірамідалне перетворення (SPT) та інші. Кожен з цих методів перетворення має свої специфічні характеристики та представлення зображення.

Нанесення водяних знаків на цифрові зображення – це процес непомітного вбудовування водяного знаку (у вигляді підпису, випадкової послідовності або якогось зображення) в зображення (носії або обкладинку), який може бути використаний для перевірки автентичності його власника. Отримане в результаті цього процесу зображення називається зображенням з водяним знаком. Методи нанесення водяних знаків можуть виконуватися як у просторовій області, так і в області перетворень. У просторовому методі водяні знаки можуть бути вбудовані в зображення шляхом зміни значень пікселів або значень найменш значущих бітів (LSB). У той час як у методі на основі домену перетворення водяний знак може бути вбудований шляхом модифікації коефіцієнтів домену перетворення. Однак, більш стійкий водяний знак може бути вбудований в область перетворення зображень шляхом модифікації коефіцієнтів області перетворення порівняно з методом водяного маркування зображень на основі просторової області

Метод нанесення водяних знаків на основі просторового доменного підходу, приховує дані водяного знаку дані водяного знаку в значеннях пікселів основного зображення. Цей клас методів вносить незначні незначні зміни в інтенсивності пікселів основного зображення.

Одним з найпоширеніших прикладів такого методу є вбудовування водяного знаку в LSB пікселів зображення. Іншими словами, значна частина низькочастотних компонентів зображення повинна бути модифікована для того, щоб вставити дані водяного знаку надійним і стійким способом. Інший приклад: зображення розбивається на однакові за розміром блоки, і до підблоків додаються певні дані водяного знаку. Непомітність даних водяного знаку досягається на основі постулату, що біти LSB є візуально незначущими. Хоча метод просторових доменних водяних знаків може бути легко реалізований і дуже швидкий, він має багато недоліків. Ці методи дуже чутливі до звичайних операцій обробки сигналів і можуть бути легко порушені та послаблені. Наприклад, стиснення з втратами може повністю знищити дані водяного знаку. Таким чином, просторовий метод нанесення водяних знаків дуже легко зруйнувати за допомогою деяких атак, таких як низькочастотна фільтрація, адитивний шум тощо. Іншими словами, методи просторового доменного водяного маркування зображень не є стійкими до звичайних операцій обробки сигналу на основному зображенні.

Області перетворення зображення – це просто інша форма представлення. Воно не змінює вміст, присутній у зображенні. Методи водяного маркування зображень на основі трансформованих доменів мають багато переваг над методами на основі просторових доменів. Як зазначено в літературі, методи водяного маркування зображень на основі трансформованих доменів є більш стійкими до різ-

них атак на водяні знаки та операцій обробки сигналів, оскільки домен перетворення не використовує вихідне зображення для нанесення даних водяного знаку.

Крім того, водяні знаки на основі домену перетворення розподіляють дані водяного знаку по всій частині основного зображення. Крім того, методи на основі перетворення доменів здатні вбудовувати більше бітів водяного знаку в основне зображення і є більш стійкими до атак.

Проблеми у сфері нанесення водяних знаків на зображення. Незважаючи на те, що було запропоновано багато різноманітних методів вбудови цифрових водяних знаків на зображення, але все ще існують певні проблеми, які потребують вирішення. Однією з головних проблем застосування водяних знаків є досягнення кращого компромісу між надійністю, прозорістю, пропускну здатністю та безпекою. Для того, щоб вирішити вищезгадану проблему (тобто знайти компроміс) для досягнення кращої продуктивності, багато дослідників представили її рішення для цієї проблеми в своїх роботах. Однак для того, щоб виправдати очікування індустрії, потрібні вдосконалення. Розглянемо деякі з найважливіших проблем, пов'язаних з нанесенням водяних знаків на зображення.

Більшість робіт у цій галузі за останнє десятиліття було присвячено захисту кольорових зображень або зображень у відтінках сірого (хост-зображень) шляхом вбудовування водяних знаків у відтінках сірого або бінарних зображень. Для вбудовування бінарного зображення або зображення у відтінках сірого потрібно перетворити його з кольорового, оскільки в мультимедіа використовується кольорові зображення. Однак, існує дуже мало методів, які вбудовують кольорові водяні знаки для захисту зображень [17-19]. З цієї точки зору, існує ще багато можливостей для вдосконалення в галузі водяних знаків зображень для вбудовування кольорового водяного знаку в кольорове основне зображення.

Висновки

За результатами огляду методів нанесення цифрових водяних знаків для захисту зображень можна зробити наступні висновки:

- дослідження показує, що існує широкий спектр методів нанесення водяних знаків, від класичних до складних алгоритмічних підходів;
- кожен метод має свої властивості та обмеження, що варто враховувати при виборі для конкретної задачі;
- важливо оцінювати стійкість кожного методу до потенційних атак, таких як ретушування, обрізання або розмиття зображень;
- вибір методу повинен базуватися на потребах та вимогах користувачів, включаючи зручність використання та якість захисту.

Загалом, огляд методів нанесення цифрових водяних знаків дає можливість краще зрозуміти доступні варіанти та вибрати оптимальний метод для конкретної ситуації або завдання.

СПИСОК ЛІТЕРАТУРИ

1. Ю. Є. Яремчук, В. В. Карпинець, І. С. Зоря, і Д. О. Козак, «ПІДВИЩЕННЯ СТІЙКОСТІ ЦИФРОВИХ ВОДЯНИХ ЗНАКІВ У ПОТОКОВИХ ВІДЕОЗАПИСАХ НА ОСНОВІ ДИФЕРЕНЦІАЛЬНОГО ВБУДОВУВАННЯ ЕНЕРГІЇ (DEW)», Вісник ВПІ, вип. 1, с. 55–64, Лют. 2023.
2. Рубан І. В. Інформаційна технологія підтвердження права власності на цифрові зображення / І. В. Рубан, Н. М. Бологова, В. О. Мартовийський // Сучасні інформаційні системи = Advanced Information Systems. – 2022. – Т. 6, № 1. – С. 118-123.
3. Ray, A., Roy, S. Recent trends in image watermarking techniques for copyright protection: a survey. *Int J Multimed Info Retr* 9, 249–270 (2020). <https://doi.org/10.1007/s13735-020-00197-9>
4. Guerriero M., Michele G. Adoption, support, and challenges of infrastructure-as-code: Insights from industry. In: 2019 IEEE international conference on software maintenance and evolution (ICSME). IEEE, 2019, p. 580-589.
5. Semenov, S., Mozhaiev, O., Kuchuk, N., Mozhaiev, M., Tiulieniev, S., Gnusov, Yu., Yevstrat, D., Chyrva, Yu. & Kuchuk, H. (2022), "Devising a procedure for defining the general criteria of abnormal behavior of a computer system based on the improved criterion of uniformity of input data samples", *Eastern-European Journal of Enterprise Technologies*, 2022, 6 (4(120)), pp. 40–49, doi: <https://doi.org/10.15587/1729-4061.2022.269128>
6. Griffin, J., Noussia, K., Nedeva, S., Zervoudakis, S., Lux, J., & McNamara, J. (2023). Artificial Intelligence (AI) and Watermarking to Transform Copyright Arbitration and Dispute Resolution for Three-Dimensional (3D) printing: An Empirical Analysis. *European Journal of Law and Technology*.
7. Бодня, М., ЄсінаМ., & Пономар, В. (2024). Дослідження можливостей застосування стеганографічних та криптографічних алгоритмів для приховування інформації. *Комп'ютерні науки та кібербезпека*, (2), 43-57
8. Л. Тарасенко. АВТОРСЬКЕ ПРАВО У ЦИФРОВУ ЕПОХУ: ОСНОВНІ ТЕНДЕНЦІЇ І ЗМІНИ. Вісник Львівського університету. Серія юридична. 2022. Випуск 75. С. 61–72.
9. Martovytskyi, Vitalii and Ruban, Igor and Bolohova, Nataliia and Sievierinov, Oleksandr and Zhurylo, Oleg and Permiakov, Oleksandr and Nosyk, Andrii and Nepokrytov, Dmytro and Krylenko, Ivan, Development of Methods for Generation of Digital Watermarks Resistant to Distortion (December 29, 2021). *Eastern-European Journal of Enterprise Technologies*, 6 (2 (114)), 103–116. doi: <https://doi.org/10.15587/1729-4061.2021.246641>
10. Рубан І. В. Модель обработки TCP-соединений для стеганографической передачи данных в информационно-телекоммуникационных сетях / И. В. Рубан, А. А. Смирнов // Сучасні інформаційні технології у сфері безпеки та оборони. - 2015. - № 3. - С. 108-112
11. Tarhouni, N., Charfeddine, M. & Ben Amar, C. Novel and Robust Image Watermarking for Copyright Protection and Integrity Control. *Circuits Syst Signal Process* 39, 5059–5103 (2020). <https://doi.org/10.1007/s00034-020-01401-1>
12. Kozina, G. L., Savchenko, I., Voskoboynik, V., & Karpukov, L. (2023). Steganographic photo protection system using fragile watermarks. *Systems and Technologies*, 64(2), 75-81. <https://doi.org/10.32782/2521-6643-2022.2-64.10>
13. F. Regazzoni, P. Palmieri, F. Smailbegovic, R. Cammarota and I. Polian, "Protecting artificial intelligence IPs: A survey of watermarking and fingerprinting for machine learning," *CAAI Trans. on Intellig. Techn.*, vol. 6, no. 2, pp. 180–191, 2021.
14. Megias, D.; Mazurczyk, W.; Kuribayashi, M. Data Hiding and Its Applications: Digital Watermarking and Steganography. *Appl. Sci.* 2021, 11, 10928. <https://doi.org/10.3390/app112210928>
15. Anand, A., Singh, A.K. Watermarking techniques for medical data authentication: a survey. *Multimed Tools Appl* 80, 30165–30197 (2021). <https://doi.org/10.1007/s11042-020-08801-0>
16. Fernandez, Pierre, et al. "Active image indexing." *arXiv preprint arXiv:2210.10620* (2022). URL: <https://arxiv.org/abs/2210.10620>
17. Wang H, Yuan Z, Chen S, Su Q. Embedding color watermark image to color host image based on 2D-DCT. *Optik (stuttg)* 2023;274:170585. <https://doi.org/10.1016/j.ijleo.2023.170585>.
18. Su, Q., Zhang, X. & Wang, G. An improved watermarking algorithm for color image using Schur decomposition. *Soft Comput* 24, 445–460 (2020). <https://doi.org/10.1007/s00500-019-03924-5>
19. Hu, F., Cao, H., Chen, S. et al. A robust and secure blind color image watermarking scheme based on contourlet transform and Schur decomposition. *Vis Comput* 39, 4573–4592 (2023). <https://doi.org/10.1007/s00371-022-02610-2>

Received (Надійшла) 08.04.2024

Accepted for publication (Прийнята до друку) 12.06.2024

Features of automatic deployment of infrastructure as code for cloud services review of digital watermarking methods for image protection

I. Zubko, V. Martovytskyi, A. Puchenko, D. Karachevtsev

Abstract. The article is devoted to the review of digital watermarking methods for protecting images from unauthorised use and copying. The object of the study is the methods of digital watermarking, their technical features, effectiveness in protecting images and the cost of implementation. The authors investigate different approaches to embedding watermarks in images and conduct a comparative analysis of them in terms of efficiency, attack resistance and implementation costs. The paper takes a closer look at various approaches to embedding digital watermarks in images, including texture-based, frequency domain, and algorithmic approaches. Each method is analysed in terms of its effectiveness, resistance to various types of attacks, such as retouching or cropping, as well as implementation and maintenance costs. The result is a comprehensive comparative study of watermarking methods, which helps to choose the best approach for the specific needs of image protection in the digital environment. The conclusions of the article may be useful for digital media professionals, content creators and software developers interested in preserving intellectual property and copyright protection in the online environment. The results of the review help to determine the most optimal methods for the specific needs of image protection in the digital environment.

Keywords: steganography, digital watermark, information security, images.

A. Kapiton¹, O. Dziuban¹, R. Baranenko², H. Sokol³

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

² Uman National University of Horticulture, Uman, Ukraine

³ National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine

SECURITY OF INFORMATION TECHNOLOGIES IN A HYBRID WORKING ENVIRONMENT

Abstract. The article discusses the main aspects of information technology security in a hybrid working environment. With the rapid rise in popularity of remote work, traditional workplace security standards have been proven to be inadequate, creating a real need for effective cyber security measures. Remote work has become the norm for many companies, but with it the threat of cyberattacks has also increased. Features and risks associated with the organization of a hybrid work environment caused by remote employment are considered, as well as methods of ensuring safety in such a work environment are investigated. The main problems that arise and need to be prevented have been identified. It has been proven that one of the key areas is to securely connect to corporate networks from remote locations. It has been established that traditional VPN methods can be ineffective or insufficiently secure in conditions of large-scale use, so it is necessary to use modern means of traffic encryption and authentication mechanisms, which aim to organize the protection of access to corporate resources. The article examines the content of the constituent components that are responsible for the security of the endpoint. Requirements are defined that allow access to update and modify the software and operating systems of remote devices, which is a special part of the proposed strategy that meets data security for operation in a hybrid environment. The determining motives of the implementation of the main structural component mechanisms of data access control and the possibility of remote shutdown of devices in the event of emotional actions have been studied. Data protection plays a crucial role in remote work. The indisputability of the need to design and implement a system for protecting all remote resources, means of recognizing and identifying suspicious messages and links, as well as the use of reliable passwords and two-factor authentication has been proven and substantiated.

Keywords: information security, hybrid working environment, cyber security, digital transformation.

Introduction

The current complex situation in the political and economic space has opened a new era of remote work, forcing organizations to rethink their cyber security strategies.

The result of the conducted research is the analysis and substantiation of ways to ensure the security of information technologies in a hybrid working environment and the study of methods of effective protection of a remote working environment. The goals of the study are to deepen the understanding of the conditions of remote work, to identify cyber security problems in this type of work, as well as to develop recommendations for ensuring the security of the remote work environment. The most important aspect is understanding cyber threats and the means to prevent them.

Attackers are constantly looking for vulnerabilities in systems and networks, especially in remote work environments. The study of methods of protection against phishing attacks, the use of unprotected Wi-Fi networks, as well as protection against the leakage of confidential information through remote access is an actual problem today.

As a result of constant modernization and rapid growth of work methods to overcome cyber threats, there is a need to improve information and communication, software and technical means. Note that today's security professionals focus on identifying and fixing software flaws, unlike users who have the ability not only to increase potential problems due to insufficient competence, but to prevent them.

Analysis of recent research and publications

Analysis of the works of leading scientists who expressed their opinions in materials published in magazines, collections of reports of conferences held in our country and abroad, gives confidence in the relevance of the researched issues. Buryachok V.L., Kyrychok R.V., the importance of information security in the conditions of remote work is investigated, various aspects of remote work are considered, in particular, security challenges and risks.

The researcher provides analytical information on the protection of remote connections, data protection and support of information security in a remote working environment [1]. In the study of Buryachok V. L., Kyrychok R. V., the authors emphasize the concept of information security as a service in the context of remote work.

The authors emphasize the importance of service level agreements, continuous monitoring and incident response in a remote work environment [2]. Kavun S. V., Nosov V. V., Manzhai O. V. gives an idea of the technical aspects of remote access and control systems. Understanding these systems is essential to implementing effective security measures in a remote work environment [3].

Vazhnytskyi B. and Tkachev V. investigate the security aspects of multi-cloud environments and, after analyzing the causes of these problems, form their own criteria for a secure cloud based on them.

In their works, the main problems and reliability criteria of multi-cloud environments are outlined for their further analysis [4].

Al-Ammouri A., M. Dekhtyar, R. Ishchenko, E. Klochan, Lebid I., Dekhtiar M., Lyaskovskii V., Popova L., Tymchenko O., Poleva N., Podlevskiy B., Rykalyuk P., Zhurakovskiy Yu., Poltorak V. Tulyakova N., Kulyk A., Kryvogubchenko S. research methods and means of information protection [5–10]. Serdyukov D., Severinov O., Sydorenko Z. consider the capabilities and deployment process of the ESET Mobile Device Connector (MDC) application in order to ensure the security and management of mobile devices in corporate environments [11].

Main part

Modern political and economic influences on the organization of employees' activities had consequences in the field of labor and contributed to the introduction of new methods in business. The rapidly growing demand for remote work forces organizations to fully or partially transfer employees to remote work with a remote access connection. Information security is of great importance for ensuring the vital interests of any state. The creation of a developed and protected environment is an indispensable condition for the development of society and the state, which must be based on the latest automated technical means. In general, the object of protection in the information system is information with limited access, which circulates and is stored in the form of data, commands, messages that have a certain limitation and value both for its owner and for a potential violator of technical information protection.

A violator is a user who has unauthorized access to information. The threat of unauthorized access is an event that qualifies as the fact of an attempt by the offender to commit unauthorized actions in relation to any part of the information in the information system. Let's consider possible channels of information leakage and options for unauthorized access to it: in the absence of a legitimate user, control and delimitation of access to the terminal, a skilled offender easily uses its functionality for unauthorized access to protected information by entering appropriate requests or commands; if there is free access to the premises, it is possible to visually observe information on the means of reflection and documentation, steal paper media, remove an extra copy, as well as steal other media with information: listings, magnetic media, etc.

A special threat is the uncontrolled downloading of software in which settings, properties, data, algorithms may be changed, a "Trojan" program may be introduced, or a computer virus may be rooted that performs destructive unauthorized actions. For example, recording information on a third-party medium, illegal transfer to communication channels, unauthorized printing of documents, violation of their integrity, unauthorized copying of important information, the importance of which is determined and limited to a very short or, on the contrary, a long time.

A dangerous situation is when the violator is an authorized user of the information system who, in connection with his functional duties, has access to one

part of the information, and uses another part outside of his authority. There are many ways for an authorized user to break into an information system and obtain, modify, distribute, or destroy protected information. For this, you can use, first of all, privileged input-output commands, uncontrollability of authorization or legality of requests and requests to databases and data banks, servers, etc. During maintenance of the equipment, remnants of information on its media (hard disk surfaces, magnetic tapes and other media) may be detected. Erasing information using conventional methods (operating system tools, special software utilities) is ineffective from the point of view of technical information protection. The violator can renew and read its remains, which is why only special means of erasing the information to be protected are needed.

One of the important parts of comprehensive protection is access at the level of the operating system. Therefore, I consider it necessary to create user accounts and pass identification, authentication and authorization of users working in this operating system. Since the enterprise does not use a system of accounts and identification and authentication, I consider it necessary to implement it for more reliable data protection. To solve this problem, it is recommended to use the access matrix for granting rights and authorities for working with official information. Access delimitation is a set of procedures that implement the verification of access requests and assessment of the possibility of providing access based on the Access Delimitation Rules. Access restriction rules are a part of the security policy that regulates the access rules of users and processes to passive objects. When considering the interaction of two objects of a computer system acting as receivers or sources of information, one should distinguish between a passive object, which is operated on, and an active object, which performs or initiates this operation.

Analyzing data from a series of surveys by researchers, the group (APWG) concluded that the volume of phishing websites has increased by 47% between 2020 and 2023 (Fig. 1). A certain number of specialists, based on their duties, due to the possibility of working in a hybrid format, in most cases the targets of these attacks. It should definitely be noted about the advantages of hybrid work, especially remote, it creates difficulties in monitoring employee behavior. A report by A Tessian report found that forty-eight percent of employees state the fact more comfortable accepting security risks when working remotely [2, 12-14]. The growth of phishing attacks for 2021–2023 is presented in Fig. 1.

Using unsecured Wi-Fi networks carries significant risk, potentially exposing sensitive information to cybercriminals. In addition, the lack of physical security controls and increased reliance on personal devices create exploitable vulnerabilities.

The analysis of threats caused by cyberattacks in recent years allowed us to conclude that the largest number of cases related to the work of specialists at home, using a home network.

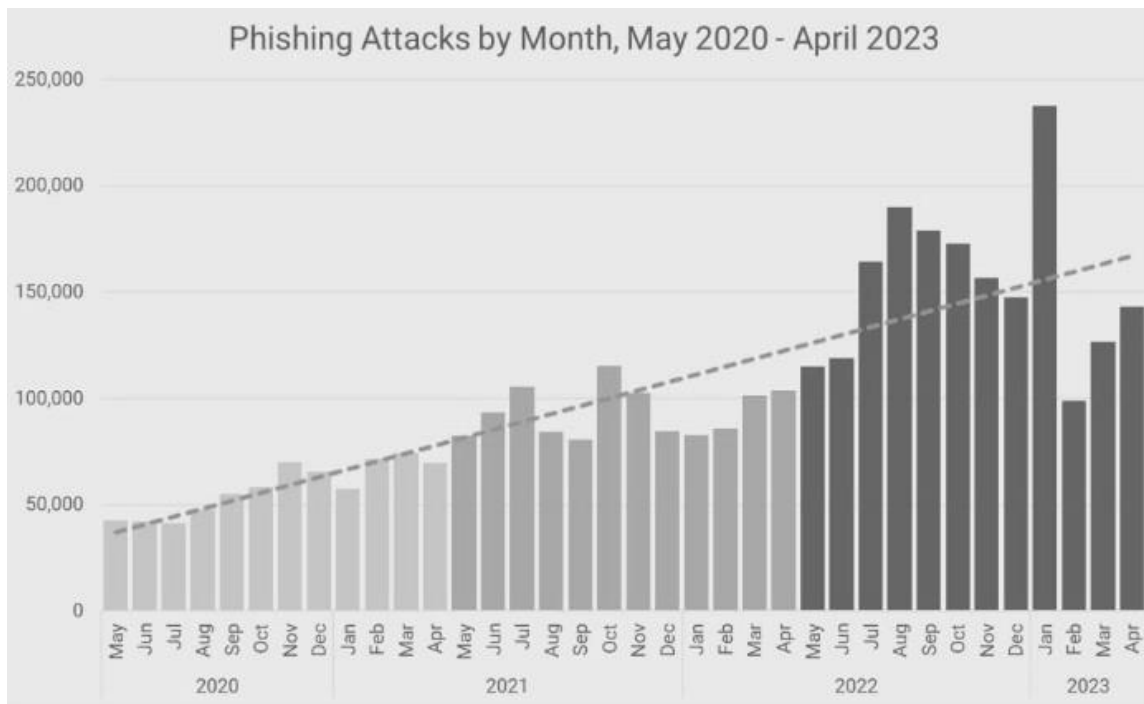


Fig. 1 Growth of phishing attacks for 2021-2023

That is why it can be argued that the security of the results of specialists' work in the conditions of hybrid work is insufficient, and measures to improve the cyber security strategy need to be strengthened. It is especially important to solve the task of implementing all possible protection methods, including the architecture of zero trust.

The minimum requirement of this concept is that users need to confirm the entered data at the time of performing a sequence of certain actions in the process of accessing resources [3-6].

Achieving work-life balance is essential to any flexible working environment. If there is a hybrid approach, it can enable employees to perform better. Additionally, home workers can schedule their days to take time off for important tasks, allowing them to be more flexible in their schedules and able to live and work wherever they want. Blended working helps people to better master their daily tasks when they need to take care of family (life or children), avoid long commutes, etc., which can increase concentration at work [12-15]. Ukrainian and global business leaders hold oppositely different opinions about what the working environment should look like in the next three years. 65% of global executives expect the office to become their primary work environment in three years. In contrast, Ukrainian company leaders continue to envision work in a hybrid format (88%), which is three times higher than the expectations of world leaders (28%).

Only 5% of Ukrainian managers talk about traditional office work. At the same time, managers both in Ukraine (5%) and in the world (7%) do not plan to transform the working environment into a format of completely remote cooperation.

The best hybrid collaboration tools for remote work presented in Table 1 [12-15].

Table 1 – The best hybrid collaboration tools for remote work

Confluence	Wrike	Trello
HighQ	InstaVC	GoVisually
Surfly	Asana	FlipBuilder
Adobe Workfront	Monday.com	FlipHTML5
Smartsheet	AnswerHub	Basecamp
Ziflow	Kissflow	Lucidspark

The application of data classification in remote work environments helps to increase the level of security and reduce the risk of privacy violations. Remote workers who clearly understand the importance of data protection and classification obligations can be more responsible in their work and compliance with security standards.

Thus, effective data classification is a key element of a cybersecurity strategy for remote work environments, enabling organizations to ensure reliable information protection and store sensitive data in a secure environment.

Conclusions

A hybrid model allows employees to participate in meetings virtually when needed and return to the in-person format for certain tasks or activities that require in-person interaction.

In addition, this approach provides employers with the necessary flexibility to plan for the future and make adjustments in changing conditions. Hybrid offices are increasingly becoming the new normal for many organizations around the world as they seek to offer

more balanced approaches to their employees in an ever-changing economic landscape. It's substantiated that ensuring IT security in remote work requires a comprehensive approach and design of advanced technologies and practices.

It is proposed to use the ISAKMP specification, which describes the mechanisms for agreeing the attributes of the used protocols and ensures a quick response to changing threats and constant updating of protection strategies, which are key to successful cyber security management.

By taking into account some precautions in advance, the company will be able to find solutions that

will allow to use the possibilities of the new hybrid office as efficiently as possible. It goes without saying that creating a hybrid office can be extremely beneficial for both employers and employees. It becomes possible to use all the conveniences of the office and at the same time work in a comfortable environment that meets the needs of security.

Due to the fact that the security of remote work also depends on the awareness of users, it has been proven that training staff on cyber security issues, explaining potential threats and periodically monitoring the entire cyber protection system significantly reduces the risk of incidents.

REFERENCES

1. Бурячок В., Киричок Р. *Основи інформаційної та кібернетичної безпеки*. Київ: КУБГ, 2019. 320 с.
2. Гребенюк А., Рибальченко Л. *Управління інформаційною безпекою*. Дніпро: ДДУВС, 2020. 144 с.
3. Кавун С., Носов В., Манжай О. *Інформаційна безпека*. посіб. Харків: ХНЕУ, 2013. 352 с.
4. Vazhynskiy V., Tkachov V. Проблематика безпеки та критерії надійності мультимарних середовищ. *Системи управління, навігації та зв'язку*. Полтава: ПНТУ, 2023, Т. 3 (73). 75-78.
5. Al-Ammouri A., M. Dekhtyar, R. Ishchenko, E. Klochan *Методи та засоби захисту інформації. Системи управління, навігації та зв'язку*. Полтава. ПНТУ, 2024. Т. 1 (75). С. 38-44.
6. Al-Ammouri A., Lebid I., Dekhtiar M., Lebid I., Al-Ammori H. Development of a mathematical model of reliable structures of information-control systems *Eastern-European Journal of Enterprise Technologies*, 2022. Vol. 5/9, Issue (119). С. 68–78.
7. Аль-Амморі. А., Ляковський В., Попова Л., Тимченко О., Полева Н. *Інформаційні системи та мережі*. К-НТУ. 2021, 194 р.
8. Подлевський Б., Рикалюк. Р. *Теорія інформації*. Львів. ЛНУ, 2016. 342 р.
9. Жураковський Ю., Полтораки В. *Теорія інформації та кодування*. К. Вища школа, 2011. 255 р.
10. Тулякова Н. *Теорія інформації*. Суми. СумДУ, 2008. 212.
11. Serdiukov D., Sievierinov O., Sydorenko Z. Особливості розгортання застосунку ESET ndm/mdc для забезпечення безпеки мобільних пристроїв. *Системи управління, навігації та зв'язку*. Полтава: ПНТУ, 2023, Т. 4 (74). С. 102-105.
12. The ultimate guide to building a hybrid desktop from scratch URL: <https://fliphtml5.com/learning-center/uk/the-ultimate-guide-to-building-a-hybrid-workplace-from-scratch/>
13. How hybrid work has affected the work of companies – research URL: <https://news.finance.ua/ua/yak-hibrydna-roboty-vplynula-na-robotu-kompaniy-doslidzhennya>
14. 20 Best Hybrid Collaboration Tools for Remote Work URL: <https://fliphtml5.com/learning-center/uk/top-20-hybrid-collaboration-tools/>
15. CEE Cybersecurity Forum URL: <https://dou.ua/calendar/46498/>

Received (Надійшла) 03.07.2024

Accepted for publication (Прийнята до друку) 21.08.2024

Безпека інформаційних технологій в гібридному робочому середовищі

А. Капітон, О. Дзюбан, Р. Бараненко, Г. Сокол

Анотація. У статті розглянуто основні аспекти безпеки інформаційних технологій в гібридному робочому середовищі. Доведено, що зі стрімким наростанням популярності віддаленої роботи традиційні стандарти безпеки на робочому місці є недостатніми, що породило реальну потребу в ефективних заходах кібербезпеки. Віддалена робота стала нормою для багатьох компаній, але разом з цим зросла й загроза кібератак. Розглянуто особливості та ризики, пов'язані із організацією гібридного робочого середовища, спричиненого віддаленим працевлаштуванням, а також досліджено методи забезпечення безпеки в такому робочому середовищі. Виявлено головні проблеми, що виникають та потребують їх попередження. Доведено, що однією з ключових областей є безпечне підключення до корпоративних мереж з віддалених місць. Встановлено, що традиційні методи VPN можуть бути неефективними або недостатньо безпечними в умовах масштабного використання, тому необхідно використовувати сучасні засоби шифрування трафіку та механізми автентифікації, які мають на меті організацію захисту доступу до корпоративних ресурсів. В роботі виконано аналіз ключових складових, що відповідають за безпеку кінцевих точок. Визначено вимоги, що уможливило забезпечення актуальних патчів та оновлень для програмного забезпечення та операційних систем віддалених пристроїв, які є невід'ємною частиною стратегії забезпечення безпеки. Досліджена провідна роль розробки та впровадження механізму контролю за доступом до даних та можливістю віддаленого відключення пристроїв у разі виявлення підозрілих дій. захист даних відіграє критичну роль у віддаленій роботі. Визначена необхідність проектування та розробка системи захисту всіх віддалених ресурсів, засоби розпізнавання та визначення у підозрілі повідомлення та посилення, а також використання сильних паролів та двофакторної автентифікації.

Ключові слова: інформаційна безпека, гібридне робоче середовище, кібербезпека, цифрова трансформація.

А. А. Кочина, К. Н. Азімов

Харківський національний автомобільно-дорожній університет, Харків, Україна

РОЗРОБКА СИСТЕМИ АВТОМАТИЗОВАНОГО ПАРСИНГУ РАЗОВИХ ЗАМОВЛЕНЬ НА ВАНТАЖНІ ПЕРЕВЕЗЕННЯ У МІЖМІСЬКОМУ СПОЛУЧЕННІ

Анотація. Ефективне переміщення товарів на великі відстані є основою сучасної торгівлі. Автоматичні механізми збору даних стимулюють розвиток системи, яка збирає і оновлює дані в режимі реального часу, використовуючи такі технології, як аналіз даних, сканування даних і сенсорні мережі для створення динамічної платформи швидкого реагування. Цей підхід, заснований на даних, має величезний потенціал як для вантажовідправників, так і для перевізників. **Мета.** Визначення теоретичних основ автоматизованого збору даних для розробки парсингової системи для автоматизованого пошуку та обробки разових замовлень на міжміські вантажні перевезення. **Методологія.** Статистичне та імітаційне моделювання. **Постановка задачі** - виявлення та оцінка потенційних замовлень на онлайн-платформі DELLA, інтеграція в бізнес-процеси компанії, пов'язані з обробкою та аналізом інформації. **Результати.** У статті проаналізовано теоретичні основи розробки програмного забезпечення для автоматизованого збору вхідних даних. Розроблено автоматизовану систему парсингу для пошуку та обробки разових замовлень на міжміські вантажні перевезення. Розробка автоматизованої системи парсингу дозволила отримати матрицю разових замовлень на регіональному рівні України. **Висновки:** Для моделювання потоку разових замовлень запропоновано використовувати методи машинного навчання для створення парсеру даних, який структурує інформацію веб-сторінки та надає її у вигляді, готовому для аналізу. Пошук інформації здійснюється на найпопулярнішому логістичному сайті серед бірж вантажних перевезень в Україні DELLA, який має найбільшу цільову аудиторію в секторі автомобільних перевезень. **Оригінальність.** На основі аналізу програмного забезпечення для автоматизованого збору даних створено парсингову систему для автоматизованого пошуку та обробки разових замовлень на міжміські вантажні перевезення у міжміському сполученні. **Практична цінність.** В результаті обробки масиву даних сформовано матрицю разових замовлень, яка визначає попит на ринку міжміських автомобільних перевезень України.

Ключові слова: парсинг, попит, разові замовлення, системи автоматизованого пошуку даних.

Вступ

Обробка та аналіз даних передбачає вивчення передових методів очищення, структурування та аналізу цих наборів даних. Механізми автоматичного збору спонукають до розробки системи, яка збирає й оновлює дані в режимі реального часу, використовуючи такі технології, як парсинг даних, сканування даних і сенсорні мережі, для створення динамічної платформи швидкого реагування. Даний підхід на основі даних має величезний потенціал як для вантажовідправників, так і для перевізників. Невеликі компанії отримують доступ до більшої кількості перевізників, оптимізованих моделей ціноутворення та відстеження відправлень у реальному часі. У результаті транспортні компанії отримують вигоду від підвищення ефективності роботи, зменшення кількості порожніх пробігів і можливості задовольнити раніше невикористаний сегмент ринку.

Аналіз публікацій. Парсинг – це автоматизований збір та структурування інформації з сайтів за допомогою програми чи сервісу. Парсинг зазвичай застосовують, коли потрібно швидко зібрати великий обсяг даних. Область застосування парсингу можна звести до двох цілей [1]:

- аналіз конкурентів, щоб краще розуміти, як вони працюють, та запозичувати у них якісь підходи;
- аналіз власного майданчика для усунення помилок, швидкого впровадження змін тощо.

Під системою автоматизованого парсингу можна розуміти програмне забезпечення (ПЗ), яке використовується для збирання даних з різних джерел.

Для створення такого ПЗ існує два підходи.

Один з них передбачає використання готових систем парсингу з відкритим кодом, таких як Scrapy або BeautifulSoup і побудову на їх основі власного парсеру. Як правило, ці системи можуть легко адаптуватися до конкретних завдань.

Другий підхід пропонує розробку власного парсеру за допомогою мов програмування, таких як Python або Java. Це дозволяє мати повний контроль за процесом парсингу та збору даних.

У порівнянні з ручним збиранням даних, парсинг дозволяє прискорити цей процес у десятки разів. Це відкриває нові можливості для дослідників та експертів. У порівнянні з ручним збиранням інформації, парсинг менш схильний до помилок. Але якщо було допущено неточності на етапі проектування (логічну помилку), це може призвести до спотворення даних та їх знецінення.

Мета та постановка задачі. Метою статті є визначення теоретичних основ розробки програмного забезпечення для автоматизованого збору даних. Розробка систем парсингу для автоматизованого пошуку та обробки разових замовлень на міжміські вантажні перевезення в міжміському сполученні. Постановка задачі – виявлення та оцінка потенційних замовлень на онлайн-майданчику DELLA, інтеграція в бізнес-процеси підприємства, пов'язані з обробкою та аналізом інформації.

Виклад основного матеріалу

Методика парсингу даних розробляється з метою автоматизації процесу пошуку разових замовлень на автомобільні міжміські перевезення. Враховуючи випадковий характер надходження разових

замовлень, процес їх обслуговування має певну складність при управлінні ланцюгами постачань. Адже цей фактор вимагає від перевізників постійного пошуку таких замовлень і швидкого реагування на їх появу. Крім того перевізник має обрати правильну для себе стратегію поведінки, а саме використовувати заздалегідь розроблені та обґрунтовані варіанти управлінських рішень або посилатись винятково на власний досвід [2].

Для полегшення даного процесу було запропоновано розробити програмний продукт, що значно скоротить час на аналіз про надходження разових замовлень, спростить його і, як результат, полегшить процес обслуговування замовлень.

Основну роль в даному процесі відіграє збирання вхідних даних.

Первинні дані – це дані, які дослідник збирає вперше. Зауважимо, що на відміну від терміну первинні дані, вхідні дані є більш загальним поняттям і використовується для опису функціонуючих систем. Так, для початку роботи програми необхідна наявність первинних даних. При цьому результати роботи програмного продукту можуть бути вхідними даними на наступних циклах її роботи, доки не виконуються умови, що дозволять завершити виконання програми.

Метод парсингу має кілька етапів для витягнення потрібної інформації із наукових сайтів:

а) збір контенту. Пошукова машина парсинг завантажує код сторінки сайту, під час такого процесу до нього долучається спеціальний скрипт, який розбиває весь код на частини та аналізує інформацію яка потрібна досліднику;

б) витяг інформації. Вся інформація, яка витягується зі сторінки, може бути не потрібна користувачеві, адже його цікавить тільки конкретне, наприклад, розділи статей, наукових робіт інших авторів в яких він може знайти для себе корисні дані та використати їх у своєму науковому дослідженні. Парсер буде знаходити ті сторінки чи розділи, де розміщена інформація про дану категорію, витягнувши все в кінцевий файл тільки тексти цих даних;

в) збереження результатів. Коли вся потрібна інформація витягнута з сайту, її потрібно зберегти. Зазвичай вносяться записи в базу даних, адже так зручніше аналізувати інформацію з даних інших наукових робіт.

Запропонована методика дасть можливість дослідити та проаналізувати отриману інформацію, чи потрібна вона дослідникові для внесення її у свою роботу. Для збору даних було запропоновано парсер для зазначеного веб-сайту за допомогою мови програмування Python та сторонніх бібліотек. В якості джерела інформації було обрано дані про вантажні зав'язки з веб-сайту della.ua.

Додаток був написаний з використанням мови програмування Python. Даний вибір пов'язаний з її популярністю в академічній сфері [3]. Python – це потужна мова програмування, яка проста у вивченні. Він має ефективні структури даних високого рівня та простий, але ефективний підхід до об'єктно-орієнтованого програмування. Елегантний синтаксис і

динамічна типізація Python разом із інтерпретованою природою роблять його ідеальною мовою для сценаріїв і швидкої розробки додатків у багатьох галузях для поширених платформ.

Інтерпретатор Python і обширна стандартна бібліотека доступні у вихідному або двійковому вигляді для всіх основних платформ на веб-сайті Python. Інтерпретатор Python легко розширюється за допомогою нових функцій і типів даних, реалізованих у C або C++ (або інших мовах, похідних від C). Python також підходить як мова розширення для настроюваних програм.

Головною перевагою Python є те, що він йде в пакеті з дуже великою бібліотекою стандартних пакетів (тобто не треба їх додатково інстальювати, щоб створювати навіть досить складні проекти) та ще більшим списком пакетів Open Source, доступних в публічних репозиторіях (для легкого встановлення). У тому числі потужні пакети для обробки зображення, створення графіків, статистичного аналізу даних, мережевої комунікації, штучного інтелекту (особливо Deep Learning) та Machine Learning, опрацювання натуральної мови, взаємодії з базами даних, створення вебдодатків, графічних алгоритмів, витягування даних з вебсайтів [4]. За замовчуванням у Python доступна стандартна бібліотека, яка містить велику кількість функцій, що багаторазово використовуються. Крім того, доступно більше 137 000 бібліотек Python для різних завдань, серед яких інтернет-розробка, наука про дані та машинне навчання (ML).

Розглянемо найбільш популярні бібліотеки, що дозволяють спростити процес написання програмного продукту: Pandas, Requests та BeautifulSoup. Вони ліцензовані BSD-ліцензією або містять її розділи. Ліцензія BSD є дозвільною, тобто вона накладає мінімальні обмеження на використання та розповсюдження ліцензійного програмного забезпечення. Основною вимогою ліцензії BSD є те, що будь-яке поширення програмного забезпечення має включати копію ліцензії та відмову від відповідальності. Багато розробників програмного забезпечення та компанії використовують цю ліцензію для того, щоб їхня робота була доступна широкому колу користувачів при збереженні прав на програмне забезпечення.

Для роботи з бібліотеками та зчитування інформації з веб-сайту необхідно розуміти як комп'ютер завантажує веб-сторінки та отримує інформацію від сервера. Коли користувач відвідує сторінку в Інтернеті, комп'ютер використовує протокол передачі гіпертексту (HTTP) для завантаження цієї сторінки.

Принцип роботи HTTP протоколу можна описати наступним чином. Коли ми хочемо переглянути веб-сторінку, ми можемо використовувати різні типи девайсів: ноутбук, стаціонарний комп'ютер або телефон. Головне, щоб на пристрої була програма браузера. Користувач або вводить уніфікований покажчик ресурсу (URL) у пошуковий рядок браузера, або переходить за посиланням з відкритої сторінки. URL-адреса починається з HTTP. Це сигнал браузеру, що йому необхідно використовувати протокол HTTP для отримання документа за цією адресою. Зазвичай IP-адреси містять зручні та читабельні для людини

назви доменів, наприклад `highload.today` або `wikipedia.org`.

На другому кроці браузер шукає потрібну IP-адресу. Для цього браузер відправляє запит до DNS-сервера для з'ясування імені домену з його IP-адресою. Як тільки браузер визначив IP-адресу комп'ютера, на якому розміщений запит URL, він відправляє HTTP-запит. На четвертому кроці, коли хост-комп'ютер отримує HTTP-запит, він відправляє клієнту відповідь із змістом та метаданими. Після виконання всіх кроків, браузер, отримує всю необхідну інформацію для відображення запитаного документа, відтворює його на екрані користувача.

Ознайомившись з принципом роботи веб-сторінки доцільно розглянути способи збереження та відтворення інформації. Зберігання даних – це спосіб організації даних на носіях інформації для її подальшої обробки. Найбільш поширеними типами даних є текстові, графічні, аудіо та відео дані. У науковій сфері найбільш поширеним є текстовий формат зберігання даних.

Текстові формати - це формати зберігання даних, які використовуються для зберігання тексту. Вони можуть бути простими або форматуваними.

До найпоширеніших текстових форматів можна віднести:

- TXT простий текстовий формат, який не підтримує форматувannya. Він є стандартним текстовим форматом, який підтримується більшістю програм;

- DOC текстовий формат, який підтримує форматувannya, створений Microsoft Word. Він є популярним форматом для зберігання документів, таких як листи, звіти та презентації;

- DOCX текстовий формат, який підтримує форматувannya, сумісний з Microsoft Word 2007 і пізнішими версіями. Він є покращеною версією формату DOC;

- RTF текстовий формат, який підтримує форматувannya, створений Microsoft. Він є універсальним текстовим форматом, який підтримується більшістю програм;

- ODT текстовий формат, створений OpenDocument для ОС Linux. Він є відкритим форматом, який підтримується більшістю офісних програм.

Прості текстові формати, такі як TXT, мають ряд переваг. Вони є простими у використанні, а їх легко відкрити та редагувати в будь-якій програмі, яка підтримує текстові формати. Однак вони не дозволяють зберігати форматувannya тексту, що може бути недоліком для деяких завдань.

Форматовані текстові формати, такі як DOC, DOCX та RTF, дозволяють зберігати форматувannya тексту, що може бути корисним для завдань, які вимагають форматувannya, наприклад, для створення документів, таких як листи, звіти та презентації. Однак вони можуть бути більш складними у використанні, ніж прості текстові формати.

Вибір текстового формату залежить від конкретних потреб. Для простих завдань, таких як зберігання тексту без форматувannya, можна використовувати простий текстовий формат, такий як TXT. Для

завдань, які вимагають форматувannya, можна використовувати форматований текстовий формат, такий як DOC, DOCX або RTF.

Серед текстових форматів розрізняють такі, що мають деякі особливості, які відрізняють його від інших. До них відносяться:

- HTML (HyperText Markup Language) використовується для створення веб-сторінок. HTML-файли складаються з тегів, які описують структуру та вигляд веб-сторінки;

- XML (Extensible Markup Language) використовується для зберігання структурованих даних. XML-файли складаються з елементів, які описують структуру даних;

- JSON (JavaScript Object Notation) використовується для зберігання об'єктів JavaScript. JSON-файли складаються з пар ключ-значення, які описують властивості об'єкта;

- YAML (YAML Ain't Markup Language) також використовується для зберігання структурованих даних. На відміну від XML, YAML-файли складаються з ключів, значень, списків та інших елементів, які описують структуру даних;

- INI (Initialization File) використовується для зберігання конфігураційних даних. INI-файли складаються з розділів, які описують різні параметри конфігурації;

- Markdown. Це полегшена мова розмітки даних, яку створено з ухилом на прочитність та зручність у публікації з подальшим перетворенням. Використовується для створення форматovanого тексту, такого як заголовки, списки, посилання та інші. Markdown-файли можуть бути перетворені на HTML, PDF та інші формати.

Окрему увагу варто приділити CSV формату. CSV (Comma-Separated Values) – це текстовий формат, який використовується для зберігання табличних даних. У таких файлах дані зберігаються у вигляді рядків, де кожний рядок представляє одну запис у таблиці. Поля в записі розділяються за допомогою роздільника, як правило, коми. CSV-файли є популярним форматом для обміну даними між різними програмами, оскільки вони є простими для обробки та зрозумілими. CSV-файли також можна легко відкрити та редагувати в будь-якому текстовому редакторі.

Крім того вони можуть використовуватися для зберігання різних типів даних, включаючи текст, числа та дати. Однак важливо пам'ятати, що CSV-файли не підтримують форматувannya даних, наприклад, шрифти, розміри шрифтів, кольори, вирівнювання та інші. Вони також можуть бути більш складними у використанні, ніж форматovanі текстові формати, такі як DOC, DOCX та RTF.

Вибір CSV-формату залежить від конкретних потреб. Для простих завдань, таких як зберігання табличних даних без форматувannya, CSV-формат є хорошим вибором. В даній роботі для зберігання проміжних даних використовувався CSV-формат.

Розглянемо більш детальний опис засобів та бібліотек для обробки інформації на Python, а також

особливості бібліотек для парсингу, які були застосовані у даній роботі.

Pandas – це бібліотека з відкритим вихідним кодом, ліцензована BSD, що надає високопродуктивні, прості у використанні структури даних та інструменти аналізу даних для мови програмування Python. Одним з основних інструментів, що був використаний у даній роботі – це клас DataFrame. Він слугує для уявлення двовимірної таблиці зі змінюваною розмірністю. Ця структура даних є основною для *pandas* дозволяє персоналізувати позначення на осях (заголовки рядків та стовпців), виконувати арифметичні операції для рядків і стовпців таблиці. У визначеннях мови Python, DataFrame можна розглядати як dict-подібний контейнер для об'єктів Series.

Для виконання основних функцій кожний клас передбачає опис конструктора класу. Він може містити чисельну кількість параметрів.

Перший параметр класу DataFrame є dict-словарем, що може містити такі структури даних, як ряди, масиви, константи, класи даних або спискові об'єкти. Якщо дані є словниками dict, порядок стовпців відповідає порядку полів словника. Якщо dict містить ряди, які мають визначений індекс, структура вирівнюється за його індексом. Це вирівнювання також відбувається, якщо дані є серією або самим DataFrame. Вирівнювання виконується на входах Series/DataFrame. Якщо дані є списком словників dicts, порядок стовпців відповідає порядку розташування словників у списку

Якщо вхідні дані є списком dicts, порядок стовпців структури буде відповідати порядку вставки.

Один з додаткових вхідних параметрів конструктора DataFrame відповідає за заголовки створюваної структури. За замовчуванням мітки стовпців для результируючого фрейма, якщо дані не містять їх, будуть числами від 0 до n. Якщо дані містять мітки стовпців, вони будуть відтворені у заголовках стовпців. В даному випадку фреймом називається таблиця, що має певний набір функцій для її обробки.

Як тільки фрейм буде заповнено даними доцільно зберегти його на локальному носії. Розглянемо як можна експортувати фрейм до CSV файлу за допомогою *to_csv()* методу пакету *pandas*. За замовчуванням, *to_csv()* метод експортує фрейм до CSV файлу з рядковим індексом як перша колонка і комою в якості сепаратора даних [5].

Для персоналізації результатів у цього методу є численна кількість параметрів. Так, щоб встановити роздільник поля для вихідного файлу слугує параметр *sep=''*, за замовчуванням рівний комі. Для відтворення заголовків у вихідному файлі можна встановити параметр *header=True*. Даний параметр варто відключити при повторному записі того самого файлу. Параметр *index=True* повідомляє програмі пронумерувати кожний рядок. Для повноцінної роботи функції першим параметром має бути назва файлу для запису інформації.

Розглянемо наступну бібліотеку – requests. Це елегантна та проста бібліотека HTTP для Python, створена для людей. Вона дозволяє надзвичайно легко надсилати запити HTTP/1.1. Немає потреби вручну додавати рядки запиту до URL-адрес або

кодувати дані PUT і POST у формі – достатньо застосувати метод *json*.

Зробити запит за допомогою Requests дуже просто. Для цього достатньо зробити імпорт модуля та створити запит GET для отримання веб-сторінки. Після того як запит виконано у програма має об'єкт Response під назвою *r*. З цього об'єкта ми можемо отримати всю необхідну інформацію.

Бібліотека забезпечує широкий спектр функцій для взаємодії з HTTP-серверами. Ось деякі з основних функцій [6]:

а) відправка HTTP-запитів будь-якого типу – означає підтримку всіх основних типів HTTP-запитів, включаючи GET, POST, PUT, DELETE та HEAD;

б) обробка відповідей HTTP – дозволяє отримувати інформацію про відповіді HTTP, включаючи статус-коди, заголовки та контент;

в) підтримка різних методів аутентифікації, таких як Basic, Digest та OAuth;

г) налаштування параметрів HTTP-запитів.

Бібліотека requests дозволяє налаштовувати параметри HTTP-запитів, такі як заголовки, cookies та тимчасові файли. Вона є потужним інструментом, який може бути використаний для вирішення різних завдань, пов'язаних з HTTP. Вона є простою у використанні та забезпечує широкий спектр можливостей.

Останньою бібліотекою, яку використано для парсингу є BeautifulSoup. Це бібліотека Python для вилучення даних із файлів HTML та XML. Вона працює з будь-яким парсером та надає прості та звичні способи навігації, пошуку та зміни дерева розбору. Вона економить програмістам години та дні роботи [7].

Beautiful Soup підтримує парсер HTML, включений до стандартної бібліотеки Python, а також ряд сторонніх парсерів на Python. Одним із них є парсер *lxml*. Інша альтернатива – написаний виключно на Python парсер *html5lib*, який розбирає HTML так само, як це робить веб-браузер. У кожного метода є свої переваги та недоліки, але найкращим з них є *lxml* парсер. Він виокремлюється від інших своєю швидкістю та гнучкістю роботи.

Варто зауважити, якщо документ є невалідним, різні парсери будуть генерувати дерево BeautifulSoup для цього документа по-різному. При роботі з веб-додатками коли говорять, що документ є невалідним, це означає невідповідність його компонування та розмітки стандартам. Ці стандарти розроблюються консорціумом Всесвітньої павутини (W3C). Найновішим стандартом є HTML5.

Перевірка на валідність HTML документів є важливою для забезпечення того, щоб HTML-код був правильно написаний і міг бути правильно інтерпретований браузерами. Валідність HTML-коду також допомагає покращити його доступність для користувачів з обмеженими можливостями.

Для перевірки на валідність HTML документів можна використовувати різні інструменти. Деякі з них є безкоштовними та доступними в Інтернеті, інші є платними та вимагають встановлення на комп'ютер. Нижче наведено деякі з найпопулярніших інструментів для перевірки на валідність HTML документів:

– W3C Markup Validation Service. Безкоштовний онлайн-інструмент, який дозволяє перевірити HTML-код на відповідність стандарту HTML5;

– HTML Tidy. Безкоштовний інструмент, який дозволяє перевірити HTML-код на відповідність стандарту HTML5 і виправити деякі помилки;

– Codesniffer. Безкоштовний інструмент, який дозволяє перевірити HTML-код на відповідність стандартам W3C і іншим стандартам кодування.

Для перевірки на валідність HTML документів можна також використовувати вбудовані засоби перевірки в таких текстових редакторах, як Atom, Sublime Text та Visual Studio Code.

Ознайомившись з компонентами проекту можемо перейти до його детального опису. Розглянемо шлях створення додатку.

На першому етапі було встановлено інтерпретатор Python для написання та запуску парсера, за допомогою *pip* системи управління пакетами Python встановлено 3 бібліотеки:

- `pip install beautifulsoup4;`
- `pip install requests;`
- `pip install pandas.`

Як було зазначено раніше, кожна бібліотека має своє призначення:

- `requests` відповідає за отримання даних (у вигляді html-сторінки) з веб-сайту на локальний комп'ютер;
- `beautifulsoup4` за інструменти обробки отриманих даних;
- `pandas` за зберігання оброблених даних на локальному носії у вигляді csv-файлу.

Встановлення пакетів здійснено через командний рядок Windows. За допомогою цих трьох бібліотек можна почати аналіз нашої веб-сторінки.

Враховавши простоту синтаксису мови Python відносно інших мов програмування та запуску програм, в якості середовища розробки обрано Notepad++, середовище перевірки коду – командний рядок Windows.

Для початку необхідно було виконати аналіз HTML-коду веб-сторінки для виявлення елементів з корисною інформацією.

Для цього було виконано пошук потрібної веб-сторінки з потрібної інформацією.

На головній сторінці сайту потрібно натиснути на вкладку «пошук вантажів і транспорту», розміщену в правій частині сайту. Перед нами відкривається нова сторінка, на якій пропонується знайти попутні вантажі, а також вільний і попутний транспорт (рис. 1).

Нас цікавить інформація про міжміські вантажні перевезення,

тому радіоперемикач має бути встановлено на пункті «Вантажі», що включено за замовчуванням. У параметрах «звідки», «куди» обираємо «Україна».

Для визначення терміну виконання заявки налаштуємо пункт «Дата». Решту параметрів можна залишити за замовчуванням.

Після виконання налаштувань пошуку натискаємо кнопку «Знайти».

В результаті відкривається нова сторінка з потрібною інформацією.

Надалі працюємо з даною сторінкою.

За допомогою інструменту «Переглянути код» в інструментах веб-браузера Chrome (рис. 2) було отримано код веб-сторінки, який збережено в текстовому форматі.

Отриманий код було «передано» до бібліотеки *beautifulsoup4*, яка виконала його структурування.

Проаналізувавши код, було виділено елементи, що містять необхідну інформацію. Результатом аналізу стала можливість здійснювати пошук інформації по ключовим елементам, що дало змогу автоматизувати пошук.

На третьому етапі було організовано запис отриманої інформації до csv-файлу.

Так було визначено назви та послідовність заголовків, формат даних тощо.

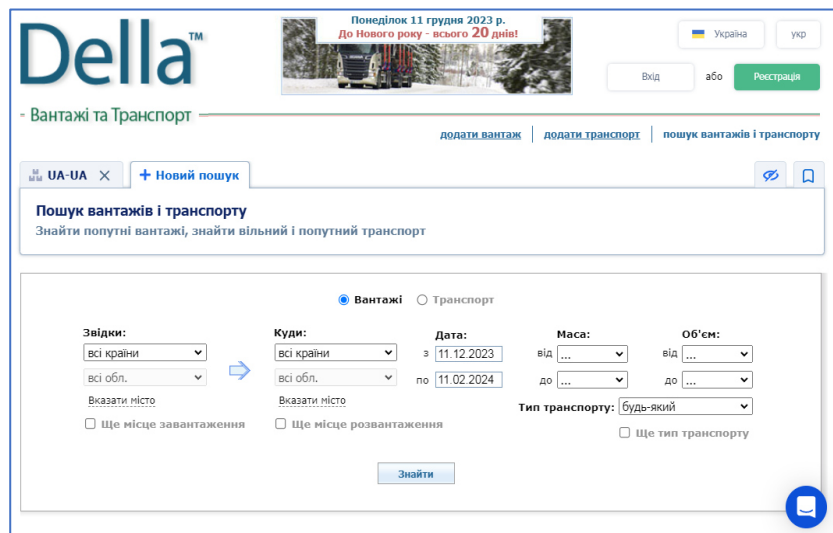


Рис. 1. Веб-сторінка «Пошук вантажів і транспорту»

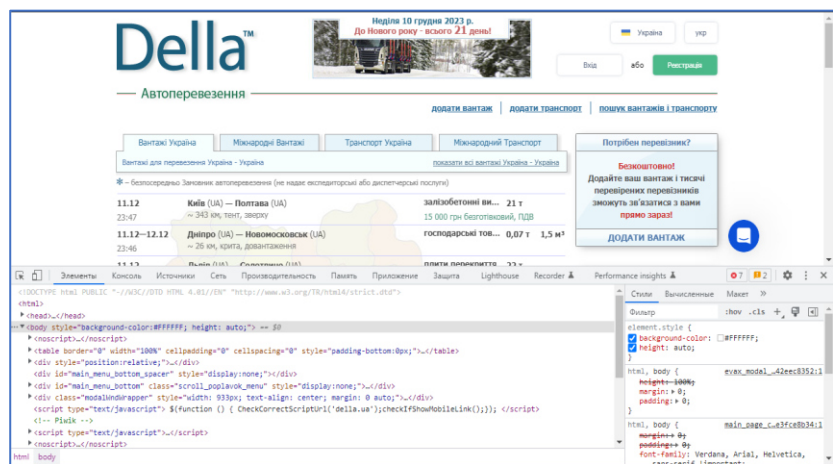


Рис. 2. Інструмент «Переглянути код» в нижній частині вікна браузера

Загалом, роботу парсера можна описати наступним чином.

На першому кроці відбувається завантаження інформації з веб-ресурсу за допомогою методів бібліотеки requests.

Отримавши відповідь на запит, завантажена інформація конвертується в текстовий формат та передається на обробку до бібліотеки *beautifulsoup4*, яка відбудовує структуру веб-сторінки локально для подальшої навігації по ключовим елементам.

На останньому кроці за допомогою методів навігації по елементам веб-сторінки завантажена інформація оброблюється та зберігається в словнику Python, після чого записується до csv-файлу.

Для запису до файлу використовується бібліотека *pandas*.

В результаті виконання парсингу веб-ресурсу було отримано інформацію про вантажні перевезення по Україні за період з 30.08.2023 по 05.09.2023.

Отримавши оброблені дані було виконано підрахунок кількості заявок, що припадають на кожну область та сформовано матрицю разових замовлень на міжміські вантажні перевезення [8].

Для автоматизації цього процесу було написано скрипт на Python.

Висновки

Для моделювання потоку разових замовлень запропоновано використовувати методи машинного навчання. Ці методи дозволяють враховувати складні залежності між різними чинниками, які впливають на попит перевезень.

У результаті пошуку способів збирання даних виявлено чисельну кількість різноманітних підходів, що ґрунтуються на роботі програмних продуктів. Одним з таких є парсер даних, який виконує структурування інформації веб-сторінки і надає їх у готовому для аналізу вигляді.

Пошук інформації відбувається на найбільш популярному серед вантажних автотранспортних бірж в Україні логістичному сайті "DELLA", що відповідно має найбільшу цільову аудиторію у сфері автомобільних перевезень. В результаті обробки масиву даних було сформовано матрицю разових замовлень, яка визначає попит на ринку міжміських автомобільних перевезень України.

СПИСОК ЛІТЕРАТУРИ

1. Мокін В. Б., Романюк О. Н., Войтко В. В., Сторчак В. Г. та Гавенко О. В. Моделювання параметрів транспортної мережі в середовищі автоматизованої системи пошуку оптимальних рішень. *Інформаційні технології та комп'ютерна інженерія*. URL: <https://www.researchgate.net>.
2. Кочина А.А. Визначення закономірностей потоку разових замовлень у міжміському сполученні при управлінні ланцюгами постачань. *Сучасні технології в машинобудуванні та транспорті*. 2023. Вип. 10. С. 151-159.
3. Систематичне накопичення дослідних матеріалів. *Підручники для вузів онлайн*: веб-сайт. URL: https://pidru4niki.com/1089030461409/pedagogika/sistematichne_nakopichennya_doslidnih_materialiv.
4. Олексій Волошин Переваги і недоліки Python. *Блог компанії Hillel*: веб-сайт. URL: <https://blog.ithillel.ua/ru/articles/preimushchestva-i-nedostatki-yazyka-python>.
5. Mrinalwalia. Saving a Pandas Dataframe as a CSV. *Geeksforgeeks*: веб-сайт. URL: <https://www.geeksforgeeks.org/saving-a-pandas-dataframe-as-a-csv/>.
6. Kenneth Reitz. Requests: HTTP for Humans. URL: <https://requests.readthedocs.io/en/latest/>.
7. Beautiful Soup Documentation. URL: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>.
8. Мосьпан Н. В. Формування стратегій автотранспортних підприємств по обслуговуванню разових замовлень на перевезення вантажів у міжміському сполученні : автореф. дис. ... канд. техн. наук : 05.22.01. Харків, 2018. 212 с.

