Національний технічний університет
«Дніпровська політехніка»

INFORMATION TECHNOLOGY: COMPUTER SCIENCE, SOFTWARE ENGINEERING AND CYBER SECURITY

Випуск 4

Видавничий дім
«Гельветика»
2024

РЕДАКЦІЙНА КОЛЕГІЯ:

ГОЛОВНИЙ РЕДАКТОР:

Удовик Ірина Михайлівна, кандидат технічних наук, доцент, декан факультету інформаційних технологій, Національний технічний університет «Дніпровська політехніка», Україна.

ЧЛЕНИ РЕДАКЦІЙНОЇ КОЛЕГІЇ:

Алексєєв Михайло Олександрович, доктор технічних наук, професор, завідувач кафедри програмного забезпечення комп'ютерних систем, Національний технічний університет «Дніпровська політехніка», Україна;

Бердник Михайло Геннадійович, доктор технічних наук, доцент, професор кафедри програмного забезпечення комп'ютерних систем, Національний технічний університет «Дніпровська політехніка», Україна;

Кабак Леонід Віталійович, кандидат технічних наук, доцент, доцент кафедри програмного забезпечення комп'ютерних систем, Національний технічний університет «Дніпровська політехніка», Україна;

Корнієнко Валерій Іванович, доктор технічних наук, професор, завідувач кафедри безпеки інформації та телекомунікацій, Національний технічний університет «Дніпровська політехніка», Україна;

Корченко Анна Олександрівна, доктор технічних наук, доцент, професор кафедри безпеки інформаційних технологій, Національний технічний університет «Дніпровська політехніка», Україна;

Лактіонов Іван Сергійович, доктор технічних наук, доцент, професор кафедри програмного забезпечення комп'ютерних систем, Національний технічний університет «Дніпровська політехніка», Україна;

Литвин Василь Володимирович, доктор технічних наук, професор, завідувач кафедри інформаційних систем та мереж, Національний університет «Львівська політехніка», Україна;

Любченко Віра Вікторівна, доктор технічних наук, професор, професор кафедри системного програмного забезпечення, Одеський національний політехнічний університет, Україна;

Маттіас Реч, Ph.D, професор, кафедра мехатроніки, Університет Ройтлінгену, Німеччина;

Молоканова Валентина Михайлівна, доктор технічних наук, професор, професор кафедри системного аналізу та управління, Національний технічний університет «Дніпровська політехніка», Україна;

Мороз Борис Іванович, доктор технічних наук, професор, професор кафедри програмного забезпечення комп'ютерних систем, Національний технічний університет «Дніпровська політехніка», Україна;

Мулеса Оксана Юріївна, доктор технічних наук, професор, професор кафедри інформаційних систем та мереж, Ужгородський національний університет, Україна;

Рак Тарас Євгенович, доктор технічних наук, доцент, професор кафедри інформаційних технологій, ПЗВО «ІТ СТЕП Університет», Україна;

Савенко Олег Станіславович, доктор технічних наук, професор, професор кафедри комп'ютерної інженерії та інформаційних систем, декан факультету інформаційних технологій, Хмельницький національний університет, Україна;

Семенов Сергій Геннадійович, доктор технічних наук, професор, професор кафедри кібербезпеки та інформаційних технологій, Харківський національний економічний університет імені Семена Кузнеця, Україна;

Сироткіна Олена Ігорівна, кандидат технічних наук, доцент, доцент школи комп'ютерних наук, Віндзорський університет, Канада;

Швачич Геннадій Григорович, доктор технічних наук, професор, професор кафедри програмного забезпечення комп'ютерних систем, Національний технічний університет «Дніпровська політехніка», Україна.

Журнал включений до Переліку наукових фахових видань України (категорія «Б») за спеціальностями 121 – Інженерія програмного забезпечення, 122 – Комп'ютерні науки, 123 – Комп'ютерна інженерія, 124 – Системний аналіз, 125 – Кібербезпека відповідно до наказів МОН України № 735 (додаток № 4) від 29.06.2021 р. та № 491 (додаток № 3) від 27.04.2023 р.

Реєстрація суб'єкта у сфері друкованих медіа: Рішення Національної ради України з питань телебачення і радіомовлення № 865 від 21.03.2024 року.
Мови видання: українська, англійська, німецька, польська, іспанська, французька, болгарська.
Офіційний сайт видання: www.journals.politehnica.dp.ua/index.php/it

Статті у виданні перевірені на наявність плагіату за допомогою програмного забезпечення StrikePlagiarism.com від польської компанії Plagiat.pl.

ISSN 2786-507X (Print)
ISSN 2786-5088 (Online)

© Національний технічний університет «Дніпровська політехніка», 2024

УДК 004.056.53
DOI https://doi.org/10.32782/IT/2024-4-1

Валентина АБЛАМСЬКА
викладач кафедри кібербезпеки, інформаційних технологій та економіки, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе, 210, м. Київ, Україна, 02121
ORCID:*0000-0003-0768-4102*

Наталія ДЯЧЕНКО
кандидат наук з державного управління,доцент кафедри кібербезпеки, інформаційних технологій та економіки, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе, 210, м. Київ, Україна, 02121
ORCID: *0000-0002-4306-7665*

Валентин ГАЛУНЬКО
доктор філософії з галузі права, доцент кафедри адміністративного права, інтелектуальної власності та цивільно-правових дисциплін, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе, 210, Київ, Україна, 02121,
ORCID: *0000-0002-8133-6766*

Бібліографічний опис статті: Абламська, В., Дяченко, Н., Галунько, В. (2024). Математичне моделювання вразливості інформаційної системи та прогнозування можливостей її захисту. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 3–8, doi: https://doi.org/10.32782/IT/2024-4-1

МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ВРАЗЛИВОСТІ ІНФОРМАЦІЙНОЇ СИСТЕМИ ТА ПРОГНОЗУВАННЯ МОЖЛИВОСТЕЙ ЇЇ ЗАХИСТУ

Проблематику статті присвячено дослідженню застосування математичних моделей для аналізу та вдосконалення роботи систем кібербезпеки та захисту інформації. Актуальність роботи зумовлена сучасними викликами цифрової епохи, коли інформаційні системи стають об'єктами численних новітніх атак (електронний скіммінг, фішинг, шкідливе програмне забезпечення (Malware), атаки типу DDoS та міжсайтовий скриптинг (XSS), спрямованих на викрадення, знищення, пошкодження або компрометацію даних. Особливу увагу приділено використанню диференціальних рівнянь та систем диференціальних рівнянь до моделювання роботи системи захисту, що дозволяє проаналізувати взаємодію атакуючих суб'єктів (зловмисників) та захисних систем, враховуючи нерівномірність кількості атак та динамічний характер загроз.

Наукова стаття обґрунтовує доцільність застосування диференціальних рівнянь для опису динамічних змін у системах кібербезпеки. Зокрема, запропоновано методику аналізу обсягу та інтенсивності атак за певний період часу (для короткотермінового прогнозу, який є залежним від початкових показників системи) для визначення продуктивності системи захисту даних. В даному контексті розглянуто застосовність диференціальних рівнянь та систем диференціальних рівнянь для оцінки ефективності поточних заходів кіберзахисту, що дозволяє прогнозувати при зміні потенційних інцидентів рівень забезпечення оперативної реакції на загрози.

Метою роботи *є огляд і оцінка місця диференціальних рівнянь та систем диференціальних рівнянь в моделюванні роботи системи захисту інформації.*

Методологія. *В рамках дослідження проведено моделювання реагування системи захисту інформації на несанкціонований вплив ззовні на об'єкти збереження інформації. Продемонстровано можливості та роль математичного моделювання у визначенні необхідного обсягу ресурсів, їх потенціалу для забезпечення стабільної роботи інформаційних систем. Розроблені підходи дозволяють адаптувати методи протидії до характеру та кількості загроз, включаючи застосування різноманітних програмних засобів, оптимізацію кількості інструментів чи методів для підвищення якісних характеристик систем.*

Наукова новизна *цього дослідження полягає у використанні математичного методу диференціальних рівнянь для оцінки ефективності системи захисту інформації та короткострокове прогнозування на основі отриманих результатів.*

Висновки. *Отримані результати мають практичне призначення (для діагностики спроможності системи кібербезпеки запобіганню несанкціонованого впливу) і можуть бути використані для подальшої розробки нових алгоритмів забезпечення кібербезпеки, оптимізації роботи захисних систем та підвищення їх ефективності в умовах постійно змінюваного кіберландшафту.*

Ключові слова: *кібербезпека, математична модель, диференціальні рівняння, системи диференціальних рівнянь, атака, протидія, прогноз.*

Valentyna ABLAMSKA
Lecturer at the Department of Cybersecurity, Information Technology and Economics, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121
ORCID: *0000-0003-0768-4102*

Natalia DIACHENKO
Candidate of Public Administration Sciences, Associate Professor at the Department of Cybersecurity, Information Technology and Economics, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121
ORCID: *0000-0002-4306-7665*
Scopus-Author ID: *57216565101*

Valentyn HALUNKO
PhD in law, Associate Professor at the Department of Administrative Law, Intellectual Property and Civil-Law Disciplines, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121, valentinvalentin0987@gmail.com
ORCID: *0000-0002-8133-6766*

To cite this article: Ablamska, V., Diachenko, N., Halunko, V. (2024). Matematychne modeliuvannia vrazlyvosti informatsiinoi systemy ta prohnozuvannia mozhlyvostei yii zakhystu [Mathematical modeling of the vulnerability of the information system and forecasting the possibilities of its protection]. Information Technology: Computer Science, Software Engineering and Cyber Security, 4, 3–8, doi: https://doi.org/10.32782/IT/2024-4-1

MATHEMATICAL MODELING OF THE VULNERABILITY OF THE INFORMATION SYSTEM AND FORECASTING OF ITS PROTECTION POSSIBILITIES

The issue of the article is devoted to the study of the application of mathematical models for the analysis and improvement of the work of cyber security and information protection systems. The relevance of the work is determined by the modern challenges of the digital age, when information systems become the objects of numerous modern attacks (electronic skimming, phishing, malicious software (Malware), DDoS attacks and cross-site scripting (XSS), aimed at theft, destruction, damage or compromise data. Special attention is paid to the use of differential equations and systems of differential equations to model the operation of the protection system, which allows analyzing the interaction of attacking subjects (criminals) and protective systems, taking into account the uneven number of attacks and the dynamic nature of threats.

The scientific article substantiates the feasibility of using differential equations to describe dynamic changes in cyber security systems. In particular, a technique for analyzing the volume and intensity of attacks for a certain period of time (for a short-term forecast that is dependent on the initial indicators of the system) is proposed to determine the performance of the data protection system.In this context, the applicability of differential equations and systems of differential equations for evaluating the effectiveness of current cyber protection measures is considered, which allows predicting the level of ensuring operational response to threats when potential incidents change.

The objective *of the work is to review and evaluate the place of differential equations and systems of differential equations in modeling the operation of the information protection system.*

Methodology. *As part of the study, a simulation of the response of the information protection system to unauthorized external influence on information storage facilities was carried out. The capabilities and role of mathematical modeling in determining the required amount of resources and their potential to ensure the stable operation of information systems are demonstrated. The developed approaches make it possible to adapt countermeasures to the nature and number of threats, including the use of various software tools, optimization of the number of tools or methods to improve the quality characteristics of systems.*

The novelty *of this study is to use the mathematical method of differential equations to evaluate the effectiveness of the information protection system and short-term forecasting based on the obtained results.*

The results. *The obtained results have a practical purpose (to diagnose the ability of the cyber security system to prevent unauthorized influence) and can be used for the further development of new algorithms for ensuring cyber security, optimizing the operation of protective systems and increasing their effectiveness in the conditions of a constantly changing cyber landscape.*

Key words: *cyber security, mathematical model, differential equations, systems of differential equations, attack, countermeasure, forecast.*

Актуальність проблеми. Кіберзагрози стають дедалі складнішими та різноманітнішими, і важливо постійно вдосконалювати системи безпеки та навчання персоналу для запобігання таким атакам. Їх існує велика кількість: ФІШИНГ (спроби шахраїв отримати конфіденційну інформацію про логіни, паролі, номери кредитних карток через обман або підроблені веб-сайти, електронні листи або спам-повідомлення); MALWARE (віруси, трояни, руткіти, шифрувальне ПЗ, які призначені для шкоди комп'ютерам та мережам); DDoS – Distributed Denial of Service (атака, при якій зловмисники намагаються перевантажити сервер або мережу величезною кількістю запитів, що призводить до відмови в обслуговуванні); RANSOMWARE (тип шкідливого програмного забезпечення, яке блокує доступ до файлів або системи і вимагає викупу за їх відновлення); INSIDER THREATS (зловживання, вчинені співробітниками, партнерами або іншими особами, які мають доступ до внутрішніх систем організації); SNIFFING (спеціальне програмне забезпечення для прослуховування мережевого трафіку з метою перехоплення чутливої інформації.) та ін.

Комплексний підхід до захисту від цих загроз включає використання шифрування, багатофакторної автентифікації, антивірусних програм, а також регулярне оновлення програмного забезпечення.

Математика широко застосовується для моделювання процесів, аналізу та прогнозування майбутніх змін у різних галузях, у тому числі й в кібербезпеці. Використання математичних методів у виявленні аномальної поведінки, загроз та вразливостей у системах дозволяє здійснювати своєчасні заходи для запобігання кібератакам.

З розвитком технологій зростає ймовірність несанкціонованого доступу до важливих даних, їх змін або крадіжки. Відповідно, швидкий розвиток вимагає від спеціалістів з кібербезпеки розробки нових, складніших алгоритмів та методів захисту. У цій статті розглядається роль математичних методів в оцінці рівня кіберзагроз і можливості покращення ефективності систем захисту.

Сучасні організації стикаються з тим, що певні дані швидко трансформуються в інформацію, яка зберігається на фізичних носіях або в хмарних сховищах. Якщо ця інформація є власністю організації, вона може бути уразливою для кібератак, крадіжки або знищення. Оскільки конкуренція на ринку постійно зростає, доступ до чужих даних чи інноваційних технологій стає важливим фактором, що привертає все більше уваги. Тому необхідно проводити систематичний моніторинг та збір даних про несанкціоновані дії в інформаційних системах організацій, аналізувати їх і прогнозувати можливі загрози. Завдяки цьому можна розробити відповідні заходи для зниження ризиків і покращення безпеки.

Аналіз останніх досліджень і публікацій. У сучасному цифровому світі кібербезпека стала критично важливою сферою досліджень, що охоплює широкий спектр тем, від аналізу загроз до розробки стратегій захисту. Існує декілька ключових напрямків досліджень у галузі кібербезпеки, які висвітлено у сучасних публікаціях.

Важливу роль відіграє забезпечення кібербезпеки, включаючи професійні та керовані послуги, а також продукти, що охоплюють функції ідентифікації, виявлення, захисту та реагування, підкреслює важливість адаптації кіберрішень до унікальних вимог організацій з урахуванням їхнього ландшафту ризиків і стратегій безпеки (IT Ukraine Association, IEEE ICT Conference, 2025).

В сучасних дослідженнях щодо захищеності даних приділяється особлива увага на необхідність підвищення обізнаності користувачів з основних принципів кібергігієни для зниження ризиків кіберзагроз, доцільність самоосвіти у галузі кібергігієни та кібербезпеки кожного користувача, що є все більш актуальним в сучасному технологічному середовищі (Федушко С., 2023).

Важливим є і аналіз сучасних загроз, такі як розширені стійкі загрози (APT), атаки програм-вимагачів, вразливості Інтернету речей (IoT) та соціальні інженерні експлойти. Тому також важливою є необхідність багаторівневого підходу до кібербезпеки, що включає надійні заходи безпеки, комплексне навчання співробітників та регулярні аудити безпеки (29th International Conference on Telecommunications (ICT), 2023).

Важливим напрямком дослідження проблем кіберзахисту є створення законодавчих актів та організаційних заходів, спрямованих на забезпечення колективної кібербезпеки, а також ознайомлення з досвідом передових практик у сфері кібербезпеки в країнах, що входять до десятки найуспішніших у цій сфері (United Nations Development Program).

В сучасних публікаціях наголошується важливість постійного ознайомлення з актуальною інформацію про останні кіберінциденти, вразливості та тенденції у сфері кібербезпеки,

постійних адаптацій систем захисту до нових загроз у кіберпросторі (Рада національної безпеки і оборони України, 2024; Український фонд безпеки, 2024).

Мета дослідження. Метою даного дослідження є: проведення аналізу сучасних математичних моделей, застосовуваних для оцінки ефективності систем кібербезпеки; дослідження застосовності диференційних рівнянь та систем диференційних рівнянь для моделювання динаміки кібератак та ефективності протидії загрозам у нерівномірних умовах атак; оцінка ролі диференціальних рівнянь у побудові динамічних моделей кіберзахисту, зокрема для прогнозування розвитку загроз і реакції системи; надання рекомендацій щодо вдосконалення системи протидії кібератакам шляхом впровадження комплексного підходу до управління ризиками.

Виклад основного матеріалу дослідження. У сучасному світі цифрові технології стали невід'ємною частиною соціальних, економічних і технічних систем. Однак зростання кількості кіберзагроз створює нові виклики для забезпечення інформаційної безпеки. Відомо, що традиційні методи аналізу не завжди дозволяють оперативно адаптувати системи захисту до умов нерівномірного розподілу атак.

У цьому дослідженні основна увага приділяється побудові моделі на застосуванні диференціальних рівнянь та систем диференціальних рівнянь, які описують зміну кількості атак та реакцію захисних механізмів.

Також представлено теоретичне обґрунтування обраного підходу, описано моделі та методи, використані для дослідження, а також наведено результати моделювання та їх інтерпретацію.

Застосування диференційних рівнянь чи систем диференційних рівнянь розглянемо в ситуації «атака-протидія», коли несанкціонований вплив виступає в якості атаки, а знешкодження атак – система протидії. Головне припущення, що лежить в основі моделі, полягає в тому, що система кібербезпеки активно обробляє атаки аж до певної межі.

Розглянемо можливі три різні випадки наступних взаємодій.

Випадок перший. При невеликому збільшенню кількості атак протидія повністю їх знищує. Тобто не дає інцидентам знищити певні ресурси чи їх пошкодити.

Випадок другий. При збільшенні кількості атак в залежності від зовнішніх умов і випадкових причин протидія може знищити ці атаки, а може і не впоратися з великою кількістю несанкціонованих дій, які створять нестабільну ситуацію.

Випадок третій. Відповідає катастрофі. Атаки діють дуже швидко, їх кількість занадто велика і ресурси зазнають втрат або взагалі знищуються (зникають). Захист не може знищити атаки, яких занадто багато, він не може використати для своєї роботи більше інструментів ніж він має в своєму арсеналі (чи є недостатня кількість фахівців), не в змозі відслідкувати і знищити всі інциденти.

Припустимо, що атаки і протидія характеризуються кількістю *P* і щільністю *E*, відповідно. У випадку, що атаки постійно існують, процес еволюції можна описати рівнянням

$$\frac{dP}{dt} = a - bP, \qquad (1)$$

де *a* – збільшення атак за одиницю часу,

b – коефіцієнт знищення відомих атак без перенапруження системи протидії. При початковій умові, за одиницю часу $t = 0$: $P(t = 0) = P_0$ рівняння (1) має вигляд

$$P_t = \frac{a}{b} + \left(P_0 - \frac{a}{b}\right)e^{-bt},$$

якщо з часом відомі атаки знищуються без перенапруження системи.

Можна припустити, що атаки і протидія знаходяться в постійній взаємодії, і процес взаємодії є замкнутим. Цей процес взаємодії можна описати наступним рівнянням

$$\frac{dP}{dt} = a - bP - f(,P), \qquad (2)$$

де функція $f(,P) \geq 0$ – описує відслідковування і знищення атаки.

Приймемо за g(ε) величину, коли система протидії не знаходить атак, $a\,h(E,P)$ – за функцію, коли атаки впливають на систему безпеки.

Тоді відслідковування і знищення атаки можна подати формулами

$$f(,P) = cP \;\; ma \;\; h(E,P) = dP,$$

де *c* і *d* – деякі коефіцієнти (кількість атак).

При відсутності атак поведінку протидії можна описати рівнянням

g()=$r\left(1 - \frac{}{K}\right)$, де *r* – постійний коефіцієнт, а *K* – відповідає найбільшому значенню, коли $\frac{d}{dt} = 0$. Тоді отримаємо систему рівнянь:

$$\frac{dP}{dt} = a - bP - cP,$$

$$\frac{d}{dt} = r\left(1 - \frac{}{K}\right) - dP \qquad (3)$$

Якщо спробуємо перейти до безрозмірних даних, отримаємо

$$P=\frac{by}{d},=\frac{bv}{c},\tau=bt,\alpha=\frac{ad}{b^2},\ u_0=\frac{r}{b},\ p=\frac{r}{cK},$$

Звідси, модель «атака-протидія» набуває виду

$$\frac{du}{dt}=\alpha-u-uv,\ \ \frac{dv}{dt}=v\left(u_0-u\right)-pv^2. \qquad (4)$$

В системі (4) параметр α – це потужність атак, u_0 – гранично допустима кількість атак (якщо $u>u_0$), $\frac{dv}{dt}<0$ – система протидії «зависає», або взагалі не може протидіяти), *p* – коефіцієнт концентрації інцидентів на систему.

Для успішного знищення популяції атак потрібно дослідити значення рівноваги в системі диференціальних рівнянь.

Рівновагу для успішного знищення атак $\frac{du}{dt}=0$, $\frac{dv}{dt}=0$, можна знайти з таких рівнянь:

$$a-u\left(1+t\right)=0,\ -v(u-u_0)-pv^2=0.$$

Тоді для *Випадку 1* положення рівноваги матиме вигляд*:*

$$A_1=\left(\alpha,0\right),$$

для *Випадку 2* положення рівноваги матиме вигляд*:*

$$A_2=\left(\frac{u_0+p+Q}{2},\frac{u_0-p-Q}{2p}\right),$$

для *Випадку 3* положення рівноваги матиме вигляд*:*

$$A_3=\left(\frac{u_0+p-Q}{2},\frac{u_0-p+Q}{2p}\right),$$

де $Q=\sqrt{\left(u_0+p\right)^2-4\alpha p}.$

Для *Випадку 2* і *Випадку 3* положення рівноваги існує, коли $\left[\left(u+p\right)^2-4\alpha p\right]>0$.

Наступна функція описує процес завантаженості системи атаками:

$$f\left(,P\right)=\frac{\left(\frac{cP}{d}\right)}{A+P}. \qquad (5)$$

Якщо врахувати цю функцію та безрозмірний вигляд, то система рівнянь (4) перепишеться

$$\frac{du}{dt}=\alpha-u-\frac{uv}{\lambda+u},\ \ \frac{dv}{dt}=v\left(u_0-u\right)-pv^2, \qquad (6)$$

де $\lambda=A\left(\frac{d}{b}\right)>0$ – це ступінь захисту системи протидії, чим більше його значення, тим менш вразлива система безпеки і захисту. Параметр λ – характеризує надійність системи безпеки. Тобто, чим більший параметр λ, тим більший захист може надати система протидії.

Висновки і перспективи подальших досліджень. Зроблено висновок, що застосування математичних моделей є ефективним підходом для вдосконалення систем протидії сучасним кіберзагрозам, зокрема в контексті забезпечення цілісності, конфіденційності та доступності інформаційних ресурсів.

Дослідження за допомогою диференціальних рівнянь аналізу у сфері кібербезпеки підкреслюють необхідність комплексного підходу до захисту інформаційних систем, що включає постійний моніторинг нових загроз та адаптація до них елементів ефективної стратегії кібербезпеки.

Отримані результати можуть слугувати основою для подальших досліджень у галузі математичного моделювання кібербезпеки, а також для вдосконалення методів і засобів захисту інформації.

У подальшій перспективі автори планують дослідити модель, що перебуває під дією атак у стані катастрофи (*Випадок 3* положення рівноваги).

ЛІТЕРАТУРА:

1. Огляд ринку кібербезпеки в Україні. IT Ukraine Association, IEEE ICT Conference, 2025. 43 с. URL: https://itukraine.org.ua/files/Ukraine-Cybersec-Market-Review.pdf.

2. Федушко С. Сучасні підходи до дослідження кібербезпеки та кібергігієни. Науковий вісник Хмельницького національного університету. Хмельницький: «ХНУ», 2023. С. 210–213. URL: https://journals.khnu.km.ua/vestnik/?p =18284

3. The New Frontier of Cybersecurity: Emerging Threats and Innovations. 29th International Conference on Telecommunications (ICT). 2023. URL: https://arxiv.org/abs/2311.02630

4. Analytical materials on cybersecurity. United Nations Development Program. URL: https://www.undp.org/ukraine/publications/analytical-materials-cybersecurity

5. Огляд подій у сфері кібербезпеки, січень 2024. Рада національної безпеки і оборони України. URL: https://www.rnbo.gov.ua /files/2024/NATIONAL_CYBER_SCC/Cyber%20digest/Cyber%20digest_Jan_2024_UA.pdf

6. Огляд подій у сфері кібербезпеки, квітень 2024. Український фонд безпеки. URL: https://ufss.com.ua/wp-content/uploads/2024/05/Cyber-digest_Apr_2024_UA.pdf

REFERENCES:

1. Ohlyad rynkhu kiberbezpeky v Ukrayini (2025). [Overview of the cyber security market in Ukraine]. IT Ukraine Association. IEEE ICT Conference. Vol. 43, Retrieved from: https://itukraine.org.ua/files/Ukraine-Cybersec-Market-Review.pdf [in Ukrainian].

2. Fedushko S. (2023). Suchasni pidhody do doslidzhennya kiberbezpeky ta kiberhiyieny [Modern approaches to cyber security and cyber hygiene research] Naukovyi visnyk Khmelnyts'koho natsional'noho universytetu [Scientific Bulletin of the Khmelnytskyi National University]. 210–213. Retrieved from: ttps://journals.khnu.km.ua/vestnik/?p=18284 [in Ukrainian].

3. The New Frontier of Cybersecurity: Emerging Threats and Innovations. (2023). 29th International Conference on Telecommunications (ICT). Retrieved from: https://arxiv.org/abs/2311.02630 [in USA].

4. Analytical materials on cybersecurity. United Nations Development Program. Retrieved from: https://www.undp.org/ukraine/publications/analytical-materials-cybersecurity

5. Ohlyad podiy u sferi kiberbezpeky, sichen 2024 [Overview of events in the field of cyber security, January 2024]. Rada natsional'noyi bezpeky i oborony Ukrayiny [National Security and Defense Council of Ukraine]. Retrieved from: https://www.rnbo.gov.uafiles/2024/NATIONAL_CYBER_SCC/Cyber%20digest/Cyber%20digest_Jan_2024_UA.pdf [in Ukrainian].

6. Ohlyad podiy u sferi kiberbezpeky, kviten 2024. [Overview of events in the field of cyber security, April 2024]. Rada natsional'noyi bezpeky i oborony Ukrayiny [National Security and Defense Council of Ukraine]. Retrieved from: https://ufss.com.ua/wp-content/uploads/2024/05/Cyber-digest_Apr_2024_UA.pdf [in Ukrainian].

УДК 004.85
DOI https://doi.org/10.32782/IT/2024-4-2

Катерина АНТІПОВА
доктор філософії, старший викладач кафедри інженерії програмного забезпечення, Чорноморський національний університет імені Петра Могили, вул. 68 Десантників, 10, м. Миколаїв, Україна, 54000
ORCID: *0000-0002-9012-5290*
Scopus Author ID: *57212609599*

Ігор КАНДИБА
доктор філософії, старший викладач кафедри інженерії програмного забезпечення, Чорноморський національний університет імені Петра Могили, вул. 68 Десантників, 10, м. Миколаїв, Україна, 54000
ORCID: *0000-0002-8589-4028*
Scopus Author ID: *57212577217*

Світлана БОРОВЛЬОВА
старший викладач кафедри інженерії програмного забезпечення, Чорноморський національний університет імені Петра Могили, вул. 68 Десантників, 10, м. Миколаїв, Україна, 54000
ORCID: *0000-0003-1994-0556*

Віктор РАЛЕНКО
викладач кафедри інженерії програмного забезпечення, Чорноморський національний університет імені Петра Могили, вул. 68 Десантників, 10, м. Миколаїв, Україна, 54000
ORCID: *0009-0009-4161-8468*

Бібліографічний опис статті: Антіпова, К., Кандиба, І., Боровльова, С., Раленко, В. (2024). Виявлення плагіату в тексті, згенерованого великими мовними моделями. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 9–15, doi: https://doi.org/10.32782/IT/2024-4-2

ВИЯВЛЕННЯ ПЛАГІАТУ В ТЕКСТІ, ЗГЕНЕРОВАНОГО ВЕЛИКИМИ МОВНИМИ МОДЕЛЯМИ

Для виявлення тексту, згенерованого за допомогою великих мовних моделей, існують різні підходи, на основі яких розроблені такі алгоритми, як DetectGPT, RADAR, Ghostbuster, GPT-Sentinel та інші для виявлення контенту, згенерованого штучним інтелектом. Хоча автоматизована перевірка може виявити певний плагіат, дослідження показують, що програмне забезпечення для пошуку текстів не тільки не знаходить весь плагіат, але й позначає оригінальний контент як плагіат, надаючи таким чином хибнопозитивні результати. Найсучасніші детектори згенерованого тексту демонструють значне погіршення продуктивності, коли стикаються з текстами, створеними людьми, які не є носіями англійської мови.

Метою дослідження є *підвищення точності і надійності виявлення тексту, створеного штучним інтелектом, особливо в освітньому середовищі, де плагіат та академічна недоброчесність стають усе більш актуальними через використання генеративних мовних моделей.*

Методологія дослідження *базується на загальнонаукових методах аналізу та синтезу, експериментальному тестуванні та кількісному аналізі ефективності мовної моделі, призначеної для перевірки тексту на наявність плагіату.*

Наукова новизна дослідження *полягає в адаптації сучасних методів виявлення плагіату для надійної класифікації текстів, створених штучним інтелектом, у контексті української мови. Для цього створено новий датасет на основі перефразованих текстових фрагментів, згенерованих ChatGPT, та налаштовано модель для класифікації тексту. Ефективність моделі оцінена за допомогою трьох різних оціночних метрик: показника F1, частоти хибно позитивних спрацювань і частоти хибно негативних спрацювань.*

Результати дослідження показують*, що налаштована модель ефективно виявляє відмінності між двома типами тексту, надають певне представлення про сильні та слабкі сторони моделі і демонструють її потенціал для застосування для практичних задач. Подальше дослідження полягає у зборі даних з іншим контекстом, щоб оцінити точність налаштованої моделі для різних задач обробки природної мови.*

Ключові слова: *ChatGPT, академічна недоброчесність, велика мовна модель, виявлення плагіату, генерація тексту, налаштування моделі.*

Kateryna ANTIPOVA
PhD, Senior Lecturer at the Department of Software Engineering, Petro Mohyla Black Sea National University, 10, 68 Desantnykiv Str., Mykolaiv, Ukraine, 54000
ORCID: *0000-0002-9012-5290*
Scopus Author ID: *57212609599*

Ihor KANDYBA
PhD, Senior Lecturer at the Department of Software Engineering, Petro Mohyla Black Sea National University, 10, 68 Desantnykiv Str., Mykolaiv, Ukraine, 54000
ORCID: *0000-0002-8589-4028*
Scopus Author ID: *57212577217*

Svitlana BOROVLOVA
Senior Lecturer at the Department of Software Engineering, Petro Mohyla Black Sea National University, 10, 68 Desantnykiv Str., Mykolaiv, Ukraine, 54000
ORCID: *0000-0003-1994-0556*

Viktor RALENKO
Lecturer at the Department of Software Engineering, Petro Mohyla Black Sea National University, 10, 68 Desantnykiv St., Mykolaiv, Ukraine, 54000
ORCID: *0009-0009-4161-8468*

To cite this article: Antipova, K., Kandyba, I., Borovlova, S., & Ralenko, V. (2024). Vyiavlennia plahiatu v teksti, zghenerovanoho velykymy movnymy modeliamy [Plagiarism detection in texts generated by large language models]. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 9–15, doi: https://doi.org/10.32782/IT/2024-4-2

PLAGIARISM DETECTION IN TEXTS GENERATED BY LARGE LANGUAGE MODELS

Various algorithms have been developed to detect text generated by large language models, and instruments such as DetectGPT, RADAR, Ghostbuster, GPT-Sentinel, etc. are based on these algorithms. While automated systems can detect some plagiarism, studies show that such software not only fails to find all plagiarism, but also marks original content as plagiarized, thus providing false positive results. State-of-the-art AI-text detectors demonstrate a significant performance degradation when faced with texts created by non-native English speakers.

The purpose of this study *is to improve the accuracy and reliability of detecting AI-generated text, especially in the educational environment, where plagiarism and academic misconduct are becoming increasingly relevant due to the use of LLMs.*

The research methodology *is based on general scientific methods of analysis and synthesis, experimental testing, and quantitative analysis of the effectiveness of a language model fine-tuned for plagiarism detection.*

The scientific originality of the study *lies in the adaptation of modern methods of plagiarism detection for reliable classification of AI-texts in the context of the Ukrainian language. To do this, a new dataset was created based on paraphrased text fragments generated by ChatGPT, and the mT5 model was fine-tuned on this dataset for a classification task. The model's performance is evaluated using three different evaluation metrics: F1 score, false positive rate, and false negative rate.*

The results of the study *show that the fine-tuned model effectively detects the differences between the two types of text. The results also provide some insight into the strengths and weaknesses of the model, and demonstrate its potential for application in practical tasks. Further research aims to collect data from different contexts to evaluate the accuracy of the fine-tuned model for different natural language processing tasks.*

Key words: *Academic misconduct, ChatGPT, fine-tuning, LLM, plagiarism detection, text generation.*

Актуальність проблеми. Впровадження ChatGPT, революційного інструменту на базі великої мовної моделі (ВММ), значно змінило різні галузі та сфери, включаючи й академічну. Можливості розвинутих ВММ вплинули на академічний світ різними способами. Наприклад, здобувачі вищої освіти використовують ChatGPT для виконання домашніх завдань і складання іспитів. Це викликало занепокоєння стосовно поточних систем оцінювання, які використовуються у закладах вищої освіти. Викладачі та університети намагаються виявити шахрайські дії здобувачів, і плагіат є однією з головних проблем. Раніше плагіат здебільшого полягав

у представленні рефератів та есе, які містили абзаци з інших джерел без посилань на них, але з появою ВММ здобувачі тепер можуть використовувати штучний інтелект (ШІ) для створення тексту і виконання своїх завдань. Акт використання здобувачами тексту, згенерованого ВММ, і ствердження, що це їхня власна робота, називається ШІ-плагіатом. Залежність здобувачів від інструментів генерації тексту призводить до втрати креативності та здатності до навчання.

Мовні моделі – це засновані на глибокому машинному навчанні моделі, призначені для різних завдань з обробки природної мови. Такі ВММ, як ChatGPT, можуть перефразовувати текст і створювати такий текст, що майже не відрізняється від тексту, написаного людиною. Ці моделі можуть впоратися як з простими завданнями, такими як створення есе на задану тему, так і зі складними, наприклад, написання наукової роботи зі складної проблеми. Поява таких генеративних інструментів ШІ та їхня здатність генерувати схожий на людський текст становить значну загрозу для академічної доброчесності.

Завдання виявлення того, чи був певний текст згенерований алгоритмами на основі ШІ, називається виявленням контенту, згенерованого штучним інтелектом (КЗШІ). Починаючи з моделі ELMO з 94 мільйонами навчальних параметрів у 2019 році до сучасної моделі GPT-4 з 1.76 трильйонами параметрів, еволюція ВММ та їхні можливості постійно зростають. З прогнозованим швидким розвитком високопродуктивних ВММ якість вихідних текстів зростає, що ускладнює їх виявлення.

Аналіз останніх досліджень і публікацій. Для виявлення тексту, згенерованого за допомогою ВММ, були розроблені різні підходи, такі як навчання класифікаторів, додавання водяних знаків та «безпрограшні» підходи (*zero-shot approaches*). На основі цих підходів розробляються такі алгоритми, як DetectGPT (Mitchell, 2023), RADAR (Hu, 2023), Ghostbuster (Verma, 2023), GPT-Sentinel (Chen, 2023) та інші для виявлення контенту, згенерованого ШІ. OpenAI представив свій інструмент виявлення КЗШІ через два місяці після випуску ChatGPT. Однак OpenAI стверджує, що детектор не є повністю надійним. Аналогічно, кілька інструментів і програм для виявлення КЗШІ, таких як CopyLeaks, Turnitin, GPTZero і Crossplag, були випущені для загального користування для виявлення контенту, згенерованого ШІ. Підходи до розпізнавання машинного тексту можна розділити на 4 категорії (Chen, 2023).

Традиційний статистичний підхід шляхом аналізу статистичних аномалій у текстовій вибірці. Деякі статистичні показники, такі як ентропія, використовуються як порогові значення для розпізнавання ШІ-тексту. Типовим прикладом є DetectGPT (Mitchell, 2023). DetectGPT перетворює вхідний текст за допомогою мовної моделі, яка заповнює маску, наприклад, T5. Потім виконується виявлення ШІ-тексту шляхом порівняння ймовірностей для тексту та його заповнених варіантів.

Підхід з неконтрольованим навчанням шляхом класифікації ВММ з нуля. Виявлення ШІ-тексту формулюється як задача бінарної класифікації, а класифікатор навчається для цільової мовної моделі (Rodriguez, 2022). Наприклад, OpenAI навчає свій класифікатор ШІ-тексту за моделлю на основі RoBERTa.

Підхід з контрольованим навчанням шляхом налаштування (*fine-tuning*) мовної моделі з додаванням або без додавання модуля класифікації.

Постфактумні методи нанесення водяних знаків, такі як методи на основі глибокого навчання (Dai, 2022), можуть бути застосовані до ВММ. У (Kirchenbauer, 2023) запропоновано м'яку схему водяного маркування, яка вбудовує водяний знак у кожне слово згенерованого речення, розділяючи словниковий запас на різні списки і диференційовано вибираючи наступний маркер. Однак, як показано в (Sadasivan, 2023), ефективність багатьох існуючих детекторів ШІ-контенту значно знижується через перефразування тексту.

Метою дослідження є підвищення точності і надійності виявлення тексту, створеного ШІ, особливо в освітньому середовищі, де плагіат та академічна недоброчесність стають усе більш актуальними через використання генеративних мовних моделей. Дослідження спрямоване на адаптацію сучасних методів виявлення плагіату для надійної класифікації текстів, створених ШІ, у контексті української мови. Для цього створено новий датасет на основі перефразованих текстових фрагментів, згенерованих ChatGPT, та налаштовано модель mT5 для класифікації тексту.

Виклад основного матеріалу. Академічна недоброчесність – це дії, які порушують оригінальність академічної роботи, такі як гострайтинг (*ghostwriting*), плагіат, фабрикація даних, генерація із використанням ШІ. Серед цих порушень найпоширенішою і найновішою формою неправомірної поведінки є генерація текстів. Дослідження в першу чергу зосереджується на

цьому аспекті. Найновіші інструменти генеративного ШІ, такі як ChatGPT, GPT-4 Vision, SORA від OpenAI, Gemini від Google та Perplexity від Perplexity AI, здатні генерувати різні типи контенту, включаючи текст, зображення, відео та код на різних мовах програмування.

Типологія плагіату може відрізнятися залежно від типу даних або рівня перетворення тексту. Фолтнек та ін. (Foltynek, 2019) представили різні типології, визначені в кількох наукових роботах, і запропонували нову типологію плагіату за рівнем перетворення: плагіат зі збереженням символів, зі збереженням синтаксису, зі збереженням семантики, зі збереженням ідеї та гострайтинг. Тип плагіату може також відрізнятися залежно від типу даних, наприклад, плагіат коду та плагіат тексту. Виявлення плагіату може бути зовнішнім або внутрішнім. Якщо плагіат виявляється лише за допомогою самого тексту, то це внутрішнє виявлення. Тоді якщо плагіат виявляється в порівнянні з іншим текстом, це називається зовнішнім виявленням. Виявлення плагіату також можна класифікувати відповідно до підходу: векторного, синтаксичного, семантичного, нечіткого, структурного та стиліметричного (Pudasaini, 2024).

Типовий алгоритм виявлення плагіату включає в себе інженерію ознак, класифікаційні моделі або метрики схожості текстів. Загальні ознаки, що використовуються в більшості алгоритмів виявлення плагіату, – це частота символів, середня довжина слова, середня довжина речення, частота N-грам у слові, частина мови, синоніми та гіперніми. Плагіат в основному оцінюється на основі текстової схожості з іншими еталонними текстами. Для обчислення такої схожості найчастіше використовують відстань Хаммінга, відстань Левенштейна та відстань найдовшої спільної підпослідовності, тоді як коефіцієнт Жаккара, косинусний коефіцієнт, відстань Манхеттена, евклідова відстань, коефіцієнт відповідності та коефіцієнт Соренсена були найпоширенішими метриками векторної схожості (Pudasaini, 2024).

Хоча автоматизована перевірка може виявити певний рівень плагіату, дослідження показують, що програмне забезпечення для пошуку текстів не тільки не знаходить весь плагіат, але й позначає оригінальний контент як плагіат, надаючи таким чином хибнопозитивні результати. Інструменти виявлення текстів, створених ШІ, дають збої, вони не є ані точними, ані надійними (усі мають точність нижче 80%). Загалом було виявлено, що вони діагностують документи, написані людиною, як створені штучним інтелектом (хибнопозитивні результати) і часто діагностують тексти, створені ШІ, як написані людиною (хибнонегативні результати). Інструменти виявлення здебільшого схильні класифікувати результати як написані людиною, а не виявляти вміст, створений ШІ. Загалом, приблизно 20% текстів, згенерованих штучним інтелектом, можуть бути помилково приписані людині.

Потреба в розробці нових алгоритмів, здатних точно відрізняти машинний текст від написаного людиною, стала як ніколи актуальною. Сучасний детектор контенту, згенерованого ВММ, повинен володіти такими ключовими характеристиками (Jawahar, 2020):

– ***точність*** означає, що модель повинна бути здатна розрізняти текст, згенерований ВММ, і текст, написаний людиною, досягаючи при цьому відповідного компромісу між точністю і швидкістю відкликання;

– ***ефективність*** використання даних означає, що детектор повинен оперувати якомога меншою кількістю прикладів з мовної моделі;

– ***узагальненість*** означає, що детектор повинен працювати послідовно, незалежно від будь-яких змін в архітектурі моделі, довжини запиту або набору навчальних даних;

– ***інтерпретованість*** означає, що детектор повинен надавати чіткі пояснення причин, що лежать в основі його рішень.

Нещодавнє дослідження (Hu, 2023) показало, що найсучасніші детектори ШІ-тексту демонструють значне погіршення продуктивності, коли стикаються з текстами, створеними людьми, які не є носіями англійської мови. Отже, незважаючи на багатообіцяючі результати, сучасні моделі мають певні обмеження. Одним з них є те, що більшість моделей навчаються лише на англомовному корпусі текстів. Як наслідок, ефективність розпізнавання текстів іншими мовами, зокрема слов'янськими, не є оптимальною. Для усунення цього обмеження може бути корисним налаштування моделей на текстах українською.

Щоб вирішити цю проблему, ми обрали підхід з контрольованим навчанням для розрізнення тексту, написаного людиною, і тексту, згенерованого ШІ. Спочатку ми зібрали дані з ChatGPT, а потім налаштували мовну модель на цьому наборі даних для класифікації. ChatGPT – це мовна модель, заснована на архітектурі GPT-4. Наскільки нам відомо, наразі не існує загальнодоступного набору даних, який би систематично збирав тексти, що генеруються ChatGPT. Тому ми поставили собі завдання створити власний набір даних згенерованих ШІ.

Набір даних GPTText складається з перефразованих фрагментів тексту, які були згенеровані мовною моделлю з використанням датасету UberText 2.0 (Chaplynskyi, 2023) як джерела. Датасет UberText був взятий із загальнодоступного ресурсу, який містить кілька наборів даних сучасною українською: новини, художня література, тексти на соціальну та судову тематику, статті з Вікіпедії. Набір даних містить 26 819 текстових фрагментів, кожен з яких відповідає фрагменту тексту з датасету UberText. Фрагменти, довжина яких перевищувала 2 000 слів, були відфільтровані. В процедурі перефразування був використаний API OpenAI на моделі gpt-4.

Ми налаштували модель mT5 (Xue, 2020) для задач класифікації, від послідовності до послідовності (*seq-to-seq*). Модель складається з двох основних компонентів: блоку кодера та блоку декодера, кожен з яких повторюється шість разів. Блок кодера обробляє вхідні токени, використовуючи механізм самоуваги (*self-attention*), після чого застосовується багатошаровий персептрон з прямим поширенням. Блок декодера аналогічно застосовує маскову багатоголову увагу (*multi-head attention*) та шари з прямим поширенням до закодованих представлень, що дозволяє йому генерувати виходи токен за токеном, передбачаючи ймовірність наступного слова. Під час навчання вхідна послідовність складається з текстової вибірки з GPTText, а вихідна представляє результат класифікації у вигляді «pos </s>» або «neg </s>», де «</s>» є маркером кінця послідовності. Остаточний результат – це ймовірнісний розподіл для наступного слова, що використовується для розрізнення послідовностей, згенерованих людиною або ChatGPT.

Ефективність моделі mT5 оцінена за допомогою трьох різних оціночних метрик: показника F1, частоти хибно позитивних спрацювань (FPR – *false positive rate*) і частоти хибно негативних спрацювань (FNR – *false negative rate*). Тут «позитивний» означає, що вхідний текст згенерований ChatGPT, тоді як «негативний» означає, що дані написані людиною. Враховуючи кількість істинно позитивних (*TP*), істинно негативних (*TN*), хибно позитивних (*FP*) та хибно негативних (*FN*) результатів, метрики обчислюються наступним чином:

$$\text{F1 Score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad \text{TPR} = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN} \quad TNR = \frac{TN}{TN + FP} \quad FNR = \frac{FN}{FN + TP}$$

За допомогою показника F1 вимірюють ступінь подібності між маркованою та передбачуваною відповіддю, отриманою від моделі. Показники F1, FPR і FNR для моделі mT5 розраховані на датасеті GPTText. Результати оцінювання наведені в таблиці 1. Всі дані подані у відсотках.

Таблиця 1

Показники точності моделей

Модель	Точність	TPR	TNR	FPR	FNR
mT5	95.98	97.3	96.22	3.78	2.7
GPT-2	41.5	22.38	95.87	4.13	77.62
GPT Zero	56.4	35.71	81.64	18.36	64.29

Результати дослідження показують, що налаштована модель ефективно виявляє відмінності між двома типами тексту, надає певне представлення про сильні та слабкі сторони моделі і демонструє її потенціал для застосування для практичних задач.

Висновки. В цій роботі за допомогою мовної моделі визначені відмінності між текстом, згенерованим за допомогою ChatGPT, і текстом, написаним людиною. Для цього був зібраний датасет, який складається з перефразованого контенту, згенерованого за допомогою ChatGPT. Після чого була налаштована модель mT5 для класифікації тексту. Ця модель досягла чудових результатів, з точністю понад 95% на тестовому наборі даних, оцінених за допомогою різних метрик. Такі результати дають важливу інформацію про ефективне використання мовних моделей для розпізнавання згенерованого тексту.

Однак, модель, навчена для задачі класифікації на такому наборі даних, як GPTText, може не дуже добре справлятися з іншими завданнями обробки природної мови, для яких ChatGPT широко використовується, наприклад, з відповідями на запитання. У майбутньому планується зібрати датасети з іншим контекстом, щоб оцінити точність налаштованої моделі mT5 для різних задач.

ЛІТЕРАТУРА:

1. DetectGPT: Zero-Shot Machine-Generated Text Detection using Probability Curvature / E. Mitchell et al. *Proceedings of Machine Learning Research :* Proceedings of the 40th International Conference on Machine Learning, 3 July 2023. Honolulu, Hawaii, USA, 2023. P. 24950-24962. DOI: 10.48550/arXiv.2301.11305

2. RADAR: Robust AI-Text Detection via Adversarial Learning / X. Hu et al. *Advances in Neural Information Processing Systems*, 10-16 December 2023. Vol. 36. New Orleans, USA, 2023. P. 15077–15095. DOI: 10.48550/arXiv.2307.03838

3. Ghostbuster: Detecting Text Ghostwritten by Large Language Models / V. Verma et al. *North American Chapter of the Association for Computational Linguistics*. USA, 2023. 16 p. (arXiv preprint). DOI: 10.48550/arXiv.2305.15047

4. GPT-Sentinel: Distinguishing Human and ChatGPT Generated Content / Y. Chen et al. 2023. 18 p. (arXiv preprint). DOI: 10.48550/arXiv.2305.07969

5. Cai S., Cui W. Evade ChatGPT Detectors via A Single Space. 2023. 12 p. (arXiv preprint). DOI: 10.48550/arXiv.2307.02599

6. Cross-domain detection of gpt-2-generated technical text / J. Rodriguez et al. *Human Language Technologies*: Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics, July 2022. Seattle, USA, 2022. P. 1213–1233. URL: https://aclanthology.org/2022.naacl-main.88 (дата звернення: 07.09.2024)

7. DeepHider: A Multi-module and Invisibility Watermarking Scheme for Language Model / L. Dai et al. 2022. 16 p. (arXiv preprint). DOI: 10.48550/arXiv.2208.04676

8. A Watermark for Large Language Models / J. Kirchenbauer et al. *Proceedings of Machine Learning Research :* Proceedings of the 40th International Conference on Machine Learning, 3 July 2023. Honolulu, Hawaii, USA, 2023. 26 p. DOI: 10.48550/arXiv.2301.10226

9. Can AI-Generated Text be Reliably Detected? / V.S. Sadasivan et al. 2023. 34 p. (arXiv preprint). DOI: 10.48550/arXiv.2303.11156

10. Large Language Models can be Guided to Evade AI-Generated Text Detection / N. Lu et al. 2023. 29 p. (arXiv preprint). DOI:10.48550/arXiv.2305.10847

11. Paraphrasing evades detectors of AI-generated text, but retrieval is an effective defense / K. Krishna et al. *Advances in Neural Information Processing Systems*, 10-16 December 2023. Vol. 36. New Orleans, USA, 2023. P. 27469-27500. DOI: 10.48550/arXiv.2303.13408

12. Authorship Obfuscation in Multilingual Machine-Generated Text Detection / D. Macko et al. 2024. 21 p. (arXiv preprint). DOI: 10.48550/arXiv.2401.07867

13. Foltynek T., Meuschke N., Gipp B. Academic plagiarism detection: a systematic literature review. *Association for Computing Machinery Computing Surveys (CSUR)*. Vol. 52(6). USA, 2019. p. 1–42. DOI: 10.1145/3345317

14. Survey on Plagiarism Detection in Large Language Models: The Impact of ChatGPT and Gemini on Academic Integrity / S. Pudasaini et al. 2024. 23 p. (arXiv preprint). URL: https://arxiv.org/pdf/2407.13105 (дата звернення: 10.10.2024)

15. Jawahar G., Abdul-Mageed M., Lakshmanan L. V. Automatic detection of machine generated text: A critical survey. *Proceedings of the 28th International Conference on Computational Linguistics,* December 2020. Barcelona, Spain, 2020. P. 2296–2309. URL: https://aclanthology.org/2020.coling-main.208 (дата звернення: 10.10.2024)

16. Chaplynskyi D. Introducing UberText 2.0: A Corpus of Modern Ukrainian at Scale. *Proceedings of the Second Ukrainian Natural Language Processing Workshop (UNLP)*, May 2023. Dubrovnik, Croatia, 2023. P. 1–10. URL: https://aclanthology.org/2023.unlp-1.1 (дата звернення: 21.09.2024)

17. mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer / L. Xue et al. *Human Language Technologies :* Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics, June 2021. USA, 2021. P. 483-498. DOI: 10.18653/v1/2021.naacl-main.41

18. GPT-2 Output Detector Demo. URL: https://openai-openai-detector.hf.space/ (дата звернення: 27.08.2024)

19. AI Detector – the Original AI Checker for ChatGPT & More. URL: https://gptzero.me/ (дата звернення: 27.08.2024)

REFERENCES:

1. Mitchell, E., Lee, Y., Khazatsky, A., Manning, C.D., & Finn, C. (2023). DetectGPT: Zero-Shot Machine-Generated Text Detection using Probability Curvature. *International Conference on Machine Learning*. https://doi.org/10.48550/arXiv.2301.11305

2. Hu, X., Chen, P., & Ho, T. (2023). RADAR: Robust AI-Text Detection via Adversarial Learning. https://doi.org/10.48550/arXiv.2307.03838

3. Verma, V. K., Fleisig, E., Tomlin, N., & Klein, D. (2023). Ghostbuster: Detecting Text Ghostwritten by Large Language Models. *North American Chapter of the Association for Computational Linguistics*. https://doi.org/10.48550/arXiv.2305.15047

4. Chen, Y., Kang, H., Zhai, V., Li, L., Singh, R., & Ramakrishnan, B. (2023). GPT-Sentinel: Distinguishing Human and ChatGPT Generated Content. https://doi.org/10.48550/arXiv.2305.07969

5. Cai, S., & Cui, W. (2023). Evade ChatGPT Detectors via A Single Space. https://doi.org/10.48550/arXiv.2307.02599

6. Rodriguez, J., Hay, T., Gros, D., Shamsi, Z., & Srinivasan, R. (2022). Cross-domain detection of gpt-2-generated technical text. In *NAACL*, pp. 1213–1233. https://aclanthology.org/2022.naacl-main.88

7. Dai, L., Mao, J., Fan, X., & Zhou, X. (2022). DeepHider: A Multi-module and Invisibility Watermarking Scheme for Language Model. https://doi.org/10.48550/arXiv.2208.04676

8. Kirchenbauer, J., Geiping, J., Wen, Y., Katz, J., Miers, I., & Goldstein, T. (2023). A Watermark for Large Language Models. *International Conference on Machine Learning*. https://doi.org/10.48550/arXiv.2301.10226

9. Sadasivan, V.S., Kumar, A., Balasubramanian, S., Wang, W., & Feizi, S. (2023). Can AI-Generated Text be Reliably Detected? https://doi.org/10.48550/arXiv.2303.11156

10. Lu, N., Liu, S., He, R., & Tang, K. (2023). Large Language Models can be Guided to Evade AI-Generated Text Detection. *Trans. Mach. Learn. Res*. https://doi.org/10.48550/arXiv.2305.10847

11. Krishna, K., Song, Y., Karpinska, M., Wieting, J., & Iyyer, M. (2023). Paraphrasing evades detectors of AI-generated text, but retrieval is an effective defense. https://doi.org/10.48550/arXiv.2303.13408

12. Macko, D., Móro, R., Uchendu, A., Srba, I., Lucas, J.S., Yamashita, M., Tripto, N.I., Lee, D., Simko, J., & Bieliková, M. (2024). Authorship Obfuscation in Multilingual Machine-Generated Text Detection. https://doi.org/10.48550/arXiv.2401.07867

13. Foltynek, T., Meuschke, N., & Gipp, B. (2019). Academic plagiarism detection: a systematic literature review. *ACM Computing Surveys (CSUR), 52(6),* pp. 1-42. https://doi.org/10.1145/3345317

14. Pudasaini, S., Miralles-Pechuán, L., Lillis, D., & Salvador, M. L. (2024). Survey on Plagiarism Detection in Large Language Models: The Impact of ChatGPT and Gemini on Academic Integrity. https://arxiv.org/pdf/2407.13105

15. Jawahar, G., Abdul-Mageed, M., & Lakshmanan, L. V. (2020). Automatic detection of machine generated text: A critical survey. https://aclanthology.org/2020.coling-main.208

16. Chaplynskyi, D. (2023). Introducing UberText 2.0: A Corpus of Modern Ukrainian at Scale. In *Proceedings of the Second Ukrainian Natural Language Processing Workshop (UNLP)*, pp. 1-10, Dubrovnik, Croatia. Association for Computational Linguistics. https://aclanthology.org/2023.unlp-1.1

17. Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A., & Raffel, C. (2020). mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer. *North American Chapter of the Association for Computational Linguistics*. https://doi.org/10.18653/v1/2021.naacl-main.41

18. GPT-2 Output Detector Demo. (n.d.). https://openai-openai-detector.hf.space/

19. AI Detector – the Original AI Checker for ChatGPT & More. (n.d.). https://gptzero.me/

УДК 004.89
DOI https://doi.org/10.32782/IT/2024-4-3

Михайло БЕРДНИК
доктор технічних наук, доцент, професор кафедри програмного забезпечення комп'ютерних систем, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, Дніпро, Україна, 49005
ORCID: *0000-0003-4894-8995*
Scopus Author ID: *57196469717*

Ігор СТАРОДУБСЬКИЙ
аспіранти кафедри програмного забезпечення комп'ютерних систем, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, Дніпро, Україна, 49005

Бібліографічний опис статті: Бердник, М., Стародубський, І. (2024). Використання методів машинного навчання для адаптації програмного забезпечення до різних обчислювальних платформ. *Technology: Computer Science, Software Engineering and Cyber Security,* 4, 16–28, doi: https://doi.org/10.32782/IT/2024-4-3

ВИКОРИСТАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ АДАПТАЦІЇ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДО РІЗНИХ ОБЧИСЛЮВАЛЬНИХ ПЛАТФОРМ

Актуальні виклики у сфері розробки програм, зумовлені різноманітністю апаратних архітектур, та запропоновані шляхи їх подолання за допомогою адаптивних компіляторів. В роботі розглядається створення адаптивних компіляторів, які використовують методи машинного навчання для автоматичної оптимізації та адаптації програмного забезпечення до різних обчислювальних платформ. В роботі також наводяться приклади застосування адаптивних компіляторів у мобільних та хмарних обчисленнях, де вони демонструють суттєве підвищення продуктивності та ефективності програмного забезпечення.

Метою роботи *є створення адаптивних компіляторів, використовуючи інтеграцію методів машинного навчання на всіх етапах компіляції – від аналізу вхідного коду до генерації машинного коду.*

Наукова новизна *полягає в інтеграції методів машинного навчання в процес компіляції, що дозволяє адаптивним компіляторам автоматично налаштовувати оптимізацію коду для різних обчислювальних платформ. Це рішення забезпечує динамічне налаштування компіляції, підвищує продуктивність програм та знижує потребу у ручному налаштуванні для нових архітектур.*

Висновки. *Проведений аналіз підкреслює ефективність адаптивних компіляторів, які завдяки машинному навчанню забезпечують автоматичну оптимізацію коду, підвищуючи продуктивність та ефективність програмного забезпечення на різних платформах. Запропонований підхід спрощує адаптацію програм, знижує витрати на розробку та дозволяє зосередитись на функціональності продуктів.*

Ключові слова: *адаптивні компілятори, машинне навчання, переносимість програмного забезпечення, обчислювальні платформи, генерація коду.*

Mykhailo BERDNYK
Doctor of Technical Sciences, Associate Professor, Professor at the Department of Computer Systems Software, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, MGB2006@ukr.net
ORCID: *0000-0003-4894-8995*
Scopus Author ID *: 57196469717*

Igor STARODUBSKYI
Postgraduate Student at the Department of Computer Systems Software, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005

To cite this article: Berdnyk, M., Starodubskyi, I. (2024). Vykorystannia metodiv mashynnoho navchannia dlia adaptatsii prohramnoho zabezpechennia do riznykh obchysliuvalnykh platform [The use of machine learning methods for adapting software to different computing platforms]. *Technology: Computer Science, Software Engineering and Cyber Security,* 4, 16–28, doi: https://doi.org/10.32782/IT/2024-4-3

THE USE OF MACHINE LEARNING METHODS FOR ADAPTING SOFTWARE TO DIFFERENT COMPUTING PLATFORMS

Current challenges in software development arise from the diversity of hardware architectures, and adaptive compilers offer solutions to address them. This paper examines the creation of adaptive compilers that utilize machine learning methods for automatic optimization and adaptation of software across various computing platforms. Examples of adaptive compiler applications in mobile and cloud computing are presented, where they demonstrate significant improvements in software performance and efficiency.

***The purpose** of this work is to create adaptive compilers through the integration of machine learning methods at all stages of compilation – from input code analysis to machine code generation.*

***Scientific Novelty** lies in the integration of machine learning methods into the compilation process, enabling adaptive compilers to automatically fine-tune code optimization for different computing platforms. This solution provides dynamic compilation adjustments, enhancing program performance and reducing the need for manual tuning for new architectures.*

***Conclusions**. The conducted analysis highlights the effectiveness of adaptive compilers that, through machine learning, enable automatic code optimization, thereby enhancing software performance and efficiency across different platforms. The proposed approach simplifies software adaptation, reduces development costs, and allows a focus on product functionality.*

***Key words:** adaptive compilers, machine learning, software portability, computing platforms, code generation.*

Актуальність проблеми. Сучасний розвиток обчислювальної техніки характеризується різноманітністю архітектур процесорів та обчислювальних платформ, які застосовуються в різних сферах – від мобільних пристроїв до суперкомп'ютерів і вбудованих систем. Така різноманітність створює значні виклики для розробників програмного забезпечення, оскільки вони змушені адаптувати свої програми до специфіки кожної платформи, витрачаючи на це багато часу і ресурсів. Традиційні компілятори не завжди можуть ефективно адаптувати код, що призводить до зниження продуктивності програм та обмеження їхньої функціональності на різних пристроях. Адаптивні компілятори, які використовують методи машинного навчання для автоматичної оптимізації та налаштування коду, стають надзвичайно актуальними. Вони дозволяють компіляторам «вчитися» на основі аналізу продуктивності програм на різних платформах та автоматично підбирати оптимальні стратегії компіляції для кожної архітектури. Це значно спрощує процес адаптації програмного забезпечення до нових платформ, знижує витрати на розробку та підвищує загальну продуктивність програмних продуктів. Такий підхід також дозволяє розробникам зосередитися на створенні функціональності, а не на вирішенні питань сумісності та оптимізації.

Зростаюча потреба в автоматизації процесів розробки програмного забезпечення, зменшення кількості помилок, пов'язаних з ручним налаштуванням компіляторів, та підвищення ефективності роботи програм роблять адаптивні компілятори важливим напрямком досліджень і розробок у галузі комп'ютерної науки та програмної інженерії.

Анализ останніх досліджень і публікацій. На сьогодні існують кілька компіляторів, які вже впроваджують елементи адаптивності (Z. Fisches, 2020). Наприклад, LLVM дозволяє створювати оптимізації для різних архітектур процесорів, але вимагає від розробників налаштовувати оптимізації вручну. GCC також має потужні інструменти для оптимізації, але обмежений у здатності автоматично адаптуватися до нових платформ без значних змін у конфігурації та коді.

Основним обмеженням поточних рішень є складність інтеграції машинного навчання в існуючі компіляційні процеси та недостатня гнучкість у адаптації до нових архітектур. Наприклад (S. Saxena, 2021; A. H. Ashouri, et al., 2018), більшість компіляторів не здатні автоматично навчатися та змінювати стратегії оптимізації на основі продуктивності програм у реальному часі. Це створює бар'єри для розробників, які прагнуть швидко адаптувати свої програми до нових платформ або покращити продуктивність на існуючих.

Таким чином, актуальним завданням є розробка адаптивних компіляторів які знижують витрати на адаптацію програмного забезпечення до нових платформ, підвищують продуктивностіь програм за рахунок оптимізованих компілятором інструкцій та автоматизація процесу налаштування компіляції, що дозволяє розробникам зосередитися на створенні функціональності замість вирішення проблем сумісності та оптимізації.

Мета статті: створення адаптивних компіляторів використовуючи інтеграцію методів машинного навчання на всіх етапах компіляції – від аналізу вхідного коду до генерації машинного коду.

Викладення основного матеріалу дослідження. Одним з ключових підходів є використання алгоритмів навчання з підкріпленням, де компілятор «вчиться» на основі результатів своїх дій, намагаючись мінімізувати час виконання коду або споживання ресурсів. Генетичні алгоритми дозволяють адаптивно налаштовувати параметри компіляції шляхом еволюційного підходу – відбір, схрещування та мутації параметрів оптимізації.

Також значну роль можуть відіграють нейронні мережі, особливо в задачах передбачення продуктивності коду на різних платформах. Нейронні мережі можуть аналізувати вхідний код і надавати рекомендації щодо оптимізацій, враховуючи специфіку архітектури процесора, наприклад, розмір кешу, конвеєризацію інструкцій чи наявність специфічних команд.

Методи машинного навчання дозволяють адаптивному компілятору створювати модель продуктивності для конкретної платформи та автоматично підлаштовувати процес компіляції відповідно до цієї моделі. Це значно підвищує ефективність адаптації програмного забезпечення до нових та специфічних апаратних платформ без необхідності ручної оптимізації. В роботі запропонований підхід до створення адаптивних компіляторів передбачає інтеграцію методів машинного навчання на всіх етапах компіляції –від аналізу вхідного коду до генерації машинного коду. Компілятор використовує алгоритми машинного навчання для вибору інструкцій, що найкраще відповідають характеристикам цільового процесора, а також для оптимізації розподілу регістрів та управління пам’яттю. Компілятор має модульну архітектуру, що дозволяє легко додавати нові алгоритми оптимізації або адаптувати існуючі під нові платформи (рис. 1).

На рис. 1 представлена структура адаптивного компілятора, включаючи основні компоненти, такі як аналізатор коду, модуль машинного навчання, селектор інструкцій, оптимізатор, планувальник команд та генератор машинного коду. Кожен компонент показаний як окремий блок, пов’язаний стрілками, що вказують на послідовність обробки коду. Діаграма також відображає інтеграцію алгоритмів машинного навчання у процес компіляції.

Наприклад, компілятор може містити окремі модулі для аналізу типів, оптимізації інструкцій, планування виконання та управління пам’яттю, кожен з яких адаптується за допомогою навчання на реальних даних виконання програм. Основною перевагою запропонованого підходу є його гнучкість і можливість автоматичного налаштування для нових архітектур без необхідності вносити ручні зміни до коду

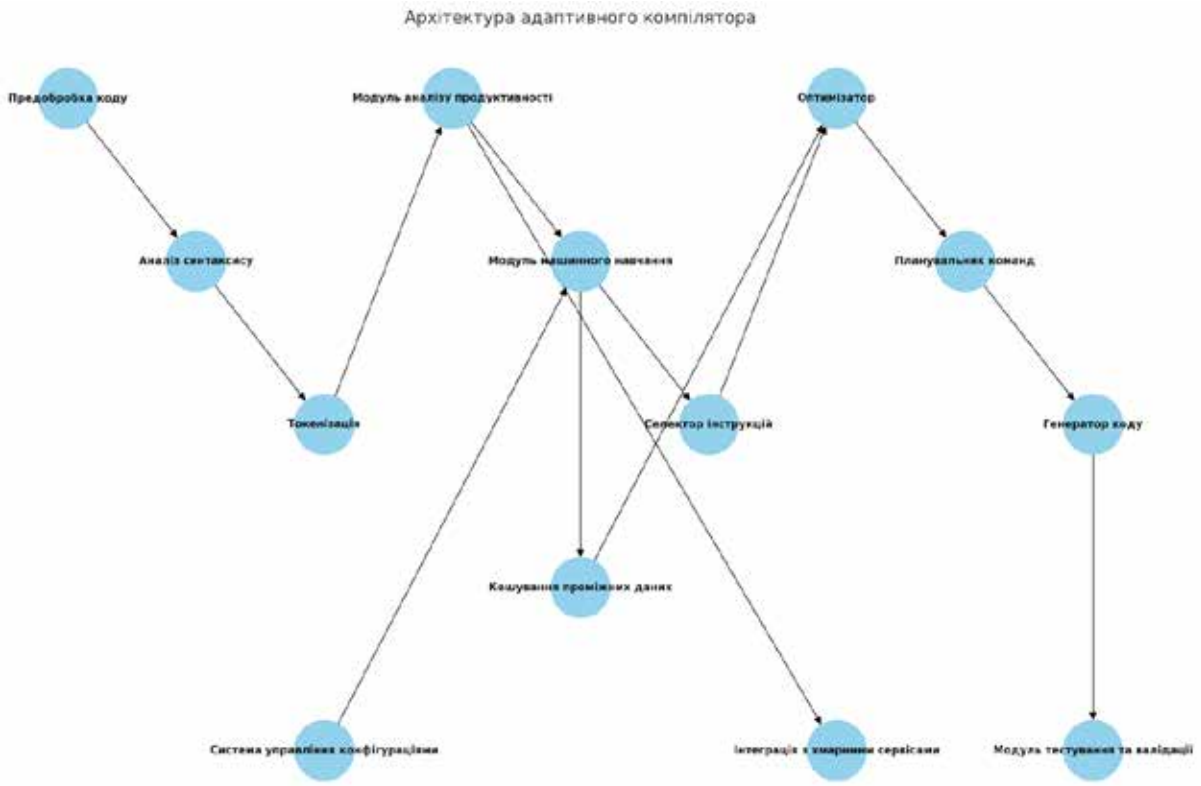


Рис. 1. Архітектура адаптивного компілятора

компілятора. Це значно знижує витрати на розробку та покращує якість і продуктивність програмного забезпечення.

На рис. 2 представлено, як адаптивний компілятор використовує методи машинного навчання для прийняття рішень на основі аналізу продуктивності. Зображено потік даних від вхідного коду через модуль аналізу продуктивності до вибору оптимальних інструкцій і налаштувань компіляції, який ілюструє циклічний процес навчання компілятора, де результати виконання коду впливають на подальші рішення компілятора.

Адаптивний компілятор можливо налаштувати для роботи за трьома сценаріями:

Сценарій 1: Оптимізація для енергозбереження.

Сценарій 2: Максимізація продуктивності.

Сценарій 3: Баланс продуктивності і енергозбереження.

В першому сценарії компілятор налаштований для максимального зниження енергоспоживання, що є критичним для мобільних пристроїв і вбудованих систем. Адаптивний компілятор використовує методи машинного навчання для вибору інструкцій, що мінімізують енергоспоживання без значної втрати продуктивності. Традиційний компілятор оптимізує інструкції за фіксованими правилами, які можуть бути менш ефективними.

Другий сценарій орієнтований на досягнення максимальної продуктивності програмного забезпечення, особливо важливий для серверних і десктопних процесорів, де швидкість виконання є пріоритетом. Адаптивний компілятор динамічно підлаштовується до архітектурних особливостей платформи, забезпечуючи оптимальний вибір інструкцій і розподіл ресурсів, тоді як традиційний компілятор використовує стандартні оптимізації. В третьому сценарії розглядається компроміс між продуктивністю та енергозбереженням, що актуально для широкого спектру застосувань, де важливо забезпечити прийнятний рівень продуктивності при збереженні ефективного енергоспоживання. Адаптивний компілятор використовує машинне навчання для знаходження оптимального балансу, а традиційний компілятор застосовує універсальні стратегії оптимізації.

Внутрішня реалізація адаптивного компілятора складається з наступних етапів:

1. Аналіз коду.
2. Використання моделі глибокого навчання.
3. Адаптивна компіляція.

На першому етапі використовуються інструменти LLVM, такі як Clang і opt, для генерації LLVM IR (Intermediate Representation) з вихідних кодів на C++.

Цей етап передбачає:

1. Генерацію LLVM IR: За допомогою Clang вихідний код компілюється в проміжне представлення LLVM, що дозволяє більш детальний

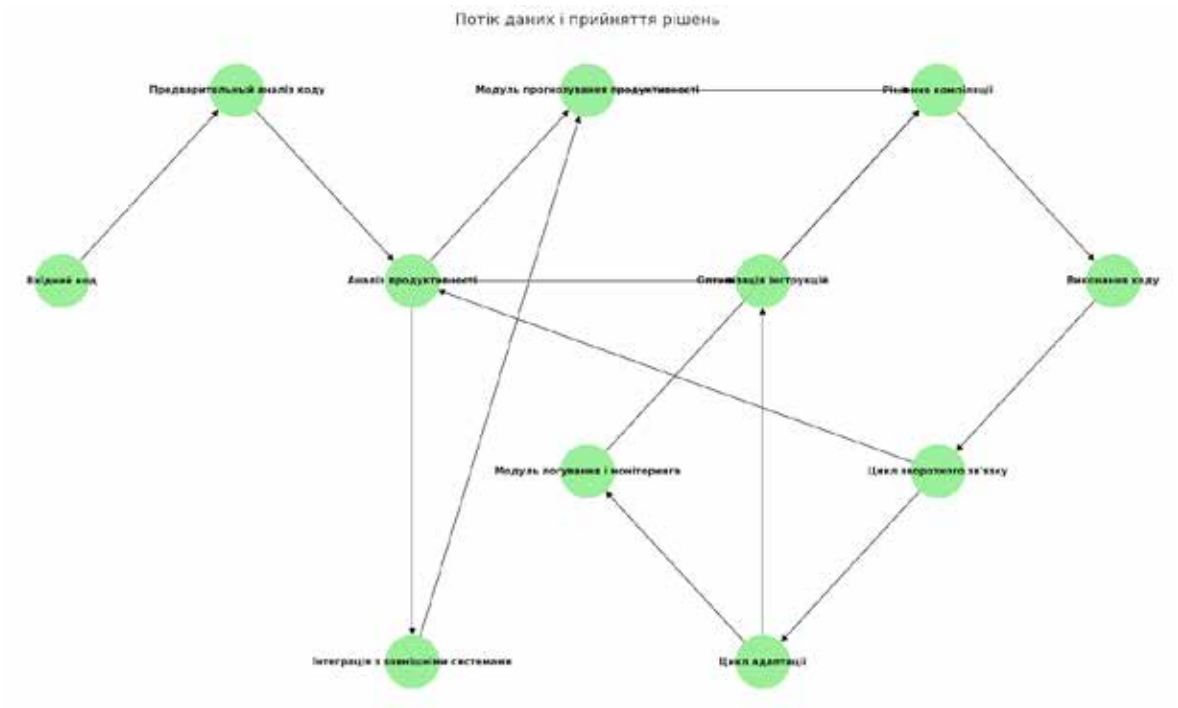


Рис. 2. Потік даних і прийняття рішень у адаптивному компіляторі

аналіз коду на рівні інструкцій і структур управління потоком.

2. Збір метрик продуктивності: Інструмент opt з прапором -stats збирає статистичні дані і метрики, такі як кількість інструкцій, цикли, використання регістрів і інші ключові показники продуктивності IR. Ці метрики є основою для подальшого навчання моделі машинного навчання, оскільки вони надають детальну інформацію про структуру та поведінку коду на рівні машинних інструкцій.

Другий етап інтегрує алгоритми машинного навчання, реалізовані на основі TensorFlow, для створення моделі, що прогнозує оптимальні параметри компіляції. Процес включає:

1. Попередня обробка і підготовка даних: Зібрані метрики обробляються і перетворюються у формат, що підходить для навчання. Це включає нормалізацію даних, вибір релевантних ознак і підготовку навчальних наборів.

2. Навчання нейронної мережі: Модель машинного навчання будується з використанням багатошарових нейронних мереж (MLP або LSTM), які здатні аналізувати складні залежності між метриками коду і бажаними оптимізаціями компіляції.

3. Адаптація і тюнінг моделі: Процес навчання включає також вибір гіперпараметрів, таких як розмір шарів, кількість нейронів, функції активації і швидкість навчання, щоб досягти максимальної точності прогнозів. Навчена модель зберігається у форматі, сумісному з TensorFlow, для подальшого використання на етапі компіляції.

Третій етап проекту фокусується на інтеграції прогнозів моделі машинного навчання в процес компіляції, використовуючи LLVM як бекенд для глибоких оптимізацій на рівні інструкцій. Цей етап включає:

1. Інтеграція з TensorFlow: Адаптивний компілятор завантажує навчений граф TensorFlow і використовує його для виконання інференсу в режимі реального часу. На основі поточних метрик IR, модель прогнозує оптимальні параметри компіляції, такі як вибір проходів оптимізації LLVM.

2. Динамічне управління оптимізаціями: Використовуючи прогнозовані оптимізації, компілятор налаштовує LLVM Pass Manager, додаючи чи видаляючи специфічні проходи оптимізації, такі як інструкційна комбінаторика, глобальне числове зниження (GVN), реасоціація виразів, спрощення графів управління потоком (CFG) тощо. Це дозволяє динамічно адаптувати компіляцію до особливостей коду, забезпечуючи оптимальний баланс між продуктивністю і ресурсозатратами. 3. Фінальна збірка і виведення артефактів: Фінальний виконуваний файл генерується на основі адаптивно оптимізованого IR. Це дозволяє отримати код, який максимально відповідає продуктивності та вимогам до специфічної обчислювальної платформи, на якій він буде виконуватися.

Для оптимізацій використовуємо нейронну мережу для аналізу коду на рівні LLVM IR та визначення оптимальних трансформацій:

1. Збір метрик з LLVM IR: використовуємо llvm-opt та аналізатор інструкцій для отримання реальних метрик.

2. Підготовка даних для навчання моделі: читаємо ці метрики і використовуємо їх для навчання моделі машинного навчання.

3. Навчання моделі з реальними даними: використовуємо TensorFlow для створення та навчання моделі.

Код на Python з TensorFlow має наступний вигляд:

```
import tensorflow as tf
from tensorflow.keras.models import
Sequential
from tensorflow.keras.layers import Dense,
LSTM, Embedding
from sklearn.model_selection import train_
test_split
import numpy as np
import pandas as pd
import subprocess
import re

def extract_llvm_metrics(ir_file):
    result = subprocess.run(
        [‘opt’, ‘-stats’, ‘-analyze’, ir_
file],
        stdout=subprocess.PIPE,
        stderr=subprocess.PIPE,
        text=True
    )

    metrics = {}
    for line in result.stderr.splitlines():
        match = re.match(r»(\d+)\s+(\w+)»,
line)
        if match:
            count, metric = match.groups()
            metrics[metric] = int(count)

    metrics_values = list(metrics.values())
    return metrics_values

def collect_real_data(ir_files):
    data = []
    for ir_file in ir_files:
        metrics = extract_llvm_metrics(ir_file)
```

```
        if metrics:
            data.append(metrics)
    return np.array(data)

ir_files = ['main.bc', 'module1.bc',
'module2.bc']
X = collect_real_data(ir_files)
y = np.random.randint(0, 10, size=(X.
shape[0], 1))

X_train, X_test, y_train, y_test = train_
test_split(X, y, test_size=0.2, random_
state=42)

def create_model(input_shape):
    model = Sequential()
    model.add(Dense(128, activation='relu',
input_shape=(input_shape,)))
    model.add(Dense(64, activation='relu'))
    model.add(Dense(32, activation='relu'))
    model.add(Dense(10,
activation='softmax'))
    return model

model = create_model(X_train.shape[1])
model.compile(optimizer='adam',
loss='sparse_categorical_crossentropy',
metrics=['accuracy'])
model.fit(X_train, y_train, epochs=20,
validation_data=(X_test, y_test))

model.save(«optimization_recommender.h5»)
```

Для інтеграції з LLVM (C++ з MLIR та моделями TensorFlow) інтегруємо модель оптимізації з процесом компіляції LLVM, щоб використовувати її рекомендації при генерації IR і наступного машинного коду:

```
#include <iostream>
#include <tensorflow/c/c_api.h>
#include <llvm/IR/LLVMContext.h>
#include <llvm/IR/IRBuilder.h>
#include <llvm/IR/Module.h>
#include <llvm/IR/PassManager.h>
#include <llvm/Transforms/Scalar.h>
#include <llvm/Support/TargetSelect.h>
#include <llvm/Support/TargetRegistry.h>
#include <llvm/IR/LegacyPassManager.h>
#include <llvm/Transforms/Scalar/GVN.h>
#include <llvm/Transforms/Scalar/Reassociate.
h>
#include <llvm/Transforms/Utils.h>
#include <llvm/Analysis/InstructionCount.h>
#include <llvm/Analysis/LoopInfo.h>
#include <llvm/Analysis/CGSCCPassManager.h>
#include <llvm/Analysis/InlineAdvisor.h>
#include <vector>
#include <fstream>

TF_Session* LoadModel(const std::string&
model_path, TF_Graph* graph) {
    TF_Status* status = TF_NewStatus();
    TF_SessionOptions* session_opts = TF_
NewSessionOptions();
    TF_Buffer* run_opts = nullptr;

    TF_Buffer* graph_def = TF_NewBuffer();
    std::ifstream file(model_path,
std::ios::binary | std::ios::ate);
    std::streamsize size = file.tellg();
    file.seekg(0, std::ios::beg);

    std::vector<char> buffer(size);
    if (file.read(buffer.data(), size)) {
        graph_def->data = buffer.data();
        graph_def->length = size;
        graph_def->data_deallocator = []
(void*, size_t, void*) {};
        TF_ImportGraphDefOptions* graph_opts
= TF_NewImportGraphDefOptions();
        TF_GraphImportGraphDef(graph, graph_
def, graph_opts, status);
        TF_DeleteImportGraphDefOptions(gr
aph_opts);
    }
    TF_DeleteBuffer(graph_def);

    if (TF_GetCode(status) != TF_OK) {
        std::cerr << «Error loading model: «
<< TF_Message(status) << std::endl;
        TF_DeleteStatus(status);
        return nullptr;
    }

    TF_Session* session = TF_
NewSession(graph, session_opts, status);
    TF_DeleteSessionOptions(session_opts);
    TF_DeleteStatus(status);
    return session;
}
std::vector<float> PredictOptimizations(TF_
Session* session, TF_Graph* graph, const
std::vector<float>& input_data) {
    TF_Status* status = TF_NewStatus();

    TF_Output input_op = {TF_
GraphOperationByName(graph, «serving_default_
input_1»), 0};
    if (input_op.oper == nullptr) {
        std::cerr << «Error: input operation
not found in the graph.» << std::endl;
        TF_DeleteStatus(status);
        return {};
    }
```

```
    int64_t dims[] = {1, static_cast<int64_
t>(input_data.size())};
    TF_Tensor* input_tensor = TF_
NewTensor(TF_FLOAT, dims, 2, input_data.
data(), input_data.size() * sizeof(float),
nullptr, nullptr);

    TF_Output output_op = {TF_
GraphOperationByName(graph,
«StatefulPartitionedCall»), 0};
    if (output_op.oper == nullptr) {
        std::cerr << «Error: output
operation not found in the graph.» <<
std::endl;
        TF_DeleteStatus(status);
        return {};
    }

    TF_Tensor* output_tensors[1] =
{nullptr};

    TF_SessionRun(session,
                  nullptr,
                  &input_op, &input_tensor,
1,
                  &output_op, output_tensors,
1,
                  nullptr, 0,
                  nullptr,
                  status);

    if (TF_GetCode(status) != TF_OK) {
        std::cerr << «Error during model
inference: « << TF_Message(status) <<
std::endl;
        TF_DeleteStatus(status);
        return {};
    }

    float* predictions = static_cast<float*>
(TF_TensorData(output_tensors[0]));
    std::vector<float> result(predictions,
predictions + TF_TensorElementCount(output_
tensors[0]));

    TF_DeleteTensor(input_tensor);
    TF_DeleteTensor(output_tensors[0]);
    TF_DeleteStatus(status);

    return result;
}
std::vector<float> ExtractMetricsFromIR(llvm:
:Module& module) {
    std::vector<float> metrics;

    llvm::legacy::FunctionPassManager
fpm(&module);
    llvm::FunctionAnalysisManager fam;
    fpm.add(llvm::createInstructionCombining
Pass());
    fpm.add(llvm::createReassociatePass());
    fpm.add(llvm::createGVNPass());
    fpm.add(llvm::createCFGSimplificationP
ass());

    fpm.doInitialization();

    int totalInstructions = 0;
    int totalLoops = 0;

    for (auto& func : module) {
        if (!func.isDeclaration()) {
            auto& instrCount = fam.getResult<
llvm::InstructionCountAnalysis>(func);
            totalInstructions += instrCount.
getTotalInsts();

            auto& loopInfo = fam.getResult<ll
vm::LoopAnalysis>(func);
            totalLoops += loopInfo.
getLoopsInPreorder().size();
        }
    }

    metrics.push_back(static_cast<float>(tota
lInstructions));
    metrics.push_back(static_
cast<float>(totalLoops));
    metrics.push_back(static_
cast<float>(module.getFunctionList().
size()));

    return metrics;
}
void ApplyAdaptiveOptimizations(llvm::Mo
dule& module, const std::vector<float>&
optimizations) {
    llvm::PassManagerBuilder builder;
    builder.OptLevel = 3;

    llvm::legacy::PassManager passManager;

    if (optimizations.size() >= 4) {
        if (optimizations[0] > 0.5)
passManager.add(llvm::createInstructionCombin
ingPass());

        if (optimizations[1] > 0.5)
passManager.add(llvm::createReassociateP
ass());
        if (optimizations[2] > 0.5)
passManager.add(llvm::createGVNPass());
        if (optimizations[3] > 0.5)
passManager.add(llvm::createCFGSimplificatio
nPass());
```

```
  } else {
        std::cerr << «Error: insufficient
number of optimization recommendations.» <<
std::endl;
    }

    passManager.run(module);
}

int main() {
    llvm::InitializeNativeTarget();
    llvm::LLVMContext context;
    llvm::Module module(«adaptive_compiler»,
context);

    std::string inputFilename = «main.bc»;
    llvm::SMDiagnostic err;
    std::unique_ptr<llvm::Module> mod = llvm:
:parseIRFile(inputFilename, err, context);

    if (!mod) {
        std::cerr << «Failed to parse LLVM IR
file: « << inputFilename << std::endl;
        return 1;
    }

    TF_Graph* graph = TF_NewGraph();
    TF_Session* session =
LoadModel(«optimization_recommender.h5»,
graph);

    if (!session) {
        std::cerr << «Failed to load the
model!» << std::endl;
        return 1;
    }

    std::vector<float> metrics =
ExtractMetricsFromIR(*mod);
    std::vector<float> optimizations =
PredictOptimizations(session, graph, metrics);

    if (!optimizations.empty()) {
        ApplyAdaptiveOptimizations(*mod,
optimizations);
        std::cout << «Adaptive compilation
completed successfully!» << std::endl;
    } else {
        std::cerr << «Failed to obtain
optimization recommendations.» << std::endl;
    }

    TF_DeleteSession(session, TF_
NewStatus());
    TF_DeleteGraph(graph);

    return 0;
}
```

Інтегруємо весь процес у пайплайн CI/CD, щоб кожна збірка автоматично виконувала адаптивну оптимізацію на основі реальних даних:

```
stages:
  - analyze
  - train
  - compile

analyze:
  stage: analyze
  image: llvm:latest
  script:
    - clang -O0 -emit-llvm -c main.cpp -o
main.bc
    - clang -O0 -emit-llvm -c module1.cpp -o
module1.bc
    - clang -O0 -emit-llvm -c module2.cpp -o
module2.bc
    - opt -stats -analyze main.bc > main.
stats 2>&1
    - opt -stats -analyze module1.bc >
module1.stats 2>&1
    - opt -stats -analyze module2.bc >
module2.stats 2>&1
  artifacts:
    paths:
      - main.bc
      - module1.bc
      - module2.bc
      - main.stats
      - module1.stats
      - module2.stats

train:
  stage: train
  image: tensorflow/tensorflow:latest
  script:
    - pip install pandas numpy scikit-learn
tensorflow
    - python code_analyzer.py
  artifacts:
    paths:
      - optimization_recommender.h5

compile:
  stage: compile
  image: llvm:latest
  script:
    - clang -O0 -emit-llvm -c main.cpp -o
main.bc
    - g++ llvm_optimizer.cpp -o llvm_
optimizer `llvm-config --cxxflags --ldflags
--system-libs --libs all` -ltensorflow
    - ./llvm_optimizer
    - clang main.bc -o main_optimized -O3
  artifacts:
    paths:
      - main_optimized
```

На рис. 3 представлено графік продуктивність коду, скомпільованого традиційним компілятором і адаптивним компілятором, на кількох різних обчислювальних платформах (наприклад, мобільні процесори, десктопні процесори, серверні процесори). По осі X зазначені різні платформи, а по осі Y – показники продуктивності, такі як час виконання або споживання енергії. Графік ілюструє переваги адаптивного компілятора в підвищенні продуктивності та ефективності роботи програмного забезпечення на різних архітектурах. Мобільний CPU: Qualcomm Snapdragon 835. Настільний CPU: AMD Ryzen 5 1600. Серверний CPU: Intel Xeon E5-2697 v4.

На рис. 4 представлено графік, який показує порівняння продуктивності традиційного та адаптивного компіляторів на різних обчислювальних платформах, а також демонструє рівень енергозбереження. Він наочно ілюструє, як адаптивний компілятор значно покращує продуктивність на всіх платформах, при цьому забезпечуючи кращу енергоефективність у порівнянні з традиційним підходом. Mobile ARM: Qualcomm Snapdragon 835. Desktop x86: AMD Ryzen 5 1600. Server x64: Intel Xeon E5-2697 v4. Embedded MIPS: Broadcom BCM3302. GPU CUDA: NVIDIA GeForce GTX 1070.

На рис. 5 показано порівняння часу компіляції та використання пам'яті традиційного та адаптивного компіляторів на різних обчислювальних платформах. Графік демонструє, що адаптивний компілятор не лише знижує час компіляції, але й суттєво оптимізує використання пам'яті, що є критичним фактором для продуктивності в умовах обмежених ресурсів.

На рис. 6. показано порівняння використання CPU та кількості помилок компіляції традиційного та адаптивного компіляторів на різних обчислювальних платформах. Він ілюструє, що адаптивний компілятор не тільки знижує використання CPU, але й суттєво зменшує кількість помилок під час компіляції, що є важливим показником для забезпечення якості та ефективності компіляції на різних архітектурах.