Received (Надійшла) 26.05.2024

Accepted for publication (Прийнята до друку) 14.08.2024

Development of an automated parsing system one-time orders for freight transport in long-distance traffic

A. Kochina, K. Asimov

Abstract. Problem. The efficient movement of goods over vast distances is the backbone of modern trade. Automatic collection mechanisms are driving the development of a system that collects and updates data in real time, using technologies such as data parsing, data scanning and sensor networks to create a dynamic, rapid response platform. This data-driven approach has huge potential for both shippers and carriers. Goal. Defining the theoretical foundations of automated data collection software for the development of a parsing system for automated search and processing of one-time orders for long-distance freight transport. Methodology. Statistical and simulation modelling. Task statement – Identification and evaluation of potential orders on the DELLA online platform, integration into the company's business processes related to information processing and analysis. Results. The article analyses the theoretical foundations of software development for automated input data collection. An automated parsing system was developed to search and process one-time orders for long-distance freight transport. The development of the automated parsing system allowed to obtain a matrix of one-time orders at the regional level of Ukraine. Conclusions: To model the flow of one-time orders, it is proposed to use machine learning methods to create a data parser that structures the information of a web page and provides it in a form ready for analysis. The information is searched for on the most popular logistics website among freight transport exchanges in Ukraine, DELLA, which has the largest target audience in the road transport sector. Originality. Based on the analysis of automated data collection software, a parsing system was created for automated search and processing of one-time orders for long-distance freight transport in intercity traffic. Practical value. As a result of processing the data set, a matrix of one-time orders was formed, which determines the demand in the Ukrainian intercity road transport market.

Keywords: parsing, demand, one-off orders, automated data retrieval systems.

О. І. Лактіонов, Н. М. Педченко, А. С. Янко

Національний університет «Полтавська політехніка імені Юрія Кондратюка», Полтава, Україна

ПРАКТИЧНІ КЕЙСИ СТВОРЕННЯ ЗГОРТКОВИХ МОДЕЛЕЙ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ЗАДАЧ РОЗПІЗНАВАННЯ ОБРАЗІВ

Анотація. Мета статті підвищення ефективності розпізнавання образів шляхом створення бінарного класифікатора об'єктів цивільної інфраструктури засобами глибокого навчання. **Результати дослідження.** Створено згорткові моделі штучного інтелекту для розпізнавання об'єктів цивільної інфраструктури. Побудовано згорткову нейронну мережу Sequential, яка нараховує три згорткових шари, шари пулінгу, шар перетворення, пов'язаний шар та вихідний шар. Оптиміальне значення помилки при навчанні/тестуванні становить 0,0650/0,4424, де точність 0,98/0,92 відповідно, у той час як результати третьої епохи демонструють помилку 0,2442/0,2595 відповідно, а точність – 0,93. Використано попередньо навчену модель мережі VGG16, котру перенавчено на досліджуваному датасеті. Вказана модель демонструє мінімальне значення помилки для навчання/тестування 0,0278/0,1538 при точності 0,99/0,96 відповідно. **Наукова новизна.** Набуло подальшого розвитку використання згорткових нейронних мереж для завдань розпізнавання рівня руйнувань цивільної інфраструктури. **Практичне значення.** Побудовано дві моделі згорткової нейронної мережі, зокрема модель Sequential й VGG16 для вирішення завдань розпізнавання зруйнованих і цілих будинків. Для виявлення вказаних об'єктів необхідною передумовою є використання камери й відповідного апаратуного забезпечення, на кшталт raspberry пі або персонального комп'ютера чи ноутбука.

Ключові слова: комп'ютерний зір, завдання класифікації, глибоке навчання, нейронні мережі.

Постановка проблеми

Використання супутникових технологій потребує створення нових інструментів штучного інтелекту, котрий дозволяє автоматизувати процеси виконувани вручну [1]. Інструменти штучного інтелекту відомі принципом постійного навчання на помилках, що дозволяє будувати моделі, точність яких сягає 90% і більше [2].

Розвідка територій й формування датасету з фото потребує швидкої обробки даних й побудови моделей штучного інтелекту. Довга за часом процедура створення моделі впливає на виконання завдань, оскільки за час навчання моделі змінюються розміщення досліджуваних об'єктів.

Існуючі тактики створення нових рішень щодо побудови моделей штучного інтелекту різноманітні й потребують низку відповідних витрат часу. Наприклад повний час створення моделі штучного інтелекту, що розглянута у попередньому дослідженні [3], сягає орієнтовно чотири-п'ять годин. При цьому рішення працює лише з числовими даними.

Крім витрат часу існуючі підходи зорієнтовані на вивчення конкретних завдань, зокрема для автоматичної оцінки типу та ступеня пошкодження дерев'яних конструкцій [4], де потрібно здійснювати перенавчання моделі на інших датасетах. З іншої сторони, більшість нейронних мереж будуються за принципом гібридизації, що покращує точність і збільшує складність моделі [5]. Втім, існуюча низка програмного забезпечення теж потребує ретельного вивчення функціоналу й можливостей.

Таким чином актуалізується питання створення класифікатора об'єктів на основі датасету цивільної інфраструктури.

Аналіз останніх досліджень і публікацій. Останні досягнення у сфері штучного інтелекту та робототехніки породили нові підходи щодо використання нейронних мереж для вирішення різних

практичних задач. Так напрацювання [6] пропонує підходи щодо використання штучного інтелекту у сфері будівництва. За висновками авторів [6], інтелектуальні технології використовуватимуться на етапах проектування, виробництва та експлуатації будівельних конструкцій.

Крім будівництва штучний інтелект активно використовують у наземній та повітряній робототехніці, зокрема для виявлення об'єктів різного типу [7]. Цікаво, що виявлення об'єктів відбувається у режимі реального часу й без затримок за допомогою моделі глибокого навчання YOLOv7. Вказана модель виявляє одночасно кілька об'єктів з точністю 0,97 й 92 кадрами в секунду при 0,02 мс на одне зображення.

Та можливість нейронних мереж на цьому не зупиняються, де у напрацюванні [8] розглянуто підходи щодо інтелектуальної діагностики несправностей, інструменти котрої характеризуються високими обчислювальними можливостями. Адекватність підходу доведено за допомогою датасету про підшипники кочення та даних несправності ротора. Пошук шляхів удосконалення обчислювальних можливостей комп'ютерних систем здійснювався також за допомогою ортогональних базисів й процесу паралельної обробки проєкцій вихідного числа [9]. Такий спосіб підвищує процес діагностування даних у 1,2 рази.

Нарощування обчислювальних можливостей комп'ютерної техніки корелюється з інструментарієм глибокого навчання. Як зазначено у напрацюванні [10], виявлення об'єктів на фото під різними кутами є складною задачею навіть й для глибоких нейронних мереж. У зв'язку з цим автори [10] пропонують метод візуальної кластеризації, що визначає різницю форми об'єкта дослідження з точністю 86,3%.

Паралельно з вивченням поведінки згорткових мереж щодо класифікації зображень досліджуються

й питання прогнозування різних подій. Наприклад у роботі [11] вивчається гібридна ймовірнісна модель на основі згорткових та рекурентних мереж для прогнозування навантаження домогосподарств. Принциповим у дослідженні є поєднання двох типів нейронних мереж й створення комбінованого варіанту мережі. Для вивчення вказаних тактик досліджень передбачається існування відповідного програмного забезпечення. Так, відома освітня платформа [12] поєднує Moodle та Jupyter для навчання у галузі STEM.

Створенням гібридних згорткових моделей також присвячено роботу [13], де обиралася оптимальна стратегія. Перевага запропонованого підходу досягається за рахунок нової ідеї оптимізації ознак. На відміну від результатів [13] у роботі [14] вивчалися питання паралельно діючих класифікаторів побудованих на основі методу опорних векторів та згорткової мережі.

Залежно від вихідних зображень, що являють собою досліджувану вибірку, одна й та сама технологія здатна вирішувати різні завдання, наприклад картографування зсувів [15]. За результатами досліджень встановлено території, схильні до зсувів ґрунту та визначено рівні ризику будівель.

Як видно з досліджень [6–15], використання нейронних мереж розглядається у різних сферах, де будуються як звичайні так і гібридні мережі. Побудова мережі зорієнтована на використання конкретного програмного забезпечення та вміння його об'єкти.

Для побудови власного класифікатора об'єктів цивільної інфраструктури, відповідно до рекомендації, розглянемо використання згорткової нейронної мережі.

Мета – підвищення ефективності розпізнавання образів шляхом створення бінарного класифікатора об'єктів цивільної інфраструктури засобами глибокого навчання.

Методологія дослідження

Процес створення нейронної мережі передбачає існування формальної постановки завдання дослідження. Нехай A та B є множинами світлин, що являють класи 0 й 1 відповідно. Завдання полягає у створенні класифікатора $y=f(x)$, де x є фотографією з множини A чи B . Створена модель класифікатора використовується у задачах розпізнавання образів й передбачає існування вихідного датасету.

У якості вихідного датасету дослідження використано зображення будинків з ресурсу [16]. Датасет включає 2735 зображень будівель, з них 1237 пошкоджених і 1498 непошкоджених. Формування навчального і тестового наборів фото здійснювалося випадковим чином у співвідношенні, де розмір навчальної вибірки складає $split_ratio = 0,8$.

Вихідні набори зображень використовувалися для побудови двох згорткових нейронних мереж. Побудова першої моделі нейронної мережі (Sequential) здійснювалася за відомою технологією [17] мовою програмування Python та з використанням бібліотек Tensor Flow й Keras. Для прискорення

пошуку гіперпараметрів та оптимальної архітектури мережі використано бібліотеку Keras Tuner [18], де перевага надавалася методу пошуку гіперпараметрів Random Search при $max_trials=10$.

Обмеженнями для пошуку архітектури нейронної мережі були значення кількості фільтрів у згорткових шарах, що досліджувалися у діапазоні [1, 4] з кроком 1.

Параметри пов'язаного шару теж вивчалися у діапазоні [1, 4] з кроком 1. При компіляції моделі швидкість навчання обмежувалася діапазоном [0,05, 1] з кроком 0,05, де функція втрати $loss='binary_crossentropy'$. З метою зменшення часу навчання моделі використовувався підхід ранньої зупинки, де $patience=8$.

Якість побудованих результатів дослідження фіксувалася метриками accuracy, precision, recall та f1 score, де здійснювався порівняльний аналіз навчального й тестового наборів. Додатково здійснювалася перевірка помилки моделі залежно від епохи.

Графічна інтерпретація отриманих результатів дослідження реалізовувалася засобами бібліотеки matplotlib, де будувалися візуалізації результатів процесу навчання моделі.

Побудовані моделі зберігалися за допомогою функціоналу бібліотеки Keras у файловий формат .h5, котрий використовувався у зовнішніх сервісах (Netron) для графічної інтерпретації моделі мережі [19].

Для порівняльного аналізу будувалася ще одна згорткова нейронна мережа VGG16 [17], котра є попередньо навченою. Це друга мережа. З метою зменшення обсягу часу навчання мережі здійснювалося перенавчання мережі на датасеті дослідження за обмежень використання верхнього шару шляхом додавання нових шарів. Моделювання мережі VGG16 проводилося протягом 3 епох.

Виклад основного матеріалу

Процес створення моделі згорткової нейронної мережі щодо класифікації об'єктів на зображеннях передбачає вибір оптимальної архітектури на основі вихідного датасету, табл. 1.

Як видно з табл. 1, згорткова нейронна мережа Sequential нараховує три згорткових шари, шари пулінгу, шар перетворення, пов'язаний шар та вихідний шар. Визначена архітектура нейронної мережі класифікації об'єктів на зображеннях використовується для побудови моделі.

Відповідно до методології дослідження процес побудови моделі нейронної мережі нараховував 10 епох, де фіксувалася точність на навчальній й тестовій вибірках, рис. 1.

Як видно з рис. 1 оптимальна точність навчальної й тестової вибірок досягається під час першої епохи та становить 0,93. Наступні епохи вказують на перенавчання моделі, оскільки збільшується різниця точності між навчальною й тестовою вибірками. Це пояснюється обмеженою кількістю вхідних зображень.

Результати визначення точності та помилка на навчальній й тестовій вибірках подано у табл. 2.

Таблиця 1 – Досліджувана архітектура згорткової нейронної мережі Sequential для класифікації об'єктів на зображеннях

Назва шару	Вихідна форма	Досліджувані параметри
Перший згортковий шар conv2d	(None, 254, 254, 2)	56
Шар пулінгу max_pooling2d	(None, 127, 127, 2)	0
Другий згортковий шар conv2d_1	(None, 125, 125, 2)	38
Шар пулінгу max_pooling2d_1	(None, 62, 62, 2)	0
Третій згортковий шар conv2d_2	(None, 60, 60, 2)	38
Шар пулінгу max_pooling2d_2	(None, 30, 30, 2)	0
Шар перетворення flatten	(None, 1800)	0
Пов'язаний шар з двома виходами dense	(None, 2)	3602
Вихідний шар dense_1	(None, 1)	3

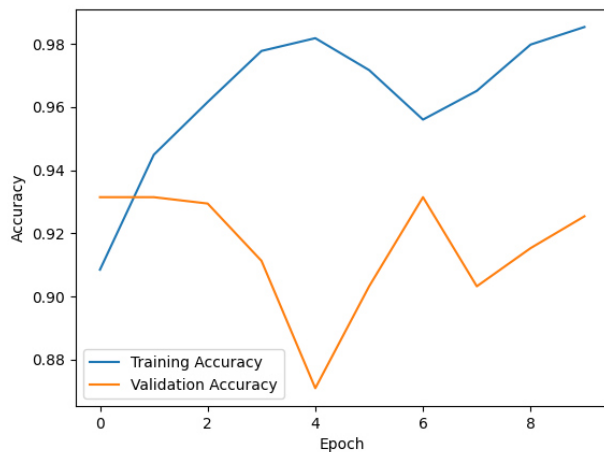


Рис. 1. Порівняльний аналіз залежності точності згорткової нейронної мережі (навчальна й тестова вибірки) Sequential від кількості епох

З точки зору балансу точності між навчальною й тестовою вибірками зображень перевагу має модель побудована під час другої епохи, де точність становить по 0,93 при помилці 0,2442/0,2595 відповідно. Але помилка під час другої епохи для навчального й тестового наборів велика. Оптимальне значення помилки при навчанні/тестуванні досяга-

Таблиця 2 – Результати навчання згорткової нейронної мережі Sequential для класифікації об'єктів на зображеннях

№ епохи	Точність на навчальній/ тренувальній вибірках	Помилка на навчальній/ тренувальній вибірках
1	0,88/0,93	0,3857/0,2843
2	0,93/0,93	0,2442/0,2595
3	0,95/0,92	0,1751/0,2628
4	0,97/0,91	0,1234/0,3922
5	0,98/0,87	0,0985/0,7616
6	0,97/0,9	0,1003/0,566
7	0,94/0,93	0,1587/0,328
8	0,96/0,9	0,1128/0,5265
9	0,97/0,91	0,0763/0,6152
10	0,98/0,92	0,0650/0,4424

ється аж на десятій епісі й становить 0,0650/0,4424 відповідно, де точність 0,98/0,92 відповідно. З одного боку, модель має ознаки перенавчання, з іншого – мінімальне значення помилки при Precision 0,93; Recall 0,91 й F1 Score 0,92.

На рис. 2 подано результати дослідження матриці плутанини найліпшої моделі мережі Sequential.

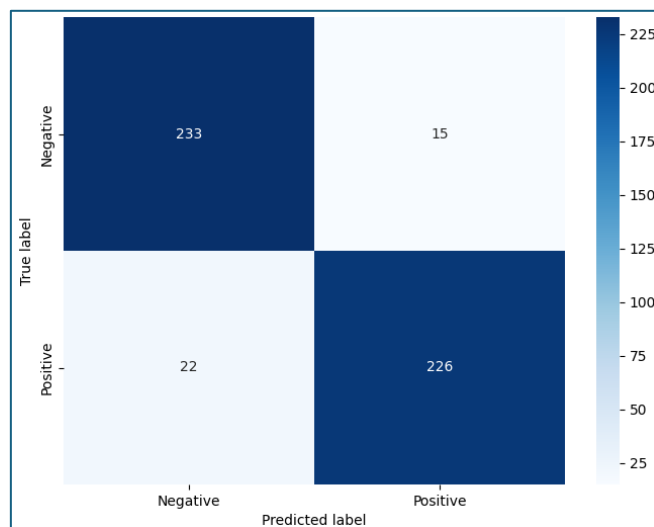


Рис. 2. Результати дослідження матриці плутанини оптимальної класифікаційної моделі мережі Sequential

З рис. 2 видно, що 233 й 226 зображень відповідних класів були вірно класифіковані, у той час як 22 й 15 зображень класифіковано хибно. Отримані хибно класифіковані зображення пояснюються обранням оптимальної моделі, котра має незначну різницю показників точності на навчальній й тестовій вибірках, а з іншої сторони мінімальне значення помилки.

Кількість хибно класифікованих зображень є різною для епох 1, 3-9, проте на вказаних епохах спостерігається процес перенавчання моделі, що доведено й на рис. 1.

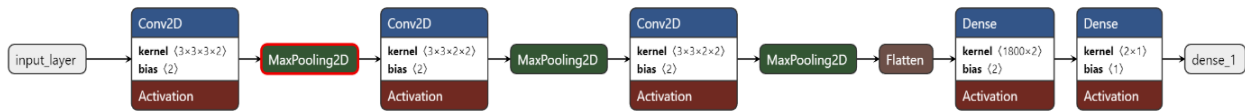


Рис. 3. Архітектура побудованої згорткової нейронної мережі для класифікації об'єктів на зображеннях створена засобами Netron

Вказана візуалізація використовувалася для демонстрації особливостей побудованої мережі й планування аспектів її майбутнього удосконалення. З метою порівняльного аналізу точності отриманих результатів першої згорткової нейронної мережі побудовано нейронну мережу VGG16. Оскільки відома мережа VGG16 попередньо навчена, але не на досліджуваному датасеті, до мережі додано досліджувану вибірку зображень – перенавчання, табл. 3.

Таблиця 3 – Результати навчання згорткової нейронної мережі VGG16 для класифікації об'єктів цивільної інфраструктури

№ епохи	Точність на навчальній/ тренувальній вибірках	Помилка на навчальній/ тренувальній вибірках
1	0,81/0,96	0,735/0,0905
2	0,97/0,96	0,0531/0,1056
3	0,99/0,96	0,0278/0,1538

З точки зору компромісу точності між навчальною й тестовою вибірками зображень цивільної інфраструктури він спостерігається за результатами другого етапу та становить 0,97/0,96 й вказує на незначні ознаки перенавчання. За критерієм помилки, мінімальне значення котрої спостерігається за результатами третьої епохи й становить для навчання/тестування 0,0278/0,1538 відповідно при точності 0,99/0,96.

Практичні рекомендації результатів дослідження

Проведене дослідження процесу побудови нейронної мережі на основі датасету будівель, за наведених у методології дослідження обмежень, дозво-

Варто відмітити, час навчання моделі такого типу складає орієнтовно до 30 хвилин, за вказаних у методології дослідження обмеженнях. У реаліях сьогодення, мінімізація витрат часу для навчання моделі є важливою з точки зору заощадження як енергоресурсів, так і швидкості впровадження моделі до використання. Остаточне рішення щодо показників якості згорткової моделі нейронної мережі схвалюється людиною, котра приймає рішення.

Додатково, для поглибленого аналізу побудованої мережі Sequential, відповідно до методології дослідження, здійснювалася її візуалізація, рис. 3.

ляє сформулювати практичні рекомендації щодо створення аналогічних моделей.

Процес побудови згорткової мережі прискорюється за рахунок використання автоматичного підбору архітектури й гіперпараметрів мережі, що реалізовано у Keras Tuner. При цьому важливим є використання Random Search, у той час як послідовний пошук збільшує час побудови моделі до 5 годин лише на одну епоху. Це важливо в умовах обмеження у часі.

Використання попередньо навченої моделі VGG16 й інтеграція до моделі датасету дослідження дозволяє перенавчити модель за 4 години, відповідно до умов вказаного дослідження.

Створені моделі можуть бути використані для бінарної класифікації зруйнованих і цілих будівель. Для автоматизації цього процесу рекомендовано здійснити інтеграцію моделі інструментами OpenCV [20] й реалізувати відповідні рішення апаратним забезпеченням, зокрема raspberry pi.

Створену модель мережі Sequential рекомендовано до використання у майбутньому, зокрема для перенавчання на нових датасетах, оскільки досвід використання VGG16 показав тенденцію зменшення помилки й підвищення точності.

Висновки

Задача побудови бінарного класифікатора об'єктів вирішується шляхом створення згорткової мережі Sequential та мережі VGG16.

Запропоновані рішення рекомендовано до використання для класифікації об'єктів цивільної інфраструктури, побудови майбутніх моделей глибокого навчання.

Крім того, запропоновані моделі у поєднанні з бібліотекою комп'ютерного зору дозволяють реалізувати принципи комп'ютерного зору.

СПИСОК ЛІТЕРАТУРИ

1. S V. Satellite Image Classification using CNN with Particle Swarm Optimization Classifier [Electronic resource] / Vidhya S, Balaji M, Kamaraj V // Procedia Computer Science. – 2024. – Vol. 233. – P. 979–987. – Mode of access: <https://doi.org/10.1016/j.procs.2024.03.287>

2. O'Shaughnessy D. Understanding Automatic Speech Recognition [Electronic resource] / Douglas O'Shaughnessy // Computer Speech & Language. – 2023. – P. 101538. – Mode of access: <https://doi.org/10.1016/j.csl.2023.101538>
3. Practical principles of integrating artificial intelligence into the technology of regional security predicting [Electronic resource] / Oleksandr Shefer [et al.] // Advanced Information Systems. – 2024. – Vol. 8, no. 1. – P. 86–93. – Mode of access: <https://doi.org/10.20998/2522-9052.2024.1.11>
4. Computing the characteristics of defects in wooden structures using image processing and CNN [Electronic resource] / Rana Ehtisham [et al.] // Automation in Construction. – 2024. – Vol. 158. – P. 105211. – Mode of access: <https://doi.org/10.1016/j.autcon.2023.105211>
5. Hybrid framework for respiratory lung diseases detection based on classical CNN and quantum classifiers from chest X-rays [Electronic resource] / G. V. Eswara Rao [et al.] // Biomedical Signal Processing and Control. – 2024. – Vol. 88. – P. 105567. – Mode of access: <https://doi.org/10.1016/j.bspc.2023.105567>
6. Artificial-intelligence-led revolution of construction materials: From molecules to Industry 4.0 [Electronic resource] / Xing Quan Wang [et al.] // Matter. – 2023. – Vol. 6, no. 6. – P. 1831–1859. – Mode of access: <https://doi.org/10.1016/j.matt.2023.04.016>
7. Yasmine G. Anti-drone systems: An Attention Based Improved YOLOv7 model for a real-time detection and identification of multi-airborne target [Electronic resource] / Ghazlane Yasmine, Gmira Maha, Medromi Hicham // Intelligent Systems with Applications. – 2023. – P. 200296. – Mode of access: <https://doi.org/10.1016/j.iswa.2023.200296>
8. Deep hypergraph autoencoder embedding: An efficient intelligent approach for rotating machinery fault diagnosis [Electronic resource] / Mingkuan Shi [et al.] // Knowledge-Based Systems. – 2022. – P. 110172. – Mode of access: <https://doi.org/10.1016/j.knsys.2022.110172>
9. Onyshchenko S. Improving the efficiency of diagnosing errors in computer devices for processing economic data functioning in the class of residuals [Electronic resource] / Svitlana Onyshchenko, Alina Yanko, Alina Hlushko // Eastern-European Journal of Enterprise Technologies. – 2023. – Vol. 5, no. 4 (125). – P. 63–73. – Mode of access: <https://doi.org/10.15587/1729-4061.2023.289185>
10. Lv X.-L. Visual clustering network-based intelligent power lines inspection system [Electronic resource] / Xian-Long Lv, Hsiao-Dong Chiang // Engineering Applications of Artificial Intelligence. – 2024. – Vol. 129. – P. 107572. – Mode of access: <https://doi.org/10.1016/j.engappai.2023.107572>
11. A hybrid CNN-GRU based probabilistic model for load forecasting from individual household to commercial building [Electronic resource] / Ming-Chuan Chiu [et al.] // Energy Reports. – 2023. – Vol. 9. – P. 94–105. – Mode of access: <https://doi.org/10.1016/j.egyr.2023.05.090>
12. Development of Shared Modeling and Simulation Environment for Sustainable e-Learning in the STEM Field [Electronic resource] / Anatolijs Zabasta [et al.] // Sustainability. – 2024. – Vol. 16, no. 5. – P. 2197. – Mode of access: <https://doi.org/10.3390/su16052197>
13. Sikakollu P. Ensemble of multiple CNN classifiers for HSI classification with Superpixel Smoothing [Electronic resource] / Prasanth Sikakollu, Ratnakar Dash // Computers & Geosciences. – 2021. – Vol. 154. – P. 104806. – Mode of access: <https://doi.org/10.1016/j.cageo.2021.104806>
14. Cross-evaluation of a parallel operating SVM – CNN classifier for reliable internal decision-making processes in composite inspection [Electronic resource] / Sebastian Meister [et al.] // Journal of Manufacturing Systems. – 2021. – Vol. 60. – P. 620–639. – Mode of access: <https://doi.org/10.1016/j.jmsy.2021.07.022>
15. Hybridizing genetic random forest and self-attention based CNN-LSTM algorithms for landslide susceptibility mapping in Darjiling and Kurseong, India [Electronic resource] / Armin Moghimi [et al.] // Quaternary Science Advances. – 2024. – P. 100187. – Mode of access: <https://doi.org/10.1016/j.qsa.2024.100187>
16. Turkiye_Earthquake_2023 [Electronic resource] // Kaggle: Your Machine Learning and Data Science Community. – Mode of access: <https://www.kaggle.com/datasets/buraktaci/turkiye-earthquake-2023?resource=download>
17. François Chollet. Deep Learning with Python / François Chollet. – Shelter Island, NY : Manning Publications Co., 2018. – 361 p.
18. Introduction to the Keras Tuner | TensorFlow Core [Electronic resource] // TensorFlow. – Mode of access: https://www.tensorflow.org/tutorials/keras/keras_tuner
19. Netron [Electronic resource] // Netron. – Mode of access: <https://netron.app/>
20. Home [Electronic resource] // OpenCV. – Mode of access: <https://opencv.org/>

Received (Надійшла) 29.05.2024

Accepted for publication (Прийнята до друку) 14.07.2024

Practical use cases for creating convolutional neural network models for image recognition tasks

O. Laktionov, N. Pedchenko, A. Yanko

Abstract. The aim of the article is to improve the efficiency of image recognition by creating a binary classifier for objects of civil infrastructure using deep learning techniques. **Research results.** Convolutional neural network models have been created for the recognition of civil infrastructure objects. A Sequential convolutional neural network has been constructed, consisting of three convolutional layers, pooling layers, a transformation layer, a fully connected layer, and an output layer. The optimal error values during training/testing are 0.0650/0.4424, with accuracies of 0.98/0.92 respectively. Results from the third epoch show errors of 0.2442/0.2595 and an accuracy of 0.93. A pre-trained VGG16 model was also utilized, fine-tuned on the dataset, demonstrating minimal error values of 0.0278/0.1538 during training/testing, with accuracies of 0.99/0.96 respectively. **Scientific novelty.** Further development of using convolutional neural networks for recognizing the level of civil infrastructure damage has been achieved. **Practical significance.** Two convolutional neural network models, Sequential and VGG16, have been built to address the recognition of damaged and intact buildings. The prerequisite for detecting these objects is the use of a camera and appropriate hardware such as a Raspberry Pi or a personal computer/laptop.

Keywords: computer vision, classification tasks, deep learning, neural networks.

О. В. Малєєва, Д. А. Набокін, Ю. А. Малєєва

Національний аерокосмічний університет імені М. С. Жуковського «ХАІ», Харків, Україна

ВИБІР НАЙБІЛЬШ ЕФЕКТИВНОГО МАРШРУТУ ПЕРЕВЕЗЕНЬ З УРАХУВАННЯМ ІНДЕКСІВ НЕБЕЗПЕЧНОСТІ ТЕРИТОРІЙ В УМОВАХ ВОЄННОГО ЧАСУ

Анотація. Предметом дослідження в статті є процеси планування вантажних перевезень в умовах воєнного часу. Мета - підвищення безпеки перевезень за рахунок побудови маршруту, що враховує умови воєнного стану. Використовуються методи системного аналізу, теорії графів та прийняття багатокритеріальних рішень. В статті проаналізовано проблеми вантажних перевезень під час воєнних дій, які зумовили збільшення частки автомобільних перевезень. Було розглянуто причини цього та можливі небезпеки під час перевезень. Визначено основні критерії безпеки маршрутів та пунктів призначення під час воєнних дій. Було створено систему індексів небезпечності, які засновані на даних про кількість артилерійських обстрілів та повідомлень про вибухи. Розроблено алгоритм формування маршруту перевезень з урахуванням вимог безпеки шляхом адаптації алгоритму Дейкстри під критерій безпеки за допомогою розроблених індексів.

Ключові слова: логістика, автомобільні перевезення, маршрут, оптимізація, багатокритеріальність, прийняття рішень, безпека перевезень

Вступ

Постановка проблеми. У сучасному світі логістика є ключовим елементом, який дозволяє товарам та послугам переміщатися від виробників до споживачів ефективно та вчасно. Важливість логістичних рішень можна пояснити кількома якими: ефективність витрат, поліпшення обслуговування, підвищення конкурентоспроможності, скорочення ризиків, стійкість до викликів.

В нинішні часи через військовий стан розподіл торговельних транспортних засобів набув значних змін. З початком війни в Україні сталися такі зміни [1]:

- повітряні перевезення відсутні зовсім через повну заборону польотів цивільної авіації;
- залізничні перевезення скоротилися на 65,3% через обмеження пропускної спроможності на пунктах пропуску;
- для водних перевезень активні лише дунайські порти, що у мирний час складали 5% від усіх подібних перевезень.

Враховуючи ці обмеження, ринок транспортних послуг почав надавати пріоритет автомобільному транспорту для перевезень вантажів.

Планування перевезень у військовий час має враховувати усі можливі небезпечні моменти. Задля безпеки краще обрати шляхи подалі від кордону чи прифронтової зони [2, 3]. Формування логістичних маршрутів під час бойових дій є складним завданням, яке потребує ретельного аналізу та урахування різних факторів. Проте, грамотне планування може допомогти забезпечити безпеку транспортних засобів та товарів, а також мінімізувати ризики.

Тому актуальним є завдання підвищення безпеки автомобільних перевезень за рахунок побудови маршруту, що враховує умови воєнного стану.

Аналіз останніх досліджень і публікацій. Питання логістичних перевезень розглядаються науковцями в різних аспектах.

В роботі [4] обґрунтовано необхідність розроблення кількох варіантів маршрутів, розкрито сут-

ність дорожніх заторів та їх багатьох негативних наслідків, таких як: марнування часу, неможливість передбачити точний час на подорож, марно витрачене паливо, забруднене повітря тощо.

Завдання побудови альтернативних маршрутів перевезень формалізовано в статті [5] з використанням технології ризик-менеджменту та оптимізаційної математичної моделі за критерієм мінімального перевищення експлуатаційних витрат. Для формування множини альтернативних маршрутів на напрямку перевезень використано положення теорії графів. Множина альтернативних маршрутів формується на основі зваженої матриці суміжності графа. В роботах [6, 7] пропонується використання класичного методу маршрутизації Кларка-Райта для формування розвізно-збірних маршрутів.

Розглядається можливість використання моделі нейронної мережі при вирішенні задачі маршрутизації перевезень вантажів та вибору виду транспорту із застосуванням змішаного методу прийняття рішення. Формування маршруту та вибір виду транспорту здійснюється із урахуванням сукупного ефекту від декількох критеріїв вибору та оцінок, вплив яких визначається методом розмитих коефіцієнтів [8]. Запропонована в роботі [9] модель дозволяє проектувати перевізний процес, враховуючи параметри транспортної мережі, учасників транспортно-го процесу, вантажів і транспортних засобів.

Особливості перевезень в умовах воєнного стану розглянуто у роботах [10, 11]. Зазначено, що нові правила перевезення вантажів, пасажирів і багажу внутрішнім транспортом, які охоплюють і правила стосовно страхування пасажирів, були внесені з метою забезпечення безпеки та регулювання перевезень в умовах воєнного стану.

Використання геоінформаційних технологій для формування в реальному часі оптимального та безпечного маршруту доставки гуманітарного вантажу розглянуто у роботах [12–14]. Для визначення критеріїв та вибору оптимального маршруту використовується метод аналізу ієрархій. Враховується довжина маршруту, час у дорозі, марка автомобіля

та інші, для визначення нових маршрутів в умовах воєнного конфлікту [15]. При доставці гуманітарного вантажу враховуються такі аспекти, як руйнування існуючої транспортної інфраструктури та інфраструктури зв'язку, відсутність кваліфікованих працівників та необхідність забезпечення безпеки гуманітарних місій [16].

З проведеного аналізу публікацій можна зробити висновок, що перспективним є застосування теорії графів для формування альтернативних маршрутів та багатокритеріальне оцінювання їх ефективності. Але фактор безпеки територій під час військових дій в існуючих публікаціях недостатньо врахований.

Мета статті: підвищення безпеки автомобільних перевезень за рахунок побудови маршруту, що враховує умови воєнного стану.

Вирішуються наступні завдання:

- аналіз проблеми вантажних перевезень під час воєнних дій;
- визначення основних критеріїв безпеки маршрутів та пунктів призначення під час воєнних дій;
- формування показника небезпечності території;
- розробка алгоритму побудови альтернативних маршрутів перевезень та багатокритеріального вибору найбільш ефективного маршруту в умовах воєнного стану.

Виклад основного матеріалу

В умовах бойових дій маршрути мають бути сплановані з урахуванням наступних факторів:

- безпека: маршрути повинні бути безпечними для пересування транспортних засобів. Вони повинні уникати місць, де можливі бойові дії, мінні поля, засідки та інші небезпеки;
- ефективність: маршрути мають бути ефективними з погляду часу та відстані. Вони мають бути короткими та швидкими, щоб мінімізувати ризики для транспортних засобів та товарів;
- гнучкість: маршрути повинні бути гнучкими та адаптованими. Вони повинні дозволяти змінювати напрямок руху в залежності від обстановки, що змінюється.

Планування маршрутів перевезень вантажів під час воєнного часу передбачає такі основні етапи:

- 1) збір даних - необхідно зібрати дані про місцевість, включаючи карти, дані про рельєф, а також інформацію про можливі небезпеки;
- 2) аналіз даних - необхідно проаналізувати зібрані дані для виявлення можливих маршрутів;
- 3) оцінка маршрутів - необхідно оцінити можливі маршрути з погляду безпеки, ефективності та гнучкості;
- 4) вибір маршруту - потрібно вибрати оптимальний маршрут.

Розглянемо особливості прийняття рішень методами багатокритеріального вибору, які дозволяють оцінити та вибрати оптимальне рішення в умовах, коли є декілька критеріїв, за якими потрібно оцінити альтернативи. Кожен критерій може мати різну вагу чи важливість. Процес постановки багатокритеріальної задачі вимагає взаємодії з експер-

тами у відповідній галузі для розробки адекватної моделі прийняття рішень.

Перелічимо основні кроки постановки багатокритеріального завдання:

- визначення мети прийняття рішення - мета має бути конкретною та вимірною, щоб забезпечити ясне розуміння того, що вважається успішним результатом;
- виділення критеріїв - критерії можуть бути різними параметрами чи характеристиками, що відображають різні аспекти проблеми;
- класифікація критеріїв - поділ критеріїв на категорії, наприклад, якісні та кількісні, або основні та другорядні;
- оцінка значущості критеріїв - це може бути зроблено за допомогою методів анкетування, експертних оцінок або інших технік;
- збір даних необхідних для оцінки значень критеріїв - аналіз статистичних даних, консультації з експертами, або використання результатів попередніх досліджень;
- формалізація задачі - формулювання цільової функції, яку необхідно оптимізувати;
- вибір методу прийняття рішень (розглянути нижче);

- валідація моделі - допомагає переконатися, що модель відображає реальні умови та може бути використана для прийняття обґрунтованих рішень.

Для вирішення завдань багатокритеріального вибору використовують різні методи, включаючи:

- метод зважених критеріїв – у цьому методі кожен критерій набуває певної ваги, яка відображає його важливість. Потім альтернативи оцінюються з урахуванням цих критеріїв, і сума оцінок дає загальну оцінку.
- аналіз ієрархії процесів – розбиває складне завдання на більш прості підзадачі, проводить попарне порівняння критеріїв та альтернатив, і потім обчислює загальну оцінку;
- методи оптимізації – деякі завдання багатокритеріального вибору можна вирішити з використанням математичних методів оптимізації.

Запропоновано для побудови раціонального та безпечного маршруту враховувати два основні критерії – оптимальність за часом та безпечність шляху. Під час воєнного стану питання логістики змінюється з «Як доставити товар швидше» на «Як доставити товар взагалі». Для такого часу питання безпеки стає важливішим за швидкість, бо краще доставити товар трохи довше, аніж не доставити взагалі. З економічної точки зору це може бути більш затратно, але загибель робітників та втрата товару, порівняно із додатковими витратами на перевезення, викличе ще більше проблем. Таким чином, безпечність займає вищу пріоритетність. Зазначимо, що довжина шляху напряму впливає на час доставлення. Відстань відповідає основним прямим маршрутам між містами. Критерії мають враховуватись одночасно для кожного проміжного шляху, щоб зберегти баланс між безпекою і часом.

Для формування альтернативних маршрутів використовується теорія графів. На графі окрім

N вершин та K ребер існують ще M фіктивних вершин, тобто загальна кількість дуг обчислюється як

$$L = (N - M)M + M(M - 1). \quad (1)$$

Від кожної реальної вершини до фіктивної існує дуга з вагою, що прагне до нескінченності. Таким чином, існує два зв'язаних графа – дійсний, де вага дорівнює відстані між містами, та фіктивний, що можна назвати «графом можливої доставки».

Припускається, що множина вершин N – це обласні центри України, множина ребер K – основні дороги між ними, множина фіктивних вершин M – міста-склади. На рис. 1 можна побачити приклад фіктивних дуг з міста Дніпро, що проведені до складів, які були обрані випадково.

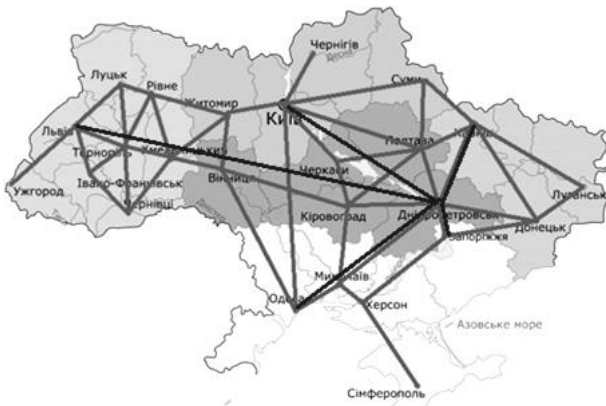


Рис. 1. Граф маршрутів та фіктивні дуги з міста Дніпро

На рисунку чорним кольором зображено фіктивні дуги. Як можна побачити, існують приклади, коли реальне ребро співпадає з дугою, але у більшості випадків дуга є лише віртуальним поєднанням двох різних вершин.

Для побудови маршруту було обрано алгоритм Дейкстри, бо він може побудувати шлях між будь-якими двома точками на карті і, на відміну від алгоритму Беллмана-Форда, не працює з від'ємною вагою ребер. При застосуванні алгоритму Дейкстри запропоновано окрім довжини шляху враховувати індекс його небезпечності.

Перед формуванням маршруту необхідно визначити індекси небезпечності для кожного міста, який може бути проміжним пунктом маршруту. Запропоновано опиратися на три фактори: кількість повідомлень про вибухи, кількість артилерійських обстрілів та чи є місто тимчасово окупованим.

Станом на 08.02.2024 ми маємо ситуацію, що зображена на рис. 2 та 3 [17].

Індекси небезпечності розраховуються як коефіцієнти, які приймають значення у діапазоні від нуля до одиниці, де 0 – безпечно, 1 – небезпечно (табл. 1).

Після аналізу даних було вирішено, що кращим варіантом для розподілення об'єктів по групам стане використання закономірності геометричної прогресії. Були спроби побудувати математичну модель розподілу областей більш точно за кількістю обстрілів, але виявилось, що через занадто велику різницю між

показниками різних областей, деякі індекси виходили неправдоподібно малими.

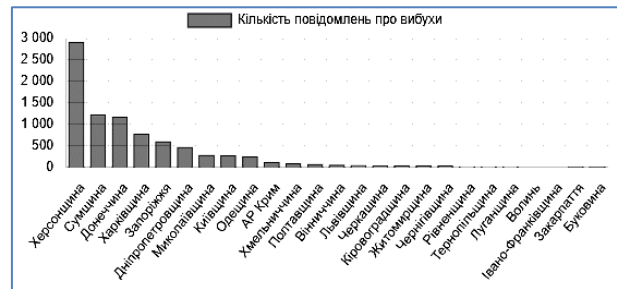


Рис. 2. Кількість повідомлень про вибухи по областях

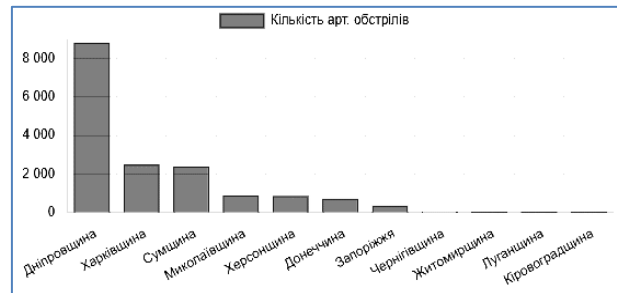


Рис. 3. Кількість артилерійських обстрілів по областях

Таблиця 1 - Дані про небезпеку по областях

Назва міста	Кількість повідомлень про вибухи	Кількість артилерійських обстрілів	Чи є окупованим
Вінниця	64	0	Ні
Дніпро	476	8 825	Ні
Донецьк	1 185	722	Так
Житомир	46	0	Ні
Запоріжжя	613	358	Ні
Івано-Франківськ	0	0	Ні
Київ	290	0	Ні
Кропивницький	47	0	Ні
Луганськ	10	0	Так
Луцьк	0	0	Ні
Львів	60	0	Ні
Миколаїв	296	891	Ні
Одеса	262	0	Ні
Полтава	84	0	Ні
Рівне	18	0	Ні
Сімферополь	128	0	Так
Суми	1233	2390	Ні
Тернопіль	18	0	Ні
Ужгород	0	0	Ні
Харків	788	2498	Ні
Херсон	2921	848	Ні
Хмельницький	105	0	Ні
Черкаси	48	0	Ні
Чернівці	0	0	Ні
Чернігів	39	0	Ні

Тому в даній роботі використовувався евристичний алгоритм. У перспективі планується для розподілу областей за індексом небезпечності використовувати статистичні методи (квантілі розподілу) та кластерний аналіз.

Розподіл обласних центрів за індексами небезпечності відображено у табл. 2.

Таблиця 2 - Розподіл міст за індексами небезпечності

Індекс небезпечності	Шкала розподілу по групах	Кількість небезпечних подій	Назва міста
0.1	<1	0	Івано-Франківськ
		0	Луцьк
		0	Ужгород
		0	Чернівці
0.3	1-10	-	-
0.5	11-100	64	Вінниця
		46	Житомир
		47	Кропивницький
		60	Львів
		84	Полтава
		18	Рівне
		18	Тернопіль
		48	Черкаси
		39	Чернігів
0.7	101-1000	971	Запоріжжя
		290	Київ
		262	Одеса
		105	Хмельницький
0.9	1001-10000	9301	Дніпро
		1187	Миколаїв
		3623	Суми
		3286	Харків
		3769	Херсон
1	Тимчасово окуповане місто	-	Донецьк Луганськ Сімферополь

Карта на рис. 4 відображає розподіл територій за індексами небезпечності за допомогою відтінків сірого кольору. Области з найбільшим індексом, тобто окуповані, відображено темно-сірим. Із зменшенням індексом небезпечності колір наближається до світлого.



Рис. 4. Розподіл областей України за зонами небезпечності

Оскільки індекс небезпечності враховується як критерій вибору шляху, то вага ребер графа береться у залежності від міста, що має більший індекс небезпечності. Якщо дорога йде до міста з індексом безпечності 0.1, то індекс такої дороги знижується на 0.2.

Таким чином, на відміну від класичного алгоритму, де відстані між містами додаються, у даній

модифікації у якості ваг використовуються індекси безпечності кожної дороги. Шлях обирається за критерієм мінімальності взваженої відстані:

$$F_{\min} = \sum_{i=1}^n S_i Ind_i, \quad (2)$$

де S_i – довжина ребра графа, Ind_i – максимальний з індексів небезпечності початкової або кінцевої вершини ребра.

Запропоновано два режими роботи алгоритму: «Безпека», що рекомендовано використовувати під час воєнного стану, та «Швидкість», що більше підходить для мирного часу. Варто зазначити, що режим «Швидкість» працює тільки з урахуванням критерію довжини шляху. У режимі «Безпека» враховується обидва критерії: довжина шляху та його безпечність. Подібний розрахунок дозволяє зберегти баланс та знайти шлях, що сумістить у собі і безпечність, і оптимальність за часом.

Опишемо етапи алгоритму формування раціонального маршруту.

1. Обрання режиму – «Безпека» чи «Швидкість». Будуються фіктивні дуги між кожним містом.

2. Залежно від вибору режиму кожній дузі присвоюється власна вага.

3. Формування маршруту. Режим «Швидкість» використовує алгоритм Дейкстри у його класичному вигляді – шукає найкоротший шлях до складу, що задовольнить потреби перевізника. Режим «Безпека» використовує модифікацію алгоритму та замість найкоротших шляхів обирає добуток найнижчих індексів небезпечності та довжини шляху для побудовання кінцевого результату.

4. Обчислення параметрів шляху: очікуваний час, ціна доставлення, рівень безпеки (у режимі «Безпека»).

5. Побудова отриманого шляху на карті та формування результату у текстовому вигляді користувачеві.

Запропонований алгоритм дозволяє будувати шлях доставки вантажів з урахуванням двох критеріїв на основі модифікації алгоритму Дейкстри, що враховує індекси небезпечності.

Висновки

В статті проаналізовано проблеми вантажних перевезень під час воєнних дій. Вказані проблеми змусили представників бізнесу зробити акцент на автомобільний транспорт. Було розглянуто можливі небезпеки під час перевезень, визначено основні критерії безпеки маршрутів та пунктів призначення під час воєнних дій.

Було запропоновано евристичний метод визначення індексів небезпечності територій, які засновані на даних про кількість артилерійських обстрілів та повідомлень про вибухи; розподілено їх між областями України та проаналізовано результати цього розподілу.

Розроблено алгоритм формування маршруту перевезень з урахуванням вимог безпеки.

Напрямок подальших досліджень є розробка реального часу на основі попередніх статистичних методу формування безпечних маршрутів у режимі даних та з урахуванням поточних геоданих.

СПИСОК ЛІТЕРАТУРИ

1. Як логістика адаптувалася до війни [Електронний ресурс] - Режим доступу: <https://www.epravda.com.ua/columns/2023/07/24/702529/>
2. Бедрій Я., Тарнавський Є. Військова логістика. - Запоріжжя: Видавничий дім "Гельветика", 2020. – 242 с. ISBN 978-966-289-100-3
3. Федорович О. Є., Уруський О. С., Чепков І. Б., Луханін М. І., Прончаков Ю. Л., Рибка К. О., Лещенко Ю. О. Моделювання транспортної логістики військових вантажів з урахуванням збитків, які виникають у зоні бойових дій через запізнення у постачанні // Радіоелектронні і комп'ютерні системи. – 2022. – № 2. – С. 63–74. DOI: <https://doi.org/10.32620/reks.2022.2.05>.
4. Гурч Л. М. Маршрутизація перевезень з використанням новітніх технологій // Вісник Національного університету "Львівська політехніка". Логістика, 2016. - № 846. - С. 48-53. - URL: http://nbuv.gov.ua/UJRN/VNULPL_2016_846_11
5. Бутко Т. В., Пархоменко Л. О., Тарасов К. О. Формалізація процедури надання альтернативного маршруту швидкісним пасажирським поїздам на основі ризик-менеджменту // Інформаційно-керуючі системи на залізничному транспорті. – 2023. – Т. 28. – №. 1. – С. 31-37. DOI: 10.18664/ikszt.v28i1.276341
6. Гужевська Л. А., Даниленко І. В. Формування розвізно-збірних маршрутів методом Кларка-Райта при доставці експрес-відправлень // Вісник Національного транспортного університету. – 2016. – №. 1. – С. 122-129. Режим доступу: http://nbuv.gov.ua/UJRN/Vntu_2016_1_17
7. Pichipul T., Kawtummachai R. A heuristic approach based on clarke-wright algorithm for open vehicle routing problem // The Scientific World Journal. – 2013. – Vol. 2013. DOI:10.1155/2013/874349
8. Шарай С. М., Сахно В. П., Поляков В. М., Рой М. П., Фадєєв М. С. Використання моделі нейронної мережі для формування маршруту перевезень вантажів та вибору виду транспорту // Матеріали XI-ої Міжнародної науково-технічної інтернет-конференції «Проблеми та перспективи розвитку автомобільного транспорту» (13-14 квітня 2023 року, м. Вінниця). – 2023. – С. 354-357. Режим доступу: <https://atmconf.vntu.edu.ua/materialy2023.pdf>
9. Куш Є. І. Розробка алгоритму формування розвізних маршрутів в логістичній системі // Вісник Східноукраїнського національного університету імені Володимира Даля. – 2017. – №. 4. – С. 128-133. Режим доступу: http://nbuv.gov.ua/UJRN/VSUNU_2017_4_27
10. Коробкова О. М. Логістична діяльність підприємств в умовах воєнного стану // The 4th International scientific and practical conference “Actual problems of modern science”(January 31–February 3, 2023), Boston, USA. International Science Group. – 2023. – С. 457-458. Режим доступу: https://web.archive.org/web/20230202220402id_/https://isg-konf.com/wp-content/uploads/2023/01/Actual-problems-of-modern-science.pdf#page=458
11. Кваша Ю. Аналіз послуг вантажних перевезень в умовах воєнного стану // Актуальні питання права та соціально-економічних відносин. Збірник наукових статей. – 2023. – С. 387-393. Режим доступу: <http://dspace.wunu.edu.ua/bitstream/316497/48146/1/Збірник-2023%20%281%29.pdf>
12. Beresford A., Pettit S. Humanitarian aid logistics: a Cardiff University research perspective on cases, structures and prospects // Journal of Humanitarian Logistics and Supply Chain Management. – 2021. – No. 11. – P. 623–638. DOI: 10.1108/JHLSCM-06-2021-0052
13. Giaglis G. M., Minis I., Tatarakis A., Zeimpekis V. Minimizing logistics risk through real-time vehicle routing and mobile technologies : Research to date and future trends // International Journal of Physical Distribution & Logistics Management. – 2004. – Vol. 34. – P. 749–764. DOI: 10.1108/09600030410567504
14. Szucs G. Decision support for route search and optimum finding in transport networks under uncertainty // Journal of Applied Research and Technology. – 2015. – Vol. 13. – P. 125–134. DOI: 10.1016/S1665-6423(15)30011-0
15. Помазков М. В., Мнацаканян М. С., Маник І. С. Моделювання інтелектуальних маршрутів перевезення гуманітарних вантажів в умовах воєнного часу // Подільський вісник: сільське господарство, техніка, економіка. – 2023. – №. 41. – С. 46-53. DOI: 10.37406/2706-9052-2023-4.7
16. Malhouni Y., Mabrouki C. C. Mitigating risks and overcoming logistics challenges in humanitarian deployment to conflict zones: evidence from the DRC and CAR // Journal of Humanitarian Logistics and Supply Chain Management. – 2023 DOI 10.1108/JHLSCM-04-2023-0031
17. Статистика повітряних тривог [Електронний ресурс] - Режим доступу: <https://air-alarms.in.ua>

Received (Надійшла) 01.04.2024

Accepted for publication (Прийнята до друку) 26.06.2024

Selection of the most effective transportation route considering territory safety indexes in wartime conditions

O. Malyeyeva, D. Nabokin, Yu. Malieieva

Abstract. The subject of research in the article is the processes of planning freight transportation in wartime conditions. The purpose of research is to increase the safety of transportation due to the construction of a route that takes into account the conditions of martial law. Methods of system analysis, graph theory and multi-criteria decision-making are used. The article analyzes the problems of cargo transportation during hostilities, which led to an increase in the share of automobile transportation. The reasons for this and possible dangers during transportation were considered. The main criteria for the security of routes and destinations during hostilities have been determined. A system of safety indexes was created, which was based on data on the number of artillery attacks and information about explosions. The algorithm for forming the transportation route with the security measures has been expanded by adapting the Dijkstra's algorithm for the security criterion with the help of fragmented indices.

Keywords: logistics, road transport, route, optimization, multicriteria, decision-making, transport safety.

Viacheslav Radchenko¹, Nina Kuchuk^{1,2}

¹ Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

² National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine

CLOUD INFRASTRUCTURE SERVICE DELIVERY SPECIFICS

Abstract. The article discusses infrastructure as a service. Cloud infrastructure includes at least basic computing resources, data storage, and network resources. Infrastructure as a service provides benefits that allow various systems to develop faster and achieve their digital transformation goals. Such deployment options for cloud systems as private cloud, public cloud, and hybrid cloud are explored. Each type of cloud has its own disadvantages and advantages. Infrastructure as a service provides four main benefits that allow various systems to develop faster and achieve their digital transformation goals. The next generation cloud is designed to run any application faster and more securely at a lower cost.

Keywords: cloud infrastructure, cloud computing, hybrid cloud.

Introduction

Infrastructure as a Service (IaaS) is a cloud computing service model where computing resources are hosted in a public cloud, private cloud, or hybrid cloud. With the IaaS model, it is possible to partially or completely move an on-premises or distributed data center infrastructure to the cloud, where it is maintained and managed by a cloud provider [1–4].

Such infrastructure elements may include computing, networking, storage hardware, and other components and software [5, 6]. A standard IaaS model uses various services, such as computing, storage, and databases, provided by a cloud provider. The service provider delivers these services by hosting hardware and software in the cloud.

The company does not need to purchase its own hardware, administer it, or allocate space for it in its data centers [7–10]. And it bears the costs on a pay-as-you-go basis. If the company needs fewer resources, the overall cost for them decreases. And as a company grows, it can provide employees with additional computing resources and other technologies in minutes.

Basic material

Cloud infrastructure includes, at a minimum, the core compute, storage, and network resources. More recently, it has also come to include higher-level services (sometimes called platform as a service [PaaS]), such as relational and NoSQL databases, real-time and bulk processing, developer pipelines and services, containers, and tooling. Unlike software as a service (SaaS), IaaS is not intended for the average end user.

IaaS is intended for [11]:

- application developers;
- IT operators;
- DevOps teams;
- systems and database administrators;
- general developers.

The cloud improves operational efficiency, quickly deploys IT services, and transforms workflows. Cloud deployments have proven useful in everything from financial reporting and supply chains to simple mobile applications. Each of the three common cloud computing models has its own unique characteristics that help solve business problems in a specific way [12–14].

The line between IaaS and PaaS is blurring because cloud solutions often offer both models in the same environment. IaaS helps you spend more efficiently on hardware and devices, while PaaS allows you to further reduce administrative costs and expand your use cases with less hardware.

Infrastructure as a Service provides four key benefits that enable various systems to evolve faster and achieve their digital transformation goals.

Cloud infrastructure reduces the time and cost of provisioning and scaling the environments needed to develop, test, and use off-the-shelf systems. It allows you to use computing services on demand, while IaaS allows you to scale your infrastructure as needed. IaaS allows the use of new, improved equipment and various services [15, 16].

The IaaS model is available in most regions of the world, and the data centers themselves are located in densely populated areas.

Compared to on-premises systems, IaaS provides higher uptime, redundancy implemented at every tier, improved security and disaster recovery options, and scalability that on-premises environments cannot.

IaaS is gaining popularity across all industries, and its user base is growing.

The cloud infrastructure user base generally consists of IT systems operators, application developers, DevOps teams, systems and database administrators, and general developers [17].

IaaS provides numerous advantages over traditional on-premises data centers. Table 1 summarizes the capabilities that IaaS provides.

Discussion and Results

The next generation cloud designed to run any application faster and more securely at a lower cost.

The OCI distributed cloud provides freedom to choose the method and way to allocate cloud services, ensuring their compliance with regulatory requirements, performance, and other needs.

There are several options for deploying cloud systems (Fig. 1):

1. **Private cloud.** Used to provide services within a single company, which is both a customer and a service provider. A private cloud is an IT model that involves placing all virtual resources on a physical server.

Table 1 – IaaS Features and Benefits

Development and testing	With IaaS, you can create and tear down test and development environments quickly and at a low cost, reducing the time to market for new applications.
Traditional applications	IaaS supports both cloud and native applications, including ERP and business intelligence applications.
Hosting websites and web applications	IaaS also supports web and mobile applications that can be quickly deployed and scaled.
Storage, backup and recovery	Storing and backing up data on-premises, as well as planning and executing disaster recovery, requires significant time and expertise. Moving infrastructure to the cloud helps reduce costs and free up time for other tasks.
High Performance Computing	With a pay-as-you-go model, IaaS makes high-performance computing and other data-intensive project-oriented workloads more accessible.

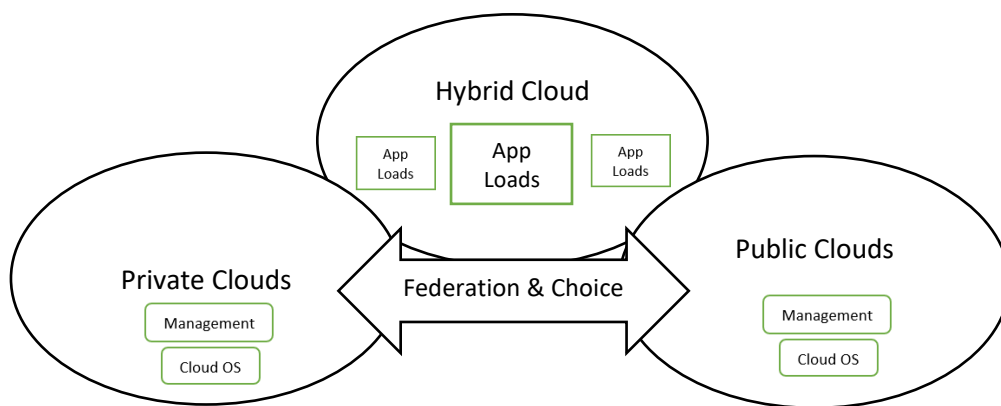


Fig. 1. Cloud Systems

2. **Public cloud.** Used by cloud providers to provide services to external customers. A public cloud is the most accessible and popular way to store data. When using this model, company information is stored in a large provider cloud along with data from other companies, but it is isolated and protected from outside access.

3. **Mixed (hybrid) cloud** is a symbiosis of the first two models, i.e. some IT resources of the company are located in a private cloud, others in a public one. This approach, especially with the use of hyperconvergent systems, provides both cost reduction due to the use of the public model and a high level of security in private.

When the demand for computing and data processing starts to exceed the capacity of the on-premises data center resources, it becomes necessary to use the cloud to quickly scale performance. Hybrid cloud infrastructure, which includes multiple public clouds from multiple cloud service providers, combines best-in-class cloud services offered by multiple cloud service providers.

The main objective of hybrid cloud and cloud integration is to facilitate the creation of new systems of participation in the cloud with their on-premises systems in a secure and reliable manner. Cloud integration helps to protect data in their old systems by securely integrating cloud applications and data with on-premises ones.

Hybrid cloud supports the widest range of cloud strategies. Hybrid cloud platforms provide the flexibility to choose infrastructure based on their workload needs, compliance, and latency requirements. A hybrid cloud consists of one or more private clouds. A critical component of a hybrid cloud that significantly improves its usability is a cloud resource management platform that allows you to use all connected clouds, both private and public, in a single interface. Without such a platform, each private and public cloud must be managed separately, which significantly increases the administration costs. At the same time, the best platforms have the functionality to issue recommendations to improve the efficiency of cloud environments by 30-50% by analyzing cloud use cases.

Conclusions

Cloud computing provides advanced computing resources that are available on-demand, scale on demand, and receive regular updates without the need to purchase and maintain on-premises infrastructure. Each type of cloud - public, private, hybrid - has its own disadvantages and advantages.

Significant benefits of a hybrid cloud are also highlighted, such as managing growing data volumes and reducing storage costs by 60%, optimizing storage and data management, and better recovery from

failures. One of the key trends in the future development of cloud services is the widespread use of artificial intelligence and machine learning. Cloud services will work with these technologies to create intuitive and automated solutions that will help companies perform their tasks faster and more efficiently. Another area of

cloud service development will be the use of IoT (Internet of Things) and smart devices. Cloud services will provide the infrastructure for processing and analyzing data received from smart devices. This will help make decisions faster and more accurately based on data received from multiple sources.

REFERENCES

1. Zhen Xiao, Weija Song, and Qu Chen, "Dynamic Resource Allocation using Virtual Machines for Cloud Computing Environment", IEEE transaction on parallel and distributed systems, Vol/ 24, Is. 6, pp. 1107 – 1117, June 2013, doi: 10.1109/TPDS.2012.283.v/
2. John J. Prevost, Kranthi Manoj Nagothu, Brian Kelley, and Mo Jamshidi, "Prediction of cloud data center networks loads using stochastic and neural models", 2011 6th International Conference on System of Systems Engineering, 12138393, 2011, doi: 10.1109/SYSOSE.2011.5966610.
3. S. Kianpisheh, and R. H. Glitho, "Cost-efficient server provisioning for deadline-constrained VNFs Chains: A parallel VNF processing approach", Proceeding of 2019 16th IEEE Annual Consumer Communications & Networking Conference, 2019, doi: 10.1109/CCNC.2019.8651799.
4. N. Kuchuk, O. Shefer, G. Cherneva, and F. A. Alnaeri, "Determining the capacity of the self-healing network segment", Advanced Information Systems, vol. 5, no. 2, pp. 114–119, Jun. 2021, doi: 10.20998/2522-9052.2021.2.16.
5. Ye. Qiang, and W. Zhuang, "Distributed and adaptive medium access control for internet-of-things-enabled mobile networks", IEEE Internet of Things Journal, 2017, vol. 4, no. 2, pp. 446–460, doi: 10.1109/JIOT.2016.2566659.
6. G. Kuchuk, S. Nechausov, and V. Kharchenko, "Two-stage optimization of resource allocation for hybrid cloud data store", International Conference on Information and Digital Technologies, Zilina, 2015, pp. 266–271, doi: 10.1109/DT.2015.7222982.
7. H. Khudov, K. Tahyan, V. Chepurnyi, I. Khizhnyak, K. Romanenko, A. Nevodnichii, and O. Yakovenko, "Optimization of joint search and detection of objects in technical surveillance systems", Advanced Information Systems, 2020, Vol. 4, No. 2, pp. 156–162, doi: 10.20998/2522-9052.2020.2.23.
8. S. Semenov, O. Sira, S. Gavrylenko, and N. Kuchuk, "Identification of the state of an object under conditions of fuzzy input data", Eastern-European Journal of Enterprise Technologies, Vol 1, No 4 (97), pp. 22–30, 2019, doi: 10.15587/1729-4061.2019.157085.
9. S. Semenov, and Cao Weilin, "Testing process for penetration into computer systems mathematical model modification", Advanced Information Systems, Vol. 4, No. 3, pp. 133–138. 2020, doi: 10.20998/2522-9052.2020.3.19.
10. H. Attar, M.R. Khosravi, S.S. Igorovich, K.N. Georgievan, and M. Alhihi, "E-health communication system with multiservice data traffic evaluation based on a G/G/1 analysis method", Current Signal Transduction Therapy, 16(2), 2021, doi: 10.2174/1574362415666200224094706.
11. P. Franti, "Efficiency of random swap clustering", Journal of Big Data, vol. 5, no. 13, 2018, pp. 1–29, doi: 10.1186/s40537-018-0122-y.
12. G. Kuchuk, A. Kovalenko, I.E. Komari, A. Svyrydov, and V. Kharchenko, "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, pp. 161–183, 2019, doi: 10.1007/978-3-030-00253-4_8.
13. A. Nechausov, I. Mamusuç, and N. Kuchuk, "Synthesis of the air pollution level control system on the basis of hyperconvergent infrastructures", Advanced Information Systems, vol. 1, no. 2, 2017, pp. 21–26. DOI: 10.20998/2522-9052.2017.2.04.
14. A. Kovalenko, H. Kuchuk, N. Kuchuk, and J. Kostolny, "Horizontal scaling method for a hyperconverged network", International Conference on Information and Digital Technologies 2021, IDT 2021, pp. 331–336, 9497534, 2021, doi: https://doi.org/10.1109/IDT52577.2021.9497534.
15. H. Attar, M.R. Khosravi, S.S. Igorovich, K.N. Georgievan, and M. Alhihi, "Review and performance evaluation of FIFO, PQ, CQ, FQ, and WFQ algorithms in multimedia wireless sensor networks", International Journal of Distributed Sensor Networks, 16(6), June 2020, doi: https://doi.org/10.1177/1550147720913233.
16. H. Kuchuk, A. Kovalenko, B.F. Ibrahim, and I. Ruban, "Adaptive compression method for video information", International Journal of Advanced Trends in Computer Science and Engineering, 8(1), pp. 66–69, 2019, doi: http://dx.doi.org/10.30534/ijatcse/2019/1181.22019.
17. N. Kuchuk, O. Mozhaiev, S. Semenov, A. Haichenko, H. Kuchuk, S. Tiulieniev, M. Mozhaiev, V. Davydov, O. Brusakova, and Y. Gnusov, "Devising a method for balancing the load on a territorially distributed foggy environment", Eastern-European Journal of Enterprise Technologies, vol. 1(4 (121)), pp. 48–55, 2023, doi: https://doi.org/10.15587/1729-4061.2023.274177

Received (Надійшла) 19.05.2024

Accepted for publication (Прийнята до друку) 21.07.2024

Особливості надання хмарних інфраструктурних послуг

В. О. Радченко, Н. Г. Кучук

Анотація. У статті розглядається інфраструктура як послуга. Хмарна інфраструктура включає принаймні базові обчислювальні ресурси, сховище даних і мережні ресурси. Інфраструктура як послуга надає переваги, які дозволяють різноманітним системам розвиватися швидше та досягати цілей цифрової трансформації. Досліджено такі варіанти розгортання хмарних систем, як приватна хмара, публічна хмара та гібридна хмара. Кожен тип хмари має свої недоліки та переваги. Інфраструктура як послуга забезпечує чотири основні переваги, які дозволяють різним системам розвиватися швидше та досягати цілей цифрової трансформації. Хмара наступного покоління розроблена для того, щоб запускати будь-яку програму швидше та безпечніше за меншу вартість.

Ключові слова: хмарна інфраструктура, хмарні обчислення, гібридна хмара.

D. Tyshchenko, T. Franchuk, R. Zakharov, V. Moskalenko

State University of Trade and Economics, Kiev, Ukraine

SUPPORTING DYNAMIC SECURITY NEEDS WITH VPN TOOLS

Abstract. The article examines the problems of supporting the dynamic security needs of a hybrid environment. It was determined that an important aspect of remote work is the training of employees on common cyber security threats. Threat types such as phishing attacks, social engineering, and malware are explored. The components of information technology security in a hybrid working environment are defined. The levels of protection of corporate information were studied. VPN features are analyzed. Security procedures for using VPN, which creates opportunities to reduce the risks of hacking a remote work environment, have been analyzed. Requirements to reduce the risks of unauthorized access to confidential information have been established. Disadvantages and limitations of VPN usage are explored. The criteria for reducing the connection speed due to additional cryptographic processing and the distance to the VPN server were analyzed. Methods and methods of blocking and limiting access to VPNs are considered, with the aim of reducing their effectiveness in some contexts. The features of a set of protocols that protect Internet Protocol (IP) communication by authenticating and encrypting all IP packets in the data stream are investigated. The principles of the VPN service are defined. Ways to prevent data misuse when working in a hybrid environment are proposed. Requirements for invulnerability of network traffic to attacks are defined. A set of protocols is explored, thanks to which it is possible to authenticate and check the integrity and encrypt IP packets. Analyzed protocols for secure key exchange on the Internet. The design of key information management protocols is proposed, requiring the development of the Internet Key Management Protocol. In the process of research, key management concepts were analyzed and improved using the ISAKMP specifications and the Oakley Key Determination Protocol. The criteria for the successful functioning of the ISAKMP specification, which describes the mechanisms for matching the attributes of the protocols used, have been established.

Keywords: hybrid working environment, cyber security, digital transformation, information security.

Introduction

The most important aspect is understanding cyber threats and the means to prevent them. The method of cyberattack prevention includes the use of VPN for a secure connection to the company's network, the installation of anti-virus software and information security monitoring systems.

The analysis of research results and the development of own means of ensuring the security of information technologies in a hybrid working environment aims not only to protect users from potential cyberattacks, but also to prepare them for the development of software products in the field of information security in the future.

Thus, cyber security management becomes a critically important aspect for basic scientific research and requires thorough analysis, particularly when working in today's digital environment.

The object of research is information technologies and their security in the context of hybrid work. The subject of the study is the analysis of the security of network connections, cloud services, virtual private networks (VPN), as well as the integration of various technological solutions used in remote work. In accordance with the stated object and subject of the research, the purpose of the work is to assess the existing security measures in the context of hybrid work, identify weak points and risks.

Analysis of recent research and publications. Examining the results of the work of modern researchers analyzing the presented issues, it should be noted that not all aspects of this topic have already been covered in scientific publications, which makes it

possible to study this issue more thoroughly. Scientists take care of the challenges of supporting the dynamic security needs of a hybrid environment, studying such types of threats as phishing attacks, social engineering and malware. Based on the received data, we offer a methodology for protecting corporate information and security procedures for VPN use.

A number of sources testify to the analysis of the long-term strategy of IT security in the conditions of remote work. In particular, O. Struuk analyzes a number of approaches and means of verifying the logic of the program developed by the user, in particular, he examines the logic of the program developed using the platform customization tool [1].

I. Arshad proposes a new framework for intelligent cyber defense, namely, studying immersion in deep learning attacks and defenses [2].

V. Pevnev considers methods and means of ensuring data integrity in information communication systems [3].

M. Ahola analyzes objective and subjective factors and errors in cyber security violations [4].

M. Rose describes the peculiarities of phishing, noting the best methods and methods of protection against it [5].

The study of the leading points and features of the implementation of VPN services on mobile devices deserves special attention, the results of research on their purpose, configuration, use, as well as the analysis of the best VPN services in Ukraine and the world are presented in numerous publications [6, 8, 15].

A. Kapiton, A. Mostova, V. Baranova, R. Baranenko, H. Sokol, M. Okhrymenko, T. Franchuk

explore the communication possibilities of digital channels, pay attention to features of adaptive combined coding of the channel network for cognitive radio networks with cooperative relay [7, 11-14].

Main part

By studying the evolving telecommuting environment and understanding the reasons for its popularity, organizations can make an informed decision about implementing a telecommuting policy. Despite the undeniable benefits, remote work also creates unique cybersecurity challenges that organizations must address to protect sensitive data and systems. Research of real examples of cyber security violations in hybrid working environments allows to analyze and improve the methods of their detection and neutralization.

In particular, the SolarWinds Supply Chain Attack, known since December 2020, came into the field of our research. Attacks block access to files or systems, demanding a ransom.

That is why methods of detecting and neutralizing attacks need to be improved. Detection of attacks can be difficult due to their cunning and distributed nature, but regular data backups, protection of systems against viruses, and improved cyber hygiene help prevent or recover from such attacks[1-3].

This will ensure the highest level of protection for corporate information, as VPN encrypts all data transmitted between devices, including on open networks, which significantly reduces the risk of unauthorized access to confidential information.

While exploring potential drawbacks and limitations.

In particular, some VPN providers may collect log files of data about some activity that violates user privacy, some countries or organizations may block or restrict access to VPNs, making them less effective in some contexts, free or low-cost VPN providers may use public servers, which may be less secure or have limited bandwidth.

Given the growing number of remote workers and the associated cyber security threats, using a VPN is becoming a necessary standard for any organization looking to keep their data private and secure. Along with the right security policies and procedures, using a VPN can significantly reduce the risks of a remote work environment being compromised. Passwords remain a weak point in cyber security.

By choosing a certain criterion, it is possible to classify groups in a certain way VPN.

Thus, according to the method of technical implementation, the following VPN groups are distinguished, which are presented on Fig. 1.

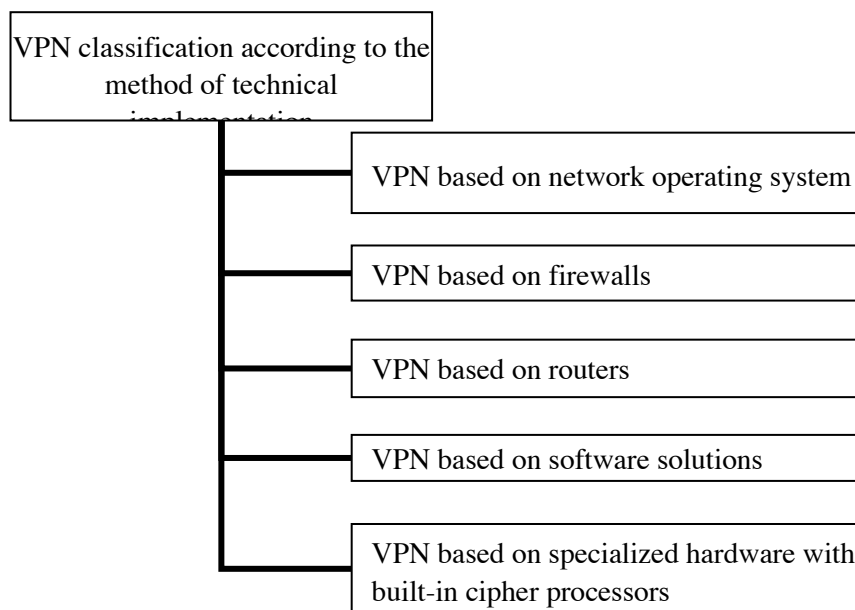


Fig. 1. VPN classification according to the method of technical implementation

In addition, an important aspect is the remote management of security policies, which allows you to set and manage security rules on remote devices, for example, by restricting access or installing mandatory software updates. Additionally, mobile device security, such as disk encryption and remote data wipe capabilities, are important aspects of protecting remote endpoints.

Organizations should establish secure endpoint configurations that include active firewalls, use of up-to-date anti-virus software, and implementation of

strong strong password policies. Educating remote employees about the importance of following these security measures on their devices is a key element of effective protection.

This helps create a consistent security system across all endpoints, providing a high level of protection against potential threats [6-8].

The user, in this case a QA engineer, is given an encryption key that must be connected to a PC for identification and access. This way of using VPN networks is used in almost all companies that have

divisions in different geographical points or remote employees, including in a hybrid work environment. Analysis of access to blocked content deserves special attention. Sometimes the desired content is available only for a certain region or vice versa - it is allowed for all but one.

This problem is solved with the help of a VPN - the user's traffic is redirected through servers located in the country that has access to the necessary resources, while bypassing all regional blocks. The service in question is also used for testing applications from another country.

If the customer needs to test the operation of the application from different parts of the world, for example, due to the specifics of the application, then in order not to hire a large number of testers from different countries, it is enough to hire one team that will test using VPN networks.

In this case, you can connect to a remote server that will receive and transmit all device traffic.

Thus, the tester will receive the external IP address of this server and will be considered a user of this country.

Inter-corporate VPN networks provide company employees with a secure exchange of information with business partners, suppliers, wholesale customers, customers, users, etc.

An inter-corporate network provides direct access from one network company to the network of another, thus contributing to the increase in the reliability of the communication that is maintained in business cooperation.

In inter-corporate networks, much attention is paid to user authentication and access control using a network screen.

By understanding the tactics used by cybercriminals, employees can become more vigilant and proactive in protecting themselves and the organization's sensitive data. In addition, it is important to keep employees constantly trained and updated on the latest trends in cybersecurity, as these threats are constantly evolving. Organizations can also use phishing attack simulations to test their employees' reactions to real threats and improve their skills in detecting and preventing phishing. Training and supporting employees in the field of cyber security are important elements of an effective strategy to protect the organization from phishing attacks. This will allow to increase the awareness and preparedness of the personnel to detect and avert potential threats, which can solve the problem of phishing attacks to a significant level.

A VPN basically creates a secure data tunnel between your local computer and another VPN server located thousands of kilometers away. When using the Internet, this VPN server becomes the source of all data, so the ISP and other third parties can no longer see the content of your Internet traffic.

At the same time, the result of the negotiation of the security context is the establishment of the security parameter index (SPI), which is a pointer to a certain element of the internal structure of the information

exchange party, which describes possible sets of security parameters. IPSec is a component of IPv6, operating at the third or network layer.

As a result, transmitted IP packets are protected in a manner that is transparent to network applications and infrastructure. Unlike SSL (Secure Socket Layer), which works on the fourth (transport) layer and is already associated with higher levels of the OSI model, IPSec is designed to provide low-level protection. An IPSec header is added to IP data ready for VPN transmission to identify protected packets.

Before transmission, these packets are encapsulated into other IP packets. IPSec supports several types of encryption, including Data Encryption Standard (DES) and Message Digest 5 (MD5). A necessary condition for establishing a secure connection is the ability to quickly agree on security parameters (authentication algorithms and keys).

IPSec supports two types of key management schemes by which participants can negotiate session parameters. This dual support at one time caused friction in the IETF Working Group.

Conclusions

Using traffic encryption, regularly updating software and network devices, and implementing monitoring and breach detection mechanisms will help prevent many types of cyberattacks.

In general, only a comprehensive approach to cyber security, which includes staff training, the use of secure technologies and the use of reliable protocols, can ensure the creation of a secure remote work environment that will effectively protect the organization's valuable assets.

Only this approach will ensure productive and safe work in a remote format.

A VPN is a good solution for establishing security and maintaining anonymity online, and it deserves the attention of anyone who uses the Internet on mobile devices or PCs. VPN networks provide a wide range of possibilities.

They are used not only by IT professionals, but also by ordinary users for various reasons.

In the process of using the possibility of limited authorization, in addition to the specified protection methods, additional special systems should be used to strengthen protection, taking into account the encryption tools used in the hybrid working environment of the corporation by means of VPN services.

Note that it is the logical network that is designed and created based on the use of the most widely used architecture model, which enables employees of the corporation and other users to work with the database during work.

When organizing work, in such a case, network administrators may implement additional special procedures to strengthen security.

Using the capabilities of the VPN service does not require additional work and does not pose a particular danger for testers.

It was determined that the design of key information management protocols is necessary, namely

the development of IKMP, an application-level key management protocol.

The emerging key information management standards provide an opportunity to support Key Distribution Centers. Data integrity and confidentiality

guarantees in the IPsec specification are provided through the use of authentication and encryption mechanisms, respectively, where the result of security context negotiation is the establishment of the SPI security parameter index.

REFERENCES

1. Striuk O. User-Designed Application Logic Verification Approach and Tools URL: <https://radics.tech/user-designed-application-logic-verification-approach-and-tools/>
2. Arshad I. A Novel Framework for Smart Cyber Defence: A Deep-Dive Into Deep Learning Attacks and Defences. *IEEE Access*. 2023. № 11. 88527-88548.
3. Pevnev V. Ensuring the Data Integrity in Infocommunication Systems *International Journal of Computing*. 2022. 228-233. URL: <https://doi.org/10.47839/ijc.21.2.2591>.
4. Ahola M. The role of human error in successful cyber security breaches. *Usecure Blog*. URL: <https://blog.usecure.io/the-role-of-human-error-in-successful-cyber-security-breaches>.
5. Rouse M. What is email phishing? Definition, examples and best practices URL: <https://searchsecurity.techtarget.com/definition/phishing>
6. VPN on mobile devices: purposes, setup, usage URL: <https://training.qatestlab.com/blog/technical-articles/vpn-mobile-device-setting/>
7. Kapiton A., Kryvoruchko O., Tyshenko D., Franchuk T., Tsiutsiura M. Modern website creation technologies Комерціалізація інновацій в умовах Індустрії 4.0. Суми, СДУ, 2023. 145-153.
8. The best VPN URL: <https://ua.cybernews.com/lp/best-vpn-ua/>
9. Найкращі служби VPN в Україні URL: <https://protonvpn.com/>
10. Шифр-VPN URL: <https://cipher.com.ua/uk/products/cipher-vpn>
11. Mostova A., Kapiton A., Baranova V. Digital transformation of business in Ukraine: current trends, challenges and prospects. *Transformation of the economic system in the context of information technology challenges*. Riga, Latvia: Baltija Publishing, 2024. 70-86.
12. Kapiton A. Communication opportunities of digital channels. *Сектор безпеки і оборони України на захисті національних інтересів: актуальні проблеми та завдання в умовах воєнного стану*. Хмельницький, НАДПСУ, 2024. 953-954.
13. Skakalina O., Kapiton A. Design and software implementation of a knowledge exchange web service. порівняльний аналіз застосування евристичних алгоритмів для розв'язання задачі TSP. *Системи управління, навігації та зв'язку*, 2024. 2 (76). 144-151.
14. Kapiton A., Baranenko R., Sokol H., Okhrymenko M., Franchuk T. Адаптивне комбіноване кодування мережі каналів для когнітивних радіомереж з кооперативною ретрансляцією. *Електронне моделювання*, 2024. 1. 78-89.
15. Bitdefender Premium VPN URL: <https://www.bitdefender.ro/solutions/vpn.html>

Received (Надійшла) 22.05.2024

Accepted for publication (Прийнята до друку) 14.08.2024

Підтримка динамічних потреб безпеки засобами VPN

Д. Тищенко, Т. Франчук, Р. Захаров, В. Москаленко

Анотація. У статті досліджено проблеми підтримки динамічних потреб безпеки гібридного середовища. Визначено, що важливим аспектом віддаленої роботи є навчання співробітників загальним загрозам кібербезпеки. Досліджено типи загроз, такі як фішингові атаки, соціальна інженерія та зловмисне програмне забезпечення. Визначено компоненти безпеки інформаційних технологій в гібридному робочому середовищі. Досліджено рівні захисту корпоративної інформації. Проаналізовано особливості VPN. Проаналізовано процедури безпеки використання VPN, що створює можливості зменшення ризиків зламу віддаленого робочого середовища. Встановлено вимоги до зниження ризиків несанкціонованого доступу до конфіденційної інформації. Вивчено недоліки та обмеження використання VPN. Проаналізовано критерії зменшення швидкості з'єднання через додаткову криптографічну обробку та відстань до сервера VPN. Розглянуто методи та способи блокування та обмеження доступу до VPN, з метою зменшення їх ефективності в деяких контекстах. Досліджено особливості набору протоколів, які захищають зв'язок Інтернет-протоколу (IP) шляхом автентифікації та шифрування всіх IP-пакетів у потоці даних. Визначено принципи роботи служби VPN. Запропоновано шляхи запобігання зловживанню даними, при роботі у гібридному середовищі. Визначено вимоги до невразливості мережевого трафіку до атак. досліджено набір протоколів, завдяки яким є можливість автентифікувати та перевіряти цілісність та шифрувати IP-пакети. Проаналізовано протоколи для безпечного обміну ключами в Інтернеті. Запропоновано проектування протоколів керування ключовою інформацією, що вимагають розробки Internet Key Management Protocol. В процесі дослідження проаналізовано та вдосконалено концепції управління ключами з використанням специфікацій ISAKMP і протоколу Oakley Key Determination Protocol. Встановлено критерії успішного функціонування специфікації ISAKMP, що описує механізми узгодження атрибутів використовуваних протоколів.

Ключові слова: гібридне робоче середовище, кібербезпека, цифрова трансформація, інформаційна безпека.

О. С. Ткаченко¹, Є. В. Мелешко¹, В. В. Міхав²

¹ Центральноукраїнський національний технічний університет, Кропивницький, Україна

² ПУ «Університет науки, підприємництва та технологій», Київ, Україна

КОМП'ЮТЕРНА МОДЕЛЬ ПОШИРЕННЯ ІНФОРМАЦІЙНИХ ВІРУСІВ У СОЦІАЛЬНІЙ МЕРЕЖІ ПРИ РІЗНІЙ ПОВЕДІНЦІ КОРИСТУВАЧІВ

Анотація. У наш час моделювання процесів поширення інформаційних вірусів є важливою задачею кібербезпеки, адже треба чітко розрізняти де правда, а де фейк, вміти виявляти джерело поширення фейкових новин і протидіяти дезінформації, щоб доносити до людей істину. Метою даної роботи було створення та дослідження комп'ютерної моделі поширення інформаційних вірусів у соціальній мережі при різній поведінці користувачів. Для виконання поставленої мети було використано епідеміологічну модель SIRS та моделі генерації структури соціальної мережі Барабаши-Альберт та Воттса-Строгаца. Модель SIRS ідеально підходить для імітації поширення комп'ютерного вірусу, адже в цій моделі людина може приймати циклічно три стани: здоровий, хворий та з імунітетом, а потім знов здоровий, а в соціальній мережі користувач заражується фейком замість вірусу і також проходить через ці стани. Для моделювання структури соціальної мережі було використано алгоритми Барабаши-Альберт і Воттса-Строгаца, які доступні в бібліотеці Networkx мови програмування Python. Запропоновано декілька різних способів поведінки користувачів для захисту від інформаційних вірусів, зокрема, видалення зв'язків між користувачами, видалення користувачів з мережі та блокування користувачів при їх підозрілій активності. Здійснено емпіричне дослідження та порівняння запропонованих методів боротьби з інформаційним вірусом за різними критеріями. Було запропоновано початкові параметри мережі, а саме, кількість користувачів, кількість зв'язків між ними та коефіцієнти моделі SIRS. З використанням мови програмування Python та бібліотек Pygame і Networkx було реалізовано запропоновану модель поширення інформаційного вірусу в соціальній мережі та змодельовано такі методи боротьби з фейком як: видалення зв'язків між користувачами, створення нового зв'язку, видалення користувачів, блокування користувачів. Найкращий результат боротьби з інформаційним вірусом ми отримуємо при комбінуванні методів видалення зв'язків і користувачів, а також блокування користувачів. При запропонованій поведінці користувачів інформаційному вірусу вдалося вдало протидіяти й виявити поширювача фейку та видалити його, при цьому кількість зв'язків між користувачами соціальної мережі зменшилась не дуже суттєво.

Ключові слова: соціальна мережа, інформаційна безпека, фейкові новини, інформаційні віруси, модель SIRS, модель Барабаши-Альберт, модель Воттса-Строгаца.

Вступ

Швидкий розвиток інформаційних технологій в останні роки суттєво вплинув на трансформацію соціальних структур, що призвело до появи інформаційного суспільства та цифрової економіки. Становлення глобальних медіа, розвиток комп'ютерних технологій, зростання Інтернету та соціальних мереж – ці та інші аспекти докорінно змінили механізми виробництва та поширення інформації, значно розширивши можливості для маніпулювання індивідуальною та суспільною свідомістю.

З одного боку, соціальні мережі дуже корисний та зручний інструмент для спілкування та бізнесу, з іншого боку, вони можуть використовуватися для поширення фейкових новин, ведення інформаційних воєн та впливу на політичні процеси [1]. Також соціальні мережі використовують складні алгоритми для персоналізації контенту, зокрема, рекомендаційні системи [2], що може призводити до створення камер відлуння [3, 4] та бульбашок фільтрів [4, 5], де користувачі стикаються лише з інформацією, що відповідає їхнім переконанням. Це сприяє утворенню ілюзії, що їхні погляди є загальноприйнятими та правдивими. В соціальних мережах інформацію зручно подавати таким чином, щоб викликати сильні емоції та використовувати суб'єктивні судження для впливу на громадську думку [6].

У наші часи розуміння унікальної епідеміології фейкових новин може мати таку ж вагу, як і вміння критичного мислення та фактчекінгу [7]. Постаць-

ники дезінформації шукають найбільш вразливих та цінних «жертв» – тих, хто найбільше поширить їх інформаційні віруси. Вже не секрет, що для реклами продукту формується група людей або залучається популярна людина, щоб коментувати чи поширювати інформацію про цей об'єкт з найкращої сторони, рекомендуючи його через Twitter, Facebook або інші соціальні мережі. Схожий принцип використовується для поширення фейків та впливу на свідомість людей.

Таким чином середовище соцмереж схильне до епідемій фейкових новин [7]. Протидія фейкам вже стала питанням національної безпеки, як і протидія епідеміям біологічних вірусів. Бізнес підірваних новин приносить видавцям значні доходи від реклами та політичний вплив. Фейкові новини здатні впливати на соціальну реальність та змінювати її. Соцмережі стали ідеальною платформою для обміну контентом й поширення фейків. Тому питання захисту від інформаційних вірусів є дуже важливим для кібербезпеки у наш часі.

Метою роботи є розробка комп'ютерної моделі поширення інформаційних вірусів у соціальній мережі при різній поведінці користувачів. Для досягнення поставленої мети було використано теорію складних мереж [8–10] та моделі поширення вірусів [11].

Дослідження моделей соціальних мереж та моделей поширення вірусів

У якості моделей для генерації структури соціальної мережі були взяті моделі Барабаши-Альберт та Воттса-Строгаца, а для поширення вірусу – епі-

деміологічна модель SIRS. Розглянемо ці моделі та модель SIR, на якій заснована SIRS.

Модель Барабаши-Альберт є однією з ключових моделей для пояснення виникнення та еволюції складних мереж. Ця модель базується на двох основних концепціях: «прив'язка» та «безшкальне структурування» [9]. Концепція прив'язки вказує на те, що нові вузли у мережі більш схильні приєднуватися до вузлів, які вже мають багато зв'язків, ніж до менш зв'язаних вузлів. Це призводить до того, що деякі вузли набувають значно більше зв'язків, ніж інші, що створює безшкальну структуру мережі.

Алгоритм побудови моделі Барабаши-Альберт наступний:

Крок 1. Створюється початковий граф з декількох вузлів.

Крок 2. Додається новий вузол до графа.

Крок 3. Вибирається випадковий існуючий вузол у графі з ймовірністю, яка залежить від кількості зв'язків цього вузла.

Крок 4. Новий вузол з'єднується з обраним вузлом.

Крок 5. Повторюються кроки 2-4, доки не буде створено достатньо вузлів у графі.

Модель Воттса-Строгаца є ще однією важливою моделлю для вивчення складних мереж. Ця модель базується на концепції «малого світу» та «згрупованості» [10]. Концепція «малого світу» вказує на те, що в багатьох реальних мережах відстань між будь-якими двома вузлами виявляється дуже короткою, незважаючи на велику кількість вузлів у мережі. Концепція «згрупованості» показує, що в реальних мережах часто спостерігається явище, коли вузли мають тенденцію групуватися разом у підграфі.

Алгоритм побудови моделі Воттса-Строгаца наступний:

Крок 1. Створюємо граф у вигляді початкової решітки з N вузлів, де кожен вузол початково з'єднаний з k найближчими сусідами.

Крок 2. Для кожного вузла i , розглядаємо кожне з його з'єднання (i, j) і видаляємо з ймовірністю p .

Крок 3. З ймовірністю p , для кожного вузла i , замість видаленого зв'язку (i, j) , додаємо новий зв'язок (i, t) , де t обирається випадково з усіх інших вузлів, за винятком себе і своїх поточних сусідів, щоб уникнути повторень. Повторюємо цей крок доки не буде створено достатньо зв'язків у мережі.

Модель SIR – загальна епідеміологічна модель, яка забезпечує спрощений спосіб опису передачі інфекційного захворювання через лю-

дей, де вони можуть проходити через наступні 3 стани: сприйнятливий – S , заразні – I та одужали – R [11]. Загальна популяція: $N = S + I + R$. У замкнутій популяції без життєвої динаміки епідемія в кінцевому підсумку згасне через недостатню кількість сприйнятливих осіб для підтримки захворювання. Інфіковані особи, додані пізніше, не почнуть нової епідемії через довічний імунітет наявної популяції. На рис. 1, а представлено графік зростання інфекції та її виснаження під час спалаху за моделлю SIR.

Модель SIRS. Попередня модель SIR припускає, що люди мають довічний імунітет до хвороби після одужання і це стосується різноманітних захворювань. Але для іншого класу повітряно-крапельних захворювань, наприклад сезонного грипу, імунітет людини може з часом ослабнути. У цьому випадку використовується модель SIRS [11], яка дозволяє моделювати той факт, що люди, які одужали від вірусу, можуть повернутися до сприйнятливого стану.

На рис. 1, б показано коливання через те, що люди втрачають імунітет і знову стають сприйнят-

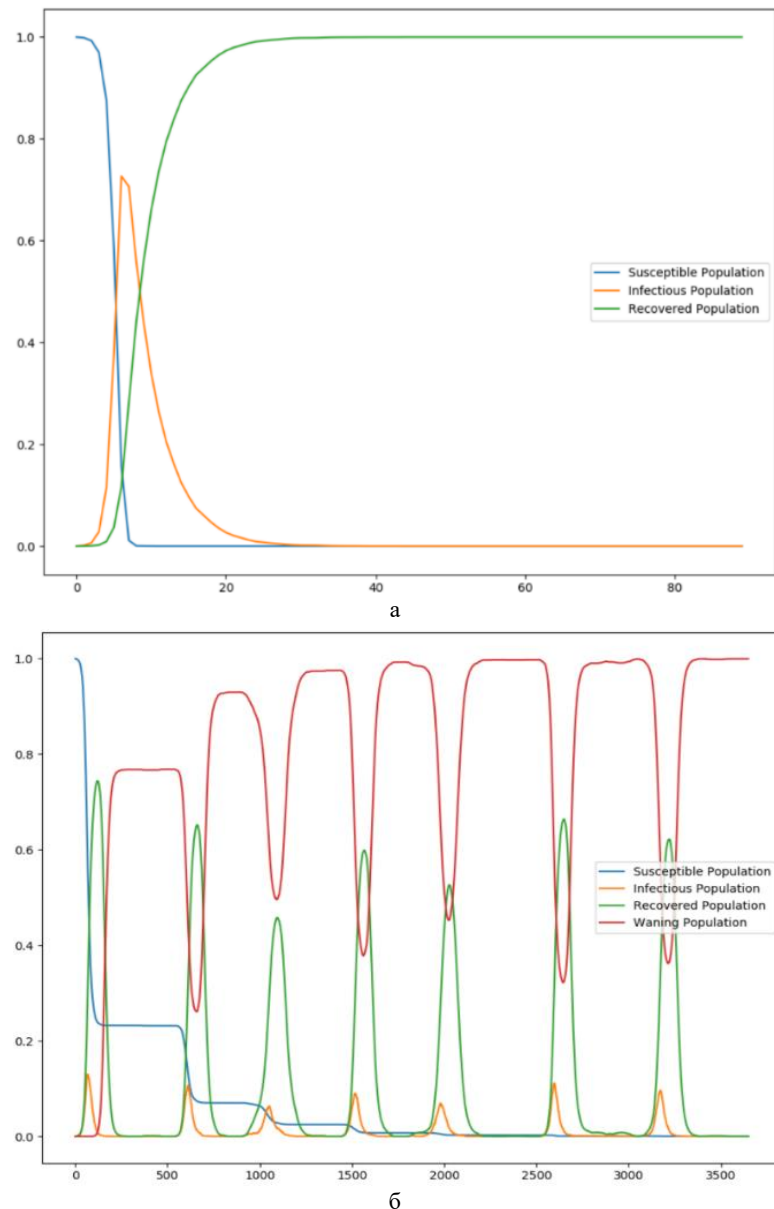


Рис. 1. Графіки SIR (а) та SIRS з «Waning Population» (б) [11]

ливими (особи, сприйнятливі через ослаблений імунітет, не класифікуються як сприйнятливі на графіку, а позначені як “Waning Population”, а ось після повної втрати імунітету особи знову стають “Susceptible”).

Наведена нижче схема SIR/SIRS моделей (рис. 2) показує, як люди змінюють свої стани під час поширення вірусу.

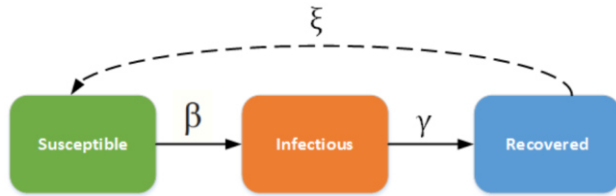


Рис. 2. Моделі SIR та SIRS [11]

Пунктирна лінія показує, як модель SIR стає моделлю SIRS (Susceptible – Infectious – Recovered –

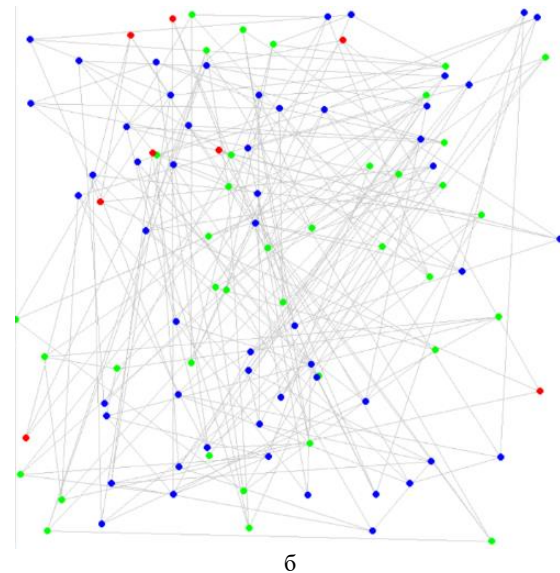
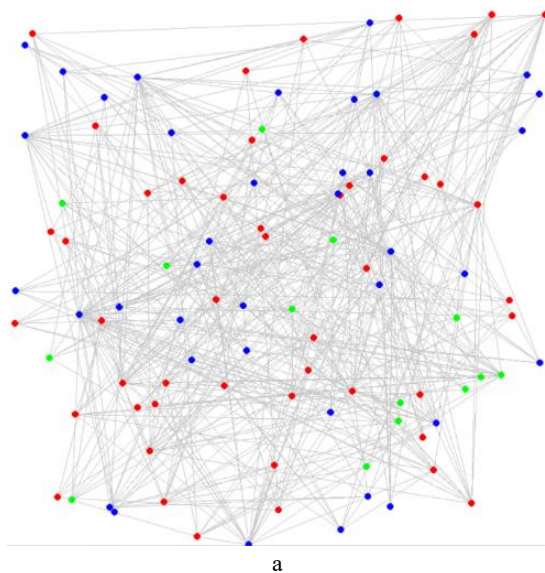


Рис. 3. Візуалізація запропонованої моделі поширення інформаційних вірусів у соціальній мережі: а) на основі моделей Барабаші-Альберт та SIRS, б) на основі моделей Воттса-Строгаца та SIRS

У моделі SIRS є такі показники: β – це показник того, на скільки користувач мережі сприятливий для прийняття фейку як правдивої інформації; γ – цей показник означає, наскільки швидко користувач перестав вірити цьому фейку; ξ – це швидкість, з якою користувачі, які перестали вірити фейку, знову починають вірити цій інформації під впливом інших користувачів.

В дослідженні розглянуті наступні можливі значення показників моделі SIRS: β – від 0.3 до 0.7, так як цілком захищених від фейків користувачів немає, але і всій інформації повністю користувач не довіряє. Вирішено обрати середнє значення $\beta = 0.5$. Показник γ – від 0.1 до 0.5, адже зазвичай, якщо людина в щось вірить, її важко переконати у іншому. Для моделювання було обрано середнє: $\gamma = 0.3$. Показник ξ – від 0.01 до 0.05, він найменший, тому що, якщо людина перестала вірити фейку, то ймовірність того, що вона знову у нього повірить дуже мала. Обрано середнє – $\xi = 0.03$.

Susceptible), де одужання не надає довічного імунітету, і люди можуть знову стати сприйнятливими.

Рівень зараження, β , контролює швидкість поширення, яка представляє ймовірність передачі захворювання між сприйнятливою та інфікованою особою. Швидкість одужання, $\gamma = 1/D$, визначається середньою тривалістю інфекції D . У моделі SIRS ξ – швидкість, з якою особи, що одужали, повертаються до чутливого стану з втратою імунітету.

Запропонована модель поширення інформаційних вірусів у соціальній мережі та її емпіричне дослідження

На основі вище наведеного дослідження було розроблено модель соціальної мережі, в якій поширюється інформаційний вірус у вигляді фейку.

Засобами мови програмування Python та бібліотеки Networkx було реалізовано розробку та візуалізацію розробленої моделі (рис. 3).

Було проведено серію експериментів з розробленою моделлю та різним комбінаціями її параметрів, результати представлені у табл. 1-2.

Ймовірність p для моделі Воттса-Строгаца обрана стала – 0.1.

Генерувалися мережі зі 100 користувачів, один з яких поширює фейки. Програма виконується певний час, на вибір користувача (обрано 10 секунд).

За допомогою запропонованої моделі ми порівнювали такі критерії: кількість сприятливих до фейку користувачів; кількість заражених (інфікованих) фейком; кількість користувачів, які перестали вірити фейку (одужали); кількість зв'язків між користувачами в мережі, які залишились, після боротьби з фейковими новинами такими способами як: видалення зв'язків між користувачами, створення нового зв'язку, якщо у користувача жодного не залишилось; видалення користувачів з мережі; блокування користувачів; поєднання різних комбінацій із запропонованих способів.

Таблиця 1 – Результати експериментів для розробленої моделі на основі моделей Барабаші-Альберт та SIRS, початкові параметри: кількість вузлів графа – 100, кількість ребер – 475, $\beta = 0.5$, $\gamma = 0.3$, $\xi = 0.03$

	Сприятливі	Інфіковані	Одужали	К-ть зв'язків	Сприятливі	Інфіковані	Одужали	К-ть зв'язків
Вид	Режим №1. Базовий режим моделювання				Режим №4. Наявні додаткові дії: видалення зв'язків і створення нового, якщо зв'язків не залишилось			
1	17	3	80	475	49	1	50	426
2	16	7	77	475	98	1	1	464
3	12	11	77	475	96	1	3	459
Сер.	15	7	78	475	81	1	18	449.67
Вид	Режим №2. Наявні додаткові дії: видалення зв'язків і вузлів (користувачів)				Режим №5. Наявні додаткові дії: блокування вузлів (користувачів)			
1	87	0	12	439	11	10	79	475
2	32	7	61	427	16	11	73	475
3	99	0	0	467	20	7	73	475
Сер.	72.67	2.33	24.33	444.33	15.67	9.33	75	475
Вид	Режим №3. Наявні додаткові дії: видалення зв'язків і вузлів, блокування вузлів (користувачів)				Режим №6. Наявні додаткові дії: видалення зв'язків, створення нового, якщо зв'язків не залишилось, і блокування			
1	92	0	7	423	70	1	29	421
2	98	1	1	470	22	8	70	421
3	99	0	0	461	84	1	15	436
Сер.	96.33	0.33	2.67	451.33	58.67	3.33	38	426

Таблиця 2 – Результати експериментів для розробленої моделі на основі моделей Воттса-Строгаца та SIRS, початкові параметри: кількість вузлів графа – 100, кількість ребер – 200, $\beta = 0.5$, $\gamma = 0.3$, $\xi = 0.03$

	Сприятливі	Інфіковані	Одужали	К-ть зв'язків	Сприятливі	Інфіковані	Одужали	К-ть зв'язків
Вид	Режим №1. Базовий режим моделювання				Режим №4. Наявні додаткові дії: видалення зв'язків і створення нового, якщо зв'язків не залишилось			
1	30	3	67	200	95	1	4	195
2	46	1	53	200	40	1	59	188
3	39	1	60	200	71	1	28	174
Сер.	38.33	1.67	60	200	68.67	1	30.33	185.67
Вид	Режим №2. Наявні додаткові дії: видалення зв'язків і вузлів (користувачів)				Режим №5. Наявні додаткові дії: блокування вузлів (користувачів)			
1	99	0	0	196	45	4	51	200
2	99	0	0	196	37	3	60	200
3	99	0	0	194	64	16	20	200
Сер.	99	0	0	195.33	48.67	7.67	43.67	200
Вид	Режим №3. Наявні додаткові дії: видалення зв'язків і вузлів, блокування вузлів (користувачів)				Режим №6. Наявні додаткові дії: видалення зв'язків, створення нового, якщо зв'язків не залишилось, і блокування			
1	99	0	0	196	96	1	3	194
2	99	0	0	196	98	1	1	194
3	99	0	0	196	98	1	1	197
Сер.	99	0	0	196	97.33	1	1.67	195

Під час базового режиму моделювання фейк поширюється мережею без будь-яких обмежень.

Видалення зв'язків – кожен вузол (користувач) в графі (мережі) має декількох сусідів, з якими утворює зв'язок, а так як фейк поширюється від одного користувача до іншого, то видалення між ними зв'язку означає, що користувач перестав вірити своєму сусіду, від якого часто отримував фейки. У дослідженні ми видаляли зв'язок, коли користувач тричі отримав фейк від сусіда.

Створення нового зв'язку – якщо користувач розірвав зв'язки з усіма своїми сусідами, то щоб не видаляти його з мережі, було вирішено додати можливість створення нового зв'язку для нього, за умови, що цей користувач зараз не має ніяких зв'язків.

Видалення вузлів – якщо зв'язків не залишилось, то користувач видаляється (можна обрати лише або

цей варіант, або попередній), тобто він повністю втратив довіру до своїх сусідів, або сусіди до нього.

Блокування вузлів – користувач, який повірив фейку деяку кількість разів (в дослідженні – 3 рази), блокується на певний час (в дослідженні – 0.0001 с) для того, щоб його сусіди на деякий час перестали з ним взаємодіяти, через те, що він часто «хворіє».

Висновки

Було розроблено та реалізовано комп'ютерну модель поширення інформаційних вірусів у соціальній мережі з використанням моделей Барабаші-Альберт та Воттса-Строгаца для генерації структури мережі та епідеміологічної моделі SIRS для моделювання поширення вірусної інформації. Запропоновано різні способи моделювання поведінки користувачів у соціальній мережі при поширенні інфор-

маційного вірусу, та різні способи боротьби з ним. Було проведено експерименти з використанням різних запропонованих способів моделювання поведінки користувачів у соціальній мережі. За результатами дослідження можна зробити такі висновки:

– чим більша кількість зв'язків у соціальній мережі, тим легше інформаційному вірусу поширюватися нею;

– інформаційний вірус поширюється хвилями у соціальній мережі;

– найкращий результат боротьби з інформаційним вірусом у розробленій моделі отримано при комбінуванні таких стратегій захисту користувачів,

як видалення зв'язків і вузлів та блокування вузлів при їх підозрілій активності, адже за час роботи моделі у такому режимі вдалося виявити поширювача фейку та видалити його, при цьому кількість зв'язків між користувачами зменшилась не дуже сильно;

– при запропонованій поведінці користувачів (комбінація видалення зв'язків і вузлів та блокування вузлів при їх підозрілій активності) інформаційному вірусу вдається успішно протидіяти;

Запропоновані способи боротьби проти поширення інформаційного вірусу працюють, і їх можна застосовувати і в реальних соціальних мережах, якщо в них є необхідні умови та механізми.

СПИСОК ЛІТЕРАТУРИ

1. Курбан О. В. Сучасні інформаційні війни в соціальних онлайн-мережах. *Інформаційне суспільство*. 2016. Вип. 23. С. 85-90. URL: http://nbuv.gov.ua/UJRN/is_2016_23_15
2. Мелешко Є. В. Проблеми сучасних рекомендаційних систем та методи їх рішення. *Системи управління, навігації та зв'язку*. 2018. Вип. 4. С. 120-124. URL: http://nbuv.gov.ua/UJRN/suntz_2018_4_25
3. Cinelli M., De Francisci Morales G., Galeazzi A., Quattrocioni W., Starnini M. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*. 2021. Vol. 118, No. 9, e2023301118. DOI: <https://doi.org/10.1073/pnas.2023301118>
4. Rhodes S. C. Filter bubbles, echo chambers, and fake news: how social media conditions individuals to be less critical of political misinformation. *Political communication*. 2022. Vol. 39, No. 1. P. 1-22. DOI: <https://doi.org/10.1080/10584609.2021.1910887>
5. Chitra U., Musco C. Analyzing the Impact of Filter Bubbles on Social Network Polarization. In *Proceedings of the 13th International Conference on Web Search and Data Mining (WSDM '20)*, Association for Computing Machinery, New York, NY, USA. 2020. P.115-123. URL: <https://doi.org/10.1145/3336191.3371825>
6. Yerlikaya T., Aslan, S. T. Social Media and Fake News in the Post-Truth Era: The Manipulation of Politics in the Election Process. *Insight Turkey*. 2020. Vol. 22, No. 2. P. 177-96. URL: <https://www.jstor.org/stable/26918129>
7. Динаміка поширення фейків в соціальному просторі. *Портал Медіаосвіти і Медіаграмотності*. 2020. URL: <https://medialiteracy.org.ua/dynamika-poshyrennya-fejkiv-v-sotsialnomu-prostorii/>
8. Newman M. E. J. *Networks: An Introduction* (1st edn). Oxford University Press. 2010. DOI: <https://doi.org/10.1093/acprof:oso/9780199206650.001.0001>
9. Barabási A.-L., Albert R. Emergence of scaling in random networks. *Science*. 1999. Vol. 286(5439), P. 509-512. DOI: <https://doi.org/10.1126/science.286.5439.509>
10. Watts D. J., Strogatz S. H. Collective dynamics of 'small-world' networks. *Nature*. 1998. Vol. 393(6684), P. 440-442. DOI: <https://doi.org/10.1038/30918>
11. SIR and SIRS models. *IDM documentation*. URL: https://docs.idmod.org/projects/emod-generic/en/2.20_a/model-sir.html#sirs-model

Received (Надійшла) 29.05.2024

Accepted for publication (Прийнята до друку) 31.07.2024

A computer model of information virus propagation in a social network with different user behavior

O. Tkachenko, Ye. Meleshko, V. Mikhav

Abstract. Nowadays, modeling the processes of the spread of information viruses is an important task of cybersecurity, because it is necessary to clearly distinguish where the truth is from where it is fake, to be able to identify the source of the spread of fake news and to counter disinformation in order to convey the truth to people. The purpose of this work was to create and research a computer model of information virus propagation in a social network with different user behavior. The SIRS epidemiological model and the Barabási-Albert and Watts-Strogatz social network structure generation models were used to achieve this goal. The SIRS model is ideal for simulating the spread of a computer virus, because in this model a person can cycle through three states: susceptible, infected, and recovered with immunity, analogous to how a user in a social network can be "infected" to and be "cured" of a fake. Barabási-Albert and Watts-Strogatz algorithms, which are available in the Networkx library of the Python programming language, were used to model the structure of the social network. Several different ways of user behavior have been proposed to protect against information viruses, including removing connections between users, removing users from the network, and blocking users for suspicious activity. An empirical research and comparison of the proposed methods of combating the information virus was carried out according to various criteria. The initial parameters of the network were proposed, namely, the number of users, the number of connections between them, and the coefficients of the SIRS model. Using the Python programming language and the Pygame and Networkx libraries, the proposed model of the spread of an information virus in a social network was implemented and such methods of combating fakes as: deleting connections between users, creating a new connection, deleting users, and blocking users – were simulated. We get the best result in the fight against the information virus when we combine the methods of deleting connections and users, as well as blocking users. With the proposed user behavior, the information virus managed to successfully counteract and detect the spreader of the fake and remove it, while the number of connections between users of the social network decreased not very significantly.

Keywords: social network, information security, fake news, information viruses, SIRS model, Barabási-Albert model, Watts-Strogatz model.

Е. Г. Фастовський¹, А. І. Роговий¹, О. Б. Ахієзер¹, А. В. Фролов², Р. В. Артюх^{2,3}

¹ Національний технічний університет "Харківський політехнічний інститут", Харків, Україна

² Харківський національний університет радіоелектроніки, Харків, Україна

³ ДП "Південний державний ПК і НД інститут авіаційної промисловості", Харків, Україна

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ АНАЛІЗУ ТА СИНТЕЗУ ПОЯСНЕНИХ МОДЕЛЕЙ ШТУЧНОГО ІНТЕЛЕКТУ НА ОСНОВІ ВЕРБАЛЬНИХ МЕТОДІВ

Анотація. Предметом дослідження є аналіз та синтез пояснених моделей штучного інтелекту. Мета роботи – розробка інформаційної технології аналізу та синтезу пояснених моделей штучного інтелекту на основі вербальних методів. У статті вирішуються такі завдання: аналіз математичних формул та методів, що використовуються для пояснення рішень, що приймають моделі штучного інтелекту, аналіз методів, класів, фреймворків та функцій програмних бібліотек, а також їх використання для пояснення рішень, що приймають моделі штучного інтелекту, синтез пояснених вербальних моделей штучного інтелекту, розробка проекту системи синтезу пояснених вербальних моделей штучного інтелекту. Використовуються такі методи: системний аналіз, вербальні методи прийняття рішень (формування системи понять визначеної предметної галузі, формування порядкової класифікації станів об'єкта/процесу, впорядкування станів об'єкта/процесу з певного класу, визначення найкращого стану об'єкта/процесу), методи моделювання та проєктування інформаційних систем (діаграми варіантів використання, діаграми діяльності). Здобуто такі результати: проведено аналіз математичних формул та методів, що використовуються для пояснення рішень, що приймають моделі штучного інтелекту. Запропоновано підхід до синтезу пояснених вербальних моделей штучного інтелекту. Проведено аналіз методів, класів, фреймворків та функцій програмних бібліотек, а також їх використання для пояснення рішень, що приймають моделі штучного інтелекту. Розроблено проєкт системи синтезу пояснених вербальних моделей штучного інтелекту. **Висновки:** методи вербального аналізу виявляються ефективними для синтезу пояснених моделей штучного інтелекту, що включає кілька етапів: визначення системи понять, створення критеріальних описів станів, їх класифікація, впорядкування та обрання найкращого стану. Вони підкреслюють важливість використання лінгвістичної інформації разом з числовими даними для комплексного аналізу складних проблем. Інтегруючи елементи вербального аналізу в пояснені моделі штучного інтелекту, можна покращити взаємодію з користувачем, його розуміння і сприйняття систем штучного інтелекту.

Ключові слова: пояснені моделі штучного інтелекту; вербальні методи прийняття рішень.

Вступ

Розробленим моделям штучного інтелекту (ШІ) сьогодні бракує можливості пояснення у багатьох важливих галузях їх застосування. Наприклад, у банківській сфері, коли модель ШІ відхиляє заявку на отримання кредиту, важливо пояснити заявнику причини відхилення та які коригувальні дії він може вжити, щоб отримати кредит. Або у медичній діагностиці, коли модель ШІ прогнозує розвиток діабету у людини, важливо пояснити чому та які фактори сприятимуть розвитку захворювання в майбутньому. Нарешті, при експлуатації автономного транспортного засобу, коли модель ШІ ідентифікує об'єкти на дорозі, важливо пояснити чому було прийнято певне рішення щодо руху цього транспорту в певній ситуації. Моделі ШІ на основі машинного або глибокого навчання розв'язують задачу пошуку функції, що задовольняє як навчальним, так і тестовим даним. Але функціональний зв'язок зазвичай настільки складний, що його не можна пояснити кінцевому користувачу моделі ШІ. Тому доречно застосувати для пояснення вербальні методи, що враховують можливість та обмеження людини в процесі обробки інформації при вирішенні складних слабо структурованих проблем, що описуються якісними факторами.

Аналіз проблеми й наявних методів

Коли рішення повністю приймаються машинами, а люди завжди на стороні одержувача, виникає гостра необхідність зрозуміти, як машини дійшли

цих рішень. Моделі, на яких засновані системи ШІ, часто називають моделями "чорної скриньки". Отже, існує необхідність у зрозумілості та інтерпретованості моделей для того, щоб пояснити зроблені ними прогнози [1].

Лінійні моделі. Лінійні моделі, такі як лінійна регресія для прогнозування реальної величини виходу або модель логістичної регресії для передбачення класу та відповідних ймовірностей, є алгоритмами контрольованого навчання. Ці лінійні моделі для контрольованого машинного навчання є дуже простими для інтерпретації. Їх також легко пояснити зацікавленим особам [2–5].

Лінійна регресія використовується для прогнозування кількісного результату цільової змінної з огляду на набір предикторів. Формула моделювання зазвичай виглядає так:

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \varepsilon. \quad (1)$$

Бета-коефіцієнти відомі як параметри, а епсилон-член відомий як похибка. Похибку можна розглядати як консолідований показник, який відображає нездатність моделі прогнозувати. Нічого не можна передбачити зі 100% точністю в реальному світі, оскільки варіації даних є реальністю. Дані постійно змінюються. Метою розробки моделі є прогнозування з якомога можливою точністю та стабільністю. Цільова змінна приймає значення члена перехоплення, коли незалежні змінні приймають нульове значення.

Моделі лінійної регресії застосовуються, коли цільова ознака є безперервною. Коли ж цільова ознака є

двійковою, як-от 0 або 1, істина чи хибність, прийняти чи відхилити, то модель лінійної регресії не застосовується. Це пояснюється тим, що прогнозоване значення для цільової функції може перевищувати діапазон 0 і 1, але очікується обмеження вихідних даних двома класами, оскільки потрібно передбачити два класи окремо. Ось чому потрібна модель логістичної регресії, яка використовує двійкові значення для обчислення логарифмічних шансів, відомих як відношення шансів. Співвідношення шансів лінійно пов'язане з характеристиками. Саме тому модель логістичної регресії відома як лінійна модель [6–9].

Модель логістичної регресії можна пояснити за допомогою наступного рівняння:

$$\text{Sigmoid}(t) = \frac{1}{1+e^{-t}}. \quad (2)$$

Наведена вище функція також відома як сигмоїдна функція.

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n. \quad (3)$$

$\ln\left(\frac{p}{1-p}\right)$ – логарифмічна ймовірність результату. Бета-коефіцієнти, зазначені у наведеному вище рівнянні, пояснюють, як ймовірність змінної результату збільшується або зменшується на кожну одиницю збільшення або зменшення пояснюючої змінної. Форма сигмоїдної функції виглядає як s-подібна крива.

Інтерпретація моделі логістичної регресії істотно відрізняється від моделі лінійної регресії. Зважена сума з правої частини рівняння перетворюється на значення ймовірності. Ліва частина рівняння називається логарифмічною ймовірністю, оскільки це логарифм відношення ймовірності того, що подія відбудеться, до ймовірності того, що не відбудеться.

Нелінійні моделі. Дерево рішень – це нелінійна модель, яка пов'язує незалежну змінну із залежною. На локальному рівні це можна розглядати як кусково-лінійну регресію, але на глобальному рівні – це нелінійна модель, оскільки немає однозначного зв'язку між залежною та незалежними змінними. На відміну від лінійної регресійної моделі немає математичного рівняння, що показує взаємозв'язок між вхідними та вихідними змінними. Якщо зберігати параметр максимальної глибини дерева на нескінченному рівні, то дерево рішень може ідеально підходити до даних, що є класичним сценарієм надлишкового припасування моделі. Незалежно від того, чи можна навчальний набір даних лінійно поділити чи ні, дерева рішень схильні до надмірного припасування. Із цим необхідно боротися. Зазвичай вдаються до обрізки дерев, щоб отримати найкращу модель. Можна розглядати дерево рішень як послідовність умовних тверджень, результатом яких є значення або клас у вихідному стовпчику. Наприклад, якщо людині 45 років, вона працює у приватному секторі і має ступінь доктора філософії, то вона безперечно заробляє понад 50 тис. дол. на рік. Дерева рішень – це алгоритм контрольованого навчання, який застосовується, коли існує безліч можливих комбінацій характеристик, здатних вплинути на цільовий стовпець. У дереві рішень генеральна сукупність розбивається на дві чи більше

підгруп з урахуванням найбільш значущого роздільника чи диференціатора у вхідних змінних [10–13].

Ансамблеві моделі. Ансамблеві моделі – це найскладніший набір моделей, які потребують детального пояснення, оскільки результат є об'єднаним результатом кількох передбачень. Ансамбль має на увазі просто групування. Що важливо у випадку ансамблевих моделей, так це те, як пояснити прогнози, який варіант моделі фактично створив прогноз, і як прочитати граничний внесок функцій у остаточний процес прогнозування [14].

Перевага моделі дерева рішень полягає в тому, що вона враховує потенційну нелінійність, яка існує в наборі даних. Змінні взаємодіють вступають у гру під час створення прогнозів моделі. Однак обмеження моделі дерева рішень полягає в тому, що вона схильна до упередженості, оскільки потужні або сильніші функції беруть участь у процесі побудови дерева, а слабкі функції не можуть увійти в процес розгалуження дерева, оскільки їм бракує передбачуваної сили. Таким чином, модель стає упередженою до кількох вибіркових і сильніших характеристик із набору даних. Це також іноді призводить до переобладнання моделі. Щоб збалансувати вплив сильних функцій, важливо регулювати введення функцій у моделі на основі дерева. Якщо залишити сильні функції поза етапом створення моделі та включити лише слабкі функції до створення дерева, все одно можна генерувати прогнози, але це знову буде упереджена модель. Таким чином, лікування зміщення моделі та водночас контроль переобладнання можна зробити лише тоді, коли використовують комбінацію сильних і слабких функцій у процесі побудови дерева. Комбінацію можна виконати виключно на основі методу завантаження, а кількість дерев можна збільшити достатньо, щоб усереднити прогнози [15].

Ансамблеві моделі бувають трьох типів: моделі упаковки (bagging, також відомі як моделі агрегування початкового завантаження), моделі підвищення (boosting) та моделі укладання (stacking). Моделі укладання можуть бути двох видів: групове укладання (same group stacking) та компонування різних варіантів (different variant model stacking). Групове укладання включає однорідні типи моделей, такі як тільки включення деревоподібних моделей і зіставлення кожного результату моделі з іншими моделями. Гетерогенна модель укладання означає накладення деревоподібних та недеревоподібних моделей одна на одну та поєднання їх прогнозів. Зрозумілість моделей укладання зовсім не складна, оскільки можна ідентифікувати конкретну модель і пояснити прогнози та параметри. Більш складними є моделі упаковки та підвищення. Модель випадкового лісу є прикладом моделі упаковки, де вирощується багато дерев, та їх прогнози поєднуються для досягнення кінцевого результату [16]. Моделі підвищення беруть базову модель, навчаються на її результатах та намагаються покращити модель ітеративним способом [17].