Тести виконувалися на кількох типах платформ, що включали:

– мобільний процесор: Qualcomm Snapdragon 835;
– десктопний процесор: AMD Ryzen 5 1600;
– серверний процесор: Intel Xeon E5-2697 v4;
– Embedded-процесор (вбудовані системи): Broadcom BCM3302;

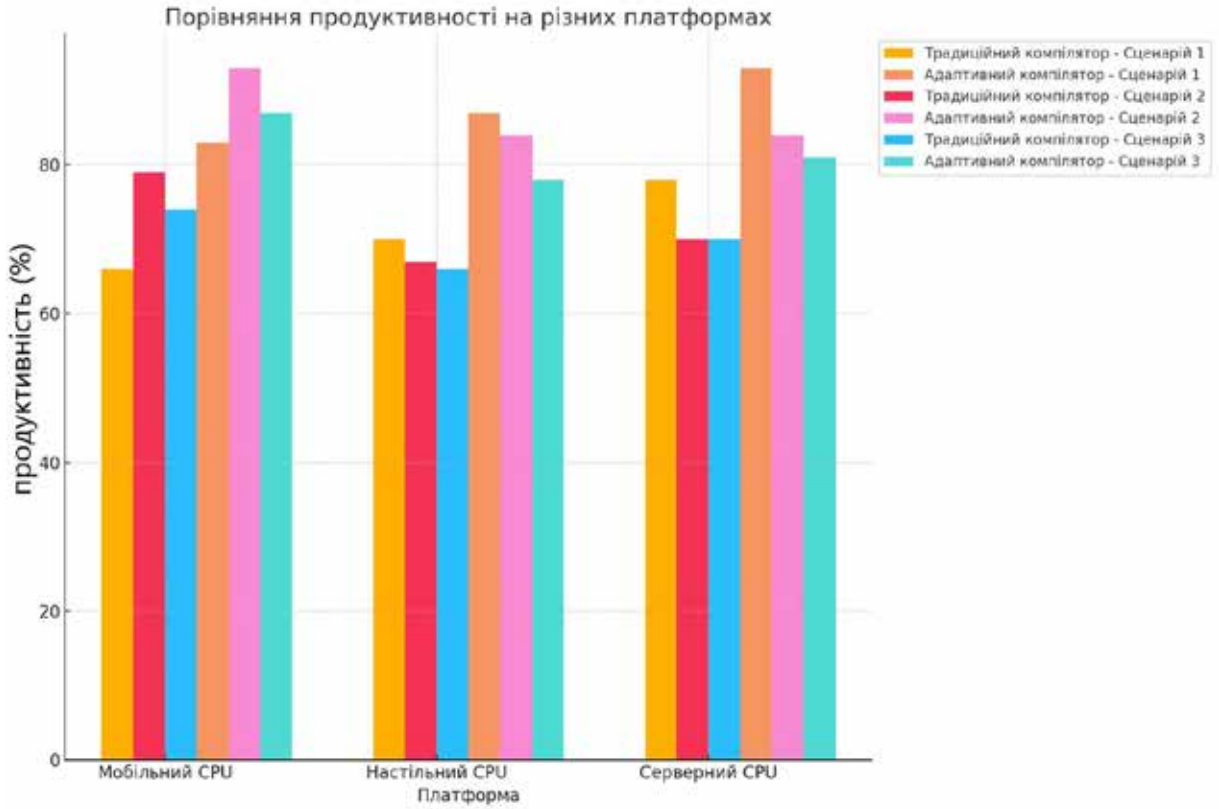


Рис. 3. Графік порівняння продуктивності коду на різних платформах

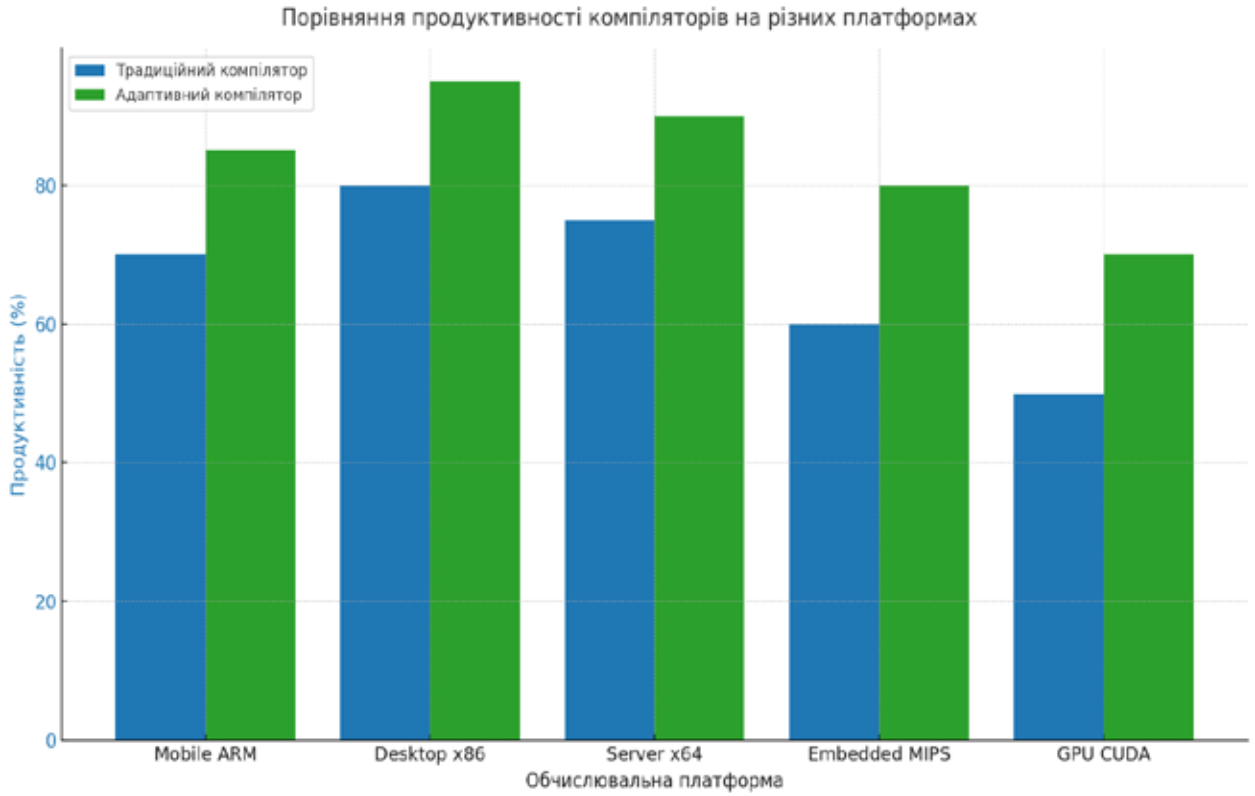


Рис. 4. Порівняння продуктивності

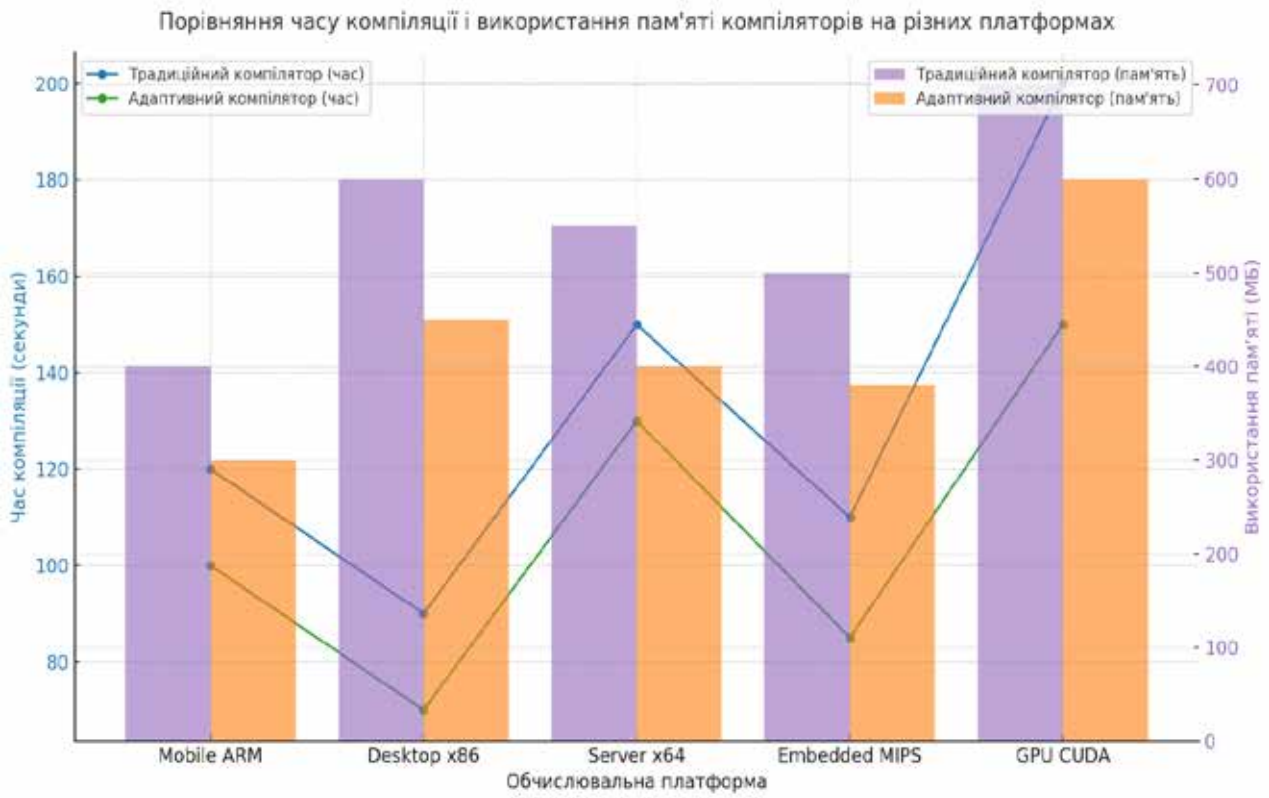


Рис. 5. Порівняння часу компіляції

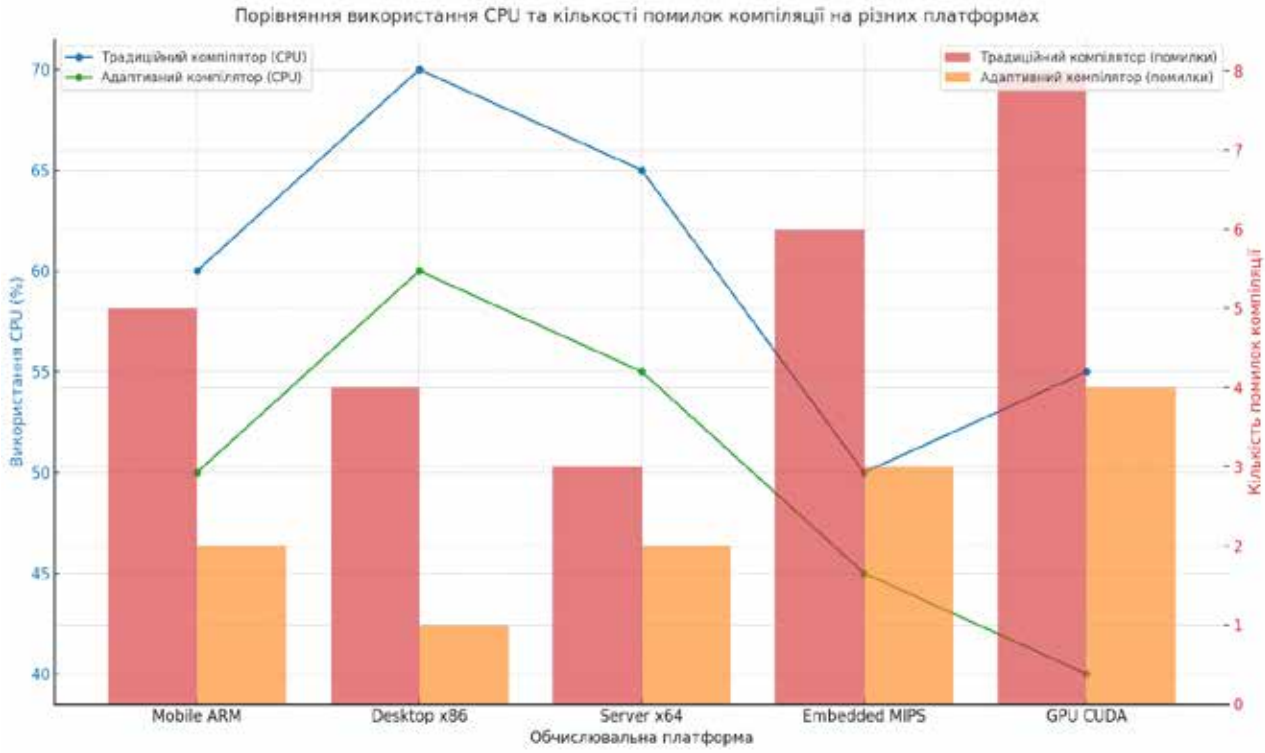


Рис. 6. Порівняння використання CPU

– графічний процесор для обчислень на CUDA: NVIDIA GeForce GTX 1070.

Тести передбачали три різні сценарії оптимізації для адаптивного компілятора:

1. Енергозбереження: адаптивний компілятор був налаштований на мінімізацію споживання енергії, що особливо важливо для мобільних та вбудованих систем.

2. Максимальна продуктивність: орієнтований на максимальну швидкість виконання, що критично для серверних та десктопних процесорів.

3. Баланс між продуктивністю та енергозбереженням: забезпечення високої продуктивності при розумному рівні енергоспоживання.

При тестуванні були використані операційні системи для тестування та інструменти для збору метрик:

1. Мобільні пристрої (Qualcomm Snapdragon 835): Android 11.0, що забезпечує стандартні умови для енергозбереження та продуктивності в мобільному середовищі.

2. Десктопні пристрої (AMD Ryzen 5 1600): Ubuntu 20.04 LTS, яка забезпечує стабільну підтримку компіляційних інструментів та оптимізацій для тестування продуктивності на настільних платформах.

3. Серверні системи (Intel Xeon E5-2697 v4): CentOS 8.0, яка часто використовується в серверних середовищах завдяки своїй стабільності та ефективності при обробці великих обчислювальних завдань.

4. Вбудовані системи (Broadcom BCM3302): OpenWRT 19.07, операційна система, оптимізована для пристроїв з обмеженими ресурсами, що дозволяє протестувати ефективність адаптивного компілятора для енергозбереження в умовах обмеженого обладнання.

5. CUDA-платформи (NVIDIA GeForce GTX 1070): Ubuntu 18.04 з підтримкою CUDA 11.0, що забезпечує оптимальні умови для обчислювальних задач, таких як тренування та використання моделей машинного навчання.

6. LLVM opt: Використовувався для збору метрик продуктивності Intermediate Representation (IR) під час компіляції, таких як кількість інструкцій, використання регістрів, кількість циклів тощо.

7. TensorFlow та нейронні мережі: Використовувалися для створення моделей прогнозування оптимальних параметрів компіляції. Нейронні мережі тренувалися на основі зібраних даних про продуктивність.

8. Профайлер енергоспоживання: Вбудоване в операційну систему програмне забезпечення для вимірювання споживання енергії, яке дозволяло оцінити ефективність компіляції в режимах енергозбереження.

9. CI/CD пайплайн: Автоматизація процесу тестування і компіляції була організована через

CI/CD пайплайн, що включав етапи аналізу, тренування моделі та компіляції з оптимізацією.

Одним із яскравих прикладів застосування адаптивних компіляторів є оптимізація програмного забезпечення для мобільних пристроїв, де обмежені ресурси (як-от пам'ять, обчислювальна потужність і енергоефективність) вимагають ретельної оптимізації коду. Використання адаптивного компілятора з методами машинного навчання дозволяє автоматично налаштовувати код для виконання на конкретних моделях мобільних процесорів, таких як ARM або Snapdragon, враховуючи особливості їх архітектури та інструкцій.

Адаптивний компілятор аналізує виконання програм у реальному часі та автоматично змінює стратегії оптимізації, підлаштовуючи розподіл ресурсів процесора, управління пам'яттю та вибір інструкцій так, щоб досягти максимальної продуктивності при мінімальному споживанні енергії. Це особливо важливо для додатків з інтенсивним використанням ресурсів, таких як ігри або мультимедійні програми, де навіть незначні оптимізації можуть суттєво вплинути на зручність користування та тривалість роботи пристрою від батареї.

Іншим прикладом є використання адаптивних компіляторів у хмарних обчисленнях, де автоматична оптимізація дозволяє знижувати витрати на обчислювальні ресурси, підвищувати швидкість обробки даних і ефективність використання серверів. Адаптивний компілятор може динамічно адаптувати виконання коду на хмарних платформах, враховуючи специфіку віртуальних машин або контейнерів, що використовуються, що дозволяє забезпечити максимальну продуктивність і гнучкість у розподілі навантаження.

Висновки і перспективи подальших досліджень. В роботі було розглянуто створення адаптивного компілятора, що використовує методи машинного навчання та автоматичної оптимізації для адаптації програмного забезпечення до різних обчислювальних платформ. Запропонований підхід дозволяє значно підвищити продуктивність і ефективність програмного забезпечення шляхом динамічного налаштування процесу компіляції на основі реальних даних про продуктивність і характеристики платформи виконання. Інтеграція таких компонентів, як модуль аналізу продуктивності, селектор інструкцій, планувальник команд, і модуль машинного навчання, забезпечує гнучкість і масштабованість у розробці програмного забезпечення.

Одним із ключових аспектів адаптивного компілятора є здатність до рефлексивного аналізу власного стану і адаптація оптимізацій у режимі реального часу. Це досягається через інтеграцію модуля машинного навчання, який використовує алгоритми навчання з підкріпленням для прогнозування оптимальних проходів компіляції на основі метрик продуктивності, зібраних під час виконання програм. Використання циклів зворотного зв'язку та адаптації дозволяє компілятору не тільки налаштовувати інструкції і плани виконання, але й адаптуватися до змін у робочих навантаженнях та апаратній архітектурі.

Додаткові компоненти, такі як модуль логування та моніторингу, забезпечують детальний контроль над процесом компіляції, дозволяючи виявляти і виправляти проблеми на ранніх стадіях. Інтеграція з зовнішніми системами і хмарними сервісами дозволяє використовувати великі обчислювальні ресурси для тренування моделей машинного навчання, що підвищує загальну адаптивність і продуктивність компілятора. Перспективи розвитку адаптивних компіляторів полягають у подальшій автоматизації процесу оптимізації з використанням передових алгоритмів глибинного навчання, таких як рекурентні нейронні мережі та трансформери, які можуть обробляти складні залежності в коді і надавати ще більш точні рекомендації щодо оптимізації. Також можливе впровадження методів активного навчання, де компілятор самостійно визначає, які додаткові дані йому необхідно зібрати для покращення своїх моделей.

Іншим важливим напрямом є розробка стандартів і специфікацій для адаптивних компіляторів, що дозволить уніфікувати підходи до адаптації програмного забезпечення на різних платформах і забезпечить сумісність між різними інструментами розробки. Це особливо актуально у контексті розвитку новітніх архітектур, таких як квантові обчислення та нейроморфні процесори, які вимагають специфічних підходів до компіляції та оптимізації. Узагальнюючи, адаптивні компілятори з інтеграцією машинного навчання представляють значний крок вперед у напрямку створення високопродуктивних та гнучких програмних систем, здатних автоматично адаптуватися до змін у технологічному середовищі та вимогах користувачів. Цей підхід сприяє зниженню витрат на розробку, підвищенню якості продуктів і розширенню можливостей програмного забезпечення в умовах швидкозмінних технологій.

ЛІТЕРАТУРА:

1. Z. Fisches. Neural Self-Supervised Models of Code. Masters thesis, ETH Zurich, 2020.

2. Saxena S. Machine Learning for Compilers and Architecture. 2021. 3. Google Research. MLGO: A Machine Learning Guided Compiler Optimizations Framework. 2021.

4. A. H. Ashouri, W. Killian, J. Cavazos, G. Palermo, and C. Silvano. A Survey on Compiler Autotuning using Machine Learning. *CSUR*, 51(5), 2018.

REFERENCES:

1. Z. Fisches. (2020). Neural Self-Supervised Models of Code. Masters thesis, ETH Zurich.

2. Saxena, S. (2021). Machine Learning for Compilers and Architecture. 3. Google Research. (2021). MLGO: A Machine Learning Guided Compiler Optimizations Framework.

4. A. H. Ashouri, W. Killian, J. Cavazos, G. Palermo, and C. Silvano. (2018). A Survey on Compiler Autotuning using Machine Learning. *CSUR*, 51(5).

УДК 004.42:658.8
DOI https://doi.org/10.32782/IT/2024-4-4

Володимир БОГОМ'Я
доктор технічних наук, професор, професор кафедри кібербезпеки, інформаційних технологій та економіки, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе 210, м. Київ, Україна, 02121
ORCID: *0000-0003-4403-3130*
Scopus Author ID: *51863292100*

Любов ЧЕРЕМІСІНА
викладач кафедри кібербезпеки, інформаційних технологій та економіки, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе 210, м. Київ, Україна, 02121
ORCID: *0009-0005-0719-0745*

Андрій ЯРМОЛАТІЙ
викладач кафедри кібербезпеки, інформаційних технологій та економіки, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе 210, м. Київ, Україна
ORCID: *0009-0004-8655-9928*

Бібліографічний опис статті: Богом'я, В., Черемісіна, Л., Ярмолатій, А. (2024). Безпечна обробка даних в бізнес-інтернет платформах та системах електронної комерції. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 29–36, doi: https://doi.org/10.32782/IT/2024-4-4

БЕЗПЕЧНА ОБРОБКА ДАНИХ В БІЗНЕС-ІНТЕРНЕТ ПЛАТФОРМАХ ТА СИСТЕМАХ ЕЛЕКТРОННОЇ КОМЕРЦІЇ

Глобалізація та розвиток мережі інтернет докорінно змінили способи електронної комерції та бізнес-платформ ставши рушійною силою сучасного суспільства. Інтернет перетворився на універсальну платформу для інтеграції бізнесів, наукових спільнот і громадян світу, забезпечуючи безпрецедентний рівень взаємодії та доступу до комерційних послуг.

Проте стрімке поширення інформаційних технологій несе із собою й суттєві ризики. Зокрема, питання захисту чутливих даних користувачів, які можуть бути викрадені, втратити конфіденційність або навіть використані для маніпуляцій, стали центральними у сфері інформаційної безпеки. Це загрожує не лише окремим користувачам, але й підриває довіру до цифрових сервісів у цілому, що перешкоджає їхньому подальшому розвитку.

Мета роботи. *Розгляд досвіду і викликів практичного впровадження стандартів резиденції даних в бізнес платформах, які мають клієнт-серверну архітектуру, що мають відповідати сучасним вимогам захисту систем електронної комерції.*

Методологія дослідження. *В роботі використані емпіричні та теоретичні методи.*

Наукова новизна. *Проаналізовано основні технічні підходи для забезпечення відповідності вимогам щодо резиденції даних в бізнес-інтернет платформах та системах електронної комерції, а сам: географічне розміщення серверів, шифрування та токенізація, обмеження доступу та проксі-сервери.*

Відсутність єдиних підходів до визначення чутливих даних і механізмів їх захисту створює ситуації, коли особиста інформація користувачів стає вразливою. Розв'язання цих проблем вимагає створення та удосконалення міжнародних стандартів регулювання захисту даних, які будуть враховувати особливості глобального цифрового середовища та сприяти гармонізації національних законодавств.

Висновки. *В статті наведено, що питання резиденції даних залишається одним з головних технічних і юридичних викликів сьогодення для сучасного бізнесу та систем електронної комерції, де основним каменем спотикання є сама клієнт-серверна архітектура сучасних веб-платформ та інших додатків цієї ж архітектури, які працюють в глобальній мережі.*

Ключові слова: *системи електронної комерції, мережа Інтернет, захист даних, обробка даних, безпека інформації, бізнес-інтернет платформа, кібербезпека.*

Volodymyr BOHOMYA
Doctor of Technical Sciences, Professor, Professor at the Department of Cyber Security, Information Technologies and Economics, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121, bog260341@gmail.com
ORCID: *0000-0003-4403-3130*
Scopus Author ID: *51863292100*

Liubov CHEREMISINA
Lecturer at the Department of Cybersecurity, Information Technology and Economics, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121, cheremisina1112@gmail.com
ORCID: *0009-0005-0719-0745*

Andrii YARMOLATII
Lecturer at the Department of Cybersecurity, Information Technology and Economics, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121, ayarmolatii@gmail.com
ORCID: *0009-0004-8655-9928*

To cite this article: Bohomya, V., Cheremisina, L., Yarmolatii, A. (2024). Bezpechna obrobka danykh v biznes-Internet platformakh ta systemakh elektronnoi komertsii [Secure data processing in business-Internet platforms and e-commerce systems]. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 29–36, doi: https://doi.org/10.32782/IT/2024-4-4

SECURE DATA PROCESSING IN BUSINESS-INTERNET PLATFORMS AND E-COMMERCE SYSTEMS

Globalization and the development of the Internet have radically changed the ways in which information is exchanged and processed, becoming a destructive force in modern society. Every day, a huge amount of data is transmitted in various formats – text, audio, video, which contributes to progress in the economy, science, education and other areas. The Internet has become a universal platform for integrating businesses, scientific communities and citizens of the world, providing an unprecedented level of interaction and access to information.

However, the rapid spread of information technologies carries significant risks. In particular, the protection of sensitive user data, which can be stolen, lost privacy, or even used for manipulation, has become central in the field of information security. This not only threatens individual users, but also undermines trust in digital services as a whole, which hinders their further development.

Purpose of the work. *Consideration of the experience and challenges of practical implementation of data residency standards in business platforms that have client-server architecture, which must meet modern requirements for the protection of e-commerce systems.*

Research methodology. *Empirical and theoretical methods are used in the work.*

Scientific novelty. *The main technical approaches to ensure compliance with the requirements for data residency in business-Internet platforms and e-commerce systems are analyzed, namely: geographical location of servers, encryption and tokenization, access restrictions, and proxy servers.*

The lack of uniform approaches to identifying sensitive data and mechanisms for protecting it creates situations where users' personal information becomes vulnerable. Solving these problems requires the creation and improvement of international standards for data protection regulation, which will take into account the peculiarities of the global digital environment and contribute to the harmonization of national legislations.

Conclusions. *The article states that the issue of data residency remains one of the main technical and legal challenges of today for modern business and e-commerce systems, where the main stumbling block is the client-server architecture of modern web platforms and other applications of the same architecture that operate on the global network.*

Key words: *e-commerce systems, Internet, data protection, data processing, information security, business Internet platform, cybersecurity.*

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями. Сучасний етап розвитку інформаційного суспільства характеризується постійним збільшенням обсягів даних, що генеруються, передаються, обробляються та зберігаються в глобальній мережі Інтернет. Ця тенденція відкриває нові можливості для ефективного використання інформації в різних сферах, таких як наука, бізнес,

електронна комерція, освіта тощо. Проте, разом з цими можливостями з'являються серйозні загрози, пов'язані з витоком чутливих даних, несанкціонованим доступом та маніпуляцією інформацією. Зокрема, відсутність єдиних стандартів щодо визначення чутливих даних та методів їх захисту спричиняє ситуації, коли особиста інформація користувачів стає вразливою.

Електронна комерція – це придбання чи продаж товару за допомогою електронних носіїв, чи через мережу, подібну до Internet. Дане поняття може включати в себе замовлення, оплату та доставку товарів або послуг. Глобалізація цифрових технологій ускладнює застосування локальних законодавств щодо регулювання обробки даних. Географічна децентралізація сучасних клієнт-серверних додатків особливо в бізнес і державній сфері часто означає те, що серверна частина системи розташована в одній країні, а користувач – в іншій. Така структура створює правову колізію між законодавствами різних держав, що обмежує можливості ефективного захисту даних.

Ця проблема є не лише теоретичною, але й має суттєве практичне значення у бізнесі та електронній комерції – формування довіри клієнтів до цифрових сервісів залежить від здатності компаній забезпечувати конфіденційність їхніх даних.

Розв'язання цієї проблем вимагає створення та удосконалення міжнародних стандартів регулювання захисту даних, які будуть враховувати особливості глобального цифрового середовища та сприяти гармонізації національних законодавств.

Аналіз останніх досліджень і публікацій. Проблема відповідності архітектури інформаційних систем вимогам щодо місця зберігання даних (data residency) є однією з основних у сучасних дослідженнях безпеки даних та регуляторної відповідності. Деякі дослідження висвітлюють різні аспекти цієї проблеми, зокрема формування стандартів, технічних підходів та правових рішень.

Найчастіше представляють фреймворк для рекомендацій щодо архітектур, що відповідають вимогам місця зберігання даних. Дослідження зосереджено на автоматизації вибору архітектурних рішень, які враховують регіональні нормативні вимоги, але не розглядають питання децентралізованих клієнт-серверних систем.

Інколи робота фокусується на переміщенні та перерозподілі веб-сервісів для забезпечення відповідності вимогам щодо резиденції даних. У статті розглянуто алгоритмічні підходи до мінімізації витрат на релокацію сервісів. Однак у роботі недостатньо розкрито питання інтеграції з існуючими правовими нормами різних країн.

У дослідженнях (Кларк, Джейкоб, 2018; Стьопочкін, Новіков, 2022) пропонується сервісний підхід до вирішення проблеми місця зберігання даних у хмарі. Описуються методи шифрування та управління доступом, однак недостатньо висвітлено організаційні аспекти впровадження таких рішень у великих розподілених системах.

Автори аналізують виклики та можливості стандартизації питань щодо резиденції даних. У роботі наголошено на важливості створення міжнародних стандартів, які враховують як технічні, так і правові аспекти (Богом'я, Кочегаров, 2023; Богом'я, Черемісіна, Ярмолатій, Бараненко, 2024).

Мета статті. Розгляд досвіду і викликів практичного впровадження стандартів резиденції даних в бізнес платформах, які мають клієнт-серверну архітектуру, що мають відповідати сучасним вимогам захисту систем електронної комерції.

Виклад основного матеріалу дослідження. Для розкриття теми варто розділити відповідь га гранульовані підпункти: об'єкти захисту, середовище захисту, методика захисту, варіанти технічної реалізації

Об'єкти захисту. Перш за все, важливо визначити, що таке чутливі дані – це інформація, яка потребує підвищеного захисту через її особисту, конфіденційну чи стратегічну цінність. Несанкціонований доступ до таких даних може мати серйозні негативні наслідки для окремих осіб, організацій або навіть для держави. До основних категорій чутливих даних належать: персональні дані (PII), які поділяються на звичайні та особливо чутливі, фінансові дані, конфіденційна інформація бізнесу та дані, що стосуються національної безпеки.

Персональні дані – це інформація, яка ідентифікує або може ідентифікувати конкретну особу. Прикладами можуть бути ім'я, адреса, номер телефону, ідентифікаційний номер (номер паспорта або соціального страхування), дата народження тощо. До особливо чутливих персональних даних, як правило, відносять більш критичні особисті дані, які потребують ще більшого захисту тобто медичні дані (медичні записи, історія захворювань – PHI, Personal Health Information), біометричні дані (відбитки пальців, скани сітківки), раса, етнічна приналежність, політичні або релігійні переконання, а також дані про сексуальну орієнтацію і гендерну ідентичність. Фінансовими даними вважаються дані, пов'язані з фінансовими

операціями, які можуть бути використані для шахрайства або крадіжки: банківські рахунки, номери кредитних карток (PCI – Payment Car Information), Інформація про доходи або кредити тощо.

Конфіденційні бізнес-дані – це дані, що стосуються комерційної діяльності компаній і є важливими для бізнесу. Серед них розрізняють – інтелектуальна власність (патенти, авторські права), стратегії розвитку, фінансові плани, комерційні таємниці, інформація про клієнтів, постачальників і партнерів. Даними національної безпеки вважається інформація, яка може поставити під загрозу безпеку держави, якщо вона потрапить до рук ворогів або конкурентів. Захист таких даних вимагає використання надійних технічних, організаційних і правових методів, які мінімізують ризики витоку, несанкціонованого доступу або інших загроз для конфіденційності й безпеки.

Середовище захисту. Сьогодні бізнеси найчастіше використовують веб-платформи для організації своєї діяльності, оскільки вони є основою для багатьох підприємств, незалежно від їхнього розміру. До найпоширеніших типів таких платформ належать: для управління ресурсами підприємства (ERP – Enterprise Resource Planning), для управління IT-послугами (ITSM – Information Technology Service Management), а також для управління продажами і взаєминами з клієнтами (CRM – Customer Relationship Management).

З точки зору власника бізнесу, веб-платформа – це комплекс програмних і апаратних компонентів, які дозволяють співробітникам і клієнтам (користувачам) взаємодіяти з даними та бізнес-логікою через інтернет, зокрема за допомогою хмарних сховищ. Однак з погляду IT-фахівця або спеціаліста з кібербезпеки, веб-платформа – це додаток з класичною клієнт-серверною архітектурою, де сам додаток складається з двох частин, між якими здійснюється зв'язок через глобальну мережу, зазвичай інтернет.

Перша частина (клієнтська) являє собою звичайний браузер. Це зручно, адже в такому випадку немає необхідності кожен раз створювати нову клієнтську частину для кожного нового додатку – достатньо лише розробити HTML-інтерфейс. В цій частині користувач має можливість переглядати інформацію від сервера в текстовому, графічному та мультимедійному форматах а також своїми діями(наприклад пошук) відправляти додаткові запити на серверну частину чи то для отримання додаткових даних чи для зміни та збереження поточних. Усі дані які бачить чи слухає користувач, будуть видалені з оперативної пам'яті клієнтської частини одразу після переходу на інший сайт чи закриття браузера. Друга частина (серверна) призначення для обробки запитів від користувачів та перманентного збереження і обробки даних. Як правило серверна частина архітектури клієнт-сервер завжди потужніша і як наслідок – більша в розмірах та не володіє такою мобільністю як клієнтська частина, тому і рідше змінює свою локацію чи майже не змінює. Зі зростанням глобалізації і цифровізації, дані клієнтів можуть розміщуватися на серверах у різних країнах. І тут постає питання «резиденції даних» – де саме знаходяться ці дані – на якому сервері і наскільки безпечно вони зберігаються і чи відповідає це вимогам регуляторів.

Дотримання вимог до резиденції даних особливо актуальне для бізнесів, що обробляють персональні дані або конфіденційну інформацію.

Все більше і більше країн приймають нові закони про контроль резиденства даних. Серед таких – GDPR (Загальний регламент про захист даних Європейського Союзу), який обмежує експорт персональних даних за межі ЄС; PDPA – Закон про захист персональних даних різних країн; HIPAA (Health Insurance Portability and Accountability Act) – американський закон, що регулює конфіденційність та безпеку медичних даних.

Україна намагається не відставати від сучасних світових тенденцій. Закон «Про інформацію», «Про захист персональних даних», «Про захист інформації в інформаційно-телекомунікаційних системах» тощо.

Резидентство даних розвивається і навіть стає пов'язане з питаннями національної безпеки та все частіше розглядається як торгова зброя і огляд іноземних інвестицій.

Такі норми встановлюють суворі вимоги до розташування та доступу до чутливих даних. Звісно у появи цих норм є природні і очевидні причини але має свої наслідки і вплив на технічну складову.

Методика захисту в питанні резиденції даних. Найбільшим фактором сьогодення, який створює технічні виклики є конфлікт стандартів та регуляцій – як вже було зазначено вище, різні країни мають різні вимоги до резиденції даних, що не лише ускладнює ведення бізнесу на міжнародному рівні де співпрацюють учасники різних країн на одній платформі але й інколи унеможливлює використання тих чи інших клієнт-серверних платформ.

Одні країни забороняють бачити дані за їх межами, інші – передавати дані, інші – зберігати

і т.д. Це призводить до необхідності максимальної гнучкості а отже до грануляції технічних операцій над даними.

Варіанти технічної реалізації резиденції. Існує кілька основних технічних підходів для забезпечення відповідності вимогам щодо резиденції даних в бізнес-інтернет платформах та системах електронної комерції. Ось деякі з найбільш поширених варіантів:

1. **Географічне розміщення серверів** – цей підхід передбачає розміщення дублікатів серверів веб-платформи в різних країнах, що дозволяє компаніям контролювати, де саме зберігаються дані користувачів. Перевагою є швидке вирішення питання, але серед недоліків – висока вартість і складність технічної підтримки та обслуговування в довгостроковій перспективі.

2. **Шифрування та токенізація** – шифрування даних при їх передачі допомагає мінімізувати ризики при збереженні або пересиланні інформації через міжнародні мережі. Цей метод часто використовується в комбінації з іншими підходами для посилення захисту.

3. **Обмеження доступу** – налаштування доступу до даних для співробітників залежно від їхньої ролі, місцезнаходження та необхідності працювати з конкретною інформацією. Це дозволяє додатково контролювати доступ до чутливих даних.

4. **Проксі-сервери** – ці сервери використовуються в сфері кіберзахисту для токенізації даних. Проксі-сервери можуть діяти як шлюзи, які токенізують чутливі дані перед їх передачею через мережу, зберігаючи оригінальні дані на локальному сервері в юрисдикції, де зберігання дозволено. Токенізація дозволяє передавати тільки зашифровані або частково зашифровані дані, що підвищує безпеку і допомагає відповідати вимогам регуляторів.

Висновки. В даній статті розглянуто досвід і виклики практичного впровадження стандартів резиденції даних в бізнес платформах які мають клієнт-серверну архітектуру, що відповідають сучасним вимогам захисту систем електронної комерції.

Стаття демонструє сучасні варіанти реалізації вимог і спроби уніфікувати будь-яку реалізацію резиденції даних. Та не дивлячись на те, що проблематика залишається відкритою до можливих нових рішень і на те, що сучасний бізнес готовий вкладати значні ресурси в питання резиденції даних, сама архітектура «клієнт-сервер» базується на фізичних реаліях і з'явилася як результат використання природної архітектури, що накладає свої обмеження, які неможливо обійти, що в свою чергу, створює сильну залежність та необхідність співпраці між світовим законодавством та інженерно-технічними реалізаціями.

ЛІТЕРАТУРА:

1. Framework for Recommending Data Residency Compliant Application Architecture. URL: https://ieeexplore.ieee.org/document/9712154 (дата звернення 09.10.2024).

2. Web Services Relocation and Reallocation for Data Residency Compliance. URL: https://ieeexplore.ieee.org/abstract/document/10181214 (дата звернення 09.10.2024).

3. The Impact of data residency on cloud computing. URL: https://ieeexplore.ieee.org/document/8418109 (дата звернення 09.10.2024).

4. Data residency as a service: a secure mechanism for storing data in the cloud. URL: https://www.inderscience.com/offers.php?id=100875 (дата звернення 09.10.2024).

5. Data Residency Challenges and Opportunities for Standardization. URL: https://www.inderscience.com/offers.php?id=100875 (дата звернення 09.10.2024).

6. Architecting for Compliance and Data Residency. URL: https://medium.com/salesforce-architects/architecting-for-compliance-and-data-residency-13e5d5d9a87f (дата звернення 09.10.2024)

7. Scale across borders: build a multi-region architecture while maintaining data residency. URL: https://community.aws/content/2dhVhtsciD5gVBlCKUlHoszrDzU/scale-beyond-borders (дата звернення 09.10.2024).

8. Managing Data Residency – concepts and theory. URL: https://blog.frankel.ch/data-residency/1/ (дата звернення 09.10.2024).

9. "How AI is transforming cybersecurity" by Gary Eastwood, " URL: https: / / www. Information – age. com/how-ai-is-transforming-cybersecurity-123478294/ (дата звернення 09.10.2024).

10. Кларк Дж., Джейкоб Дж. ШІ та кібербезпека: загрози та рішення. *Журнал кібербезпеки*, 2018. 4 (1), С. 1–14.

11. І. В. Стьопочкін, О. М. Новіков. Методи штучного інтелекту в кібербезпеці: навч. посіб. для здобувачів спец. 125 «Кібербезпека», 2022. 82 с.

12. Богом'я В. І., Кочегаров В. С. Кібербезпека в хмарних сервісах за допомогою застосування криптографічних методів. *Водний транспорт.* № 1 (37). 2023. С. 239–246. doi.org/10.33298/2226-8553.2023.1.37.27

13. Богом'я В. І., Черемісіна Л. О, Ярмолатій А. В, Бараненко О. О. Резиденція даних сучасних бізнес платформ, їх систем аутентифікації та особливості технічної реалізації в глобальній мережі internet. *Водний транспорт.* № 3 (41). 2024. С. 240–246.

REFERENCES:

1. Framework for Recommending Data Residency Compliant Application Architecture. Retrieved from: https://ieeexplore.ieee.org/document/9712154 (accessed 09.10.2024).

2. Web Services Relocation and Reallocation for Data Residency Compliance. Retrieved from: https://ieeexplore.ieee.org/abstract/document/10181214 (accessed 09.10.2024).

3. The Impact of Data Residency on Cloud Computing. Retrieved from: https://ieeexplore.ieee.org/document/8418109 (accessed 09.10.2024).

4. Data Residency as a Service: A Secure Mechanism for Storing Data in the Cloud. Retrieved from: https://www.inderscience.com/offers.php?id=100875 (accessed 09.10.2024).

5. Data Residency Challenges and Opportunities for Standardization. Retrieved from: https://www.inderscience.com/offers.php?id=100875 (accessed 09.10.2024).

6. Architecting for Compliance and Data Residency. Retrieved from: https://medium.com/salesforce-architects/architecting-for-compliance-and-data-residency-13e5d5d9a87f (accessed 09.10.2024).

7. Scale Across Borders: Build a Multi-Region Architecture While Maintaining Data Residency. Retrieved from: https://community.aws/content/2dhVhtsciD5gVBlCKUlHoszrDzU/scale-beyond-borders (accessed 09.10.2024).

8. Managing Data Residency – Concepts and Theory. Retrieved from: https://blog.frankel.ch/data-residency/1/ (accessed 09.10.2024).

9. "How AI is Transforming Cybersecurity" by Gary Eastwood. Retrieved from: https://www.information-age.com/how-ai-is-transforming-cybersecurity-123478294/ (accessed 09.10.2024)

10. Clark, J., Jacob, J. (2018). SHI ta kiberbezpeka: zahrozy ta rishennya. Zhurnal kiberbezpeky, 4 (1), 1–14.

11. Stiopochkina, I. V., Novikov, O. M. (2022). Metody shtuchnoho intelektu v kiberbezpetsi [Methods of artificial intelligence in cyber security]: navch. posib. dlya zdobuvachiv spets. 125 "Kiberbezpeka" / KPI im. Ihorya Sikors'koho; uklad.:. Kyiv: KPI im. Ihorya Sikors'koho, 82 [in Ukrainian].

12. Bohomyia, V. I., Kocheharov, V. S. (2023). Kiberbezpeka v khmarnykh servisasakh za dopomohoyu zastosuvannya kryptohrafichnykh metodiv [Cyber security in cloud services using cryptographic methods]. *Vodnyy transport.* 1(37), 239–246 [in Ukrainian].

13. Bohomya, V. I., Cheremisina, L. O., Yarmolatii, A. V., Baranenko, O. O. (2024). Rezydentsiya danykh suchasnykh biznes platform, yikh system autentyfikatsiyi ta osoblyvosti tekhichnoyi realizatsiyi v holovnii merezhi internet [Data Residence of Modern Business Platforms, Their Authentication Systems and Features of Technical Implementation in the Global Internet]. *Vodnyy transport.* 3 (41), 240–246 [in Ukrainian].

УДК 004.42:005.8
DOI https://doi.org/10.32782/IT/2024-4-5

Володимир БОГОМ'Я
доктор технічних наук, професор, професор кафедри кібербезпеки, інформаційних технологій та економіки, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе, 210, Київ, Україна, 02121
ORCID: *0000-0003-4403-3130*

Валентин ГАЛУНЬКО
доктор філософії з галузі права, доцент кафедри адміністративного права, інтелектуальної власності та цивільно-правових дисциплін, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе, 210, Київ, Україна, 02121
ORCID: *0000-0002-8133-6766*

Бібліографічний опис статті*:* Богом'я, В., Галунько, В. (2024). Правове регулювання кібербезпеки в контексті захисту критичної інфраструктури. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 35–42, doi: https://doi.org/10.32782/IT/2024-4-5

ПРАВОВЕ РЕГУЛЮВАННЯ КІБЕРБЕЗПЕКИ В КОНТЕКСТІ ЗАХИСТУ КРИТИЧНОЇ ІНФРАСТРУКТУРИ

В статті здійснено аналіз дослідження забезпечення кібербезпеки об'єктів критичної інфраструктури (далі – КІ), яке є одним з найважливіших викликів для сучасності. Кібератаки на об'єкти критичної інфраструктури можуть мати руйнівні наслідки для національної безпеки, економіки та суспільства. З огляду на зростання кількості та витонченості кіберзагроз, особливого значення набуває ефективне правове регулювання у цій сфері.

Мета роботи. *Комплексний аналіз національного правового регулювання кібербезпеки з точки зору захисту критичної інфраструктури, виявлено проблеми в чинному законодавстві та надано конкретні рекомендації щодо його вдосконалення з урахуванням сучасних викликів та міжнародного досвіду. Для досягнення поставленої мети було використано наступний комплекс наукових методів Зокрема, аналіз та синтез нормативно-правових актів, проаналізовано тексти законів, стратегій та стандартів у сфері кібербезпеки та захисту інформаційної безпеки.*

Методологія. *Розгляд правового регулювання як складної системи, що складається із взаємопов'язаних елементів методом узагальнення формулюючи загальність висновків на основі аналізу та синтезу. Основні закони, що регулюють питання кібербезпеки КІ в Україні. Особливу увагу приділено аналізу визначень ключових термінів, розподілу повноважень між державними органами, встановленню вимог до кіберзахисту та визначенню відповідальності за порушення у сфері кібербезпеки.*

Наукова новизна. *комплексний аналіз правового регулювання кібербезпеки критичної інфраструктури в Україні з урахуванням, зокрема, поточних викликів, спричинених збройним нападом російської федерації, та міжнародного досвіду, і виявляє низку проблем у чинному законодавстві. Деякі ключові терміни не мають чіткого визначення, що ускладнює правозастосування, що існують проблеми з координацією між різними державними органами, відповідальними за кібербезпеку інформації та комунікацій, а також що існуючі механізми контролю за дотриманням законодавства є неефективними. Підвищення ефективності захисту об'єктів критичної інфраструктури від кіберзагроз, а також розглядають необхідність диференціації регуляторних підходів відповідно до критичності об'єкта.*

Висновки. *Проведене дослідження дозволило зробити низку важливих висновків щодо поточного стану правового регулювання кібербезпеки критичної інфраструктури в Україні та визначити основні напрями його вдосконалення.*

Зокрема, необхідно уточнити термінологію, оптимізувати розподіл повноважень між органами державної влади, посилити вимоги до кібербезпеки та підвищити відповідальність за порушення. Реалізація запропонованих заходів дозволить підвищити рівень захисту критичної інфраструктури України від кіберзагроз.

Ключові слова: *кібербезпека, критична інфраструктура, правове регулювання, кіберзагрози, законодавство, відповідальність, державні органи, міжнародний досвід, кіберзахист.*

Volodymyr BOHOMIA
Doctor of Technical Sciences, Professor, Professor at the Department of Cyber Security, Information Technology and Economics, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121, bog2603@ukr.net
ORCID: *0000-0003-4403-3130*

Valentyn HALUNKO
PhD in law, Associate Professor at the Department of Administrative Law, Intellectual Property and Civil-Law Disciplines, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121, valentinvalentin0987@gmail.com
ORCID: *0000-0002-8133-6766*

To cite this article: Bohomia, V., Halunko, V. (2024). Pravove rehuliuvannia kiberbezpeky v konteksti zakhystu krytychnoi infrastruktury [Legal regulation of cybersecurity in the context of critical infrastructure protection]. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 35–42, doi*:* https://doi.org/10.32782/IT/2024-4-5

LEGAL REGULATION OF CYBERSECURITY IN THE CONTEXT OF CRITICAL INFRASTRUCTURE PROTECTION

The article analyses the study of ensuring cybersecurity of critical infrastructure (hereinafter – CI), which is one of the most important challenges for today. Cyberattacks on critical infrastructure facilities can have devastating consequences for national security, the economy and society. Given the growing number and sophistication of cyber threats, effective legal regulation in this area is of particular importance.

Objective. *To provide a comprehensive analysis of the national legal regulation of cybersecurity in terms of critical infrastructure protection, to identify problems in the current legislation and to provide specific recommendations for its improvement, taking into account current challenges and international experience. To achieve this goal, the following set of scientific methods was used In particular, the analysis and synthesis of legal acts, the texts of laws, strategies and standards in the field of cybersecurity and information security protection were analysed.*

Scientific novelty. *a comprehensive analysis of the legal regulation of critical infrastructure cybersecurity in Ukraine, taking into account, in particular, the current challenges caused by the armed attack of the Russian Federation and international experience, and identifies a number of problems in the current legislation. Some key terms are not clearly defined, which complicates law enforcement, that there are problems with coordination between different government agencies responsible for cybersecurity of information and communications, and that existing mechanisms for monitoring compliance with the law are ineffective. Improving the effectiveness of protecting critical infrastructure against cyber threats, and considering the need to differentiate regulatory approaches according to the criticality of the facility.*

Conclusions. *This study has made it possible to draw a number of important conclusions about the current state of legal regulation of critical infrastructure cybersecurity in Ukraine and to identify the main areas for its improvement.*

In particular, it is necessary to clarify the terminology, optimise the distribution of powers between public authorities, strengthen cybersecurity requirements and increase liability for violations. Implementation of the proposed measures will increase the level of protection of Ukraine's critical infrastructure from cyber threats.

Key words: *critical infrastructure, cyber defence, cyber threats, cybersecurity, government agencies, international experience, legal regulation, legislation, liability.*

Актуальність проблеми. В умовах стрімкого розвитку інформаційних технологій та глобальної цифровізації, кібербезпека набуває стратегічного значення для забезпечення стабільного функціонування та захисту національних інтересів. Кіберпростір становить невід'ємною частиною сучасного світу, забезпечуючи функціонування критично важливих сфер суспільного життя, таких як наприклад енергетика, транспорт, фінанси, телекомунікації, охорона здоров'я тощо. Сукупність цих об'єктів визначається як критична інфраструктура (далі – КІ). Критична інфраструктура є особливо вразливою до кібератак через свою складність, взаємозалежність та потенційно катастрофічні наслідки від збоїв у її роботі. Кібератаки на КІ можуть призвести не лише до задання економічного дисбалансу надаючи значущих збитків та, і порушення надання важливих послуг, і, до загрози життя та здоров'я громадян, тобто дестабілізації ситуації в країні в цілому. Каскадний ефект від ураження одного елемента КІ може призвести до збоїв у роботі інших секторів, що підкреслює необхідність комплексного та надійного захисту в кіберпросторі. Збройна агресія російської федерації проти України у 2022 році суттєво змінила ландшафт кіберзагроз. Кіберпростір став ареною активних

бойових дій, де кібератаки використовуються як інструмент гібридної війни для досягнення військових та політичних цілей. Збільшилась кількість цілеспрямованих та скоординованих атак, спрямованих на дестабілізацію роботи органів виконавчої влади, порушення функціонування об'єктів КІ та поширення дезінформації. Враховуючи зазначені фактори, ефективне правове регулювання кібербезпеки КІ є критично важливим для забезпечення національної безпеки та стабільного функціонування держави в цілому. Необхідно створити чітку та дієву правову базу, яка б визначала обов'язки суб'єктів КІ, повноваження, механізмів тощо.

Аналіз останніх досліджень і публікацій. Визначено та класифіковано у яких присвячені роботи таких українських вчених, як: О. Алексеєва (Алексеєва О.А., 2023), Б. Ворович, С. Єсімов, Х. Кайдрович, Ю. Кемська (Ковалів М., Скриньковський Р., Назар Ю., та ін., 2021), С. Князь, М. Ковалів, І. Красницький, Ю. Назар, В. Пядишев (Пядишев В. Г., 2022), П. Рогов, О. Скіцько, Р. Скриньковський, В. Ткаченко (Рогов П. Д. Ворович Б. О., Ткаченко В. А., 2017), Р. Ширшов (Скіцько О. І., Ширшов Р. А., 2024), і також які досліджували іноземний досвід: Б. Леонов, В. Серьогін (Леонов Б. Д., Шостак Р. М., Серьогін В. С., 2020), Р. Шостак із зарубіжних дослідників можна зазначити Д. Фідлера (Fidler D., 2015) та ін.

Метою статті є аналіз ключових нормативно-правових актів України, що регулюють кібербезпеку критичної інфраструктури, з метою виявлення прогалин та колізій у законодавчому регулюванні та формулювання рекомендацій щодо їх усунення.

Виклад основних положень. Особливого значення захисту критичної інфраструктури набув в 1990-х роках у США відповідь на тероризм і зростаючий зв'язок між критичною інфраструктурою і кібертехнологіями та глобальне поширення інтернету, в той же час які розробляли міжнародне право для регулювання цього процесу. Поширення попередніх комунікаційних технологій підштовхнуло до створення міжнародного права та інституцій щодо посилення захисту КІ. Як наслідок, держави світу почали використовувати міжнародно-правові механізми для зміцнення кіберзахисту своєї національної критичної інфраструктури (Fidler D., 2015).

Виявимо розуміння до понять кібербезпеки в контексті захисту критичної інфраструктури потрібно дамо виділити особливу роль термінувань таких понять як критична інфраструктура, кібербезпека, кібератака, кіберзагроза та кіберінцидент. Так, Закон України від 16.11.2021 року «Про критичну інфраструктуру» визначає дефініцію що, критична інфраструктура – це сукупність об'єктів національної критичної інфраструктури і критичної інформаційної інфраструктури. Що, є об'єктами національної критичної інфраструктури – це об'єкти, системи, служби та послуги, необхідні для підтримання життєдіяльності суспільства та функціонування держави, порушення функціонування яких призведе до негативних наслідків для національної безпеки, економіки та/або соціальної сфери. Критична інформаційна інфраструктура – це сукупність інформаційних, телекомунікаційних та інформаційно-телекомунікаційних систем та електронних комунікаційних мереж, що забезпечують функціонування об'єктів критичної інфраструктури (Про критичну інфраструктуру: Закон України від 16.11.2021 № 1882-IX, 2025). З міжнародних стандартів таких як Директива ЄС 2022/2555 (NIS2) та NIST Cybersecurity Framework (США) охоплюють широкий спектр секторів, включаючи енергетику, транспорт, охорону здоров'я, цифрову інфраструктуру, публічного адміністрування та інші. Визначаються за критеріями розміру та важливості послуг фокусується на системах та активах, порушення яких може мати руйнівні наслідки для національної безпеки, економіки, громадського здоров'я та безпеки (NIST, 2025; NIS 2, 2025).

Кібербезпека є невід'ємною складовою національної безпеки та сталого розвитку інформаційного суспільства. Згідно з Законом України «Про основні засади забезпечення кібербезпеки України», кібербезпека визначається як захищеність життєво важливих інтересів людини і громадянина, суспільства та держави під час використання кіберпростору, за якої забезпечуються сталий розвиток інформаційного суспільства та національна безпека України. Це визначення кореспондується з міжнародними підходами, зокрема зі стандартами ISO/IEC 27000, які визначають інформаційну безпеку як захист конфіденційності, цілісності та доступності інформації (Антимонопольний комітет України. 2021; Про основні засади забезпечення кібербезпеки України : Закон України від 05.10.2017). В цьому контексті критичної інфраструктури кібербезпека набуває особливого значення. Збої в роботі інформаційних систем, що забезпечують функціонування об'єктів КІ, можуть призвести до серйозних наслідків, таких як наприклад: блокування фінансових операцій; порушення функціонування систем охорони здоров'я; інші критичні ситуації, що загрожують життю та здоров'ю громадян, а також національній безпеці.

Отже, кібербезпека в КІ тісно пов'язана з іншими видами безпеки, такими як фізична безпека (захист об'єктів від фізичного доступу), інформаційна безпека (захист інформації від несанкціонованого доступу, використання, розкриття, зміни або знищення) та операційна безпека (забезпечення безпеки операційних процесів). Ефективний захист КІ можливий лише за умови комплексного підходу, який враховує всі ці аспекти.

Кібератака являє собою навмисну дію в кіберпросторі, спрямовану на порушення конфіденційності, цілісності, доступності інформації, що обробляється в інформаційно-комунікаційних системах, або на порушення безпеки, сталого та штатного режиму функціонування цих систем. Кіберзагроза, згідно з Законом України «Про основні засади забезпечення кібербезпеки України», визначається як наявні та потенційні явища і чинники, що створюють небезпеку життєво важливим інтересам людини і громадянина, суспільства та держави під час використання кіберпростору. Це визначення охоплює широкий спектр потенційних негативних впливів, від технічних вразливостей до навмисних дій зловмисників. Міжнародні стандарти також приділяють увагу цьому поняттю. Зокрема, ISO/IEC 27005:2022 (Information security, cybersecurity and privacy protection – Guidance on managing information security risks) визначає загрозу як «потенційну причину небажаної події, яка може призвести до шкоди системам або організації», підкреслюючи ймовірнісний характер загрози та її потенційні наслідки (Будстандарт online, 2025). Відповідно до Закону України «Про основні засади забезпечення кібербезпеки України» кіберінцидент визначається як порушення нормального функціонування інформаційно-телекомунікаційних систем, електронних комунікаційних мереж та об'єктів критичної інформаційної інфраструктури або доступності, цілісності, конфіденційності та автентичності електронних інформаційних ресурсів, що обробляються в них. Визначається як подія або серія кіберподій, що призвела або може призвести до порушення. Це визначення охоплює широкий спектр подій, від індивідуальних спроб несанкціонованого доступу до складних організованих атак, і відповідає міжнародним стандартам, хоча й має певні відмінності (Про основні засади забезпечення кібербезпеки України: Закон України від 05.10.2017).

Розглянемо детальніше основні положення вищезазначених нормативно-правових актів: Закон України «Про основні засади забезпечення кібербезпеки України»: визначає кібербезпеку як захищеність людини і громадянина під час використання кіберпростору. Визначаючи її як захист життєво важливих інтересів суспільства та держави. Встановлює основні принципи державної політики у сфері кібербезпеки: законність, пріоритетність захисту прав і свобод людини і громадянина, комплексність, безперервність, пропорційність та міжнародне співробітництво. Суб'єкти забезпечення кібербезпеки є: Президент України, Верховна Рада України, Кабінет Міністрів України, Рада національної безпеки і оборони України, Служба безпеки України, ДССЗЗІ, НКРЗІ, інші державні органи, органи місцевого самоврядування, підприємства тощо. ДССЗЗІУ відповідає за три основні сфери кібербезпеки: запобігання кіберзагрозам, виявлення та нейтралізація кіберзагроз, реагування на кіберінциденти та відновлення об'єктів кіберзахисту (Державна служба спеціального зв'язку та захисту інформації України, 2025). Важливо, що він визначає мандат Національного координаційного центру кібербезпеки як головного органу, що координує діяльність національної системи кібербезпеки.

Аналіз закону показує, що він створює загальні рамки для регулювання кібербезпеки, але не містить детальних вимог щодо захисту інформації та комунікацій.

Закон України «Про критичну інфраструктуру» визначає «критичну інфраструктуру» як сукупність об'єктів, систем, служб і споруд, необхідних для забезпечення життєдіяльності суспільства і виконання державою своїх функцій (Про критичну інфраструктуру : Закон України від 16.11.2021). Він визначає критерії віднесення об'єктів до критичної інфраструктури з урахуванням їх важливості для національної безпеки, економіки та соціальної сфери. Енергетика, транспорт, інформаційно-комунікаційні технології, банківська справа та фінанси, охорона здоров'я, водопостачання та водовідведення, питне водопостачання, цивільний захист, хімічна промисловість, космічна діяльність, електронні комунікації, поштовий зв'язок та паливно-енергетичний комплекс. Закон також визначає обов'язки розвідувальних органів щодо забезпечення кібербезпеки, включаючи проведення оцінки ризиків, здійснення заходів кіберзахисту, реагування на кіберінциденти та інформування компетентних органів про кіберінциденти. Передбачається створення Національної системи захисту розвідувальних органів, яка включає розвідувальні органи, державні установи та інші організації.

Отже, закон встановлює конкретні вимоги до захисту інформаційної інфраструктури та відіграє важливу роль у регулюванні кібербезпеки інформаційної інфраструктури.

Стратегія кібербезпеки України (Про затвердження Порядку проведення огляду стану кіберзахисту критичної інформаційної інфраструктури, державних інформаційних ресурсів та інформації, вимога щодо захисту якої встановлена законом : Постанова Кабінету Міністрів України; Порядок від 11.11.2020): визначає стратегічні цілі та завдання держави у сфері кібербезпеки, включаючи забезпечення захисту національних інтересів у кіберпросторі, розвиток національної системи кібербезпеки, протидію кіберзлочинності та міжнародне співробітництво. Стратегія визначає пріоритетні напрями державної політики у сфері кібербезпеки, в тому числі у сфері захисту інформаційно-комунікаційних технологій (КІ). Стратегія є важливим документом стратегічного планування, який визначає загальний напрям розвитку системи кібербезпеки України.

Постанови Кабінету Міністрів України: розвивають положення закону та визначають конкретні механізми реалізації державної політики у сфері кібербезпеки інформаційно-комунікаційної інфраструктури (Про затвердження плану заходів на 2023-2024 роки з реалізації Стратегії кібербезпеки України : Розпорядження Кабінету Міністрів України; План, Заходи від 19.12.2023).

Наприклад: Постанова Кабінету Міністрів України від 19 червня 2019 року № 512 «Про затвердження Порядку формування переліку об'єктів критичної інфраструктури» (Про затвердження Загальних вимог до кіберзахисту об'єктів критичної інфраструктури : Постанова Кабінету Міністрів України; Вимоги, Перелік від 19.06.2019).

В контексті захисту критичної інфраструктури вчена Алексєєва О.А. комплексно підійшла до правового забезпечення кібербезпеки об'єктів критичної інфраструктури, враховуючи як національне законодавство, так і зарубіжний досвід, зокрема, США та ЄС. Акцентується увага на актуальності питання в умовах кібервійни з росією та необхідності вдосконалення нормативно-правової бази. Важливим аспектом є аналіз законопроекту № 8087, який спрямований на посилення кіберзахисту, але автор вказує на необхідність внесення змін відповідно до європейських стандартів та інтересів бізнесу. Авторка обґрунтовано зазначає питання відсутності узгодженої державної політики та недостатньої міжвідомчої координації у сфері захисту ІКТ та наголошує на важливості державно-приватного співробітництва. Водночас наголошується на необхідності адаптації іноземного досвіду до українських реалій та врахуванні історичних особливостей управління технічною безпекою. Загалом у дослідженні представлено системний підхід до аналізу проблеми, визначено ключові виклики та запропоновано конкретні напрями вдосконалення системи кіберзахисту КІ України (Алексеєва О. А., 2023; Про внесення змін до деяких законів України щодо невідкладних заходів посилення спроможностей із кіберзахисту державних інформаційних ресурсів та об'єктів критичної інформаційної інфраструктури: проєкт закону України).

Можна погодитися з вченими Скіцьком О.І. та Ширшовим Р.А. автори справедливо підкреслюють, що проблема кіберзахисту об'єктів критичної інфраструктури в Україні значно загострилася в умовах тотальної агресії російської федерації. Зокрема, підкреслюється, що об'єкти критичної інфраструктури, такі як енергетичні, телекомунікаційні, медійні та фінансові компанії, є основними цілями кібератак, що підтверджує їх стратегічну важливість в умовах військових дій. Автори обґрунтовано доводять, що з огляду на мінливу динаміку нормативно-правової бази необхідний комплексний підхід до управління кібербезпекою інформаційно-комунікаційних об'єктів. Запропонований комплексний підхід, спрямований на зменшення вразливостей, пом'якшення наслідків інцидентів та ефективну протидію кіберзагрозам, є логічним і відповідає сучасним вимогам у сфері кібербезпеки. Так, автори чітко окреслюють ключові виклики та пропонують практичні шляхи їх подолання. Це важливо для забезпечення стійкості критичної інфраструктури України в умовах гібридної війни (Скіцько О. І., Ширшов Р. А., 2024).

В умовах глобалізації та транснаціонального характеру кіберзагроз вивчення міжнародного досвіду та активна участь у міжнародному співробітництві мають важливе значення для забезпечення ефективної кібербезпеки КІ. Директива (ЄС) 2022/2555 (NIS2) (6) є ключовим законодавчим актом ЄС у сфері кібербезпеки. Вона замінила попередню Директиву NIS та значно розширила її сферу дії, охоплюючи більшу кількість секторів та суб'єктів. Основні положення NIS2: розширена сфера застосування: охоплює такі критичні сектори, як енергетика, транспорт, охорона здоров'я, цифрова інфраструктура, державне управління, космос, виробництво та харчова промисловість; диференційований підхід: різні вимоги до різних суб'єктів господарювання, залежно від розміру суб'єкта господарювання та важливості його послуг. Розрізняють «основні» та «критичні» організації; посилені вимоги до управління ризиками: вимагає впровадження комплексних заходів з управління ризиками кібербезпеки, включаючи оцінку

ризиків, впровадження технічних та організаційних заходів безпеки та реагування на інциденти; повідомлення про інциденти: встановлює зобов'язання повідомляти про кіберінциденти, що дозволяє швидко реагувати на загрози та обмінюватися інформацією; нагляд та санкції. Передбачає механізми нагляду та санкції за недотримання Директиви NIS2 є важливим кроком у гармонізації законодавства ЄС у сфері кібербезпеки та сприятиме підвищенню рівня інформаційної безпеки в Європі.

Таким чином, Директива (ЄС) 2022/2555 (NIS2) є ключовим інструментом регулювання кібербезпеки в ЄС та значно розширює сферу застосування до секторів та суб'єктів у порівнянні з попередньою версією. Основними аспектами NIS2 є розширення сфери застосування на ключові сектори економіки, диференційований підхід до вимог відповідно до розміру та важливості суб'єкта.

NIST CSF – це стандарт, розроблений Національним інститутом стандартів і технологій США (NIST), який надає організаціям структурований підхід до управління ризиками кібербезпеки, включаючи захист КІ (NIST CSF 2.0, 2024).

Держави-члени НАТО вважають кібербезпеку важливим елементом колективної оборони і приділяють їй значну увагу. НАТО розробила низку документів та ініціатив у цій сфері. План дій з кіберзахисту: визначає конкретні заходи і завдання з посилення кіберсил і засобів Альянсу. Центр передового досвіду з кіберзахисту (CCDCOE): проводить дослідження, навчання та обмін досвідом у сфері кіберзахисту; країни НАТО активно співпрацюють у сфері кібербезпеки, обмінюючись інформацією про загрози та проводячи спільні навчання (About us, 2025).

Українське законодавство має певні спільні риси з міжнародними стандартами, але є й відмінності. Акцент на управлінні ризиками, вимоги до звітності про інциденти. Недостатня деталізація вимог до кіберзахисту для різних секторів інформаційної безпеки, відсутність диференційованого підходу до різних типів суб'єктів. Для вдосконалення українського законодавства доцільно врахувати кращі практики ЄС та США. Зокрема: запровадження диференційованого підходу до регулювання кібербезпеки КІ відповідно до рівня важливості об'єкта регулювання. Запровадити диференційований підхід до регулювання кібербезпеки КІ залежно від рівня важливості об'єкта. Посилити вимоги до інформування про інциденти та обміну інформацією. Посилення міжнародного співробітництва у сфері кібербезпеки. Україна бере активну участь у міжнародному співробітництві у сфері кібербезпеки та співпрацює з такими міжнародними організаціями, як ООН, ОБСЄ, Рада Європи, НАТО та ЄС. Розробка та імплементація міжнародних стандартів і норм у сфері кібербезпеки.

Висновки. По-перше, в умовах глобальної діджиталізації та зростання кіберзагроз, особливо в умовах збройної агресії російської федерації, захист критичної інфраструктури має стратегічне значення для національної безпеки та стабільного функціонування держави. Ефективне правове регулювання є ключовим елементом у забезпеченні такого захисту.

Аналіз чинного законодавства України свідчить про наявність базової правової бази. Водночас були виявлені прогалини та суперечності, які потребують вирішення. Серед них відсутність чіткого диференційованого підходу до різних типів організацій КІ, а також необхідність посилення міжвідомчої координації та обміну інформацією щодо кіберінцидентів.

Вивчення міжнародного досвіду може допомогти визначити найкращі практики та сфери для вдосконалення українського законодавства. Диференціювати підхід до регулювання кібербезпеки відповідно до критичності об'єкта.

ЛІТЕРАТУРА:

1. Алексеєва О. А. Правове забезпечення кібербезпеки об'єктів критичної інфраструктури. *Інформація і право*. 2023. № 4 (47). С. 168–176. DOI: https://doi.org/10.37750/2616-6798.2023.4(47).291633.

2. АМКУ долучається до автоматизованого обміну індикаторами кіберзагроз та інформацією про кіберінциденти. *Антимонопольний комітет України*. 2021. URL: https://amcu.gov.ua/news/amku-doluchayetsya-do-avtomatizovanogo-obminu-indikatorami-kiberzagroz-ta-informaciyeyu-pro-kiberincidenti.

3. Діяльність Адміністрації Держспецзв'язку у сфері кіберзахисту. *Державна служба спеціального зв'язку та захисту інформації України*. 2025. URL: https://cip.gov.ua/ua/statics/cyber-protection.

4. ДСТУ ISO/IEC 27005:2023 Інформаційна безпека, кібербезпека та захист конфіденційності. Настанова керування ризиками інформаційної безпеки (ISO/IEC 27005:2022, IDT). *Будстандарт online*. 2025. URL: https://online.budstandart.com/ua/catalog/doc-page.html?id_doc=104400.

5. Ковалів М., Скриньковський Р., Назар Ю., Єсімов С., Красницький І., Кайдрович Х., Князь С., Кемська Ю. Правове забезпечення кібербезпеки критичної інформаційної інфраструктури України. *Traektoriâ Nauki = Path of Science*. 2021. No 4, Vol. 7. С. 2011–2018. DOI: https://10.22178/pos.69-12

6. Комісія УСПП з питань науки та IT зробила переклад Директиви NIS2. *I-UA.tv*. 2023. URL: https://i-ua.tv/tech/82418-komisiia-uspp-z-pytan-nauky-ta-it-zrobyla-pereklad-dyrektyvy-nis2.

7. Леонов Б. Д., Шостак Р. М., Серьогін В. С. Розвиток методичного забезпечення антитерористичної захищеності об'єктів критичної інфраструктури (на прикладі США). *Інформація і право*. 2020. № 3 (34). С. 88–95. URL: https://ippi.org.ua/sites/default/files/12_18.pdf.

8. Про внесення змін до деяких законів України щодо невідкладних заходів посилення спроможностей із кіберзахисту державних інформаційних ресурсів та об'єктів критичної інформаційної інфраструктури: проєкт закону України (реєстр. № 8087 від 29.09.22 р.). 2025. URL: https://itd.rada.gov.ua/billInfo/Bills/pubFile/1490881

9. Про затвердження Загальних вимог до кіберзахисту об'єктів критичної інфраструктури : Постанова Кабінету Міністрів України; Вимоги, Перелік від 19.06.2019 № 518. База даних «Законодавство України». Верховна Рада України. 2025. URL: https://zakon.rada.gov.ua/go/518-2019-%D0%BF.

10. Про затвердження плану заходів на 2023-2024 роки з реалізації Стратегії кібербезпеки України : Розпорядження Кабінету Міністрів України; План, Заходи від 19.12.2023 № 1163-р. База даних «Законодавство України». Верховна Рада України. 2025. URL: https://zakon.rada.gov.ua/go/1163-2023-%D1%80.

11. Про затвердження Порядку проведення огляду стану кіберзахисту критичної інформаційної інфраструктури, державних інформаційних ресурсів та інформації, вимога щодо захисту якої встановлена законом : Постанова Кабінету Міністрів України; Порядок від 11.11.2020 № 1176. База даних «Законодавство України». Верховна Рада України. 2025. URL: https://zakon.rada.gov.ua/go/1176-2020-%D0%BF.

12. Про критичну інфраструктуру : Закон України від 16.11.2021 № 1882-IX. База даних «Законодавство України». Верховна Рада України. 2025. URL: https://zakon.rada.gov.ua/go/1882-20.

13. Про основні засади забезпечення кібербезпеки України : Закон України від 05.10.2017 № 2163-VIII. База даних «Законодавство України». Верховна Рада України. 2025. URL: https://zakon.rada.gov.ua/go/2163-19.

14. Пядишев В. Г. Кібербезпека критичних інфраструктур: закордонний досвід та українські реалії. *Південноукраїнський правничий часопис*. 2022. № 4. Ч. 3. С. 229–234. URL: http://www.sulj. oduvs.od.ua/archive/2022/4/part_3/38.pdf.

15. Рогов П. Д. Ворович Б. О., Ткаченко В. А. Шляхи Забезпечення кібернетичної безпеки об'єктів критичної інформаційної інфраструктури держави у воєнній сфері. *Збірник наукових праць Центру воєнно-стратегічних досліджень Національного університету оборони України імені Івана Черняховського*. 2017. № 1. С. 64–72. URL: http:// nbuv.gov.ua/UJRN/Znpcvsd_2017 _1_13.

16. Скіцько О. І., Ширшов Р. А. Нормативно-правове забезпечення кібербезпеки об'єктів критичної інфраструктури. *Науковий вісник Львівського державного університету внутрішніх справ. Серія юридична*. 2024. № 2. С. 73–79. DOI: https://doi.org/10.32782/2311-8040/2024-2-11.

17. About us. *The NATO Cooperative Cyber Defence Centre of Excellence*. 2025. URL: https://ccdcoe.org/about-us/.

18. Cybersecurity Framework. NIST. 2025. URL: https://www.nist.gov/cyberframework.

19. Fidler D. Whither the Web?: International Law, Cybersecurity, and Critical Infrastructure Protection. *16 Georgetown Journal of International Affairs 8*. 2015. P. 8–20. URL: https://www.repository.law.indiana.edu/facpub/2452.

20. NIS 2: Огляд Директиви ЄС про кібербезпеку. *GigaCloud*. 2025. URL: https://gigacloud.ua/blog/navchannja/nis-2-ogljad-direktivi-es-pro-kiberbezpeku.

21. NIST CSF 2.0 шість функцій кібербезпеки. *Itspecialist*. 2024. URL: https://my-itspecialist.com/nist-csf-2.0-six-cybersecurity-functions.

REFERENCES:

1. Alexeyeva, O. A. (2023). Pravove zabezpechennia kiberbezpeky obiektiv krytychnoi infrastruktury [Legal support of cybersecurity for critical infrastructure objects]. *Information and Law*, 4(47), 168–176. https://doi.org/10.37750/2616-6798.2023.4(47).291633 [in Ukrainian].

2. Antimonopoly Committee of Ukraine. (2021). AMKU doluchaietsia do avtomatyzovanoho obminu indykatoramy kiberzahroz ta informatsiieiu pro kiberintsydenty [AMCU joins the automated exchange of cyber threat indicators and cyber incident information]. Retrieved from: https://amcu.gov.ua/news/amku-doluchayetsya-do-avtomatizovanogo-obminu-indikatorami-kiberzagroz-ta-informaciyeyu-pro-kiberincidenti [in Ukrainian].

3. Diialnist Administratsii Derzhspetszviazku u sferi kiberzakhystu [Activities of the State Special Communications Service of Ukraine in the field of cyber defense]. (2025). *State Service of Special*

Communications and Information Protection of Ukraine. Retrieved from: https://cip.gov.ua/ua/statics/cyber-protection [in Ukrainian].

4. Budstandard Online. (2025). DSTU ISO/IEC 27005:2023 *Information security, cybersecurity, and privacy protection. Guidelines for information security risk management (ISO/IEC 27005:2022, IDT).* Retrieved from: https://online.budstandart.com/ua/catalog/doc-page.html?id_doc=104400.

5. Kovaliv, M., Skrynkovskyi, R., Nazar, Y., Yesimov, S., Krasnytskyi, I., Kaidrovych, K., Kniaz, S., & Kemska, Y. (2021). Pravove zabezpechennia kiberbezpeky krytychnoi informatsiinoi infrastruktury Ukrainy [Legal support of cybersecurity for critical information infrastructure of Ukraine]. Traektoriâ Nauki = Path of Science, 7 (4), 2011–2018. https://doi.org/10.22178/pos.69-12 [in Ukrainian].

6. I-UA.tv. (2023). *The USPP Commission on Science and IT has translated the NIS2 Directive*. Retrieved from: https://i-ua.tv/tech/82418-komisiia-uspp-z-pytan-nauky-ta-it-zrobyla-pereklad-dyrektyvy-nis2 [in Ukrainian].

7. Leonov, B. D., Shostak, R. M., & Seryogin, V. S. (2020). Rozvytok metodychnoho zabezpechennia antyterorystychnoi zakhyshchenosti obiektiv krytychnoi infrastruktury (na prykladi SShA). [Development of methodological support for anti-terrorist protection of critical infrastructure objects (on the example of the USA)]. *Information and Law*, 3(34), 88–95. Retrieved from: https://ippi.org.ua/sites/default/files/12_18.pdf [in Ukrainian].

8. Verkhovna Rada of Ukraine. (2025). *On amendments to some laws of Ukraine regarding urgent measures to strengthen the capabilities of cybersecurity of state information resources and critical information infrastructure objects: Draft Law of Ukraine (Reg. No. 8087 of 29.09.22).* Retrieved from: https://itd.rada.gov.ua/billInfo/Bills/pubFile/1490881 [in Ukrainian].

9. Cabinet of Ministers of Ukraine. (2019). *On approval of the General requirements for cybersecurity of critical infrastructure objects: Resolution No. 518 of 19.06.2019*. Retrieved from: https://zakon.rada.gov.ua/go/518-2019-%D0%BF [in Ukrainian].

10. Cabinet of Ministers of Ukraine. (2023). *On approval of the action plan for 2023-2024 for the implementation of the Cybersecurity Strategy of Ukraine: Order No. 1163-r of 19.12.2023*. Retrieved from: https://zakon.rada.gov.ua/go/1163-2023-%D1%80 [in Ukrainian].

11. Cabinet of Ministers of Ukraine. (2020). *On approval of the Procedure for conducting a review of the state of cybersecurity of critical information infrastructure, state information resources, and information, the protection of which is required by law: Resolution No. 1176 of 11.11.2020*. Retrieved from: https://zakon.rada.gov.ua/go/1176-2020-%D0%BF [in Ukrainian].

12. Verkhovna Rada of Ukraine. (2021). *On critical infrastructure: Law of Ukraine No. 1882-IX of 16.11.2021*. Retrieved from: https://zakon.rada.gov.ua/go/1882-20 [in Ukrainian].

13. Verkhovna Rada of Ukraine. (2017). *On the basic principles of ensuring cybersecurity of Ukraine: Law of Ukraine No. 2163-VIII of 05.10.2017*. Retrieved from: https://zakon.rada.gov.ua/go/2163-19 [in Ukrainian].

14. Pyadyshev, V. G. (2022). Kiberbezpeka krytychnykh infrastruktur: zakordonnyi dosvid ta ukrainski realii [Cybersecurity of critical infrastructures: Foreign experience and Ukrainian realities]. *Pivdennoukrainskyi Pravnychyi Chasopys*, 4(3), 229–234. Retrieved from: http://www.sulj.oduvs.od.ua/archive/2022/4/part_3/38.pdf [in Ukrainian].

15. Rohov, P. D., Vorovych, B. O., & Tkachenko, V. A. (2017). Shliakhy Zabezpechennia kibernetychnoi bezpeky obiektiv krytychnoi informatsiinoi infrastruktury derzhavy u voiennii sferi [Ways to ensure cyber security of critical information infrastructure objects of the state in the military sphere]. *Collection of Scientific Works of the Center for Military-Strategic Studies of the National Defense University of Ukraine named after Ivan Chernyakhovskyi*, 1, 64–72. Retrieved from: http://nbuv.gov.ua/UJRN/Znpcvsd_2017_1_13 [in Ukrainian].

16. Skitsko, O. I., & Shirshov, R. A. (2024). Normatyvno-pravove zabezpechennia kiberbezpeky obiektiv krytychnoi infrastruktury [Regulatory and legal support of cybersecurity for critical infrastructure objects]. *Scientific Bulletin of Lviv State University of Internal Affairs. Legal Series*, 2, 73–79. https://doi.org/10.32782/2311-8040/2024-2-11 [in Ukrainian].

17. NATO Cooperative Cyber Defence Centre of Excellence. (2025). *About us*. Retrieved from: https://ccdcoe.org/about-us/.

18. NIST. (2025). *Cybersecurity Framework*. Retrieved from: https://www.nist.gov/cyberframework.

19. Fidler, D. (2015). Whither the Web?: International Law, Cybersecurity, and Critical Infrastructure Protection. *Georgetown Journal of International Affairs*, 16, 8–20. Retrieved from: https://www.repository.law.indiana.edu/facpub/2452.

20. GigaCloud. (2025). *NIS 2: Overview of the EU Cybersecurity Directive*. Retrieved from: https://gigacloud.ua/blog/navchannja/nis-2-ogljad-direktivi-es-pro-kiberbezpeku.

21. Itspecialist. (2024). *NIST CSF 2.0: Six cybersecurity functions*. Retrieved from: https://my-itspecialist.com/nist-csf-2.0-six-cybersecurity-functions.

УДК 004.056.5
DOI https://doi.org/10.32782/IT/2024-4-6

Андрій БОРИСЕНКО
доктор філософії, молодший науковий співробітник, Інститут транспортних систем і технологій НАН України, вул. Писаржевского, 5, м. Дніпро, Україна, 49005
ORCID: *0009-0001-5083-5255*
Scopus Author ID: *58158469800*

Бібліографічний опис статті: Борисенко, А. (2024). Безпечне програмування: автоматизовані інструменти для виявлення вразливостей в коді. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 43–52, doi: https://doi.org/10.32782/IT/2024-4-6

БЕЗПЕЧНЕ ПРОГРАМУВАННЯ: АВТОМАТИЗОВАНІ ІНСТРУМЕНТИ ДЛЯ ВИЯВЛЕННЯ ВРАЗЛИВОСТЕЙ В КОДІ

З кожним роком виявляється все більше вразливостей програмного забезпечення та вони стають все більш шкідливими для організацій у всьому світі, впливаючи на їхні фінанси, діяльність і репутацію.

Мета роботи. *Вразливості можуть становити серйозний ризик та призвести до зламу системи, витоку інформації або відмови в обслуговуванні. Ця проблема погіршується через нестачу фахівців з кібербезпеки. Автоматизовані інструменти та технології можуть оптимізувати процес виявлення вразливостей, зробивши його ефективнішим. Метою роботи є аналіз ефективності існуючих рішень для виявлення різних типів вразливостей на різних етапах розробки програмного забезпечення, а також визначення переваг і недоліків кожного з інструментів. Робота спрямована на підвищення точності та ефективності виявлення уразливостей в реальних умовах програмування.*

Методологія. *Досліджено наявні автоматизовані інструменти для виявлення вразливостей у коді в контексті безпечного програмування. Проведено порівняльний аналіз популярних рішень, таких як статичний аналізатор коду, динамічний аналіз та інструменти для аналізу залежностей. Ретельно вивчено принципи роботи, функціональні можливості та особливості кожного інструменту. Виокремлено основні переваги та обмеження існуючих рішень, зокрема щодо їх інтеграції у процеси, швидкості виявлення помилок та точності результатів. На основі результатів проведеного аналізу надано рекомендації щодо оптимального вибору та налаштування інструментів для виявлення, а також пропозиції щодо їх вдосконалення для забезпечення більш ефективного виявлення та усунення уразливостей.*

Наукова новизна. *Великі мовні моделі та інші інструменти штучного інтелекту швидко привертають увагу та застосовуються в усьому світі. Наукова новизна цієї статті полягає в їх комплексному аналізі. У статті вперше проведено детальне порівняння різних типів інструментів для виявлення вразливостей, зокрема з акцентом на новітні рішення, які забезпечують більш точне виявлення вразливостей у сучасних технологіях програмування. Визначено актуальні проблеми, що виникають при використанні автоматизованих інструментів для виявлення вразливостей, зокрема, питання фальшивих спрацьовувань та інтеграційні складнощі, що не завжди висвітлюються в існуючих дослідженнях. Це дозволяє більш глибоко зрозуміти обмеження існуючих рішень і визначити напрямки їх вдосконалення.*

Висновки. *Результати цієї роботи демонструють значний потенціал використання машинного навчання для виявлення вразливостей програмного забезпечення безпосередньо на рівні вихідного коду. Створений набір даних, а також застосування методів машинного навчання, дозволили досягти високої точності в класифікації потенційних вразливостей. Було доведено, що можна значно покращити ефективність аналізу, скорочуючи час, необхідний для виявлення вразливостей, і дозволяючи швидше фокусуватися на потенційних проблемах.*

Ключові слова: *безпека систем, кібербезпека, вразливості, програмне забезпечення з відкритим кодом, автоматизація, тестування безпеки, методологія оцінки.*

Andrii BORYSENKO
Doctor of Philosophy, Junior Research Fellow, Institute of Transport Systems and Technologies of Ukrainian National Academy of Science, 5, Pisarzhevskogo Str., Dnipro, Ukraine, 49005
ORCID: *0009-0001-5083-5255*
Scopus Author ID: *58158469800*

To cite this article: Borysenko, A. (2024). Bezpechne prohramuvannia: avtomatyzovani instrumenty dlia vyiavlennia vrazlyvostei v kodi [Secure programming: automated tools for detecting vulnerabilities in code]. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 4, 43–52, doi: https://doi.org/10.32782/IT/2024-4-6

SECURE PROGRAMMING: AUTOMATED TOOLS FOR DETECTING VULNERABILITIES IN CODE

The aim. *Vulnerabilities can pose a serious risk and lead to system breaches, information leaks or denial of service. This problem is exacerbated by the shortage of cybersecurity professionals. Automated tools and technologies can optimize the vulnerability detection process, making it more efficient. The purpose of the work is to analyze the effectiveness of existing solutions for detecting different types of vulnerabilities at different stages of software development, as well as to identify the advantages and disadvantages of each of the tools. The work is aimed at improving the accuracy and efficiency of vulnerability detection in real programming conditions.*

The methodology. *The existing automated tools for detecting vulnerabilities in code in the context of secure programming are studied. A comparative analysis of popular solutions, such as static code analyzers, dynamic analysis and dependency analysis tools, is conducted. The principles of operation, functionality and features of each tool are carefully studied. The main advantages and limitations of existing solutions are highlighted, in particular regarding their integration into processes, speed of error detection and accuracy of results. Based on the results of the analysis, recommendations are provided for the optimal selection and configuration of detection tools, as well as suggestions for their improvement to ensure more effective detection and elimination of vulnerabilities.*

Scientific novelty. *Large language models and other artificial intelligence tools are quickly attracting attention and are used worldwide. The scientific novelty of this article lies in their comprehensive analysis. The article provides the first detailed comparison of different types of vulnerability detection tools, in particular with an emphasis on the latest solutions that provide more accurate detection of vulnerabilities in modern programming technologies. Current problems that arise when using automated vulnerability detection tools are identified, in particular, the issue of false positives and integration difficulties that are not always covered in existing studies. This allows for a deeper understanding of the limitations of existing solutions and to identify areas for their improvement.*

Conclusions. *The results of this work demonstrate the significant potential of using machine learning to detect software vulnerabilities directly at the source code level. The dataset we created, as well as the application of machine learning methods, allowed us to achieve high accuracy in classifying potential vulnerabilities. It was proven that the efficiency of the analysis can be significantly improved, reducing the time required to detect vulnerabilities and allowing us to focus on potential problems faster.*

Key words: *systems security, cybersecurity, vulnerabilities, open-source software, automation, security testing, assessment methodology.*

Постановка проблеми. Події, пов'язані з кібербезпекою, часто висвітлюються в ЗМІ. Протягом багатьох років спостерігається помітне зростання глобальних кібератак. Дослідження, проведене компанією IBM (International Business Machines – один з найбільших в усьому світі виробників та постачальників апаратного та програмного забезпечення, а також IT-сервісів та консалтингових послуг) показало декілька цікавих ключових висновків. У 2023 році середня вартість одного порушення становила для організацій 4,45 мільйона доларів США. Це на 2,3% більше порівняно з 4,35 мільйона минулого року (Booth, 2023). Майбутні прогнози, отримані в іншому дослідженні, вказують на різке збільшення загальних глобальних витрат, пов'язаних з інцидентами кібербезпеки, за оцінками, глобальні витрати досягнуть 10,5 трильйонів доларів США щорічно до 2025 року. Це становить приблизно 28,8 мільярдів доларів США на добу, або 333 тисячі доларів США щосекунди (Check Point Blog, 2022). Завдання надання організаціям досвіду з кібербезпеки виходить за рамки простого пошуку спеціалістів із безпеки – це також передбачає пошук професіоналів, які мають необхідний досвід. Непривілейовані країни особливо сприйнятливі до викликів через недостатню інфраструктуру кібербезпеки, відсутність міжвідомчої координації та реагування на надзвичайні ситуації, обмежені навички використання інформаційно-комунікаційних технологій (ІКТ) і неадекватний захист критичної національної інфраструктури.

У боротьбі з вищезазначеними проблемами може допомогти автоматизація перевірок безпеки. Автоматизовані технології пропонують методичний підхід до цих перевірок і усувають потреби в підвищенні знань. Дуже важливо, щоб процес автоматизації призначав різні рівні серйозності для різних вразливостей безпеки. Характеристика вразливостей програмного забезпечення є критично важливим кроком у виявленні першопричин вразливості, розуміння її наслідків, механізмів атак і відповідних методів пом'якшення. Національний інститут стандартів та технологій США (НІСТ) розробив стандартизовану онтологію опису вразливостей (ООВ), яка визначає атрибути опису, необхідні для ефективного надання оперативної інформації розробникам програмного забезпечення з метою усунення/зменшення вразливості.

Рисунок 1 (Booth, 2016) демонструє різні атрибути онтології ООВ, що використовуються для характеристики вразливостей програмного забезпечення. Наприклад, категорія

характеристики методу впливу представляє, як можна використати вразливість. Характеристики в категорії відображають конкретні методи, які зловмисник може використати, щоб скористатися вразливістю: обхід автентифікації, збій довіри, вихід із контексту, атака «людина посередині» та виконання коду. Інші характеристики вказують наслідки, домен, місце та засоби пом'якшення використання вразливостей. Зацікавлені сторони та постраждалі користувачі вразливого продукту покладаються на характерні дані, щоб визначити, як експлойт впливає на їхні системи та як запобігти експлуатації.

На жаль, визначення характеристик уразливості, як описано в ООВ, є ручним, трудомістким і асинхронним процесом.

Правильна характеристика вразливості вимагає перегляду її описів з достатнім досвідом безпеки та знайомства з онтологією. Цей інтенсивний процес призвів до проблем із якістю звітів про вразливості. На рисунку 2 показано графічне представлення зв'язку між скануванням та ідентифікацію всіх пристроїв, оцінкою вразливостей та генеруванням детальних звітів, що підсумовують результати. Дослідження виявили, що звіти про вразливості НІСТ часто залишаються неповними або не мають конкретних характеристик.

Аналіз останніх досліджень і публікацій. Останні дослідження в галузі безпечного програмування та автоматизованих інструментів для виявлення вразливостей в коді показують значний прогрес у розробці інструментів, які не тільки покращують ефективність виявлення вразливостей, але й інтегруються в сучасні методи розробки програмного забезпечення, забезпечуючи більш високий рівень автоматизації і точності. Зокрема, наукові публікації останніх років зосереджуються на кількох ключових напрямках, зокрема на застосуванні машинного навчання, статичному та динамічному аналізі коду, а також на інтеграції інструментів у CI/CD процеси (процеси неперервної інтеграції та доставки).

Використання машинного навчання в контексті виявлення вразливостей набуває популярності. Дослідження, показують, що машинне навчання може значно підвищити точність автоматизованих інструментів для виявлення вразливостей, здатних адаптуватися до нових типів загроз. Алгоритми на основі нейронних мереж та класифікації, навчання з підкріпленням та кластеризації дозволяють значно поліпшити прогнозування потенційних точок входу для атак. Вони здатні виявляти не лише відомі, але й нові, невідомі вразливості, які раніше могли залишатися поза увагою традиційних інструментів. Особливо популярними є методи, які застосовуються до статичного аналізу: створення моделей на основі вихідного коду, які можуть передбачити потенційні вразливості, аналізуючи синтаксис і логіку програм.

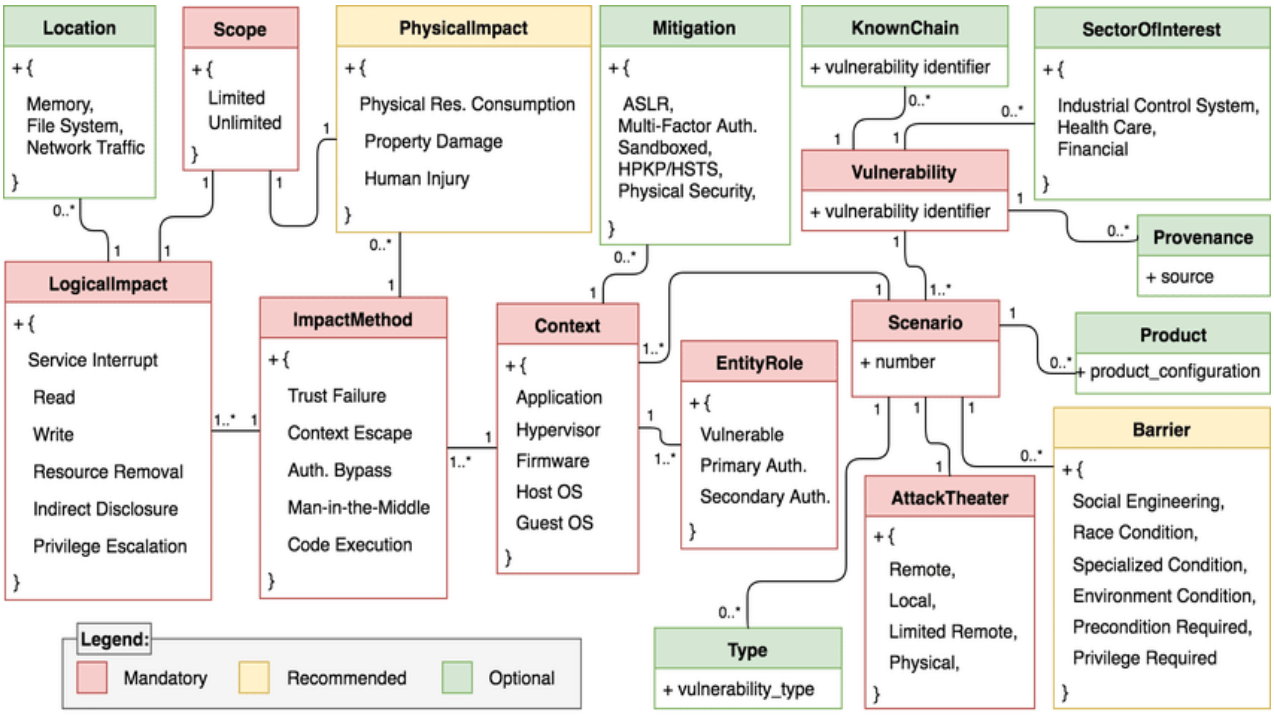


Рис. 1. Онтологія опису вразливості НІСТ (ООВ)

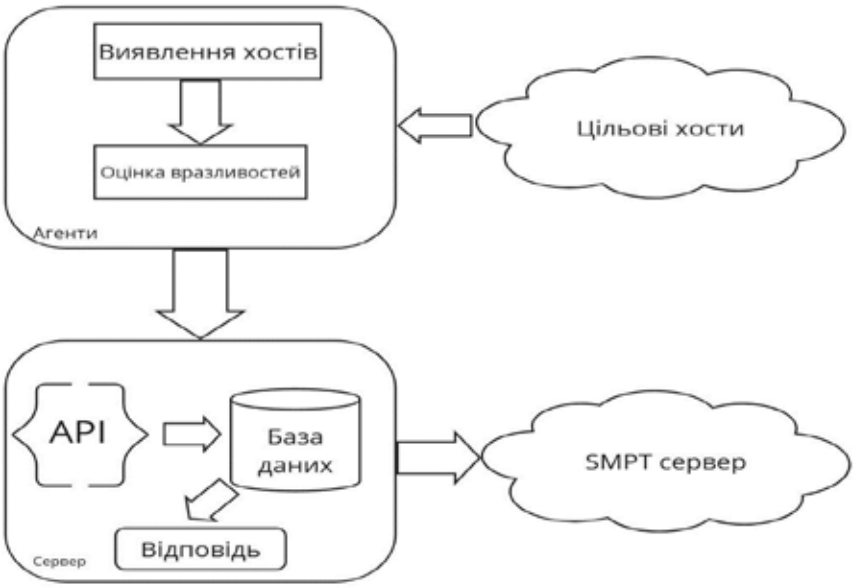


Рис. 2. Графічне представлення зв'язку між модулями оцінки вразливості безпеки (стрілки зображують потік даних)

Використання неструктурованих даних, таких як репозиторії коду, разом з технологіями глибокого навчання дозволяє створювати системи, що більш ефективно виявляють патерни, схожі на вразливості, ніж стандартні алгоритми аналізу (Elmrabit, 2020).

Статичний та динамічний аналіз коду залишаються двома основними підходами для виявлення вразливостей. Останні дослідження демонструють ефективність комбінованого використання обох підходів для більш повного аналізу програмного забезпечення. Статичні інструменти, такі як SonarQube, Checkmarx, та Clang Static Analyzer, стали стандартом для виявлення помилок на етапі розробки. Вони фокусуються на пошуку вразливостей в синтаксисі та структурі коду, що дозволяє розробникам своєчасно виправляти потенційно небезпечні фрагменти.

Водночас, динамічний аналіз, зокрема з використанням інструментів для тестування на проникнення, таких як OWASP ZAP та Burp Suite, зосереджує увагу на перевірці коду в процесі його виконання. Останні дослідження також показують, що інтеграція цих інструментів у сучасні пайплайни CI/CD забезпечує більш швидку верифікацію коду в реальному часі, дозволяючи автоматично виявляти уразливості в тестових середовищах.

Дослідження доводять, що поєднання статичного і динамічного аналізу дає змогу знижувати ймовірність виявлення помилок лише на одному етапі тестування, забезпечуючи багатогранний підхід до безпеки (Acharya, 2023).

Дослідження показують, що інтеграція інструментів виявлення вразливостей у процеси CI/CD стає стандартом практики для безпечного програмування. Наприклад, використання таких інструментів, як GitLab CI/CD, разом із автоматизованими системами для перевірки коду, дає можливість здійснювати регулярні перевірки безпеки на всіх етапах розробки, зменшуючи ймовірність помилок, що потрапляють у кінцевий продукт.

У статтях дослідники вказують, що автоматизація перевірок на етапах розробки й деплою не тільки підвищує якість програмного забезпечення, а й дозволяє виявляти вразливості ще до потрапляння коду у продуктивне середовище. Інтеграція таких інструментів дозволяє швидше реагувати на загрози, знижуючи час, необхідний для виявлення та виправлення вразливостей (Smith, 2023).