Моделі часових рядів. Основною метою моделі часових рядів є оцінка значення цільової змінної з використанням часу як незалежної змінної. Цільовою змінною може бути вартість акцій, кількість одиниць

продукту, сума доходу, який надходить на рахунок компанії або кількість унікальних відвідувачів певного веб-сайту. Прогнозовані значення є багатокеровими, оскільки, використовуючи модель часових рядів, зазвичай, прогноують кілька часових кроків. Модель часових рядів генерує прогнозні значення. Вони мають певні рівні довіри. Чим вищий рівень достовірності, тим краща модель. На нижчих рівнях довіри моделі бракує стабільності у створенні прогнозних значень. Довірчий інтервал можна розрахувати як прогнозоване значення очікуваного значення плюс-мінус 1,96 (стандартизоване значення зі статистичної таблиці, що відповідає 95% довіри), помножене на стандартну помилку залишкового терміну, розрахованого на основі моделі. Це засновано на умові нормального розподілу члена помилки.

Модель часових рядів вимагає, щоб дані записувалися через часті проміжки часу, без будь-яких перерв у часовому кроці. Дані часових рядів є впорядкованими за своєю природою, оскільки порядок визначає неявну послідовність часу. Інженеру з машинного навчання важливо створювати корисні функції з даних, щоб робити правильні прогнози. У моделі часових рядів час є незалежною змінною. В однофакторній моделі часових рядів є лише одна змінна. У моделі причинного прогнозування використовують модель, подібну до регресійної моделі. В однофакторній моделі часових рядів ознаками є авторегресійні терміни, такі як терміни з відставанням, терміни ковзного середнього (наприклад, триперіодне або п'ятиперіодне ковзне середнє) і різниці терміни. Найпопулярніші моделі прогнозування часових рядів спираються виключно на історичні значення цільової змінної [18-22].

Моделі обробки природної мови. Сьогодні весь світ пов'язаний між собою Всесвітнім павутинням (World Wide Web). Неструктуровані текстові дані трапляються всюди. Вони знаходяться в соціальних мережах, у розмовах електронною поштою, у чатах через різні програми, на HTML-сторінках, у текстових документах, у службах підтримки клієнтів, відповідях на онлайн-опитування та багато іншого. Нижче наведено деякі варіанти використання неструктурованих даних:

- класифікація документів, коли на вході знаходиться текстовий документ, а виходом може бути бінарний клас або багатокласовий ярлик;
- іноді якщо настрої позначені, вони також відповідають сценарію класифікації документів. Інакше класифікація настроїв буде класифікацією з урахуванням лексику;
- моделі розпізнавання іменованих сутностей (named entity recognition – NER), де на вході перебуває текстовий документ, але в виході – клас іменованих сутностей;
- оцінка якості текстового опису задач спринту, ідентифікації та класифікації їх на «зрозумілі» та «не зрозумілі»;
- резюме тексту, коли великий текст стиснутий і представлений в компактній формі.

У класифікації текстів звичайними є моделі NER і передбачення наступного слова, де вхід – це

пропозиція чи вектори слів, а вихід – мітка, яку потрібно класифікувати чи передбачити. До настання ери машинного навчання завдання класифікації тексту вирішувалася вручну, коли група анотаторів читала, розуміла зміст тексту, що є у документі, і відносила його до певного класу. З масштабним зростанням обсягу неструктурованих даних ручний спосіб класифікації став дуже складним. Тепер деякі анотовані дані можуть бути введені в комп'ютер і застосований алгоритм навчання, щоб навчену модель можна було в майбутньому використовувати для прогнозування.

Неструктуровані документи чи вхідні текстові вектори дуже багатовимірні. Прогностичні моделі, що використовуються для класифікації документів, повинні бути пояснені, тому що причини, що лежать в основі прогнозу, або особливості, що лежать в основі передбаченого класу, мають бути показані кінцевому користувачеві. У випадку класифікації тексту модель передбачає клас 1 проти класу 2, тому важливо знати ключові слова, які є позитивними та негативними для класу 1. У багатокласовій класифікації це стає більш складним, оскільки необхідно пояснити всі ключові слова, які призводять до передбачення певного класу [23-29].

Існують різні інструменти та механізми для створення зрозумілості моделей. Бібліотеки Python з відкритим вихідним кодом мають деякі переваги та недоліки [30].

SHAP. Бібліотека SHAP (SHapley Additive exPlanations) – це уніфікований підхід на базі Python для пояснення результатів будь-якої моделі машинного навчання. Бібліотека SHAP Python заснована на теорії ігор із локальними поясненнями. Підхід теорії ігор – це спосіб отримати прогнози за наявності одного чинника проти його відсутності. Якщо відбувається значна зміна в очікуваному результаті, то фактор є дуже важливим для цільової змінної. Цей метод поєднує кілька попередніх методів для пояснення результатів, що генеруються моделями машинного навчання. Фреймворк SHAP може бути використаний для різних типів моделей, крім моделей на основі часових рядів. Бібліотека SHAP може бути використана для осмислення моделей [31].

LIME. LIME розшифровується як локальні інтерпретовані пояснення, що не залежать від моделі (Local Interpretable Model-Agnostic Explanations). Локальне відноситься до пояснення локальності класу, що був передбачений моделлю. Поведінка класифікатора при локальності дає розуміння прогнозу. Інтерпретованість означає, що коли передбачення не може бути інтерпретовано людиною, то в ньому немає сенсу. Отже, передбачення класів мають бути інтерпретованими. Незалежність від моделі передбачає, що замість розуміння конкретного типу моделі система та метод мають бути здатними генерувати інтерпретації.

Проблема класифікації тексту (наприклад, аналіз настрою) – коли на вхід подаються тексти документів, а на виході отримується клас. Коли модель передбачає позитивний настрій для тексту, необхідно знати, які слова змусили модель передбачити клас як

позитивний. Ці вектори слів іноді дуже прості, наприклад, окремі слова. Іноді вони складні (наприклад, вкраплення слів), і в цьому випадку потрібно знати, як модель інтерпретувала вкраплення слів і як це впливає на класифікацію. У цих сценаріях бібліотека LIME є надзвичайно корисною для розуміння сенсу моделей машинного навчання та глибокого навчання [32-35].

EL15. ELI5 – це бібліотека на базі Python, призначена для створення зрозумілого конвеєра ШІ, який дозволяє візуалізувати та налагоджувати різні моделі машинного навчання за допомогою уніфікованого API. Вона має вбудовану підтримку кількох фреймворків та надає можливість пояснювати "чорні ящики" моделей. Мета бібліотеки – зробити пояснення простими для всіх видів моделей "чорного ящика" [36-42].

Мета статті - проаналізувати математичні форми та методи, що використовуються для пояснення рішень, які приймають моделі штучного інтелекту; провести аналіз методів, класів, фреймворків та функцій програмних бібліотек, а також їх використання для пояснення рішень, що приймають моделі ШІ; запропонувати підхід до синтезу пояснених вербальних моделей штучного інтелекту; розробити проєкт системи синтезу пояснених вербальних моделей штучного інтелекту.

Матеріали та методи дослідження

Інформаційна технологія аналізу та синтезу пояснювальних моделей штучного інтелекту на основі вербальних методів передбачає використання обробки природної мови й машинного навчання для створення розмовних систем ШІ, які можуть розуміти людську мову та реагувати на неї. Ці системи можна використовувати в різних додатках, таких як чат-боти, віртуальні асистенти та голосові помічники, для покращення розуміння рекомендації, що надаються користувачеві [1, 43, 44].

Методи вербального аналізу використовуються для розв'язування складних неструктурованих проблем. Дослідження в цій галузі зосереджені на розробці методів підтримки прийняття рішень, які включають як числові, так і вербальні аспекти. Вони звертають увагу на важливість включення вербальних елементів у процеси прийняття рішень, підкреслюючи цінність лінгвістичної інформації поряд з числовими даними. Такий підхід дозволяє проводити більш комплексний аналіз складних проблем, які не можуть бути повністю охоплені лише кількісними методами [45]. Ці розробки мають значення для навчання та аналізу комунікативних патернів за допомогою технологій ШІ [46].

Синтез пояснювальних вербальних моделей ШІ починається з формування системи понять певної предметної галузі. Виявляються критерії, якими користуються фахівці при аналізі об'єктів або процесів в певній предметній галузі. Для кожного критерію формується шкала впорядкованих значень: від кращих до гірших або від більш характерних для об'єкта/процесу в певній ситуації до менш характерних. Це дозволяє формувати критеріальні описи всіх

гіпотетично можливих станів об'єкта/процесу, порівнювати їх через використання відповідних графів домінування станів, що будуються на основі шкал критеріїв, та визначати й пояснювати чому певний стан об'єкта/процесу кращий/гірший, ніж інший [47].

На другому етапі відбувається формування порядкової класифікації станів об'єкта/процесу. Визначається назва та лінійний порядок класів. Класифікація відбувається за принципами "кращий стан не може потрапити у гірший клас" та "гірший стан не може потрапити у кращий клас". Це дозволяє сформувати повну класифікацію через аналіз обмеженої кількості формально визначених найінформативніших станів, їх пряму та непряму класифікацію кращих/гірших станів, яка обмежує для них відповідні класи. Через це обмеження можна контролювати класифікацію на наявність протиріч та пояснювати, чому певний стан об'єкта/процесу належить до визначеного класу. Особливе значення мають граничні стани кожного класу, для яких зміна значення за певним критерієм призводить до переходу об'єкта/процесу в кращий/гірший клас. Це дозволяє визначати та пояснювати ефективність певної траєкторії покращення стану об'єкта/процесу через зміни значень за певними критеріями.

На третьому етапі обирається певний клас станів об'єкта/процесу та формується їх впорядкування. Для цього будуються єдині порядкові шкали для всіх можливих пар критеріїв, які потім об'єднуються в єдину порядкову шкалу всіх критеріїв, що використовується для порівняння станів об'єкта/процесу з визначеного класу, які неможливо було порівняти через відповідні графи домінування станів. Можливість побудови єдиної порядкової шкали всіх критеріїв без циклів є ознакою відсутності протиріч впорядкування станів, а її використання пояснює чому певний стан об'єкта/процесу кращий/гірший, ніж інший.

На четвертому етапі відбувається вибір найкращого стану серед тих, що не вдалося впорядкувати на попередньому етапі. Для цього обираються кращі стани з кожної впорядкованої гілки. Випадковим чином обирається пара станів для порівняння та визначаються їх відносні недоліки за кожним критерієм. Ці недоліки впорядковуються за важливістю для користувача та відбувається процес їх взаємної компенсації. За результатами порівняння визначається кращий стан з цієї пари, який порівнюється з іншим випадковим станом за такою ж процедурою. Наприкінці залишається тільки один найкращий стан об'єкта/процесу.

Отже, методи вербального аналізу виявляються ефективними для синтезу пояснених моделей ШІ, що включає кілька етапів: визначення системи понять; створення критеріальних описів станів; їх класифікація, впорядкування та обрання найкращого стану. Вони підкреслюють важливість використання лінгвістичної інформації разом з числовими даними для комплексного аналізу складних проблем.

Для розробки системи синтезу пояснених моделей штучного інтелекту (ШІ) необхідно врахувати кілька ключових аспектів [48–50]:

– зіставлення вербальної поведінки людини з поведінкою системи ШІ, враховуючи особистісні

критерії, щоб генерувати відповідні вербальні та невербальні комбіновані моделі поведінки;

– покращення пояснення систем ШІ в користувацьких інтерфейсах, зосередившись на зрозумілості, персоналізації та візуалізації пояснень, визнанні компромісів, усуненні потенційних помилкових уявлень і прив'язці пояснень до ментальних моделей користувачів задля прозорості та сприйняття користувачами рекомендацій систем ШІ;

– перевірка результатів дослідження в реальних умовах через вивчення персоналізованих пояснень, візуалізацію та забезпечення відображення пояснень у користувацькому інтерфейсі.

Вербальний аналіз рішень [45] – це ефективний інструмент, який може вдосконалити системи пояснюваного ШІ, надаючи словесні пояснення моделей

машинного навчання та міркувань за допомогою слів, тексту або природної мови. Наприклад, цей підхід може допомогти людям краще зрозуміти процес прийняття рішень системами ШІ і потенційно зменшити кількість пацієнтів, яким необхідно пройти тривалий і багатоетапний процес діагностики [51].

Синтез пояснювальних вербальних моделей ШІ передбачає чотири основні етапи [52] (рис. 1):

- формування системи понять визначеної предметної галузі;
- формування порядкової класифікації станів об'єкта/процесу;
- впорядкування станів об'єкта/процесу з певного класу;
- визначення найкращого стану об'єкта/процесу.

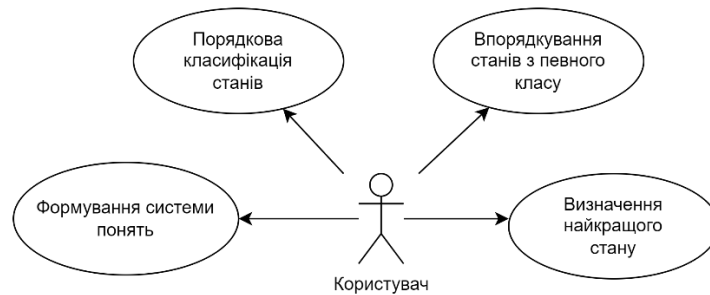


Рис. 1. Діаграма варіантів використання системи

Формування системи понять визначеної предметної галузі складається з таких діяльностей (рис. 2):

- виявлення критеріїв порівняння об'єктів/процесів певної предметної галузі;
- формування порядкових шкал для критеріїв;
- побудова графа домінування станів об'єктів/процесів;
- порівняння визначених станів та пояснення типу відношення між ними (краще, гірше, непорівнянні).

Порядкова класифікація гіпотетично можливих станів складається з таких діяльностей (рис. 3):

- визначення назв та лінійного порядку класів;
- визначення найінформативнішого стану;
- пряма класифікація найінформативнішого стану;

- перевірка несуперечливості прямої класифікації;
- непряма класифікація станів;
- класифікація реальних станів;
- визначення та пояснення ефективності певної траєкторії покращення стану.

Впорядкування станів об'єкта/процесу з певного класу складається з таких діяльностей (рис. 4):

- побудова єдиної порядкової шкали (ЄПШ) для всіх пар критеріїв;
- побудова єдиної порядкової шкали для всіх критеріїв;
- перевірка несуперечливості ЄПШ;
- порівняння визначених станів та пояснення типу відношення між ними (краще, гірше, непорівнянні).

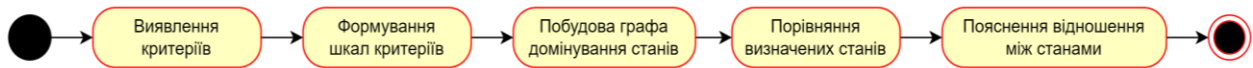


Рис. 2. Діаграма діяльностей формування системи понять

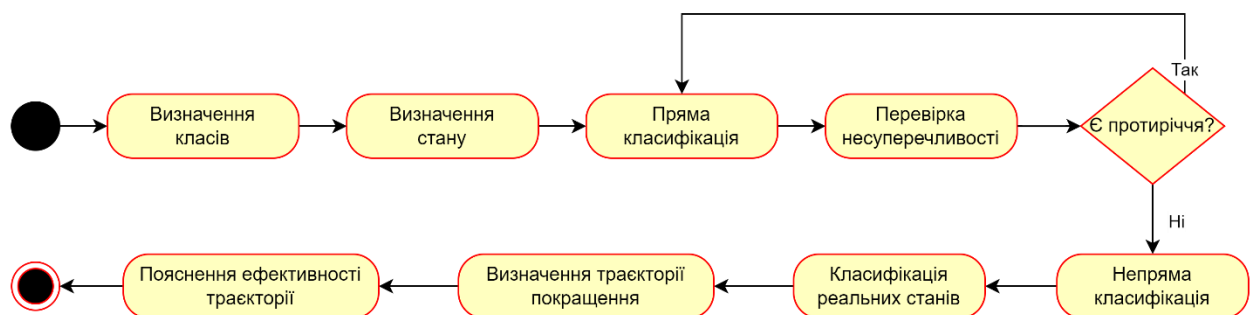


Рис. 3. Діаграма діяльностей порядкової класифікації станів

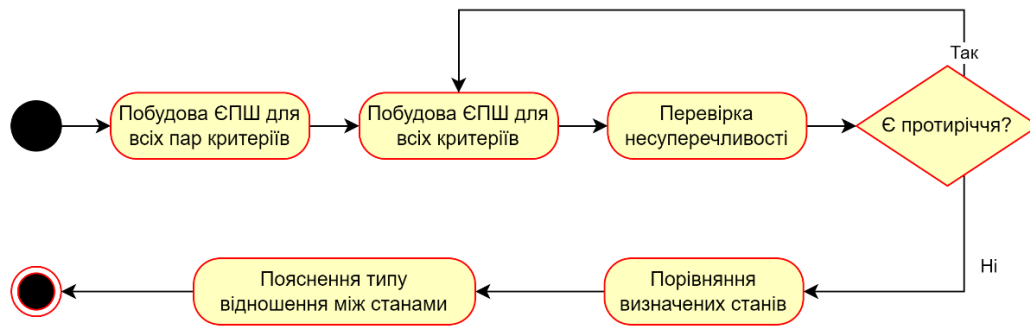


Рис. 4. Діаграма діяльностей впорядкування станів

Визначення найкращого стану об'єкта/процесу складається з таких діяльностей:

- вибір пари станів для порівняння;
- визначення відносних недоліків станів за кожним критерієм;
- впорядкування недоліків за важливістю для користувача;
- взаємна компенсація недоліків;
- визначення кращого стану.

Результати

Проведено аналіз математичних формул та методів, що використовуються для пояснення рішень, що приймають моделі штучного інтелекту:

- лінійні моделі;
- нелінійні моделі;
- ансамблеві моделі;
- моделі часових рядів;
- моделі обробки природної мови.

Проведено аналіз методів, класів, фреймворків та функцій програмних бібліотек, а також їх використання для пояснення рішень, що приймають моделі штучного інтелекту:

- SHAP;
- LIME;
- ELI5.

Запропоновано підхід до синтезу пояснених вербальних моделей штучного інтелекту.

Розроблено проєкт системи синтезу пояснених вербальних моделей штучного інтелекту.

Висновки

Отже, методи вербального аналізу виявляються ефективними для синтезу пояснених моделей штучного інтелекту, що включає кілька етапів: визначення системи понять, створення критеріальних описів станів, їх класифікація, впорядкування та обрання найкращого стану.

Вони підкреслюють важливість використання лінгвістичної інформації разом з числовими даними для комплексного аналізу складних проблем.

Інтегруючи елементи вербального аналізу в пояснені моделі штучного інтелекту, можна покращити взаємодію з користувачем, його розуміння і сприйняття систем штучного інтелекту.

СПИСОК ЛІТЕРАТУРИ

1. Mishra, P. Model Explainability and Interpretability. *Practical Explainable AI Using Python*. Apress, Berkeley, CA. 2022. P. 1–22. DOI: https://doi.org/10.1007/978-1-4842-7158-2_1
2. Montesinos López, O. A., Montesinos López, A., Crossa, J. Linear Mixed Models. *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Springer, Cham. 2022. P. 141–170. DOI: https://doi.org/10.1007/978-3-030-89010-0_5
3. Starbuck, C. Linear Regression. *The Fundamentals of People Analytics*. Springer, Cham. 2023. P. 181–206. DOI: https://doi.org/10.1007/978-3-031-28674-2_10
4. Elements of Generalized Linear Mixed Models / Salinas Ruíz, J. et al. *Generalized Linear Mixed Models with Applications in Agriculture and Biology*. Springer, Cham. 2023. P. 1–42. DOI: https://doi.org/10.1007/978-3-031-32800-8_1
5. Linear Regression / James, G. et al. *An Introduction to Statistical Learning*. Springer Texts in Statistics. Springer, Cham. 2023. P. 69–134. DOI: https://doi.org/10.1007/978-3-031-38747-0_3
6. Wüthrich, M. V., Merz, M. Generalized Linear Models. *Statistical Foundations of Actuarial Learning and its Applications*. Springer Actuarial. Springer, Cham. 2023. P. 111–205. DOI: https://doi.org/10.1007/978-3-031-12409-9_5
7. Kleinbaum, D. G., Klein, M. Important Special Cases of the Logistic Model. *Logistic Regression. Statistics for Biology and Health*. Springer, New York, NY. 2010. P. 41–71. DOI: https://doi.org/10.1007/978-1-4419-1742-3_2
8. Logistic Regression / Weisburd, D. et al. *Advanced Statistics in Criminology and Criminal Justice*. Springer, Cham. 2022. P. 127–185. DOI: https://doi.org/10.1007/978-3-030-67738-1_4
9. Logistic Regression / Backhaus, K. et al. *Multivariate Analysis*. Springer Gabler, Wiesbaden. 2021. P. 267–354. DOI: https://doi.org/10.1007/978-3-658-32589-3_5
10. Grąbczewski, K. Techniques of Decision Tree Induction. *Meta-Learning in Decision Tree Induction*. Studies in Computational Intelligence. 2014. Vol. 498. P. 11–117. DOI: https://doi.org/10.1007/978-3-319-00960-5_2
11. Barros, R.C., de Carvalho, A.C.P.L.F., Freitas, A.A. Decision-Tree Induction. *Automatic Design of Decision-Tree Induction Algorithms*. SpringerBriefs in Computer Science. Springer, Cham. 2015. P. 7–45. DOI: https://doi.org/10.1007/978-3-319-14231-9_2
12. Kozak, J. Ant Colony Decision Tree Approach. *Decision Tree and Ensemble Learning Based on Ant Colony Optimization*. Studies in Computational Intelligence. 2019. Vol. 781. P. 45–80. DOI: https://doi.org/10.1007/978-3-319-93752-6_3

13. Construction of Optimal Decision Trees and Deriving Decision Rules from Them / Azad, M. et al. *Decision Trees with Hypotheses. Synthesis Lectures on Intelligent Technologies*. Springer, Cham. 2022. P. 41–53. DOI: https://doi.org/10.1007/978-3-031-08585-7_4
14. Seni, G., Elder, J. F. Importance Sampling and the Classic Ensemble Methods. *Ensemble Methods in Data Mining. Synthesis Lectures on Data Mining and Knowledge Discovery*. Springer, Cham. 2010. P. 53–87. DOI: https://doi.org/10.1007/978-3-031-01899-2_4
15. Seni, G., Elder, J. F. Predictive Learning and Decision Trees. *Ensemble Methods in Data Mining. Synthesis Lectures on Data Mining and Knowledge Discovery*. Springer, Cham. 2010. P. 14–27. DOI: https://doi.org/10.1007/978-3-031-01899-2_2
16. Cutler, A., Cutler, D. R., Stevens, J. R. Random Forests. Zhang, C., Ma, Y. (eds) *Ensemble Machine Learning*. Springer, New York, NY. 2012. P. 157–175. DOI: https://doi.org/10.1007/978-1-4419-9326-7_5
17. Ferreira, A. J., Figueiredo, M. A. T. Boosting Algorithms: A Review of Methods, Theory, and Applications. Zhang, C., Ma, Y. (eds) *Ensemble Machine Learning*. Springer, New York, NY. 2012. P. 35–85. DOI: https://doi.org/10.1007/978-1-4419-9326-7_2
18. Štulajter, F. Random Processes and Time Series. *Predictions in Time Series Using Regression Models*. Springer, New York, NY. 2002. P. 51–71. DOI: https://doi.org/10.1007/978-1-4757-3629-8_2
19. Das, M., Ghosh, S. K. Standard Bayesian Network Models for Spatial Time Series Prediction. *Enhanced Bayesian Network Models for Spatial Time Series Prediction. Studies in Computational Intelligence*. 2020. Vol. 858. P. 11–22. DOI: https://doi.org/10.1007/978-3-030-27749-9_2
20. Estimation Under Normal Mixture Models for Financial Time Series Data /Sun, L.H. Et al. *Copula-Based Markov Models for Time Series. SpringerBriefs in Statistics()*. Springer, Singapore. 2020. P. 55–72. DOI: https://doi.org/10.1007/978-981-15-4998-4_4
21. Brockwell, P. J., Davis, R. A. ARMA Models. *Introduction to Time Series and Forecasting. Springer Texts in Statistics*. Springer, Cham. 2016. P. 73–96. DOI: https://doi.org/10.1007/978-3-319-29854-2_3
22. Interval Evaluation of Stationary State Probabilities for Markov Set-Chain Models / O. Akhiezer et al. *10th International Conference on Advanced Computer Information Technologies (ACIT)*. 2020. P. 82–85. DOI: <https://doi.org/10.1109/ACIT49673.2020.9208932>
23. Petrov, S. Latent Variable Grammars for Natural Language Parsing. *Coarse-to-Fine Natural Language Processing. Theory and Applications of Natural Language Processing*. Springer, Berlin, Heidelberg. 2011. P. 7–46. DOI: https://doi.org/10.1007/978-3-642-22743-1_2
24. Goyal, P., Pandey, S., Jain, K. Introduction to Natural Language Processing and Deep Learning. *Deep Learning for Natural Language Processing*. Apress, Berkeley, CA. 2018. P. 1–74. DOI: https://doi.org/10.1007/978-1-4842-3685-7_1
25. Kulkarni, A., Shivananda, A. Advanced Natural Language Processing. *Natural Language Processing Recipes*. Apress, Berkeley, CA. 2019. P. 97–128. DOI: https://doi.org/10.1007/978-1-4842-4267-4_4
26. Harris, I. G., Harris, C. B. Generation of Verification Artifacts from Natural Language Descriptions. Soeken, M., Drechsler, R. (eds) *Natural Language Processing for Electronic Design Automation*. Springer, Cham. 2020. P. 37–70. DOI: https://doi.org/10.1007/978-3-030-52273-5_3
27. Catta, D., Moot, R., Retoré, C. Dialogical Argumentation and Textual Entailment. Loukanova, R. (eds) *Natural Language Processing in Artificial Intelligence—NLPinAI 2020. Studies in Computational Intelligence*. 2021. Vol. 939. P. 191–226. DOI: https://doi.org/10.1007/978-3-030-63787-3_7
28. Rodrigues, M., Teixeira, A. Extracting Relevant Information Using a Given Semantic. *Advanced Applications of Natural Language Processing for Performing Information Extraction. SpringerBriefs in Electrical and Computer Engineering()*. Springer, Cham. 2015. P. 37–50. DOI: https://doi.org/10.1007/978-3-319-15563-0_4
29. Гринченко, М., Роговий, М. Модель ідентифікації задач спринту проєкту на основі їх опису. Сучасний стан наукових досліджень та технологій в промисловості, 2023. 4(26), P. 33–44. <https://doi.org/10.30837/ITSSI.2023.26.033>
30. Gianfagna, L., Di Cecco, A. Explainable AI: Needs, Opportunities, and Challenges. *Explainable AI with Python*. Springer, Cham. 2021. P. 27–46. DOI: https://doi.org/10.1007/978-3-030-68640-6_2
31. Mishra, P. Explainability for Ensemble Supervised Models. *Explainable AI Recipes*. Apress, Berkeley, CA. 2023. P. 119–206. DOI: https://doi.org/10.1007/978-1-4842-9029-3_4
32. Handling Missing Values in Local Post-hoc Explainability / Cinquini, M. et al. Longo, L. (eds) *Explainable Artificial Intelligence. xAI 2023. Communications in Computer and Information Science*. 2023. Vol. 1902. P. 256–278. DOI: https://doi.org/10.1007/978-3-031-44067-0_14
33. Explainable AI Methods - A Brief Overview / Holzinger, A. et al. Holzinger, A., Goebel, R., Fong, R., Moon, T., Müller, KR., Samek, W. (eds) *xxAI - Beyond Explainable AI. xxAI 2020. Lecture Notes in Computer Science()*. 2022. Vol. 13200. P. 13–38. DOI: https://doi.org/10.1007/978-3-031-04083-2_2
34. Kamath, U., Liu, J. Post-Hoc Interpretability and Explanations. *Explainable Artificial Intelligence: An Introduction to Interpretable Machine Learning*. Springer, Cham. 2021. P. 167–216. DOI: https://doi.org/10.1007/978-3-030-83356-5_5
35. Kumar, D., Mehta, M. A. An Overview of Explainable AI Methods, Forms and Frameworks. Mehta, M., Palade, V., Chatterjee, I. (eds) *Explainable AI: Foundations, Methodologies and Applications. Intelligent Systems Reference Library*. 2023. Vol. 232. P. 43–59. DOI: https://doi.org/10.1007/978-3-031-12807-3_3
36. Vuppapalapati, C. ML Models: Food Security and Climate Change. *Artificial Intelligence and Heuristics for Enhanced Food Security. International Series in Operations Research & Management Science*. 2022. Vol. 331. P. 395–518. DOI: https://doi.org/10.1007/978-3-031-08743-1_6
37. Sonar Signal Prediction Using Explainable AI for IoT Environment / Yadav, T. et al. Bansal, H.O., Ajmera, P.K., Joshi, S., Bansal, R.C., Shekhar, C. (eds) *Next Generation Systems and Networks. BITS-EEE-CON 2022. Lecture Notes in Networks and Systems*. 2023. Vol. 641. P. 209–222. DOI: https://doi.org/10.1007/978-981-99-0483-9_19
38. Brad, S., Ştetco, E. An Interactive Artificial Intelligence System for Inventive Problem-Solving. Nowak, R., Chrzyszcz, J., Brad, S. (eds) *Systematic Innovation Partnerships with Artificial Intelligence and Information Technology. TFC 2022. IFIP Advances in Information and Communication Technology*. 2022. Vol. 655. P. 165–177. DOI: https://doi.org/10.1007/978-3-031-17288-5_15

39. Yan, W. Q. Convolutional Neural Networks and Recurrent Neural Networks. *Computational Methods for Deep Learning. Texts in Computer Science*. Springer, Singapore. 2023. P. 69–124. DOI: https://doi.org/10.1007/978-981-99-4823-9_3
40. Paaß, G., Giesselbach, S. Foundation Models for Text Generation. *Foundation Models for Natural Language Processing. Artificial Intelligence: Foundations, Theory, and Algorithms*. Springer, Cham. 2023. P. 227–311. DOI: https://doi.org/10.1007/978-3-031-23190-2_6
41. Vuppapalati, C. The Role of the Government and the AI Readiness. *Machine Learning and Artificial Intelligence for Agricultural Economics. International Series in Operations Research & Management Science*. 2021. Vol. 314. P. 479–548. DOI: https://doi.org/10.1007/978-3-030-77485-1_7
42. Ye, A., Wang, Z. Data Preparation and Engineering. *Modern Deep Learning for Tabular Data*. Apress, Berkeley, CA. 2023. P. 95–179. DOI: https://doi.org/10.1007/978-1-4842-8692-0_2
43. Фастовський Е. Г., Сльчанінов Д. Б. Інформаційна технологія аналізу та синтезу пояснених моделей штучного інтелекту. *Теоретичні та практичні дослідження молодих вчених [Електронний ресурс] : зб. тез доп. 17-ї Міжнар. наук.-практ. конф. магістрантів та аспірантів, 28-30 листопада 2023 р. С. 85. URL: <https://repository.kpi.kharkov.ua/handle/KhPI-Press/71485>*
44. Conversational AI: What Is It, How Does It Work, and Why Does It Matter?. URL: <https://www.247.ai/insights/conversational-ai-what-it-and-how-does-it-work> (дата звернення: 29.02.2024)
45. Moshkovich H, Mechitov A, Olson D. Verbal Decision Analysis. *Multiple Criteria Decision Analysis: State of the Art Surveys. International Series in Operations Research & Management Science*. 2005. Vol. 78. P. 609–633. DOI: https://doi.org/10.1007/0-387-23081-5_15
46. Butow P, Hoque E. Using artificial intelligence to analyse and teach communication in healthcare. *The Breast*. 2020. Vol. 50; P. 49–55. DOI: <https://doi.org/10.1016/j.breast.2020.01.008>
47. Фастовський Е. Г. Синтез пояснених вербальних моделей штучного інтелекту. *Інформаційні технології: теорія і практика. I (VII) міжнародна науково-практична конференція здобувачів вищої освіти і молодих учених "Інформаційні технології: теорія і практика". Тези доповідей (Дніпро 20 – 22 березня 2024)*. 2024. С. 225–227, URL: <https://ir.nmu.org.ua/handle/123456789/166564>
48. AKM Bahalul Haque, A. K. M. Najmul Islam, Patrick Mikalef. Explainable Artificial Intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research. *Technological Forecasting & Social Change*. 2023. Vol. 186, Part A. Article 122120. DOI: <https://doi.org/10.1016/j.techfore.2022.122120>
49. How to explain AI systems to end users: a systematic literature review and research agenda / Laato S. et al. *Internet Research*. 2022. Vol. 32, No 7. P. 1–31.
50. Aly A., Tapus A. A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction. *8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 2013. P. 325–332. DOI: <https://doi.org/10.1109/HRI.2013.6483606>
51. A Protocol for the Diagnosis of Autism Spectrum Disorder Structured in Machine Learning and Verbal Decision Analysis / Evandro Andrade et al. *Computational and Mathematical Methods in Medicine*. 2021. Vol. 2021. Article ID 1628959. DOI: <https://doi.org/10.1155/2021/1628959>
52. Фастовський Е. Г. Проект системи синтезу пояснених вербальних моделей штучного інтелекту. *28-й Міжнародний молодіжний форум "Радіоелектроніка та молодь у XXI столітті". Зб. матеріалів форуму (Харків 16 – 18 квітня 2024)*. 2024. Т. 6. С. 74–76. DOI: <https://doi.org/10.30837/IYF.IIS.2024.074>

Received (Надійшла) 08.04.2024

Accepted for publication (Прийнята до друку) 19.06.2024

Information technology of analysis and synthesis of explained models of artificial intelligence based on verbal methods

Eduard Fastovskyi, Anton Rogovyi, Olena Akhiezer, Andrii Frolov, Roman Artiukh

Abstract. The subject matters of the article is the analysis and synthesis of explained models of artificial intelligence. The goal of the work is to develop an information technology for analyzing and synthesizing explained models of artificial intelligence based on verbal methods. The following tasks were solved in the article: analysis of mathematical formulas and methods used to explain decisions made by artificial intelligence models, analysis of methods, classes, frameworks, and functions of software libraries, as well as their use to explain decisions made by artificial intelligence models, synthesis of explained verbal models of artificial intelligence, development of a system for synthesizing explained verbal models of artificial intelligence. The following methods are used: system analysis, verbal decision-making methods (formation of a system of concepts in a particular subject area, formation of an ordinal classification of object/process states, ordering object/process states from a particular class, determining the best state of an object/process), methods of modeling and designing information systems (use case diagrams, activity diagrams). The following results were obtained: the mathematical formulas and methods used to explain decisions made by artificial intelligence models are analyzed. An approach to the synthesis of explained verbal models of artificial intelligence is proposed. The methods, classes, frameworks, and functions of software libraries, as well as their use to explain decisions made by artificial intelligence models, are analyzed. A project of a system for synthesizing explained verbal models of artificial intelligence was developed. **Conclusions:** verbal analysis methods prove to be effective for synthesizing explained artificial intelligence models, which includes several stages: defining a system of concepts, creating criterion descriptions of states, classifying them, organizing them, and selecting the best state. They emphasize the importance of using linguistic information together with numerical data for a comprehensive analysis of complex problems. By integrating elements of verbal analysis into explained artificial intelligence models, it is possible to improve user interaction, understanding, and perception of artificial intelligence systems.

Keywords: explained models of artificial intelligence; verbal decision-making methods.

Minglei Zhou^{1,2}, Nina Kuchuk²

¹ Shanghai Polyether Rui Information Technology Co., Ltd., Shanghai, China

² National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine

A ROBUST METHOD FOR EMBEDDING AND EXTRACTING WATERMARKS IN VIDEO USING DISCRETE COSINE TRANSFORM

Abstract. This paper presents a robust method for embedding and extracting watermarks in video frames using Discrete Cosine Transform (DCT). The method ensures high imperceptibility and maintains the quality of the video while embedding the watermark in the luminance component of each frame. Extensive experiments demonstrate the method's robustness against common video processing operations, including compression, noise addition, and cropping. The proposed approach is computationally efficient, making it suitable for real-time applications. Future work aims to optimize the algorithm for faster performance and enhance robustness against more sophisticated attacks. Practical applications include digital rights management, content verification, and forensic analysis.

Keywords: Video, Discrete, Transform, Digital watermarking, Embedding Algorithm.

Introduction

In the digital age, the proliferation of multimedia content has necessitated robust methods for protecting intellectual property rights. Digital watermarking has emerged as a pivotal technology for embedding imperceptible information into multimedia data, such as images, audio, and video, to assert ownership and prevent unauthorized use. Among various watermarking techniques, video watermarking is particularly crucial due to the widespread distribution and consumption of video content across various platforms. The rapid advancement of video compression techniques and the increasing complexity of video manipulation pose significant challenges to the robustness and imperceptibility of watermarks. Traditional watermarking methods often struggle to maintain watermark integrity when subjected to lossy compression, noise, cropping, and other common video processing operations. Therefore, there is a compelling need for a watermarking method that can withstand these challenges while ensuring the watermark remains imperceptible to human viewers. The Discrete Cosine Transform (DCT) has been extensively used in image and video compression standards, such as JPEG and MPEG, due to its energy compaction properties. This makes DCT an ideal candidate for embedding watermarks, as modifications in the DCT domain can be made less perceptible. Leveraging DCT for watermarking can enhance robustness and maintain the quality of the video.

To address these challenges, we have developed a robust method for embedding and extracting watermarks in video frames using Discrete Cosine Transform (DCT), and implemented this method in a Python library called video-invisible-watermark. This library is publicly available and can be accessed from the Python Package Index (PyPI) at <https://pypi.org/project/video-invisible-watermark/>.

This thesis aims to develop a robust method for embedding and extracting watermarks in video frames using Discrete Cosine Transform (DCT). The specific objectives of this study are to design a watermark embedding algorithm that utilizes DCT coefficients for robust and imperceptible watermarking, to implement a watermark extraction algorithm capable of accurately

retrieving the embedded watermark from watermarked video frames, to evaluate the robustness of the proposed watermarking method against various video processing operations, including lossy compression, noise addition, and cropping, and to compare the performance of the proposed method with existing video watermarking techniques.

1 System Framework

The proposed system for embedding and extracting watermarks in video frames using Discrete Cosine Transform (DCT) consists of several key components and processes. The system architecture includes the following steps:

1. **Watermark Image Preparation:** Converting the watermark image to grayscale and resizing it.
2. **Video Frame Extraction:** Splitting the video into individual frames.
3. **Watermark Embedding:** Embedding the watermark image into the luminance (Y) component of each frame using DCT.
4. **Watermarked Video Reconstruction:** Combining the watermarked frames back into a video.
5. **Watermark Extraction:** Extracting the watermark from the watermarked video frames.

Here is a detailed diagram illustrating the system architecture (Fig. 1).

2 Detailed Sections

2.1 Watermark Image Preparation

To prepare the watermark image, we convert it to grayscale and resize it to a suitable size (e.g., 100x100 pixels). This ensures the watermark can be effectively embedded within the video frames.

2.2 Video Frame Extraction

We extract individual frames from the video, ensuring we process enough frames for the desired duration. This involves reading the video file and extracting each frame sequentially.

2.3 Watermark Embedding Algorithm

The watermark embedding algorithm uses DCT to embed the watermark image into the luminance (Y) component of each video frame.

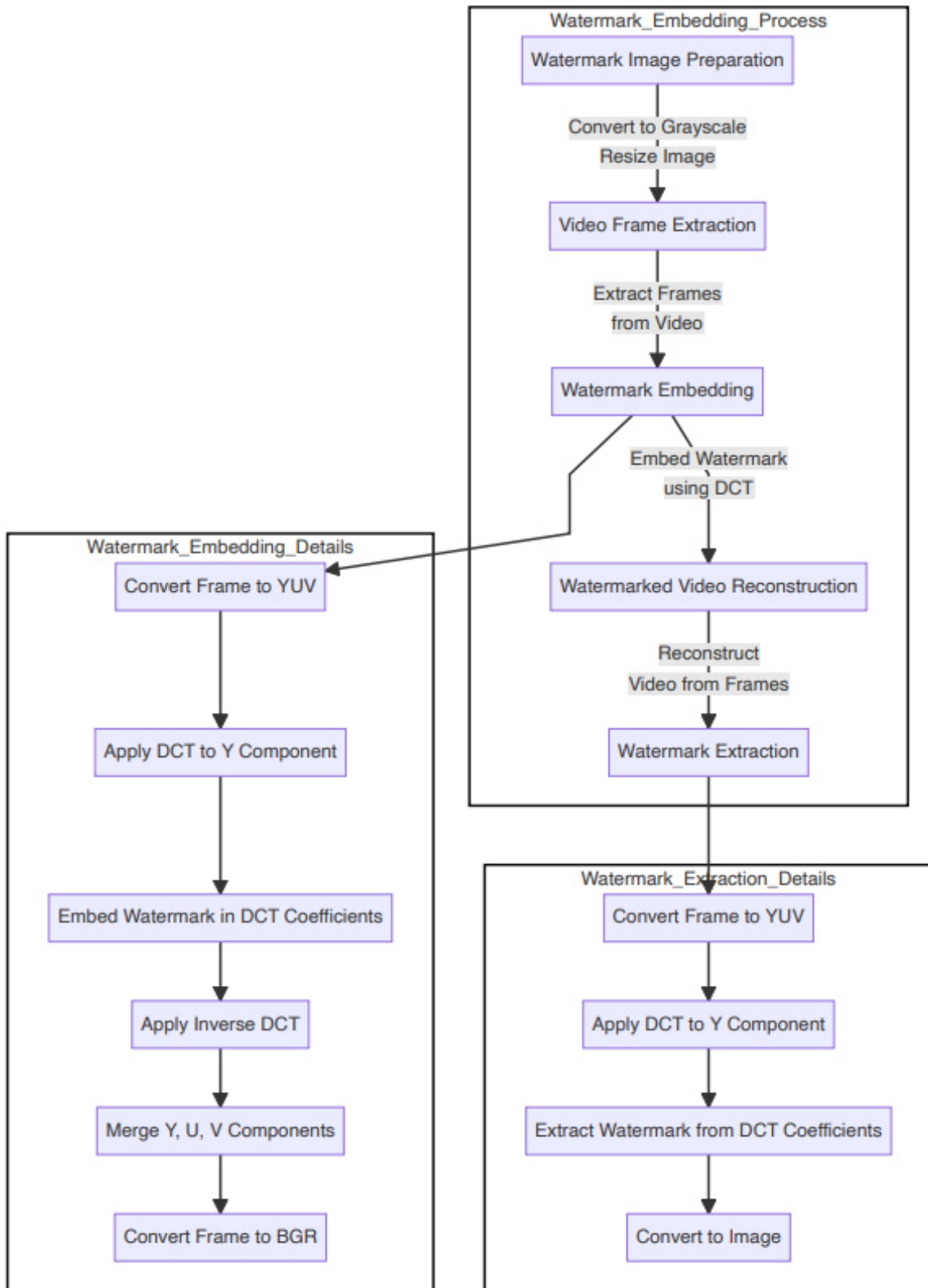


Fig. 1. Detailed diagram illustrating the system architecture

The process involves the following steps:

1. Convert Frame to YUV:

$$yuv_frame = convert_to_yuv(frame), \tag{1}$$

where $yuv_frame = (Y,U,V)$

2. Apply DCT to Y Component:

$$dct_y = DCT(Y). \tag{2}$$

3. Embed Watermark in DCT Coefficients:

The choice of DCT coefficients to embed the watermark is critical. We select mid-frequency DCT coefficients as they offer a balance between robustness and imperceptibility. The embedding process is defined by:

$$dct_y_with_watermark[i,j] = dct_y[i,j] + \alpha \cdot watermark[i,j], \tag{3}$$

where α is a scaling factor to control the strength of the watermark.

4. Apply Inverse DCT:

$$idct_y = IDCT(dct_y_with_watermark). \quad (4)$$

5. Merge Y, U, V Components:

$$watermarked_frame = merge_yuv(idct_y, U, V). \quad (5)$$

6. Convert Frame to BGR:

$$bgr_frame = convert_to_bgr(watermarked_frame). \quad (6)$$

2.4 Watermark Extraction Algorithm

The extraction algorithm reverses the embedding process by transforming the luminance component (Y) of the watermarked frames back into the DCT domain, extracting the watermark, and reconstructing the image.

1. Convert Frame to YUV:

$$yuv_frame = convert_to_yuv(frame), \quad (7)$$

where $yuv_frame = (Y,U,V)$

2. Apply DCT to Y Component:

$$dct_y = DCT(Y) \quad (8)$$

3. Extract Watermark from DCT Coefficients:

$$= \frac{Extracted_watermark[i,j] = dct_y_with_watermark[i,j] - dct_y[i,j]}{\alpha}. \quad (9)$$

4. Convert to Image:

$$Watermark_image = convert_to_image(extracted_watermark) \quad (10)$$

2.5 Algorithm Complexity Analysis

The complexity of the watermark embedding and extraction algorithms is primarily determined by the DCT and inverse DCT operations. Given an $N \times N$ image, the computational complexity of the DCT is:

$$O(N^2 \log N) \quad (11)$$

The same complexity applies to the inverse DCT. Therefore, the overall complexity of embedding or extracting a watermark from a single frame is:

$$O(N^2 \log N) \quad (12)$$

This section provides a comprehensive overview of the technical implementation and principles underlying the proposed watermarking method. The subsequent section will present the experimental results and discuss the effectiveness and robustness of the proposed method.

3 Results

This section presents the results of embedding and extracting watermarks from videos of different resolutions, formats, sizes, and durations. The robustness of the watermarking method is evaluated by examining the quality of the embedded and extracted watermarks under various conditions.

3.1 Experimental Setup and Environment

- Hardware: Intel Core i7-10700K CPU, 16GB
- Software: Python 3.8, OpenCV 4.5.3, Pillow 8.2.0
- Videos: Different videos with varying resolutions (720p, 1080p, 4K), formats (MP4, AVI),

sizes (short clips, full-length movies), and durations (30 seconds, 2 minutes, 10 minutes)

3.2 Dataset Selection

The dataset consists of a diverse collection of video samples to evaluate the robustness of the watermarking method.

The selected videos vary in resolution, format, and length to ensure comprehensive testing.

3.3 Experimental Procedure

1. Watermark Embedding: The watermark image is embedded into each video frame using DCT.

2. Watermark Extraction: The watermark is extracted from the watermarked video frames.

3.4 Evaluation Metrics

- PSNR (Peak Signal-to-Noise Ratio): Measures the quality of the watermarked video.
- SSIM (Structural Similarity Index): Assesses the similarity between the original and watermarked video frames.
- Embedding Time: The time taken to embed the watermark in the video.
- Extraction Time: The time taken to extract the watermark from the video.
- Robustness: The resilience of the watermark against compression, noise addition, and cropping.

3.5 Results for Different Resolutions, Formats, and Durations

The results are summarized in the following tables, showing the comparison between the original watermark, watermarked video frame and the extracted watermark image for different scenarios (Table 1–3, where 1 – Resolution, 2 – Format, 3 – Test, 4 – Original Watermark, 5 – Extracted Watermark, 6 – PSNR(dB), 7 – SSIM, 8 – Embedding Time(s), 9 – Extraction Time(s)).

Table 1 – Embedding and Extraction Results for Different Resolutions

1	4	5	6	7	8	9
720p			38.5	0.98	0.12	0.08
1080p			36.2	0.97	0.18	0.14
4K			34.7	0.95	0.35	0.30

Table 2 – Embedding and Extraction Results for Different Formats

2	4	5	6	7	8	9
MP4			37.5	0.97	0.15	0.12
AVI			36.9	0.96	0.16	0.13

Table 3 – Robustness Testing Results

	3	6	7	8	9
Compression	35.2	0.94	0.14	0.11	
Noise	33.8	0.92	0.13	0.10	
Cropping	32.1	0.90	0.12	0.09	

3.6 Summary of Results

The results demonstrate that the proposed watermarking method maintains high embedding quality and extraction accuracy across different video resolutions, formats, and durations.

The watermark remains robust against common video processing operations, such as compression, noise addition, and cropping.

Conclusions

The proposed watermarking method, which utilizes the Discrete Cosine Transform (DCT), demonstrates notable efficacy and resilience across various video resolutions, formats, and durations. By employing DCT, the method ensures high imperceptibility, maintaining video quality while embedding the watermark in the luminance component of each frame. Extensive experimentation has shown the method's resilience against common video processing operations, such as compression, noise addition, and cropping. Its computational efficiency makes it suitable for real-time applications, achieving a balance between robustness, quality, and performance.

This method holds significant potential in the fields of digital rights management and content verification. The ability to embed imperceptible yet robust watermarks into video content provides a valuable tool for protecting intellectual property and verifying the authenticity and integrity of digital media.

The method ensures that watermarks remain intact and recoverable even after various processing operations, making it a reliable solution for content creators and distributors.

Furthermore, the implementation of this method in a publicly available Python library, named video-invisible-watermark, enhances accessibility and ease of use for developers and researchers. This facilitates the adoption of robust watermarking techniques in various digital media applications, promoting wider use and further advancements in the field.

The DCT-based watermarking method offers a robust, efficient, and high-quality solution for embedding and extracting watermarks in digital video content. Future work could focus on optimizing the method to reduce embedding and extraction times and enhancing its robustness against more advanced attacks. Integrating this method into real-time video processing systems could provide significant benefits for content creators and distributors in protecting their digital assets. This study lays a solid foundation for future research and development in digital watermarking, contributing to ongoing efforts to safeguard digital content in an increasingly multimedia-driven world.

REFERENCES

- Cox, I. J., Miller, M. L., Bloom, J. A., Fridrich, J. and Kalker, T. (2007), *Digital Watermarking and Steganography*, Morgan Kaufmann, 594 p., doi: <https://doi.org/10.1016/B978-0-12-372585-1.X5001-3>
- Potdar, V., Han, S. and Chang, E. (2005), "A survey of digital image watermarking techniques", *2005 3rd IEEE International Conference on Industrial Informatics, INDIN*, pp 709–716, doi: <https://doi.org/10.1109/INDIN.2005.1560462>
- Nikolaidis, N. and Pitas, I. (1998), "Robust image watermarking in the spatial domain", *Signal Processing*, vol. 66, is. 3, pp. 385–403, doi: [https://doi.org/10.1016/S0165-1684\(98\)00017-6](https://doi.org/10.1016/S0165-1684(98)00017-6)
- Kuchuk, H., Kovalenko, A., Ibrahim, B.F. and Ruban, I. (2019), "Adaptive compression method for video information", *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8(1), pp. 66-69, doi: <http://dx.doi.org/10.30534/ijatcse/2019/1181.22019>
- Mozhaev, O., Kuchuk, H., Kuchuk, N., Mykhailo, M. and Lohvynenko, M. (2017), "Multiservice network security metric", *2nd International Conference on Advanced Information and Communication Technologies, AICT 2017 – Proceedings*, pp. 133–136, doi: <https://doi.org/10.1109/AIACT.2017.8020083>
- Swanson, M. Zhu, B. and Tewfik, A. (1998), "Transparent Robust Image Watermarking", *Proceedings of 3rd IEEE International Conference on Image Processing*, Lausanne, Switzerland, doi: <https://doi.org/10.1109/ICIP.1996.560421>
- Langelaar, G. C., Setyawan, I. and Lagendijk, R. L. (2000), "Watermarking digital image and video data. A state-of-the-art overview", *IEEE Signal Processing Magazine*, vol. 17, no. 5, pp. 20–46, doi: <https://doi.org/10.1109/79.879337>

Received (Надійшла) 10.06.2024

Accepted for publication (Прийнята до друку) 07.08.2024

Метод вбудовування та вилучення водяних знаків у відеокадри за допомогою дискретного косинусного перетворення

М. Чжоу, Н. Кучук

Анотація. У цьому документі представлено надійний метод вбудовування та вилучення водяних знаків у відеокадри за допомогою дискретного косинусного перетворення (DCT). Цей метод забезпечує високу непомітність і зберігає якість відео при вбудовуванні водяного знаку в компонент яскравості кожного кадру. Масштабні експерименти демонструють стійкість методу до звичайних операцій обробки відео, включаючи стиснення, додавання шуму та кадрування. Запропонований підхід є обчислювально ефективним, що робить його придатним для програм реального часу. Майбутня робота спрямована на оптимізацію алгоритму для швидшої роботи та підвищення надійності проти більш складних атак. Практичні застосування включають керування цифровими правами, перевірку вмісту та криміналістичний аналіз.

Ключові слова: відео, дискрет, перетворення, цифровий водяний знак, алгоритм вбудовування.

Є. В. Шевчук, В. М. Федорченко

Харківський національний технічний університет радіоелектроніки, Харків, Україна

АНАЛІЗ ОСНОВНИХ ВРАЗЛИВОСТЕЙ І СПОСОБІВ ЗАХИСТУ МЕХАНІЗМУ КОНСЕНСУСУ В ДЕЦЕНТРАЛІЗОВАНИХ БЛОКЧЕЙН СИСТЕМАХ

Анотація. Актуальність. Захист механізму консенсусу в блокчейн системах – один з найважливіших напрямків у вдосконаленні децентралізованих блокчейн систем, цей механізм відповідає за валідацію транзакцій й підтвердженням що транзакції аутентичні, тобто цей механізм відповідає за захист ресурсів і грошей системі, також за захист від різноманітних вразливостей які створює децентралізація, тож аналіз таких вразливостей і захист механізму консенсусу є ключовою темою захисту децентралізованого блокчейну. **Метою даної роботи** є опис основних вразливостей та надання рекомендації щодо вибору підходу до захисту та побудови механізму консенсусу в децентралізованих блокчейн системах. **Об'єктом дослідження** є блокчейн. **Предметом дослідження** є механізм консенсусу децентралізованих блокчейн систем. **Результати.** В роботі були проаналізовані механізму консенсусу децентралізованих блокчейн систем, їх вразливості та методи захисту від них. **Висновок.** Механізм консенсусу децентралізованого блокчейну вірогідно не зможе бути повністю захищений ніколи через його децентралізовану природу, але можна звести ці ризики до мінімуму, наведені методи, а саме їх розумне комбінування допоможе знизити типові ризики будь яких загроз до мінімуму.

Ключові слова: блокчейн, консенсус, PoS, PoW, DPoS, VRF, double spending attack, selfish mining, censorship attack.

Вступ

Сьогодні блокчейн використовується в багатьох сферах, таких як:

- фінанси: децентралізовані фінанси (DeFi), смартконтракти, криптовалюти;
- логістика: відстеження ланцюгів поставок, запобігання контрафакту;
- охорона здоров'я: зберігання медичних записів, обмін даними;
- державне управління: цифрові ідентифікатори, голосування, прозорість урядових процесів.

Ця технологія має величезний потенціал для зміни світу. Вона може зробити нашу економіку більш ефективною, нашу систему управління більш прозорою, а наше життя – більш безпечним.

Блокчейн – це не просто криптовалюта. Це децентралізована база даних, що ведеться спільною мережею комп'ютерів. Завдяки цьому вона стає прозорою, безпечною та стійкою до цензури. Блокчейн – це не просто технологія, це нова парадигма довіри. Він дає нам можливість створювати децентралізовані системи, які не підлягають контролю жодної особи чи організації. І ця система була б неможливою без механізму консенсусу, технології, яка надає можливості створення чесної децентралізації яку як мінімум складно зламати якимось чином. Консенсус є ключовою технологією для децентралізованих блокчейн систем, він потрібен для того щоб система залишалася чесною для всіх учасників, ця система відповідає за створення нових блоків в блокчейні і за підтвердження цих блоків, отже вона відповідає за всі транзакції, всі операції в блокчейн, а отже його захист від злону є ключовою темою для захисту даних та користувачів в мережі блокчейн

Результати досліджень

1. Що таке блокчейн? Blockchain зобов'язаний своєю назвою тому, як він зберігає дані транзакцій – у блоках, пов'язаних разом, щоб утворити ланцюжок

[1–5]. З ростом кількості транзакцій зростає і блокчейн. Блоки записують і підтверджують час і послідовність транзакцій, які потім реєструються в блокчейні в межах окремої мережі, що регулюється правилами, погодженими учасниками мережі. «Кожен блок містить хеш (цифровий відбиток або унікальний ідентифікатор), пакети останніх дійсних транзакцій із мітками часу та хеш попереднього блоку. Попередній хеш блоку пов'язує блоки разом і запобігає зміні будь-якого блоку або вставці блоку між двома існуючими блоками». Теоретично цей метод робить блокчейн захищеним від втручання.

Чотири ключові концепції блокчейну:

Спільна книга. Спільна книга – це розподілена система записів, яка «тільки для додавання» використовується в бізнес-мережі. «Завдяки спільній книзі транзакції реєструються лише один раз, усуваючи дублювання зусиль, типове для традиційних бізнес-мереж».

Дозволи. Дозволи забезпечують безпеку транзакцій, їх автентифікацію та можливість перевірки. «Завдяки можливості обмежувати участь у мережі організації можуть легше дотримуватися правил захисту даних, таких як ті, що передбачені в Законі про перенесення та підзвітність медичного страхування (HIPAA)» та Загальному регламенті ЄС щодо захисту даних (GDPR).

Розумні контракти. Розумний контракт – це «угода або набір правил, які регулюють бізнес-операцію; він зберігається в блокчейні та виконується автоматично як частина транзакції».

Консенсус. Завдяки консенсусу всі сторони погоджуються на транзакцію, перевірену мережею. Блокчейни мають різні механізми консенсусу, включаючи підтвердження частки, мультипідпис і PBFT (практична візантійська відмовостійкість).

У кожній блокчейн-мережі є різні учасники, які виконують такі ролі, зокрема:

Користувачі блокчейну. Учасники (зазвичай бізнес-користувачі) з дозволами приєднуватися до мережі

блокчейн і проводити транзакції з іншими учасниками мережі. **Регулятори.** Користувачі блокчейну зі спеціальними дозволами для контролю за транзакціями, що відбуваються в мережі. **Оператори мережі блокчейн.** Особи, які мають спеціальні дозволи та повноваження на визначення, створення, керування та моніторинг мережі блокчейн. **Центри сертифікації.** Особи, які

видають і керують різними типами сертифікатів, необхідних для запуску дозволеного блокчейну. На рис. 1 можна побачити просту схему блокчейна, кожен блок має в собі хеш і хеш минулого блоку, цей хеш збирається із кореня дерева меркла яке в свою чергу перевіряє всі минулі транзакції, також воно слугує у вигляді ідентифікатора блоку транзакцій.

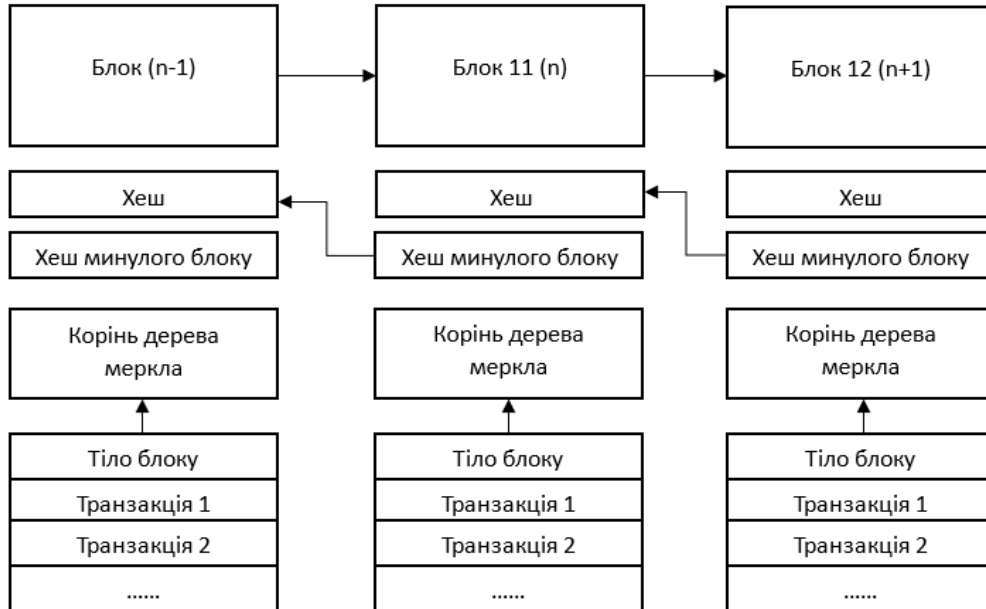


Рис. 1. Схема простого блокчейна

2. Що таке механізм консенсусу? Механізм консенсусу — це саморегульований стек програмних протоколів, записаних у код блокчейну, який синхронізує мережу для узгодження стану цифрової книги [6, 7, 8]. Це робиться шляхом підтримки єдиного набору даних — взаємно узгодженої версії історії транзакцій блокчейну — замість використання кожного вузла або комп'ютера в мережі для окремого збереження власної копії бази даних у повному обсязі. Незважаючи на те, що під час програмування мережевого стандарту верифікації слід враховувати різноманітні консенсусні механізми, кожен підхід спрямований на дискредитацію шахраїв у їхніх спробах заперечити записи.

Консенсус працює таким чином що вузли вводять дані з транзакції, що очікує на розгляд, а потім звітують із статусом схвалення або відхилення, коли запит буде перехресно перевірено з його записами. Наприклад, якщо користувач намагається обробити транзакцію, використовуючи раніше витрачені монети, які вже були враховані, цей запит буде легко відхилено щодо незмінної книги, що підтверджується несхваленням більшості. Користувачів, які не дотримуються консенсусу, часто блокують у мережі. У випадку, якщо вузол хоче оскаржити запис, йому доведеться подати запит на відкликання всієї мережі. Якщо більше ніж дві третини однорангових вузлів схвалюють, транзакція підтверджується, розповсюджується та постійно записується в блокчейн.