У Україні також активно проводяться дослідження в сфері безпечного програмування та автоматизованих інструментів для виявлення вразливостей в коді. Наприклад, українські науковці в своїх дослідженнях використовують нейронні мережі для аналізу великих обсягів коду в реальному часі, що допомагає виявляти нові типи вразливостей і навіть передбачати можливі атаки. Окрім академічних досліджень, низка українських IT-компаній займається розробкою та вдосконаленням автоматизованих інструментів для виявлення вразливостей у програмному забезпеченні. Ці дослідження свідчать про значний прогрес в Україні в галузі безпечного програмування і автоматизованого

виявлення вразливостей. Вчені й фахівці вітчизняних університетів і компаній не тільки розробляють нові методи та інструменти для покращення безпеки, але й активно інтегрують їх у реальні процеси розробки програмного забезпечення.

Таким чином, сучасні методи виявлення вразливостей включають комбінований підхід з використанням різних технік, інтегрованих у розробницькі пайплайни, але, враховуючи думки різних авторів, доводиться констатувати, що в умовах постійного впливу зовнішніх та внутрішніх чинників необхідно приділити більше уваги дослідженню проблем безпеці і це питання є актуальним.

Мета дослідження. Метою статті є надання детального аналізу існуючих інструментів і методів автоматизованого виявлення вразливостей у програмному забезпеченні, охоплюючи як класичні підходи (статичний та динамічний аналіз коду), так і новітні технології, зокрема ті, що використовують машинне навчання і штучний інтелект для підвищення точності виявлення вразливостей. Дана стаття має на меті показати, як інструменти для автоматизованого виявлення вразливостей можуть бути інтегровані у процеси розробки програмного забезпечення. Також важливим є допомогти підвищити рівень кібербезпеки в Україні через популяризацію кращих практик безпечного програмування і розвиток автоматизованих інструментів для виявлення вразливостей.

Для досягнення поставленої мети треба виконати наступні завдання:

– Провести детальний огляд сучасних наукових публікацій, досліджень та технічних звітів з теми безпечного програмування та автоматизованих інструментів для виявлення вразливостей.

– Оцінити ефективність основних підходів до виявлення вразливостей у коді, зокрема статичний і динамічний аналіз, а також нові технології, такі як машинне навчання.

– Проаналізувати майбутні тенденції в розвитку автоматизованих інструментів для виявлення вразливостей і визначити напрямки, які потребують подальшого розвитку.

– Підсумувати основні висновки та зробити рекомендації для практичного використання автоматизованих інструментів для виявлення вразливостей.

Виклад основного матеріалу дослідження. Наразі існує широкий спектр інструментів статичного аналізу, які намагаються виявити типові вразливості. Вони варіюються від реалізацій з відкритим кодом, таких як статичний аналізатор Clang, до пропрієтарних рішень (Kremenek, 2018). Статичні аналізатори роблять це без необхідності виконувати програми. Динамічні аналізатори неодноразово виконують програми з багатьма тестовими входами на реальних або віртуальних процесорах, щоб виявити слабкі місця. Як статичні, так і динамічні аналізатори є інструментами, заснованими на правилах, тому вони обмежені правилами, створеними вручну, і не можуть гарантувати повне тестове покриття кодових баз.

Щоб виявити вразливості в коді було розглянуто два взаємодоповнюючі підходи. Перший використовує функції, отримані з проміжного представлення програм, створених під час процесу компіляції та збірки. Другий працює безпосередньо з вихідним кодом. Ці два підходи пропонують незалежні джерела інформації. Моделі, засновані на вихідному коді, можуть вивчати статистичні кореляції в тому, як пишеться код, використовуючи методи обробки природної мови, але не мають жодних внутрішніх знань про структуру чи семантику вбудованої мови (все це потрібно вивчати з даних). Моделі на основі збірки використовують переваги функцій, отриманих від компілятора, який розуміє структуру мови, але ці функції можуть абстрагувати певні властивості коду, корисні для виявлення вразливостей. Застосовуючи обидва підходи, ми дозволяємо зрештою об’єднати кілька моделей для підвищення ефективності виявлення вразливостей.

Враховуючи складність і різноманітність програм, потрібна велика кількість навчальних прикладів для навчання моделей машинного навчання, які можуть ефективно вивчати шаблони вразливостей безпеки безпосередньо з коду (Gonzalez, 2019). Було зібрано величезний набір даних із мільйонів прикладів функціонального рівня коду C і C++ із SATE IV, дистрибутива Debian Linux і загальнодоступних репозиторіїв Git на GitHub. У таблиці 1 показано зведені дані про кількість функцій, які було зібрано та використано з кожного джерела в наборі даних із понад 12 мільйонів функцій.

Хоча набір даних SATE IV надає позначені приклади багатьох типів вразливостей, він складається з фрагментів синтетичного коду, які недостатньо охоплюють простір природного коду, щоб забезпечити лише відповідний навчальний набір. Випуски пакетів Debian надають вибір дуже добре керованого та підібраного коду, який сьогодні використовується в багатьох системах. Набір даних GitHub надає більшу кількість і більш різноманітний (часто нижчої якості) код.

Таблиця 1

Зведені дані аналізу функцій

	SATE IV	GitHub	Debian
Всього	121,353	9,706,269	3,046,758
Пройшли перевірку	11,896	782,493	491,873
Не вразливі	6,503 (55%)	730,160 (93%)	461,795 (94%)
Вразливі	5,393 (45%)	52,333 (7%)	30,078 (6%)

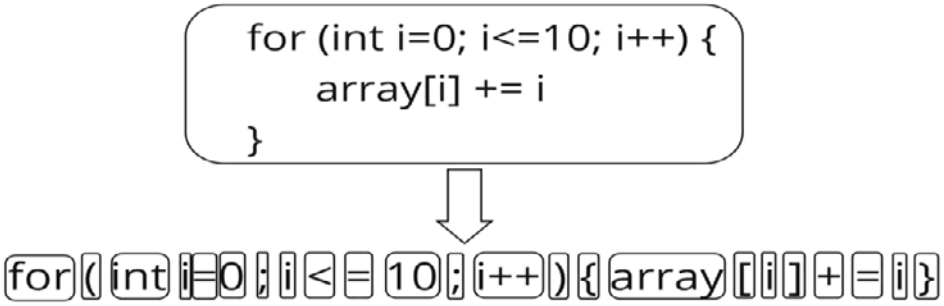


Рис. 3. Приклад, що ілюструє процес лексичного складання

Було реалізовано спеціальний лексер C/C++, який аналізує код і класифікує елементи в різні контейнери: коментарі, рядкові літерали, літерали з одним символом, літерали з кількома символами, числа, оператори, директиви попереднього компілятора та імена. Назви далі поділені на ключові слова (такі як for, if тощо), виклики системних функцій (такі як malloc), типи (такі як bool, int тощо) або імена змінних (Aminu, 2024). Усі назви змінних зіставляються з тим самим загальним ідентифікатором, але кожне унікальне ім'я змінної в одній функції отримує окремий індекс (з метою відстеження того, де змінні знову з'являються). Це відображення змінюється від функції до функції залежно від позиції, у якій ідентифікатор з'являється вперше. Унікальні рядкові літерали, з іншого боку, відображаються на один і той самий ідентифікатор. Нарешті, цілі літерали зчитаються цифра за цифрою, тоді як літерали з плаваючою речовиною відображаються на загальний маркер. На рисунку 2 показаний приклад процесу лексики.

Розроблений лексер зміг скоротити код C/C++ до представлень, використовуючи загальний розмір словника лише 156 токенів. Також одним з дуже важливих кроків підготовки даних було видалення потенційних дублюючих функцій. Сховища з відкритим вихідним кодом часто мають функції, що дублюються в різних пакетах. Таке дублювання може штучно завищити показники продуктивності та приховати повторні визначення, оскільки навчальні дані можуть витікати в тестові набори. Так само є багато функцій, які є майже дублікатами, містять тривіальні зміни у вихідному коді, які суттєво не впливають на виконання функції. Ці майже дублікати важко видалити, оскільки вони часто можуть з'являтися в дуже різних частинах коду та можуть виглядати зовсім по-різному на рівні вихідного коду.

Щоб захиститися від цих проблем, було виконано надзвичайно суворий процес видалення дублікатів. Була видалена будь-яка функцію з дубльованим лексичним представленням вихідного коду або дубльованим вектором функцій на рівні компіляції. Цей вектор функцій на рівні компіляції було створено шляхом вилучення графа потоку керування функції, а також операцій, що відбуваються в кожному базовому блоці (opcode vector, чи op-vec), а також визначення та використання змінних (матриця use-def).

Рядок «Прохідна перевірка» у таблиці 1 відображає кількість функцій, що залишилися після процесу видалення дублікатів, приблизно 10,8% від загальної кількості вилучених функцій. Хоча наш жорсткий процес видалення дублікатів відфільтровує значну кількість даних, цей підхід забезпечує найбільш консервативні результати продуктивності, точно оцінюючи, наскільки добре наш інструмент працюватиме з кодом, якого він ніколи раніше не бачив.

Позначення вразливості коду на функціональному рівні було серйозною проблемою. Щоб створити мітки, було використано три підходи: статичний аналіз, динамічний аналіз і додавання тегів до повідомлень/звітів про помилки.

Незважаючи на те, що динамічний аналіз здатний виявляти тонкі недоліки, виконуючи функції з широким діапазоном можливих вхідних даних, він надзвичайно ресурсомісткий, тому цей підхід не був реалістичним для надзвичайно великого набору даних. Виявилося, що маркування на основі комітів-повідомлень є дуже складним завданням, оскільки надає мітки низької якості. Також був опробований простий пошук за ключовими словами, шукаючи такі слова фіксації, як «buggy», «broken», «error», «fixed», тощо, щоб позначити пари функцій до та після, що дало кращі результати з точки зору релевантності. Однак цей підхід значно зменшив кількість функцій-кандидатів і все ще вимагав значної перевірки вручну, що робило його невідповідним для величезного набору даних. У результаті було вирішенно використовувати три статичні аналізатори з відкритим вихідним кодом, Clang, Cppcheck і Flawfinder (Zhou, 2019), щоб генерувати мітки. Кожен статичний аналізатор різниться за сферою пошуку та виявлення. Наприклад, сфера дії Clang дуже широка, але також включає синтаксис, стиль програмування та інші висновки, які навряд чи призведуть до вразливості. Сфера дії Flawfinder орієнтована на загальний перелік вразливостей і не зосереджується на інших аспектах, таких як стиль. Тому було включено кілька статичних аналізаторів і скоротили їхні результати, щоб виключити знахідки, які зазвичай не пов'язані з уразливими місцями безпеки, щоб створити надійні мітки «вразливий» і «не вразливий» залежно від загального переліку вразливостей. Із 390 загальних типів результатів статичних аналізаторів було визначено, що 149 призводять до потенційної вразливості безпеки. Приблизно 6,8% наших підібраних видобутих функцій C/C++ викликали виявлення вразливості. Таблиця 2 показує статистику частих вразливостей у цих функціях.

При побудові моделей машинного навчання для виявлення помилкових функцій було розглянуто дві цілі. По-перше, оцінка загально можливої продуктивності – наскільки добре кожна модель може передбачити наявність небезпечних методів у коді. По-друге, порівняння продуктивності функцій на основі складання та на основі вихідного коду, щоб виявити, який набір краще прогнозує якість коду.

Реалізований підхід машинного навчання до виявлення вразливостей поєднує представлення нейронних функцій вихідного коду лексичної функції з потужним класифікатором ансамблю, випадковим лісом.

Використано підходи вилучення ознак, подібні до тих, що використовуються для класифікації настрою речення за допомогою згорткових нейронних мереж (CNN) і рекурентних нейронних мереж (RNN) для класифікації вразливості джерела на функціональному рівні:

1) Вбудовування функцій: маркери, що утворюють лексифіковані функції, спочатку вбудовуються у фіксоване k-вимірне представлення (обмежене діапазоном [–1, 1]), яке вивчається під час навчання класифікації через зворотне поширення до лінійного перетворення одноразового вбудовування. Оскільки розроблений словниковий запас набагато менший, ніж словниковий запас природних мов, була змога використати набагато менше вбудовування, ніж це типово для додатків NLP. Експерименти показали, що k = 13 є найкращим для контрольованих розмірів вбудовування, врівноважуючи виразність вбудовування та надмірне припасування.

2) Вилучення функцій: було досліджено як CNN, так і RNN для вилучення функцій із вбудованих представлень джерела. Вилучення згорткових ознак: використовуємо n згорткових фільтрів із формою m × k, тому кожен фільтр охоплює повний простір вбудовування маркера. Розмір фільтра m визначає кількість послідовних маркерів, які розглядаються разом, і було виявлено, що досить великий розмір фільтра m = 9 працює найкраще. Загальна кількість n = 512 фільтрів у поєднанні з пакетною нормалізацією з наступною ReLU була найбільш ефективною.

3) Об'єднання: оскільки довжина функцій C/C++, може різко змінюватися, як згорткові, так

Таблиця 2

Статистика щодо вразливостей, виявлених у аналізованому наборі даних C/C++

ID вразливості	Опис вразливості	Частота, %
120/121/122	Переповнення буфера	38.2%
119	Неправильне обмеження операцій у межах буфера пам'яті	18.9%
476	Розіменування покажчика NULL	9.5%
469	Використання віднімання вказівника для визначення розміру	2.0%
20, 457, 805, тощо	Неналежна перевірка введення, використання неініціалізованої змінної, доступ до буфера з неправильним значенням довжини, тощо	31.4%

і рекурентні функції максимально об'єднуються вздовж довжини послідовності, щоб створити представлення фіксованого розміру (n або n відповідно). У цій архітектурі рівні вилучення функцій повинні навчитися визначати різні сигнали вразливості, тому наявність будь-якого з них у послідовності є важливою.

4) Щільні шари: за шарами вилучення ознак слідує повний зв'язаний класифікатор. Під час навчання було використано 50% випадання з'єднань максимального об'єднаного представлення функцій до першого прихованого шару. Було виявлено, що використання двох прихованих шарів 64 і 16 перед остаточним вихідним шаром softmax дало найкращу ефективність класифікації.

5) Навчання: для зручності пакетування даних було проведено тренування лише на функціях із довжиною маркера 10 ≤ ≤ 500, доповненою до максимальної довжини 500. Оскільки набір даних був сильно незбалансованим, уразливі функції були важливіші у функції втрати.

Було виявлено, що використання нейронних особливостей дало найкращі результати, окрема оптимізація функцій і класифікатора, здається, допомогла протистояти переобладнанню. Цей підхід також робить зручнішим швидко перенавчати класифікатор на нові набори ознак або комбінації ознак.

Перевірка важливості функції пакета слів показує, що класифікатор використовує кореляції міток із індикаторами довжини та складності джерела та комбінаціями викликів, які зазвичай зловживають і призводять до вразливостей. Покращення в порівнянні з базовою лінією можна інтерпретувати як результат більш складних і специфічних моделей індикації вразливості. Загалом розроблені моделі CNN показали кращі результати, ніж моделі RNN як автономні класифікатори та генератори ознак. Крім того, CNN швидше навчалися і вимагали набагато менше параметрів.

Розроблені методи машинного навчання мають деякі додаткові переваги порівняно з традиційними інструментами статичного аналізу. Розроблені користувальницькі лексери та моделі машинного навчання можуть швидко переробляти та оцінювати великі сховища та вихідний код без необхідності компілювати код. Крім того, оскільки машинне навчання обробляє всі вихідні ймовірності, порогові значення можна налаштувати для досягнення бажаної точності та запам'ятовування.

З іншого боку, статичні аналізатори повертають фіксовану кількість результатів, яка може бути надзвичайно великою для величезних кодових баз або занадто малою для критичних програм. У той час як статичні аналізатори можуть краще локалізувати вразливості, які вони знаходять, нами також було використано методи візуалізації, такі як карта активації функції, показана на рисунку 4, щоб допомогти зрозуміти, чому наші алгоритми приймають свої рішення.

Висновки і перспективи подальших досліджень. Було продемонстровано потенціал використання машинного навчання для виявлення вразливостей програмного забезпечення безпосередньо з вихідного коду. Для цього було створено обширний набір даних вихідного коду C/C++, видобутий із сховищ Debian і GitHub, позначений підібраними

```
wchar_t * data;
unionType myUnion;
data = new wchar_t[100];
wmemset(data, L 'A', 100 - 1);
data[100 - 1] = L '\0';
myUnion.unionFirst = data;
{
    wchar_t * data = myUnion.unionSecond;
    {
        wchar_t dest[50] = L "";
        memcpy(dest, data, wcslen(data) * sizeof(wchar_t));
        dest[50 - 1] = L '\0';
        printWLine(data);
        delete[] data;
    }
}
```

Рис. 4. Знімок екрана для інтерактивної демонстрації виявлення вразливостей

виявленими вразливостей із набору інструментів статичного аналізу, і об'єднанням його з набором даних SATE IV. Було розроблено власний лексер C/C++, щоб створити просте загальне представлення вихідного коду функції, ідеальне для машинного навчання. Було застосовано різні методи машинного навчання, які були точно налаштовані для розробленої програми та досягли найкращих загальних результатів, використовуючи функції, отримані за допомогою нейронної мережі. Пройдено до висновків, що підхід, керований даними, може доповнити існуючі інструменти статичного аналізу та скоротити кількість часу, необхідного для виявлення потенційних вразливостей у програмному забезпеченні. Ранжуючи функції за ймовірністю помилок, розробники зможуть швидше завершити перевірку коду та виділити потенційні проблеми, які, можливо, не були виявлені інакше.

Майбутня робота має бути зосереджена на вдосконаленні міток, таких як мітки з інструментів динамічного аналізу або видобуті з патчів безпеки. Це дозволить оцінкам, створеним на основі моделей машинного навчання, більше доповнювати інструменти статичного аналізу. Методи автоматизації, розроблені в цій роботі для навчання безпосередньо на вихідному коді функції, також можуть бути застосовані до будь-якої проблеми класифікації коду, такої як виявлення порушень стилю, категоризація комітів або класифікація алгоритмів/завдань. У міру розробки більших і краще позначених наборів даних глибоке навчання для аналізу вихідного коду стане більш практичним для широкого спектру важливих проблем.

ЛІТЕРАТУРА:

1. B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, «Learning deep features for discriminative localization,» in Computer Vision and Pattern Recognition (CVPR), pp. 2921–2929, IEEE, 2019.

2. Check Point Blog. Check Point Research: Third Quarter of 2022 Reveals Increase in Cyberattacks and Unexpected Developments in Global Trends. checkpoint.com. URL: https://blog.checkpoint.com/2022/10/26/third-quarter-of-2022-revealsincrease-in-cyberattacks/

3. D. Gonzalez, F. Alhenaki, and M. Mirakhorli. Architectural security weaknesses in industrial control systems (ICS) an empirical study based on disclosed software vulnerabilities. In 2019 IEEE International Conference on Software Architecture (ICSA), pages 31–40, March 2019.

4. G. Smith. The intelligent solution: Automation, the skills shortage and cyber-security. Comput. Fraud. Secur, 2018, 6–9.

5. H. Booth, D. Rike H. Booth, D. Rike, and G. Witte. The National Vulnerability Database (NVD): Overview. Technical report, National Institute of Standards and Technology (NIST), 2021.

6. H. Booth. Draft NISTIR 8138, Vulnerability Description Ontology (VDO). Technical report, National Institute of Standards and Technology (NIST), 2016.

7. M. Aminu, S. Anawansedo, Y. Sodiq. Driving Technological Innovation for a Resilient Cybersecurity Landscape. *International Journal of Latest Technology in Engineering, Management & Applied Science*, 13(4), 126–133, 2024.

8. N. Elmrabit, F. Zhou, F. Li. Evaluation of Machine Learning Algorithms for Anomaly Detection. In Proceedings of the 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), Dublin, Ireland, 15–19 June 2020; pp. 1–8.

9. S. Acharya, U. Rawat, R. Bhatnagar. A Comprehensive Review of Security: Treats, Vulnerabilities, MalwareDetection, and Analysis. In Security and Communication Net-works, 34 pages, 2023.

10. Z. Kremenek, T. Zhang. A memory model for static analysis of c programs. In International Symposium On Leveraging Applications of Formal Methods, Verification and Validation (2018), Springer, pp. 535–548.

REFERENCES:

1. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. and Torralba, A. (2019). «Learning deep features for discriminative localization,» in Computer Vision and Pattern Recognition (CVPR), pp. 2921–2929, IEEE,

2. Check Point Blog. Check Point Research: Third Quarter of 2022 Reveals Increase in Cyberattacks and Unexpected Developments in Global Trends. checkpoint.com. Available online: https://blog.checkpoint.com/2022/10/26/third-quarter-of-2022-revealsincrease-in-cyberattacks/ (accessed on 2 February 2024).

3. Gonzalez, D., Alhenaki, F. and Mirakhorli, M. (March 2019). Architectural security weaknesses in industrial control systems (ICS) an empirical study based on disclosed software vulnerabilities. In 2019 IEEE International Conference on Software Architecture (ICSA), pages 31–40,

4. Smith, G. (2018) The intelligent solution: Automation, the skills shortage and cyber-security. Comput. Fraud. Secur, 6–9.

5. Booth, H., Rike, D., Booth, H., Rike, D. and Witte, G. (2021). The National Vulnerability Database (NVD): Overview. Technical report, National Institute of Standards and Technology (NIST),

6. Booth, H. (2016). Draft NISTIR 8138, Vulnerability Description Ontology (VDO). Technical report, National Institute of Standards and Technology (NIST),

7. Aminu, M. Anawansedo, S. Sodiq, Y. (2024). Driving Technological Innovation for a Resilient Cybersecurity Landscape. *International Journal of Latest Technology in Engineering, Management & Applied Science*, 13(4), 126–133,

8. Elmrabit, N. Zhou, F. Li, F. (2020). Evaluation of Machine Learning Algorithms for Anomaly Detection. In Proceedings of the 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), Dublin, Ireland, 15–19 June; pp. 1–8.

9. Acharya, S. Rawat, U. Bhatnagar, R. (2023). A Comprehensive Review of Security: Treats, Vulnerabilities, MalwareDetection, and Analysis. In Security and Communication Net-works, 34 pages,

10. Kremenek, Z., Zhang, T. (2018). A memory model for static analysis of c programs. In International Symposium On Leveraging Applications of Formal Methods, Verification and Validation, Springer, pp. 535–548.

УДК 004.93:633.854
DOI https://doi.org/10.32782/IT/2024-4-7

Станислав ВЕДМЕДЄВ
аспірант кафедри системного аналізу та обчислювальної математики, Національний університет «Запорізька політехніка», вул. Жуковського, 64, м. Запоріжжя, Україна, 69063
ORCID: *0009-0005-9635-8879*
Scopus Author ID: *58298235800*

Бібліографічний опис статті: Ведмедєв, С. (2024). Розробка програмного забезпечення для збирання фенотипічних даних насіння соняшнику. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 4, 53–60, doi: https://doi.org/10.32782/IT/2024-4-7

РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ЗБИРАННЯ ФЕНОТИПІЧНИХ ДАНИХ НАСІННЯ СОНЯШНИКУ

Головним напрямком в селекції соняшнику є підвищення якості насіння й промислових характеристик, що є важливими для споживача. Для оцінки якості насінини може бути використано оцінку за фенотипом. Фенотипування дозволяє створити формалізований підхід до оцінювання, що включає перелік ознак, форму точного опису та засоби для вимірювання цих характеристик.

Метою представленого дослідження *є розробка програмного забезпечення вивчення морфологічних ознак насінини соняшнику за фотозображеннями, а саме метричних характеристик, форми та забарвлення насінини.*

Методологія. *Розроблені програми базуються на точному математичному обрахунку, що забезпечує точність створених описів. Програмні продукти аналізу зображень розроблено зі застосуванням мови програмування Python. Програма розрахунків розроблена в середовищі PyCharm Community Editor 2022.2.4 з використанням бібліотек: os, cv2, imutils, numpy, time, Pool. Cv2-для обробки зображення. Дослідження проводилося на колекції фотозображень насінини соняшнику, що містить 3000 фотографій 250 ліній соняшнику та створена фахівцями Інституту олійних культур Національної Академії Аграрних Наук України.*

Наукова новизна *полягає в тому що вперше було розроблено програми для аналізу зображення насіння соняшнику з урахуванням біологічного різноманіття насіння. Вони вперше дали змогу описати досі якісні ознаки насіння у якості таблиць кількісних показників.*

Висновки. *У результаті проведеного дослідження розроблено модель для вимірювання метричних характеристик та форми насінини, а також програмне забезпечення, яке складається з програм розпізнавання метричних характеристик насіння, форми та смугастого забарвлення насінини. Розробки можуть бути використані для аналізу фотозображень насінини, яке зроблено на стаціонарному приладі з детермінованими параметрами в умовах лабораторії. Отриманні результати є важливою складовою для селекційної роботи по створенню нових високопродуктивних гібридів, сортів та ліній соняшнику. Зараз розробка використовується селекціонерами та генетиками Інституту олійних культур Національної Академії Аграрних Наук України в ході наукової селекційної роботи.*

Ключові слова: *фотозображення, насіння, програмне забезпечення, розпізнавання, фенотипування*

Stanyslav VEDMEDEV
Postgraduate Student at the Department of System Analysis and Computational Mathematics, Zaporizhia Polytechnic National University, 64, Zhukovsky Str., Zaporizhzhia, Ukraine, 69063, vedmedev_s@ukr.net
ORCID: *0009-0005-9635-8879*
Scopus Author ID: *58298235800*

To cite this article: Vedmedev, S. (2024). Rozrobka prohramnoho zabezpechennia dlia zbyrannia fenotypichnykh danykh nasinnia soniashnyku [Development of software for collecting phenotypic data of sunflower seeds]. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 4, 53–60, doi: https://doi.org/10.32782/IT/2024-4-7

DEVELOPMENT OF SOFTWARE FOR COLLECTION OF PHENOTYPIC DATA OF SUNFLOWER SEEDS

The main direction in sunflower breeding is the improvement of seed quality and industrial characteristics, which are important for the consumer. Phenotypic evaluation can be used to assess seed quality. Phenotyping allows for a formalized approach to assessment that includes a list of traits, a form of precise description, and means to measure those characteristics.

The purpose of the presented *research is the development of software for the study of morphological features of sunflower seeds based on photo images, namely metric characteristics, shape and color of seeds.*

Methodology. *The developed programs are based on accurate mathematical calculation, which ensures the accuracy of the created descriptions. Image analysis software products are developed using the Python programming language. The calculation program was developed in the PyCharm Community Editor 2022.2.4 environment using the following libraries: os, cv2, imutils, numpy, time, Pool. Cv2-for image processing. The study was conducted on a collection of sunflower seed photographs, which contains 3,000 photographs of 250 sunflower lines and was created by specialists of the Institute of Oil Crops of the National Academy of Agrarian Sciences of Ukraine. Developments can be used for the analysis of photo images of seeds, which is made on a stationary device with deterministic parameters in laboratory conditions.*

The scientific novelty *is that, for the first time, programs were developed for the analysis of the image of sunflower seeds, taking into account the biological diversity of the seeds. For the first time, they made it possible to describe the qualitative characteristics of seeds in the form of tables of quantitative indicators.*

Conclusions. *As a result of the research, a model was developed for measuring the metric characteristics and shape of the seed, as well as software, which consists of programs for recognizing the metric characteristics of the seed, shape and striped color of the seed. The obtained results are an important component for selection work to create new high-yielding hybrids, varieties and lines of sunflower. Now the development is used by breeders and geneticists of the Institute of Oil Crops of the National Academy of Agrarian Sciences of Ukraine in the course of scientific selection work.*

Key words *photo images, seeds, software, recognition, phenotyping.*

Актуальність проблеми. Серед господарських культур України найбільший економічний ефект зумовлює культура соняшнику. Насьогодні актуальними проблемами для аграрного виробництва соняшнику визначено забезпечення якості соняшникової продукції, механізація процесів вирощування та створення нових сортів. Одним із ключових напрямків у селекційній роботі є фенотипування рослин, що здійснюється як у польових умовах, так і в лабораторіях. Традиційні методи оцінки фізичних характеристик рослин упродовж їхнього вегетативного розвитку передбачають ручне спостереження та фіксацію, що характеризується суб'єктивністю оцінок, низькою точністю, відсутністю стандартизації та потребує значних часових витрат (Plants guidelines). Сучасні селекційні експерименти потребують використання новітніх технологій, що включають як створення пристроїв та устаткування, так й методів комп'ютерного обробки інформації та агрегації біометричних даних.

Аналіз предметної галузі та наявних рішень. Сучасним трендом по вирішенню цієї задачі є фенотипування (Pieruschka, 2019). Термін позначає процес переведення візуальної інформації про об'єкт у цифрове сприйняття яке забезпечує класифікацію цих об'єктів. Цей процес можна розкласти на кілька етапів: вибір ознак, що підлягає вивченню (Ведмедєва,2023), підбір та створення умов і обладнання для отримання зображень, збір великої кількості зображень з урахуванням факторів впливу на фенотип об'єкту, розробка програм для аналізу зображень та переведення їх у табличні редактори, створення бази даних отриманої інформації, онтологій (Bakurova, 2023), використання бази даних для класифікації зразків, прогнозу якостей та добору. На кожному з цих етапів вирішують свої проблеми. Серед них є загальні, характерні для усіх об'єктів. Наприклад задачі підпису та поєднання фото з підписом (Triki, 2022), або збереження точності вимірів при зменшенні розмірів фотографій (Borges, 2020). Дуже важливою є проблема оптимізації обсягу інформації при збереженні її цінності (Walker, 2022). Вже зроблені дослідження на окремих рослинних об'єктах з метою фенотипування. За проектом виконаним у Юліхському дослідницькому центрі використано позитронно-емісійну томографію і вивчено зображення живих коренів у ґрунті (Fiorani, 2013). У дослідженнях Китайського сільськогосподарського університету 3D-моделі 20 генотипів цукрових буряків були реконструйовані методом SFM на трьох етапах їх росту в польових умовах (Xiao, 2020). Було розроблено автоматичний конвеєр для обробки даних, включаючи обробку хмари точок окремих рослин та вилучення їх характерних ознак. Дослідженнями Yang, S. з співавторами (Yang, 2021) розроблено метод фенотипування насіння сої. Запропоновано новий синтетичний метод генерації та доповнення зображень та метод навчання переносу шляхом точного налаштування вагових коефіцієнтів попередньо навченої моделі, що забезпечує зниження витрат на обчислення. В роботі (Олешко, 2020) запропоновано застосування алгоритму Е. Ростена & Т. Драммонда (Rosten 2006) в задачах розпізнавання хвороб рослин у поєднанні з детекторами кольору. В роботі (Гетьман, 2022) побудовано систему розпізнавання образів по фотографії на основі навченої нейронної мережі за категоріями в залежності від частини

рослини. Для аугментації обиралися ті перетворення, які відбуваються в реальному житті.

Відомо багато досліджень з розпізнавання образів та вимірювання об'єктів на фото, однак у селекційному значенні використовується не велика кількість розробок. Відомі лише початкові публікації з розробки цих технологій (Старчак 2023). Зокрема на коренях (Kumar, 2014) та листках (Shibayama, 2011) рослин. Розробки перших пристроїв автоматичного фенотипування насіння (Алієв, 2022) ставили на меті масове вилучення насінин, що різко відрізнялися за кольором. Застосування на насінні соняшнику з селекційною метою, яке проводиться с залученням представлених нами програм знаходиться у стадії розробки та отримання перших результатів (Махова, 2023).

Автор цієї статті брав участь в розробці онтологічного підходу для управління створеними сховищами даних по вирощуванню, селекції, генетики та фенотипуванню соняшника (Bakurova, 2023). Також результати розробки The Helianthus phenotype digital model system для реалізації селекційних програм соняшнику кондитерського було представлено на VI International Scientific Congress Society of Ambient Intelligence 2023 November 20–25, 2023 (Society of ambient intelligence, 2023).

Метою представленого дослідження є розробка програмного забезпечення вивчення морфологічних ознак насінини соняшнику за фотозображеннями, а саме метричних характеристик, форми та забарвлення насінини.

Матеріали та методи. Колекція фотозображень насінини соняшнику містить 3000 фотографій 250 ліній соняшнику та створена фахівцями Інституту олійних культур НААН України. Для фотографування в лабораторних умовах відбувалося зі застосуванням стаціонарного приладу, що має детерміновані налаштування, розробленого Е.Б. Алієвим (Алієв, 2019).

Колекційні зразки насіння бувають різної форми: трикутної, овальної, закругленої, близької до прямокутної. За розміром змінюються від 0,5 до 1,5 мм довжини насінини. Для отримання тривимірних параметрів насінини використовуються фотографії, що містять дві проєкції кожної насінини. Забарвлення має як різний колір, так й різний малюнок. Колір змінюється від білого до чорного з різними відтінками, від жовтого до темно-коричневого, або бордового. За малюнком бувають однокольорові, смугасті з невеликою або великою кількістю смужок (рис. 1а) (Gorohivets, 2016).

Програмні продукти аналізу зображень розроблено зі застосуванням мови програмування Python. Програма розрахунків розроблена в середовищі PyCharm Community Editor 2022.2.4 з використанням бібліотек: os, cv2, imutils, numpy, time, Pool. Cv2-для обробки зображення.

Результати і обговорення. Розроблено та апробовано програмне забезпечення для аналізу фотозображень насіння соняшнику, що виконано в лабораторних умовах на стаціонарному приладі з детермінованими налаштуваннями.

Програмне забезпечення складається з двох блоків: визначення метричних розмірів та визначення забарвлення насінини.

Програмне забезпечення для розпізнавання метричних розмірів насінини. Оригінальна фотографія містила зображення насінини, що центрувалося по довжині вздовж діагоналі знімку. Це викликано розташуванням предметного столику стаціонарного приладу, на якому розміщувалася насінина для фотографування. Тому, на першому кроці зображення насінини (рис. 1 а) повертається на кут в 315 градусів та обрізаються зайві ділянки зображення. Кут повернення постійний. Координати обрізки можна змінювати у разі необхідності для кожного виду насінини. Приклад отриманого фото на рис. 1 б. На другому етапі проводимо обробку зображення. Приймемо модель для вимірювання метричних характеристик та форми насінини у формі ромбу, що вписаний у контури насінини (рис. 1 в). Далі знаходимо розміри діагоналей ромбу, що визначають довжину та ширину насінини. Далі визначаємо кути ромбу, повну площу ромбу та площі окремих трикутників, для визначення форми насінини.

Розрахунок відрізків довжини та ширини відбувається за формулами:

line_a = y – y_min_r_coordinate;

line_b = y_max_r_coordinate – y;

line_c = x – x_min_c_coordinate;

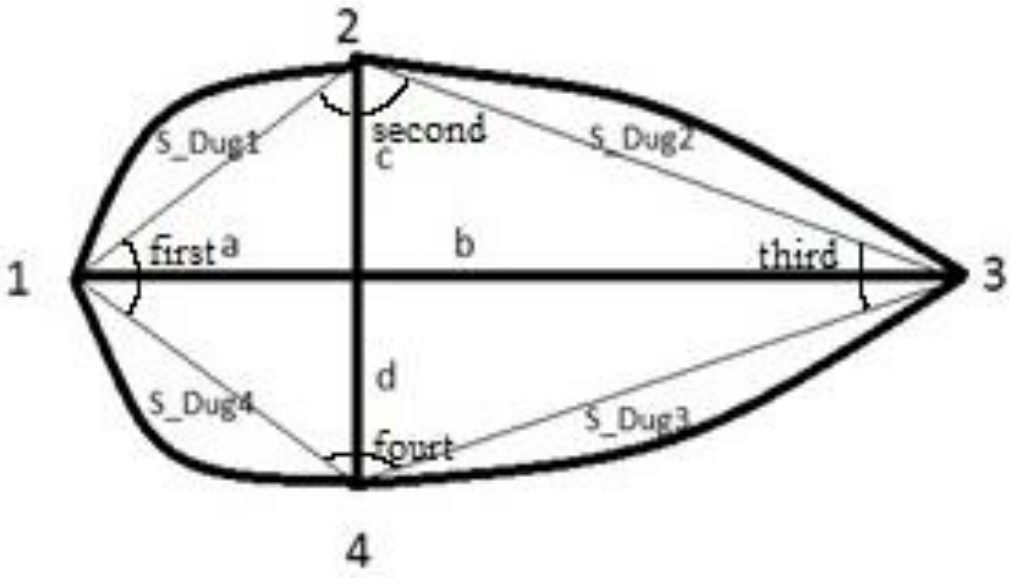
line_d = x_max_r_coordinate – x,

де

line_a та line_b -складові довжини насінини, lineC та lineD – складові ширини насінини (рис. 1б).

Результати розрахунків параметрів організовано у вигляді таблиці, де рядок містить інформацію по кожному фото. Перший стовпчик – це назва фото. Далі розташовано виміри кутів, площі трикутників та площі ділянок за межами трикутників. Вони позначені: firstAngle – перший кут; secondAngle – другий

а) необроблене фото б) фото смугастої насінини



в) модель для вимірювання метричних характеристик та форми насінини

Рис. 1. Фото насінини соняшнику та модель для розрахунків метричні характеристики насінини

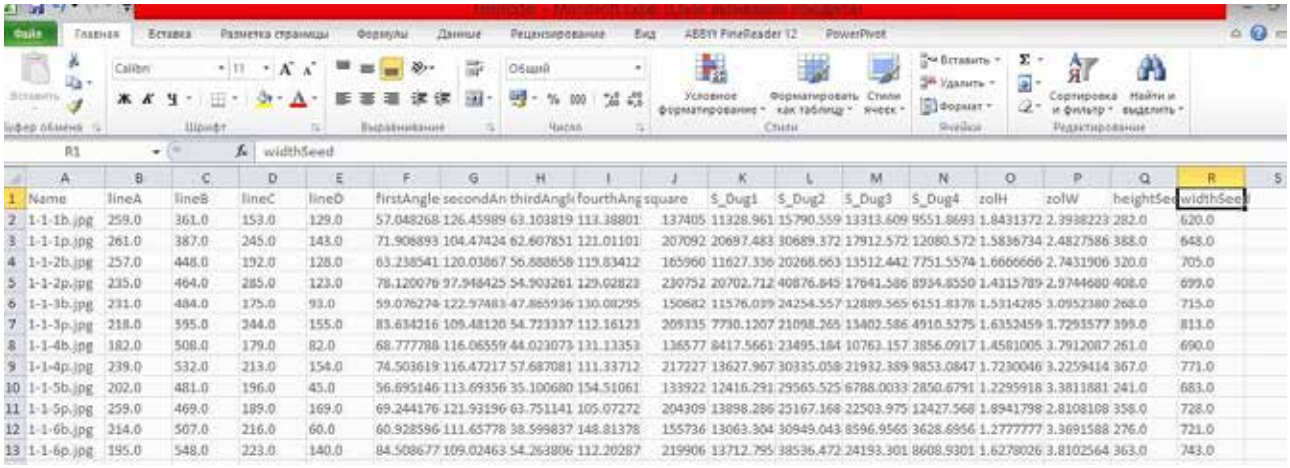
кут; thirdAngle – треттй кут; fourthAngle – четвертий кут; square – площа насінини в пікселях; S_Dug1 – площа 1 ділянки за межами трикутника; S_Dug2 – площа 2 ділянки за межами трикутника; S_Dug3 – площа 3 ділянки за межами трикутника; S_Dug4 – площа 4 ділянки за межами трикутника; zolH- Співвідношення половинок ширини; zolW- співвідношення половинок довжини; heightSeed – ширина насінини; widthSeed – довжина насінини.

Обчислення відбуваються у циклі, для обробки усіх фото, поміщених у теку програми. Фрагмент таблиці, що містить інформацію про метричні характеристики насінини, представлено на рис. 2.

Для отримання адекватної роботи програми усі наявні фото повинні мати формат JPG, назву латиницею, або у вигляді цифр. Фотозображення розміщуються в одну теку з програмою. Файл результатів result.csv розміщується в цій же теці. Цей результат можна відкривати та обробляти у будь яких табличних редакторах та статистичних програмах, зокрема у Excel.

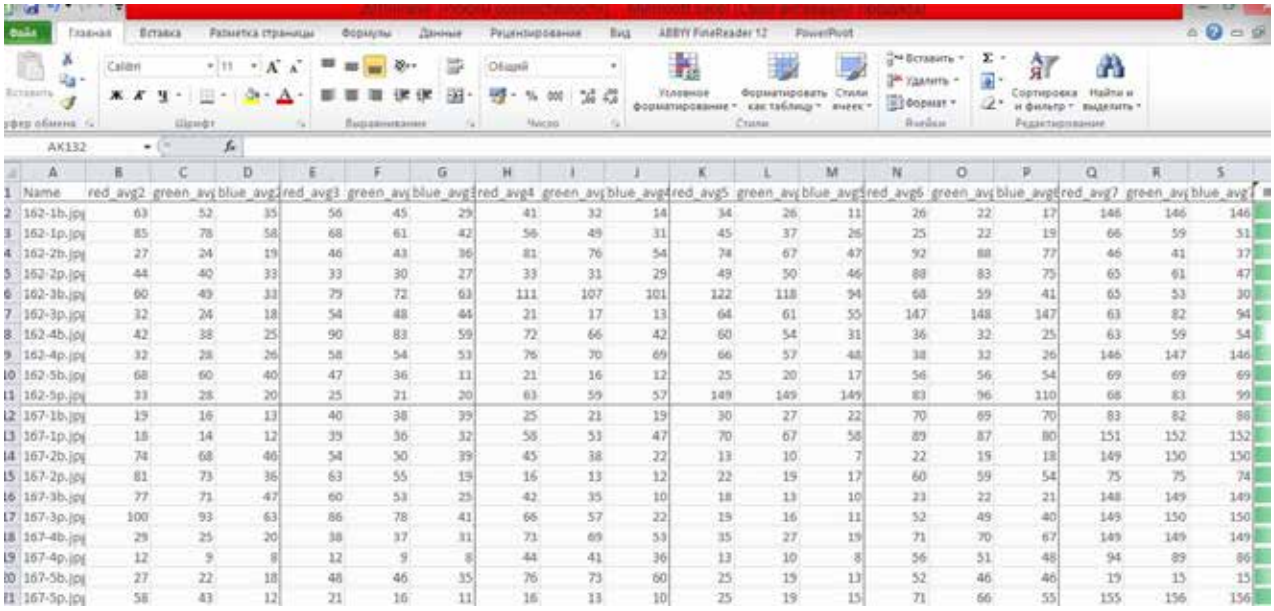
Програма визначення забарвлення насіння з можливістю розпізнавати кількість смуг.

Насіння соняшнику, в основному, має смугасте забарвлення. Смуги бувають дуже контрастні і близькі за забарвленням. На фото зазвичай спостерігається невелика кількість смуг від 3 до 7. Для оцінки було прийнято рішення аналізувати вісім діапазонів. Принцип визначення смуг полягає в аналізі забарвлення смужки пікселей, розташованих за шириною насінини. З наявної смужки пікселей виділяють найбільші різниці між двома сусідніми пікселями. Згідно



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Name	lineA	lineB	lineC	lineD	firstAngle	secondAn	thirdAngl	fourthAng	square	S_Dug1	S_Dug2	S_Dug3	S_Dug4	zolH	zolW	heightSee	widthSee
2	1-1-1b.jpg	259.0	361.0	153.0	129.0	57.048268	126.45989	63.103819	113.38801	137405	11328.961	15790.559	13313.609	9551.8693	1.8431372	2.3938223	282.0	620.0
3	1-1-1p.jpg	261.0	387.0	245.0	143.0	71.908893	104.47424	62.607851	121.01101	207092	20697.483	30689.372	17912.572	12080.572	1.5836734	2.4827586	388.0	648.0
4	1-1-2b.jpg	257.0	448.0	192.0	128.0	63.238541	120.03867	56.888658	119.83412	165960	11627.336	20268.663	13512.442	7751.5574	1.6666666	2.7431906	320.0	705.0
5	1-1-2p.jpg	235.0	464.0	285.0	123.0	78.120076	97.948425	54.903261	129.02823	230752	20702.712	40876.845	17641.586	8934.8550	1.4315789	2.9744680	408.0	699.0
6	1-1-3b.jpg	231.0	484.0	175.0	93.0	59.076274	122.97483	47.865936	130.08295	150682	11576.039	24254.557	12889.565	6151.8378	1.5314285	3.0952380	268.0	715.0
7	1-1-3p.jpg	218.0	595.0	244.0	155.0	83.634216	109.48120	54.723337	112.16123	209335	7730.1207	21098.265	13402.586	4910.5275	1.6352459	3.7293577	399.0	813.0
8	1-1-4b.jpg	182.0	508.0	179.0	82.0	68.777788	116.06559	44.023073	131.13353	136577	8417.5661	23495.184	10763.157	3856.0917	1.4581005	3.7912087	261.0	690.0
9	1-1-4p.jpg	239.0	532.0	213.0	154.0	74.503619	116.47217	57.687081	111.33712	217227	13627.967	30335.058	21932.389	9853.0847	1.7230046	3.2259414	367.0	771.0
10	1-1-5b.jpg	202.0	481.0	196.0	45.0	56.695146	113.69356	35.100680	154.51061	133922	12416.291	29565.525	6788.0033	2850.6791	1.2295918	3.3811881	241.0	683.0
11	1-1-5p.jpg	259.0	469.0	189.0	169.0	69.244176	121.93196	63.751141	105.07272	204309	13898.286	25167.168	22503.975	12427.568	1.8941798	2.8108108	358.0	728.0
12	1-1-6b.jpg	214.0	507.0	216.0	60.0	60.928596	111.65778	38.599837	148.81378	155736	13063.304	30949.043	8596.9565	3628.6956	1.2777777	3.3691588	276.0	721.0
13	1-1-6p.jpg	195.0	548.0	223.0	140.0	84.508677	109.02463	54.263806	112.20287	219906	13712.795	38536.472	24193.301	8608.9301	1.6278026	3.8102564	363.0	743.0

Рис. 2. Табличний результат роботи програми визначення метричних складових насіння соняшнику



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Name	red_avg2	green_avg	blue_avg2	red_avg3	green_avg	blue_avg3	red_avg4	green_avg	blue_avg4	red_avg5	green_avg	blue_avg5	red_avg6	green_avg	blue_avg6	red_avg7	green_avg	blue_avg7
2	162-1b.jpg	63	52	35	56	45	29	41	32	14	34	26	11	26	22	17	146	146	146
3	162-1p.jpg	85	78	58	68	61	42	56	49	31	45	37	26	25	22	19	66	59	51
4	162-2b.jpg	27	24	19	46	43	36	81	76	54	74	67	47	92	88	77	46	41	37
5	162-2p.jpg	44	40	33	33	30	27	33	31	29	49	50	46	88	83	75	65	61	47
6	162-3b.jpg	60	49	33	79	72	63	111	107	101	122	118	94	68	59	41	65	53	30
7	162-3p.jpg	32	24	18	54	48	44	21	17	13	64	61	55	147	148	147	63	82	94
8	162-4b.jpg	42	38	25	90	83	59	72	66	42	60	54	31	36	32	25	63	59	54
9	162-4p.jpg	32	28	26	58	54	53	76	70	69	66	57	48	38	32	26	146	147	146
10	162-5b.jpg	68	60	40	47	36	11	21	16	12	25	20	17	56	56	54	69	69	69
11	162-5p.jpg	33	28	20	25	21	20	63	59	57	149	149	149	83	96	110	68	83	99
12	167-1b.jpg	19	16	13	40	38	39	25	21	19	30	27	22	70	69	70	83	82	88
13	167-1p.jpg	18	14	12	39	36	32	58	53	47	70	67	58	89	87	80	151	152	152
14	167-2b.jpg	74	68	46	54	50	39	45	38	22	13	10	7	22	19	18	149	150	150
15	167-2p.jpg	81	73	36	63	55	19	16	13	12	22	19	17	60	59	54	75	75	74
16	167-3b.jpg	77	71	47	60	53	25	42	35	10	18	13	10	23	22	21	148	149	149
17	167-3p.jpg	100	93	63	86	78	41	66	57	22	19	16	11	52	49	40	149	150	150
18	167-4b.jpg	29	25	20	38	37	31	73	69	53	35	27	19	71	70	67	149	149	149
19	167-4p.jpg	12	9	8	12	9	8	44	41	36	13	10	8	56	51	48	94	89	86
20	167-5b.jpg	27	22	18	48	46	35	76	73	60	25	19	13	52	46	46	19	15	15
21	167-5p.jpg	58	43	12	21	16	11	16	13	10	25	19	15	71	66	55	155	156	156

Рис. 3. Табличний результат роботи програми визначення забарвлення смужок насінини соняшнику

цієї різкої зміни значень забарвлення виділяють вісім діапазонів, для кожного з яких вираховується середнє значення в кольоровій моделі RGB. Фрагмент таблиці, що містить інформацію про забарвлення смужок насінини соняшнику, представлено на рис. 3.

Застосування розробленого програмного забезпечення дозволяє отримати інформацію про фенотипові ознаки насінини за фотозображенням, в який описані розміри, форма та забарвлення кожної насінини. В рамках роботи лабораторії генетики та генетичних ресурсів фахівцями Інституту олійних культур НААН України отримана інформація доповнюється даними за іншими вимірами, більша частина яких проводиться і записується при зважені насінини, зважені після обрушення лушпиння та ін. (ДСТУ 8836:2019). На даний час таблиця, яку отримано за допомогою цих програм, містить 34 виміри насінини. До цієї інформації долучають таблиці морфологічного опису рослин з ще 34 вимірами. Долучення нових параметрів, визначення їх зав'язків та подальше використання цих програм дозволить проводити успішний селекційний добір рослин соняшнику та створювати нові лінії, сорти та гібриди соняшнику різних якостей. Цей інструмент допоможе вирішувати завдання пошуку та створення генотипів з крупним насінням, завданого забарвлення, форми та інші.

Наукова новизна полягає в тому що вперше було розроблено програми для аналізу

зображення насіння соняшнику з урахуванням біологічного різноманіття насіння. Вони вперше дали змогу описати досі якісні ознаки насіння у якості таблиць кількісних показників.

Проведені експерименти продемонстрували високу роздільність розробленого методу. Виділено контрастні зразки за розмірами та забарвленням насіння. Представлені розробки мають обмеження пов'язані з об'єктами дослідження та обладнанням для отримання фото. Ці обмеження можна зняти додатковим прилаштуванням при виникненні необхідності як програмного так і технічного характеру.

Проведені дослідження і створені інструменти базуються на точних математичних обчисленнях, без використання штучного інтелекту, як у інших дослідників (Rosten, 2006; Олешко, 2020; Triki, 2022; Гетьман, 2022; Старчак, 2023). У дослідженні Інституту олійних культур НААН України використовується колекція різного генетичного матеріалу з встановленими генами, які обумовлюють забарвлення насіння і з відомими донорами великого розміру насінин. Це дає змогу виявити закономірності морфології до генотипу, встановити межі генетичної мінливості з застосуванням точного математичного розрахунку. Лише подальше накопичення інформації і точних даних про насінини дасть змогу залучати штучний інтелект. Тому що навчати не зрозумівши причинно наслідкові зв'язки і межі мінливості, не встановивши дійсну різницю між фенотипами різних генотипів для селекціонерів не має сенсу. Таке навчання усього лише створить ще одного штучного «селекціонера» з власною думкою, яка обґрунтована лише досвідом (кількістю фото), а не знаннями (математичним доведенням відмінності якостей).

Висновки. У результаті проведеного дослідження розроблено модель для вимірювання метричних характеристик та форми насінини, а також програмне забезпечення для фенотипування насінини соняшнику за фотозображеннями, а саме визначення метричних характеристик, форми та забарвлення насінини. Розроблене програмне забезпечення базується на точному математичному обрахунку морфологічних ознак насінини соняшнику за фотозображеннями, що забезпечує точність створених описів. Зараз розробки вже використовуються селекціонерами та генетиками Інституту олійних культур НААН України для наукової селекційної роботи.

ЛІТЕРАТУРА:

1. Алієв Е. Б. Автоматичне фенотипування насіннєвого матеріалу соняшнику: монографія/ Алієв Е. Б. Київ: Аграрна наука, 2022. 104 с.

2. Алієв Е. Б. Розробка пристрою для автоматичного фенотипування насіннєвого матеріалу соняшнику. *Machinery & Energetics. Journal of Rural Production Research,* 2019. Вип.10. № 1. С. 11–17. DOI: 10.31548/machenergy.2019.01.011-017

3. Ведмедєва К. В. Маркерні морфологічні ознаки соняшнику: ідентифікація, успадкування та ефекти алелів генів: монографія/ Ведмедєва К. В. Аграрна наука, Київ. 2023. 240 с. ISBN 978-966-540-572-6. DO I: 10.31073/978-966-540-572-6.

4. Гетьман І., Держевецька М., Бауліна Т., Кухтік Т., Соломко Т. Проект програмного комплексу для реалізації додатку для розпізнавання лікарських рослин. *ITSynergy*, 2022. № 1. С. 6–25. https://doi.org/10.53920/ITS-2022-1-1

5. ДСТУ 8836:2019 Насіння олійних культур. Методи визначення вмісту лушпиння [Чинний від 2019-26-11]. Нац. Стандарт України. Кіїв: Держспоживстандарт України. 2019: 7. http://online.budstandart.com/ua/catalog/doc-page?id_doc=82148 (In Ukrainian)

6. Махова Т. В., Ведмедєв С. Р., Поляков О. І. Створення баз даних фенотипових ознак та добір ліній соняшнику особливого призначення. *Науково-технічний бюлетень Інституту олійних культур НААН.* 2023. № 35, P.51–62. DOI: 10.36710/IOC-2023-35-05.

7. Олешко Т. І., Квашук Д. М., Якименко А. М. Сучасні підходи до аналізу зображень в системах ідентифікації захворювань рослин із застосуванням детектору FAST. *Наукоємні технології.* 2020. № 1(45). С. 85–91.

8. Старчак О. В., Коротєєва Т. О. Класифікація якості насіння соняшника за його зображеннями із використанням нейронних мереж. *Науковий вісник НЛТУ України.* 2023. № 33(6), С. 69–75.

9. Bakurova A., Vedmedeva K., Vedmedev S., Tereschenko E. Ontological Model of Helianthus Cultivation in Ukrainian Conditions. *CEUR Workshop Proceedings This link is disabled.* 2023. № 3396, P. 130–140.

10. Borges L. M., Reis V. C., Izbicki R. Schrodinger's phenotypes: Herbarium specimens show two-dimensional images are both good and (not so) bad sources of morphological data. *Methods in Ecology and Evolution*, 2020. 11(10), P. 1296–1308. doi:10.1111/2041-210x.13450.

11. Fiorani F., Schurr U. Future scenarios for plant phenotyping. *Annu. Rev. Plant Biol.* 2013. V.64. № 1, P. 267–291.

12. Gorohivets N. A., Vedmedeva E. V. Inheritance of epidermis pigmentation in sunflower achenes. *Cytology and Genetics*, 2016. Вип. *50. №* 2. С. 116–120.

13. Kumar P., Huang C., Cai J., Miklavcic S. J. Root phenotyping by root tip detection and classification through statistical learning. *Plant Soil.* 2014. № 380. P.193–209. doi: 10.1007/s11104-014-2071-3.

14. Pieruschka R., Schurr U. Plant Phenotyping: Past, Present, and Future. *Plant Phenomics.* 2019. № 7507131. P. 1–6. doi: 10.34133/2019/7507131

15. Plants guidelines for the conduct of tests for distinctness, uniformity and stability common sunflower UPOV Code(s): HLNTS_ANN *Helianthus annuus L*. URL: https://www.upov.int/edocs/mdocs/upov/en/twa_47/tg_81_7_ proj_1.pdf.

16. Rosten E., Drummond T. Machine Learning for High-Speed Corner Detection. ECCV (1). *Springer. of Lecture Notes in Computer Science*. 2006. Vol. 3951. P. 430–443.

17. Shibayama M., Sakamoto T., Takada E., Inoue A., Morita K., Takahashi W., Kimura, A. Estimating paddy rice leaf area index with fixed point continuous observation of near infrared reflectance using a calibrated digital camera. *Plant Product. Sci.*, 2011. № 14, P. 30–46. doi: 10.1626/pps.14.30

18. SOCIETY OF AMBIENT INTELLIGENCE 2023. URL: https://www.isc-sai.org/

19. Triki A., Bouaziz B., Mahdi W. A deep learning-based approach for detecting plant organs from digitized herbarium specimen images. *Ecological Informatics*, 2022, № 69. doi:10.1016/j.ecoinf.2022.101590.

20. Walker B. E., Tucker A., Nicolson N. Harnessing Large-Scale Herbarium Image Datasets Through Representation Learning. *Frontiers in Plant Science.* 2022. № 12. doi:10.3389/fpls.2021.806407.

21. Xiao S., Chai H., Shao K., Shen M., Wang Q., Wang R., Sui Y., Ma Y. Image-Based Dynamic Quantification of Aboveground Structure of Sugar Beet in Field. *Remote Sens.* 2020. 12, P. 269. doi: 10.3390/rs12020269.

22. Yang S., Zheng L., He P., Wu T., Sun S., Wang M. High-throughput soybean seeds phenotyping with convolutional neural networks and transfer learning. *Plant Methods.* 2021. V.17. № 50, P. 1–17. doi: 10.1186/s13007-021-00749-y.

REFERENCES:

1. Aliyev, E. B. (2022). Avtomatychne fenotypuvannya nasinnyevoho materialu sonyashnyku. Kyyiv: Ahrarna nauka, 104 c. [in Ukrainian]

2. Aliyev, E. B. (2019). Rozrobka prystroyu dlya avtomatychnoho fenotypuvannya nasinnyevoho materialu sonyashnyku. *Machinery & Energetics. Journal of Rural Production Research,* 10, 1, 11–17. DOI: 10.31548/machenergy.2019.01.011-017 [in Ukrainian]

3. Vedmedeva, K. V. (2023). Markerni morfolohichni oznaky sonyashnyku: identyfikatsiya, uspadkuvannya ta efekty aleliv heniv. Ahrarna nauka, Kyyiv, 240 c. ISBN 978-966-540-572-6. DOI: 10.31073/978-966-540-572-6. [in Ukrainian]

4. Het'man, I., Derzhevets'ka, M., Baulina, T., Kukhtik, T., & Solomko, T. (2022). Proekt prohramnoho kompleksu dlya realizatsiyi dodatku dlya rozpiznavannya likars'kykh roslyn. *ITSynergy*, (1), 6–25. https://doi.org/10.53920/ITS-2022-1-1 [in Ukrainian]

5. DSTU 8836:2019 Nasinnya oliynykh kul'tur. Metody vyznachennya vmistu lushpynnya [Chynnyy vid 2019- 26-11]. Nats. standart Ukrayiny. Kyyiv: Derzhspozhyvstandart Ukrayiny. 2019: 7. Retrieved from: http://online.budstandart.com/ua/catalog/doc-page?id_doc=82148 [in Ukrainian]

6. Makhova, T. V., Vedmedyev, S. R., Polyakov, O. I. (2023). Stvorennya baz danykh fenotypovykh oznak ta dobir liniy sonyashnyku osoblyvoho pryznachennya. *Naukovo-tekhnichnyy byuleten' Instytutu oliynykh kul'tur NAAN*, 35, 51–62. DOI: 10.36710/IOC-2023-35-05 [in Ukrainian]

7. Oleshko, T. I., Kvashuk, D. M., Yakymenko, A. M. (2020). Suchasni pidkhody do analizu zobrazhen' v systemakh identyfikatsiyi zakhvryuvan' roslyn iz zastosuvannyam detektoru FAST. Naukoyemni tekhnolohiyi NAU 1(45), 85–91. [in Ukrainian]

8. Starchak, O. V., Korotyeyeva, T. O. (2023). Klasyfikatsiya yakosti nasinnya sonyashnyka za yoho zobrazhennyamy iz vykorystannyam neyronnykh merezh. Naukovyy visnyk NLTU Ukrayiny, 33(6), 69–75. [in Ukrainian]

9. Bakurova, A., Vedmedeva, K., Vedmedev, S., Tereschenko, E. (2023). Ontological Model of Helianthus Cultivation in Ukrainian Conditions CEUR Workshop ProceedingsThis link is disabled., 3396, pp. 130–140

10. Borges, L. M., Reis, V. C., Izbicki, R. (2020). Schrodinger's phenotypes: Herbarium specimens show two-dimensional images are both good and (not so) bad sources of morphological data. Methods in Ecology and Evolution, 11(10), 1296–1308. doi:10.1111/2041-210x.13450.

11. Fiorani, F., Schurr, U. (2013). Future scenarios for plant phenotyping. Annu. Rev. Plant Biol, 64(1), 267–291.

12. Gorohivets, N. A., Vedmedeva, E. V. (2016). Inheritance of epidermis pigmentation in sunflower achenes. Cytology and Genetics, 50(2), 116–120.

13. Kumar, P., Huang, C., Cai, J., Miklavcic, S. J. (2014). Root phenotyping by root tip detection and classification through statistical learning. Plant Soil, 380, 193–209. doi: 10.1007/s11104-014-2071-3.

14. Pieruschka, R., Schurr, U. (2019). Plant Phenotyping: Past, Present, and Future. Plant Phenomics, 7507131: 1–6. doi: 10.34133/2019/7507131

15. Plants guidelines for the conduct of tests for distinctness, uniformity and stability common sunflower UPOV Code(s): HLNTS_ANN Helianthus annuus L. Retrieved from: https://www.upov.int/edocs/mdocs/upov/en/twa_47/tg_81_7_ proj_1.pdf.

16. Rosten, E., Drummond, T. (2006). Machine Learning for High-Speed Corner Detection. ECCV (1). Springer of Lecture Notes in Computer Science, 3951, 430–443.

17. Shibayama, M., Sakamoto, T., Takada, E., Inoue, A., Morita, K., Takahashi, W., Kimura, A. (2011). Estimating paddy rice leaf area index with fixed point continuous observation of near infrared reflectance using a calibrated digital camera. Plant Product. Sci., 14, 30–46. doi: 10.1626/pps.14.30

18. SOCIETY OF AMBIENT INTELLIGENCE (2023). Retrieved from: https://www.isc-sai.org/

19. Triki, A., Bouaziz, B., Mahdi, W. (2022). A deep learning-based approach for detecting plant organs from digitized herbarium specimen images. Ecological Informatics, 69. doi:10.1016/j.ecoinf.2022.101590.

20. Walker, B. E., Tucker, A., Nicolson, N. (2022). Harnessing Large-Scale Herbarium Image Datasets Through Representation Learning. Frontiers in Plant Science, 12. doi:10.3389/fpls.2021.806407.

21. Xiao, S., Chai, H., Shao, K., Shen, M., Wang, Q., Wang, R., Sui, Y., Ma, Y. (2020). Image-Based Dynamic Quantification of Aboveground Structure of Sugar Beet in Field. Remote Sens, 12, 269. doi: 10.3390/rs12020269.

22. Yang, S., Zheng, L., He, P., Wu, T., Sun, S., Wang, M. (2021). High-throughput soybean seeds phenotyping with convolutional neural networks and transfer learning. Plant Methods, 17, 50, 1–17. doi: 10.1186/s13007-021-00749-y.

UDC 004.94
DOI https://doi.org/10.32782/IT/2024-4-8

Vyacheslav GOREV
Candidate of Physical and Mathematical Sciences, Associate Professor, Head of the Department of Physics, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, Gorev.V.M@nmu.one
ORCID: *0000-0002-9528-9497*
Scopus-Author ID: *55047688400*

Yana SHEDLOVSKA
Candidate of Technical Sciences, Associate Professor at the Department of Information Technology and Computer Engineering, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, shedlovska.y.i@nmu.one
ORCID: *0000-0003-4931-4070*
Scopus Author ID: *57191853306*

Ivan LAKTIONOV
Doctor of Technical Sciences, Associate Professor, Professor at the Department of Software Engineering, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, laktionov.i.s@nmu.one
ORCID: *0000-0001-7857-6382*
Scopus Author ID: *57194557735*

Grygorii DIACHENKO
Candidate of Technical Sciences, Associate Professor at the Department of Electric Drive, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, diachenko.g@nmu.one
ORCID: *0000-0001-9105-1951*
Scopus Author ID: *57201252081*

To cite this article: Gorev, V., Shedlovska, Ya., Laktionov, I., Diachenko, G. (2024). LMS prohnozuvannia ta filtratsiia Kalmana dlia ymovirnosti vynyknennia khvorob kukurudzy [LMS Prediction and Kalman Filtering for the Probability of the Maize Disease Occurrence]. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 61–67, doi: https://doi.org/10.32782/IT/2024-4-8

LMS PREDICTION AND KALMAN FILTERING FOR THE PROBABILITY OF THE MAIZE DISEASE OCCURENCE

The paper is devoted to an urgent agriculture problem such as the maize disease occurrence prediction. The data for the probability of the maise disease occurrence used in the paper are taken from the dataset of professional weather stations from Metos by Pessl Instruments using the FieldClimate IoT platform, access to which is provided by Metos Ukraine LLC. The data are collected from September 2022 to September 2023 with a 1-hour interval for the Dnipropetrovsk region. In our recent papers the data prediction is made on the basis of the neural networks. It should be stressed that the data contain discontinuities, so the question occurs whether the data smoothing may enhance the prediction. In our recent papers we used the smoothing algorithm which is based on an artificial exponential decay which is used in order to smooth the data discontinuities. However, such an approach is rather artificial one. So a question occurs whether a standard smoothing algorithm, for example such as the Kalman one, may enhance the data prediction. For simplicity, in this paper we restrict ourselves to the investigation of the LMS prediction of the smoothed and non-smoothed data. It is shown that the Kalman filtering may slightly enhance some metrics of the LMS prediction. The corresponding investigation for the neural network prediction may be a plan for the future.

The aim of the work *is to investigate the LMS prediction of the probability of the maize disease occurrence in the case where the data are non-smoothed and in the case where the data are smoothed on the basis of the Kalman algorithm.*

The methodology *consists in the Kalman filtering and the LMS prediction of the data for the probability of the maize disease occurrence.*

The scientific novelty *consists in the use of the Kalman filtering of the data for the probability of the maize disease occurrence in order to enhance the data prediction.*

The conclusions *are as follows. The data Kalman filtering may slightly enhance some metrics of the LMS prediction.*

Key words: *Kalman filtering, LMS prediction, probability of the maize disease occurrence.*

В'ячеслав ГОРЄВ
кандидат фізико–математичних наук, доцент, завідувач кафедри фізики, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького 19, Дніпро, Україна, 49005
ORCID: *0000-0002-9528-9497*
Scopus-Author ID: *55047688400*

Яна ШЕДЛОВСЬКА
кандидат технічних наук, доцент кафедри інформаційних технологій та комп'ютерної інженерії, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького 19, Дніпро, Україна, 49005
ORCID: *0000-0003-4931-4070*
Scopus Author ID: *57191853306*

Іван ЛАКТІОНОВ
доктор технічних наук, доцент, професор кафедри програмного забезпечення комп'ютерних систем, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького 19, Дніпро, Україна, 49005
ORCID: *0000-0001-7857-6382*
Scopus Author ID: *57194557735*

Григорій ДЯЧЕНКО
кандидат технічних наук, доцент кафедри електропривода, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького 19, Дніпро, Україна, 49005
ORCID: *0000-0001-9105-1951*
Scopus Author ID: *57201252081*

Бібліографічний опис статті: Горєв, В, Шедловська, Я., Лактіонов, І., Дяченко, Г. (2024). LMS прогнозування та фільтрація Калмана для ймовірності виникнення хвороб кукурудзи. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 61–67, doi: https://doi.org/10.32782/IT/2024-4-8

LMS ПРОГНОЗУВАННЯ ТА ФІЛЬТРАЦІЯ КАЛМАНА ДЛЯ ЙМОВІРНОСТІ ВИНИКНЕННЯ ХВОРОБ КУКУРУДЗИ

Стаття присвячена такій актуальній проблемі сільського господарства, як прогнозування захворювання кукурудзи. Використані в роботі дані для ймовірності виникнення хвороби кукурудзи взяті з набору даних професійних метеостанцій Metos від Pessl Instruments за допомогою платформи FieldClimate IoT, доступ до якої надає ТОВ «Метос Україна». Дані збирались з вересня 2022 року по вересень 2023 року з інтервалом в 1 годину для Дніпропетровської області. У наших останніх роботах прогнозування даних зроблено на основі нейронних мереж. Слід підкреслити, що дані містять розриви, тому виникає питання, чи може згладжування даних покращити прогнозування. У наших нещодавніх роботах ми використовували алгоритм згладжування, який базується на штучному експоненційному спаді, який використовується для згладжування розривів даних. Однак такий підхід є досить штучним. Тому виникає питання, чи може стандартний алгоритм згладжування, наприклад, такий як алгоритм Калмана, покращити прогнозування даних. Для простоти в цій статті ми обмежимося дослідженням LMS прогнозуванням для згладжених та незгладжених даних. Показано, що фільтрація Калмана може трохи покращити деякі показники LMS прогнозування. Відповідне дослідження для прогнозування на основы нейронних мереж може бути планом на майбутнє.

Метою роботи є *дослідити LMS прогнозування ймовірності виникнення хвороби кукурудзи у випадку незгладжених даних та у випадку згладжених даних на основі алгоритму Калмана.*

Методологія *полягає у фільтрації Калмана та LMS прогнозуванні даних для ймовірності виникнення хвороби кукурудзи.*

Наукова новизна *полягає у використанні фільтрації Калмана даних для ймовірності виникнення хвороби кукурудзи з метою покращення прогнозування даних.*

Висновки *є такими. Фільтрація даних на основі алгоритму Калмана може дещо покращити деякі показники LMS прогнозування.*

Ключові слова: *Фільтрація Калмана, LMS прогнозування, ймовірність виникнення хвороби кукурудзи.*

Introduction. An urgent agriculture problem devoted to the prediction of the probability of the maize disease occurrence is considered in this paper. We use the corresponding data taken from the dataset of professional weather stations from Metos by Pessl Instruments using the FieldClimate IoT platform, access to which is provided by Metos Ukraine LLC, see (Laktionov et al, 2023; Diachenko et al, 2024). The data under investigation contain discontinuities, so the data may be smoothed in order to enhance the prediction. In paper (Diachenko et al, 2024) the use of exponentially decreasing functions is proposed to describe the data smoothing in the vicinity of the discontinuities. However, such a trick is rather artificial one, so in this paper the data smoothing based on the Kalman filtering is considered. Kalman filtering is rather popular for the prediction improvement; see, for example, (Lai et al, 2019). In papers (Laktionov et al, 2023; Diachenko et al, 2024) neural networks were used in order to obtain the data prediction. For simplicity, in this paper we restrict ourselves to the use of the adaptive LMS prediction method, which is rather popular in the literature, see, for example (Tajdari, 2021; Prasetyowati et al, 2021). Both the prediction of smoothed and non-smoothed data are investigated.

Investigation of the LMS prediction of smoothed and non-smoothed data. The data under consideration, see (Laktionov et al, 2023; Diachenko et al, 2024), is shown in Fig. 1 where p is the probability of the maize disease occurrence and n is the number of a point; the data are collected from September 2022 to September 2023 with a 1-hour interval for the Dnipropetrovsk region. The data contains more than 8000 points.

As is known, the algorithm of the Kalman filter is as follows, see (Lai et al, 2019; Gusev et al, 2019). Let us suppose that physically the data obey the equation

$$p_n = ap_{n-1} + bu_{n-1} + \xi_{n-1} \tag{1}$$

where ξ_n is a random process which describes the random character of system evolution, u_n is the known quantity which governs the system evolution, a, b are known constants. Let us suppose that the data are measured by a noisy device which readings are

$$z_n = cp_n + \eta_n \tag{2}$$

where η_n is a random process which describes the device error and c is the gain coefficient. The filter error, ξ_n and η_n are supposed to be independent stationary processes.

Let us denote the smoothed data as p^{opt}. At the first step we put

$$p_1^{\text{opt}} = p_1, \ \left\langle e_1^2 \right\rangle = \sigma_\eta^2 \tag{3}$$

where $\left\langle e_n^2 \right\rangle$ is the mean square error of the Kalman filter and σ_η^2 is the variance of the process η_n.

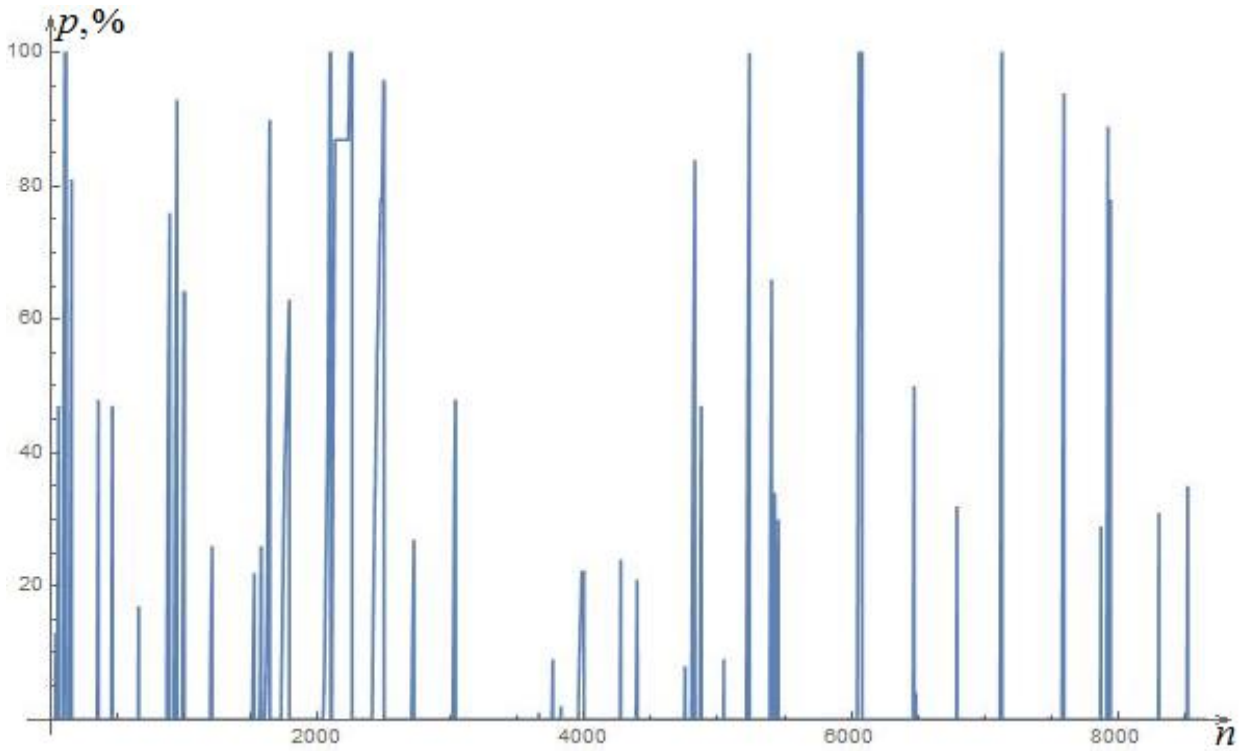


Fig. 1. The data for the probability of the maize disease occurrence

Then at each step the following recurrent formulas are used:

$$p_n^{\text{opt}} = K_n p_n + \left(1 - cK_n\right)\left(ap_{n-1}^{\text{opt}} + bu_{n-1}\right),$$

$$K_n = c\frac{a^2\left\langle e_{n-1}^2\right\rangle + \sigma_\xi^2}{\sigma_\eta^2 + c^2a^2\left\langle e_{n-1}^2\right\rangle + c^2\sigma_\xi^2},$$

$$\left\langle e_n^2\right\rangle = \sigma_\eta^2\frac{a^2\left\langle e_{n-1}^2\right\rangle + \sigma_\xi^2}{\sigma_\eta^2 + c^2a^2\left\langle e_{n-1}^2\right\rangle + c^2\sigma_\xi^2} \quad (4)$$

where σ_ξ^2 is the is the variance of the process ξ_n. According to (Lai et al, 2019), the following parameter choice should be made: the gain coefficient $c = 1$, the physically process does not change its state at each step, so $a = 1$, $b = 0$. In such a case the expressions (4) read as

$$p_n^{\text{opt}} = K_n p_n + \left(1 - K_n\right)p_{n-1}^{\text{opt}},$$

$$K_n = \frac{\left\langle e_{n-1}^2\right\rangle + \sigma_\xi^2}{\sigma_\eta^2 + \left\langle e_{n-1}^2\right\rangle + \sigma_\xi^2},\ \left\langle e_n^2\right\rangle = \sigma_\eta^2\frac{\left\langle e_{n-1}^2\right\rangle + \sigma_\xi^2}{\sigma_\eta^2 + \left\langle e_{n-1}^2\right\rangle + \sigma_\xi^2}. \quad (5)$$

In (Lai et al, 2019) it is stressed that σ_η^2 should be known on the basis of the device characteristics and σ_ξ^2 should be artificially chosen in order to obtain the lowest prediction errors. But in our case the data for the probability of the maize disease occurrence is calculated on the basis of a couple of measurements which are considered to be rather accurate. So, in this paper we artificially put $\sigma_\xi^2 = 1$ and we investigate under which values of σ_η^2 the prediction of the smoother process leads to the most reliable results.

In this paper we investigate a one-point-forward prediction based on the LMS adaptive algorithm. In fact, we provide the prediction algorithm for the smoothed data, but we consider the obtained result as a prediction for the non-smoothed data. The LMS algorithm is as follows, see, for example, (Tajdari, 2021; Prasetyowati et al, 2021). The step size μ and the algorithm depth N are chosen. Let us denote the predicted process as $\bar{p}$. The first N points of the predicted and actual process are considered to be the same and at the zero step the algorithm weights are considered to be zero ones:

$$\bar{p}_n = p_n^{\text{opt}},\ n = \overline{1,N};\ h_j^{(0)} = 0,\ j = \overline{1,N}. \quad (6)$$

Then a one-point-forward prediction at each step $n > N$ is as follows:

$$\bar{p}_n = \begin{cases} 0, \sum_{j=1}^{N} h_j p_{n-j}^{\text{opt}} < 0 \\ 100, \sum_{j=1}^{N} h_j p_{n-j}^{\text{opt}} > 100 \\ \sum_{j=1}^{N} h_j p_{n-j}^{\text{opt}}, \text{otherwise} \end{cases},$$

$$h_i^{(n)} = h_i^{(n-1)} + 2\mu\left(p_n^{\text{opt}} - \bar{p}_n\right)p_{n-i}^{\text{opt}},\ i = \overline{1,N}, \quad (7)$$

here we take into account the fact that obviously $\bar{p}_n \in [0,100]$. The following prediction metrics are investigated, see (Laktionov et al, 2023; Diachenko et al, 2024):

$$\text{RMSE} = \sqrt{\frac{1}{M-N}\sum_{j=N+1}^{M}\left(p_j - \bar{p}_j\right)^2},$$

$$\text{MAE} = \frac{1}{M-N}\sum_{j=N+1}^{M}\left|p_j - \bar{p}_j\right|,$$

$$R^2 = 1 - \frac{\sum_{j=N+1}^{M}\left(p_j - \bar{p}_j\right)^2}{\sum_{j=N+1}^{M}\left(p_j - \langle p\rangle\right)^2}, \quad (8)$$

where $M = 8658$ is the number of data points, and here it is taken into account that only the points with numbers $N+1$, $N+2$,…, M are predicted.

First of all, we investigate the case where $N \geq 2$. An exhaustive search of the parameters σ_η (in the range $\sigma_\eta \in [0,1]$ with the step equal to 10^{-1}), μ (in the range $\mu \in [10^{-7}; 2\cdot 10^{-5}]$ with the step equal to 10^{-7}) and N (in the range $N \in [2;20]$ with the step equal to 1) is provided in order to obtain the minimal MAE value. The corresponding results are given in Table 1.

Table 2

Minimal MAE values in the case N=1

σ_η	**Minimal MAE**	**Corresponding μ value**
0	0.823	$9.1\cdot 10^{-6}$
0.1	0.825	$9.2\cdot 10^{-6}$
0.2	0.828	$9.5\cdot 10^{-6}$
0.3	0.836	$9.9\cdot 10^{-6}$
0.4	0.847	$1.06\cdot 10^{-5}$
0.5	0.863	$1.12\cdot 10^{-5}$
0.6	0.883	$1.23\cdot 10^{-5}$
0.7	0.905	$1.38\cdot 10^{-5}$
0.8	0.930	$1.58\cdot 10^{-5}$
0.9	0.956	$1.89\cdot 10^{-5}$
1	0.982	$2.00\cdot 10^{-5}$

The value $\sigma_\eta = 0$ describes the non-smoothed data. So, as can be seen, the data smoothing enhances the corresponding prediction, for $\sigma_\eta = 0.7$ we obtain the minimal MAE value during all the above-mentioned exhaustive search, and the corresponding MAE value 0.884 is less than the MAE in the case of the non-smoothed data. In the case of the non-smoothed data ($\sigma_\eta = 0$, $N = 10$, $\mu = 8.8\cdot 10^{-6}$) we have $\text{MAE} = 0.919$, $\text{RMSE} = 5.22$ and $R^2 = 0.912$ while in the case where $\sigma_\eta = 0.7$, $N = 9$, $\mu = 1.45\cdot 10^{-5}$ we have $\text{MAE} = 0.884$, $\text{RMSE} = 5.14$ and $R^2 = 0.915$, so

the MAE and RMSE metrics are better for the considered smoothed data rather than for the non-smoothed data. The corresponding actual and predicted values for the probability of the maize disease occurrence are shown in Fig. 2

Let us investigate the case where $N = 1$. A similar exhaustive search of the parameters σ_η and μ is provided in order to obtain the minimal MAE value. The corresponding results are given in Table 2. As can be seen, the data smoothing does not enhance the MAE value. However, for example, for the case where $\sigma_\eta = 0.1$, $N = 1$, $\mu = 9.2 \cdot 10^{-6}$ we have $\mathrm{MAE} = 0.825$, $\mathrm{RMSE} = 4.995$ and $R^2 = 0.920$ while for the non-smoothed data ($\sigma_\eta = 0$, $N = 1$, $\mu = 9.1 \cdot 10^{-6}$) we have $\mathrm{MAE} = 0.823$, $\mathrm{RMSE} = 4.997$ and $R^2 = 0.920$, so the RMSE parameter is slightly better in the case where the data is smoothed. It should be stressed that in the paper we write the investigated values rounded off up to three significant digits except for the above-mentioned RMSE values. The corresponding values are given rounded off up to four significant digits in order to demonstrate their difference.

The actual and predicted data values for the case $\sigma_\eta = 0.1$, $N = 1$, $\mu = 9.2 \cdot 10^{-6}$ are shown in Fig. 3. As can be seen, the MAE and RMSE metrics are slightly better for the case $N = 1$ rather than for the case where $N \geq 2$. However, some peaks are better predicted in the case where $N \geq 2$ rather than in the case where $N = 1$.

Table 1

Minimal MAE values in the case N≥2

σ_η	**Minimal MAE**	**Corresponding μ value**	**Corresponding N value**
0	0.919	$8.8 \cdot 10^{-6}$	10
0.1	0.918	$8.9 \cdot 10^{-6}$	10
0.2	0.914	$9.3 \cdot 10^{-6}$	10
0.3	0.908	10^{-5}	10
0.4	0.900	$1.14 \cdot 10^{-5}$	9
0.5	0.893	$1.26 \cdot 10^{-5}$	9
0.6	0.886	$1.41 \cdot 10^{-5}$	9
0.7	0.884	$1.45 \cdot 10^{-5}$	9
0.8	0.888	$1.5 \cdot 10^{-5}$	9
0.9	0.899	$1.5 \cdot 10^{-5}$	9
1	0.919	$1.49 \cdot 10^{-5}$	9

Conclusions. The paper is devoted to an urgent agriculture problem of the prediction of the probability of the maize disease occurrence. In this paper for simplicity we restrict ourselves to the LMS prediction algorithm, and it shown that the Kalman smoothing of initial data may lead to the enhancement of some prediction metrics, in particular, the RMSE one. By comparison of the obtained prediction metrics we can conclude that the LMS prediction may work not worse than some

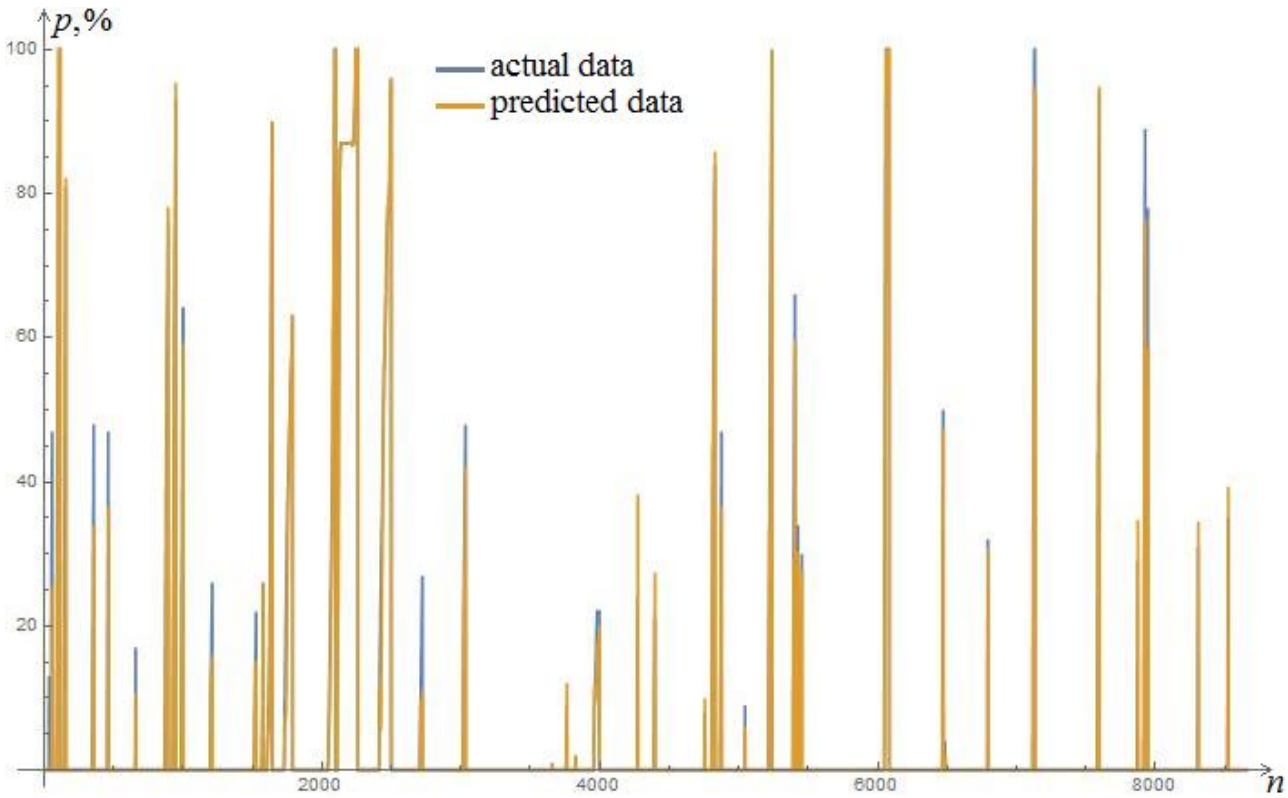


Fig. 2. Actual and predicted data for $\sigma_\eta = 0.7$, $N = 9$, $\mu = 1.45 \cdot 10^{-5}$

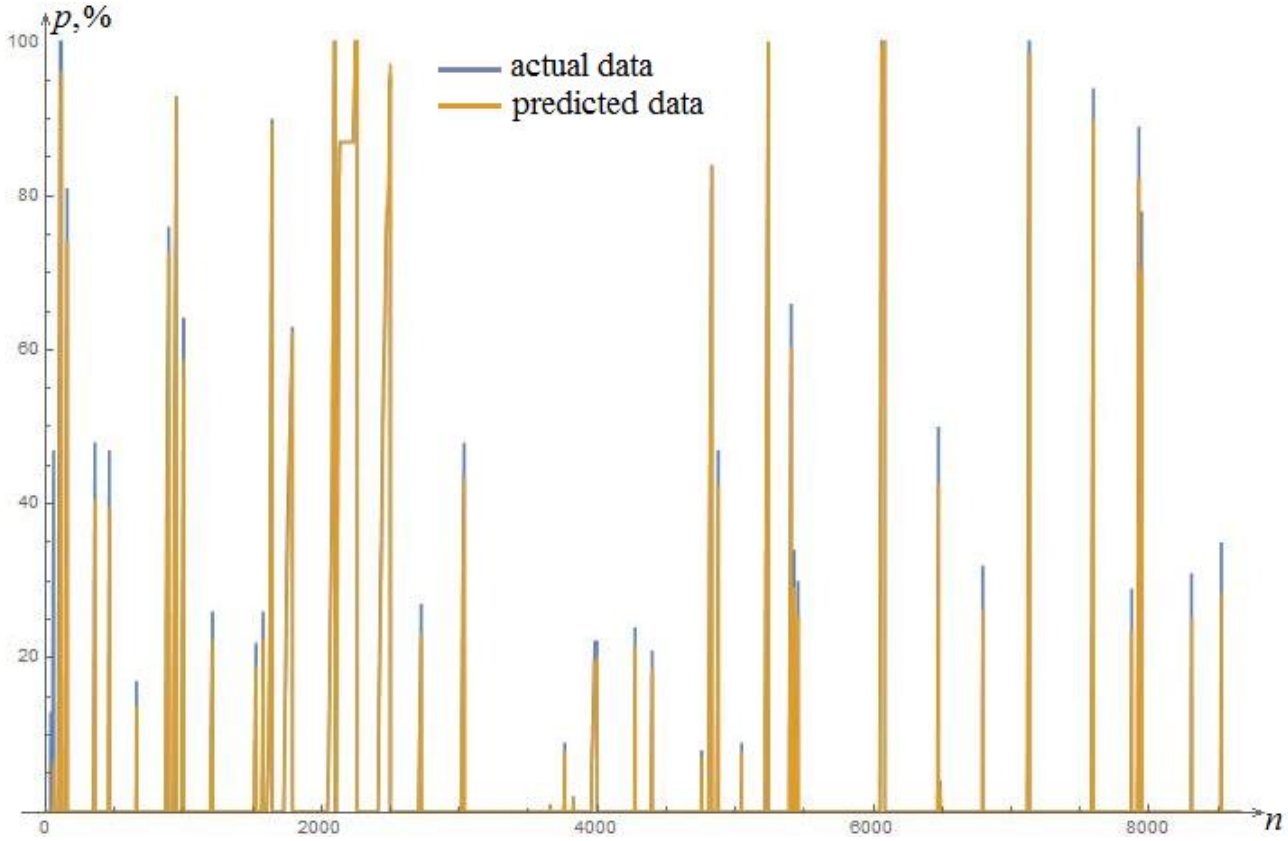


Fig. 3. Actual and predicted data for $\sigma_\eta = 0.1$**,** $N = 1$**,** $\mu = 9.2 \cdot 10^{-6}$

neural networks in some cases, see, for example, (Diachenko et al, 2024). However, the Random Forest neural network leads to better prediction results for the data under consideration rather than the LMS prediction, see (Diachenko et al, 2024). The corresponding investigation of the Random Forest prediction on the basis of the Kalman smoothing may be a plan for the future.

Acknowledgements. This research is carried out as part of the scientific project «Development of software and hardware of intelligent technologies for sustainable cultivation of agricultural crops in war and post-war times» funded by the Ministry of Education and Science of Ukraine at the expense of the state budget (State Registration No. 0124U000289).

BIBLIOGRAPHY:

1. Laktionov I., Diachenko G., Rutkowska D., Kisiel-Dorohinick M., An Explainable AI Approach to Agrotechnical Monitoring and Crop Diseases Prediction in Dnipro Region of Ukraine. *Journal of Artificial Intelligence and Soft Computing Research*. 2023. Vol. 13, issue 4, p. 247–272. doi: 10.2478/jaiscr-2023-0018

2. Diachenko G., Laktionov I., Vizniuk A., Gorev V., Kashtan V., Khabarlak K., Shedlovska Y., An Improved Approach to Prediction of Maize Disease Occurrence Based on Weather Monitoring and Machine Learning: a Case of the Forest-Steppe and Northern Steppe of Ukraine. *Engineering Applications of Artificial Intelligence*, 2024, under review.

3. Lai X., Yang T., Wang Z., Peng C. IoT Implementation of Kalman Filter to Improve Accuracy of Air Quality Monitoring and Prediction. *Applied Sciences*. 2019. Vol. 9, issue 9, 1831 (23 pages). doi: 10.3390/app9091831

4. Tajdari T. Adaptive method to predict an track unknown system behaviors using LMS and RLS algorithms. *Facta Universitatis, Series: Electronics and Energetics*. 2021. Vol. 34, issue 1, pp. 133–140. doi: 10.2298/FUEE2101133T

5. Prasetyowati S. A. D., Arifin B., Ardalli J., Ismail M., Subroto I. M. I., Purnomo M. H. A Hybrid FLC-LMS Algorithm for Predicting Sediment Volume in the River. *International Journal of Intelligent Engineering and Systems*. 2021. Vol. 14, issue 2, pp. 395–409. doi: 10.22266/ijies2021.0430.36

6. Gusev O. Yu., Gorev V. M., Korniienko V. I. Theory of adaptive filtration: tutorial. Dnipro: NTU “DP”, 156 p, 2019.

REFERENCES:

1. Laktionov, I., Diachenko, G., Rutkowska, D. & Kisiel-Dorohinick, M. (2023). An Explainable AI Approach to Agrotechnical Monitoring and Crop Diseases Prediction in Dnipro Region of Ukraine. *Journal of Artificial Intelligence and Soft Computing Research*, 13(4), 247–272. doi: 10.2478/jaiscr-2023-0018

2. Diachenko, G., Laktionov, I., Vizniuk, A., Gorev, V., Kashtan, V., Khabarlak, K. & Shedlovska, Y. (2024). An Improved Approach to Prediction of Maize Disease Occurrence Based on Weather Monitoring and Machine Learning: a Case of the Forest-Steppe and Northern Steppe of Ukraine. *Engineering Applications of Artificial Intelligence*, under review.

3. Lai, X., Yang, T., Wang, Z. & Peng, C. (2019). IoT Implementation of Kalman Filter to Improve Accuracy of Air Quality Monitoring and Prediction. *Applied Sciences*, 9(9), 1831 (23 pages). doi: 10.3390/app9091831

4. Tajdari, T. (2021). Adaptive method to predict an track unknown system behaviors using LMS and RLS algorithms. *Facta Universitatis, Series: Electronics and Energetics*, 34(1), 133–140. doi: 10.2298/FUEE2101133T

5. Prasetyowati, S. A. D., Arifin, B., Ardalli, J., Ismail, M., Subroto I. M. I. & Purnomo, M. H. (2021). A Hybrid FLC-LMS Algorithm for Predicting Sediment Volume in the River. *International Journal of Intelligent Engineering and Systems*, 14 (2), 395–409. doi: 10.22266/ijies2021.0430.36

6. Gusev, O. Yu., Gorev, V. M. & Korniienko, V. I. (2019). *Theory of adaptive filtration: tutorial*. Dnipro: NTU “DP”, 156 p.