3. Основні вразливості механізму консенсусу блокчейн систем. На даний момент, існують три основні механізми консенсусу які блокчейни використо-

вують для створення і валідації блоків транзакцій. PoW – proof of work, широко відомий механізм в якому для того щоб підтверджувати транзакції потрібно використовувати потужності процесорів або відеокарт [9, 10], найвідоміший його представник це Bitcoin. PoS – proof of stake механізм який активно набирає свою популярність, в цьому механізмі консенсусу вузол який створює блоки, відомий як валідатор, вибирається в залежності від того яку частину активів блокчейну вони вносять в якості застави, але й це тільки дає їм шанс бути вибраними, вузли вибираються випадково, чим більше застава ти більше шанс бути вибраним і отримати винагороду [11, 12], і як можна побачити на рис. 2 вибирається не одна нода для обробки а одразу декілька.

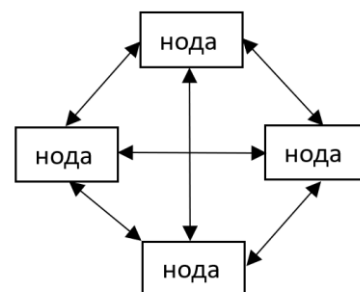


Рис. 2. PoS

Також вони перевіряють результати один одної, цей рисунок також можна віднести до PoW єдиною різницею слугує спосіб, по якому ця нода вибирається. DPoS – delegated proof of stake це механізм консенсусу який трохи нагадує PoS, але в випадку з DPoS замість

того щоб всі вузли брали участь, власники валюти обирають делегатів, які відповідають за додавання нових блоків [13, 14].

Приклад цього DPoS можна побачити на рис. 3 тут також можна побачити що існують слідувачі ноди, ці ноди нічого не записують в блок транзакцій, але сліdkують за тим щоб ці транзакції були правильними (працюють без винагородження).

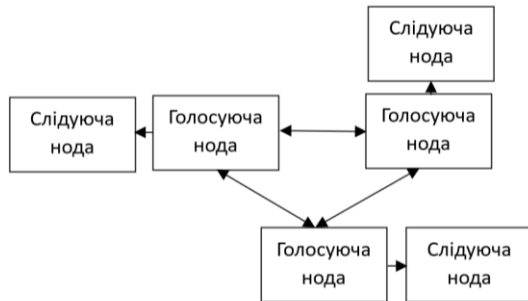


Рис. 3. DPoS

Не зважаючи на те, що в обох методах консенсусу ми отримуємо однаковий результат – валідований блок транзакції, шлях яким вони приходять до нього сильно відрізняється, і відповідно до цього шляху вони мають різні важливі вразливості.

Основні вразливості PoW:

- Атака 51% - якщо зловмисник контролює більше 50% обчислювальної потужності мережі, він може спробувати атаку 51%. Маючи мажоритарний контроль, зловмисник може переписати історію транзакцій, скасувати транзакції або подвійно витратити монети, підриваючи незмінність і безпеку блокчейну [15–17];

- Selfish mining - майнери можуть брати участь у егоїстичному майнінгу, де вони приховують розкриті блоки в мережі, щоб отримати перевагу. Це порушує справедливість процесу консенсусу та може призвести до ризиків централізації [15, 17];

- Double spending attacks - хоча PoW ускладнює подвійне витрачання, це не неможливо. Зловмисник із достатньою обчислювальною потужністю може спробувати видобути дві конфліктні транзакції одночасно, що потенційно призведе до подвійних витрат у разі успіху [17].

Основні вразливості PoS:

- Nothing at stake problem - у PoS валідаторам не потрібно витрачати ресурси (як у PoW), щоб брати участь у процесі консенсусу. Це відкриває двері до проблеми «Nothing at stake», коли валідатори можуть підтримувати кілька конфліктуючих ланцюжків, полегшуючи виконання мережних розгалужень і атак [17];

- Sensorship attacks - заможні валідатори потенційно можуть вступити в змову, щоб цензурувати певні транзакції або надавати перевагу певним учасникам, оскільки вони мають більшу частку в безпеці мережі [15, 17];

- Stake Grinding - зловмисники потенційно можуть маніпулювати процесом відбору для створення блоку, змінюючи свою ставку або участь, підриваючи справедливість і безпеку механізму консенсусу [15, 16].

Основні вразливості DPoS:

- Централізація - влада зосереджена в руках невеликої кількості делегатів. Це може зробити систему більш сприятливою до різних типів атак [18];

- Зловживання владою – делегати можуть зловживати своєю владою, наприклад цензуруючи транзакції (ця проблема є як в PoS так і в DPoS) [19];

- Атака змови – делегати можуть змовитися з метою маніпулювання системою [20].

4. Способи захисту механізму консенсусу блокчейн систем Одним із варіантів для захисту механізму консенсусу є зміна його алгоритму, наприклад з PoW на PoS, але потрібно пам'ятати що такий підхід достатньо складно реалізувати, і можна завжди отримати інші вразливості нового механізму консенсусу [21]. Також важливим підходом до захисту систем є створення великих покарань для учасників системи які провели або намагаються провести атаку на систему, це сильно збільшить ризики для зловмисників, і відіб'є в них бажання атакувати систему, адже навіть неуспішна атака, а просто спроба зробити її призведе до колосальних збитків і зловмисники втратять всі свої ресурси. Також важливим аспектом є постійний аудит системи, він допомагає виявляти виникаючі вразливості, атаки які готуються, а також можливі маніпуляції в системі в минулому [22]. Створення централізованого нагляду, призведе до більшої централізації, але й до більшої безпеки, якщо блокчейн сильно пов'язаний з банківською справою або фінансами це може бути необхідним, це рішення можна використати для запобігання майже всіх існуючих вразливостей.

Захист PoW механізму консенсусу від атак 51% є збільшення часу затримки підтвердження блокчейну, цей варіант збільшує кількість часу для виявлення таких атак, і збільшує вартість таких атак, але також це збільшує час обробки всіх транзакцій, що призведе до падіння продуктивності блокчейна і збільшення вартості транзакції [23]. Для систем які щойно розвернулися і ще не набрали велику кількість валідаторів є можливість взяти на себе частину валідації, можливо навіть тимчасово взяти 51%, це призведе до більшої централізації, але на ранніх етапах до більшої безпеки.

Одним із найкращих способів захисту від selfish mining є зміна систем винагородження майнерів таким чином, що вигоди від використання цієї вразливості буде менше ніж в звичайних чесних майнерів.

Для того щоб захиститися від double spending атаки можна додати одноразові криптографічні записи які називаються попсе, значення цих записів має бути створене до того як блок буде добутий, його значення можна використати тільки один раз, такі записи слугують для того щоб переконатися в унікальності блока, так само можна додати часові мітки на блоки, вони також слугують для підтвердження унікальності добутого блока [24].

Захиститися від nothing at stake можна впровадивши систему репутації валідатора, наприклад збільшення нагород для чесних валідаторів з гарною репутацією, або збільшення шансу що їх оберуть для валідації блоку, і протилежно для цього знижувати

нагороди і шанс на валідацію валідаторам з низькою репутацією, також можна додати штрафи для валідаторів які будуть помічені в конфліктуючих форках. Створення чекпоінтів це періодичне збереження блокчейну в певному моменті, всі валідатори мають будувати блокчейн далі тільки з цього чекпоінту, що унеможливить підтримку конфліктних форків.

Захистом від *sensorship attacks* є використання спеціальних нод стеження, які будуть слідкувати за валідаторами і створювати звітність вірогідних порушень, створення методів валідації які будуть порівнювати транзакції з найдовшим ланцюжком блоків також допоможе уникнути *long range attacks*.

Від *stake grinding* захистом може стати використання перевіреної випадкової функції (VRF) для генерації випадковості блоків. VRF дозволяє власникам монет генерувати випадкові значення так, що ніхто крім власника не зможе створити їх за допомогою його приватного ключа, при цьому всі інші зможуть перевірити правильність згенерованого значення.

Захист від централізації в DPoS механізмі консенсусу можна впровадити використанням алгоритмів з множинними делегатами, це розподіляє владу між більшою кількістю учасників, і робить систему більш стійкою до атак, таким чином навіть учасники з найменшою долею зможуть приймати певну участь у голосуванні.

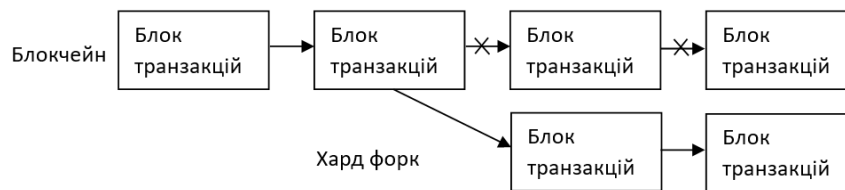


Рис. 4. Хард форк

Треба пам'ятати що головна вразливість механізму консенсусу це той факт що він контролюється не однією організацією, а великою групою людей, це як його головна особливість так і головний недолік, не зважаючи на те що абсолютного захисту на разі не існує, можна впровадити комбінацію захисних мір які зможуть не впливаючи на швидкість блокчейна збільшити його безпеку.

Важливо зменшити кількість вразливостей до мінімуму, бо кожна така вразливість збільшує можливість для створення інших атак, роблячи атаки таких типів дешевшими, а вразливість блокчейну більшою, найважливішими мірами для захисту від всіх типів атак є збільшення покарань в системі, наприклад втрата всіх вкладених монет зловмисником, проведення постійного аудиту системи, для того щоб виявляти атаки і вразливості завчасно і могли покарати зловмисника який ще не встиг завершити атаку, але вже втрапить всі ресурси, що сильно збільшить ризики для них, якщо ж блокчейн пов'язаний з банківською інфраструктурою та фінансами можливим рішенням є більша централізація в руках керуючої компанії.

З методів захисту наведених вище можна вивести комбінацію технологій захисту, необхідно додати максимальні покарання за зловмисну діяльність, або за підготовку такої діяльності, в випадку з PoW

Захиститися від зловживання владою можна за допомогою створення механізмів відкликання, це дозволить користувачам видаляти делегатів, які зловживають своїми повноваженнями і використати смарт контракти для автоматизації виконання правил, і ускладнення зловживання повноваженнями.

Від атак змови можна захиститися використовуючи наведені вище системи репутації і порівнянням транзакцій з найдовшим кодом, це допоможе виявити зловмисних валідаторів завчасно.

Якщо все ж таки зловмисники змогли успішно атакувати ваш блокчейн, вихід все ще існує і це хард форк, як зображено на рис. 4. Хард форк можна зробити якщо валідатори погодяться в випадку з PoW це майнери, а у випадку з PoS і DPoS це стейк холдери, що чинний ланцюг має бути перетертий, то можна перестати вважати дійсним ланцюжок транзакцій в якому є сліди активності зловмисників, і почати новий, чистий, але якщо ми зробимо так, то всі транзакції в минулих блоках просто перетруться, а там розміщені не тільки транзакції зловмисників, а всі транзакції які були створені в цьому блокчейні в певний час, для цього також є вихід, можна взяти транзакції не пов'язані зі зловмисниками, і інтегрувати їх вже в новий ланцюжок, репутаційні збитки і можливо реальні залишаться, але будуть значно зменшені.

це буде втрата всіх винагород і повна ануляція балансу пула зловмисників, також заборона працювати в вашому блокчейні для цього пула, в випадку з PoS і DPoS втрата всього стейку. Постійний аудит є дещо дорогою але необхідною мірою для виявлення і запобігання атак які готуються, для виявлення вразливостей до того як ними почнуть користуватися, а також маніпуляцій в системі в минулому.

Висновки

Механізм консенсусу децентралізованого блокчейну вірогідно не зможе бути повністю захищений ніколи через його децентралізовану природу, але можна звести ці ризики до мінімуму, наведені методи, а саме їх розумне комбінування допоможе знизити типові ризики будь яких загроз до мінімуму. Важливо пам'ятати що кожна загроза залишена в блокчейні несе ризики спрощення створення інших загроз, наприклад *sensorship attacks* можуть запобігати виявленню як *stake grinding* так і *nothing at stake*.

Потрібно пам'ятати що через децентралізацію можна створити разом з учасниками хард форк, і вичистити всі зміни які зробив зловмисник, але це не означає що не постраждає як репутація цієї блокчейн системи так і вірогідно зловмисники зможуть заробити на ній, що призведе до збитків всіх інших

учасників системи. Важливо пам'ятати що в таких системах ідеальний захист, принаймні поки що неможливий, але варто робити все можливе для збільшення захисту блокчейну від якомога більшої кількості атак.

Треба також відмітити що блокчейн потрібно постійно вдосконалювати, адже постійно з'являють-

ся нові вразливості і методи захисту від них, адже ідеального захисту від атак не існує.

І ще однією важливою темою є праця розробників разом з спільнотою учасників децентралізованої системи, це допоможе забезпечити максимальну швидкість реакції на атаку, або швидко знайти вразливість.

СПИСОК ЛІТЕРАТУРИ

1. What is blockchain URL: <https://www.ibm.com/topics/blockchain>
2. Blockchain Facts: What Is It, How It Works, and How It Can Be Used URL: <https://www.investopedia.com/terms/b/blockchain.asp>
3. Understanding Blockchain Technology URL: <https://www.forbes.com/advisor/investing/cryptocurrency/what-is-blockchain/>
4. What is Blockchain Technology URL: <https://www.coindesk.com/learn/what-is-blockchain-technology/>
5. How does blockchain work URL: <https://online.stanford.edu/how-does-blockchain-work>
6. What Is a Consensus Mechanism URL: <https://builtin.com/blockchain/consensus-mechanism>
7. What Are Consensus Mechanisms in Blockchain and Cryptocurrency URL: <https://www.investopedia.com/terms/c/consensus-mechanism-cryptocurrency.asp>
8. Consensus Mechanisms In Blockchain: A Deep Dive Into The Different Types URL: <https://hacken.io/discover/consensus-mechanisms/>
9. What Is Proof-of-work (PoW)? All You Need to Know URL: <https://blockworks.co/news/what-is-proof-of-work>
10. What Is Proof of Work (PoW) in Blockchain URL: <https://www.investopedia.com/terms/p/proof-work.asp>
11. What Does Proof-of-Stake (PoS) Mean in Crypto URL: <https://www.investopedia.com/terms/p/proof-stake-pos.asp>
12. What is proof of stake URL: <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-proof-of-stake>
13. What is Delegated Proof of Stake (DPoS)? Your Comprehensive Guide to DPoS URL: <https://medium.com/unicorn-ultra/what-is-delegated-proof-of-stake-dpos-your-comprehensive-guide-to-dpos-07fd5185b108>
14. What Is Delegated Proof-of-Stake (DPoS) URL: <https://www.ledger.com/academy/what-is-delegated-proof-of-stake-dpos>
15. Blockchain Common Vulnerability List URL: <https://github.com/slowmist/Cryptocurrency-Security-Audit-Guide/blob/main/Blockchain-Common-Vulnerability-List.md>
16. Blockchain Security: Common Vulnerabilities and How to Protect Against Them URL: <https://hacken.io/insights/blockchain-security-vulnerabilities/>
17. Blockchain Vulnerabilities and Attacks URL: <https://www.linkedin.com/pulse/blockchain-vulnerabilities-attacks-yeshwanth-n/>
18. Semenov, S., Mozhaiev, O., Kuchuk, N., Mozhaiev, M., Tiulieniev, S., Gnusov, Yu., Yevstrat, D., Chyrva, Yu. & Kuchuk, H. (2022), "Devising a procedure for defining the general criteria of abnormal behavior of a computer system based on the improved criterion of uniformity of input data samples", *Eastern-European Journal of Enterprise Technologies*, 2022, 6 (4(120)), pp. 40–49, doi: <https://doi.org/10.15587/1729-4061.2022.269128>
19. Kovalenko, A., Kuchuk, H., Kuchuk, N. and Kostolny, J. (2021), "Horizontal scaling method for a hyperconverged network", 2021 International Conference on Information and Digital Technologies (IDT), Zilina, Slovakia, doi: <https://doi.org/10.1109/IDT52577.2021.9497534>
20. Petrovska, I. and Kuchuk, H. (2023), "Adaptive resource allocation method for data processing and security in cloud environment", *Advanced Information Systems*, vol. 7(3), pp. 67–73, doi: <https://doi.org/10.20998/2522-9052.2023.3.10>
21. Datsenko, S., and Kuchuk, H. (2023), "Biometric authentication utilizing convolutional neural networks", *Advanced Information Systems*, vol. 7, no. 2, pp. 67–73. Doi: <https://doi.org/10.20998/2522-9052.2023.3.10>
22. 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
23. 51% Attack: The Concept, Risks & Prevention URL: <https://hacken.io/discover/51-percent-attack/>
24. Understanding Double-Spending and How to Prevent Attacks URL: <https://www.investopedia.com/terms/d/doublespending.asp>

Received (Надійшла) 11.04.2024

Accepted for publication (Прийнята до друку) 12.06.2024

Analysis of the main vulnerabilities and ways of protection of the consensus mechanism in decentralized blockchain systems

Eugene Shevchuk, Volodymyr Fedorchenko

Abstract. Topicality. The protection of the consensus mechanism in blockchain systems is one of the most important directions in the improvement of decentralized blockchain systems, this mechanism is responsible for validating transactions and confirming that transactions are authentic, that is, this mechanism is responsible for protecting resources and money in the system, as well as protecting against various vulnerabilities created by decentralization. so analyzing such vulnerabilities and protecting the consensus mechanism is a key topic of decentralized blockchain security The goal of this work is a description of the main implications and recommendations for choosing an approach to protection and promoting a consensus mechanism in decentralized blockchain systems. The subject of the research is the impact of the consensus mechanism in decentralized blockchain systems. The subject of research is the consensus mechanism of decentralized blockchain systems. **Results.** The work analyzed the consensus mechanism of decentralized blockchain systems, their differences and methods of protection from them. **Conclusions.** The consensus mechanism of a decentralized blockchain is unlikely to be susceptible to theft through its decentralized nature, but it is possible to reduce the risks to a minimum by introducing methods, and their reasonable combination to help reduce the type of risks and any threats to a minimum.

Keywords: blockchain, consensus, PoS, PoW, DPoS, double spending attack, selfish mining, censorship attack.

О. В. Шматко¹, Д. В. Кулініч¹, Т. В. Горбач²

¹ Харківський національний технічний університет радіоелектроніки, Харків, Україна

² Національний технічний університет «Харківський політехнічний інститут», Харків, Україна

РОЗРОБКА ТА ДОСЛІДЖЕННЯ АРХІТЕКТУРНОЇ МОДЕЛІ СИСТЕМИ ОБМІНУ ПЕРСОНАЛЬНИМИ ДАНИМИ НА ОСНОВІ БЛОКЧЕЙН

Анотація. Актуальність. Сучасне суспільство стикається з зростаючою потребою у безпечному, на-дійному та прозорому обміні персональними даними пацієнтів у сфері охорони здоров'я. Захист конфіденційності та цілісності медичної інформації є пріоритетом для забезпечення якісного та ефективного медичного догляду. Блокчейн-технології надають обіцяючий інструмент для вирішення цієї проблеми, дозволяючи створити децентралізовану та безпечну систему обміну персональними даними пацієнтів. **Метою даної роботи** є забезпечення високого рівня безпеки та конфіденційності медичних даних, а також підвищення ефективності процесів у сфері охорони здоров'я за рахунок розробки програмних компонентів системи обміну персональними даними пацієнтів на основі блокчейн-технологій. **Об'єктом дослідження** є система обміну персональними даними пацієнтів у сфері охорони здоров'я. **Предметом дослідження** є програмні компоненти, що базуються на блокчейн-технологіях, призначені для забезпечення безпеки, прозорості та ефективності обміну медичною інформацією. **Результати.** У даній роботі запропоновано архітектурну модель безпечної та ефективної системи обміну медичними даними, яка може бути широко впроваджена у сфері охорони здоров'я. **Висновок.** Впровадження системи безпечного обміну персональними даними на основі технологій блокчейн у сфері охорони здоров'я допоможе покращити якість медичного обслуговування та забезпечити швидкий доступ до важливих даних для медичного персоналу. Теоретична значимість полягає у розширенні знань щодо застосування блокчейн-технологій у галузі охорони здоров'я та питань безпеки та конфіденційності медичної інформації. Це дослідження може бути основою для подальших досліджень у цій галузі та сприяти розвитку нових методів та підходів до обміну медичними даними.

Ключові слова: блокчейн, персональні дані пацієнтів, IoT, смарт-контракти, Ethereum, модель системи обміну медичними даними.

Вступ

Медичні дані містять багато записів даних про пацієнтів, які важливі для подальшого лікування та майбутніх досліджень. Однак для захисту конфіденційності даних їх необхідно надійно зберігати і надавати спільний доступ до них. Блокчейн широко використовується в управлінні медичними даними завдяки своїм децентралізованим функціям і захисту від несанкціонованого доступу.

Медичні дані актуальні для всіх. У них записується фізична інформація про наш організм. Це важливо для діагностики та лікування захворювань [1]. З швидким розвитком штучного інтелекту медичні дані стали великим надбанням. Це може допомогти нам створити діагностичні моделі зі штучним інтелектом та допомогти лікарям у діагностиці. Хоча запис медичної інформації еволюціонував від початкових паперових записів до електронних медичних записів (EMR), які є більш зручними для доступу та зберігання даних, необхідно приділяти більше уваги захисту конфіденційності даних [2]. Багато лікарень та установ зменшили передачу та обмін даними, щоб уникнути витоку конфіденційних даних, що призвело до утворення розрізнених даних, оскільки медичні дані розкидані по різних закладах охорони здоров'я [3].

Конфіденційність та безпека медичних даних також призводять до інших проблем. Наприклад, для безпеки пацієнтів необхідно повторно обстежувати кожного разу, коли вони потрапляють до нової лікарні. Така поведінка призводить до марної втрати енергії і грошей. З метою захисту конфіденційності пацієнтів медичні дані не можуть передаватися науковим установам, що перешкоджає розвитку медицини. Це

спонукало до пошуку безпечних методів зберігання і передачі даних, і блокчейн широко використовується, завдяки своїй децентралізованій природі, захищеної від несанкціонованого доступу, для обміну медичними даними [4]. Сучасне суспільство стикається з зростаючою потребою у безпечному, надійному та прозорому обміні персональними даними пацієнтів у сфері охорони здоров'я. Захист конфіденційності та цілісності медичної інформації є пріоритетом для забезпечення якісного та ефективного медичного догляду. Блокчейн-технології надають обіцяючий інструмент для вирішення цієї проблеми, дозволяючи створити децентралізовану та безпечну систему обміну персональними даними пацієнтів.

Об'єктом дослідження є система обміну персональними даними пацієнтів у сфері охорони здоров'я. Предметом дослідження є програмні компоненти, що базуються на блокчейн-технологіях, призначені для забезпечення безпеки, прозорості та ефективності обміну медичною інформацією.

Метою цієї роботи є забезпечення високого рівня безпеки та конфіденційності медичних даних, а також підвищення ефективності процесів у сфері охорони здоров'я за рахунок розробки програмних компонентів системи обміну персональними даними пацієнтів на основі блокчейн-технологій.

Основна частина

У зростаючому світі технологій речі навколо нас стають розумніші, ніж ми думаємо. Такі галузі, як охорона здоров'я, також є революціонерами завдяки новітнім технологіям. У міру розвитку технологій якість та ефективність галузі охорони здоров'я також швидко ростуть. І лікарі, і пацієнти отримують переваги від

технологічного прогресу в галузі охорони здоров'я. Тепер ми отримуємо лабораторні звіти, МРТ і комп'ютерну томографію за менший час і є більш ефективними і точними, ніж раніше. Цифрові рентгенівські знімки-революційний спосіб поглянути на переломи та пухлини в кістках, а також Цифрове зберігання медичних записів відкривають новий спосіб догляду за пацієнтами з використанням технологій глибокого навчання і штучного інтелекту. Крім того, завдяки технологічним досягненням можливий постійний віддалений моніторинг пацієнтів та збір даних від пацієнтів у режимі реального часу за допомогою датчиків ІОТ, а також виконання аналізу без затримок [1].

Тепер ми можемо точніше прогнозувати важкі захворювання (наприклад, рак) і призначати ліки на дуже ранній стадії. Хоча зберігання медичних даних в цифровому вигляді дає багато переваг, воно також відкриває двері для забезпечення безпеки загрози та втрата даних. Як ми знаємо, медичні дані є критично важливими даними, вони складаються з конфіденційної і чутливої інформації, що відноситься до пацієнтів.

Отже, нам потрібен надійний механізм для підтвердження цілісності, конфіденційності безпеки медичних даних. Інтеграція технологій блокчейн з галузю охорони здоров'я може вирішити проблеми, пов'язані з цілісністю та безпекою даних. Тепер ми можемо більш ефективно і безпечно обмінюватися даними про пацієнтів, пов'язаними зі здоров'ям, з лікарями і постачальниками медичних послуг.

Спочатку (у 70-х роках) система охорони здоров'я називалася healthcare 1.0. В охороні здоров'я відчувалася гостра нестача ресурсів і обмежувалася можливість взаємодії з цифровими системами. Витрати і час були збільшені через відсутність вбудованих біомедичних датчиків, коли медичні компанії протягом цього періоду перейшли на паперові рецепти та звітність. Концепція систем охорони здоров'я виникла в 1991 році до 2005 року з healthcare 2.0. На цьому етапі використовувалося цифрове відстеження, що дозволяє лікарям використовувати обладнання для візуалізації для вивчення стану здоров'я пацієнта. З впровадженням інтернет-платформи постачальники медичних послуг почали створювати онлайн-спільноти і використовувати хмарні сервери для зберігання інформації про пацієнтів, що забезпечило повсюдний доступ як для пацієнта, так і для практикуючого лікаря. Healthcare 3.0, породила концепцію користувальницької настройки медичних карт пацієнтів. Нові користувальницькі інтерфейси забезпечують індивідуальний і оптимізований досвід роботи. На додаток до цих досягнень були впроваджені системи медичної документації, які дозволяють відстежувати медичні дані пацієнтів в режимі реального часу і на універсальному рівні. Аналогічним чином, поряд із системами EHR, такими як HL7, які були інтегровані для зберігання інформації про пацієнтів, почали з'являтися автономні мережеві системи, такі як канали соціальних мереж. Це скоротило обмін медичними даними, будь то в мережі або між клініцистами, які використовують HL7. Цей методи також покращили здатність взаємодіяти та комунікувати з пацієнт. Ера охорони здоров'я 4.0 почалася в 2016 році і триває по сьогодні [3]. За цей час було застосовано ряд різних

технологій, включаючи туманні обчислення, прикордонні обчислення, Хмарні обчислення, Інтернет речей, просунуту аналітику, штучний інтелект і машинне навчання, а також блокчейн, щоб перетворити його в інтелектуальну систему охорони здоров'я або Індустрія охорони здоров'я 4.0. Основна увага була приділена носимим датчикам стану здоров'я.

Innoplexus поєднує в собі штучний інтелект і блокчейн для забезпечення безперервного сканування глобальних даних науки про життя [5]. Система надає дані науково-дослідним інститутам і фармацевтичним компаніям. BlockRx - це платформа, яка успішно використовується в реальних додатках [6]. Платформа поєднує в собі технологію блокчейн і передову технологію цифрової бухгалтерської книги iSolve. Платформа об'єднує медичні дані з біомедичних та науково-дослідних інститутів. BlockRx був застосований на практиці і домігся значного розвитку.

Було опубліковано кілька статей, в яких узагальнюються моделі, засновані на блокчейне. Джин та ін. аналізують конфіденційність обміну медичними даними за допомогою типу блокчейна, використовуваного в моделі [7]. Огляд ділить блокчейни на дві категорії: без дозволів і з дозволенним доступом. Потім аналізуються переваги та недоліки залежно від типів блокчейнів. Лейлі та ін. проаналізували ряд робіт, опублікованих у період з 2016 по 2020 рік [8]. Ця стаття присвячена ситуаціям застосування в охороні здоров'я і не фокусується в першу чергу на порівнянні та узагальненні моделей. Саха та ін. узагальнили деякі підходи до охорони здоров'я, засновані на блокчейне, але вони не порівнюють ці підходи [9]. Ісраа та ін. провели аналіз моделі з унікальної точки зору, розглядаючи як переваги, так і загрози, які технологія представляє пацієнтам [10]. Хассельгрена та ін. провели статистичний аналіз опублікованих робіт. Однак у цьому огляді не було узагальнено методів [11]. Сюй та ін. в основному аналізують застосування блокчейна в медичних даних про онкологію, таких як відстежуваність ліків і обмін даними про онкологію [12].

Блокчейн-це децентралізована розподілена технологія (DDT) [16]. У блокчейні колекція записів, які закривають обмін або передачу цінностей та цифрових активів, таких як транзакції, товари та послуги, розробляється та управляється розподіленою системою обчислювальних вузлів у одноранговій мережі. Блокчейн походить від біткоіни, технології, що представляє собою розподілену базу даних з постійно зростаючими записами, що розглядаються як блок, і ці записи не можуть бути змінені [19]. Основна ідея блокчейна полягає в стабілізації цілісності, відстежуваності і підзвітності спільно використовуваних даних. Розподілена книга обмежує методи, включаючи збереження та автентифікацію, які виконуються в мережі взаємодіючих вузлів. Ці вузли впроваджують програмне забезпечення для аудиту, яке узгоджує зображення спільної книги між одноранговою мережею акціонерів, представляючи всі підзвітні дії за допомогою цифрових відбитків пальців або хеш-кодів. Книга класифікується як розповсюджена та визначається під час запису даних. У блокчейні у кожного учасника вузла є своя загальна бухгалтерська книга. Він генерує

прозорий, незмінний запис [20]. Журнали блокчейна забезпечують точність для прийняття повідомлень в ІТ-середовищі health, а журнали аудиту - для подальших запитів про такі дозволи і продуктивності моделей доступу. Виходячи з цієї функціональності, фреймворк працює як послідовний опис авторизації доступу до електронної медичної інформації (ЕІІ). За останнє десятиліття дослідники впровадили кілька систем управління охороною здоров'я, заснованих на блокчейні, для забезпечення різних цілей безпеки [21, 22]. Блокчейн гарантує, що дані не були підірвані в результаті шкідливих атак, і перевіряє безліч аспектів Походження даних [23]. Ця технологія використовує криптографічні методи, а розподілене середовище мережі блокчейн забезпечує поширення всієї інформації, що забезпечує видимий, заслуговує довіри цифровий відбиток пальця і перевіряються шляху [24].

Існує два основних види блокчейна: безстроковий і дозволений блокчейн. Публічний блокчейн також називають блокчейном без прав доступу. Першим винаходом блокчейна без прав доступу є біткоїн. Блокчейн без дозволів легкодоступний і відкритий для дій з читання і запису всіма учасниками системи [25]. Це означає, що кожен може брати участь у системі з псевдонімною ідентифікацією. Користувач також може читати інформацію або транслювати її в ефір і ідентифікується як частина механізму консенсусу [26, 27]. Ethereum також застосовує блокчейн без дозволів, і будь-який бажаний може розробляти і комбінувати смарт-контракти по мережі без будь-яких обмежень з боку розробників. Дозволений блокчейн також називають приватним блокчейном. Окрема організація використовує дозволений блокчейн [28]. На відміну від блокчейна без дозволів, блокчейн з дозволами розроблений таким чином, що учасники мережі заздалегідь визначені для дій читання/запису і назавжди ідентифікуються всередині системи. Отже, основна відмінність між блокчейном без прав доступу та блокчейном з дозволами полягає в тому, як користувач може отримати доступ до мережі. У дозволений блокчейн-мережі впровадьте візантійську відмовостійкість (BFT) [29]. Структура Hyperledger розроблена таким чином, щоб забезпечити безпеку технології Загального реєстру і розширити можливості дозволених користувачів.

Hyperledger Fabric-це тип дозволеної блокчейн-технології, яка працює на основі блокчейн-підприємства з відкритим вихідним кодом, підтримуваного Linux Foundation [30]. Hyperledger-це постійно поширений колективний або приватний блокчейн, який намагається вдосконалити технологію блокчейн за допомогою галузевих додатків. Як правило, Hyperledger Fabric - це розподілена мережа, що формулює однорангову систему, де кожен одноранговий вузол має репліковану, узгоджену копію структури даних блокчейна, зокрема, ланцюговий Індекс транзакції, що описує виклик і виконання ланцюгових кодів. Hyperledger Fabric дає можливість розширити спектр застосування технології блокчейн за межі кріптовалютних угод, які розрізняють різні області застосування реляційних баз даних, включаючи управління медичною інформацією [31].

Linux Foundation підтримувала проекти Hyperledger Fabric, одним з таких прикладів є Hyperledger

Composer. Архів бізнес-мережі (BNA) - це функціональна розробка Hyperledger Composer, яка успадкована від блокчейна Hyperledger Fabric [15].

Бізнес-мережа включає в себе учасників, і вони об'єднуються за допомогою їх ідентифікації, а також активів, які генеруються в системі; транзакції визначають обмін активами. Ці правила передбачають виконання транзакцій, званих смарт-контрактами, і в кінцевому підсумку всі транзакції зберігаються в бухгалтерській книзі. Малюнок 1 ілюструє загальну архітектуру Hyperledger Composer. Файл моделі містить три основні компоненти: учасники, активи та транзакції. Учасники є кінцевими користувачами системи і можуть мати справу з активами та взаємодіяти з іншими за допомогою транзакцій. Активи, як правило, є змінними, збереженими в мережі. Транзакції є цілями системи і викликаються для оновлення налаштувань. Файл сценарію в бізнес-мережі визначає багато функцій транзакцій у системі. Він складається з Java Script (JS) і має справу з бізнес-логікою, яка визначає, які стандарти діють для користувачів і які типи ресурсів є загальними. Список контролю доступу (ACL) описує різні діапазони доступу, якими володіють учасники в мережі. У файлі ACL фіксується мета учасників, що визначає їх ефективність при створенні, читанні, оновленні або видаленні ресурсів. Файл запиту пояснює склад та використання запитів із системи. Вони залишаються фіксованими для екстраполяції транзакцій журналу, який містить записи всіх попередніх транзакцій у мережі. Архівний запис-це список реєстру, наданий архівним записом, який включає історію транзакцій та подій, виконаних у системі. Поки транзакція обробляється, запис журналу оновлюється, зберігаючи історію всіх транзакцій всередині бізнес-мережі. Учасники з їх ідентифікаційними даними беруть участь у відправці транзакцій, а ресурси записів журналу можуть бути вилучені за допомогою запитів composer для запиту певних записів.

Блокчейн можна розділити на публічний ланцюжок, приватний ланцюжок і консорціумний ланцюжок в залежності від способу участі [15]. Публічна мережа, як впливає з назви, є повністю загальнодоступною та доступною для всіх. Оскільки дані в ланцюжку не можуть бути змінені, публічні ланцюги вважаються повністю децентралізованими. Ланцюжок консорціуму обмежений лише авторизованими учасниками для участі, а дозволи на читання та запис та дозволи на облік участі в блокчейні формулюються відповідно до правил Альянсу. Приватний ланцюжок використовується тільки в приватних організаціях, а дозволи на читання і запис в блокчейн і дозволу на участь в бухгалтерському обліку формулюються відповідно до правил приватної організації. Беруть участь вузлів небагато, і вони строго обмежені [16]. У табл. 1 порівнюються різні типи блокчейнів.

Як децентралізована однорангова система, вузли отримують транзакції в іншому порядку [17]. Отже, необхідні послідовні алгоритми для забезпечення того, щоб вузли узгоджували транзакції. Proof of work (POW) - це перший успішний децентралізований блокчейн-алгоритм консенсусу. Для вирішення візантійської проблеми було запропоновано практичний алгоритм Візантійська відмовостійкість

(PBFT) [18]. Це гарантує, що блокчейн все ще може нормально функціонувати з деякими несправними або шкідливими вузлами.

Таблиця 1 – Порівняння різних типів блокчейнів

Тип блокчейну	Децентралізація	Пропускна здатність	Витрати	Масштабованість
Державна мережа	Висока	Низька	Високі	Погана
Мережа консорціумів	Середня	Середня	Середні	Відмінна
Приватна мережа	Низька	Висока	Низькі	Відмінна
Гібридна мережа	-	-	Низькі	Велика

Смарт-контракти - це комп'ютерні протоколи, які поширюють, перевіряють або приводять у виконання контракти інформаційним способом [19]. Смарт-контракти не вимагають аутентифікації третьою стороною, а успішні транзакції відстежуються і незворотні. Для складання юридично дійсного контракту використовується комп'ютерна програма, і цей контракт може бути виконаний автоматично. Смарт-контракт-це код, розгорнутий на блокчейні, який гарантує безпеку транзакцій без нагляду Третіх Сторін [20]. Процес укладення смарт-контракту показаний на рис. 1.

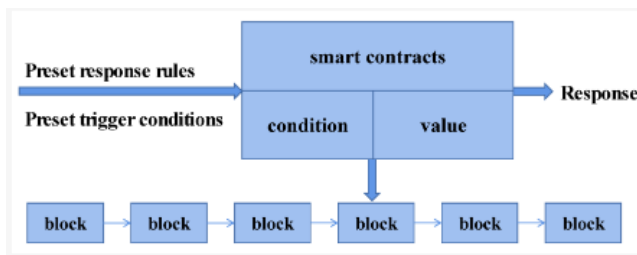


Рис. 1. Системна модель. Джерело: [20]

1. Децентралізація: виконання смарт-контрактів не повинно залежати від участі або втручання сторонніх організацій, а нагляд і арбітраж контрактів здійснюються комп'ютерами;

2. Незмінність: як тільки смарт-контракт розгорнутий, весь вміст не може бути змінено. Це чимось схоже на контракт в традиційному світі, який не може бути змінений після його підписання;

3. Низька вартість: оскільки смарт-контракти не вимагають контролю з боку стороннього посередника, як тільки відбувається порушення контракту, код вводиться в дію і має набагато меншу вартість у порівнянні з традиційними контрактами;

4. Відкритість і прозорість: після успішного розгортання смарт-контракт буде працювати відповідно до дизайну-кодом і може бути переглянутий будь-яким користувачем з високим ступенем прозорості [21].

Існує багато ситуацій, в яких блокчейн використовується для обміну медичними даними, і зараз існує три типи в залежності від сценаріїв застосування. Перший-це безпечне зберігання даних і доступ до них на основі блокчейна. Другий-використання блокчейна в поєднанні з ІОМТ. Третє-використання

блокчейна для заміни центрального федерального навчального закладу.

Поява EMR принесла зручність, а також проблеми конфіденційності. Через проблеми безпеки медичні дані не можуть передаватися вільно. Було запропоновано кілька моделей, заснованих на блокчейні [22].

Заснована на блокчейні модель 'medichain' була запропонована Rahul et al. [23]. Ця модель використовує блокчейн як базу даних для зберігання повної інформації про випадок пацієнта в блоці. Записи транзакцій хешуються для зберігання отриманих хеш-значень у дереві Merkle, щоб забезпечити безпеку даних та запобігти підробці, тим самим зменшуючи помилки при прийнятті клінічних рішень. Щоб вирішити проблему широкого спектру джерел та різноманітних структур медичних даних, дані з усіх полів об'єднуються в єдиний гіперпростір, що зберігається у запропонованій структурі. У цьому методі використовується ланцюгове зберігання. Однак блокчейн менш масштабований. Зберігання даних у мережі також коштує дорого. Wu впроваджує орієнтовану на пацієнта модель контролю доступу із збереженням конфіденційності в процесі контролю доступу до приватної інформації в системах охорони здоров'я [24]. Потім технологія блокчейн використовується для створення приватної платформи зберігання інформації, а для реалізації передачі інформації використовуються стандартні криптографічні алгоритми. У цьому процесі конфіденційна інформація також захищається договором авторизації файлів для подальшого запобігання крадіжці медичної конфіденційної інформації. Модель пропонує детальний метод контролю доступу із збереженням конфіденційності, який надає користувачам різні привілеї на основі оцінки їх типів. Інформація про EMR зберігається в хмарній базі даних і розміщується сторонньою організацією, що надає хмарні сервіси. Коли дані зберігаються в хмарі, генерується хеш цих даних. Потім хеш зберігається в блокчейні. Коли дані в хмарі підробляються, їх можна порівняти за хеш-значенням у ланцюжку. У цій моделі консенсусним алгоритмом є POW, який вимагає великої кількості неприпустимих обчислень вузлами. Liu et al. запропонували полегшену модель на основі блокчейна для обміну та захисту медичних даних [25]. Модель використовує технологію повторного шифрування через проксі для забезпечення обміну даними між лікарями в різних лікарнях. Використовувану хеш-функцію важко зіставити. Таким чином, збережену медичну інформацію практично неможливо підробити. Традиційне делеговане підтвердження зацікавленості вдосконалено для отримання нового алгоритму консенсусу, який є більш безпечним і надійним. Розроблено механізм зіставлення захворювань, що дозволяє пацієнтам, які страждають одним і тим же захворюванням, спілкуватися один з одним. Після взаємної аутентифікації сеансові ключі можуть бути встановлені між пацієнтами. Цей механізм може допомогти пацієнтам обмінюватися інформацією про захворювання. Приватна мережа швидка в транзакціях, але менш децентралізована. Вона більше підходить для додатків всере-

дині компаній або установ. Це не застосовується, коли багато пацієнтів і лікарень.

Схема спільного використання EHR на основі гібридного ланцюга запропонована Yu et al. зберігає приватну частину електронного обігу у федеративному ланцюжку, а не приватну частину - у публічному ланцюжку [26]. Тільки ліцензовані користувачі можуть отримати доступ до закритої частини, А до закритої частини можна надати доступ науковим установам для розвитку медицини. Модель також використовує автономне сховище, і в ланцюжку зберігаються тільки хеші даних, щоб запобігти підробці даних, а смарт-контракти можуть автоматично управляти запитом EMR, процесом затвердження і використання. Гібридний ланцюговий підхід, що застосовується в моделі, є дуже новим. Однак вузлам не надаються атрибути, і використовується грубий контроль доступу.

Zou та ін. розробили нову структуру ланцюжка, щоб уникнути проблеми розгалуження, і запропонували заснований на довірі механізм консенсусу для протидії візантійським атакам [27]. Медичні установи можуть накопичувати бали довіри за рахунок безперервного майнінгу в обмін на EMR. Пропонована система репутації повинна накопичувати Репутаційні бали за рахунок великої кількості невірних обчислень. Для отримання права голосу споживається велика кількість енергії. Шахназ та ін. пропонують засновану на блокчейне дрібнозернисту систему доступу, яка надає різні права доступу пацієнтам, лікарям, медсестрам і адміністраторам [28]. Доступ до електронних звернень реєструється в моделі, запропонованій в [29], і для пошуку інформації без дешифрування даних в ланцюжку використовується метод шифрування з можливістю пошуку. Цей метод захищає конфіденційність даних і забезпечує швидкість виконання запити. Метод також використовує управління доступом на основі ролей. Порівняння різних моделей показано в табл. 2. Як видно з таблиці, майже всі ці моделі використовують автономне сховище. Це пов'язано з невеликою ємністю блокчейна, яка обмежує ємність сховища даних. Це проблема, яку потрібно вирішити в майбутньому

Таблиця 2 – Порівняння систем зберігання даних на основі блокчейн

Посилання	Тип блокчейну	Методи зберігання	Шифрування даних
[23]	публічний	мережеве сховище	ні
[24]	публічний	автономне сховище	так
[25]	приватний	автономне сховище	так
[26]	гібридне	автономне сховище	ні
[27]	публічний	автономне сховище	так
[28]	публічний	автономне сховище	ні
[29]	публічний	автономне сховище	так

ІОМТ включає різні медичні пристрої, які використовують комп'ютерні мережі для підключення та

визначення параметрів симптомів пацієнтів. ІОМТ має великі переваги для лікування пацієнтів із захворюваннями, а виявлення фізичних ознак дозволяє якомога швидше виявити захворювання та звернутися за медичною допомогою [30]. Однак на ринку існує безліч продуктів ІОМТ без єдиних стандартів управління, і це загрожує витоком інформації [31]. Блокчейн пропонує рішення для забезпечення безпеки медичної ІОМТ [32]. Чен та ін. розробили систему збору даних на основі ІОМТ для забезпечення безпечного зберігання та обміну медичними даними [33]. Система може збирати дані з декількох медичних пристроїв одночасно, щоб забезпечити збір медичних записів пацієнта в режимі реального часу під час операції. Система спроектована як схема анонічного обміну медичними даними на базі хмарного сервера з алгоритмом повторного шифрування через проксі. Такий підхід підвищує безпеку обміну приватними медичними даними. Система реалізована на основі Hyperledger Fabric, дозволеної блокчейн-архітектури, з архітектурою розгортання двоканальної структури і кодом медичного ланцюжка, призначеним для управління даними і контролю доступу. Цей метод використовує алгоритм консенсусу kafka. Цей послідовний алгоритм може призвести до збою половини вузлів, але він не може дозволити зловмисним вузлам. Це робить систему більш вразливою до атак.

Нова технологія безпечної аутентифікації на основі блокчейна була запропонована Джафаром для підвищення безпеки конфіденційних медичних даних, що передаються між пацієнтами та лікарнями [34]. Цифровий підпис Lamport Merkle (LMDS) виконує тут генерацію та перевірку підпису, щоб забезпечити безпечну передачу конфіденційних медичних даних у медичних мережах Інтернету речей на основі хмарних серверів. Розумні контракти дозволяють сторонам-учасникам (тобто пацієнтам та лікарям) встановлювати умови та автоматизувати операції через хмарний сервер, зменшуючи роботу третіх сторін. Смарт-контракти також мають різні адреси і облікові записи в блокчейне, так що кожен пристрій Інтернету речей може переглядати і виконувати свої інструкції, тим самим знижуючи накладні витрати на зв'язок. Алькаралле та ін. представили нову модель захищеної передачі зображень та діагностики за допомогою глибокого навчання та блокчейну для ІоМТ [35]. Запропонована модель включає кілька процесів, зокрема збір даних, захищені транзакції, шифрування хеш-значення та класифікацію даних. На початковому етапі дані про пацієнта збираються за допомогою інструментів Інтернету речей, а потім шифруються за допомогою алгоритму GO-FFO. Крім того, хеші в блокчейні шифруються і стискаються за допомогою технології NIS-BWT. Нарешті, процес класифікації виконується за допомогою моделі DBN. Покращений алгоритм шифрування, хоча і більш безпечний, вимагає більше часу для шифрування та дешифрування, ніж інші алгоритми. У системі, запропонованій Suyel [36], передбачений API-інтерфейс. Цей інтерфейс генерує та підтримує дані про стан здоров'я між медичним працівником та пацієнтом. Крім того, смарт-контракти в повній мірі використовуються в пропонуваній системі для запобігання

шкідливої поведінки шляхом встановлення безпечних правил за допомогою смарт-контрактів. Метод використовує лише просту автентифікацію. Якщо вузлів можна присвоїти деталізовані властивості. Це могло б зробити модель більш досконалою. Ху та ін. припускають, що багато досліджень iomt, засновані на блокчейне, в даний час зосереджені на перевірці криптографічних алгоритмів. Час від часу слід приділяти більше уваги

недійсним підписам, щоб зменшити ймовірність відмови перевірки. Порівняння моделей iomt на основі блокчейну показано в табл. 3. Смарт - контракти на блокчейні відіграють важливу роль в ІОМТ. Смарт-контракти не вимагають участі третіх осіб і можуть автоматично виконувати поставлені завдання при виконанні умов. Зазвичай модель використовує криптографічні алгоритми для підвищення безпеки.

Таблиця 3 – Порівняння моделей ІОМТ на основі блокчейна

Посилання	Тип блокчейн	шифрування даних	смарт-контракт	Основа блокчейн
[33]	приватний	так	немає	Повторне шифрування, анонімний обмін
[34]	загальнодоступний	так,	так	Цифровий підпис Лампорта Меркла
[35]	загальнодоступне	так,	так	Шифрування після класифікації даних
[36]	загальнодоступне	так	так	Сховище сертифікатів унікальних даних
[37]	приватний	так	немає	Механізм цифрової верифікації

Конфіденційність медичних даних перешкоджає роботі машинного навчання на основі даних. Федеративне навчання-це новий метод штучного інтелекту, який захищає конфіденційність даних при побудові моделей штучного інтелекту. Федеративне навчання дозволяє декільком вузлам спільно вивчати модель публічно, і між вузлами передаються лише градієнти та втрати, а не самі дані, що може забезпечити хороший захист даних. Однак вузлам потрібно передати дані до центральної установи для наступного обчислення. Блокчейн може бути хорошою альтернативою центральній установі і дозволяє уникнути нечесності центральної структури.

В рамках цієї роботи для управління ключами пропонується метод повторного шифрування через проксі (PRE). Повторне шифрування проксі - сервера-це метод, тоді як проксі-сервер перетворює зашифрований текст у (CA), який зашифрований за допомогою pkA, до зашифрованого тексту B (CB), який можна розшифрувати за допомогою skB, що використовують ключ повторного шифрування (rkA→B). Проксі вимагає лише зашифрованого тексту A та ключа шифрування, який створюється за допомогою skA і pkB поза проксі. Таким чином, власник тексту A може ділитися секретними даними, не розкриваючи секретний ключ або секретні дані. Ключова концепція полягає в тому, щоб розкрити проксі якомога менше даних, оскільки це ненадійна платформа, і дозволити йому виконати зміну ключа з skA на skB для розшифровки зашифрованого тексту A.

Наведений нижче алгоритм пояснює алгоритм повторного шифрування проксі-сервера, який може бути використаний нашої роботі.

1) Генерація ключа:

Нехай $G_1 = \langle g \rangle$ циклічна група простого порядку q . Приватний ключ пацієнта $sk_a = a \in Z_q^*$ вибраний випадковим чином, відкритий ключ $pk_a = g^a$. Особистий ключ лікаря $sk_b = b \in Z_q^*$ вибраний випадковим чином, відкритий ключ $pk_b = g^b, r \in Z_q^*$, обрано випадковим чином, $Z = e(g, g)$:

$$rk_{A \rightarrow B} = (g^b)^{1/a} = g^{b/a} \in Z_q^*.$$

2) Шифрування:

Нехай $m \in G_2$. Зашифрований текст

$$C_a = (Z^r \cdot m, g^{ra}).$$

Розшифровка (пацієнт):

$$m = \frac{Z^r \cdot m}{e(g^{ra}, g^{1/a})} = \frac{Z^r \cdot m}{Z^r}.$$

Повторне шифрування:

$$C_a \rightarrow ProxyServer \rightarrow C_b ;$$

$$(Z^r \cdot m, g^{ra}) \rightarrow (Z^r \cdot m, e(g^{ra}, rk_{A \rightarrow B}));$$

$$C_b = \left(Z^r \cdot m, e \left(g^{ra}, g^{\frac{b}{a}} \right) \right);$$

$$C_b = (Z^r \cdot m, Z^{rb}).$$

Розшифровка (лікар):

$$m = \frac{Z^r \cdot m}{(Z^{rb})^{1/b}}.$$

В роботі пропонується використання Hyperledger Fabric release 2.2, для створення блокчейн-мережі з N одноранговими вузлами (P1, P2,..., PN), де N більше або дорівнює 3, і вузлом обслуговування замовлень. Вузли є основними елементами мережі, оскільки вони зберігають книги (L) та розумні контракти (S). В ідеалі кожна однорангова інфраструктура повинна керуватися іншою корпорацією. У цьому сенсі вони можуть представляти N зацікавлених сторін таких як: уряд, організації охорони здоров'я, інститути громадянського суспільства, лікарні та інші, — що діють в інтересах підтримки і розвитку сфери охорони здоров'я. Таким чином, вузли надають мережеві послуги, такі як запис та читання книг для адміністраторів та користувачів, що стосуються цих сторін. Теоретично для N не існує верхньої межі, відмінної від тієї, що накладається апаратним та програмним забезпеченням, що використовує протокол консенсусу.

Вузли пов'язані зі своїми відповідними клієнтськими вузлами (CL_1, CL_2, \dots, CL_N)-елементами поза мережею, які дозволяють додатку підключатися до блокчейну, тобто зовнішній додаток отримує доступ до реєстру і смарт-контрактів через з'єднання клієнт-одноранговий вузол. Крім того, Hyperledger Fabric розглядає канал (C) як основний канал зв'язку, за допомогою якого однорангові вузли та клієнти можуть створити консорціум із чітко визначеною політикою, забезпечуючи таким чином механізм ізоляції активів та транзакцій від решти мережі. У цьому контексті кожен смарт-контракт і відповідний реєстр можуть бути окремо викликани по певному каналу тільки користувачами,

раніше зареєстрованими в консорціумі, тим самим забезпечуючи сумісність і конфіденційність.

Однорангові вузли призначаються консорціуму: уряду, організаціям охорони здоров'я, інститутам громадянського суспільства та лікарням у нашому прикладі — відповідними центрами сертифікації (CA_1, CA_2, \dots, CA_N), елементами, які генерують інфраструктуру відкритих та приватних ключів для видачі посвідчень особи за допомогою цифрових сертифікатів. Всякий раз, коли один з членів Консорціуму вступає до ресурсів блокчейна, органи сертифікації підтверджують каналу цифрову ідентифікацію заявника і її/його права на використання необхідного смарт-контракту. Нарешті, вузол обслуговування замовлень забезпечує взаємодію між одноранговими вузлами під час надсилання транзакції та забезпечує узгодженість книги після виконання узгодженого протоколу. У Hyperledger Fabric політика схвалення здійснюється в результаті три етапного процесу:

- 1) пропозиція
- 2) замовлення та упаковка
- 3) перевірка та фіксація.

На рис. 2 представлений архітектура системи, що пропонується в роботі.

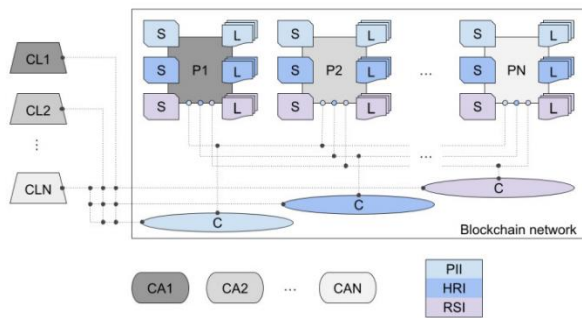


Рис. 2. Архітектура системи, що пропонується:

- HRI: інформація про медичну документацію;
 PII: інформація про особу; C: канал;
 RSI: інформація про обмін записами;
 P: одноранговий вузол; S: розумний контракт;
 L: книга; CL: клієнт; CA: центр сертифікації
 (джерело: власна розробка)

Розробка нашої блокчейн-мережі з урахуванням N партнерів по схваленню і їх відповідних клієнтів і центрів сертифікації. Кожен канал пов'язаний з певним набором реєстрів і смарт-контрактів, які відповідно називаються особистою інформацією, інформацією про стан здоров'я та інформацією про обмін записами. В ідеалі кожен потрібний одноранговий центр сертифікації-клієнт-центр сертифікації повинен керуватися іншою організацією чи установою.

Для підтримки розробки архітектурної моделі буде реалізовано прототип системи заснованої на блокчейні для обміну медичними даними пацієнтів.

При використанні 3-однорангової мережі наш перший тест налаштований на виконання робочого навантаження від 100 до 2500 одночасних відправлень метаданих про стан здоров'я з кроком в 100 кроків по кожному смарт-контрактом шаблонів PII, HRI і RSI. Ми обмежуємо наш тест 2500 запитами,

оскільки Hyperledger Fabric стандартно налаштований на виконання максимум 2500 одночасних запитів. Сценарії написання налаштовані на використання 5 працівників, які відправляють одночасно 10 000 транзакцій, загальна сума кожної з яких становить 50 000. Сценарії читання налаштовані на паралельне використання одних і тих же 5 працівників, але для випадкового запиту записів протягом 600 секунд безперервної роботи. Контролер швидкості підтримується в режимі фіксованого навантаження, починаючи з 50 tps і 500 tps для транзакцій запису і читання відповідно, і збільшуючись до досягнення максимальних швидкостей. Оскільки PII, HRI та RSI призначені для зберігання лише зашифрованих текстів, у нашому тесті всі змодельовані подання метаданих про стан здоров'я генеруються випадковим чином у вигляді рядків фіксованої довжини для кожного поля смарт-контракту. Порожня блокчейн-мережа створюється в кожному навантажувальному тесті, щоб гарантувати рівні умови. Наше тестове середовище складається з комп'ютера з процесором Intel Xeon E-2246g (12 МБ кеш-пам'яті, 3,60 ГГц, 6 ядер, 12 потоків), графічним адаптером NVIDIA Quadro P1000 і оперативною пам'яттю об'ємом 16 ГБ, що працює під управлінням 64-розрядної операційної системи Ubuntu 18.04.5 LTS.

Результат роботи прототип системи відображає не тільки час виконання кожного компонента в запропонованій нами системі на основі блокчейна, але також відображає час виконання всієї системи одним користувачем. Час, необхідний для операції збереження та вилучення, обчислюється на основі синтетичного навантаження МД розміром 128, 512 КБ, 2, 8, 32 і 128 МБ. Для зберігання МД в запропонованій системі на основі блокчейна клієнт-власник МД і сервер шлюзу управляють операцією сховища. Середній час виконання кожного з підпроцесів двох елементів показано в табл. 4 і 5 відповідно.

У табл. 4 представлені детальні дані про час виконання запитів клієнта-власника МД. Клієнт-власник МД виконує п'ять процесів, включаючи хешування, шифрування, генерацію ключів повторного шифрування, підписання та надсилання даних на сервер шлюзу для збереження МД. Згідно з табл. 4, час виконання трьох процесів, включаючи хешування, шифрування та надсилання даних, залежить від розміру даних МД, в той час як час виконання процесів генерації ключа повторного шифрування і підпису залишається незмінним. Таким чином, час виконання для трьох процесів (тобто хешування, шифрування та надсилання даних) можна визначити за допомогою щільності ймовірності Функція (PDF) та час виконання решти двох процесів (Генерація ключа повторного шифрування і підписання) можуть бути представлені у вигляді константи.

Сервер шлюзу також виконує п'ять основних процесів, включаючи Перевірка підпису, завантаження даних, локальне зберігання даних, підписання та зберігання файлу журналу в блокчейні. Як показано в табл. 5, час виконання лише одного процесу, який є завантаженням даних, залежить від розміру МД, в той час як час виконання інших процесів

залишається практично постійним. Щоб отримати МД в запропонованій системі на основі блокчейна, користувальницький клієнт і сервер шлюзу виконують

операцію вилучення. Середній показник час виконання кожного з підпроцесів цих двох елементів також показано в табл. 6 і 7 відповідно.

Таблиця 4 – Час виконання операцій клієнтом-власником МД

Розмір даних	Час хешування	Час шифрування	Час генерації ключа повторного шифрування	Час підпису	Час відправки даних
128 КБ	10.29	91.18	24.16	1.16	152.73
512 КБ	18.24	94.01	24.66	1.15	173.87
2 МБ	40.63	101.19	26.15	1.18	268.95
8 МБ	65.60	142.03	26.88	1.16	421.67
32 МБ	241.80	303.79	27.00	1.31	645.70
128 МБ 1	946.10	1828.21	27.10	1.42	2200.36

Таблиця 5 – Час виконання операцій сервером-шлюзу

Розмір даних	Час перевірки підпису	Час завантаження	Час збереження локальної копії	Час входу на сервер	Час блокування
128 ГБ	0,07	157,41	31,57	1,24	3372,79
512 ГБ	0,07	221,65	24,66	1,15	3173,87
2 МБ	0,07	273,65	38,40	1,30	3365,34
8 МБ	0,08	457,51	33,82	1,49	3238,70
32 МБ	0,06	654,280	28,62	1,60	2935,02
128 МБ	0,07	2150,87	38,71	1,58	3381,05

Таблиця 6 – Час виконання операцій в процесі отримання МД користувачем-клієнтом

Розмір даних	Час пошуку МД по блокчейн	Час підтвердження підпису (власник, сервер)	Час підпису користувача	Час відправки запиту	Час розшифровки
128 КБ	785,69	0,07, 0,04	1,33	115,36	3,20
512 КБ	820,48	0,07, 0,04	1,29	124,30	6,04
2 МБ	751,15	0,07, 0,04	1,31	110,77	16,63
8 МБ	770,61	0,07, 0,04	1,23	136,25	59,41
32 МБ	823,37	0,07, 0,04	1,75	127,79	238,90
128 МБ	796,67	0,07, 0,04	1,39	128,77	1814,79

Таблиця 7 – Час виконання операцій на сервері-шлюзу для процесу вилучення МД

Розмір даних	Час перевірки підпису користувача	Час збереження журналу в блокчейн	Час повторного шифрування	Час завантаження даних
128 КБ	0.11	3304.62	30.59	38.02
512 КБ	0.10	3288.55	31.50	78.27
2 МБ	0.10	3308.91	34.28	152.31
8 МБ	0.11	3398.48	58.75	214.84
32 МБ	0.13	3367.66	79.70	469.33
128 МБ	0.12	3372.62	80.65	1093.03

Згідно з табл. 6 час виконання розшифровки залежить тільки від розміру даних. Сервер шлюзу також виконає чотири основних процеси, включаючи перевірку підпису, збереження файлу журналу на блокчейні, відбувається завантаження даних і повторне шифрування. Час виконання двох процесів, включаючи завантаження даних і повторне шифрування, залежить від розміру даних, як показано в табл. 7. Щоб отримати МД, Користувач-клієнт виконує шість основних процесів, включаючи пошук в блокчейне, перевірку підпису власника, Перевірка підпису сервера, підписання, надсилання запиту на сервер шлюзу і розшифровку отриманих МД. Згідно з табл. 4-7 для оцінки середнього часу роботи використовуються найближчі середні дані (32 МБ даних). Час, необхідний клієнту-власнику МД для виконання операції зі сховищем, становить 1219,606 мс, а час обслуговування сервера шлюзу при виконанні операції зі сховищем становить 3619,578 мс.

Таким чином, середній час роботи системи для операції зберігання становить приблизно 4839,184 мс або 4,84 с. Час, необхідний для користувача клієнту для виконання операції вилучення, становить 1191,919 мс, а час обслуговування сервера шлюзу для виконання операції вилучення становить 3916,822 мс. В результаті час роботи системи для операції вилучення становить приблизно 5108,741 мс або 5,19 с. ці результати показують середню продуктивність для одного користувача.

Архітектурна модель моделюється з урахуванням робочого навантаження, і результат моделювання порівнюється з результатом, що спостерігається в системі-прототипі. При виконанні прототипу середній час відгуку для операції зберігання становить 4,84 сек, тоді як середній час відгуку для операції зберігання становить 5,0 с.

Таким чином, моделювання передбачало середній час відгуку з відносною похибкою 3,3% для

операції зберігання. Середній час відгуку на операцію вилучення, що виконується системою-прототипом, становить 5,11 с, в той час як середній час відгуку для операції вилучення, яке оцінюється архітектурною моделлю, становить 5,3 с. моделювання передбачає середній час відгуку з відносною похибкою 3,7% для операції вилучення. Таким чином, моделювання передбачало час відгуку, близький до результату, що спостерігається в системі-прототипі.

Висновки

У роботі пропонується архітектурна модель системи обміну медичними даними на основі блокчейн. В роботі реалізовано прототип системи для дослідження ключових параметрів запропонованої архітектури. Робота прототипу системи була перевірена на пакетах медичних даних, що складаються з даних різних розмірів, включаючи 128КБ, 512 КБ, 2, 8, 32 і 128 МБ. Результуючий час виконання ділиться на дві групи. У першій групі час кожного виконання змінюється залежно від розміру пакету МД та часу виконання. У другій групі час виконання залишається майже таким же. Таким чином, час виконання першої

групи моделюється за допомогою функції PDF, тоді як час виконання другої групи моделюється з їх початковими значеннями як константа. Архітектурна модель оцінює, що запропонована модель системи може реагувати протягом 4 хвилин на 165000 звернень в день. Однак результат моделювання зі швидкістю надходження 3,8 запити в секунду показує, що час відгуку для всіх операцій становить <20 хв, і 50% цих відгуків знаходяться в межах 8 хвилин від аварійних вимог. Результат моделювання зі швидкістю прибуття 15,2 в секунду показує, що тільки 30% часу реагування укладається в 8 хвилин, відведених на екстрену допомогу, однак використання системи кожну 1 годину для всіх людей може бути дуже рідкісним випадком.

Запропонований прототип системи має деякі обмеження. Блокчейн hyperledger виконується з фіктивним консенсусом. Мережа між кожним комп'ютером моделюється без істотних мережевих затримок. Однак, запропонована архітектурна модель забезпечує основу для майбутніх досліджень оптимальної конфігурації системи, таких як нефункціональні властивості.

СПИСОК ЛІТЕРАТУРИ

1. Häyrynen Kristiina, Saranto K, Nykänen Pirkko. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *Int J Med Inform.* 2008 May;77(5):291–304. doi: 10.1016/j.ijmedinf.2007.09.001.S1386-5056(07)00168-2
2. Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. *J Am Med Inform Assoc.* 2013 Jan 01;20(1):117–21. doi: 10.1136/amiajnl-2012-001145. <http://europepmc.org/abstract/MED/22955496>. amiajnl-2012-001145
3. Ludwick DA, Doucette J. Adopting electronic medical records in primary care: lessons learned from health information systems implementation experience in seven countries. *Int J Med Inform.* 2009 Jan;78(1):22–31. doi: 10.1016/j.ijmedinf.2008.06.005.S1386-5056(08)00092-0
4. Zahabi M, Kaber DB, Swangnetr M. Usability and Safety in Electronic Medical Records Interface Design: A Review of Recent Literature and Guideline Formulation. *Hum Factors.* 2015 Aug;57(5):805–34. doi: 10.1177/0018720815576827.0018720815576827
5. Mikkelsen G, Aasly J. Concordance of information in parallel electronic and paper based patient records. *International Journal of Medical Informatics.* 2001 Oct;63(3):123–131. doi: 10.1016/s1386-5056(01)00152-6
6. Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. *BMJ.* 2003 May 17;326(7398):1070. doi: 10.1136/bmj.326.7398.1070. <http://europepmc.org/abstract/MED/12750210>. 326/7398/1070
7. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. *J Am Med Inform Assoc.* 2006;13(2):121–6. doi: 10.1197/jamia.M2025. <http://europepmc.org/abstract/MED/16357345>. M2025
8. Archer N, Fevrier-Thomas U, Lokker C, McKibbin KA, Straus SE. Personal health records: a scoping review. *J Am Med Inform Assoc.* 2011;18(4):515–22. doi: 10.1136/amiajnl-2011-000105. <http://europepmc.org/abstract/MED/21672914>. amiajnl-2011-000105
9. Roehrs A, da Costa Cristiano André, Righi RDR, de Oliveira Kleinner Silva Farias. Personal Health Records: A Systematic Literature Review. *J Med Internet Res.* 2017 Jan 06;19(1):e13. doi: 10.2196/jmir.5876.
10. Rudin RS, Motala AR, Goldzweig CL, Shekelle PG. Usage and Effect of Health Information Exchange. *Ann Intern Med.* 2014 Dec 02;161(11):803. doi: 10.7326/m14-0877
11. Williams C, Mostashari F, Mertz K, Hogin E, Atwal P. From the Office of the National Coordinator: the strategy for advancing the exchange of health information. *Health Aff (Millwood)* 2012 Mar;31(3):527–36. doi: 10.1377/hlthaff.2011.1314.31/3/527
12. Cimino JJ, Frisse ME, Halamka J, Sweeney L, Yasnoff W. Consumer-mediated health information exchanges: the 2012 ACMI debate. *J Biomed Inform.* 2014 Apr;48:5–15. doi: 10.1016/j.jbi.2014.02.009. [https://linkinghub.elsevier.com/retrieve/pii/S1532-0464\(14\)00046-X](https://linkinghub.elsevier.com/retrieve/pii/S1532-0464(14)00046-X). S1532-0464(14)00046-X
13. Zhuang Y, Sheets LR, Chen Y, Shae Z, Tsai JJ, Shyu C. A Patient-Centric Health Information Exchange Framework Using Blockchain Technology. *IEEE J Biomed Health Inform.* 2020 Aug;24(8):2169–2176. doi: 10.1109/jbhi.2020.2993072.
14. Gordon WJ, Catalini C. Blockchain Technology for Healthcare: Facilitating the Transition to Patient-Driven Interoperability. *Comput Struct Biotechnol J.* 2018;16:224–230. doi: 10.1016/j.csbj.2018.06.003. [https://linkinghub.elsevier.com/retrieve/pii/S2001-0370\(18\)30028-X](https://linkinghub.elsevier.com/retrieve/pii/S2001-0370(18)30028-X). S2001-0370(18)30028-X
15. Zhang P, White J, Schmidt DC, Lenz G, Rosenbloom ST. FHIRChain: Applying Blockchain to Securely and Scalably Share Clinical Data. *Comput Struct Biotechnol J.* 2018;16:267–278. doi: 10.1016/j.csbj.2018.07.004. [https://linkinghub.elsevier.com/retrieve/pii/S2001-0370\(18\)30037-0](https://linkinghub.elsevier.com/retrieve/pii/S2001-0370(18)30037-0). S2001-0370(18)30037-0
16. Murphy DR, Satterly T, Rogith D, Sittig DF, Singh H. Barriers and facilitators impacting reliability of the electronic health record-facilitated total testing process. *Int J Med Inform.* 2019 Jul;127:102–108. doi: 10.1016/j.ijmedinf.2019.04.004.S1386-5056(18)31386-8

17. Tanwar S, Parekh K, Evans R. Blockchain-based electronic healthcare record system for healthcare 4.0 applications. *Journal of Information Security and Applications*. 2020 Feb;50:102407. doi: 10.1016/j.jisa.2019.102407
18. Dagher GG, Mohler J, Milojkovic M, Marella PB. Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. *Sustainable Cities and Society*. 2018 May;39:283–297. doi: 10.1016/j.scs.2018.02.014.
19. Zhang A, Lin X. Towards Secure and Privacy-Preserving Data Sharing in e-Health Systems via Consortium Blockchain. *J Med Syst*. 2018 Jun 28;42(8):140. doi: 10.1007/s10916-018-0995-5.10.1007/s10916-018-0995-5
20. Cao S, Zhang G, Liu P, Zhang X, Neri F. Cloud-assisted secure eHealth systems for tamper-proofing EHR via blockchain. *Information Sciences*. 2019 Jun;485:427–440. doi: 10.1016/j.ins.2019.02.038.
21. Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. *Bitcoin: Open Source P2P Money*. 2008. [2021-04-23]. <https://bitcoin.org/bitcoin.pdf>.
22. Ferdous MS, Chowdhury MJM, Hoque MA. A survey of consensus algorithms in public blockchain systems for cryptocurrencies. *Journal of Network and Computer Applications*. 2021 May;182:103035. doi: 10.1016/j.jnca.2021.103035.
23. Kuo T, Zavaleta Rojas H, Ohno-Machado L. Comparison of blockchain platforms: a systematic review and healthcare examples. *J Am Med Inform Assoc*. 2019 May 01;26(5):462–478. doi: 10.1093/jamia/ocy185.
24. McGhin T, Choo KR, Liu CZ, He D. Blockchain in healthcare applications: Research challenges and opportunities. *Journal of Network and Computer Applications*. 2019 Jun;135:62–75. doi: 10.1016/j.jnca.2019.02.027.
25. Vazirani AA, O'Donoghue O, Brindley D, Meinert E. Implementing Blockchains for Efficient Health Care: Systematic Review. *J Med Internet Res*. 2019 Feb 12;21(2):e12439. doi: 10.2196/12439. <https://www.jmir.org/2019/2/e12439/> v21i2e12439
26. Hussien HM, Yasin SM, Udzir SNI, Zaidan AA, Zaidan BB. A Systematic Review for Enabling of Develop a Blockchain Technology in Healthcare Application: Taxonomy, Substantially Analysis, Motivations, Challenges, Recommendations and Future Direction. *J Med Syst*. 2019 Sep 14;43(10):320. doi: 10.1007/s10916-019-1445-8.10.1007/s10916-019-1445-8 [PubMed: 31522262] [CrossRef: 10.1007/s10916-019-1445-8]
27. Azaria A, Ekblaw A, Vieira T, Lippman A. *MedRec: Using blockchain for medical data access and permission management*. 2016 2nd International Conference on Open and Big Data (OBD); August 22–24; Vienna, Austria. 2016. pp. 25–30
28. Yue X, Wang H, Jin D, Li M, Jiang W. Healthcare Data Gateways: Found Healthcare Intelligence on Blockchain with Novel Privacy Risk Control. *J Med Syst*. 2016 Oct;40(10):218. doi: 10.1007/s10916-016-0574-6.10.1007/s10916-016-0574-6
29. Roehrs A, da Costa Cristiano André, da Rosa Righi Rodrigo. OmniPHR: A distributed architecture model to integrate personal health records. *J Biomed Inform*. 2017 Jul;71:70–81. doi: 10.1016/j.jbi.2017.05.012.
30. Ichikawa D, Kashiya M, Ueno T. Tamper-Resistant Mobile Health Using Blockchain Technology. *JMIR Mhealth Uhealth*. 2017 Jul 26;5(7):e111. doi: 10.2196/mhealth.7938. <https://mhealth.jmir.org/2017/7/e111/> v5i7e111
31. Mannaro K, Baralla G, Pinna A, Ibba S. A Blockchain Approach Applied to a Teledermatology Platform in the Sardinian Region (Italy) Information. 2018 Feb 23;9(2):44. doi: 10.3390/info9020044
32. 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
33. Ji Y, Zhang J, Ma J, Yang C, Yao X. BMPLS: Blockchain-Based Multi-level Privacy-Preserving Location Sharing Scheme for Telecare Medical Information Systems. *J Med Syst*. 2018 Jun 30;42(8):147. doi: 10.1007/s10916-018-0998-2.10.1007/s10916-018-0998-2 [PubMed: 29961160] [CrossRef: 10.1007/s10916-018-0998-2]
34. Kleinaki A, Mytis-Gkometh P, Drosatos G, Efraimidis PS, Kaldoudi E. A Blockchain-Based Notarization Service for Biomedical Knowledge Retrieval. *Comput Struct Biotechnol J*. 2018;16:288–297. doi: 10.1016/j.csbj.2018.08.002. [https://linkinghub.elsevier.com/retrieve/pii/S2001-0370\(18\)30040-0](https://linkinghub.elsevier.com/retrieve/pii/S2001-0370(18)30040-0) .S2001-0370(18)30040-0
35. Jamil F, Hang L, Kim K, Kim D. A Novel Medical Blockchain Model for Drug Supply Chain Integrity Management in a Smart Hospital. *Electronics*. 2019 May 07;8(5):505. doi: 10.3390/electronics8050505.
36. Patel V. A framework for secure and decentralized sharing of medical imaging data via blockchain consensus. *Health Informatics J*. 2019 Dec;25(4):1398–1411. doi: 10.1177/1460458218769699.
37. Jamil F, Ahmad N, Iqbal N, Kim D. Towards a Remote Monitoring of Patient Vital Signs Based on IoT-Based Blockchain Integrity Management Platforms in Smart Hospitals. *Sensors (Basel)* 2020 Apr 13;20(8):2195. doi: 10.3390/s20082195. <https://www.mdpi.com/resolver?pii=s20082195> .s2008215.

Received (Надійшла) 02.04.2024

Accepted for publication (Прийнята до друку) 05.06.2024

Development and research of an architectural model of a blockchain-based personal data exchange system.

Olexander Shmatko, Dmytro Kulinich, Tetiana Gorbach

Abstract. **Abstract.** **Relevance.** Modern society is facing a growing need for secure, reliable and transparent exchange of personal data of patients in the field of healthcare. Protecting the confidentiality and integrity of medical information is a priority to ensure quality and efficient medical care. Blockchain technologies provide a promising tool for solving this problem by enabling the creation of a decentralized and secure system for the exchange of personal patient data. **The purpose** of this work is to ensure a high level of security and confidentiality of medical data, as well as to increase the efficiency of healthcare processes by developing software components of a system for exchanging personal data of patients based on blockchain technologies. **The object of research** is the system of exchange of personal data of patients in the healthcare sector. **The subject of the study** is software components based on blockchain technologies designed to ensure the security, transparency and efficiency of medical information exchange. **Results.** This paper proposes an architectural model of a secure and efficient health data exchange system that can be widely implemented in the healthcare sector. **Conclusion.** The introduction of a secure personal data exchange system based on blockchain technology in the healthcare sector will help improve the quality of medical care and provide faster access to important data for medical personnel. Theoretical significance lies in expanding knowledge about the use of blockchain technologies in the healthcare sector and the security and confidentiality of medical information. This study can serve as a basis for further research in this area and contribute to the development of new methods and approaches to the exchange of medical data.

Keywords: blockchain, personal data of patients, IoTM, smart contracts, Ethereum, model of medical data exchange system.

О. В. Шматко, Д. В. Сафанков

Харківський національний технічний університет радіоелектроніки, Харків, Україна

РОЗРОБКА ТА ВПРОВАДЖЕННЯ ХМАРНОЇ ПЛАТФОРМИ ДЛЯ ТОКЕНІЗАЦІЇ ПЕРСОНАЛЬНИХ ДАНИХ СТУДЕНТІВ

Анотація. Актуальність. Сучасний світ переживає важливий перехідний період з етапу інформатизації важливих аспектів людської активності до етапу їхньої цифровізації. Інформатизація являє собою процес модернізації різних сфер діяльності за допомогою інформаційно-комунікаційних технологій, в той час як цифрова трансформація означає глибоке переосмислення і переформатування цих сфер на основі цифрових рішень і технологій. Цей процес вимагає не лише підтримки існуючих умов для розвитку передових цифрових платформ і технологій, але й створення нових можливостей для зародження інноваційних рішень. Ключовими технологіями у цій цифровій епісі є великі дані, нейротехнології та штучний інтелект, блокчейн, квантові технології, новітні виробничі методики, промисловий інтернет, елементи робототехніки та сенсорики, технології бездротового зв'язку та інновації у сфері віртуальної та доповненої реальності. У даній роботі, в контексті цифрової трансформації освітньої сфери, проведено дослідження з використання блокчейн-технологій для токенизації освітніх ресурсів та аналізуються потенціал їх застосування у освіті. **Метою даної роботи** є підвищення конфіденційності освітніх даних, їх цілісності та доступності, а також забезпечення надійного обміну цими даними між всіма зацікавленими сторонами за рахунок проектування, розробки та впровадження програмних компонентів хмарної платформи для токенизації персональних даних студентів на основі технології блокчейн. **Об'єкт дослідження** включає в себе системи передачі та обробки освітніх даних, які забезпечують обмін даними між всіма зацікавленими сторонами в освітньому процесі. **Предметом дослідження** є методи та засоби проектування та розробки програмних компонентів, необхідні для створення і підтримки платформ для токенизації персональних даних студентів. Ці компоненти включають в себе програмне забезпечення для шифрування, аутентифікації, авторизації, а також механізми для забезпечення відмовостійкості та відновлення даних. **Результати.** У даній роботі запропоновано модель хмарної децентралізованої системи для токенизації персональних даних студентів. **Висновок.** Інтеграція блокчейн технологій в освітній сектор має великий потенціал для зміцнення довіри та підвищення якості освітніх послуг, а також для створення нової екосистеми освітніх активів, доступних і перевірених на глобальному рівні. Реалізація блокчейн технологій у сфері освіти вимагає розробки спеціалізованого програмного забезпечення для видачі та перевірки цифрових дипломів, заснованих на блокчейні. Впровадження хмарної платформи для токенизації персональних даних студентів обіцяє значне покращення в обробці та зберіганні освітніх документів, сприяючи більшій довірі та прозорості у сфері освіти.

Ключові слова: блокчейн, персональні дані студентів, хмарні сховища даних, смарт-контракти, Ethereum, архітектурна модель хмарної платформи токенизації даних

Вступ

Технологія блокчейн пропонує децентралізовану структуру, що дозволяє здійснювати транзакції між двома сторонами без необхідності посередництва третьої сторони [1]. Ця система використовує загальнодоступну книгу для полегшення транзакцій між сторонами, що не довіряють, при цьому цілісність книги забезпечується алгоритмом консенсусу. Існує багато алгоритмів [2], кожен з яких призначений для підтримки узгодженості записів у книзі, перевірки автентичності права учасників вносити записи чи зміни та захисту від потенційних порушень.

Невід'ємні властивості блокчейна, такі як децентралізація, прозорість та безпека, зробили його дуже привабливим у багатьох секторах, включаючи фінанси [3], управління ланцюгами поставок [4–6], Інтернет речей, охорону здоров'я [7] та освіти [8]. У сфері освіти блокчейн впроваджує інновації в декількох областях, включаючи академічні дослідження, системи репутації, електронні портфоліо, управління інтелектуальною власністю, інтеграцію безперервної освіти з платформами навчальної аналітики і перевірку залікових одиниць, облікових даних і сертифікатів [9–10].

Існує безліч наукових оглядів [11–13], присвячених застосуванню блокчейна в освіті, в яких підкреслюється широкий спектр рішень, що вирішують ключові проблеми в цій галузі. Наприклад, деякі роботи

розглядали роль блокчейна в управлінні академічними сертифікатами та інформацією про акредитацію [14–18], хоча початковий аналіз базувався переважно на теоретичних пропозиціях, а не на практичних реалізаціях. В інших оглядах були виявлені нові проекти, що використовують блокчейн в освіті, і виявлені різні потенційні додатки.

Грунтуючись на попередніх систематичних оглядах освітніх додатків на основі блокчейна, в цій статті пропонується хмарна платформа для токенизації персональних даних студентів. Такий підхід дозволяє більш детально зрозуміти вплив блокчейна і його впровадження в освітніх установах, демонструючи здатність технології впроваджувати інновації та вдосконалювати освітні процеси і системи.

Метою роботи є підвищення конфіденційності освітніх даних, їх цілісності та доступності, а також забезпечення надійного обміну цими даними між всіма зацікавленими сторонами за рахунок проектування, розробки та впровадження програмних компонентів хмарної платформи для токенизації персональних даних студентів на основі технології блокчейн.

Основна частина

Аналізуючи класифікацію застосування блокчейна відповідно [19], можна виділити трьохрівневий підхід: Блокчейн 1.0 фокусується на валютних операціях, Блокчейн 2.0 розширюється до смарт-контрак-

тів та фінансових інструментів, тоді як Блокчейн 3.0 виходить за рамки фінансових транзакцій, адаптуючись до різноманітних секторів, включаючи освіту.

Ця технологія відкриває нові можливості для реформування системи видачі та зберігання освітніх документів, надаючи переваги надійності, прозорості та ефективності. Використання блокчейна дозволяє знизити ризики пов'язані з підробкою документів та забезпечити безпечне зберігання освітніх активів без необхідності втручання третіх сторін.

Розгляд поточної системи видачі та зберігання дипломів (рис. 1) виявляє її недоліки, такі як складність процесу, високі витрати та вразливість до фізичних та юридичних ризиків. Пропонується рішення,

яке полягає у токенизації освітніх активів за допомогою блокчейн, що дозволить цифровізувати процес видачі та зберігання дипломів, забезпечити їх автентичність та надійне зберігання в децентралізованому реєстрі [20]. Це не лише спростить доступ до освітніх документів і їх перевірку, але й забезпечить їх збереження навіть у випадку припинення діяльності навчальних закладів.

Для подолання викликів, пов'язаних із фальсифікацією документів та їхнім збереженням, у даному дослідженні пропонується застосування технології блокчейн для токенизації освітніх матеріалів.

На рис. 2 демонструється метод токенизації освітніх матеріалів.

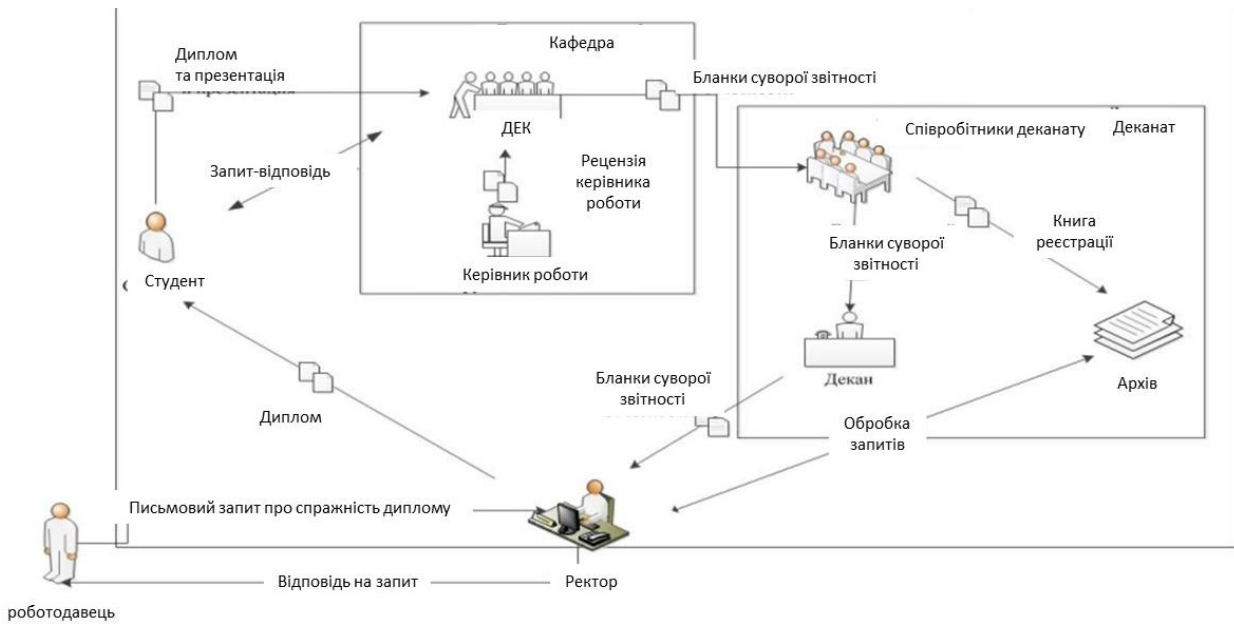


Рис. 1. Існуючий порядок видачі дипломів (джерело: власна розробка)

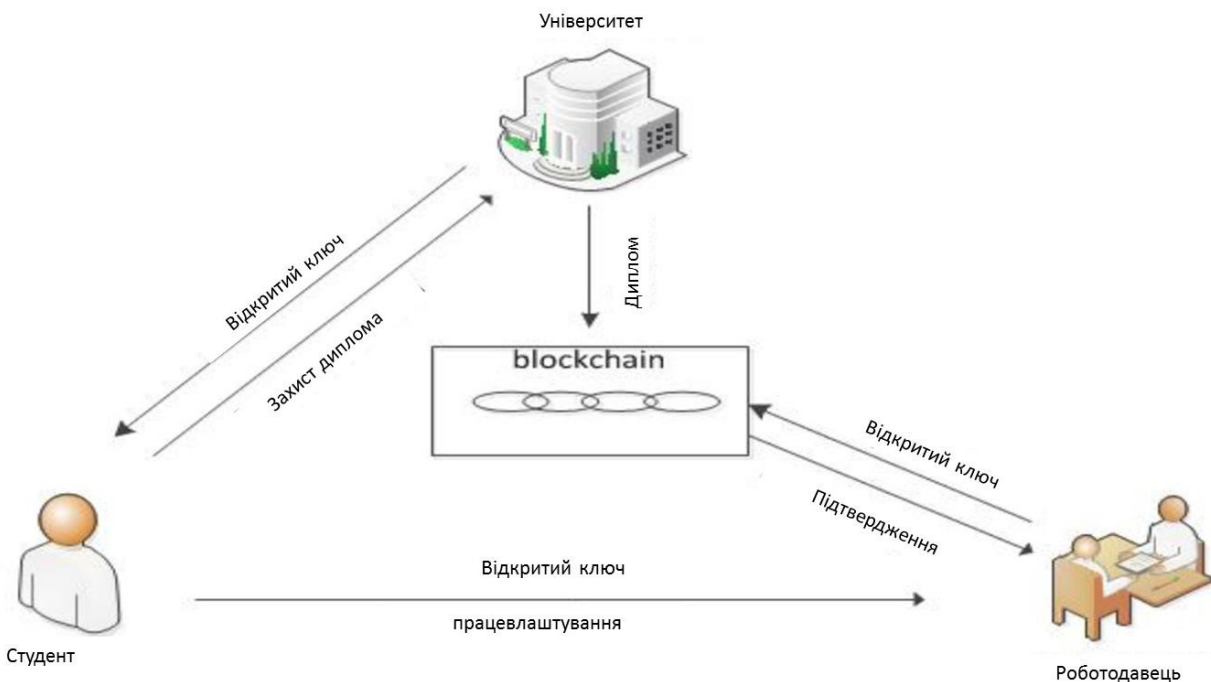


Рис. 2. Модель токенизації освітніх активів (джерело: власна розробка)

За цим підходом, університети, що видають цифрові версії дипломів, використовують спільний децентралізований реєстр для їх зберігання. Цифрові дипломи, підписані за допомогою унікального приватного ключа, надаються безпосередньо роботодавцям, забезпечуючи можливість перевірки автентичності диплома через порівняння з хешем у блокчейні. Це сприяє вирішенню проблеми фальсифікації документів та збереження важливих даних.

У випадку закриття установи, що видала диплом, інформація про диплом залишається у блокчейні, усуваючи потребу в додаткових ресурсах для перевірки документів через третіх осіб, оскільки роботодавці мають прямий доступ до перевірки через блокчейн.

Для глибшого розуміння процесу в блокчейні, важливо ознайомитися з його робочим механізмом, представленим на рис. 3.



Рис. 3. Механізм роботи технології блокчейн (джерело: власна розробка)

Спочатку створюється цифровий токен, що містить ключову інформацію, таку як назва університету, ім'я випускника, дату видачі та інші деталі. Університет застосовує свій приватний ключ для підпису диплома, що підтверджується мережею і записується у блок. Для валідації диплома створюється хеш-файл, унікальний для кожного документу, що дозволяє перевіряти його автентичність без змін у контенті.

Основною перевагою цього методу є надійне збереження сертифікатів у блокчейні, що забезпечує доступність та автентичність документів навіть у випадку закриття установи. Це вимагає від університетів лише наявності програмного забезпечення для створення та верифікації підписаних цифрових сертифікатів у блокчейні, спрощуючи процес перевірки дипломів без залучення зовнішніх сервісів.

Щоб скористатися унікальними технологічними можливостями блокчейна для системи токенизації студентських даних, пропонується хмарна платформа на основі блокчейн-системи Ethereum зі смарт-контрактами. Приватний блокчейн також називають "дозволяючим блокчейном", який обмежує доступ певним користувачам. Архітектура системи, показана на рис. 4, містить два модулі:

1) модуль прив'язки: Системний адміністратор освітнього закладу створить контактну точку для студента після того, як цифровий токен диплома студента буде готовий, і введе відповідну первинну інформацію в смарт-контракт для майбутньої індексації (як показано на рис. 5);

2) модуль запити: студенти надають роботодавцям дозвіл на доступ до своїх даних, додаючи роботодавців до "дозволеного списку" у смарт-контракті.

Роботодавці можуть вибирати записи через точки доступу після отримання доступу до записів студентів. Подальший обмін даними між освітніми установами, що беруть участь у системі токенизації освітніх даних, включатиме шифрування даних і використання блокчейн-системи для відправки та отримання ключів розшифровки (як показано на рис. 6).

Для інтеграції запропонованої платформи вкрай важливо спочатку створити надійну систему аутентифікації і комунікації, яка лежить в основі функціональності системи. Ця платформа повинна включати різні криптографічні методи, включаючи криптографію з відкритим ключем і рівень захищених сокетів (SSL), поряд з різними протоколами обміну ключами, для перевірки системи і забезпечення безпеки, конфіденційності та ідентифікації користувача.

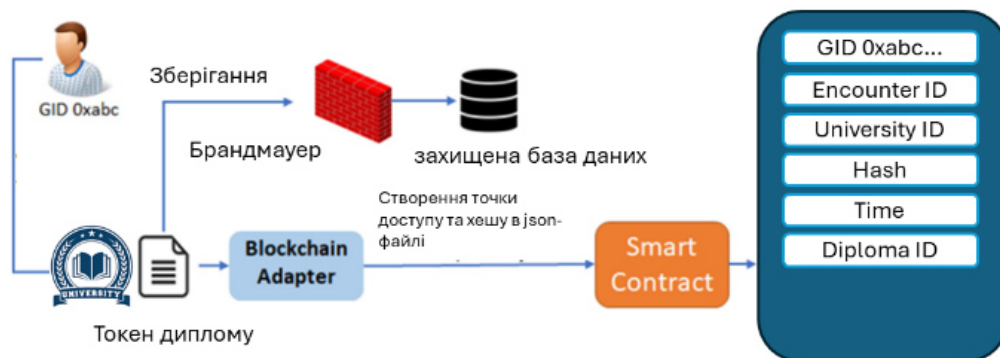


Рис. 4. Системна модель (джерело: власна розробка)



Рис. 5. Схема індексації смарт-контрактів (джерело: власна розробка)



Рис. 6. Схема обробки даних (джерело: власна розробка)

Дані, зібрані за допомогою цього фреймворку, передаються в модуль API, який не тільки зберігає дані в модулі бази даних, але і генерує з'єднання з модулем блокчейна, пов'язуючи його з базою даних медичних записів. Крім того, він записує ідентифікатор транзакції в окрему систему баз даних.

Транзакція на блокчейні ініціюється активацією функції Chaincode CreateIntervention. Ця функція реєструє ідентифікатор транзакції, ідентифікатори навчальних закладів та цифрових гаманців студента, а також шлях до модуля бази даних, який зберігає інформацію про персональні дані студента.

Коли роботодавцю потрібно вивчити освітні дані студента (дані про диплом та виписки оцінок), він повинен пройти автентифікацію в додатку. Після автентифікації він може отримати необхідний ідентифікатор транзакції (Encounter ID) з модуля бази даних. Ідентифікатор обробляється шляхом запису події доступу та повернення транзакції, яка включає посилання у вигляді рядка бази даних на необхідні записи даних студента.

Студенти, які отримують доступ до своїх записів, дотримуються процесу, подібного до процесу роботодавця. Студенти повинні пройти автентифікацію та підключитися до бази даних своїх особистих записів через зовнішню програму. Потім вони знаходять в базі даних ідентифікатор конкретної операції запису даних.

Цей ідентифікатор використовується для виклику функції Chaincode GetInterventionByID, яка отримує транзакцію, що містить посилання на необхідний рядок бази даних студентських записів.

Надання студентом доступу до своїх персональних записів покращує їх розуміння та контроль над цими записами. У сценаріях аудиту адміністратор

мережі може отримати доступ до всіх записів транзакцій блокчейну як незмінний журнал, хоча і без доступу до фактичних студентських даних. Однак вони можуть складати журнали доступу та втручань, включаючи ідентифікатори користувачів.

Для дослідницьких цілей архітектура системи, яка розділяє студентські та особисті записи, якщо не надається доступ до модуля blockchain, забезпечує доступ до бази даних студентських записів без шкоди для особистих даних. Таке налаштування забезпечує відстеження даних та анонімність, спрощуючи програми, які надають студентам право власності на їхні студентські записи.

Оцінка продуктивності системи включала в себе кілька тестів для перевірки налаштування мережі Hyperledger і функціональності всієї системи в цілому. Щоб гарантувати, що результати тестування можуть бути послідовно відтворені, всі параметри середовища тестової мережі були задокументовані відповідно до рекомендацій з Білої книги робочої групи Hyperledger по продуктивності і масштабованості.

Для тестування запропонованої платформи була створена на налаштована тестова інфраструктура на віддаленому віртуальному сервері на основі Amazon Lightsail, оснащеному 4 ГБ оперативної пам'яті, 2 процесорними ядрами та процесором 3000 МГц, що працює під управлінням Ubuntu 20.04.2 LTS. Це налаштування означає, що всі вузли в тестовому середовищі були розміщені на одній машині. Мережа була структурована як проста установка, що включала три організації, кожна з яких мала по одному вузлу, що складало загалом три вузли, де проводилися всі транзакції. Протокол консенсусу RAFT був обраний через його простоту впровадження та повну підтримку в документації Hyperledger Fabric. Для потреб

бази даних використовувалася CouchDB через її здатність виконувати складні запити, А Amazon RDS MySQL використовувався для управління SQL-з'єднаннями, пов'язаними з транзакціями. Розроблений смарт-контракт мав дві основні функції: одну для створення нових транзакцій та вставки даних, а іншу для отримання даних JSON за допомогою певного ідентифікатора.

Спочатку розмір блоку був встановлений на максимум 99 МБ, але для тестування кращі використовувати менші блоки по 512 КБ, що містять до десяти транзакцій у кожному. Для створення значного навантаження при тестуванні був використаний скрипт на Python, а інструмент Postman використовувався для оцінки здатності API обробляти передачу даних через HTTP POST-запити.

Hyperledger Caliper був встановлений як інструмент для тестування та оцінки продуктивності мережі Hyperledger Fabric, зосереджуючись на таких показниках, як пропускна здатність, затримка та

масштабованість.

Вплив на продуктивність оцінювали шляхом зміни розміру блоку та кількості транзакцій у блоці. Затримка вимірювалася часом, витраченим на надсилання транзакції до отримання її результату, а пропускна здатність вимірювалася кількістю дійсних транзакцій, оброблених протягом заданого періоду часу. Перевіреними функціями смарт-контрактів були CreateIntervention для операцій запису та GetInterventionByID для операцій читання, при цьому Hyperledger Caliper надсилав 1000 транзакцій зі швидкістю 100 в секунду.

Результати, як показано на рис. 7 і 8, показали стабільну середню затримку і пропускну здатність при різних розмірах блоків без істотних коливань. Така узгодженість, ймовірно, обумовлена налаштуванням тестового середовища, при якому всі вузли були розташовані на одному комп'ютері, що зводило до мінімуму затримки розповсюдження по мережі.

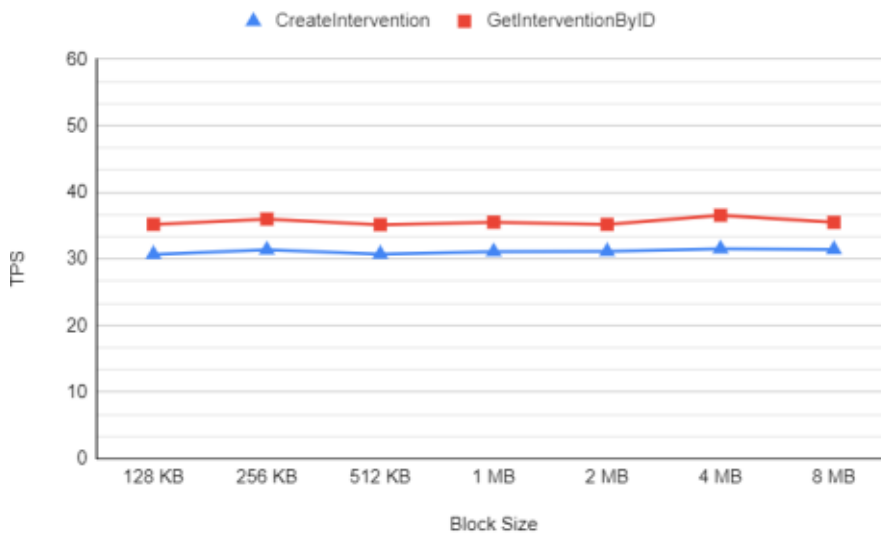


Рис. 7. Результати оцінки пропускної здатності (виміряної в TPS) при варіюванні розміру блоку (з використанням 10 транзакцій на блок)

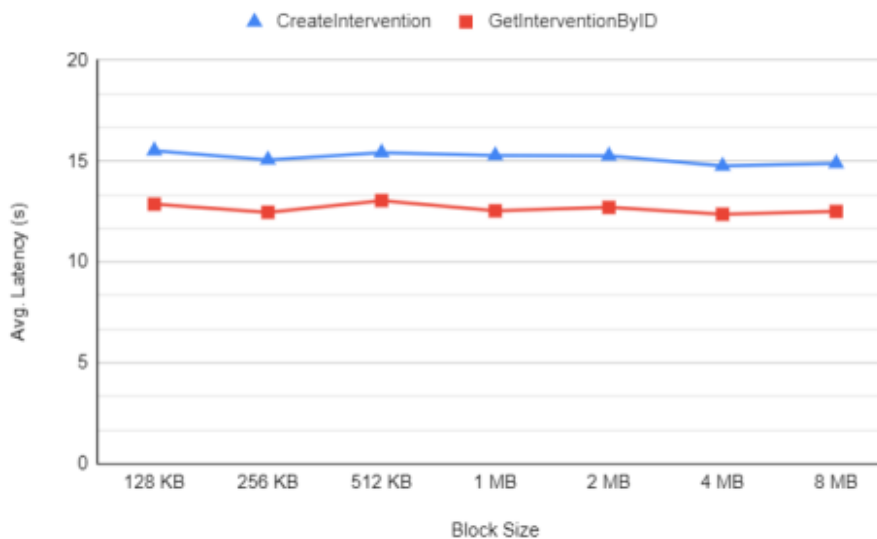


Рис. 8. Результати затримки (вимірюється в секундах) при зміні розміру блоку (з використанням 10 транзакцій на блок)

На рис. 9 та 10 показано, як збільшується пропускна здатність при збільшенні кількості транзакцій у блоці, що узгоджується з очікуваними результатами та подібними оцінками Hyperledger Fabric.

Однак затримка досягла стабільного рівня після досягнення 50 транзакцій в блоці, не показуючи істотних змін після цього моменту.

Важливо розуміти, що конкретне обладнання тестового середовища та конфігурація мережі суттєво впливають на ці показники продуктивності.

Таким чином, результати можуть відрізнятися в різних установках, що підкреслює важливість врахування характеристик тестового середовища при оцінці продуктивності.

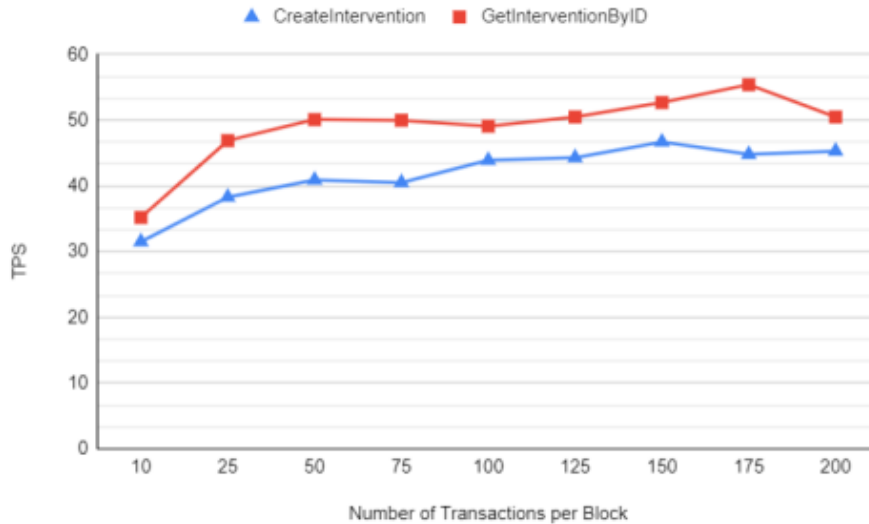


Рис. 9. Результати пропускної здатності (виміряної в TPS) Hyperledger, використовувані функції Fabric network і chain при варіюванні кількості транзакцій в блоці (з використанням розміру блоку 2 МБ).

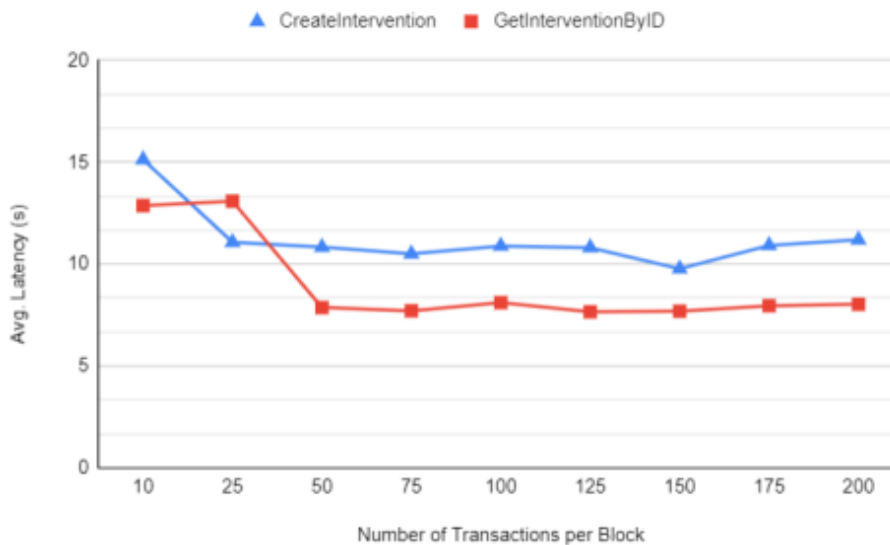


Рис. 10. Результати затримки (вимірюється в секундах) при варіюванні кількості транзакцій в блоці (з використанням розміру блоку 2 МБ)

Крім того, оцінка додаткового сховища, необхідного для системи, була проведена шляхом моніторингу використання дискового простору після пакетів транзакцій. Як показано на рис. 11, кожна партія з 2000 транзакцій призвела до збільшення дискового простору в середньому на 11%. Ці результати підкреслюють компроміс між покращеною конфіденційністю, цілісністю та незмінністю, пропонованими блокчейном, та підвищеними вимогами до сховища, які це тягне за собою, що є важливим фактором для систем освіти, які розглядають для впровадження

блокчейна.

В результаті було прийнято рішення зберегти в Ethereum тільки окремі хмарні адреси, що тягне за собою невелику плату за газ. З іншого боку, наша система є більш ефективною, ніж будь-яка інша система. Графік залежності плати за газ від розміру вхідних даних у байтах показано на рис. 9. На цьому графіку можна наочно бачити, що при збільшенні розміру даних плата за газ автоматично збільшується. Таким чином, розмір даних, що зберігаються в блокчейн, повинен бути якомога менше.

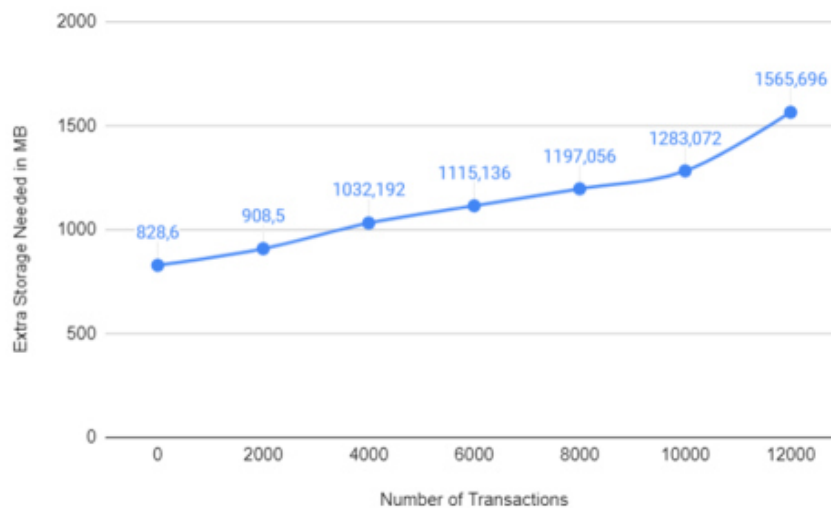


Рис. 11. Додаткове сховище, необхідне для впровадження технології блокчейн для токенизації студентських даних

Висновки

Це дослідження вносить значний внесок у вдосконалення блокчейн-системи для токенизації персональних даних студентів шляхом впровадження практичних функцій. Налаштування адаптера блокчейна для полегшення взаємодії з блокчейном, що дозволяє обробляти і візуалізувати студентські записи за допомогою зручних графічних інтерфейсів. Підвищена безпека, яка досягається завдяки дворівневим налаштуванням, гарантує, що тільки перевірені користувачі зможуть отримати доступ до певних функцій смарт-контрактів, тим самим зменшуючи потенційні проблеми. Механізм хешування, що гарантує цілісність і узгодженість даних в системі. Настроюваний контроль даних, що дозволяє студентам керувати своїми записами та ділитися лише інформацією, яку вони вибирають. Функція вибіркового доступу роботодавців, що дозволяє їм переглядати конкретні студентські записи, пов'язані з їх дипломами, без доступу до всього файлу персональних даних.

Впровадження та оцінка моделі для перевірки здійсненності, стабільності та надійності пропонованої блокчейн-системи в контексті токенизації персональних даних студентів. Важливо відзначити, що блокчейн не є єдиним рішенням для сфери освіти, але це дослідження демонструє життєздатність і надійність використання унікальних атрибутів блокчейна.

Отримані результати свідчать про те, що інтеграція блокчейна з платформою для токенизації персональних даних студентів не тільки здійсненна, але і

пропонує суттєві переваги, включаючи безпечне зберігання транзакцій і незмінний, що піддається перевірці журнал транзакцій для учасників мережі. Конструкція цієї системи також дозволяє розділити особисті та медичні дані, забезпечуючи відстеження даних та анонімність.

Хоча інтеграція блокчейна і систем управління освітніми даними все ще знаходиться на стадії становлення, незаперечні переваги розподіленого, незмінного журналу транзакцій очевидні. Оскільки технологія блокчейн продовжує розвиватися завдяки зниженню витрат і спрощенню розгортання, а також у міру розробки нових автономних методів зберігання даних, інтегровані в блокчейн ІОТ-системи, ймовірно, стануть нормою для рішень обміну персональними даними студентів. Ключовими висновками є:

Результати дослідження підтверджують доцільність використання блокчейна для токенизації персональних даних студентів. Конфіденційність можна ефективно підтримувати, поєднуючи хмарне сховище з технологією блокчейн, щоб відповідати стандартам конфіденційності.

У міру зростання мережі, що включає все більше об'єктів (організацій, вузлів, користувачів), підтримка продуктивності для забезпечення якості обслуговування для всіх учасників має вирішальне значення. Вирішення потенційних проблем масштабованості, таких як підвищена затримка при більшій кількості транзакцій і великих розмірах блоків, може зажадати впровадження протоколу сегментації для допустимих блокчейнів..

СПИСОК ЛІТЕРАТУРИ

1. Nair, M. M., & Tyagi, A. K. (2023). Blockchain technology for next-generation society: current trends and future opportunities for smart era. *Blockchain Technology for Secure Social Media Computing*, 10.
2. Dash, S. P. (2023). An introduction to blockchain technology: Recent trends. *Recent Advances in Blockchain Technology: Real-World Applications*, 1-24.
3. Magomadova, M. M., Semenova, L. V., & Artamonov, V. S. (2023). The Analysis of Blockchain Technologies Usage in Global Financial Markets. In *SHS Web of Conferences (Vol. 172, p. 05010)*. EDP Sciences.
4. Salvador, B. S., & Simon, A. T. (2024). Perspectives for the implementation of blockchain technology in the supply chain: research, trends and future. *Exacta*, 22(1), 119-143.
5. Oleksandr, Shmatko, Gorbach Tetiana, and Zherzherunov Pavlo. "Innovative model for supply chain management using blockchain technology." The 9th International scientific and practical conference "Theoretical and practical aspects of the development of science and education" (March 05–08, 2024) Prague, Czech Republic. Int. Science Group. 2024. 349 p.. 2024.

6. Shmatko, O., Litvinova, J., & Shokin, M. (2023). New blockchain-based supply chain management system model. *Scientific Collection «InterConf+»*, (33 (155)), 470-479.
7. Wenhua, Z., Qamar, F., Abdali, T. A. N., Hassan, R., Jafri, S. T. A., & Nguyen, Q. N. (2023). Blockchain technology: security issues, healthcare applications, challenges and future trends. *Electronics*, 12(3), 546.
8. Shuhaimi, J., Awang, H., & Jafar, M. F. (2024). Worldwide research history and trends on Blockchain applications in education: A bibliometric analysis, 2017-2022. *Multidisciplinary Reviews*, 7(3), 2024053-2024053.
9. Molopa, S. T., & Cronje, J. (2024, March). Research on Blockchain Adoption in Higher Education: A Systematic Review and Conceptual Model. In *Future of Information and Communication Conference* (pp. 110-130). Springer Nature Switzerland.
10. Haque, M., Kumar, V. V., Singh, P., Goyal, A. A., Upreti, K., & Verma, A. (2023). A systematic meta-analysis of blockchain technology for educational sector and its advancements towards education 4.0. *Education and Information Technologies*, 28(10), 13841-13867.
11. Jain, R., Seth, N., Sood, K., & Grima, S. (2023). Mapping the Research on Blockchain in Education: A Systematic Review and Bibliometric Analysis. *Digital Transformation, Strategic Resilience, Cyber Security and Risk Management*, 111, 53-66.
12. Razia, B. (2021, November). A systematic review of the use of blockchain in higher education. In *International Conference on Business and Technology* (pp. 631-648). Cham: Springer International Publishing.
13. De Alwis, T., Dewasiri, N. J., & Sood, K. (2023). Persistent Fiscal Deficit as a Determinant of Inflation in Sri Lanka: An ARDL Approach. In *Digital Transformation, Strategic Resilience, Cyber Security and Risk Management* (pp. 27-51).
14. Al Hemairy, M., Abu Talib, M., Khalil, A., Zulficar, A., & Mohamed, T. (2024). Blockchain-based framework and platform for validation, authentication & equivalency of academic certification and institution's accreditation: UAE case study and system performance (2022). *Education and Information Technologies*, 1-30.
15. Petrovska, I. and Kuchuk, H. (2023), "Adaptive resource allocation method for data processing and security in cloud environment", *Advanced Information Systems*, vol. 7(3), pp. 67–73, doi: <https://doi.org/10.20998/2522-9052.2023.3.10>
16. Biswas, A. K., Dasgupta, M., & Ray, S. (2023, January). Secure Management of Digital Academic Certificates Using Blockchain Technology. In *Machine Learning, Image Processing, Network Security and Data Sciences: Select Proceedings of 3rd International Conference on MIND 2021* (pp. 805-813). Singapore: Springer Nature Singapore.
17. Rustemi, A., Dalipi, F., Atanasovski, V., & Risteski, A. (2023). A systematic literature review on blockchain-based systems for academic certificate verification. *IEEE Access*.
18. Petrovska, I., Kuchuk, H., Mozhaiev, M. (2022), Features of the distribution of computing resources in cloud systems, 2022 IEEE 4th KhPI Week on Advanced Technology, KhPI Week 2022 - Conference Proceedings, 03-07 October 2022, Code 183771, doi: <https://doi.org/10.1109/KhPIWeek57572.2022.9916459>
19. Pu, S., & Lam, J. S. L. (2023). The benefits of blockchain for digital certificates: A multiple case study analysis. *Technology in Society*, 72, 102176.
20. Maestre, R. J., Bermejo Higuera, J., Gámez Gómez, N., Bermejo Higuera, J. R., Sicilia Montalvo, J. A., & Orcos Palma, L. (2023). The application of blockchain algorithms to the management of education certificates. *Evolutionary Intelligence*, 16(6), 1967-1984.
21. Swan, M. (2019). Blockchain economic networks: Economic network theory—Systemic risk and blockchain technology. *Business Transformation through Blockchain: Volume I*, 3-45.
22. Shmatko, O., Borova, T., Yevseiev, S., & Milov, O. (2021). Tokenization of educational assets based on blockchain technologies. *Journal «ScienceRise: Pedagogical Education»* №, 3, 42.

Received (Надійшла) 03.04.2024

Accepted for publication (Прийнята до друку) 05.06.2024

A method of simulation of computer network traffic with fractal properties

Olexander Shmatko, Danylo Safankov

Abstract. Topicality.. The modern world is going through an important transition period from the stage of informatization of important aspects of human activity to the stage of their digitalization. Informatization is a process of modernization of various fields of activity through information and communication technologies, while digital transformation means a deep rethinking and reformation of these areas based on digital solutions and technologies. This process requires not only maintaining the existing conditions for the development of advanced digital platforms and technologies, but also creating new opportunities for the emergence of innovative solutions. The key technologies in this digital age are big data, Neurotechnologies and artificial intelligence, blockchain, quantum technologies, the latest manufacturing techniques, the industrial internet, elements of robotics and sensors, wireless communication technologies and innovations in the field of virtual and augmented reality. In this paper, in the context of digital transformation of the educational sphere, a study on the use of blockchain technologies for tokenization of educational resources is conducted and the potential of their application in education is analyzed. The goal of this work is to increase the confidentiality of educational data, their integrity and availability, as well as to ensure reliable exchange of this data between all stakeholders through the design, development and implementation of software components of the cloud platform for tokenization of students' personal data based on blockchain technology. The object of research includes systems for transmitting and processing educational data that ensure data exchange between all interested parties in the educational process. The subject of research is methods and tools for designing and developing software components necessary for creating and maintaining platforms for tokenizing students' personal data. These components include software for encryption, authentication, authorization, as well as mechanisms for ensuring fault tolerance and data recovery. Results In this paper, a model of a cloud-based decentralized system for tokenizing students' personal data is proposed. **Conclusions.** The integration of blockchain technology into the education sector has great potential to build trust and improve the quality of educational services, as well as to create a new ecosystem of educational assets that are accessible and proven globally. The implementation of blockchain technology in the field of Education requires the development of specialized software for issuing and verifying digital documents based on the blockchain. The introduction of a cloud-based platform for tokenizing students' personal data promises significant improvements in the processing and storage of educational documents, contributing to greater trust and transparency in the field of Education.

Keywords: blockchain, personal data of students, cloud data storage, smart contracts, Ethereum, architectural model of the cloud data tokenization platform.

А. В. Шостак

Національний аерокосмічний університет імені М.Є. Жуковського, Харків, Україна

ПРО ОСОБЛИВОСТІ ФОРМУВАННЯ ВХІДНИХ ДАНИХ У СІАМСЬКІЙ НЕЙРОННІЙ МЕРЕЖІ

Анотація. Проаналізовано різні способи формування вхідних даних та оцінок дескрипторів сіамської нейронної мережі (СНМ) для порівняння зображень рукописних цифр. Запропоновано спосіб використання квазівипадкової N -вимірної послідовності векторів, сформованих відповідно до методу Соболя, для формування дескрипторів, які, спільно із зображеннями, є вхідними даними для навчання СНМ та її подальшого використання. Виконано тестування сіамської нейронної мережі з використанням отриманих оцінок дескрипторів зображень рукописних цифр. Під час тестування СНМ використовувався набір MNIST. Результат тестування моделі СНМ дав значення показника асигасу, що дорівнює 0.9706. Результати тестування засвідчили, що розглянуті оцінки дескрипторів h_1 і h_2 зменшують кількість помилок під час тестування порівняно з використанням дескриптора h_0 на основі квазівипадкової N -вимірної послідовності векторів.

Ключові слова: сіамська нейронна мережа, дескриптор, тестування нейронної мережі.

Вступ

Сіамська нейронна мережа (СНМ) зазвичай містить три підмережі [1, 2]. Перші дві підмережі однакові, їхні входи є входами всієї СНМ і вони обчислюють представлення входів мережі. Остання підмережа виконує порівняння представлень і видає результат роботи

СНМ у вигляді числа в діапазоні від 0 до 1. Сферою використання СНМ є системи, у яких виконується порівняння вхідних даних на ідентичність [1–3].

СНМ формує значення показника, яке оцінює ступінь близькості вхідних даних [1–4].

У відомих виданнях не повною мірою описано особливості формування вхідних даних і використання представлень для навчання і тестування СНМ.

Основна частина

Узагальнену структуру СНМ представлено й описано в [1, 2].

Вхідні дані x_1 і x_2 за допомогою перших двох підмереж перетворюються на N -вимірні представлення (або дескриптори) $h(x_1)$ та $h(x_2)$. Перші дві підмережі ідентичні за своїми параметрами W . Третя підмережа СНМ обчислює евклідову відстань між представленнями вхідних даних і на підставі неї формує значення оцінки ступеня подібності в діапазоні від 0 до 1.

Пропонується для навчання СНМ та її тестування використовувати архітектуру тільки з однією вхідною підмережею (рис. 1). На один вхід СНМ надходить N -вимірний дескриптор j -ї цифри h_0j , на

другий вхід – відображення цифри x з бібліотеки MNIST [5].

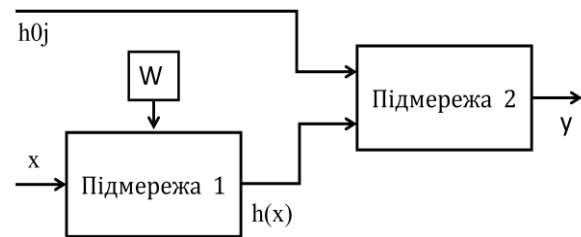


Рис. 1. Архітектура СНМ під час навчання і тестування

Підмережі 1 і 2 на рисунку 1 аналогічні відповідним підмережам узагальненої структури СНМ [2]. Дескриптор $h(x)$ є поданням вхідного зображення x . Дескриптори для десяти цифр h_0 є елементами квазівипадкової N -вимірної послідовності з десяти векторів, сформованої відповідно до методу Соболя [6, 7].

Було проаналізовано квазівипадкові N -вимірні послідовності векторів, що генеруються за методами Холтона [8], Соболя [6, 7] і R-послідовність [7]. N -вимірні послідовності векторів за методом Соболя показала найбільше значення за показником середня відстань між векторами. Тому послідовність векторів за методом Соболя і була використана для формування десяти дескрипторів цифр, які подаються на перший вхід СНМ (рис. 1) і використовуються для навчання СНМ. Перші десять 10-вимірних векторів квазівипадкової послідовності Соболя, згенеровані з використанням бібліотеки SciPy, мають такий вигляд:

- 0) [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
- 1) [0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5]
- 2) [0.75 0.25 0.25 0.25 0.75 0.75 0.25 0.75 0.75 0.75]
- 3) [0.25 0.75 0.75 0.75 0.25 0.25 0.75 0.25 0.25 0.25]
- 4) [0.375 0.375 0.625 0.875 0.375 0.125 0.375 0.875 0.875 0.625]
- 5) [0.875 0.875 0.125 0.375 0.875 0.625 0.875 0.375 0.375 0.125]
- 6) [0.625 0.125 0.875 0.625 0.625 0.875 0.125 0.125 0.125 0.375]
- 7) [0.125 0.625 0.375 0.125 0.125 0.375 0.625 0.625 0.625 0.875]
- 8) [0.1875 0.3125 0.9375 0.4375 0.5625 0.3125 0.4375 0.9375 0.9375 0.3125]
- 9) [0.6875 0.8125 0.4375 0.9375 0.0625 0.8125 0.9375 0.4375 0.4375 0.8125]

Навчання СНМ та її тестування проводилося за допомогою Google Colaboratory. Основні складові підмереж 1 і 2 (рис. 1) докладно описані в [2, 4].