UDC 004.056:004.75
DOI https://doi.org/10.32782/IT/2024-4-9

Serhii DOROZHYNSKYI
PhD, Associate Professor at the Department of Cybersecurity, Information Technology and Economics, Dean of the Faculty of Information Technology and Social Sciences and Humanities, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121, dorozhun1706@gmail.com
ORCID: *0000-0002-5395-6423*
Scopus author ID: *57222147991*

Vitaliy TUPKALO
Doctor of Technical Sciences, Professor, Professor at the Department of Cybersecurity, Information Technologies and Economics, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121
ORCID: *0000-0002-6594-530X*

Yuriy SHCHERBYNA
Candidate of Technical Sciences, Associate Professor at the Department of Information Technologies, National University «Odessa Law Academy», Odessa, Ukraine, shcherbyna@onua.edu.ua
ORCID: *0000-0003-3885-6747*

To cite this article: Dorozhynskyi, S., Tupkalo, V., Shcherbyna, Yu. (2024). Vazhlyvi pytannia obrobky ta zakhystu danykh u dystrybutovanykh systemakh [Important issues of data processing and protection in distributed systems]. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 4, 68–74, doi: https://doi.org/10.32782/IT/2024-4-9

IMPORTANT ISSUES OF DATA PROCESSING AND PROTECTION IN DISTRIBUTED SYSTEMS

The purpose *of this article is to explore the critical aspects of big data processing in distributed networks as a contemporary challenge in ensuring cybersecurity. Special attention is given to issues of confidentiality, integrity, and availability of data in multi-user environments with high interaction complexity.*

The methodology *of the study involves a comprehensive approach, including the analysis of threats and vulnerabilities specific to distributed architectures such as cloud technologies, peer-to-peer networks, and decentralized infrastructures. Key issues such as asynchronous data access, limited computational resources, operation synchronization complexity, and risks of data interception during transmission are examined. The study evaluates modern cryptographic approaches, including homomorphic encryption, distributed key management mechanisms, and privacy-preserving protocols. Additionally, the application of machine learning and deep learning algorithms is explored for anomaly detection in system behavioral models in real-time.*

The scientific novelty *of the research lies in the systematization of threats and vulnerabilities inherent to distributed data processing systems and the development of effective approaches to neutralize them. For the first time, a comprehensive approach to the use of blockchain technologies is proposed as a means of ensuring transparent auditing of events and tracking transactions in distributed environments.*

The conclusions *provide practical recommendations for integrating modern cryptographic methods, distributed key management mechanisms, blockchain technologies, and machine learning algorithms to build resilient cybersecurity systems. The findings are of practical significance for ensuring the protection of big data in distributed networks and improving the effectiveness of responses to potential threats.*

Key words: *big data, distributed networks, cybersecurity, information protection, cryptography, homomorphic encryption, anomaly detection, machine learning, blockchain, risk management, privacy, data integrity, cloud technologies.*

Сергій ДОРОЖИНСЬКИЙ
PhD, доцент кафедри кібербезпеки, інформаційних технологій та економіки, декан факультету інформаційних технологій та соціально-гуманітарних наук, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе, 210, м. Київ, Україна, 02121
ORCID: *0000-0002-5395-6423*
Scopus Author ID: *57222147991*

Віталій ТУПКАЛО
доктор технічних наук, професор, професор кафедри кібербезпеки, інформаційних технологій та економіки, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе 210, м. Київ, Україна, 02121
ORCID: *0000-0002-6594-530X*

Юрій ЩЕРБИНА
кандидат технічних наук, доцент кафедри інформаційних технологій, Національний університет «Одеська юридична академія», Одеса, Україна, 02121
ORCID: *0000-0003-3885-6747*

Бібліографічний опис статті: Дорожинський, С., Тупкало, В., Щербина, Ю. (2024). Важливі питання обробки та захисту даних у дистрибутованих системах. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 4, 68–74, doi: https://doi.org/10.32782/IT/2024-4-9

ВАЖЛИВІ ПИТАННЯ ОБРОБКИ ТА ЗАХИСТУ ДАНИХ У ДИСТРИБУТОВАНИХ СИСТЕМАХ

Метою статті *є дослідження критичних аспектів обробки великих даних у розподілених мережах як сучасного виклику забезпечення кібербезпеки. Особлива увага приділена проблемам конфіденційності, цілісності та доступності даних у багатокористувацьких середовищах із високим рівнем складності взаємодії.*

У ***методології*** *роботи застосовано комплексний підхід, що включає аналіз загроз і вразливостей, характерних для розподілених архітектур, таких як хмарні технології, однорангові мережі та децентралізовані інфраструктури. Розглянуто ключові проблеми, зокрема асинхронний доступ до даних, обмежені обчислювальні ресурси, складність синхронізації операцій і ризики перехоплення інформації під час передачі. Проведено аналіз сучасних криптографічних підходів, включаючи гомоморфне шифрування, механізми розподіленого управління ключами та протоколи збереження конфіденційності. Додатково досліджено алгоритми машинного та глибинного навчання для виявлення аномалій у поведінкових моделях систем у режимі реального часу.*

Наукова новизна *роботи полягає у систематизації загроз та вразливостей, притаманних розподіленим системам обробки даних, і розробці ефективних підходів для їх нейтралізації. Вперше запропоновано комплексний підхід до використання блокчейн-технологій як засобу забезпечення прозорості аудиту подій і відстеження транзакцій у розподілених середовищах.*

У ***висновках*** *наведено практичні рекомендації щодо інтеграції сучасних криптографічних методів, механізмів розподіленого управління ключами, блокчейн-технологій та алгоритмів машинного навчання для побудови стійких до загроз кібербезпекових систем. Результати дослідження мають практичне значення для забезпечення захисту великих даних у розподілених мережах та підвищення ефективності реагування на потенційні загрози.*

Ключові слова: *великі дані, розподілені мережі, кібербезпека, захист інформації, криптографія, гомоморфне шифрування, виявлення аномалій, машинне навчання, блокчейн, управління ризиками, конфіденційність, цілісність даних, хмарні технології.*

Introduction. In today's era of digital transformation, big data has become a key resource for organizations across industries. Distributed networks, such as cloud computing, peer-to-peer networks, and decentralized platforms, provide scalability, accessibility, and efficiency in processing this data. However, the growth in the volume of information and its criticality is accompanied by the emergence of new threats to its security, due to the dynamism of distributed environments, asymmetry of computing power, and the complexity of integrating various system components.

Cyberattacks aimed at compromising the confidentiality, integrity and availability of data pose a particular danger. Vulnerabilities in network infrastructure, unreliable data transmission protocols, and flaws in data access systems create favorable conditions for attacks. In particular, attackers use the specifics of distributed networks to launch denial-of-service (DDoS) attacks, steal data and unauthorized modification. In addition, the problem is complicated by the requirements for real-time data processing speed, which imposes restrictions on the use of traditional security methods that are too resource-intensive. In this context, it is critical to develop effective, scalable and adaptive security mechanisms that take into account the specifics of big data and distributed network architectures.

The purpose of the article is to identify, analyze, and develop effective methods and approaches to ensuring the protection of big data in distributed networks that take into account the

specifics of such environments, in particular their dynamism, asynchrony, and scalability.

To achieve this goal, the following **main tasks** have been set:

1. To review and systematize the current threats arising from the processing of big data in distributed networks, with a focus on the problems of confidentiality, integrity and availability of information.
2. Investigate existing cryptographic security methods, their effectiveness and adaptability to distributed environments.
3. To develop and analyze threat detection algorithms using machine learning technologies, in particular deep learning, to identify anomalies in real time.
4. Evaluate the potential of blockchain technologies to ensure the security of big data in distributed networks, in particular for event auditing and transaction tracing.

One of the key challenges in securing big data in distributed networks is the need to balance the efficiency of information processing with its protection. Multi-user environments are often characterized by high dynamic data access, which makes it difficult to manage access rights, ensure transparency of operations, and prevent unauthorized interference. In addition, the nature of big data processing requires the integration of various technologies, such as containerization, compute orchestration, and workflow automation, which increases the number of possible vulnerabilities.

Of particular importance are the problems of ensuring data integrity, when changing even a small part of the information can lead to significant disruptions in the system's functioning or to the appearance of unreliable results. At the same time, the need for fast real-time data processing makes it difficult to implement comprehensive security checks at every stage of computing.

Modern challenges also relate to the unification of security protocols and standards in the context of the globalization of the information space. Distributed networks often span multiple jurisdictions with different laws and regulatory requirements, making it difficult to create universal approaches to data protection. The issue of system compatibility and ensuring their uninterrupted functioning in the event of threats or attacks is another important aspect.

Another challenge is the development of decentralized platforms where data processing takes place without clear centralized control. In such systems, there is a growing risk of disagreements between participants, distrust in the results of data processing, and the spread of manipulation in transactions. These challenges emphasize the need to develop not only technical solutions but also regulatory and organizational approaches that ensure comprehensive protection of big data in distributed environments.

Threats to big data in distributed networks can be categorized into three main aspects: confidentiality, integrity, and availability, which form the basic triad of cybersecurity – the CIA model. Each of these aspects has its own specific threat characteristics and consequences.

Confidentiality means ensuring that only authorized entities have access to data. The formalization of a confidentiality breach can be described by the formula:

$$B(C) = P(U_i \in U_a),$$

where U_i – access subject, U_a – set of authorized entities, P – the possibility of unauthorized access. The main types of privacy attacks are information interception and metadata analysis, including man-in-the-middle attacks and attacks on encrypted data channels.

Integrity ensures the integrity and accuracy of data during its processing, storage, or transmission. A breach of integrity can be expressed by the following formula:

$$B(I) = \exists x_i, x_j \in D, x_i \neq x_j, h(x_i) \neq h(j),$$

where D – dataset, $h(x)$ – a hash function used for data verification. Integrity threats include unauthorized data modification and data injection. In particular, data overwriting attacks affect critical systems, reducing their reliability.

Availability means ensuring that data and resources can be legally accessed at any time. An availability violation is described by the following formula:

$$B(A) = \frac{D_t}{T_t} > \delta,$$

where δ – acceptable level of downtime, D_t – system unavailability time, T_t – total system operation time.

Classification of threats by sources and methods of attack

Threats to big data in distributed networks can be categorized according to various attributes, including attack sources and methods. Defining these categories allows you to create more accurate models for protecting and responding to threats. The main sources of threats are internal and external attacks, as well as unauthorized use of resources. Attack methods can be divided into attack and exploitation methods used to violate each of the security aspects: confidentiality, integrity, and availability.

This classification allows us to take into account not only the types of attacks, but also their sources, which allows us to create more flexible and effective defense strategies. For example, to combat internal threats, it is necessary to apply strict control over access to data and monitor user activity. For external threats, such as DDoS attacks, it is necessary to use traffic filtering and network-level protection methods. In addition, developing methods to quickly identify fake certificates or attack verification mechanisms will be an important step in ensuring security in the face of the constant development of attack techniques.

To effectively protect big data in distributed networks, it is necessary to implement comprehensive methods focused on ensuring confidentiality, integrity, and availability. These methods should take into account the specifics of distributed environments, such as decentralized access, a large number of entry points, and high volumes of data to be processed.

Confidentiality protection is aimed at protecting information from unauthorized access. The main methods are:

1. Data encryption, where symmetric (AES) and asymmetric algorithms (RSA, ECC) are used to protect confidentiality. The encryption process can be described by the formula:

$$C = E_k(P),$$

where C – ciphertext, P – plain text, E_k – encryption function with a key.

2. Data anonymization, where techniques such as differential privacy add noise to the data to preserve user privacy.

3. Access control, where multi-level policies, such as Role-Based Access Control, are implemented.

Data integrity ensures that information remains unchanged and reliable. The main methods:

- Hashing – uses algorithms such as SHA-256 to verify changes to data.
- Digital signatures – provide authentication of the data source using asymmetric encryption.
- Version control – storing previous versions of data to ensure that it can be restored after modification.

Data availability guarantees continuous access to information, even in the event of a threat. The main methods:

- Distributed storage – the use of systems such as RAID or decentralized networks that guarantee data redundancy:
- Protection against DDoS attacks – the use of traffic filtering mechanisms, for example, through content delivery networks (CDNs).
- Backup automation – regular backups for quick recovery in case of loss.

These methods should interact in a single system to provide a full range of data protection. For example, encryption guarantees confidentiality, but digital signatures or hashing are required to verify data.

Distributed networks that process large amounts of data are characterized by a number of specific features that significantly affect the implementation and effectiveness of information security methods. These features include a geographically distributed structure, high dynamics of the network topology, a large amount of processed data, and a multi-user environment.

Table 1

Classification of threats by sources and methods of attack

Threat category	Threat source	Attack methods	Attack goals	Attack example
Internal threats	Employees or users of the system	1.Illegal access to data 2.Modification of data without authorization	Confidentiality Integrity	Insider attacks Data leakage
External threats	External attackers or organizations	Man-in-the-middle or distributed denial of service attacks	Confidentiality Integrity Availability	DDoS attacks Data interception
Threats from technology	Software and hardware vulnerabilities	Exploitation of OS or software vulnerabilities; Exploitation of flaws in network protocols	Confidentiality Integrity Availability	Database hacking; Network attacks
Variable threats	Attacks on verification mechanisms	Fake certificates or metadata tampering; Attacks on identity systems	Confidentiality Integrity	Forgery of digital signatures; Use of fake certificates
Threats caused by external factors	**Unforeseen circumstances**	**Natural disasters; Energy losses**	**Availability**	**Loss of access to data due to natural disasters**

One of the key features is **decentralization**, which creates additional challenges for access control and security monitoring. Distributed networks do not have a single data center, which makes it difficult to use traditional centralized security systems. In such conditions, it is necessary to use methods that allow you to dynamically adapt access policies depending on changes in the network topology. For example, role-based and attribute-based access control systems should provide a quick response to changes in user roles or resource status in real time. It is also important to ensure effective protection in terms of scalability. Big data processing requires computing resources that far exceed the capabilities of traditional cybersecurity approaches. Therefore, modern security methods must take into account the limitations of computing power, network bandwidth, and processing delays. This leads to a growing demand for lightweight cryptographic algorithms, data compression methods, and traffic flow optimization that minimize the impact on system performance. Another feature is the increased vulnerability to attacks due to the multi-user environment. The interaction of numerous users and devices in a distributed network creates risks to data confidentiality and integrity. These risks can be mitigated by using multi-level encryption, mandatory multi-factor authentication, and integrating intrusion detection systems with machine learning to analyze anomalies in user behavior. An important challenge is to ensure uninterrupted data availability in distributed storage networks. In such systems, any damage or failure of individual nodes should not affect the functioning of the entire network. This is achieved through data redundancy, regular backups, and the implementation of auto-checking and self-healing mechanisms.

Distributed computing is the fundamental basis for working with big data, as it provides parallel processing of information, optimized resource utilization, and scalability of systems. However, this architecture imposes significant requirements and creates new challenges for building data protection systems. They are characterized by a large number of nodes that can be geographically distributed. This requires the implementation of decentralized security systems that can operate without centralized control. In particular, such systems use blockchain, which provides reliable verification of transactions, or cryptographic algorithms that allow the exchange of keys without the need for centralized trusted parties.

The second critical aspect is the scalability of the security architecture. Security systems must be able to dynamically expand to handle growing data volumes and the number of requests. In such an environment, traditional encryption and authentication methods may not be effective due to processing delays. This requires the use of lightweight protocols, such as elliptic curve cryptography, as well as the integration of artificial intelligence-based threat detection methods. In addition, distributed computing creates risks associated with infrastructure heterogeneity. In most cases, distributed systems consist of nodes running on different platforms and using different technologies. This increases the likelihood of exploiting compatibility vulnerabilities in system components. To solve this problem, it is necessary to implement standardized security protocols and mutual authentication mechanisms between nodes. It is also important to manage access to resources in a distributed architecture. Due to the large number of users and requests, there is a risk of system overload and an increased likelihood of attacks such as DDoS. To prevent this, access restriction policies, load balancing between nodes, and dynamic resource scaling are used. Another challenge is to ensure transparency and control over security in distributed computing. Дntegrated monitoring systems are used to analyze the behavior of nodes and requests in real time. Such systems are able to automatically detect anomalies that may indicate a possible attack and take measures to neutralize it.

Effective protection of big data in distributed networks requires the implementation of modern monitoring and threat detection systems that can operate in real time, given the dynamic nature of such networks. These systems are based on analyzing large volumes of traffic, detecting anomalies in the behavior of users and network nodes, and being able to respond quickly to security incidents.

One of the main components of modern monitoring systems is mechanisms for collecting and analyzing telemetry data from all network nodes. In distributed networks, these mechanisms must be decentralized and scalable to process data from many sources simultaneously. Streaming data processing technologies, such as Apache Kafka or Apache Flink, play an important role in this, allowing for efficient real-time processing of events. Threat detection systems, in turn, use machine learning and artificial intelligence algorithms to analyze anomalies. For example, clustering methods or neural networks to detect complex patterns in behavior. These algorithms allow not only to identify threats that match known attack patterns, but also to predict potentially new types of attacks that have no historical analogues.

Integrating monitoring with incident response tools is also an important aspect. Systems such as SIEM automate detection and response processes,

providing analytical information for security operators. For example, if suspicious activity is detected in network traffic, the system can automatically isolate the host or block access, preventing the threat from spreading. Another critical component is ensuring transparency of the monitoring system's actions for administrators. This is achieved by building a multi-level data visualization system that allows you to get an idea of the network status and possible threats. The use of graphical interfaces, for example, based on systems such as Kibana, facilitates data analysis even on a large scale. The importance of protecting the monitoring systems themselves, which also become a target for attackers, should be noted separately. For this purpose, methods of protection against man-in-the-middle attacks, multi-level encryption of telemetry data, and the use of containerization technologies to isolate components are being implemented.

Integrating security into the architecture of distributed networks is a key aspect of big data security. This process requires taking into account the specifics of distributed environments, such as decentralization, infrastructure heterogeneity, and the dynamic nature of data flows. An integrated approach to integration should provide protection at all levels of the architecture, including the infrastructure, network, and application layers.

At the infrastructure level, isolation and segmentation technologies are implemented to minimize the risks of threats spreading in the event of compromise of individual nodes. For example, the use of virtualization and containerization (Docker, Kubernetes) provides logical isolation of processes and execution environments. This makes it impossible for an attacker to access other system components even if an attack on a single node is successful.

At the network level, the integration of security means involves the implementation of mechanisms for encrypting data transmission channels (for example, TLS/SSL protocols) and protection against attacks at the routing level (for example, BGP-Sec). It is also important to use network segmentation using virtual local area networks (VLANs) to limit the interaction between different groups of nodes. Distributed systems are increasingly using software-defined networks (SDN), which provide centralized control and adaptation of security policies to dynamic changes in the network.

At the application level, integration includes the implementation of multi-factor authentication, role-based and attribute-based access control models, as well as control and audit tools. The uniqueness of distributed environments requires these mechanisms to be highly adaptive: they must take into account the distribution of resources and ensure policy consistency across different nodes. For example, when a user's role changes, the access policy should be automatically updated across all system components.

Automation is of particular importance for security integration. In distributed networks, large amounts of data and high process dynamics make manual security management difficult. Therefore, the integration of automated systems, such as SOAR (Security Orchestration, Automation, and Response), allows you to automatically respond to security incidents. For example, if abnormal activity is detected on a particular node, the system can initiate isolation of that node, change access policies, or redirect traffic to other nodes. Another important component is the integration of self-diagnostic and self-healing mechanisms. This allows the system to proactively detect vulnerabilities, check the compliance of settings with modern security standards, and automatically restore functionality in the event of a failure. For example, the use of data replication and consensus mechanisms (RAFT, Paxos) in distributed data warehouses guarantees their availability even if some nodes are lost.

Conclusions. The article investigates the critical aspects of big data processing in distributed networks, taking into account the challenges associated with ensuring their protection. By analyzing modern methods of data processing, monitoring and protection, it is established that distributed systems are an important component of big data infrastructure, but their decentralization, scalability and dynamic nature significantly complicate the construction of effective security systems. In particular, the structure of distributed networks, characterized by geographical remoteness of nodes, heterogeneity of execution environments, and high traffic intensity, imposes additional requirements for security tools. The study determined that traditional approaches to ensuring the confidentiality, integrity, and availability of information are becoming insufficient, especially in the context of system scaling. In this regard, it is proposed to use lightweight cryptographic algorithms, in particular elliptic curve cryptography, and blockchain technologies for decentralized data storage and transaction verification. Threat monitoring and incident management are critical to protecting big data in distributed networks. We propose to integrate real-time data analysis mechanisms, use artificial intelligence to detect anomalies, and automate response processes. Systems such as SIEM and SOAR, which are focused on automatically performing threat prevention operations, have proven effective in reducing response times and minimizing the impact of

incidents on system functionality. Integration of security tools into the architecture of distributed networks should take into account a multi-level approach. At the infrastructure level, it is recommended to use node isolation through virtualization and containerization. At the network level, the use of software-defined networks (SDN) allows you to dynamically adapt security policies to changes in topology. At the application level, it is important to implement multi-factor authentication, adaptive access control, and audit mechanisms. A high level of automation of security processes is a mandatory component of modern systems. The introduction of self-healing technologies and self-diagnostic mechanisms allows for a proactive approach to eliminating threats. Data replication and consensus tools, such as RAFT and Paxos, ensure data availability and integrity even in the event of the loss of individual nodes. The study results confirm that effective protection of big data in distributed networks is only possible with the use of innovative technologies, adaptive methods, and an integrated approach to security. Integration of advanced technologies such as blockchain, artificial intelligence, software-defined networks, and automated incident management systems are key to creating resilient and dynamic security systems that meet the demands of modern information environments.

BIBLIOGRAPHY:

1. M. S. Rahman and H. Reza, «A Systematic Review Towards Big Data Analytics in Social Media,» in Big Data Mining and Analytics, vol. 5, no. 3, pp. 228–244, September 2022, doi: 10.26599/BDMA.2022.9020009.

2. D. Syed, A. Zainab, A. Ghrayeb, S. S. Refaat, H. Abu-Rub and O. Bouhali, «Smart Grid Big Data Analytics: Survey of Technologies, Techniques, and Applications,» in IEEE Access, vol. 9, pp. 59564–59585, 2021, doi: 10.1109/ACCESS.2020.3041178.

3. X. Sun, Y. He, D. Wu and J. Z. Huang, «Survey of Distributed Computing Frameworks for Supporting Big Data Analysis,» in Big Data Mining and Analytics, vol. 6, no. 2, pp. 154–169, June 2023, doi: 10.26599/BDMA.2022.9020014.

4. J. Liao and J. Lin, «A Distributed Deep Reinforcement Learning Approach for Reactive Power Optimization of Distribution Networks», в IEEE Access, vol. 12, p. 113898–113909, 2024, doi: 10.1109/ACCESS.2024.3445143.

5. D. Park, S. Kang and C. Joo, «A learning-based distributed algorithm for scheduling in multi-hop wireless networks,» *Journal of Communications and Networks*, vol. 24, no. 1, pp. 99–110, Feb. 2022, doi: 10.23919/JCN.2021.000030.

6. M. W. S. Atman and A. Gusrialdi, «Finite-Time Distributed Algorithms for Verifying and Ensuring Strong Connectivity of Directed Networks,» in IEEE Transactions on Network Science and Engineering, vol. 9, no. 6, pp. 4379–4392, 1 Nov. -Dec. 2022, doi: 10.1109/TNSE.2022.3200466.

7. T. Lenard, A. Collen, M. Benyahya, N. A. Nijdam and B. Genge, «Exploring Trust Modeling and Management Techniques in the Context of Distributed Wireless Networks: A Literature Review,» in IEEE Access, vol. 11, pp. 106803–106832, 2023, doi: 10.1109/ACCESS.2023.3320945.

REFERENCES:

1. Rahman, M. S., & Reza, H. (2022). A systematic review towards big data analytics in social media. Big Data Mining and Analytics, 5(3), 228–244. https://doi.org/10.26599/BDMA.2022.9020009

2. Syed, D., Zainab, A., Ghrayeb, A., Refaat, S. S., Abu-Rub, H., & Bouhali, O. (2021). Smart grid big data analytics: Survey of technologies, techniques, and applications. IEEE Access, 9, 59564–59585. https://doi.org/10.1109/ACCESS.2020.3041178

3. Sun, X., He, Y., Wu, D., & Huang, J. Z. (2023). Survey of distributed computing frameworks for supporting big data analysis. Big Data Mining and Analytics, 6(2), 154–169. https://doi.org/10.26599/BDMA.2022.9020014

4. Liao, J., & Lin, J. (2024). A distributed deep reinforcement learning approach for reactive power optimization of distribution networks. IEEE Access, 12, 113898–113909. https://doi.org/10.1109/ACCESS.2024.3445143.

5. Park, D., Kang, S., & Joo, C. (2022). A learning-based distributed algorithm for scheduling in multi-hop wireless networks. *Journal of Communications and Networks*, 24(1), 99–110. https://doi.org/10.23919/JCN.2021.000030

6. Atman, M. W. S., & Gusrialdi, A. (2022). Finite-time distributed algorithms for verifying and ensuring strong connectivity of directed networks. IEEE Transactions on Network Science and Engineering, 9(6), 4379–4392. https://doi.org/10.1109/TNSE.2022.3200466

7. Lenard, T., Collen, A., Benyahya, M., Nijdam, N. A., & Genge, B. (2023). Exploring trust modeling and management techniques in the context of distributed wireless networks: A literature review. IEEE Access, 11, 106803–106832. https://doi.org/10.1109/ACCESS.2023.3320945

УДК 007:004
DOI https://doi.org/10.32782/IT/2024-4-10

Валентин ДЯЧЕНКО
кандидат економічних наук, доцент кафедри кібербезпеки, інформаційних технологій та економіки, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе 210, м. Київ, Україна, 02121
ORCID: *0000-0002-0055-9256*
Scopus Author ID: *57994193000*

Наталія ДЯЧЕНКО
кандидат наук з державного управління, доцент кафедри кібербезпеки, інформаційних технологій та економіки, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе 210, м. Київ, Україна, 02121
ORCID: *0000-0002-4306-7665*
Scopus Author ID: *57216565101*

Бібліографічний опис статті: Дяченко, В., Дяченко, Н. (2024). Врахування ризиків при використанні інформаційних технологій. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 75–80, doi: https://doi.org/10.32782/IT/2024-4-10

ВРАХУВАННЯ РИЗИКІВ ПРИ ВИКОРИСТАННІ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

У статті досліджено ризики та загрози при використанні інформаційно-комунікаційних технологій.

Метою дослідження *є ідентифікація основних видів ризиків та загроз при хмарному збереженні даних. На виконання мети ідентифіковано комплекс заходів щодо формування засад кібербезпеки. Акцентовано, що з урахуванням викликів сьогодення потрібні не лише кіберграмотність користувачів інформаційно-комунікаційних технологій, а й координація та переорієнтація досліджень у сфері комп'ютерних розробок, адже нові покоління програмно-апаратного забезпечення повинні гарантувати безпеку діяльності та дотримання принципу конфіденційності приватної, підприємницької чи державної інформації.*

Методологія дослідження. *Методологічні аспекти дослідження враховують концепції виявлення закономірностей появи нових ризиків та загроз інформаційній та кібернетичній безпеці в сучасних умовах стрімкого розвитку інформаційних технологій. У роботі було застосовано комплекс взаємопов'язаних наукових методів, зокрема: діалектичний, компаративного аналізу та логічний під час дослідження сутності загроз інформаційній безпеці у сучасних умовах невизначеності та ризику; метод емпіричного дослідження при порівнянні особливостей надання хмарних послуг ІТ-сервісами; системний підхід при ідентифікації комплексу заходів щодо формування засад кібербезпеки.*

Наукова новизна *полягає в інтегрованому підході до дослідження сутності ризиків та загроз інформаційній та кібернетичній безпеці в умовах стрімкого розвитку інформаційно-комунікаційних технологій та при використанні хмарних сервісів.*

Висновки. *У результаті теоретичного дослідження було здійснено ідентифікацію ризиків та загроз кібернетичній безпеці. Наголошено, що оперативна ідентифікація нових типів загроз та ризиків, визначення їх ключових характеристик сприятиме окресленню їх сутнісних властивостей, розробці алгоритмів їх аналізу та методів управління ними.*

Ключові слова: *ризики, загрози, інформаційні технології, хмарні технології.*

Valentyn DIACHENKO
Candidate of Economic Sciences, Associate Professor at the Department of Cybersecurity, Information Technology and Economics, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121, diachenko.v@ukr.net
ORCID: *0000-0002-0055-9256*
Scopus Author ID: *57994193000*

Natalia DIACHENKO
Candidate of Public Administration Sciences, Associate Professor at the Department of Cybersecurity, Information Technology and Economics, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121, n.diachenko@ukr.net
ORCID: *0000-0002-4306-7665*
Scopus Author ID: *57216565101*

To cite this article: Diachenko, V., Diachenko, N. (2024). Vrakhuvannia ryzykiv pry vykorystanni informatsiinykh tekhnolohii [Considering risks when using information technology]. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 75–80, doi: https://doi.org/10.32782/IT/2024-4-10

CONSIDERING RISKS WHEN USING INFORMATION TECHNOLOGY

The article examines risks and threats when using information and communication technologies.

The purpose of the study *is to identify the main types of risks and threats when storing data in the cloud. To achieve this goal, a set of measures to form the foundations of cybersecurity has been identified. It is emphasized that, taking into account today's challenges, not only cyber literacy of information and communication technology users is required, but also coordination and reorientation of research in the field of computer development, because new generations of software and hardware must guarantee the security of activities and compliance with the principle of confidentiality of private, business or state information.*

Research methodology. *The methodological aspects of the study take into account the concept of identifying patterns of the emergence of new risks and threats to information and cyber security in modern conditions of rapid development of information technologies. The work used a set of interrelated scientific methods, in particular: dialectical, comparative analysis and logical when studying the essence of threats to information security in modern conditions of uncertainty and risk; empirical research method when comparing the features of the provision of cloud services by IT services; systematic approach when identifying a set of measures to form the foundations of cybersecurity.*

The scientific novelty *lies in an integrated approach to studying the essence of risks and threats to information and cyber security in the context of the rapid development of information and communication technologies and when using cloud services.*

Conclusions. *As a result of the theoretical study, risks and threats to cyber security were identified. It is emphasized that the operational identification of new types of threats and risks, the determination of their key characteristics will contribute to the delineation of their essential properties, the development of algorithms for their analysis and methods for managing them.*

Key words: *risks, threats, information technologies, cloud technologies.*

Актуальність проблеми. Інтернет, інформаційно-комунікаційні технології (далі – ІКТ), надання послуг у цифровому форматі, що стали невід'ємною умовою сьогодення, від електронного документообігу, інтернет-магазинів, онлайн-банкінгу, інтелектуальних систем управління бізнесом передбачають необхідність не лише наявності знань, умінь та навичок використання ІКТ, а й наявність досвіду з управління системами захисту інформації та кібернетичної безпеки, оскільки кіберзагрози інтенсивно еволюціонують, кіберзлочини постійно вдосконалюються, набуваючи рис транснаціонального характеру.

Аналіз останніх досліджень і публікацій. Дослідженню ризиків, загроз та небезпек при використанні інформаційно-комунікаційних технологій присвятили увагу ряд відомих вітчизняних науковців, зокрема О. Гайдук та В. Звєрєв здійснили аналіз кіберзагроз в умовах стрімкого розвитку інформаційних технологій (Гайдук, Звєрєв, 2024, с. 229). А. Лисеюк та Т. Свінцицька дослідили правове забезпечення кібербезпеки України в умовах воєнного стану та євроїнтеграції (Лисеюк, Свінцицька, 2024). К. Мовчан ідентифіковано ризики кібербезпеки в епоху робототехніки (Мовчан, 2023).

Метою дослідження є ідентифікація основних видів ризиків та загроз при хмарному збереженні даних.

Виклад основного матеріалу дослідження. За умов сучасного правового режиму воєнного стану в Україні та стрімкого зростання кіберризиків і кіберзагроз гостро постає потреба у підготовці висококваліфікованих фахівців у цій сфері, які могли б здійснювати як аналіз уже реалізованих заходів у сфері захисту комп'ютерних і комунікаційних мереж від кібератак, так і передбачати можливі ризики й загрози та оперативно реагувати на їх прояви.

Системні перебої з електропостачанням, що мали місце у певні періоди, спричиняли збої у наданні електронних послуг. А кібератаки спричиняли блокування діяльності органів державної влади та роботи важливих для економіки та формування засад національної безпеки підприємств, установ та організацій. З метою формування засад кібербезпеки реалізовано ряд заходів для вирішення базових правових, суспільних, політичних та організаційних питань (рис. 1):

В умовах сьогодення вкрай необхідно посилити кіберзахищеність підприємств, установ, організацій та критичної інфраструктури, дотримуючись принципу «безпека понад усе» (англ. security-first thinking).

З метою формування базових знань з кібербезпеки ряд закладів освіти пропонують безоплатні курси, зокрема, заклад вищої освіти

Комплекс заходів з безпечного функціонування кіберпростору

Політика кібербезпеки	Законодавство	Глобальне партнерство	Просвітницькі програми з кібербезпеки
– Національна стратегія кібербезпеки; – Реалізація планів з виконання Стратегії; – Координація та контроль у сфері кібербезпеки	– Доктрина інформаційної безпеки України; – Закони України; – Положення та підзаконні й нормативні акти	– Ратифікація Конвенції про кіберзлочинність; – За підтримки НАТО створено ситуаційні центри; – Участь у міжнародних спецопераціях. –	– CERT-UA при Держспецзв'язку на сайті надає консультації громадянам; – оновлення освітніх програм у закладах вищої освіти. –

Рис. 1. Комплекс заходів щодо формування засад кібербезпеки

Вища школа публічної освіти пропонує загальну короткострокову програму «Основи забезпечення кібербезпеки та протидії дезінформації», яка передбачає дистанційну форму навчання, два напрями – інформаційна безпека та кібербезпека. Партнером-укладачем програми є «Міжнародна академія інформації». Метою програми є розвиток у слухачів компетентностей у сфері кібербезпеки, у тому числі застосування інструментів, що забезпечують конфіденційність, цілісність чи доступність даних, запобігання кібератакам та протидія дезінформації шляхом набуття професійних навичок роботи з інформацією, опанування цифрової грамотності та основ кібербезпеки.

З урахуванням викликів сьогодення потрібні не лише кіберграмотність користувачів ІКТ, а й координація та переорієнтація досліджень у сфері комп'ютерних розробок, адже нові покоління програмно-апаратного забезпечення повинні гарантувати безпеку діяльності та дотримання принципу конфіденційності приватної, господарської чи державної інформації.

Передумовою формування кола висококваліфікованих спеціалістів у сфері інформаційних технологій є підготовка здобувачів вищої освіти відповідно до освітньої програми «Управління системами захисту інформації та кібернетичної безпеки», метою якої є розвиток соціального та інтелектуального капіталу шляхом підготовки у галузі інформаційних технологій висококваліфікованих, соціально відповідальних фахівців кібербезпеки з високим рівнем етичних стандартів та професійної гідності, які, в контексті сприяння реалізації стратегій кібербезпеки та інформаційної безпеки, відповідають сучасним потребам ринку праці, суспільства та держави, сфокусовані на захисті прав, свобод і законних інтересів громадян в інформаційному просторі, володіють теоретико-практичними знаннями та вміннями, необхідними для розуміння принципів управління системами та комплексами інформаційної та/або кібербезпеки держави в цілому або окремих суб'єктів їх інфраструктури, здатні до постійного навчання і самовдосконалення, до застосування методів та засобів технічного та криптографічного захисту інформації від ризику стороннього кібернетичного впливу, що передбачає розробку нових, удосконалення або подальшого розвитку наявних розробок та досліджень.

Освітньо-професійна програма «Управління системами захисту інформації та кібернетичної безпеки» базується на загальновідомих наукових результатах в галузі інформаційних технологій та зорієнтована на формування комплексу знань, умінь та навичок щодо моделювання, проєктування, розробки, інтеграції та супроводження систем та комплексів інформаційної та/чи кібербезпеки на базі сучасних інформаційно-комунікаційних технологій, які надають здобувачами вищої освіти широкі можливості для самореалізації у сфері зайнятості та кар'єрного зростання.

Програма передбачає формування у здобувачів вищої освіти компетентностей щодо сучасних методів, методик, інформаційно-комунікаційних технологій та технологій забезпечення інформаційної та/або кібербезпеки, необхідних для розв'язання базових задач та практичних проблем в ІТ-сфері з урахуванням методології системного ІТ-аудиту щодо виявлення кібернетичних загроз та вторгнень, а також сучасних тенденцій розробки нових, удосконалення

або подальшого розвитку наявних розробок та досліджень у сфері інформаційних технологій.

Серед особливостей освітньої програми «Управління системами захисту інформації та кібернетичної безпеки» варто виокремити поєднання комплексу освітніх компонентів, що формують компетентності за фахом, враховують стратегічні напрями забезпечення кібербезпеки та інформаційної безпеки України, зорієнтовані на розвиток інтелектуального капіталу особистості, з акцентом на здатність здобувачів до розуміння змісту управління системами та комплексами інформаційної та/чи кібербезпеки, застосування сучасних концепцій та інструментарію забезпечення інформаційної безпеки, методологію системного IT-аудиту протидії кібернетичним загрозам та проєктування нових, удосконалення чи подальшого розвитку наявних розробок та досліджень.

Програмою передбачено наявність інструментів та обладнання, зокрема сучасного програмно-апаратного забезпечення інформаційно-комунікаційних технологій, яке передбачає у рамках інтерактивного навчання використовувати розроблене Microsoft лабораторне середовище SimuLand із відкритим вихідним кодом, розроблене з метою допомогти розгорнути лабораторне середовище, що відтворює добре відомі методи, які використовуються в реальних сценаріях атак, активно тестувати та перевіряти ефективність відповідних виявлень Microsoft 365 Defender, Azure Defender і Microsoft Sentinel, а також розширити дослідження загроз за допомогою телеметрії та криміналістичних артефактів, створених після кожної вправи моделювання, протестувати та поліпшити захист.

Одним з актуальних питань сьогодення є використання Інтернету, зокрема, хмарних сервісів. Хмарні технології (англ. cloud technologies) – сервіс, що дозволяє віддалено використовувати засоби накопичення, обробки та зберігання даних.

Базова концепція хмарного збереження та обробки даних базується на різних моделях надання IT-послуг, серед яких: CaaS, WaaS, SaaS, DaaS, PaaS, EaaS (табл. 1). Серед пропонованих сервісів варто виокремити SaaS (Software as a Service) – програмне забезпечення як послуга), як вигідну альтернативу придбання програмного забезпечення, адже SaaS дозволяє отримувати програмне забезпечення як послугу, а не дорого купувати ліцензійні програми.

Серед популярних продуктів SaaS є Salesforce.com (https://www.salesforce.com/) – найбільший у світі SaaS-провайдер (Salesforce, 2024), що надає доступ до власної CRM-системи (системи управління відносинами із клієнтами).

Набір додатків компанії Google Inc. (https://www.google.com/) містить поштовий сервіс з розширеними можливостями та інший ефективний функціонал для оптимізації діяльності, як приватної, так і підприємницької чи державної (Google Inc., 2024).

Серед популярних та зручних у використанні хмарних сховищ довільно виокремити:

– Google Drive – сховище даних, яке належить компанії Google Inc., дає змогу користувачам зберігати інформацію на хмарних серверах та обмінюватися нею з іншими користувачами в Інтернеті.

– OneDrive (офіційно Microsoft OneDrive) – сховище файлів, засноване на хмарній організації інтернет-сервісу зберігання файлів з додатковими функціями файлообміну (Microsoft OneDrive, 2024);

Таблиця 1

Особливості надання IT-послуг хмарними сервісами

Назва послуги	Опис
CaaS (Communication as a Service – комунікація як послуга)	Надає послуги зв’язку: IT-телефонія, поштові послуги тощо.
WaaS (Workplace as a Service – робоче місце як послуга)	Спеціалізується на наданні віртуальних робочих місць
SaaS (Software as a Service – програмне забезпечення як послуга)	Провайдер розміщує у себе додаток, а користувачі оплачують послугу за використання програми
DaaS (Data as a Service – дані як послуга)	Надання даних на вимогу користувача незалежно від його розташування
PaaS (Platform as a Service – платформа як послуга)	Надання інтегрованої IT-платформи для створення, розгортання, тестування та підтримки додатків
EaaS (Everythinge as a Service – все як послуга)	Комплекс хмарних сервісів, що задовольняє великий спектр потреб користувачів

– Dropbox файлообмінник та синхронізатор файлів від компанії Dropbox Inc., розташованої у Сан-Франциско, США (Dropbox Inc., 2024).

При використанні хмарних сервісів необхідно враховувати ризики:

– хмарні послуги надаються певною компанією, тому й збереження приватної інформації повністю залежить від неї;

– для перегляду власної інформації чи її опрацювання необхідно бути в мережі інтернет;

– зникнення напруги може спричинити тимчасову відсутність можливості доступу до власних баз даних, що зберігаються у хмарних сервісах.

Наведені вище постачальники хмарних послуг забезпечують захист персональних даних, зокрема:

– Google Drive забезпечують захист при передачі даних;

– Dropbox забезпечує захист через аутентифікацію;

– OneDrive здійснює подвійну аутентифікацію.

Однак, серед основних загроз хмарного збереження даних варто виокремити (табл. 2):

У рамках інтерактивного та практико-орієнтованого навчання здобувачів вищої освіти при опануванні ними освітніх компонентів програми «Управління системами захисту інформації та кібернетичної безпеки» доцільно використовувати кіберполігон.

Розгортання на базі кафедри кібербезпеки та інформаційних технологій *Київського університету інтелектуальної власності та права* кіберполігону – сукупності спеціальних програмно-апаратних комплексів, які об'єднані провідними та безпровідними засобами комунікацій, що можуть бути інтегрованими в мережу Інтернет та застосовуються для здійснення моніторингу впливу на системи управління, які можуть становити інтерес, для захисту власних систем від несанкціонованого доступу – сприяє формуванню у здобувачів вищої спеціальності 125 Кібербезпека навичок використання тактик передбачення кібератак, методів ідентифікації симуляції кібератак, відпрацювання методик їх відбиття. Навчальний кіберполігон сприяє формуванню у здобувачів вищої освіти системи професійних здатностей, адже дозволяє імітувати кібератаки, кібернапади на сервери, які обслуговують інфраструктури підприємства, установи чи організації для пошуку вразливих місць, усунення їх вразливості, налагодження ефективної системи захисту наявних комп'ютерних та інформаційно-комунікаційних ресурсів, відновлювати штатне їх функціонування. Програмне забезпечення та системи візуалізації кіберполігону сприяють відпрацюванню кібердій, що здійснюються у віртуальному середовищі. Системами візуалізації передбачено можливість моделювання кібератак, які можуть здійснюватись на комп'ютерні мережі, що передбачає зменшення чи й зовсім уникнення витрат на придбання ресурсів хмарних технологій.

Навчальний кіберполігон дозволяє імітувати кібератаки, кібернапади на сервери, які обслуговують інфраструктури закладу вищої освіти для пошуку вразливих місць, усунення їх вразливості, налагодження ефективної системи захисту наявних комп'ютерних та інформаційно-комунікаційних ресурсів, відновлювати штатне функціонування інформаційних, інформаційно-телекомунікаційних (автоматизованих) систем після реалізації загроз, здійснення кібератак, збоїв чи відмов різного рівня та походження.

Висновки та перспективи подальших досліджень. За умов сучасного правового режиму воєнного стану в Україні та стрімкого зростання кіберризиків і кіберзагроз гостро постає потреба аналізу та визначення трендів і тенденцій перебігу кіберзагроз, їх взаємного впливу та чинників, що не пов'язані напряму з кіберпростором, що здатні сприяти появі суттєвого негативного впливу на процеси розвитку комп'ютерних та інформаційно-комунікаційних систем і мереж. Ідентифікація нових типів загроз та ризиків, визначення їх ключових характеристик сприятиме окресленню їх сутнісних властивостей, розробці алгоритмів їх аналізу та методів управління ними.

Таблиця 2

Основні види загроз хмарного збереження даних

Загроза	Опис
Крадіжка даних	Можлива при атаці на сервер, коли отримують доступ до бази email-адрес
Втрата, псування даних	Дані можуть бути втраченими чи пошкодженими внаслідок системних помилок, програмної недосконалості хмарних сервісів
Незахищеність інтерфейсів	Помилки в проєктуванні хмарних сервісів роблять їх уразливими до кібератак
Суміжна вразливість	Спільний доступ до одних і тих же ресурсів створює повторну вразливість

ЛІТЕРАТУРА:

1. Гайдук О., Звєрєв В. Аналіз кіберзагроз в умовах стрімкого розвитку інформаційних технологій. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка»*.2024, № 3 (24). С. 225–236. https://doi.org/10.28925/2663-4023.2024.23.225236.

2. Лисеюк А., Свінцицька Т. Правове забезпечення кібербезпеки України в умовах воєнного стану та євроїнтеграції. *Право та інновації*, 2024, № 4 (48), с. 32–38. https://doi.org/10.37772/2518-1718-2024-4(48)-4.

3. Мовчан К. О. Ризики кібербезпеки в епоху робототехніки. *Вчені записки ТНУ імені В. І. Вернадського. Серія: Технічні науки*. 2023, Том 34 (73). с. 79–83. https://doi.org/10.32782/2663-5941/2023.4/13.

4. Salesforce.com. URL: https://www.salesforce.com/.

5. Google.Inc. URL: https://www.google.com/.

6. Microsoft OneDrive. URL: https://www.microsoft.com/uk-ua/microsoft-365/onedrive/online-cloud-storage.

7. Dropbox Inc. URL: https://www.dropbox.com/uk_UA/.

REFERENCES:

1. Haiduk, O., Zvieriev, V. (2024). Analysis of cyber threats in the context of rapid development of information technologies [Analysis of cyber threats in conditions of rapid development of information technologies]. *Elektronne fakhove naukove vydannia «Kiberbezpeka: osvita, nauka, tekhnika»*. https://doi.org/10.28925/2663-4023.2024.23.225236 [in Ukrainian].

2. Lyseiuk, A., Svintsytska, T. (2024). Pravove zabezpechennia kiberbezpeky Ukrainy v umovakh voiennoho stanu ta yevrointehratsii [Legal support for Ukraine's cybersecurity in the context of martial law and European integration]. *Pravo ta innovatsii*. https://doi.org/10.37772/2518-1718-2024-4(48)-4 [in Ukrainian].

3. Movchan, K. O. (2023). Ryzyky kiberbezpeky v epokhu robototekhniky [Cybersecurity risks in the age of robotics]. *Vcheni zapysky TNU imeni V. I. Vernadskoho. Seriia: Tekhnichni nauky.* https://doi.org/10.32782/2663-5941/2023.4/13 [in Ukrainian].