При компіляції моделі СНМ використовувалася функція контрастних втрат [3, 4], основні параметри функції fit навчання СНМ - розмір пакета рівний 16 і число епох рівне 10.

Результат тестування моделі СНМ за допомогою методу evaluate такий - значення показника assiguas дорівнює 0.9706, а loss - 0.0226.

Після навчання СНМ з тренувального набору зображень рукописних цифр на виході підмережі 1 (рисунок 1) було отримано кластери 10-вимірних дескрипторів цифр, а також на основі цих кластерів обчислено оцінки дескрипторів цифр.

Оцінки дескрипторів цифр виконувались в такий спосіб [2]:

1. h_1 – дескриптор, що забезпечує мінімум суми евклідових відстаней на відповідному кластері,

2. h_2 – дескриптор, який є рішенням методу Нелдера-Міда на відповідному кластері.

Наприклад, дескриптор h_{10} має такий вигляд:

$h_{10} = [-0.9991518 \quad -0.99878484 \quad -0.9936094 \quad -0.9974742 \quad 0.99989784 \quad -0.9944306 \quad -0.9998869 \quad -0.97155184 \quad -0.9997868 \quad -0.97745085]$.

Для пошуку дескрипторів h_2 було використано метод Нелдера-Міда з Python-бібліотеки SciPy. Дескриптор h_{20} для цифри 0 має такий вигляд:

$h_{20} = [-0.99668455 \quad -0.9942764 \quad -0.99217147 \quad -0.9915273 \quad 0.9961964 \quad -0.9922386 \quad -0.9978741 \quad -0.9734324 \quad -0.9987575 \quad -0.9803164]$.

У табл. 1 наведено розраховані з використанням дескриптора h_2 характеристики кластерів дескрипторів цифр - мінімальна відстань (r), максимальна відстань (радіус кластера) між дескрипторами

кластера і дескриптором h_2 (R), середня всерединікластерна відстань (ssR) і середнє квадратичне відхилення (sD). Кількість даних у табл. 1 для цифр від 0 до 9 відповідно така - 2961, 3423, 2948, 3073, 2926, 2709, 2975, 3107, 2875, 3003.

Кластер дескрипторів зображень цифри 9 найбільш некомпактний за показником R (за радіусом кластера). Кластер дескрипторів зображень цифри 3 найнекомпактніший за показником ssR (за середньою всерединікластерною відстанню). Найкомпактніший за показником ssR кластер дескрипторів цифри 0. Найкомпактніший за показником R кластер дескрипторів зображень цифри 1.

Тестування роботи дескрипторів h_0 , h_1 і h_2 виконувалося в архітектурі СНМ з рисунка 1 за методикою, наведеною в [2].

Таблиця 1 – Характеристики кластерів дескрипторів цифр

Цифра	r	R	ssR	sD
0	0.00996	5.17015	0.30645	0.58298
1	0.23036	4.00553	0.97232	0.55612
2	0.03950	4.60219	0.89012	0.97916
3	0.09604	5.32569	1.16170	0.87623
4	0.07548	4.04438	0.92428	0.70884
5	0.08129	4.81103	1.05936	0.92620
6	0.04481	4.22272	0.69917	0.74386
7	0.03546	4.87599	1.00177	0.84371
8	0.15787	4.46806	1.10556	0.82220
9	0.03347	5.87126	0.73154	0.86083

У табл. 2 представлені підсумки роботи СНМ при подачі на другий вхід даних для цифри 0 з тестового набору з використанням дескрипторів h_0 , h_1 і h_2 для цифр від 0 до 9.

Таблиця 2 – Результати тестування на тестовому наборі зображень із цифрою 0 з використанням дескрипторів h_0 , h_1 і h_2 для цифр від 0 до 9

Цифра	0	1	2	3	4	5	6	7	8	9
Помилки для h_0	22	18	5	8	3	0	1	5	0	0
y	3.86e-05	0.96	1.00	0.99	0.94	0.42	0.91	0.99	0.43	0.38
Помилки для h_1	16	32	1	0	0	1	0	1	0	0
y	5.43e-06	1.00	0.92	0.05	0.07	0.74	0.49	0.94	0.02	0.00
Помилки для h_2	15	35	1	0	0	1	0	1	0	0
y	5.58e-06	0.99	0.93	0.07	0.06	0.76	0.48	0.95	0.04	0.00

У стовпчику з цифрою 0 показано результат тестування, коли на перший вхід СНМ (рис. 1) подають значення дескриптора h_{00} для цифри 0 (рядки 2, 3), а на вхід x – дані для цифри 0. Цифра 22 у другому рядку стовпчика 0 означає, що для h_{00} вихід СНМ для 22 зображень цифри 0 з 980 був меншим за 0.5, причому 3.86e-05є мінімальне значення виходу СНМ. Тобто для 22 зображень цифри 0 винесено рішення, що вони не відповідають дескриптору h_{00} цифри 0. На рис. 2 показана цифра з міткою 0, що відповідає значенню $y=0.00020055$, коли на першому вході дескриптор h_{00} .

Цифра 18 у другому рядку стовпчика з цифрою 1 показує, що для h_{01} вихід СНМ для 18 зображень цифри 0 з 980 був більшим за 0.5, максимальне значення y дорівнює 0.96. Тобто для 18 зображень ви-

несено рішення, що вони відповідають дескриптору h_{01} цифри 1.

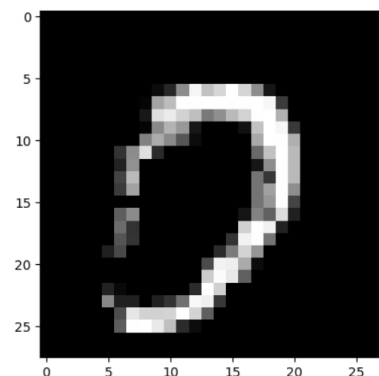


Рис. 2. Цифра з міткою 0 (для $h_{00} y = 0.00020055$)

На рис. 3 показана цифра з міткою 0 для значення $y=0.9560281$, коли на першому вході дескриптор h_{01} .

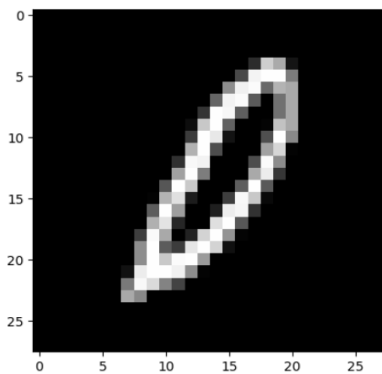


Рис. 3. Цифра з міткою 0 (для $h_{01} y = 0.9560281$)

Таким чином, за використання дескриптора h_0 було 62 помилки на 980 зображень цифри 0, що становить 6.3%, за використання дескриптора h_1 - 51 помилка (5.2%) і за використання дескриптора h_2 - 53 помилки (5.4%).

Наведені в табл. 2 результати тестування зображень із цифрою 0 показують, що розглянуті оцінки дескрипторів h_1 і h_2 значно зменшують кількість помилок під час тестування.

Висновки

У роботі запропоновано спосіб використання квазिवипадкової N -вимірної послідовності векторів, сформованих відповідно до методу Соболя, для

формування дескрипторів, які, спільно із зображеннями, є вхідними даними для навчання СНМ та її подальшого використання.

У роботі також проаналізовано способи побудови оцінок дескрипторів для порівняння і розпізнавання зображень для СНМ.

Використана для навчання і тестування архітектура СНМ складається з двох підмереж. Перша підмережа формує уявлення вхідного зображення, що аналізується. На перший вхід другої підмережі подають дескриптор, який використовують під час навчання або отримані оцінки дескрипторів, а на другий вхід - дескриптор зображення, що аналізується.

Під час тестування СНМ використовувався набір MNIST. Результат тестування моделі СНМ дав значення показника accuracy, що дорівнює 0.9706.

Результати тестування засвідчили, що розглянуті оцінки дескрипторів h_1 і h_2 зменшують кількість помилок під час тестування порівняно з використанням дескриптора h_0 на основі квазिवипадкової N -вимірної послідовності векторів.

Під час випробування 980 зображень із цифрою 0 у разі використання дескриптора h_0 кількість помилок становила 62, за використання дескриптора h_1 - 51 помилка і за використання дескриптора h_2 - 53 помилки.

У подальших дослідженнях слід приділити увагу вибору та аналізу впливу показників квазिवипадкової N -вимірної послідовності дескрипторів на якість навчання та на точність порівняння зображень у сіамській нейронній мережі.

СПИСОК ЛІТЕРАТУРИ

1. Chicco D. Siamese Neural Networks: An Overview. *Artificial Neural Networks*. 2021. MIMB, vol. 2190. P. 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 ШП, 2021. Вип. 4(66). С. 91–96. DOI: <https://doi.org/10.26906/SUNZ.2021.4.079>
3. Contrastive loss for Siamese networks with Keras and TensorFlow. URL: <https://www.pyimagesearch.com/2021/01/18/contrastive-loss-for-siamese-networks-with-keras-and-tensorflow/>
4. Image similarity estimation using a Siamese Network with a contrastive loss. URL: https://keras.io/examples/vision/siamese_contrastive/
5. The Mnist database of handwritten digits. URL: <http://yann.lecun.com/exdb/mnist/>
6. 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
7. Roberts M. The Unreasonable Effectiveness of Quasirandom Sequences. 2020. URL: <https://extremelarning.com.au/unreasonable-effectiveness-of-quasirandom-sequences/>
8. Halton J. H. On the efficiency of certain quasi-random sequences of points in evaluating multidimensional integrals. *Numer. Math.* 1960. Vol. 2. P. 84–90. DOI: <https://doi.org/10.1007/BF01386213>

Received (Надійшла) 29.05.2024

Accepted for publication (Прийнята до друку) 21.08.2024

On the features of the formation of input data in the Siamese neural network

A. Shostak

Abstract. Various methods of forming input data and descriptor estimates of the Siamese neural network (SNM) for comparing images of handwritten digits are analyzed. A method of using a quasi-random N -dimensional sequence of vectors, formed according to Sobol's method, is proposed for the formation of descriptors, which, together with images, are input data for training SNM and its further use. Siamese neural network testing was performed using the received evaluations of image descriptors of handwritten digits. The MNIST set was used during SNM testing. The result of testing the SNM model gave a value of the accuracy indicator equal to 0.9706. The test results showed that the considered estimates of the descriptors h_1 and h_2 reduce the number of errors during testing compared to the use of the descriptor h_0 based on a quasi-random N -dimensional sequence of vectors.

Keywords: Siamese neural network, descriptor, neural network testing.

Yuliia Andrusenko

Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

IAAS PERFORMANCE ASSESSMENT WITH SERVICE LEVELS

Abstract. The paper proposes a modification of the IaaS cloud model. To demonstrate the practicality and competitiveness of the method, a comprehensive performance study is conducted using simulation. Workloads based on real production runs of heterogeneous HPC systems are used to evaluate the practicality of the scheduling method. An online scheduling problem is considered. Jobs arrive one after another, and after a new job arrives, the scheduler must decide whether to reject this incoming job or schedule it on one of the machines. The problem is online, since the scheduler must solve it without information about the next jobs.

Keywords: IAAS, service levels, distribution, jobs.

Introduction

Infrastructure as a Service (IaaS) is a cloud computing service model where computing resources are hosted in a public cloud, private cloud, or hybrid cloud. With the IaaS model, it is possible to partially or completely move an on-premises or distributed data center infrastructure to the cloud, where it is maintained and managed by a cloud provider [1–4].

The main problem of the paper is online scheduling. Jobs arrive one after another [5, 6]. When a new job arrives, the scheduler must decide whether to reject this new job or schedule it on one of the machines. The problem is online. Because the scheduler must solve it without information about the next jobs [7–9]. For this problem, the performance of the algorithms is evaluated by a set of metrics. They include the contention factor and the number of accepted jobs [10].

In the beginning, two greedy algorithms with one service level are investigated [11]. In both cases, the key properties of the service level must be met to provide benefits for real-world settings. Since the service level is often considered as a successor to the real-time paradigm, the service-oriented one with deadlines [12–14].

Basic material

Consider a set of service levels offered by an service level agreement (SLA). Let it be $S = [S_1, S_2, \dots, S_l, \dots]$. For a given SL_j^l , a job j requires a throughput of s_j^l , which is guaranteed by the provisioning of the corresponding virtual machine VM, and incurs a cost Sl per unit of execution time depending on the required urgency. This urgency is expressed through the slack factor $f_j^l \geq 1$; $u_{max} = \max \{u_j^l\}$ denotes the maximum cost for all $l=1 \dots k$ and $j=1 \dots n$. The total number of jobs submitted to the system is n_r .

Each job j is described by a tuple $\langle r_j, w_j, d_j, S_j, SL_j^l \rangle$ containing the release date r_j , the amount of work w_j describing the computational load that must be executed before the required response time, the deadline d_j , and the service level SL_j^l .

Let $p_j = w_j / s_j^l$ be the guaranteed time that the system will spend processing the job before the deadline, according to the service level SL_j^l .

Let d_j be the latest time by which the system must execute job J_j if it is accepted. This value is calculated when a job is accepted as $d_j = r_j + f_j^l p_j$. The maximum

time for a job to complete is $d_{max} = \max \{d_j^l\}$. When a job is submitted for execution, its characteristics are known.

The revenue that the system will receive for executing job J_j is calculated as $u_j^l p_j$. Once a job has been submitted, the scheduler must decide whether to accept the job or not before other jobs arrive.

In order to accept job J_j , the scheduler must ensure that some machine in the system is capable of executing it before its deadline. If accepted, later submitted jobs cannot cause job J_j not to be executed before its deadline.

Once a job is accepted, the scheduler uses some rule to generate an execution plan. The set of accepted jobs $J = [J_1, J_2, \dots, J_n]$ is a subset of the incoming jobs, where $n \leq n_r$ is the number of accepted jobs.

Consider a set of heterogeneous machines $M = [M_1, M_2, \dots, M_m]$. Each machine M_i is described by a tuple $\langle s_i, eff_i \rangle$ indicating its relative processing speed s_i and its energy efficiency eff_i .

At a time, only a subset of all machines can accept a job. Let $M_a(t) = [M_1, M_2, \dots]$ be such a set of admissible machines. This set is defined for each job as the subset of available machines that can execute the job without missing deadlines and can guarantee the processing power s_j for processing. Machines whose processing speed is less than the speed guaranteed by the SLA cannot accept the job.

The value of s_i is conservatively chosen such that the speedup of all applications exceeds s_i . Thus, users receive the same guarantees regardless of which processors are used. Deadlines are calculated based on the service level and cannot be changed, and the guaranteed processing time is not violated by slower processing. C_{max} denotes the execution time of all jobs.

Next, we consider a two-level scheduling approach (Fig. 1). At the upper level, the system checks whether the job can be accepted or not using the Greedy acceptance policy. If the job is accepted, the system selects a machine from the set of admissible machines to execute it at the lower level.

A greedy acceptance policy is used at the upper level. It is based on the EDD algorithm, which prioritizes jobs according to their deadlines. When a job J_j is submitted to the system, in order to determine whether to accept or reject it, the system searches for a set of machines that can execute job J_j before its deadline, ensuring that no job in a machine is late.

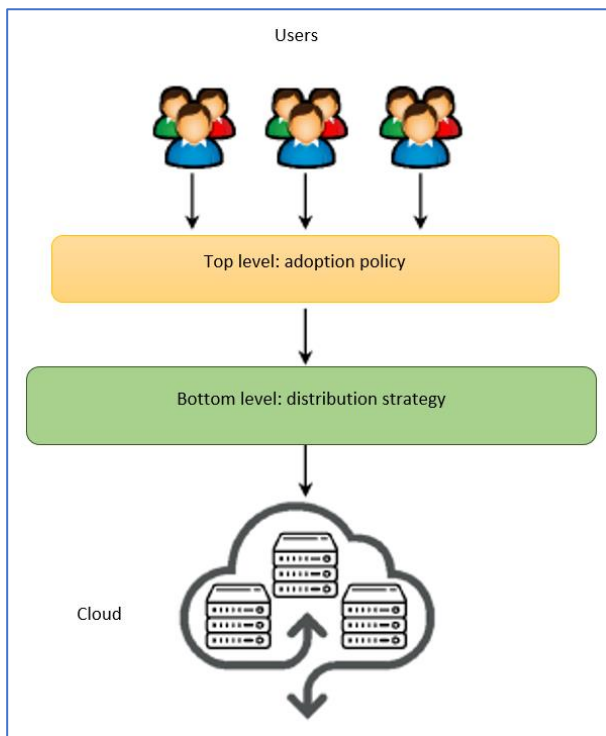


Fig. 1. Two-level scheduling approach using upper-level acceptance policy and lower-level allocation strategies

If the set of available machines is not empty ($|M_a(r_j)| \geq 1$), then job J_j is accepted, otherwise it is rejected. This completes the first stage of scheduling.

At the lower level, the Preemptive EDD algorithm is used with preemptions for each machine separately. This algorithm is easy to implement and gives an optimal

solution to the problem 1 |prmp, r_j , online| L_{max} . In general, the delay L_j of job J_j is defined as $(c_j - d_j, 0)$. Then we get $L_{max} = \max \{L_j\}$. For all plans in our problems, L_{max} must be satisfied, since none of the jobs can be late. Moreover, the Preemptive EDD algorithm creates a schedule without delays greedy scheduling and therefore does not postpone the use of resources to the future, when they may be needed by yet unknown jobs. Preemptive EDD verifies that all already accepted jobs whose due date is greater than the due date of the incoming job will be completed before their due date.

The machine for job distribution can be determined taking into account various criteria. They are characterized by the type and amount of information used to make the distribution decision.

Two levels of available information are distinguished. At level 1, it is assumed that the job execution time, machine speed and acceptance policy are known. At level 2, in addition, the machine energy efficiency and the energy consumed during the job execution are known.

Table 1 provides detailed information about the distribution strategies used in this work. The proposed strategies can be divided into three groups:

- 1 – without knowledge of the system, without information about jobs and resources;
- 2 – taking into account energy consumption, with information about energy consumption;
- 3 – with information about the speed of machines.

The jobs are ordered by decreasing deadlines. To execute them, jobs are taken from the head of the queue. When a new job is released, it is placed in the queue according to its deadline.

Table 1 – Job Distribution Strategies

Type	Strategy	Level	Description
Knowledge Free	Rand	1	Assigns work to a suitable machine chosen at random using uniform distribution over the range $[1 \dots m]$
	FFit	1	Distributes work j to the first available machine capable of performing it
	MLp	1	Distributes work j to the machine with the least load at time r_j : $\min\{n_i\}$
Energy aware	Max-eff	2	Distributes work to the machine with the greatest energy efficiency $\{eff_i\}$
	Min-e	2	Distributes work to the machine with the minimum total energy consumption at time r_j
	MCT-eff	2	Allocates work j to the machine with the best balance between execution time and energy efficiency, the execution time and the completion time of job k on machine i
Speed aware	Max-seff	2	Distributes work j to the machine
	Max-s	2	Distributes work j to the fastest machine: $\max\{s_i\}$

EDD is an optimal algorithm for minimizing delays in a single-machine system. In our case, this corresponds to minimizing the number of rejected jobs.

Since IaaS clouds are supposed to be a promising alternative to data centers, it can be expected that the workload transferred to clouds will have similar characteristics to those transferred to real parallel and grid systems.

It is well known that the distribution of jobs is uneven in time and depends on the time of day and day of the week.

Moreover, each individual log shows a different distribution. In addition, they are recorded in different time zones.

Therefore, it is necessary to normalize the used workloads by shifting the workloads by a certain time

interval in order to represent a more realistic situation. The workloads are transformed in such a way that all traces start on the same weekday and at the same time of day.

For this purpose, all jobs until the first Monday at midnight are deleted. Time zone normalization, profiled time interval normalization, and invalid job filtering should be considered.

Conclusions

Cloud computing provides advanced computing. The paper proposes a modification of the IaaS cloud model. To

demonstrate the practicality and competitiveness of the method, a comprehensive performance study is conducted using simulation.

Workloads based on real production runs of heterogeneous HPC systems are used to evaluate the practicality of the scheduling method.

An online scheduling problem is considered. Jobs arrive one after another, and after a new job arrives, the scheduler must decide whether to reject this incoming job or schedule it on one of the machines. The problem is online, since the scheduler must solve it without information about the next jobs.

REFERENCES

1. John J. Prevost, Kranthi Manoj Nagothu, Brian Kelley, and Mo Jamshidi, "Prediction of cloud data center networks loads using stochastic and neural models", 2011 6th International Conference on System of Systems Engineering, 12138393, 2011, doi: 10.1109/SYSOSE.2011.5966610.
2. S. Kianpisheh, and R. H. Glitho, "Cost-efficient server provisioning for deadline-constrained VNFs Chains: A parallel VNF processing approach", Proceeding of 2019 16th IEEE Annual Consumer Communications & Networking Conference, 2019, doi: 10.1109/CCNC.2019.8651799.
3. N. Kuchuk, O. Shefer, G. Cherneva, and F. A. Alnaeri, "Determining the capacity of the self-healing network segment", Advanced Information Systems, vol. 5, no. 2, pp. 114–119, Jun. 2021, doi: 10.20998/2522-9052.2021.2.16.
4. Ye. Qiang, and W. Zhuang, "Distributed and adaptive medium access control for internet-of-things-enabled mobile networks", IEEE Internet of Things Journal, 2017, vol. 4, no. 2, pp. 446–460, doi: 10.1109/JIOT.2016.2566659.
5. G. Kuchuk, S. Nechausov, and V. Kharchenko, "Two-stage optimization of resource allocation for hybrid cloud data store", International Conference on Information and Digital Technologies, Zilina, 2015, pp. 266–271, doi: 10.1109/DT.2015.7222982.
6. H. Khudov, K. Tahyan, V. Chepurnyi, I. Khizhnyak, K. Romanenko, A. Nevodnichii, and O. Yakovenko, "Optimization of joint search and detection of objects in technical surveillance systems", Advanced Information Systems, 2020, Vol. 4, No. 2, pp. 156–162, doi: 10.20998/2522-9052.2020.2.23.
7. S. Semenov, O. Sira, S. Gavrylenko, and N. Kuchuk, "Identification of the state of an object under conditions of fuzzy input data", Eastern-European Journal of Enterprise Technologies, Vol 1, No 4 (97), pp. 22–30, 2019, doi: 10.15587/1729-4061.2019.157085.
8. S. Semenov, and Cao Weilin, "Testing process for penetration into computer systems mathematical model modification", Advanced Information Systems, Vol. 4, No. 3, pp. 133–138. 2020, doi: 10.20998/2522-9052.2020.3.19.
9. H. Attar, M.R. Khosravi, S.S. Igorovich, K.N. Georgievan, and M. Alhihi, "E-health communication system with multiservice data traffic evaluation based on a G/G/1 analysis method", Current Signal Transduction Therapy, 16(2), 2021, doi: 10.2174/1574362415666200224094706.
10. P. Franti, "Efficiency of random swap clustering", Journal of Big Data, vol. 5, no. 13, 2018, pp. 1–29, doi: 10.1186/s40537-018-0122-y.
11. A. Nechausov, I. Mamuscu, and N. Kuchuk, "Synthesis of the air pollution level control system on the basis of hyperconvergent infrastructures", Advanced Information Systems, vol. 1, no. 2, 2017, pp. 21–26. DOI: 10.20998/2522-9052.2017.2.04.
12. A. Kovalenko, H. Kuchuk, N. Kuchuk, and J. Kostolny, "Horizontal scaling method for a hyperconverged network", International Conference on Information and Digital Technologies 2021, IDT 2021, pp. 331–336, 9497534, 2021, doi: https://doi.org/10.1109/IDT52577.2021.9497534.
13. H. Attar, M.R. Khosravi, S.S. Igorovich, K.N. Georgievan, and M. Alhihi, "Review and performance evaluation of FIFO, PQ, CQ, FQ, and WFQ algorithms in multimedia wireless sensor networks", International Journal of Distributed Sensor Networks, 16(6), June 2020, doi: https://doi.org/10.1177/1550147720913233.
14. N. Kuchuk, O. Mozhaiev, S. Semenov, A. Haichenko, H. Kuchuk, S. Tiulieniev, M. Mozhaiev, V. Davydov, O. Brusakova, and Y. Gnosov, "Devising a method for balancing the load on a territorially distributed foggy environment", Eastern-European Journal of Enterprise Technologies, vol. 1(4 (121)), pp. 48–55, 2023, doi: https://doi.org/10.15587/1729-4061.2023.274177

Received (Надійшла) 19.05.2024

Accepted for publication (Прийнята до друку) 21.07.2024

Оцінка продуктивності Інтернету хмарної моделі IaaS

Ю. О. Андрусенко

Анотація. У статті пропонується метод модифікації хмарної моделі IaaS. Щоб показати практичність та конкурентоспроможність методу, проведено комплексне дослідження продуктивності запропонованого методу за допомогою моделювання. Для оцінки практичності методу планування використовуються робочі навантаження, що ґрунтуються на реальних виробничих трасах гетерогенних систем. Розглядається проблема онлайн-планування. Роботи надходять одна за одною, і після надходження нової роботи планувальник повинен вирішити, чи відхилити йому поточну роботу, чи запланувати її на одній з віртуальних машин хмарної моделі IaaS. Проблема вирішується онлайн, оскільки планувальник повинен вирішувати її без інформації щодо наступних робіт.

Ключові слова: IAAS, рівні обслуговування, розподіл, роботи.

Я. І. Бірук, А. В. Климчук

Київський національний університет будівництва і архітектури, Київ, Україна

ЗАСАДИ ПРОЄКТУВАННЯ НЕОДНОРІДНИХ ЕЛЕКТРОМАГНІТНИХ ЕКРАНІВ

Анотація. Розглянуто два класи неоднорідних електромагнітних екранів – екранів на основі регулярних екрануючих елементів та матеріалів з хаотичним розташуванням металомістких частинок. Показано, що точне прогнозування захисних властивостей композиційних матеріалів з хаотичним розташуванням включень неможливе через емпіричний характер відомих співвідношень Лорентца, Максвелла-Гарнета, Оделевського. Запропоновано розроблення будівельних та оздоблювальних матеріалів з регулярним розташуванням захисного матеріалу у будівельних виробках. Головною проблемою надання будівельним і оздоблювальним матеріалам захисних властивостей є зниження основних параметрів цих матеріалів (пружних модулів) через додавання екрануючої субстанції. Для уникнення такого ефекту запропоновано використання технологічних порожнин у будівельних матеріалах. З цієї метою можливо використовувати сажу теплових електростанцій та залізородний концентрат. Теоретичні міркування свідчать, що додавання у вихідну суміш піно- та газобетону екрануючої добавки великої дисперсності у процесі піноутворення сприяє розподілу частинок добавки на внутрішніх поверхнях порожнин. Це значно підвищує захисні властивості матеріалів без зниження пружних модулів виробів. Надано розрахунковий апарат щодо ефективності екранування електромагнітних випромінювань сітчастими металевими структурами. Для надання конструкції широкосмуговості запропоновано застосування двох паралельних структур, налаштованих на максимальну та мінімальну довжину хвиль випромінювання, які визначаються експериментально. Показано, що застосування розв'язку хвильового рівняння для вектора Герца дозволяє отримати прості і зручні співвідношення для визначення мінімальних частот випромінювань, які проникають крізь отвори (частоти зрізу). Це дозволило визначити такі параметри для прямокутних та круглих отворів. Отримано співвідношення для розрахунку ефективності захисту за рахунок поглинання електромагнітної енергії та відбиття електромагнітних хвиль. При цьому показана можливість врахування площі і кількості отворів одиницю площі екрана. Виконані розрахунки стосуються суто геометричних характеристик захисних конструкцій з отворами. Перспективним напрямом досліджень є визначення захисних властивостей неоднорідних електромагнітних екранів з урахуванням електрофізичних та магнітних властивостей матеріалів, з яких конструкції виготовлені.

Ключові слова: електромагнітний екран, регулярна структура, будівельні матеріали, частота зрізу вектор Герца.

Вступ

Екранування є найбільш ефективним методом зниження рівнів електричних, магнітних та електромагнітних полів усіх частотних діапазонів. Традиційними матеріалами для виготовлення електромагнітних екранів є метали і сплави. Але вони мають суттєві недоліки і не завжди задовольняють сучасні вимоги. Головними недоліками металевих матеріалів є неможливість керування захисними властивостями та великі коефіцієнти відбиття електромагнітних хвиль. Цих недоліків позбавлені композиційні матеріали, у яких керування захисними властивостями здійснюється зміною вмісту металевих або металомісних компонентів у діелектричній матриці. Але проектування таких матеріалів пов'язане з низкою проблем. Головними показниками, які обумовлюють захисні властивості композиційних екрануючих матеріалів є електрофізичні та магнітні властивості – питома електропровідність, відносні діелектрична та магнітна провідності. Ці показники для металів і сплавів є довідковими величинами, але для композиційних матеріалів вони різняться у залежності від складу вихідної суміші. Наявні розрахункові методи визначення необхідних параметрів є емпіричними і мають недостатню точність. Крім того вони враховують тільки об'ємний вміст екрануючих добавок у діелектричну матрицю за умови, що такої добавки менше за третину об'єму. При цьому структура

композиту вважається однорідною. Тому постає задача визначення основних підходів до попередньої оцінки захисних властивостей матеріалу, що дозволить скоротити витрати часу й коштів для проектування матеріалу потрібних захисних властивостей.

Огляд публікацій і розробок у галузі проектування електромагнітних екранів. Більшість досліджень і прикладних розробок з проектування композиційних матеріалів для екранування електричних, магнітних та електромагнітних полів широкого частотного діапазону є експериментальним. Здебільшого вони стосуються матеріалів з максимально рівномірним розподілом екрануючих включень у діелектричній матриці.

У дослідженнях [1, 2] показано залежність коефіцієнтів екранування від вмісту дрібнодисперсних включень полімерну матрицю. При цьому встановлена залежність захисних властивостей від дисперсності частинок наповнювача. Робота [3] свідчить, що коефіцієнти екранування певним чином залежать від морфології екрануючих включень. Тобто, наявність певної неоднорідності за рахунок видовженої форми включень впливає на коефіцієнти екранування. Найбільш ефективний захист забезпечують наповнювачі нанорозмірів [4, 5]. Але застосування наноматеріалів автоматично призводить до високої собівартості кінцевого продукту. При цьому рівномірність розподілу частинок наповнювача забезпечується ультразвуковою обробкою вихідної суміші у рідкому стані або

застосуванням дисольверів з великими обертами [6, 7]. Такі технології виготовлення композитів доцільно використовувати для вироблення захисних матеріалів спеціального призначення – забезпечення електромагнітної сумісності чутливого електронного обладнання, технічного захисту інформації, військових потреб. Для широкого застосування вони мають високу вартість. Крім того, вони не можуть бути використані у якості оздоблювальних або будівельних матеріалів. Загальним недоліком усіх металополімерних композитів є ймовірність деградації під впливом фізичних або хімічних чинників [8]. Певним чином метастабільними є усі багатокомпонентні матеріали, на основі мікро- та наноматеріалів [9]. Неоднорідні матеріали сітчастого типу розглядають у роботі [10]. Зокрема показано, що в області дуже високих і вищих частот параметри екрана (відстань між дротами, їх діаметри) повинні однозначно відповідати частотним характеристикам екранованого електромагнітного поля, тобто такі екрани фактично монохромні. Аналогічний екран пропонується для екранування магнітного поля промислової частоти [11]. На сьогодні практично не приділяється уваги проблематиці розроблення і дослідження захисних властивостей композиційних матеріалів на основі макроstruktur, які можуть бути використані у якості оздоблювальних або будівельних матеріалів. Певним чином це обумовлене відсутністю методології розроблення таких матеріалів.

Мета роботи – розроблення загальних засад проєктування неоднорідних електромагнітних екранів на основі макроstruktur.

Викладення основного матеріалу

Макроскопічність екрануючих елементів у діелектричній матриці відносна. Вона залежить від довжини хвилі електромагнітного поля, яке екранується. Для електромагнітних випромінювань надзвичайно високих частот навіть частинки розмірами 150–200 мкм можна вважати макроскопічними. Це підтверджується експериментальними дослідженнями. У роботі [12] показано, що зі зростанням розмірів частинок відбувається різке підвищення коефіцієнта відбиття електромагнітних випромінювань ультрависокої частоти, що свідчить про суцільність екрануючої поверхні.

У загальному випадку існують два класи композицій – матеріали з регулярними екрануючими елементами та матеріали з хаотично розташованими частинками (округленими або видовженими). Ефективність екранування електромагнітних полів регулярними структурами можливо прогнозувати аналітичними методами, виходячи з електрофізичних властивостей матриці і матеріалу регулярної структури, відстаней між елементами цієї структури та її габаритних параметрів.

Для хаотично розміщених частинок наповнювача розрахувати прогнозовані параметри – діелектричну та магнітну проникності, які визначають захисні властивості, можливо із застосуванням формул Лорентца, Максвелла-Гарнета, Оделевського. При цьому формула Оделевського містить коефіцієнт

деполяризації, який враховує співвідношення довжини та ширини частинок. Але усі ці співвідношення емпіричні, тому їх застосування дає великі похибки, що за малих об'ємних вмістах екрануючого наповнювача є критичним.

Найбільш точно можна розрахувати захисні властивості металевих сітчастих регулярних структур. Коефіцієнт екранування сітки за нормального падіння електромагнітної сітки визначається співвідношенням:

$$K_s = 10 \lg \left(4 \left(\frac{d}{\lambda} \ln \frac{d}{2\pi r_0} \right)^2 / \left(1 + 4 \left(\frac{d}{\lambda} \ln \frac{d}{2\pi r_0} \right)^2 \right) \right),$$

де r_0 – радіус металевого дроту, d – середня відстань між дротами, λ – довжина падаючої електромагнітної хвилі.

У випадку падіння електромагнітної хвилі під деяким кутом до екрана:

$$K_r = 10 \lg \frac{4 \left((d \cos \varphi / \lambda) \cdot \ln (d / (2\pi r_0)) \right)^2}{1 + 4 \left((d \cos \varphi / \lambda) \cdot \ln (d / (2\pi r_0)) \right)^2},$$

$$K_s = 10 \lg \frac{(1 - \cos \varphi)^2 + 4 \left((d / \lambda) \cdot \ln (d / (2\pi r_0)) \right)}{1 + 4 \left((d / \lambda) \cdot \ln (d / (2\pi r_0)) \right)},$$

де φ – кут між напрямком розповсюдження електромагнітної хвилі і нормаллю до поверхні екрана.

Але, як зазначалося вище, недоліком таких структур є монохромність поглинання. Для подолання цього недоліку доцільно застосувати дві сітчасті регулярні структури, розташовані паралельно. Одна з них розраховується, виходячи з мінімально можливої довжини хвилі, друга – максимально можливої. Ці довжини хвиль визначаються або з технічної документації потенційних джерел електромагнітних полів, або експериментально. Такий підхід робить конструкцію широкосмуговою. Доцільно розглянути також можливість підвищення ефективності конструкції певним зсувом однієї сітки відносно іншої.

Найбільш привабливим способом забезпечення екранування електромагнітних полів широкого частотного діапазону є екранування полів будівельними матеріалами. Але будь-які екрануючі домішки до сировини змінюють міцнісні характеристики будівельних матеріалів (пружні модулі), що небажано, принаймні для стінових конструкцій та перекриттів. Тому для забезпечення захисних властивостей будівельних матеріалів екрануючий матеріал доцільно вмішувати у існуючі порожнини, як конструктивно закладені у цеглу, шлакоблоки, панелі тощо. Такі матеріали використовуються у великих обсягах, тому екрануючі добавки повинні мати мінімально можливу вартість. Зазвичай, високі коефіцієнти екранування непотрібні, тому для визначених цілей доцільно використовувати відходи або побічні продукти виробництва. Попередні дослідження показали, що коефіцієнти екранування електромагнітних полів на рівні 2,0–2,5 може забезпечити сажа, яка у великих обсягах утворюється на теплових електростанціях.

Для екранування стаціонарних та квазістаціонарних магнітних полів доцільно використовувати залізорудний концентрат, який у великих обсягах виробляється на гірничо-збагачувальних комбінатах України. Найбільш перспективним напрямом отримання високих коефіцієнтів екранування електромагнітних полів наднизької частоти (промислової) та ультрависоких і вищих частот є використання порожнин піно- та газобетону для розміщення екрануючих матеріалів. Виходячи з теоретичних міркувань, можна передбачити, що доданий у вихідну бетонну суміш залізорудний концентрат у процесі піноутворення буде осідати на внутрішніх поверхнях порожнин піно- та газобетону. Відомі експерименти з пінолатексом [13] показали, що додавання пилу із вмістом дрібнодисперсної залізної руди, отриманого в процесі подрібнення руди, навіть за товщини матеріалу 10 мм забезпечує коефіцієнти екранування магнітного поля промислової частоти – до 38, електромагнітного поля ультра високої частоти – до 52 (вміст екрануючого матеріалу – до 20 % за масою).

Значна частина будівельних матеріалів, особливо тих, що використовуються для оздоблювальних робіт є перфорованими. Використовується велика кількість перетинок перфорованої структури. Такі матеріали можливо застосовувати для цілей електромагнітної безпеки. Існують кілька простих співвідношень для визначення ефективності екранування електромагнітних полів матеріалами з отворами. Вони вважаються емпіричним і не мають широкого застосування. Але прості у застосуванні співвідношення можна строго отримати з теорії хвильоводів. Причому це можливо з будь-якою точністю у залежності від потреб. Для цього треба застосовувати хвильове рівняння для вектора Герца:

$$\frac{\partial^2 \Pi_z}{\partial x^2} + \frac{\partial^2 \Pi_z}{\partial y^2} + \frac{\partial^2 \Pi_z}{\partial z^2} + R^2 \Pi_z = 0$$

або

$$\frac{\partial^2 \Pi_z}{\partial x^2} + \frac{\partial^2 \Pi_z}{\partial y^2} + (k^2 - h^2) \cdot \Pi = 0,$$

де $\Pi = c \sin(m\pi x/a) \sin(n\pi y/b)$ (хвильовий вектор Герца), k – хвильове число, h – повздовжнє хвильове число, яке характеризує розповсюдження хвилі уздовж осі Z , m , n – натуральні коефіцієнти, a , b – ширина і довжина хвильоводу.

Наявність можливості розкладання функції Π у ряд забезпечує необхідну точність обчислень. Будь-який отвір можна вважати хвильоводом малої довжини, тому відповідні розрахунки є коректними. Розв'язуючи хвильове рівняння, обирається співвідношення k та h , яке відповідає найбільшій довжині хвилі, яка проникає крізь отвір. Таку довжину можна визначити як критичну. Якщо r – радіус отвору, то для прямокутного отвору $\lambda_{кр} = 2/\sqrt{1/a^2 + 1/b^2}$, для круглого отвору $\lambda_{кр} = 2\pi r/1,84 = 3,4r$.

Такі отвори є непроникними для електромагнітного випромінювання з довжиною хвилі $\lambda_{кр}$. У практичному застосуванні зручніше користуватися критичною частотою, так званою «частотою зрізу».

Для прямокутного отвору $f_3 = 1,5 \cdot 10^8 / b$, Гц, де b – найбільший розмір прямокутного отвору. Для круглого отвору $f_3 = 1,75 \cdot 10^8 / D$, Гц, де D – діаметр отвору. Тоді, застосовуючи стандартні визначення ефективності поглинання електромагнітних хвиль A , отримуємо для прямокутного отвору $A = 27,3 \cdot l/b$, дБ; для круглого отвору $A = 32 \cdot l/D$, дБ, де l – товщина стінки екрана.

Для розрахунку втрат на поглинання зони з отвором (апертурою), обирається частота, яка за наближеними співвідношеннями нижча за f_3 .

З наведених співвідношень можна зробити висновки, що для підвищення ослаблення поля необхідно збільшувати глибину апертури l . Втрати на відбиття у загальному випадку визначаються як:

$$R = 100 - 20 \lg(bf) + 20 \lg(1 + \ln(b/a)), \text{ дБ},$$

де b – ширина отвору, a – висота отвору, f – частота випромінювання.

Відповідно до круглого отвору втрати на відбиття розраховуються зі співвідношення

$$R = 102 - 20 \lg(df), \text{ дБ}$$

де d – діаметр отвору.

Кількість отворів (щільність розташування) враховується співвідношенням $K_n = 10 \lg(Sn)$, дБ, де S – площа одного отвору, n – кількість отворів на одиницю площі. Добуток Sn надає частку площі отворів на одиницю площі екрана у відносних одиницях. Тобто, коефіцієнт K_n має від'ємне значення ($Sn < 1$). Фізичне тлумачення цього факту полягає у тому, що ефективність екранування знижується зі збільшенням загальної площі отворів.

Було розглянуто суто геометричні критерії ефективності екранування неоднорідних провідних структур. У подальшому доцільно визначити ефективності екранування неоднорідних структур з урахуванням електрофізичних властивостей матеріалів з яких вони виготовлені.

Висновки

1. Розглянуто два класи неоднорідних матеріалів для екранування електромагнітних полів. Показано, що матеріали з хаотичним розташуванням екрануючих елементів складні щодо прогнозування їх захисних властивостей через емпіричний характер відповідних математичних співвідношень. Надано розрахунковий апарат щодо розрахунку ефективності сітчастих металевих структур. Показано, що застосування паралельних сітчастих структур робить конструкцію широкосмуговою.

2. Надано обґрунтовані рекомендації щодо підвищення екрануючих властивостей стандартних будівельних та оздоблювальних матеріалів. Для запобігання зниженню пружних модулів матеріалів екрануючі добавки розміщуються у існуючих порожнинах шлакоблоків, вироби із піно- та з газобетону. Для цього доцільно використовувати відходи теплових електростанцій та залізорудний концентрат.

3. Встановлено, що прості і зручні у використанні співвідношення для визначення захисних властивостей екрануючих поверхонь з отворами можна строго математично отримати з теорії хвильоводів. Для цього здійснюється розв'язок хвильового

рівняння для вектора Герца. Надано співвідношення для визначення критичних довжин хвиль (частот зрізу), які пропускаються отвором. Визначені коефіцієнти поглинання та відбиття для екранів з отворами прямокутної та круглої форм.

СПИСОК ЛІТЕРАТУРИ

- Glyva, V., Kasatkina, N., Nazarenko, V., Levchenko, L., Panova, O., Tykhenko, O., Khodakovskyy, O. Development and study of protective properties of the composite materials for shielding the electromagnetic fields of a wide frequency range. *Eastern-European Journal of Enterprise Technologies*, 2020, 2(12-104), pp. 40–47. <https://doi.org/10.15587/1729-4061.2020.201330>.
- Glyva V., Kovalenko V., Levchenko L., Tykhenko O. Research into protective properties of electromagnetic screens based on the metal-containing nanostructures. *Eastern-European Journal of Enterprise Technologies*. 2017. Vol. 3, № 12 (87). P. 50–56. <https://doi.org/10.15587/1729-4061.2017.103167>
- Касаткіна Н.В., Тихенко О.М., Панова О.В., Бірук Я.І. Підвищення ефективності композиційних електромагнітних екранів регулюванням морфології феромагнітного наповнювача. «Системи управління навігації та зв'язку», – 2020. Вип. № 3(61), С. 115-119. <https://doi.org/10.26906/SUNZ.2020.3.115>
- 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., Sucheа 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>
- Glyva, V., Podkopaev, S., Levchenko, L., Karaieva, N., Nikolaiev, K., Tykhenko, O., Khodakovskyy, O., & Khalmuradov, B. (2018). Design and study of protective properties of electromagnetic screens based on iron ore dust. *Eastern-European Journal of Enterprise Technologies*, 1(5 (91)), 10–17. <https://doi.org/10.15587/1729-4061.2018.123622>
- Glyva, V., Bakharev, V., Kasatkina, N., Levchenko, O., Levchenko, L., Burdeina, N., Guzii, S., Panova, O., Tykhenko, O., Biruk, Y. Design of liquid composite materials for shielding electromagnetic fields. *Eastern-European Journal of Enterprise Technologies*, 2021, 3(6-111), pp. 25–31. <https://doi.org/10.15587/1729-4061.2021.231479>
- 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. <https://doi.org/10.15330/pccs.21.4.771-778>
- Alina Ruxandra Caramitu, Ioana Ion, Adriana Mariana Bors, Violeta Tsakiris, Jana Pinteа, Ana-Maria Daniela Caramitu. 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>
- Butenko O., Boychuk V., Savchenko B., Kotsyubynsky V., Khomenko V., Barsukov V. Pure ultrafine magnetite from carbon steel wastes. *Materials Today: Proceedings*. 2019. V. 6, pp. 270–278.
- Панова О.В. Дослідження захисних властивостей металевих електромагнітних екранів та визначення умов їх максимальної ефективності. Системи управління навігації та зв'язку. 2020. 2(60), с. 127–130. <https://doi.org/10.26906/SUNZ.2020.2.127>
- Grinchenko V.S. Mitigation of three-phase power line magnetic field by grid electromagnetic shield. *Tekhnichna Elektrodynamika*. 2018. Vol. 2018, Issue 4. P. 29–32. <https://doi.org/10.15407/techned2018.04.029>.
- 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/techned2018.04.014>.
- 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.

Received (Надійшла) 24.04.2024

Accepted for publication (Прийнята до друку) 10.07.2024

Principles of designing inhomogeneous electromagnetic screens

Y. Biruk, A. Klymchuk

Abstract. Two classes of inhomogeneous electromagnetic shields are considered: shields based on regular shielding elements and materials with a chaotic arrangement of metal-containing particles. It is shown that an accurate prediction of the protective properties of composite materials with a chaotic arrangement of inclusions is impossible due to the empirical nature of the known Lorentz, Maxwell-Garnet, and Odelevsky relations. The development of building and finishing materials with a regular arrangement of protective material in construction products is proposed. The main problem of imparting protective properties to building and finishing materials is the reduction of the main parameters of these materials (elastic modules) due to the addition of a shielding substance. To avoid this effect, it is proposed to use technological cavities in building materials. For this purpose, it is possible to use soot from thermal power plants and iron ore concentrate. Theoretical considerations indicate that the addition of a screening additive of large dispersion to the initial mixture of foam and aerated concrete during foaming promotes the distribution of additive particles on the inner surfaces of the cavities. This significantly increases the protective properties of materials without reducing the elastic modulus of products. A calculation apparatus for the efficiency of shielding electromagnetic radiation by mesh metal structures is presented. To make the structure broadband, it is proposed to use two parallel structures tuned to the maximum and minimum wavelengths of radiation, which are determined experimentally. It is shown that the use of the solution of the wave equation for the Hertz vector allows obtaining simple and convenient relations for determining the minimum frequencies of radiation penetrating through the holes (cutoff frequencies). This made it possible to determine such parameters for rectangular and circular openings. The relations for calculating the effectiveness of protection due to the absorption of electromagnetic energy and the reflection of electromagnetic waves were obtained. The possibility of taking into account the area and number of holes per unit area of the screen is shown. The calculations performed concern purely geometric characteristics of protective structures with holes. A promising area of research is to determine the protective properties of heterogeneous electromagnetic shields, taking into account the electrophysical and magnetic properties of the materials from which the structures are made.

Keywords: electromagnetic shield, regular structure, building materials, cutoff frequency, Hertz vector.

В. А. Глива¹, Г. Ю. Краснянський¹, М. О. Довгановський¹, Т. М. Краснянський²

¹Київський національний університет будівництва і архітектури, Київ, Україна

²Український католицький університет, Львів, Україна

РОЗРОБЛЕННЯ І ДОСЛІДЖЕННЯ ЗАХИСНИХ ВЛАСТИВОСТЕЙ МАТЕРІАЛІВ ДЛЯ БЛОКУВАННЯ ЕЛЕКТРОМАГНІТНИХ ВИПРОМІНЮВАНЬ ІНФРАЧЕРВОНОГО ДІАПАЗОНУ

Анотація. Розроблено матеріали для блокування електромагнітних випромінювань інфрачервоного діапазону. Застосування нітриду бору, оксиду хрому, магнетиту, які у великих кількостях виробляються промисловістю, дозволило отримати матеріали, придатні для широкого застосування. Основою матеріалу є серійний синтетичний текстильний матеріал. Носієм наповнювача є синтетичний лак. Встановлено, що за мінімального вмісту нітриду бору (25 % за масою у одному шарі), оксиду хрому (4,5 %), магнетиту (30 %) зниження рівнів інфрачервоного випромінювання за колірними температурами складають 25–28 %. За максимальної кількості оксиду хрому (7,5 %) цей показник складає 40–45 %. Для зразків із вмістом магнетиту 40 % і мінімальним вмістом нітриду бору і оксиду хрому зниження рівнів інфрачервоних випромінювань складало 75–80 %. За максимального вмісту нітриду бору цей показник зростає до 85–90 %, що є прийнятним. Показано, що колірні температури, визначені за допомогою теплової камери відрізняються від фактичних температур зразка, що є особливостями конструкцій камер. Врахувати цей факт можливо розрахунковим методом. У відповідні функції закладено робочі діапазони інфрачервоних камер, коефіцієнт випромінювання об'єкта, потужність випромінювання об'єкта тощо. Це спрощує процес проектування захисних матеріалів. Розроблені матеріали, придатні для використання у якості теплового камуфляжу та вирішення задач теплоізоляції. Застосування у дослідних зразках магнетиту у вигляді окремого шару дозволило отримати коефіцієнт екранування електромагнітного випромінювання ультрависокої частоти порядку 6,0–6,5, що є прийнятним для вирішення задач електромагнітної сумісності електронного обладнання. Додавання до магнетиту графіту у кількостях 1,5–3,0 % підвищує коефіцієнти екранування до 8,2–8,5.

Ключові слова: інфрачервоне випромінювання, камуфляж, теплоізоляція, електромагнітна сумісність.

Вступ

Розроблення і впровадження матеріалів для блокування електромагнітних випромінювань інфрачервоного діапазону зберігає актуальність упродовж багатьох років. Це обумовлено необхідністю вирішення задач енергозбереження, що постійно вимагає створення матеріалів з високими коефіцієнтами відбиття та поглинання інфрачервоного (теплового) випромінювання. На сьогодні це особливо актуально для України, що пояснюється як обмеженістю енергоресурсів, так і наявністю великої кількості будівель і споруд, частково ушкоджених внаслідок бойових дій, що вимагає їх ремонту із забезпеченням прийнятної теплоізоляції. Крім того, в умовах війни потрібні матеріали, які ускладнюють виявлення об'єктів детекторами інфрачервоного випромінювання. Останні неперервно удосконалюються, що вимагає удосконалення відповідних матеріалів для зменшення теплових сигнатур об'єктів різної температури. Як в усьому світі, так і в Україні розроблено велику кількість матеріалів для блокування (поглинання та відбиття) електромагнітних випромінювань інфрачервоного діапазону. Але більшість з них є багатокомпонентними або багат шаровими і мають досить складні технології вироблення. Це підвищує вартість кінцевих виробів і обмежує їх масове застосування.

Тому постає завдання розроблення простого у виготовленні і прийняттого за властивостями матеріали для вирішення задач термоізоляції та теплокамouflажу.

Аналіз останніх досліджень і розробок. Розвиток досліджень і розробок щодо блокування інфра-

червоних випромінювань відбувається у двох напрямках.

Перший – матеріалів для забезпечення малопомітності нагрітих об'єктів при використанні тепловізорів. Технічні характеристики більшості тепловізорів відомі (довжини хвиль, які виявляються), а також характеристики інфрачервоних підсвіток. Наявність останніх вимагає особливих характеристик захисних покриттів – балансу відбиття та поглинання, характерного для оточуючих предметів.

Другий напрям – підвищення теплоізоляційних характеристик будівель і споруд у широкому частотному діапазоні, тобто для усього інфрачервоного випромінювання.

У дослідженні [1] показані результати розробки камуфляжного матеріалу для змінних фонових температур. Він є плівкою з напильням сплаву германію, сурми та телуру. За високих функціональних можливостей цей матеріал має високу вартість. При цьому найбільшій ефективності він досягає за температур 30–60 °С, що не завжди прийнятно. Його перевагою є можливість змінювати теплову сигнатуру за різних кутів відбиття випромінювання. Матеріал, описаний у [2], прозорий у видимій області електромагнітного спектра та непрозорий у інфрачервоній. Він складається з моношару полістиролових мікрокульок, нанесеного на потрібну поверхню. Технологія отримання моношару дуже складна і включає дворазове розчинення полімеру, оброблення ультразвуком тощо. Перенесення моношару на поверхню великої площі та складної гнучкої поверхні проблематичне. У будь-якому випадку у процесі розроблення захисних покриттів слід враховувати умови їх експлу-

атації. Наприклад, при застосуванні у зоні бойових дій матеріал повинен витримувати багаторазові деформації та високі механічні навантаження. У дослідженні [3] показана можливість створення покриття на основі поліуретану. Виготовлялися композиційні волокна, де матрицею є поліуретан, а наповнювачем – сурма та оксид олова. Ефективність такого матеріалу дуже висока (до 90 %) у широкому діапазоні температур – 20 °С–80 °С. Але собівартість також дуже висока, що унеможливило його застосування принаймні у цивільних цілях.

У відкритих джерелах представлено багато результатів виготовлення матеріалів для блокування інфрачервоних випромінювань з використанням нанотехнологій [4–6]. Вони ефективні і мають малі масогабаритні параметри, але у сучасних умовах України розглядати такі технологічні підходи недоцільно.

Таким чином, полягає задача розроблення і дослідження захисних властивостей матеріалу, виготовленого з простих і дешевих компонентів достатньої ефективності у інфрачервоній області електромагнітного спектра.

Мета роботи – розроблення і дослідження властивостей матеріалів для блокування інфрачервоних випромінювань.

Викладення основного матеріалу

Для виготовлення матеріалів, які блокують електромагнітні випромінювання інфрачервоного діапазону, у якості основи було обрано стандартну синтетичну тканину. У якості рідкої субстанції для нанесення на тканину екрануючих елементів використовувався стандартний лак ХС. Теоретичні міркування свідчать, що матеріалами, які поглинають інфрачервоне випромінювання можуть бути нітрид бору та оксид хрому. Ці матеріали широко використовуються у промисловості. Нітрид бору використовується для вогнетривких покриттів, а оксид хрому (тривалентний) у якості зеленого пігменту та абразиву. Додатково у зразки додавався магнетит для поглинання електромагнітних випромінювань надзвичайно високих та ультрависоких частот. У частину зразків додавався дрібнодисперсний графіт, що підвищувало питому провідність матеріалу. Ступені блокування інфрачервоного випромінювання контролювався за допомогою серійного тепловізора.

Молярні теплоємності компонентів склали:
нітрид бору – 298 Дж/моль·К,
магнетит – 230 Дж/моль·К,
оксид хрому – 119 Дж/моль·К.

Послідовність нанесення компонентів на підкладку: перший шар – нітрид бору, другий шар – магнетит, третій шар – оксид хрому. У частині зразків третім шаром був графіт. Усі компоненти розчинялися у лаку ХС. Вміст нітриду бору за масою складав 16–25 %, оксиду хрому – 4,5–7,5 %, магнетиту – 30–40 %, графіту – 1,5–3,0 %. За різних кількостей наповнювачів ступені блокування інфрачервоного випромінювання наведені на рис. 1, а, б, в. Ефективність екранування інфрачервоного випромінювання матеріалами з додаванням магнетиту та великої кількості нітриду бору наведено на рис. 1, г, д, е.

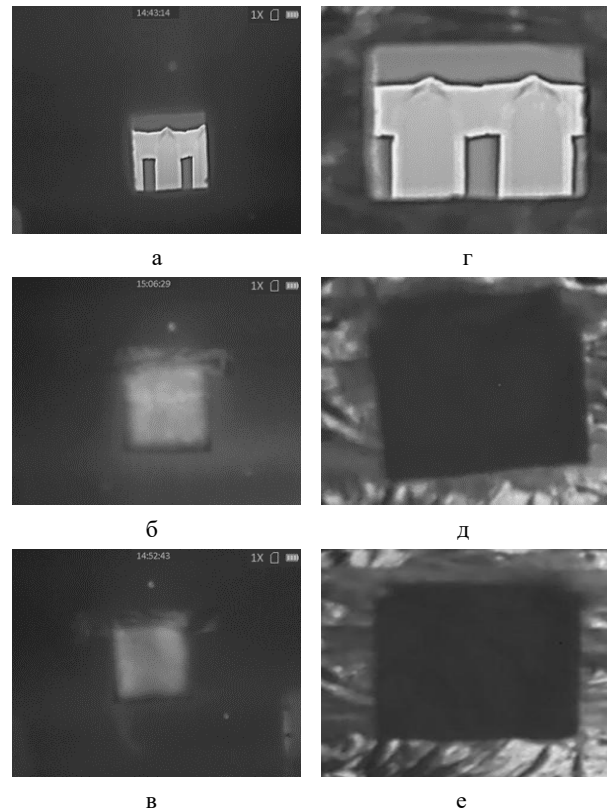


Рис. 1. Блокування інфрачервоного випромінювання:
а, г – зображення джерела випромінювання без захисного екрана;
б – мінімальні кількості наповнювачів;
в – максимальна кількість оксиду хрому.
д, е – зображення зразка з вмістом магнетиту з мінімальним (д) та максимальним (е) вмістом нітриду бору

Як видно з рис. 1 має місце зниження теплової сигнатури об'єкту, перекритого дослідними зразками.

Розглядаючи колірні температури, можна визначити, що матеріал (рис. 1, б) знижує інтенсивність інфрачервоного випромінювання на 25–28 %, на рис. 1, в – 40–45 %.

Найбільш прийнятні результати надають зразки, зображення яких наведено на рис. 1, г – е.

Зразок на рис. 1, д знижує інтенсивність інфрачервоного випромінювання на 75–80 %, а на рис. 1, е знижує інтенсивність інфрачервоного випромінювання 85–90 %. Такі результати можна вважати цілком задовільними. Але у реальних умовах експлуатації матеріалів ефективність може бути значно нижчою. Це пояснюється можливістю нагрівання самого захисного екрана внаслідок тривалого теплового впливу. Однією з можливостей уникнення такого ефекту є нанесення на основу у якості першого шару металовмісного матеріалу, наприклад, дрібнодисперсного алюмінієвого порошку. При цьому слід враховувати, що алюмінієвий порошок погано розчиняється у синтетичних лаках. Такий процес вимагає застосування швидкісного дисольвера.

Ефект різниці температури об'єкта, зафіксованої на інфрачервоному зображенні й фактичної температури об'єкта можна врахувати розрахунком:

$$T_k = CP_{об}(T) = C \times \int_{\lambda_1}^{\lambda_2} I_{об}(\lambda_1 T) d\lambda = C \int_{\lambda_1}^{\lambda_2} \varepsilon_{об}(\lambda) \cdot \frac{2\pi hc^2}{\lambda^5} \cdot \frac{1}{e^{hc/\lambda kT} - 1} \cdot d\lambda,$$

де T_k – температура об'єкта, визначена інфрачервоною камерою, C – питома константа інфрачервоної камери, $P_{об}$ – потужність випромінювання об'єкта, $I_{об}$ – спектральна яскравість об'єкта, $\varepsilon_{об}$ – коефіцієнт випромінювання об'єкта, (коефіцієнт «сірості» об'єкта), λ_1, λ_2 – спектральний діапазон інфрачервоної камери.

За допомогою таких розрахунків можливо порівняти зображення об'єкта з температурними зображеннями природних об'єктів.

Наприклад, коефіцієнти випромінювання ґрунту, трави, кущів відомі, і складають 0,91; 0,95; 0,98 і близькі до показників абсолютно чорного тіла (1,0).

Було визначено, що розроблені покриття мають коефіцієнти екранування електромагнітних полів ультрависокої частоти на рівні 6,0–6,5, що прийнятно для вирішення задач електромагнітної сумісності електронного обладнання.

Додавання графіту у кількостях до 3,0 % підвищує коефіцієнти екранування до 8,2–8,5.

Висновки

1. Розроблені матеріали для блокування електромагнітних випромінювань інфрачервоного діапазону та випромінювань надзвичайно високих та ультрависоких частот базуються на компонентах масового виробництва (нітрид бору, оксид хрому, магнетит) й прийнятні для широкого застосування для вирішення задач теплового камуфляжу та теплоізоляції.

2. Зниження рівнів інфрачервоного випромінювання за колірними температурами складають для зразків з мінімальними кількостями наповнювачів 25–28 %, з максимальними кількостями оксиду хрому – 40–45 %. Для зразків із вмістом магнетиту за мінімальної кількості нітриду бору та оксиду хрому цей показник складає 75–80 %, а з максимальним вмістом нітриду бору – 85–90 %.

3. Застосування у дослідних зразках магнетиту у вигляді окремого шару дозволило отримати коефіцієнт екранування електромагнітного випромінювання ультрависокої частоти порядку 6,0–6,5, що є прийнятним для вирішення задач електромагнітної сумісності електронного обладнання. Додавання до магнетиту графіту у кількостях 1,5–3,0 % підвищує коефіцієнти екранування до 8,2–8,5.

СПИСОК ЛІТЕРАТУРИ

1. 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>
2. 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>
3. Jeong, S.M., Ahn, J., Choi, Y.K. et al. Development of a wearable infrared shield based on a polyurethane–antimony tin oxide composite fiber. *NPG Asia Mater* 12, 32 (2020). <https://doi.org/10.1038/s41427-020-0213-z>
4. 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>
5. Luo, Hao & Li, Qiang & Du, Kaikai & Xu, Ziquan & Zhu, Huanzheng & Liu, Dongli & Cai, Lu & Ghosh, Pintu & Qiu, Min. (2019). An ultra-thin colored textile with simultaneous solar and passive heating abilities. *Nano Energy*. 65. 103998. [10.1016/j.nanoen.2019.103998](https://doi.org/10.1016/j.nanoen.2019.103998).
6. Fang, Fang & Kennedy, John & Carder, Damian & Futter, John & Rubanov, S. (2014). Investigations of near infrared reflective behaviour of TiO2 nanopowders synthesized by arc discharge. *Optical Materials*. 36. 1260-1265. [10.1016/j.optmat.2014.03.010](https://doi.org/10.1016/j.optmat.2014.03.010).

Received (Надійшла) 22.05.2024

Accepted for publication (Прийнята до друку) 10.07.2024

Development and research of protective properties of materials for blocking electromagnetic radiation in the infrared range

V. Glyva, G. Krasnianskyi, M. Dovhanovskiy, T. Krasnianskyi

Abstract. Materials for blocking electromagnetic radiation in the infrared range have been developed. The use of boron nitride, chromium oxide, magnetite, which are produced in large quantities by industry, has made it possible to obtain materials suitable for wide application. The basis of the material is a commercially available synthetic textile material. The filler is a synthetic varnish. It has been established that with a minimum content of boron nitride (25 % by weight in one layer), chromium oxide (4.5 %), and magnetite (30 %), the reduction in infrared radiation levels by color temperature is 25-28 %. With the maximum amount of chromium oxide (7.5 %), this figure is 40-45 %. For samples with a magnetite content of 40 % and a minimum content of boron nitride and chromium oxide, the reduction in infrared radiation levels was 75-80 %. At the maximum boron nitride content, this figure increases to 85-90 %, which is acceptable. It is shown that the color temperatures determined by the thermal chamber differ from the actual temperatures of the sample, which is a feature of the chamber design. It is possible to take this fact into account by the calculation method. The corresponding functions include the operating ranges of infrared cameras, the radiation coefficient of the object, the radiation power of the object, etc. This simplifies the process of designing protective materials. Materials have been developed that can be used as thermal camouflage and to solve thermal insulation problems. The use of magnetite as a separate layer in the prototypes made it possible to obtain a shielding coefficient of ultrahigh frequency electromagnetic radiation of the order of 6.0-6.5, which is acceptable for solving problems of electromagnetic compatibility of electronic equipment. The addition of graphite to magnetite in amounts of 1.5-3.0 % increases the shielding coefficients to 8.2-8.5.

Keywords: infrared radiation, camouflage, thermal insulation, electromagnetic compatibility.

Зв'язок, телекомунікації та радіотехніка

UDC 621.373.54

doi: 10.26906/SUNZ.2024.3.206

A. Serkov, N. Dzheniuk, O. Kasilov, G. Sokol, M. Tolkachov, D. Arutiunian

National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine

INTELLIGENT WIRELESS COMMUNICATION SYSTEM

Abstract. The subject of the study is models and methods for organizing a semantic communication system to increase the throughput of a wireless communication system. The goal is to develop recommendations for the implementation of multiple access with the distribution of information space according to the model. The development is based on artificial intelligence technology, which allows us to extract general and personalized semantic information from the multidimensional semantic space separately, allowing the communication system to move from the traditional transmission of information bits to the transmission of a model. The task is to ensure stable and reliable operation of a wireless network with an extended communication channel bandwidth. **Methods** used: methods of analytical modeling and time-position pulse coding. The following **results** were obtained. It is shown that in order to ensure the high quality of the wireless network, it is necessary to expand its bandwidth, which is limited by the physical resource of the radio frequency spectrum. It is proposed to overcome this contradiction by applying the technology of multiple access with distribution over the model. In this case, the information signal is emitted to the base station without a carrier frequency simultaneously in the entire frequency band, provided that the signal level is lower than the noise level. In this case, the method of positional-time coding is used, in which each information bit is encoded by hundreds of ultra-short chip pulses arriving in a certain sequence. In such wireless communication systems, the use of autocorrelation reception of modulated ultra-wideband signals is proposed. At the same time, the best reliability and noise immunity is provided by a wireless network using the time separation of the reference and information signals. **Conclusions.** Semantic communication systems with autocorrelation reception and separate transmission of reference and information signals provide an increase in the bandwidth of a wireless communication system, a high level of structural signal concealment, and reliable transmission of digital information, especially in conditions of interference.

Keywords: Semantic communication; Multiple access technology; Intelligent wireless network; Ultra-wideband signaling technology.

Introduction

The evolution of existing 5G wireless communication technologies and the development of 6G technologies are driven by the need to support new applications with various increased requirements for system performance and capacity. At the same time, modern wireless communications require enhanced security, high reliability and low signal latency. When organizing wireless communication, the physical limitation of the frequency spectrum requires the use of ultra-wideband (UWB) signaling technology, the essence of which is the transmission of low-power coded pulses in a very wide frequency band without a carrier frequency [1, 2]. To transmit information in NSF systems, pulse signals with very short pulse durations are used. Such a signal with a small space-time volume allows transmitting a large amount of information per unit of time, has a high level of noise immunity, and the ability to overcome the phenomena of multipath propagation of radio waves.

The ultra-short duration of the information sequence of the chips is due to the broadband of their spectrum in the radio frequency range (0...10 GHz), which is characterized by the lowest radio signal attenuation.

This makes it possible to use the radio frequency spectrum quite efficiently when operating in a congested frequency range. Thus, the use of NSS signals allows to

transmit information at a speed significantly exceeding the speed of traditional communication means with high noise immunity.

However, the broadband of information and reference signals impose significant restrictions on the efficiency of radio frequency spectrum use, especially when organizing multiple access, when system resources must be distributed among different users. Thus, the organization of multiple access requires an increase in the efficiency of radio frequency spectrum use.

The purpose of this work is to develop recommendations for the implementation of multiple access with the efficient use of the radio frequency spectrum.

Method of encoding an information signal

In wireless communication channels, the transmission medium is the physical path between the transmitter and the receiver. The most optimal range is between 1 and 10 GHz. This is because frequencies below 1 GHz are subject to significant interference from various industrial electronic devices. At the same time, at frequencies above 10 GHz, there is a large absorption of the useful signal by the transmission medium.

The frequency band of the radiation pulses is determined by the relation $B = WT_s$, where: W is the frequency bandwidth, and T_s is the transmission time of a bit of information. With the expansion of the signal base $B \gg 1$ base, it becomes possible to increase the

information transmission rate by reducing the duration of the transmitted signal.

In 1974, Massey J.L. [3] proposed to consider noise immune coding and modulation as a single unit. This ensures higher efficiency and energy gain from coding than the sequential application of noise immune coding and modulation. For example, to transmit information, one information bit is encoded by a sequence of many pulses (chips) per bit, which are shifted in time relative to the reference pulse-chip sequence. Depending on the value of the information bit "0" or "1", the sequence of information pulses-chips is shifted in time by 0.25 of the chip duration forward "zero" or backward "one" (Fig. 1.) [4].

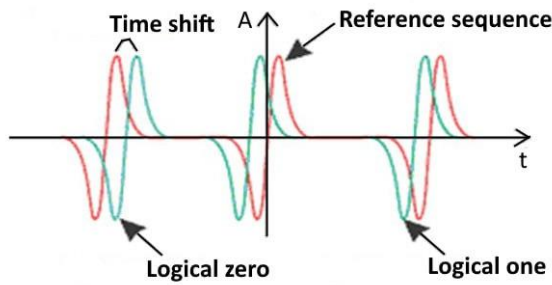


Fig. 1. The method of forming an ensemble of complex NSNS information signal

During encoding, the minimum chip arrival period is determined by the time for the chip energy to dissipate before the next chip in the information sequence arrives. On the other hand, an increase in the number of chips in the information bit sequence allows for an increase in the signal-to-noise ratio at the receiver input during decoding and correlation reception. The accumulation of a certain number of ultra-short pulses encoding each of the information bits in the receiver's correlator makes it possible to significantly increase the signal-to-noise ratio at the receiver's input, providing the ability to transmit information in a wide frequency range well below the white noise level.

Model of information signal ensemble formation

The formation of an ensemble of a complex NSN of an information signal radiated to free space is realized using the model, the scheme of which is shown in Fig. 2. The ultra-wideband signal is formed in the form of a normal random process $n(t)$ with a zero mean, uniform spectrum $S(f)$ and a frequency band Δf [5].

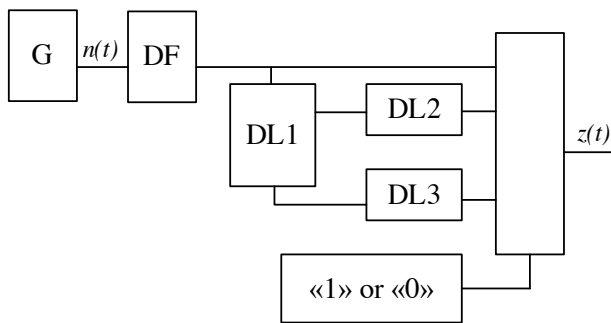


Fig. 2. Model of separate formation of reference and information signals

Fig. 2 shows: G – oscillator, which in self-oscillatory mode generates a sequence of ultrashort pulses- chips; DF – digital bandpass filter, which generates a signal $n(t)$; DL – delay lines, which are shifted by time T_1 or T_0 form information signals according to the flow of binary bits "one" or "zero" from the information source.

The oscillator (G) in self-oscillatory mode generates a sequence of ultra-short pulses-chips with a period of t_D which is fed to the input of the digital bandpass filter (DF), which generates the signal $n(t)$. From the output of the DF, the ultra-wideband signal $n(t)$ is fed to the modulator, where it is divided into information and reference signals. The three-position switch closes the transmitter output directly to the NSF signal generator (G) during the first half of the bit interval.

Thus, during the period of time $T_s/2$ the reference NSS signal is generated. In the middle of the bit interval, the switch is switched to one of two possible positions, depending on the flow of binary bits "one" or "zero" from the information source.

At the same time, the DL1 delay line provides a delay of the signal $n(t)$ by half a bit interval $T_s/2$ and delay lines DL2 and DL3 are used directly to form a stream of binary bits "one" or "zero". As a result, an information signal is formed, separated in time from the reference signal, which at one interval T_s has the following form:

$$y(t) = \begin{cases} x(t), & 0 \leq t \leq T/2; \\ x(t - T/2 - T_0), & T/2 \leq t \leq T; \\ x(t - T/2 - T_1), & T/2 \leq t \leq T. \end{cases}$$

Thus, the transmitter generates and emits a complex signal to free space, which includes an information pulse NSN signal in the form of an encoded pulse sequence and a separate reference pulse sequence of the synchronization signal, which doubles the overload of the radio frequency spectrum when organizing multiple access.

NSF signal generators emit ultrashort duration pulses with ultra-fast rise and fall. The discovery of the effect of voltage recovery and ultra-fast reverse breakdown in high-voltage transients has contributed to the creation of fundamentally new semiconductor devices capable of switching high power in short time intervals, including devices based on a delayed shock-ionization wave and drift devices with sharp recovery. Such generators allow generating ultrashort pulses of nano- and pico-second duration with power up to tens of MW and repetition rate up to tens of MHz, while controlling their temporal position with an accuracy of more than 10 ps. The maximum pulse voltage reaches tens of kilovolts, and pulse currents reach thousands of amperes [6].

As the repetition rate increases, the peak power decreases, but remains much higher than that obtained with any other semiconductor devices. This makes it possible to use the developed oscillators in UWB communication systems (Table 1), and the service life of these devices is extremely high.

Table 1 – Technical characteristics of NSS generators

Model	Growth time, ns	Pulsewidth, ns	Amplitude, V	Repetition rate, MHz
HFBG-1-0.5	< 0,7	< 2	500	2(3)
HFBG-1-2.5	< 0,7	< 2	2500	(2)
HFBG-1-5	< 0,7	< 2	5000	(0,5)
HFBG-5-0.05	< 0,14	< 0,3	50	10(20)
HFBG-7-0.15	< 0,3	< 0,7	150	10(20)
HFBG-7-0.5	< 0,3	< 0,7	500	2
HFBG-7-1	< 0,3	< 0,7	1000	1
HFBG-4-0.5	< 0,1	< 0,5	500	(0,2)
HFBG-4-1-10	< 0,1	< 0,25	1100	0,01
HFBG-4-1-50	< 0,1	< 0,25	1100	0,05

Generators of powerful ultrashort pulses are created using diode avalanche sharpeners as the final stage of a circuit based on drift diodes and transistors with sharp recovery. By connecting them in series, high switching voltages are achieved. [7]. A chain of series-connected drift diodes with a sharp recovery works as a single diode. Its voltage decay time is 50-200 ps, and the maximum allowable voltage is up to 15 kV. The advantages of such circuits are the high rate of voltage rise, which is combined with a high pulse repetition rate and their precise positioning in the time sequence, which undoubtedly makes them the basis for the implementation of the NSW technology in wireless communications.

The useful signal is extracted from the noise by correlating the received and reference signals. The correlator convolves the received signal with the reference signal to determine the time shifts of the received pulses relative to the reference. Thus, when receiving a one, the correlation function is equal to +1, and when receiving a zero, it takes the value -1. In all other cases, the correlation function is equal to 0.

At the same time, reliable detection of the information signal is carried out at any small signal-to-noise ratio.

Thus, a wireless network with the use of time separation of the reference and information signals has the best reliability and noise immunity. [8]. However, the constant transmission of the reference signal by users leads to inefficient use of the radio frequency spectrum, especially when organizing multiple access

Multiple access technology

Multiple access (MA) technology seeks to use shared information to eliminate bandwidth losses due to redundant transmissions. Maximizing the throughput of a wireless communication system and minimizing semantic errors during information recovery is based on artificial intelligence (AI) technology, which separately extracts general and personalized semantic information. [9, 10]. AI technology is used to create a knowledge base in both the receiver and the transmitter to ensure high efficiency of semantic information extraction methods and the use of prior knowledge and information about the communication channel.

The semantic communication system creates a new distribution resource, namely the model of information space, which became the basis of a new technology of multiple access. Traditional MD technology takes into account physical resources such as time (TDMA), frequency (FDMA), power and space (SDMA). Thus, building a semantic communication system at the physical level has become an important element in the transmission of semantic information.

The channel capacity, which combines traditional physical resources and the information space of the model [11], is as follows.

$$C = TB \sum_{k=0}^n \log_2 \left(I + \frac{P_k}{\sigma^2} H \Phi_k H^H \right),$$

where T, B, P, H, and Φ are the time domain resource, frequency resource, energy resource, space resource, and information space resource of the model, respectively, I is the identity matrix, and σ^2 is the noise variance.

At the same time, the new CBM technology offers multiple access with model-based distribution. Semantic features extracted from the source information using an AI model allow modeling personal characteristics of the user, which provide a knowledge base for differentiating multi-user semantic information.

In this way, CBM technology sequentially identifies, analyzes, and extracts general and personalized information contained in semantic information extracted from the model from multiple users. Then, the signals containing general information are overlaid and reused. This allows to obtain the CBM effect and increase the transmission channel capacity.

Algorithm of CBM technology implementation

The information in the semantic signals of a certain number of users consists of general and personalized information. In the case of simultaneous transmission of information from several users who initiated the uplink commands, the base station receives and compares it. When requesting a connection, the base station receives general information - the chip reference sequence in the form of a model - which is compared with the existing models in the base station database using AI. If the received model is not available in the database, it is added

to the database, and during subsequent signal transmissions from the user, the general information is not transmitted, but is removed from the knowledge base. Continuing the communication session, users send only the synchronization signal and personal semantic information to the base station. After receiving the synchronization signal and personalized information from the users, the base station recovers the primary semantic signals of each user by correlating the received and reference signals. And the reference chip sequence, which is a common information model, is removed from the base station's knowledge base, thus reducing the load on the wireless communication channel by half when organizing multiple access. Thus, the information model created with the help of AI allows modeling the personal characteristics of the user, which provides a knowledge base for differentiating multi-user semantic information.

In addition, the base station simultaneously distributes semantic signals from multiple data sources to multiple users. The signals that carry general information are combined and sent together. At the same time, signals carrying personalized information are transmitted to the user separately. In this way, each user receives his or her own information, which is recovered using a semantic decoder based on the combination of general and personalized information.

The useful signal is extracted from the noise by correlating the received and reference signals. The block diagram of the correlator is shown in Fig. 3.

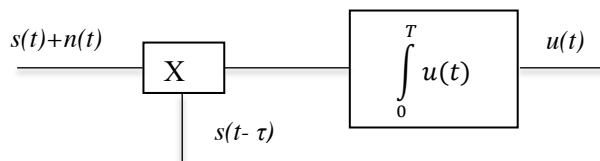


Fig. 3. Schematic diagram of the correlator of NSF signals

The correlator includes a multiplication device, to the other input of which a specially generated reference signal carrying general information is fed [11]. This signal completely repeats the useful signal in form, but lags behind it by a time τ , which allows to distinguish information. Subsequently, the signal that occurs at the output of the multiplication device is integrated, and the integration interval is equal to the duration of the signal $s(t)$.

The correlator performs convolution of the received signal with the reference signal. It is a detector that determines the time shifts of the received pulse signals relative to the reference signals. For example, when receiving a binary one, the correlation function is equal to +1, and when receiving a binary zero, it takes the value -1. In all other cases, the correlation function is zero. The accumulation of a certain number of ultra-short pulses encoding each of the information bits in the receiver's correlator makes it possible to significantly increase the signal-to-noise ratio, providing the ability to transmit information in a wide frequency range well below the white noise level. This makes it possible to meet the requirements for the level of electromagnetic compatibility of wireless mobile communication systems.

The use of ultra-wideband signals in the organization of multiple access requires the use of special antenna systems that allow them to be evenly matched to the antenna system in a wide frequency band [12]. For this purpose, an antenna that has the shape of an open slot is usually used, which determines the frequency band during reception/transmission. Moreover, the energy pattern of such an antenna is characterized by a narrow main beam and the practical absence of side lobes. However, the preliminary formation of a Gaussian monocycle entering the antenna system causes difficulties in matching in a wide frequency band. This manifests itself in the form of re-reflection of individual signal components, which distorts the shape of the Gaussian monocycle.

Therefore, the radiation pulse is formed directly in the antenna opening.

To do this, the information monopulse signal is split in half. A portion of the signal is sequentially inverted and delayed for a period of time equal to half the duration of the monopulse. Both mono-pulse signals are used to excite two TSA antennas located side by side on a single dielectric base. The electromagnetic fields of both monopulse signals interfere with the equivalent common aperture space of both antennas, creating a bipolar pulse electromagnetic field in it, while eliminating the time gap between the two parts of the radiated field, which is typical for a TSA antenna. Thus, such an antenna in an ultra-wideband communication system is capable of emitting and receiving both an ultra-short unipolar monopulse and a bipolar pulse information signal, which allows to significantly increase (by 3-10 times) the range of propagation of pulsed electromagnetic signals.

Design options for special antenna systems for receiving/transmitting ultra-wideband signals are shown in Fig. 4.

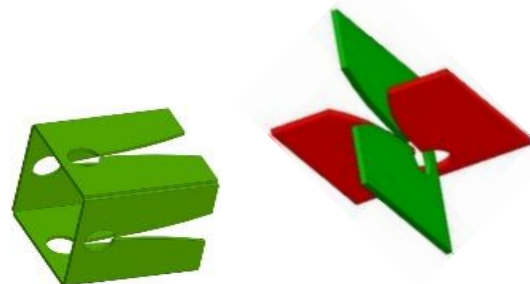


Fig. 4. Design options for special antenna systems

Conclusions

Recommendations for the implementation of multiple access with the distribution of information space according to the model are developed. The development is based on artificial intelligence technology, which allows us to extract general and personalized semantic information from the multidimensional semantic space separately, which allows the communication system to move from the traditional transmission of information bits to the transmission of a model.

It is shown that in order to ensure the high quality of the wireless network, it is necessary to expand its bandwidth, which is limited by the physical resource of

the radio frequency spectrum. It is proposed to overcome this contradiction by applying the technology of multiple access with distribution over the model.

In this case, the information signal is emitted to the base station without a carrier frequency simultaneously in the entire frequency band, provided that the signal level is lower than the noise level.

It is shown that semantic communication systems with autocorrelation reception and separate transmission of reference and information signals provide almost a doubling of the wireless communication system capacity, a high level of structural signal concealment, and reliable transmission of digital information, especially in the presence of interference.

REFERENCES

1. Ultra-Wideband Signals in Control Systems of Unmanned Aerial Vehicles / Serkov A., Kravets V., Yakovenko I., Churyumov G., Tokariev V., Namnan W. // (DESSERT'2019), pp. 25-28, - Leeds, United Kingdom, June 5-7, 2019, doi:10.1109/DESSERT.2019.8770039.
2. Security Improvement Techniques for mobile applications of Industrial Internet of Things / A.A. Serkov, B.A. Lazurenko, K.A. Trubchaninova, A.E. Horiushkina // IJCSNS International Journal of Computer Science and Network Security. VOL, 20 No. 5, p.p. 145-149. http://paper.ijcsns.org/07_book/202005/20200519.pdf.
3. J.L. Massey. Coding and Modulation in Digital Communications / Proc. Int. Zurich Seminar on Dig Comm. pp. E2(1)-E2(4), Zurich, Switzerland, 1974.
4. Serkov, O.A., Lazurenko, B.O., Pevnev, V.Y., Tkachenko, V.A., Kharchenko, V.S. Method of transmitting information by ultra-wideband pulse signals // Ukrainian Patent for Invention No. 123519 U IPC H04B 1/69, H04B 7/00, Published on April 14, 2021, Bulletin No. 15.
5. The Order of Formation of Information Signals in IIoT / Alla Jammine, Serkov Alexandr, Bogdan Lazurenko, Nait-Abdesselam Farid // IJCSNS International Journal of Computer Science and Network Security, VOL.23 No.3, pp. 139-143. http://paper.ijcsns.org/07_book/202303/20230314.pdf.
6. Kardo-Sysoev A.F., Brylevsky V.I. et al. Powerful Sources of Ultrawide Band Pulsed Coherent Signals. "EUROEM 2000" Euro Electromagnetics, Edinburgh, May 30-June 2, 2000.
7. Kardo-Sysoev A.F. et al. Ultra-Wide Band Solid State Pulsed Antenna Array. "EUROEM 2000" Euro Electromagnetics, Edinburgh, May 30-June 2, 2000.
8. Zhang P, Peng MG, Cui SG, et al., 2022b. Theory and techniques for "intellicise" wireless networks. *Front Inform Technol Electron Eng*, 23(1):1-4. <https://doi.org/10.1631/FITEE.2210000>
9. Zhang P, Xu XD, Dong C, et al., 2022c. Intellicise communication system: model-driven semantic communications. *J China Univ Posts Telecommun*, 29(1):2-12. <https://doi.org/10.19682/j.cnki.1005-8885.2022.2002>
10. Zhang P, Xu XD, Dong C, et al., 2023c. Model division multiple access for semantic communications / *Front Inform Technol Electron Eng* 2023 24(6):801-812. <https://doi.org/10.1631/FITEE.2300196>
11. Panchenko, S. V., Lazurenko, B. O., Serkov, O. A., Trubchaninova, K. A., Goryushkina, A. E. Method of receiving digital binary signals in noise conditions // Patent of Ukraine for utility model No. 145319 U IPC H04B 1/06, Published on November 25, 20, Bulletin No. 22.
12. Ultra-wideband antenna with flickering polarization and a method of its excitation / Panchenko S.V., Serkov O.A., Trubchaninova K.A., Kurtzev M.S., Lazurenko B.O. // Patent of Ukraine for invention No. 126475 U IPC H01Q 21/06, H01Q 13/08, Published on October 13, 22, Bulletin No. 41, application No. a 2019 08720 dated July 19, 2019.

Received (Надійшла) 07.03.2024

Accepted for publication (Прийнята до друку) 10.07.2024

Інтелектуальна безпровідна система зв'язку

А. Серков, Н. Дженюк, О. Касілов, Г. Сокол, М. Толкачов, Д. Арутюнян

Анотація. Предметом дослідження є моделі і методи організації системи семантичної комунікації для підвищення пропускної здатності безпровідної системи зв'язку. **Мета** – розробка рекомендацій щодо реалізації множинного доступу з розподіленням інформаційного простору за моделлю. В основу розробки покладено технологію штучного інтелекту, завдяки якій із багатовимірного семантичного простору вилучаємо окремо загальну та персоналізовану семантичну інформацію, що дозволяє системі зв'язку перейти від традиційної передачі бітів інформації до передачі моделі. **Задача** – забезпечення усталеної та надійної роботи безпровідної мережі з розширеною смугою пропускання каналу зв'язку. Використані **методи**: методи аналітичного моделювання та часового позиційно-імпульсного кодування. Отримані наступні **результати**. Показано, що для забезпечення високої якості роботи безпровідної мережі слід розширювати її пропускну спроможність, яка обмежена фізичним ресурсом радіочастотного спектру. Запропоновано долати це протиріччя шляхом застосування технології множинного доступу з розподіленням по моделі. При цьому випромінювання інформаційного сигналу до базової станції здійснюють без несучої частоти водночас у всій смузі частот за умов, що рівень сигналу нижче за рівень шуму. При цьому застосовано метод позиційно-часового кодування при якому кожен інформаційний біт кодують сотнями надкоротких імпульсів-чипів, що надходять за визначеною послідовністю. В таких системах безпровідного зв'язку запропоновано застосування автокореляційного прийому модульованих надширококузових сигналів. При цьому найкращу надійність і завадостійкість має безпровідна мережа із застосуванням часового розділення опорного та інформаційного сигналів. **Висновки.** Системи семантичної комунікації з автокореляційним прийомом та роздільною передачею опорного та інформаційного сигналів забезпечують розширення пропускної здатності безпровідної системи зв'язку, високий рівень структурної прихованості сигналу, а також надійну передачу цифрової інформації, особливо в умовах дії завад.

Keywords: семантична комунікація; технологія множинного доступу; інтелектуальна безпровідна мережа; технологія надширококузових сигналів.

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

МОБІЛЬНІ ЦИФРОВІ ТРОПОСФЕРНІ СТАНЦІЇ З КОМБІНУВАННЯМ ПРОСТОРОВО-РОЗНЕСЕНИХ СИГНАЛІВ

Анотація. В статті досліджується мобільні цифрові тропосферні станції та мобільні комбіновані цифрові радіотехнічні станції, які включають системи просторового рознесення сигналів. Це відомий прийом підвищення стійкості зв'язку та отримання необхідних якісних показників при передачі сигналів в багатопробієвих каналах. Методи прийому просторово-рознесених сигналів можна класифікувати як рознесений прийом із комбінуванням сигналів за надвисокою частотою, проміжною частотою, основною робочою смугою частот. Методи комбінування просторово-рознесених сигналів: автовибір; лінійне складання; оптимальне складання. В статті детально досліджено схему складання сигналів в лінійному тракту до демодулятора і після демодулятора. Результати дослідження показали, що кращі можливості у лінійного складання. При цьому відмічено, що важливу роль грають методи автофазування: автоматичне керування фазами гетеродинних НВЧ сигналів та автоматичне фазування складаних сигналів по проміжній частоті. При оптимальному складанні теоретично виходить найбільший вигравш щодо відношення сигнал/шум. Лінійне складання порівняно з оптимальним дає менший вигравш щодо сигнал/шум на виході пристрою комбінування всього на 1 дБ. Тому, для вузлових та високошвидкісних мобільних цифрових тропосферних станцій рекомендується застосування лінійного складання як методу комбінування рознесених сигналів. В роботі проаналізовано декілька структурних схем приймального тракту для просторово-рознесених сигналів з комбінуванням по проміжній частоті в мобільних цифрових тропосферних станціях. Відмітимо, що прийом із просторовим рознесенням антен продовжує залишатися основним засобом підвищення стійкості зв'язку на тропосферних лініях зв'язку під час передачі цифрових сигналів. В статті розглянуто методи прийому і комбінування просторово-рознесених сигналів, у яких немає перерв зв'язку через перемикання. Лінійне складання, знаходить широке застосування в сучасних мобільних цифрових тропосферних і комбінованих радіотехнічних системах, складовими яких є тропосферна компонента.

Ключові слова: просторове – рознесення сигналів; мобільні цифрові тропосферні станції; автовибір; лінійне складання; оптимальне складання; частково заповнений діалектиком хвилевод.

Вступ

Мобільні цифрові тропосферні станції (МЦТРС) і комбіновані мобільні радіотехнічні системи, такі як мобільна цифрова тропосферно-радіорелейна станція (МЦТрРРС) [1,2] та мобільна цифрова тропосферно-іоносферна станція (МЦТрІС) [3] включають у своєму складі системи просторового рознесення сигналів. Це відомий прийом підвищення стійкості зв'язку та отримання необхідних якісних показників при передачі сигналів в багатопробієвих каналах. Методи прийому просторово-рознесених сигналів можна класифікувати як рознесений прийом із комбінуванням сигналів за надвисокою частотою (НВЧ), проміжною частотою (ПЧ), основною робочою смугою частот. Перевагою методу комбінування по НВЧ є застосування лише одного приймача для перетворення НВЧ-ПЧ при n -кратному рознесенні. Недоліком є підвищення коефіцієнта шуму гілок рознесення з входів НВЧ за рахунок включення пристрою комбінування. У питанні реалізації малогабаритних хвилеводних елементів на НВЧ на допомогу приходить науково-технічний прийом реалізації трактів на частково заповнених діелектриком хвилеводах [4, 5]. Перевагою рознесеного прийому по ПЧ є відсутність демодуляції сигналу на МЦТРС і можливість ство-

рення більш ефективного за енергетичними параметрами пристрою комбінування порівняно з методом комбінування по НВЧ.

Аналіз літературних джерел. Техніка рознесеного прийому широко використовується у радіозв'язку. Найбільшого поширення набув просторово-рознесений прийом, який застосовується в системах МІМО [6-8], під час передачі сигналів OFDM [9], у системах міліметрового діапазону хвиль [10], у системах зв'язку різного призначення [11-18]. Також використовується частотно-рознесений прийом в системах спеціального призначення [19-22] і кутовий рознесений прийом на високошвидкісних лініях зв'язку [23-25]. Зазначимо, що просторово-поляризаційне та поляризаційне рознесення сигналів використовується як у системах радіозв'язку [26, 27], так і у радіолокації [28].

Метою роботи є варіанти побудови схем комбінування просторово-рознесених сигналів у МЦТРС, МЦТрРРС та МЦТрІС.

Основна частина

Методи комбінування просторово-рознесених сигналів: автовибір; лінійне складання; оптимальне складання. При складанні сигналів на прийомі розрізняють так само різні варіанти: складання до демодулятора і складання після демодулятора (рис. 1).

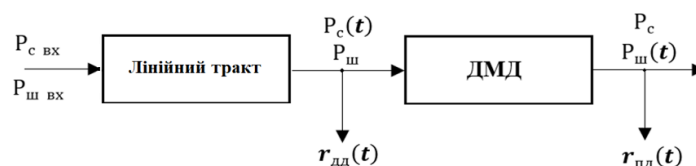


Рис. 1. Схема складання сигналів в лінійному тракту

До демодулятора ДМД складання здійснюється по ПЧ, після демодулятора ДМД – по груповому сигналу (низькій частоті). До демодулятора величина потужності шуму $P_{ш}$ постійна в часі, а змінюється відповідно до завмирання потужності сигналу $P_c(t)$. На виході демодулятора величина потужності сигналу P_c постійна, а змінюється потужність шуму $P_{ш}(t)$. Схема на рис. 1 відповідає кожній гілці рознесеного прийому, тому при швидких завмираннях сигналу на виході лінійного тракту до демодулятора і на виході демодулятора будуть змінюватися співвідношення сигнал/шум:

$$r_{дд}(t) = P_c(t)/P_{ш}, \quad (1)$$

$$r_{пд}(t) = P_c/P_{ш}(t). \quad (2)$$

В формулі (1) величина $r_{дд}(t)$ змінюється при зміні потужності $P_c(t)$, а в формулі (2) - величина $r_{пд}(t)$ змінюється при зміні потужності $P_{ш}(t)$. Ця різниця і визначає спосіб технічної реалізації системи складання – до демодулятора або після демодулятора.

Однак, незалежно від того, де проводитиметься складання суміші сигналу з шумом, наприклад, у k -гілці, співвідношення сигнал/шум $r_k(t)$ пропускається через суматор з ваговим коефіцієнтом $a_k(t)$ на його вихід, де формується сумарний сигнал/шум $r_{\Sigma}(t)$. Керуючі елементи суматора в кожній гілці змінюють за певним алгоритмом вагові коефіцієнти $a_1(t), \dots, a_k(t)$, чим досягається певний ефект від складання сигналів.

При лінійному складанні вагові коефіцієнти вибираються рівними, $a_k = const$. У цьому випадку маємо [29]:

$$r_{\Sigma}(t) \leq \sum_{k=1} r_k(t).$$

Лінійне складання виявляється тим більш ефективним, ніж менш відносна різниця між величинами $r_k(t)$ в гілках, коли завмирання неглибокі.

При автовиборі в кожен момент часу відбувається порівняння відношення сигнал/шум $r_k(t)$ всіх гілок і вибирається та, в якій в даний момент часу $r_k(t)$ найбільше, порівняно з іншими гілками. Керуючі сигнали відключають ці гілки від виходу суматора і вагові коефіцієнти стрибком приймають нульові значення:

$$a_k = \begin{cases} 1, & \text{при } k = j \ (r_j > r_k), \\ 0, & \text{при } k \neq j. \end{cases}$$

У цьому випадку маємо [29]:

$$r_{\Sigma}(t) = r_j(t) \leq \sum_{k=1} r_k(t).$$

Отже, автовибір ефективний при наявності істотних за величиною некорельованих завмирань в гілках.

При оптимальному (квазіоптимальному) складанні на вагові коефіцієнти безперервно впливають керуючі сигнали.

Максимальний ефект складання досягається за умови [29]:

$$r_{\Sigma}(t) = \sum_{k=1} r_k(t).$$

В цьому випадку, вагові коефіцієнти змінюються відповідно до правила:

$$a_k(t) \sim \sqrt{r_k(t)}.$$

В роботі [30] наводяться вирази залежності від застосованого способу складання сигналів. Використовуючи результати цієї роботи отримуємо, що енергетичний вигравш при чотирикратному складанні (дві антени на прийомі, кожна приймає сигнали горизонтальної та вертикальної поляризації) різними способами складають: автовибір – 3 дБ, лінійне складання – 5,2 дБ, оптимальне складання – 6 дБ.

Основною перевагою автовибору по ПЧ є простота побудови пристрою комбінування. Але застосування комутатора в якості пристрою комбінування при автовиборі має наступний істотний недолік. Через кінцевий час перехідних процесів при перемиканні виходу пристрою комбінування на інший приймач виникає короточасна перерва зв'язку. Тривалість цієї перерви становить ~ 1 мкс. Такі перерви зв'язку неприпустимі під час передачі високошвидкісних сигналів більше 8 Мбіт/с. При швидкостях передачі цифрової інформації 8,448 Мбіт/с і більше Мбіт/с у момент перемикання відбувається втрата кількох символів цифрової послідовності. У тропосферних системах Comtech та Raytheon досягається швидкість ~ 50 Мбіт/с та 100 Мбіт/с.

При оптимальному складанні теоретично виходить найбільший вигравш щодо сигнал/шум. Однак, лінійне складання порівняно з оптимальним дає менший вигравш щодо сигнал/шум на виході пристрою комбінування всього на 1 дБ. Тому, для вузлових та високошвидкісних МЦТРС рекомендується застосування лінійного складання як методу комбінування рознесених сигналів [31,32].

При лінійному складанні рознесених сигналів відношення сигнал/шум $r_{\Sigma}(t)$ на виході пристрою комбінування за умови, що сигнали складаються з довільними фазами, дорівнює:

$$r_{\Sigma}(t) = \frac{\sum_{i=1}^N \sum_{j=1}^N U_i(t) U_j(t) \cos[\psi_i(t) - \psi_j(t)]}{2 \sum_{k=1}^N \overline{\sigma_k^2}}, \quad (3)$$

де N – кількість гілок рознесення; U_{ij} – напруга, що огинає на вході i -ї та j -ї гілок; ψ – фази складаних сигналів; $\overline{\sigma_k^2}$ – середньоквадратичне значення напруги теплового шуму на вході k -ї гілки рознесення.

З формули (3) випливає, що для отримання максимального відношення сигнал/шум при складанні просторово-рознесених сигналів до демодуляції необхідно попередньо зробити їх фазування, тобто забезпечити рівність: $\psi_i(t) = \psi_j(t)$.

При застосуванні комбінування з лінійним складанням застосовуються два методи автофазування:

- автоматичне керування фазами гетеродинних НВЧ сигналів;
- автоматичне фазування складаних сигналів по ПЧ.

Приймачі з лінійним складанням та автоматичним фазуванням рознесених сигналів мають такі особливості:

- наявність загального гетеродину для всіх гілок рознесення;
- застосування загальної системи автоматичного регулювання пристрою для всіх гілок рознесення;
- включення у гілки рознесення елементів, що компенсують залишкову різницю часу запізнення, яка виникає через неоднакову довжину антенно-хвильових трактів гілок рознесення (у фідері однієї з антен запропоновано передбачати додатковий відрізок хвильоводу для вирівнювання електричних довжин гілок рознесення);
- застосування пристрою автофазування, що здійснює стеження за фазами складаних сигналів та підстроювання фаз сигналів у гілках рознесення.

На рис. 2 наведено структурну схему пристрою комбінування з керуванням фазами гетеродинних сигналів для випадку просторово-рознесеного прийому.

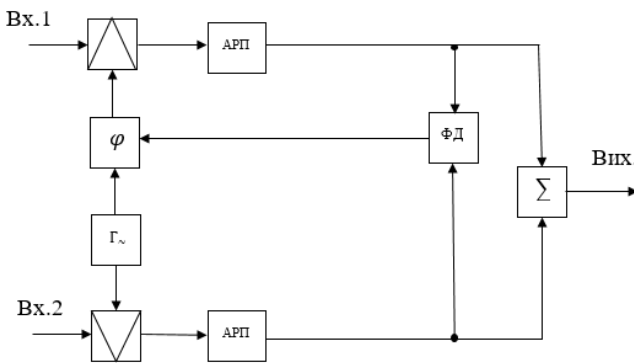


Рис. 2. Схема пристрою комбінування з керуванням фазами гетеродинних сигналів для просторово-рознесеного прийому з лінійним складанням

За допомогою фазового детектора ФД та електричного фазообертача φ здійснюється управління фазою гетеродина Г і забезпечується синфазне складання сигналів гілок рознесення.

Слід зазначити, що пристрій комбінування має бути побудовано так, щоб здійснювалося стеження тільки за фазами несучих частот сигналу, що передається, зміна яких на прольотах тропосферної лінії зв'язку відбувається порівняно повільно. Модуючий сигнал не повинен впливати на роботу пристрою комбінування. Тому в ланцюзі управління електричними фазообертачами, які входять до складу пристрою комбінування, вводяться інтегруючі ланцюжки.

Пристрої комбінування по ПЧ для приймачів просторово-рознесених сигналів з лінійним складанням показано на рис.3.

Принцип роботи цього пристрою комбінування полягає в управлінні фазою гетеродина Г, за допомогою якого відбувається перетворення частоти сигналу в одній із гілок рознесення. Потім проводиться зворотне перенесення спектра сигналу на частоту 70 МГц і сигнали обох гілок рознесення синфазно складаються.

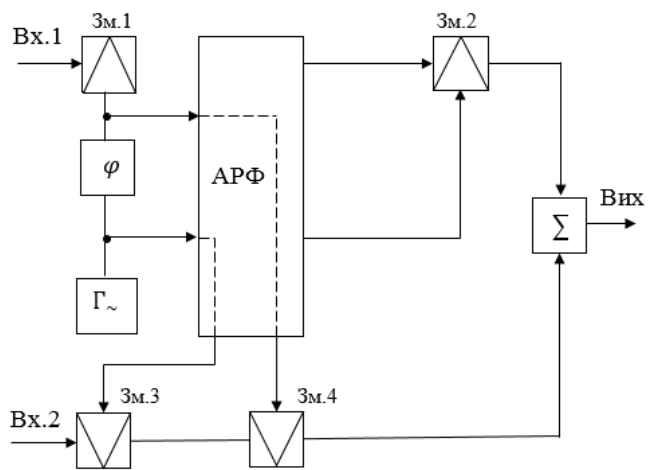


Рис. 3. Схема пристрою комбінування просторово-рознесених сигналів з лінійним складанням по ПЧ

Гетеродин Г (рис.3) працює на частоті 262 МГц. Сигнал з одного з його виходів через електричний фазообертач φ подається на змішувач Зм.1. На інший вхід Зм.1 змішувача надходить сигнал ПЧ (70МГц) з виходу ПЧ одного із приймачів. Після фільтрації на виході змішувача Зм.1 утворюється сигнал із частотою 192 МГц, який надходить на систему автоматичного регулювання фази АРФ і далі на змішувач Зм.2. На цей змішувач надходить інший сигнал від системи АРФ на частоті 262 МГц, що дозволяє на суматор Σ подати сигнал ПЧ 70 МГц. На інший вхід суматора Σ надходить також сигнал ПЧ 70 МГц, який пройшов «розфазування» в системі АРФ. За допомогою змішувачів Зм.3 і Зм.4 сигнал з частотою 262 МГц від гетеродина Г перетворюється на сигнал ПЧ (70 МГц) з фазою необхідної для синфазного складання на суматорі Σ.

Пристрій комбінування з фазуванням шляхом керування фазою НВЧ коливання гетеродина однієї з гілок просторового рознесення показано на рис. 4.

Структурну схему такого пристрою комбінування наведено на рис. 4.

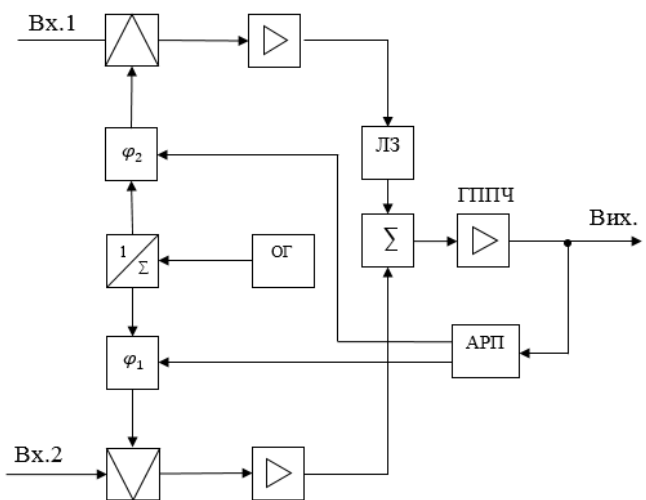


Рис. 4. Схема пристрою комбінування з фазуванням шляхом керування фазою НВЧ коливань гетеродина однієї з гілок просторового рознесення

В цьому випадку складання рознесених сигналів проводиться на виході попередніх підсилювачів ПЧ (ПППЧ), завдяки чому виконуються умови лінійного складання (рівність коефіцієнтів передачі гілок рознесення). Опорні коливання виробляються опорним генератором ОГ. Попереднє вирівнювання часу запізнення гілок рознесення здійснюється лінією затримки ЛЗ. Після складання на суматорі Σ сигнал ПЧ посилюється головним підсилювачем ПЧ ГУПЧ, загальним для обох приймачів.

За наявності «розфазування» складених сигналів на виході суматора Σ виникає додаткова АМ сигналу ПЧ. У цьому випадку АРП видає сигнал помилки на електричні фазообертачі φ_1 та φ_2 , які забезпечують фазування рознесених сигналів.

Висновки

Прийом із просторовим рознесенням антен продовжує залишатися основним засобом підви-

щення стійкості зв'язку на тропосферних лініях зв'язку під час передачі цифрових сигналів. У цій статті розглянуті методи комбінування просторово-рознесених сигналів, у яких немає перерв зв'язку через перемикання.

Відмітимо, що лінійне складання, знаходить широке застосування в сучасних мобільних цифрових тропосферних і комбінованих радіотехнічних системах, складовими яких є тропосферна компонента.

Крім того, в мобільній вузловій цифровій тропосферній станції [33], що працює за схемою «точка-багатоточка» закладено:

- просторово-рознесена передача [29, 34],
- автоматичне регулювання потужності передавача [35],
- кутовий рознесений прийом [36],
- використання фазованої антенної решітки [37].

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; опублік. 11.11.2019// Бюл.№21. <https://iprop-ua.com/?qi=120288>.
3. Патент №127524 Україна. С2. Мобільна цифрова тропосферно-іоносферна станція/ Почерняєв В.М., Сивкова Н.М., Повхліб В.С., Магомедова М.С. // заяв.07.07.2021; опублік. 20.09.2023. <https://iprop-ua.com/?qi=127524>.
4. Почерняєв В.М. Устройства на частично заполненных диэлектриком волноводах/ В.М. Почерняєв. – Киев: УКНИПСК, 2000. – 224с.
5. Почерняєв В.Н., Цибизов К.Н. Теория сложных волноводов, Науковий світ, 2003, с.223.
6. Weimin Kang. A Probabilistic Shaping Scheme for MIMO Systems with Signal Space Diversity. IEEE Wireless Communications and Networking Conference, USA. 10-13 April 2022.
7. Xiang Gao, Zhanji Wu, Sheng He, Yongtao Shi, Xiaoping Zhou. Enhanced MIMO cooperative communication based on signal space diversity. 9th Int. Conference on Communications and Networking in China, China. 14-16 August 2014.
8. Ghasemzadeh P., Hempel M., Banerjee S., Sharif H. A Spatial-Diversity MIMO Dataset for RF Signal Processing Research. IEEE Transactions on Instrumentation and Measurement. 2021. Vol. 70. pp. 1-10.
9. Aekkarat Lorphichian, Auttapon Pomsathit, Jintana Nakasuwan, Thanupong Srikalsin, Chawalit Benjangkaprasert. Performance analysis of space diversity for OFDM transmission. International Conference on Control, Automation and Systems. Korea. 14-17 October 2008.
10. Chang-Soon Choi, Yocho Shoji, Hiroyo Ogawa. Analysis of receiver space diversity gain for millimeter-wave self-heterodyne transmission techniques under two-path channel environments. IEEE Radio and Wireless Symposium, USA. January 2007.
11. Mehmet Akif Kurt, Ali Tugberk Dogukan, Ertugrul Basar. Spatial Modulation Using Signal Space Diversity. IEEE Communications Letters. 11 January 2023. pp.1020 – 1024.
12. Tingting Song, Ke Wang, Ampalavanapillai Nirmalathas, Christina Lim, Elaine Wong, Kamal Alameh. Demonstration of Optical Wireless Communications using Spatial Modulation with Signal Space Diversity. IEEE Photonics Conference (IPC). USA. 29 September 2019 - 03 October 2019.
13. Lagunas M. A., Perez Neira A. I., Amin M. G., Vidal J. Spatial processing for frequency diversity schemes. IEEE Transactions on Signal Processing, Feb. 2000. Vol. 48. No. 2. pp. 353-362. <https://doi.org/10.1109/78.823963>.
14. Kurt M. A., Dogukan A. T., Basar E. Spatial Modulation Using Signal Space Diversity. IEEE Communications Letters. March 2023. Vol. 27. No. 3. pp. 1020-1024. <https://doi.org/10.1109/LCOMM.2023.3236025>.
15. Giri, R.K., Patnaik, B. BER analysis and capacity evaluation of FSO system using hybrid subcarrier intensity modulation with receiver spatial diversity over log-normal and gamma-gamma channel model. Opt Quant Electron. 2018. No. 50. p.231. <https://doi.org/10.1007/s11082-018-1499-8>.
16. Kim S.-J., Han S.-K. Power Efficiency Analysis of Spatial Diversity Based Vertical FSO Links With Pointing Error in Multiple Beam Transmissions. IEEE Access, 2022. Vol. 10. pp. 25-31. <https://doi.org/10.1109/ACCESS.2022.3228528>.
17. Gorbunov S., Rashich A. Spatial receive diversity for SEFDM based system. 42nd International Conference on Telecommunications and Signal Processing, Budapest. 2019. pp. 265-268. <https://doi.org/10.1109/TSP.2019.8769085>.
18. Jha P.K., Kachare N., Kalyani K., Kumar D.S. Performance analysis of FSO using relays and spatial diversity under log-normal fading channel. 4th International Conference on Electrical Energy Systems, India. 2018. pp. 121-125. <https://doi.org/10.1109/ICEES.2018.8442405>.
19. Zhao X., Yu L., He F. Wide Range Swath SAR Imaging Method Based On Coherent Frequency Diverse Array Radar. 7th International Conference on Signal and Image Processing, Suzhou, China. 2022. pp. 576-580. <https://doi.org/10.1109/ICSIP55141.2022.9886965>.
20. Carvajal Mora H., Orozco N.G., Chiliguano P.T., De Almeida C. Enabling Signal Space Diversity for MU-MIMO/OFDMA Cellular Systems That Employ Frequency Diversity. IEEE Access, 2019. Vol. 7. pp. 111204-111221. <https://doi.org/10.1109/ACCESS.2019.2934638>.

21. Salucci M., Gelmini A., Poli L., Oliveri G., Massa A. Progressive compressive sensing for exploiting frequency-diversity in GPR imaging. *Journal of Electromagnetic Waves and Applications*. 2018. Vol. 9. No. 32. p. 1164-193, <https://doi.org/10.1080/09205071.2018.1425160>.
22. Imani M. F. et al. Review of Metasurface Antennas for Computational Microwave Imaging. *IEEE Transactions on Antennas and Propagation*, 2020. Vol. 68. No. 3. p. 1860-1875. <https://doi.org/10.1109/TAP.2020.2968795>.
23. Vipul Dixit, Atul Kumar. Performance Analysis of Indoor Visible Light Communication System with Angle Diversity Transmitter. *IEEE 4th Conference on Information & Communication Technology*. India. 03-05 December 2020.
24. Abrar S. Alhazmi, Sanaa H. Mohamed, Osama Z. Aletri, T. E. H. El-Gorashi, Jaafar M. H. Elmirghani. Angle Diversity Trasmmitter For High Speed Data Center Uplink Communications. *IEEE Conference on Standards for Communications and Networking*, Greece, 15-17 December 2021.
25. Roua Muwafaq Younus, Mahmood Ahmed Al Zubaidy, Safwan Hafeedh Younus Angle Diversity Receiver for Indoor Optical Wireless Communication Systems. *4th International Conference on Advanced Science and Engineering*, 21-22 September 2022.
26. Weitzen J., Wallace M. Analysis of diversity performance of space diversity and cross polarization for PCS base stations. *Ninth IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*. USA. 06 August 2002.
27. Ullah U., Mabrouk I. B., Koziel S., Al-Hasan M.. Implementation of Spatial/Polarization Diversity for Improved-Performance Circularized Polarized Multiple-Input-Multiple-Output Ultra-Wideband Antenna. *IEEE Access*. 2020. Vol. 8. p. 64112-64119. <https://doi.org/10.1109/ACCESS.2020.2984697>.
28. Mengqi Zhang, Lin Peng, Yan Liang, Yufei Wang, Jie Cui. Radar Polarization Diversity Technology for Low-altitude Targets. *CIE International Conference on Radar*, China. 15-19 December 2021.
29. Почерняев В.М., Сивкова Н.М., Повхліб В.С. Просторово - рознесена передача сигналів в цифрових тропосферних станціях // Наукові праці ОНАЗ ім. О.С. Попова, 2020. – №2. – С.92-99.
30. Гусятинский И.А. Дальняя тропосферная радиосвязь / И.А. Гусятинский, А.С. Немировский, А.В. Соколов, В.Н. Троицкий – М.: Связь, – 1968. – 248 с.
31. Патент України на винахід №122168 Мобільна високошвидкісна цифрова тропосферна станція/ Почерняев В.М., Повхліб В.С., Наритник Т.М.; заявл. 01.08.2018; опубл. 25.09.2020. <https://iprop-ua.com/?qi=122168>.
32. Патент України на винахід 126206 Мобільна вузлова цифрова тропосферна станція/ Почерняев В.М., Повхліб В.С., Наритник Т.М.; заявл. 10.03.2020; опубл. 31.08.2022. <https://iprop-ua.com/?qi=126206>.
33. Почерняев В.М., Сивкова Н.М., Магомедова М.С. Мобільна вузлова цифрова трансферна станція//Системи озброєння і військова техніка, №4(76) 2023, С. 6-15. <https://doi.org/10.30748/soivt.2023.76.01>
34. Сайко В.Г., Наконечний В.С., Баховський П.Ф., Сивкова Н.М. Алгоритм реалізації методу завадостійкого прийому сигналів, які випромінюються просторово рознесеними передавачами // Abstracts of I International Scientific and Practical Conference Stockholm, Sweden 5 7 April 2020. – С.247-253.
35. Почерняев В.М., Сивкова Н.М., Повхліб В.С. Система автоматичного регулювання потужності передавача НВЧ для комбінованих мобільних цифрових тропосферно-радіорелейних станцій // Вісник НТУУ «КПІ». Серія - Радіотехніка. Радіоапаратобудування, 2021. – №84 – С.40-47.
36. Почерняев В.М., Магомедова М.С., Сивкова Н.М., Шефер О.В. Кутовий рознесений прийом на лініях тропосферного та іоносферного зв'язку// Системи управління, навігації та зв'язку, №4 (74), 2023.- С.167-170. <https://doi.org/10.26906/SUNZ.2023.4.167>.
37. Почерняев В.М., Магомедова М.С., Сивкова Н.М. Плоска фазована антенна решітка для мобільних цифрових станцій зв'язку «точка-багатоточка» НВЧ діапазону// Системи управління, навігації та зв'язку, №2 (76), 2024. – С.187–190. <https://doi.org/10.26906/SUNZ.2024.2.187>.

Received (Надійшла) 12.05.2024

Accepted for publication (Прийнята до друку) 31.07.2024

Mobile digital troposcatter stations with combination of spatially diversity signals

V. Pochernyaev, M. Mahomedova, N. Syvkova¹

Abstract. The article examines mobile digital troposcatter stations and mobile combined digital radio engineering stations, which will include systems of spatially diversity signals. This is a well-known method of increasing the stability of communication and obtaining the necessary quality indicators when transmitting signals through multipath channels. The methods of receiving spatially diversity signals can be classified as diversity reception with combining signals by microwave frequency, intermediate frequency and the main working frequency band. Methods of combining spatially diversity signals: automatic selection; linear assembly; optimal assembly. The article examines in detail the signal composition scheme in the linear path before the demodulator and after the demodulator. The results of the research showed that the best opportunities are in linear assembly. At the same time, it was noted that autophasing methods play an important role: automatic control of the phases of heterodyne microwave signals and automatic phasing of composite signals at an intermediate frequency. With optimal composition, theoretically the greatest gain in terms of signal – to – noise ratio is obtained. However, the linear combination compared to the optimal gives a lower signal-to-noise ratio at the output of the combiner by only 1 dB. Therefore, for nodal and high-speed mobile digital troposcatter stations, the use of linear combination as a method of combining diversity signals is recommended. The article analyzes several structural schemes of the receiving path for spatially diversity signals with intermediate frequency combination in mobile digital troposcatter stations. Note that reception with spatially diversity of antennas continues to be the main means of improving communication stability on troposcatter communication lines during the transmission of digital signals. The article discusses the methods of receiving and combining spatially diversity signals in which there are no interruptions in communication due to switching. Linear combination are widely used in modern mobile digital troposcatter and combined radio engineering systems, the components of which is the troposcatter component.

Keywords: spatially diversity signals; mobile digital troposcatter stations; autoselection; linear assembly; optimal assembly; waveguides partially filled by dielectric.

АЛФАВІТНИЙ ПОКАЖЧИК

Agbeyangi A. O.	55	Заковоротний О. Ю.	29	Орлова Т. О.	29
Behlić A.	55	Заполовський М. Й.	34	Очеретенко С. В.	16
Fellner R.	55	Захаров Р.	149	Пашченко Р. Е.	5
Hashimov E.G.	55	Збежховська У. Р.	70	Педченко Н. М.	136
Petrovski A.	55	Зубко І. С.	121	Плехова Г. А.	16
Radovanović M.	55	Калімулін Т. М.	70	Поповкін М. М.	97
Zhou M.	168	Капітон А.	126	Почерняєв В. М.	211
Zied Chaari M.	55	Карачевцев Д. Д.	121	Пророк М. Ю.	25
Азімов К. Н.	130	Касілов О.	206	Прудіус М. А.	93
Андрусевич А. О.	20	Качанов П. О.	41	Пунченко А. В.	121
Андрусенко Ю. О.	196	Кібірев Д. О.	101	Радченко В. О.	146
Артюх Р. В.	158	Климчук А. В.	199	Роговий А. І.	158
Арутюнян Д.	206	Колісник Є. Б.	116	Сальніков С. С.	84
Ахієзер О. Б.	158	Комаричев А. В.	80	Сафанков Д. В.	185
Бараненко Р.	126	Кочина А. А.	130	Серков А.	206
Бараней Д. І.	80	Краснянський Г. Ю.	203	Сивкова Н. М.	211
Бешта В. С.	80	Краснянський Т. М.	203	Скиба О. В.	101
Білоконь А. С.	84	Кулініч Д. В.	175	Сокол Г.	126, 206
Бірук Я. І.	199	Кучук Н. Г.	146, 168	Соломоненко Ю. С.	70
Бовчалюк С. Я.	89	Лактіонов О. І.	25, 136	Стародубцев М. Г.	20
Бологова Н. М.	101	Лисиця Д. О.	196	Стефанюк Я. Г.	93
Борисов С. О.	84	Любий Є. В.	11	Стяглик Н. І.	106
Боряк Б. Р.	25	Ляшова А. О.	116	Тарасенко Д. А.	113
Бурда А. С.	93	Магомедова М. С.	211	Тарасенко М. В.	41
Вечірська А. Д.	106	Мазур Є. В.	106	Тищенко Д.	149
Волк М. О.	97	Макаренко Г. С.	20	Ткаченко О. С.	153
Гайдай Я. А.	89	Малєєва О. В.	141	Толкачов М.	206
Галицька О. О.	101	Малєєва Ю. А.	141	Фастовський Е. Г.	158
Глива В. А.	203	Мартовицький В. О.	121	Федорченко В. М.	84, 170
Горбач Т. В.	175	Марюшко М. В.	5	Філімончук Т. В.	80
Гороховатський В. О.	106	Мезенцев М. В.	34	Фомічов О. О.	93
Гриньов Д. В.	166	Мелешко Є. В.	153	Франчук Т.	149
Грідасов І. Ю.	70	Міхав В. В.	153	Фролов А. В.	20, 158
Даценко С. С.	113	Можаєв О.О.	116	Хижняк І. А.	70
Демська Н. П.	20	Москаленко В.	149	Худов Г. В.	70
Дженюк Н.	206	Набокін Д. А.	141	Шевчук Є. В.	170
Дзюбан О.	126	Неронов С. М.	16	Шматко О. В.	175, 185
Довгановський М. О.	203	Овсій Д.	25	Шостак А. В.	193
Дорофєєв К. І.	11	Олімпієва Ю. І.	46	Юзова І. Ю.	70
Дяченко В. О.	116	Оліфір М. В.	34	Янко А. С.	25, 136

Наукове видання

СИСТЕМИ УПРАВЛІННЯ, НАВІГАЦІЇ ТА ЗВ'ЯЗКУ

Збірник наукових праць

Випуск 3 (77)

Відповідальна за випуск *К. С. Нестеренко*Технічний редактор *Т. В. Уварова*Коректор *О. В. Морозова*Комп'ютерна верстка *Н. Г. Кучук, І. Ю. Петровська*Оформлення обкладинки *І. В. Львіна*Ідентифікатор медіа R30-04135 згідно з рішенням Національної ради України
з питань телебачення і радіомовлення від 25.04.2024 № 1416

Підписано до друку 10.09.2024. Формат 60×84/8. Ум.-друк. арк. 24,0. Тираж 130 прим. Зам.910-24

Адреса редакції: Україна, 36011, м. Полтава, Першотравневий проспект, 24, тел. (050) 302-20-71

Національний університет «Полтавська політехніка імені Юрія Кондратюка»

Віддруковано з готових оригінал-макетів у цифровій друкарні Impress

61002, м. Харків, вул. Пушкінська, 56, тел. + 38 (057) 714-52-11

e-mail: irina@impress.biz.ua