4. Salesforce.com. Retrieved from: https://www.salesforce.com/.

5. Google.Inc. Retrieved from: https://www.google.com/.

6. Microsoft OneDrive. Retrieved from: https://www.microsoft.com/uk-ua/microsoft-365/onedrive/online-cloud-storage.

7. Dropbox Inc. Retrieved from: https://www.dropbox.com/uk_UA/.

УДК 622.17: 004.383.8
DOI https://doi.org/10.32782/IT/2024-4-11

Тімур ЖЕЛДАК
кандидат технічних наук, доцент, завідувач кафедри системного аналізу та управління, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, Дніпро, Україна, 49005
ORCID: *0000-0002-4728-5889*
Scopus Author ID: *55602208300*

Олександр ВЛАДИКО
кандидат технічних наук, доцент кафедри системного аналізу та управління, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, Дніпро, Україна, 49005
ORCID: *0000-0001-9779-9565*
Scopus Author ID: *55522741200*

Бібліографічний опис статті: Желдак, Т., Владико, О. (2024). Використання машинного навчання для кластеризації з використанням індексу небезпеки хвостосховищ та їх впливу на навколишнє середовище. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 4, 81–91, doi: https://doi.org/10.32782/IT/2024-4-11

ВИКОРИСТАННЯ МАШИННОГО НАВЧАННЯ ДЛЯ КЛАСТЕРИЗАЦІЇ З ВИКОРИСТАННЯМ ІНДЕКСУ НЕБЕЗПЕКИ ХВОСТОСХОВИЩ ТА ЇХ ВПЛИВУ НА НАВКОЛИШНЄ СЕРЕДОВИЩЕ

В статті виконано аналіз небезпечного впливу хвостосховищ утворених від промислового виробництва на навколишнє середовище методами машинного навчання з метою їх кластеризації. При різних розмірах хвостосховищ, їх токсичності та параметрів була визначена їх положення в кластерах, що дозволило класифікувати стан рівня їх небезпеки. Зроблено порівняльний аналіз підходів машинного навчання та інших методів пошуку аналогів для визначення потенційної небезпеки хвостосховищ. За допомогою методу DBSCAN було визначено хвостосховища, які за своїми параметрами потрапляли до кореневого, граничного та шумового кластерів. Всі хвостосховища, що потрапили до кореневого та граничного кластерів, було класифіковано наступним методом кластеризації – k-середніх, і вони були розподілені до одного з чотирьох кластерів. Кластеризація методом k-середніх будувалася на вибірках даних хвостосховищ шляхом порівняння метрики відстані між витягнутими в одновимірний вектор матрицями суміжності. Наведено результати оцінки роботи алгоритму, що підтверджують, що моделювання та пошук центрів кластеризації за допомогою k-середніх є більш комплексним рішенням задачі в порівнянні з методом агломеративної кластеризації, який було приведено до порівняння k-середніх. Таким чином точність використаного підходу кластеризації виявився вищою, ніж у існуючих класичних алгоритмах кластеризації.

Метою роботи *є кластеризація хвостосховищ, що утворені від діяльності промислового виробництва, за рівнем їх небезпечного впливу на навколишнє середовище з використанням методів машинного навчання.*

Методологія. *Для досягнення мети використані методи машинного навчання та аналізу даних, що реалізуються в рамках фреймворку Sklearn мови програмування Python, використаний пакет роботи з даними pandas, що дозволяє обробляти великі масиви даних з використанням гнучкої системи форматування та запитів до них. Групування даних здійснювалося з використанням групи методів навчання без вчителя (unsupervised learning): алгоритмів кластеризації DBSCAN, кластерного аналізу методом ієрархічної деревоподібної кластеризації та методом k-середніх, методу головних компонентів.*

Наукова новизна. *Наукова новизна полягає у виборі способів кластеризації хвостосховищ за допомогою методів кластерного аналізу.*

Висновки. *За допомогою методів машинного навчання виконана кластеризація хвостосховищ, в залежності від ступеня їх небезпеки для навколишнього середовища.*

Ключові слова: *машинне навчання, кластеризація, метод k-середніх, хвостосховища, метод DBSCAN, індекс небезпеки, метод агломеративної кластеризації, навколишнє середовище.*

Timur ZHELDAK
Candidate of Technical Sciences, Associate Professor, Head of the System Analysis and Control Department, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, zheldak.t.a@nmu.one
ORCID: *0000-0002-4728-5889*
Scopus Author ID: *55602208300*

Alexander VLADYKO
Candidate of Technical Sciences, Associate Professor at the Department of System Analysis and Management, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, vladyko.o.b@nmu.one
ORCID: *0000-0001-9779-9565*
Scopus Author ID: *55522741200*

To cite this article: Zheldak, T., Vladyko A. (2024). Vykorystannia mashynnoho navchannia dlia klasteryzatsii z vykorystanniam indeksu nebezpeky khvostoskhovyshch ta yikh vplyvu na navkolyshnie seredovyshche [Using machine learning for clustering using the hazard index of tailings and their environmental impact]. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 4, 81–91, doi: https://doi.org/10.32782/IT/2024-4-11

USING MACHINE LEARNING FOR CLUSTERINGUSING THE HAZARD INDEX OF TAILINGSAND THEIR ENVIRONMENTAL IMPACT

The article analyzes the hazardous impact of industrial tailings on the environment using machine learning methods for their clustering. With different sizes of tails, their toxicity and parameters, their location in clusters was determined, which made it possible to classify the condition according to the degree of their danger. A comparative analysis of machine learning approaches and other methods of finding analogues for determining the potential danger of tailings storage facilities was carried out. Using the DBSCAN method, the tails were identified, which according to their parameters fell into the root, marginal, and noise clusters. All tails that fell into the root and border clusters were classified by the following clustering method – k-means, and they were allocated to one of four clusters. K-means clustering was constructed on tail data samples by comparing the distance metric between adjacency matrices compiled into a one-dimensional vector. The results of the evaluation of the algorithm are given, which confirm that the modeling and search for clustering centers using k-means is a more difficult solution to the problem compared to the method of agglomerative clustering, which was reduced to the comparison of k-means. Thus, the accuracy of the used clustering approach turned out to be higher than that of existing classical clustering algorithms.

The purpose *of the work is the clustering of tailings, formed by the activity of industrial production, according to the level of their dangerous impact on the environment using machine learning methods.*

The methodology. *To achieve the goal, the methods of machine learning and data analysis implemented within the Sklearn framework of the Python programming language were used, and the pandas data package was used, which allows processing large data sets using a flexible system of formatting and queries to them. Data grouping was carried out using a group of unsupervised learning methods: DBSCAN clustering algorithms, cluster analysis using the hierarchical tree clustering method and k-means method, principal component method.*

The scientific novelty. *The scientific novelty consists in the choice of methods of clustering of tailings storage facilities using methods of cluster analysis.*

Conclusions. *With the help of machine learning methods, clustering of tailings storage facilities was performed, depending on the degree of their danger to the environment.*

Key words: *machine learning, clustering, k-means method, tailings, DBSCAN method, hazard index, agglomerative clustering method, environment.*

Актуальність проблеми. Аварії на хвостосховищах у світі за останні три десятиліття показали, що вони призводять до великих катастроф, негативно впливають на здоров'я людей і спричиняють різноманітну шкоду навколишньому середовищу (Nikolaieva, 2015). Незважаючи на те, що умови безпеки значно покращилися, завдяки суворим вимогам і технічним заходам, безпека певної кількості хвостосховищ все ще залишається на незадовільному рівні через нестачу грошей для їх постійного захисту.

Крім того, очікується різке збільшення кількості хвостосховищ внаслідок збільшення видобутку кольорових металів, оскільки розумні та передові технології спричиняють різке зростання попиту на метали (кобальт, мідь, літій, нікель та ін.).

Як відомо видобуток корисних копалин супроводжується значним потоком відходів, що утворюються в результаті такої діяльності. Одним із багатьох типів відходів гірничодобувної промисловості є рідкі відходи, шлами, хвости, дрібнозернисті відходи, отримані з гірничо-збагачувальних фабрик, які потім зберігаються певним чином. В ідеальних умовах хвостосховища повинні забезпечувати безпечне тривале зберігання отриманих відходів. Однак на хвостосховищах можуть виникати аварії через несприятливі природні умови, недоліки проектування та будівництва, невідповідну практику експлуатації та управління.

Таким чином, суспільство може згодом зіткнутися зі зростаючим ризиками аварій на хвостосховищах із потенційними жертвами

та екологічними збитками. Якщо безпекою хвостосховища не почати керувати належним чином, відповідно до нових технологічних стандартів з урахуванням впливу водного балансу супутніх річок, тоді можуть виникати додаткові непередбачувані наслідки на хвостосховищах. У зв'язку з цим у керівництва країнами зростає занепокоєння щодо погіршення стану навколишнього середовища, спричиненого неконтрольованими великомасштабним переміщеннями небезпечних матеріалів, пов'язаних із аваріями на хвостосховищах. Такі аварії можуть призвести до непередбачуваних розливів і викидів небезпечних матеріалів, що зберігаються на хвостосховищах у навколишнє середовище. Це забруднення може мати серйозні наслідки та спричиняти збитки для здоров'я людини, побудованої інфраструктури, економічної діяльності, і викликати негативний вплив на довкілля та природні ресурси. Викиди з хвостосховищ дуже часто мають негативний широкомасштабний або транскордонний вплив на ресурси довкілля. Крім того, аварії на хвостосховищах можуть призвести до довгострокового забруднення води та ґрунту, мати негативні хронічні та накопичувальні наслідки для здоров'я людини та навколишнього середовища.

Тому особливої актуальності набувають питання створення для хвостосховищ адекватних умов та заходів безпеки разом з прогнозування вірогідних аварій. Аварії в на хвостосховищах як в Європейському союзі так і в Україні яскраво продемонстрували, наскільки серйозним може бути вплив неправильної їх експлуатації на людей, навколишнє середовище та водні ресурси. Ці події вимагають розробки та впровадження нових методів аналізу хвостосховищ і нових управлінських рішень по недопущенню випадків аварії.

Одним із сучасних методів прогнозування небезпеки хвостосховищ є методи штучного інтелекту до яких відноситься машинне навчання, яке дозволяє розділити хвостосховища на кластери по їх небезпечності та прогнозувати їх стан на певний проміжок часу.

Аналіз останніх досліджень і публікацій. Для кластеризації хвостосховищ активно використовуються підходи, що ґрунтується на використанні технологій машинного навчання. Зокрема, можливе застосування різних алгоритмів кластеризації з метою отримання груп зі схожими параметрами хвостосховищ. Для визначення їх впливу на навколишнє середовище використовують моделі, що описують різні процеси в яких найчастіше застосовують ієрархічну агломеративну кластеризацію та кластеризацію методом *k*-середніх. Далі виконується навчання моделі, в результаті якого вихідна база даних поділяється на групи родовищ зі схожими властивостями (Astakhova, 2015).

За допомогою метода *k*-середніх можна виділити кластер з подібними параметрами хвостосховищ, де пропонується використовувати даний вид кластеризації, попередньо знизивши розмірність вихідної бази даних. Потім для оцінки якості отриманих кластерів з аналогами можна використати регресійні моделі. Проте слід зазначити, що ці підходи мають певні недоліки. Обидва види кластеризації виявляються досить чутливими до вхідних даних великої розмірності. Зниження розмірності вхідних даних може призвести до втрат вихідної інформації. Інший недолік – складність застосування таких підходів у завданнях, де потрібно відновити один або кілька параметрів досліджуваних хвостосховищ. Хвостосховище, частина параметрів якого пропущена, неспроможна брати участь у процесі навчання моделі, і потребує додаткової обробки щодо його кластера.

Таким чином, запропонований алгоритм повинен ефективно працювати як з безперервними, так і з дискретними параметрами необхідної розмірності. Виходячи з пред'явлених вимог, був зроблений вибір на користь реалізації із застосуванням технологій машинного навчання, таких як метод *k*-середніх, які дозволяють при середній обчислювальній складності та незначному застосуванні вхідних даних отримати хорошу інтерпретованість навченої моделі. За допомогою формування умовних розподілів ймовірностей у вузлах мережі з'являється можливість не лише оцінювати невідомі змінні, а й виконувати пошук аномальних значень та проводити аналіз достовірності та інформативності отриманих даних.

Мета дослідження: кластеризація хвостосховищ, що утворені від діяльності промислового виробництва, за рівнем їх небезпечного впливу на навколишнє середовище з використанням методів машинного навчання. Запропоновані методи машинного навчання дозволяють більш точно визначити хвостосховища з великим рівнем небезпеки для навколишнього середовища.

Виклад основного матеріалу. Машинне навчання є великим розділом в галузі вивчення штучного інтелекту, яке включає методи побудови різних алгоритмів, здатних до самонавчатися. Виділено три групи методів такого навчання, що часто використовуються: навчання з учителем (регресія, класифікація); навчання

без вчителя (пошук правил, зменшення розмірності, кластеризація) та навчання з підкріпленням (генетичний алгоритм, *Q*-learning та ін.) (IMES, 2019). З цих методів найкраще вирішує задачу визначення небезпеки впливу хвостосховищ на навколишнє середовище є кластеризація. В свою чергу в кластеризації застосовують наступні алгоритми і методи машинного навчання (Kravchenko, 2020):

– евристичні графові алгоритми (алгоритм виділення зв'язкових компонентів, алгоритм найкоротшого незамкнутого шляху, FOREL алгоритм);

– статистичні алгоритми, засновані на розбиття (метод *k*-середніх, алгоритм DBSCAN, що ґрунтується на щільності розподілів характеристик, які вивчаються);

– ієрархічні методи (агломеративні та дивізіонні: алгоритми CURE, ROCK, Chameleon, метод Варда);

– алгоритми нечіткої кластеризації (FCM, FCS та MM алгоритми).

Кожна група з цих методів кластеризації має свої переваги і недоліки. Зокрема, статистичні алгоритми, такі як метод *k*-середніх та алгоритм DBSCAN, засновані на розбитті, ефективно працюють із великими обсягами даних, що завжди можна застосувати для графових методів кластеризації.

Обробка статистичних даних для вирішення задач кластеризації з використанням індексу небезпеки хвостосховищ можливо за рахунок баз даних аварій на хвостосховищах, яку сформуємо з використанням бібліографічних та відкритих джерел. Для повноти аналізу аварій та проривів у навколишнє середовище на хвостосховищах необхідно мати статистику цих подій за 20 – 70 років, тому знайдено відповідну інформацію у відповідних джерелах. І це необхідна статистика за останні 60 років в Європі, що повністю покриває нашу потребу в статистиці на 100%. Отже, за цей період зареєстровано 323 аварії. Кількість відмов залишалася незмінною протягом перших трьох десятиліть, а потім зменшилася після 1990 року протягом двох десятиліть, імовірно, відображаючи скорочення видобувної діяльності в країнах колишнього Радянського Союзу. Однак за останнє десятиліття кількість відмов повернулася до рівня початку накопичення статистики. Тенденція до аварій зростає, тому необхідно вжити заходів, щоб уникнути велику їх кількість (Winkelmann-Oei G., 2015).

Проведемо оцінку індексу потенціалу небезпеки хвостосховища та індексу ризику відходів. В межах попередніх пілотних проектів Німецького агентства з навколишнього середовища була розроблена методологія хвостосховища для підтримки регіональної та місцевої оцінки безпеки. Яка включає оцінку потенціалу небезпеки на основі індексу для великої кількості хвостосховищ, так званий індекс небезпеки хвостосховищ *THI*, і детальний контрольний список для аналізу безпеки окремих хвостосховищ. Спираючись на сильні сторони методики, а також покращуючи та адаптуючи її на основі сучасних технічних знань і найкращих доступних технологій отримуємо набір практичних інструментів для покращення умов безпеки хвостосховищ. *THI* вже довів свою корисність у спрямуванні обмежених фінансових і кадрових ресурсів країни на хвостосховища, що представляють найвищий потенціал небезпеки. Критерії, що лежать в основі підходу *THI*, були покращені завдяки використанню результатів історичного аналізу відмов хвостосховища. Оскільки *THI* враховує лише потенціал небезпеки, потенційні наслідки окремих аварій хвостосховищ, що становлять різні загрози для навколишнього середовища та населення, не враховуються. Цю проблему було вирішено шляхом визначення зони потенційного ризику поблизу хвостосховища на основі розмірів попередніх аварій для оцінки навколишнього середовища (водної екосистеми) та населення, що перебувають у зоні ризику. Результат цього підходу перетворює *THI* на індекс ризику хвостосховища – *TRI*, який навіть краще відображає найнебезпечніші хвостосховища в одній країні з точки зору потенційно постраждалого населення та навколишнього середовища (https://www.sendaiplatform.org).

За допомогою цього методу велика кількість хвостосховищ може бути відсортована та розставлена за пріоритетністю відповідно до розрахованого потенціалу небезпеки.

Метод *THI* враховує такі параметри, які були визначені як найважливіші:

– загальна місткість хвостосховища;

– токсичність речовин хвостів, що зберігаються;

– статус управління хвостосховищем;

– природні умови, характерні для ділянки хвостосховища;

– параметри безпеки дамби.

Для підготовки даних для машинного навчання розкриємо більш детально ці параметри.

Загальна місткість хвостосховища THI_{Cap}. Припускається, що цей параметр зростає зі збільшенням об'єму за логарифмічним співвідношенням з основою 10. Таким чином, збільшення об'єму хвостосховищ у 10 разів буде

означати збільшення індексу на одиницю, що розраховується за формулою $THI_{Cap} = Log_{10}[V_t]$, де V_t – загальний об'єм хвостосховищ у хвостосховища, м^3.

Токсичність речовин хвостів, що зберігаються THI_{tox}. Для інтегрованої характеристики токсичності вкрай важливо мати параметр, який представляє всі потенційні загрози для водної екосистеми в короткостроковій і довгостроковій перспективі. Методи об'єднують всі потенційні загрози для водних екосистем, включаючи гостру та хронічну токсичність, а також біологічні накопичення та накопичує небезпеки для різних організмів (риб, ракоподібних, бактерій та ін.). Ці дані доступні в мережі Інтернет для, приблизно, 7000 речовин (https://www.sendaiplatform.org).

Статус управління хвостосховища **–** це статус хвостосховища, який слід ідентифікувати з чотирьох варіантів. Статистика аварій на хвостосховищах (Rico et al., 2008a, 2008b) показує, що закриті та відновлені хвостосховища безпечніші з точки зору частоти аварій. На цих хвостосховищах аварій не зафіксовано. З цієї причини передбачається, що параметр, пов'язаний з управлінням хвостосховищами, є нижчим для закритих або реконструйованих споруд порівняно з активними хвостосховищами.

Природні умови, характерні для ділянки хвостосховища. Параметр «Природні умови» THI_{Nat} пов'язаний з екологічними ризиками, які дуже часто пов'язані з аварією хвостосховища. Особливо землетруси, сильні зливи та повені багато разів класифікувалися як причини аварій на хвостосховищах. Тому, відповідний потенціал небезпеки розраховується за таким рівнянням: $THI_{Nat} = THI_{Seism} + THI_{Flood}$, де THI_{Seism} – індекс небезпеки для сейсмічної активності, а THI_{Flood} – індекс небезпеки для затоплення на основі геологічних і гідрологічних умов майданчика хвостосховища. Вплив повеней THI_{Flood} пов'язаний із зонами, схильними до повеней, зі статистичним параметром *HQ*-500, який кількісно визначає частоту повеней із повторюваністю один раз у п'ятсот років.

Параметри безпеки дамби. Стійкість дамби є, мабуть, найбільш критичним параметром у оцінці небезпеки. Вважається, що параметр «Умови греблі» THI_{Dam} пов'язаний із проектним параметром греблі «Коефіцієнт безпеки» *FoS*, який має бути розрахований вже на етапі проектування хвостосховища та відноситься до стійкості схилу греблі (Coduto, 1998; Круз та ін., 2008; Фредлунд та ін., 2012). Термін *FoS* зазвичай використовується для вираження запасу міцності схилів на насипних дамбах.

Метод індексу ризику відходів *TRI*. Оцінка *TRI* враховує загальний потенціал небезпеки, а також населення та водні тіла нижче за течією як потенційні об'єкти ризику впливу в разі аварії.

Оскільки соціально-економічні цінності ризику та вразливості потенційних рецепторів можна оцінити лише шляхом детальної оцінки, підхід *TRI* не включає ці аспекти. Будь-яка подальша детальна оцінка ризику для окремих хвостосховища для підтримки планування на випадок надзвичайних ситуацій або специфічної оцінки безпеки повинна включати більш конкретні аспекти та інформацію безпосередньо на ділянці та навколо неї.

TRI розраховується на основі значень *THI* і *TEI* за є їх додатком, тому розрахунок виглядаю так: $TRI = THI + TEI$. Подібно до *THI*, *TEI* та *TRI* також слід оцінювати за логарифмічною шкалою і визначається за такою формулою: $TEI = TEI_{Pop} + TEI_{Env}$. Параметр TEI_{Pop} є фактором, що враховує населення нижче за течією, розташоване до 10 км від хвостосховища *PAR*. Коефіцієнт TEI_{Pop} визначається простою класифікацією.

Вплив на навколишнє середовище TEI_{Env} – це коефіцієнт, який враховує розмір найближчої до хвостосховища водойми, яка розташована нижче за течією в межах 10 км від хвостосховища і може бути забруднена аварією на хвостосховища. Коефіцієнт TEI_{Env} визначається на основі середнього значення річкового стоку або площі поверхні озера.

Отже вхідний набір даних за 2022 рік містить 346 записів за 13 показниками хвостосховищ Європи. Вибрані перших п'ять даних для аналізу та кластеризації регіонів показники наведено у таблиці 1.

На цьому формування вхідних даних для проведення кластеризації з використанням індексу небезпеки хвостосховищ є достатнім. Тому перейдемо до самих методів, що забезпечують таке визначення.

Наступним виконаємо кластеризацію з використанням індексу небезпеки хвостосховищ методами машинного навчання. Незважаючи на різні підходи до вирішення завдань кластеризації, всі методи ґрунтуються на поданні об'єктів, які пов'язані між собою, що дозволяє оцінити їх близькі властивості. Тому при простій кластеризації об'єктів за декількома числовими параметрами найбільше поширеними у спеціалізованому програмному забезпеченні залишаються ієрархічна кластеризація та метод *k*-середніх, що відрізняється відносною простотою, високою якістю одержуваних результатів, їх інтерпретованістю та широкими можливостями

Таблиця 1

Вхідні дані хвостосховищ на основі індексу потенціалу небезпеки та оцінки ризиків

код	Склад відходів	Індекс небезпеки хвостосховищ							Індекс ризику відходів			
		THI_{Cap}	THI_{Tox}	THI_{Man}	THI_{Nat}	THI_{Dam}		THI	TEI_{Pop}	TEI_{Env}	TEI	TRI
					THI_{Seism}	THI_{Flood}	THI_{Dam}					
AT1	шлам	6,26	1,00	3,00	0,00	0,00	1,00	11,26	4,00	2,00	6,00	17,26
AT2	шлам	5,70	1,00	0,00	1,00	1,00	1,00	9,70	4,00	3,00	7,00	16,70
AT3	шлам	5,48	3,00	0,00	1,00	1,00	1,00	11,48	4,00	2,00	6,00	17,48
AT4	шлам	5,70	2,00	0,00	1,00	0,00	1,00	9,70	4,00	2,00	6,00	15,70
AT5	шлам	5,45	1,00	0,00	1,00	0,00	1,00	8,45	4,00	2,00	6,00	14,45

з налаштування правил групування. Відмінність даних методів полягає лише в тому, що перші починають алгоритм *n* елементів (класів) і далі об'єднують близькі на відстані групи об'єктів, поки не залишиться всього один клас, а другі навпаки починають алгоритм з одного класу та поділяють далекі групи, доки буде досягнуто групування *n* об'єктів на *n* класів. З чого стає зрозуміло, що до беззаперечних переваг цього методу належить можливість побудови дендограм, тобто дерев, на яких чітко видно етапи класифікації та відстань між класами.

Розглянемо алгоритм DBSCAN для кластеризації хвостосховищ. Як відомо, цей алгоритм заснований на щільності для заданої множини точок у деякому просторі, який відносить в одну групу точки, що розташовані найбільш щільно (точки з багатьма сусідами) та розмічає точки, які лежать в областях з малою щільністю (чиї сусіди розташовані занадто далеко) як викиди (шум), тому визначимо епсилон-околицю об'єкта, де для будь-якого вектору *x* в метричному його просторі ознак визначається як безліч точок, що віддаляються від *x* не більше ніж на відстань ε, тому отримаємо вираз: $U_z(x)=\{u\in U:\rho(x,u)\leq\varepsilon\}$, де $\rho(x, u)$ – евклідова відстань, як метрика простору ознаки. З виразу зрозуміло, що параметр $\varepsilon > 0$, тому задаємо його як параметр роботи алгоритму DBSCAN. Далі, на основі отриманої епсилон-околиці об'єктів згрупуємо їх на три типи: кореневий (що містить не менше *m* об'єктів у своїй епсилон-околиці, $|U_z(x)|\geq m$); граничний (не кореневий, але в околиці кореневого); шумовий (не кореневий і не граничний). З чого зрозуміло, що даний алгоритм заснований на евристиках, а не на математиці, де розрахунки більш точні та чіткі. При виконанні алгоритму отримаємо результати кластеризації, що зазначені на рис. 1, де зеленим кольором позначено основний кластер, синім та голубим – граничний та червоним – шумовий.

Аналізуючи отримані результати, можна зробити висновок, що цей алгоритм дозволяє нам отримати необхідне рішення в межах розглянутого завдання та виділити 5% шумових точок. Однак алгоритм стає нестійким і час від часу видаляє частину вибірки, вважаючи визначені точки шумом. Тому цей алгоритм може застосовуватися для рішення розглянутої задачі за умов, що у ньому зазначено, що точка не повинна мати сусідів для того, щоб потрапити в кластер, який допомагає розділити основні кластери та шумові точки.

Наступним розглянемо агломеративну ієрархічну кластеризацію, яка є сукупністю алгоритмів упорядкування даних, вкладених у створення ієрархії, що складається з груп точок, що спостерігаються. Вихідними для проведення кластерного аналізу служить матриця відстаней між об'єктами, сформована з використанням тієї чи іншої метрики. Агломеративна кластеризація починається з *n* кластерів, де *n* – число спостережень і передбачається, що кожне з них є окремим кластером. Потім алгоритм намагається знайти і згрупувати найбільш схожі між собою точки даних та з цього починається формування груп.

Отже, маємо набір даних і кожен об'єкт – це незалежний кластер. Потім, відповідно до заданої метрики між об'єктами: $\rho(x_i, x_j)$, $i, j = 1, \ldots, l$, вибираємо два найближчих таксони *U*, *V* і об'єднуємо їх в один W:$W = U\cup V$. Тепер у нас з'явилася група з двох об'єктів. Тут є безліч евристики: відстань ближнього сусіда, відстань далекого сусіда, групова середня відстань, відстань між центрами, відстань Уорда. Так от, виявляється, всі ці варіанти можна описати одним виразом: $R_{WS} = \alpha_U R_{US} + \alpha_V R_{VS} + \beta R_{UV} + \gamma|R_{US} - R_{VS}|$, де α_U, α_V, β, γ – деякі параметри, що визначають вид метрики між кластерами (Wozniak M., 2021).

Цей вираз називається формулою Ланса-Уільямса і для зазначених способів обчислень відстаней між таксонами, коефіцієнти набувають вигляду:

$$\alpha_U = \frac{|S|+|U|}{|S|+|W|},\ \alpha_V = \frac{|S|+|V|}{|S|+|W|},\ \beta = \frac{-|S|}{|S|+|W|},\ \gamma=0. \quad (1)$$

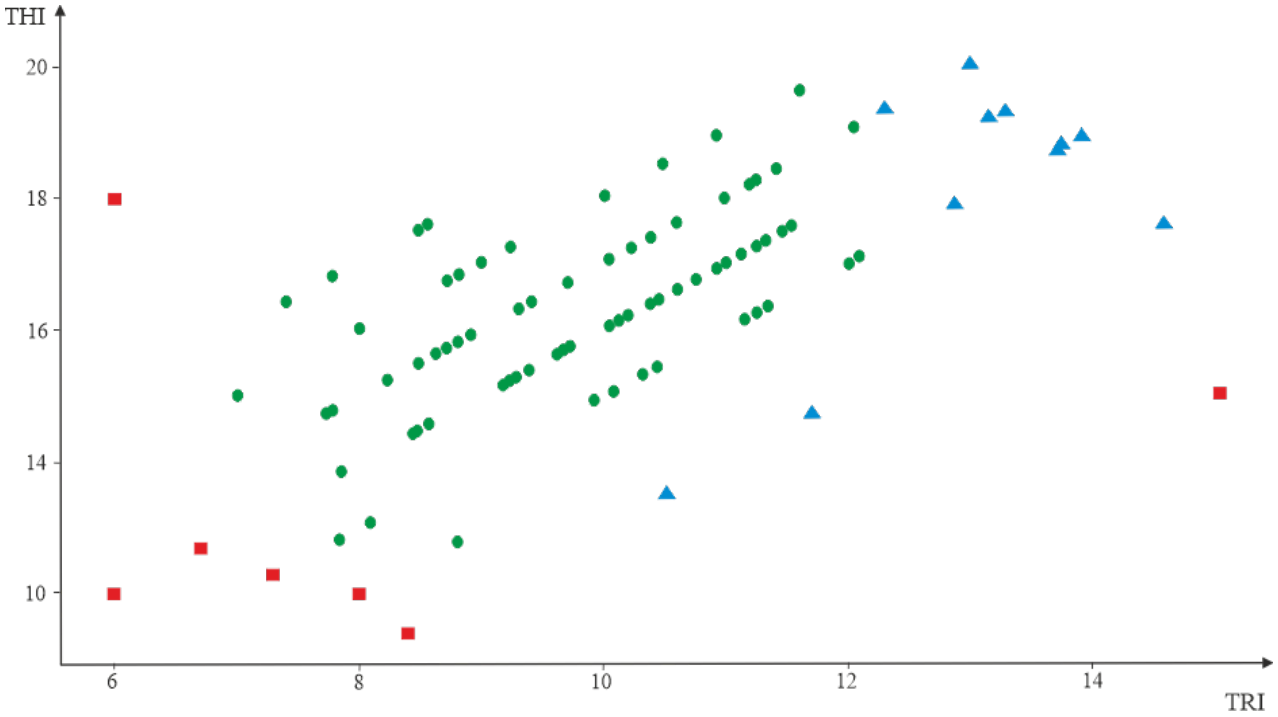


Рис. 1. Візуалізація кластеризації даних з допомогою алгоритму DBSCAN

Для проведення агломеративної ієрархічної кластеризації необхідно виконати наступний порядок дій. По перше, визначимо кількість кластерів, де $C_1=\{\{x_i\},\ldots,\{x_j\}\}$, – безліч одноелементних кластерів, $R\{x_i\}\cdot\{x_j\}= \rho(x_i, x_j)$ – метрика (відстань) між окремими об'єктами для всіх t = 2, …, l, де t – номер ітерації. Знайти в C_{t-1} пару кластерів U, V з мінімальною відстанню R_{UV}. Далі поєднуємо їх в один кластер: $W = U\cup V$. $C_t = C_{t-1}\cup\{W\}/\{U, V\}$. Тому для всіх $S\in C_t$. Наступним обчислимо R_{WS} за формулою Ланса-Уільямса: $R_{WS}=\alpha\cup R_{US} + \alpha_V R_{VS} + \beta R_{UV} + \gamma|R_{US}-R_{VS}|$.

Для візуалізації процесу ієрархічної кластеризації побудуємо дендрограмму. Де по вертикалі відкладемо мінімальну відстань R_t між кластерами, а по горизонталі – об'єкти (хвостосховища). Такий графік дозволяє побачити в якому порядку відбувалося об'єднання даних у групи і наскільки кластери відокремлені один від одного по мінімальній відстані (рис. 2а). Інший вигляд цієї ж інформації наведено на рис. 2б, де показані результати розбиття хвостосховищ в плані.

На рис. 2а можна в обох випадках спостерігати тупикові гілки дендрограми (хвостосховища 43, 42, 40; 34, 35), які закінчуються на рівнях побудови дерева більше або рівних 5 і 7, що є аномаліями класифікації.

Виходячи з рис. 2а ми бачимо, що найефективніше можемо поділити кластери на три групи. На основі оцінки дендрограми та відстані між кластерами на співвідношенні додавання кластера до кластера та граничної відстані на якому відбувається їх об'єднання можна поділити вибірку на три основі кластери. Виходячи з такого ділення можемо сказати, що ці хвостосховища достатньо чітко об'єднуються в один таксон і це проходить до верху крізь всі рівні, крім аномальних значень, які були розглянуті вище.

На рис. 2б представлені хвостосховища, що поділені на три групи та в яких показаний чіткий поділ кольорами. До синього кластеру (на рисунку позначено трикутниками) класифікуємо хвостосховища, що показують найбільшу ємність разом з достатньо низьким рівнем токсичності. Такі хвостосховища мають низький рівень небезпеки у випадку аварії. До зеленого (на рисунку позначено колами) та червоного (на рисунку позначено квадратами) кластерів відносяться група хвостосховищ, що характеризуються ємністю менше середньої та більш високим рівнем токсичності. Такі хвостосховища мають достатньо високу токсичність речовин, і відповідно високий рівень небезпеки.

Наступним методом кластеризації застосуємо метод k-середніх. Вхідні дані з табл. 1 за результатами аналізу кореляції факторів були скориговані та спрямовані на проведення кластерного аналізу з метою визначення наявності у структурі даних кластерів, де виконаємо перебір кількості кластерів (Astakhova, 2015).

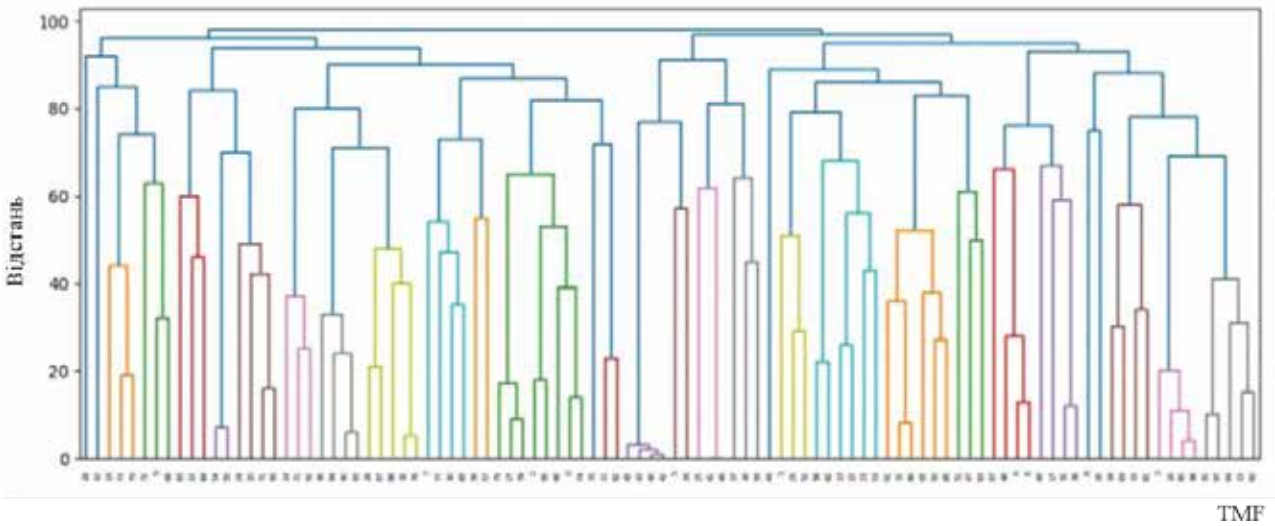


Рис. 2а. Агломеративна ієрархічна кластеризація (дендрограма)

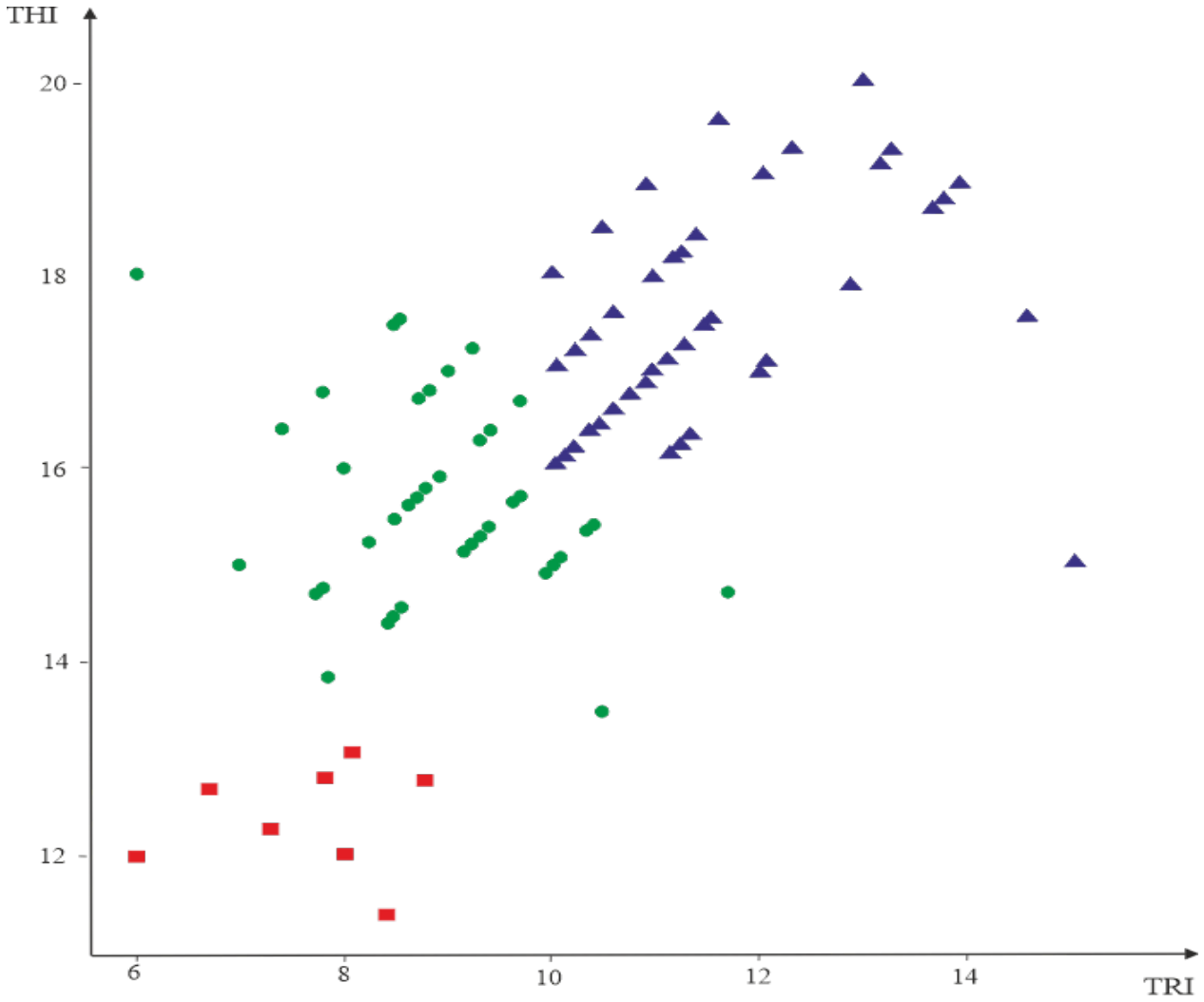


Рис. 2б. Агломеративна ієрархічна кластеризація

Як відомо, метод *k*-середніх полягає у відшуканні такого розбиття набору з *n* точок $\{x_1, x_2, \ldots, x_n\}$ на *k* непустих кластерів, що не перетинаються, а також щоб була мінімальною сума квадратів відстаней від кожної точки до центра відповідного кластера. Класичний метод *k*-середніх шукає найкращу множину кластерів $\{S_1, S_2, \ldots, S_k\}$ шляхом мінімізації цільової функції: $\sum_{h=1}^{k}\sum_{x_i \in S_h} x_i - c_h^{\ 2} \rightarrow \min_{S_1, S_2, \ldots, S_k}$, де c_h – центр *h*-го кластера, застосуємо це для пошуку кластерів у хвостосховищах. Такий підхід дозволить графічним способом визначити кількість кластерів, яке буде відповідати точці, де спад своїх значень уповільнюється найбільш сильно. Така точка знаходиться на рівні чотирьох кластерів. Отже, було визначено, що в заданій структурі даних можна виділити чотири кластери, що відрізняються між собою за якимись, поки що не встановленими, факторами.

Отриманий розподіл на кластери говорить про наявність факторів, що впливають на структуру даних, і підтверджують гіпотезу про можливість кластеризації хвостосховищ, проте не розкривають сутність цих факторів. Для визначення факторів, які впливають структуру даних чинників, використаємо метод алгоритму кластеризації Ллойда. Цей метод дозволяє знизити розмірність багатовимірних даних та отримати їх візуальну структуру без істотної втрати якості інформації.

Для покращення інтерпретації структури даних вони були зображені на рис. 3.

Як осі координат використані значення головних компонент (1 і 2 відповідно). У цих координатах були представлені хвостосховища (кожен кластер від 0 до 3-го показаний різним кольором). На рис. 3 спостерігається досить чіткий поділ на чотири кластери – темно-червоні квадрати праворуч (перший за зростанням кластер), чорні зірки вгорі по центру (другий за чисельністю кластер), зелені кола зліва (третій кластер), сині трикутники в нижній частині графіка (найменший за чисельністю кластер). Кількість кластерів, що візуально спостерігається на рис. 3, збігається з кількістю отриманих іншими способами.

Таким чином, за допомогою методу кластерного аналізу вихідний набір даних може бути згрупований на чотири кластери. Візуалізація вихідних даних шляхом основних компонентів підтверджує наявність чотирьох груп точок даних. З урахуванням отриманих основних компонентів було визначено межі технічних показників кластерів.

На дисперсію даних впливають дві основні компоненти, перша – ємність хвостосховищ, та друга – оцінка токсичності. Друга компонента пов'язана з оцінкою ризику впливу на населення та навколишнє середовище. Розглянемо детальніше отримані кластери.

До першого кластера відноситься група хвостосховищ, яку можна охарактеризувати як

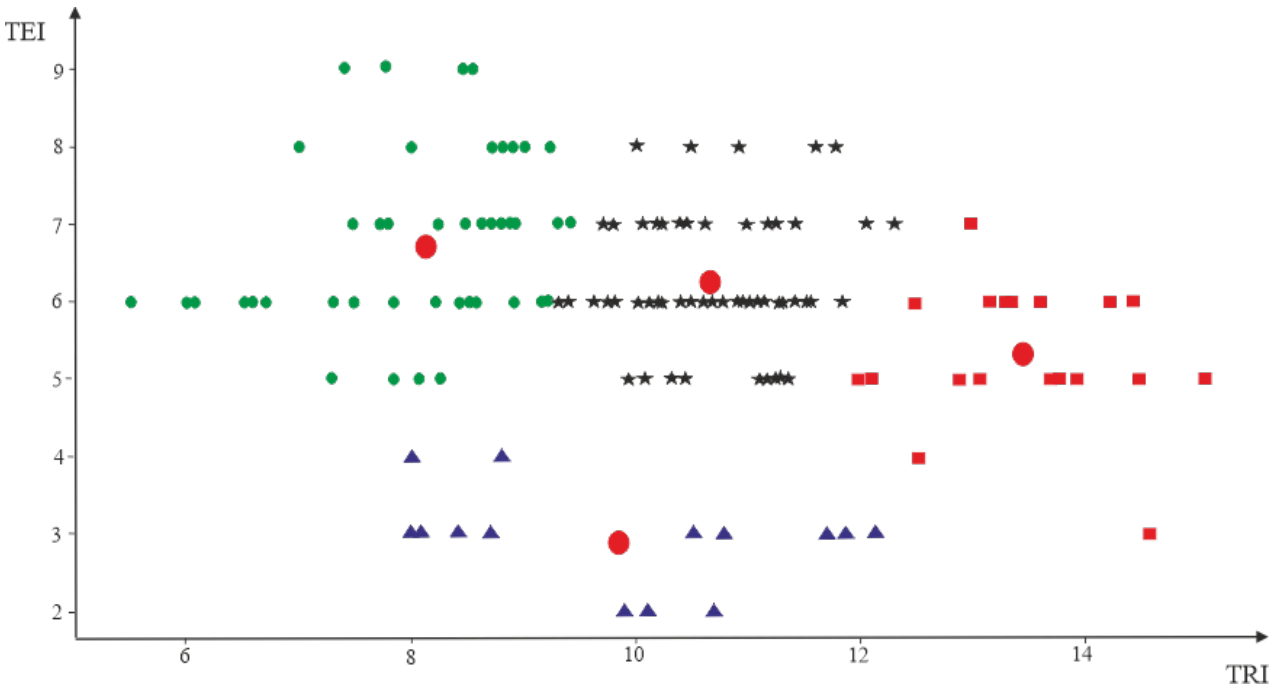


Рис. 3. Візуалізація даних з допомогою методу кластеризації Ллойда

середні за ємністю та низькою токсичністю. Ці хвостосховища можуть складатися з шламу різної крупності із загальним об'ємом близько 1650000 м^3. Такі хвостосховища мають низьку токсичність речовин хвостів, а відповідно і низький рівень небезпеки на навколишнє середовище та на населення. Як правило такі хвостосховища з низьким рівнем аварії з нестійкою дамбою, що відноситься до сприятливих природних умов.

До другого кластеру відносяться група хвостосховищ, що характеризуються найбільшою ємністю разом з низьким рівнем токсичності. Ці хвостосховища можуть складатися з шламу із загальною місткістю більше 10580000 м^3. Такі хвостосховища мають низьку токсичність речовин відходів, а відповідно і низьким рівнем небезпеки для навколишнього середовища і населення. Як правило такі хвостосховища мають низький рівень аварій з нестійкою дамбою, що відноситься до сприятливих природних умов.

До третього кластеру відносяться група хвостосховищ, що характеризуються найменшою ємністю разом та високим рівнем токсичності. Ці хвостосховища можуть складатися з червоного мулу або глини із відносно невеликою загальною місткістю близько 780000 м^3. Такі хвостосховища мають високу токсичність речовин відходів, а відповідно високим рівнем небезпеки на навколишнє середовище та відповідно – на населення з нестійкою дамбою. У випадку аварії на такому хвостосховищі наслідки для навколишнього середовища та населення несприятливі.

До четвертого кластеру відносяться група хвостосховищ, що характеризуються середньою ємністю разом та високим рівнем токсичності. Ці хвостосховища можуть складатися з шламів із загальним об'ємом близько 3880000 м^3. Такі хвостосховища мають високу токсичність речовин хвостів, а відповідно і високий рівень небезпеки на навколишнє середовище та на населення. У випадку аварії на такому хвостосховищі наслідки для навколишнього середовища та населення має неприємні наслідки, але не несприятливі.

Таким чином, встановлені за допомогою методу *k*-середніх індексу потенціалу небезпеки хвостосховищ в Європейському союзі дозволять їх розподілити на групи за їх небезпечністю для навколишнього середовища так наслідками у випадку їх аварії.

Порівнюючи метод агломеративної кластеризації та метод кластеризації Ллойда можна дійти до висновку, що другий метод є більш точний, що дозволяє більш якісно здійснити аналіз хвостосховищ.

Висновки і перспективи подальших досліджень. В статті виконано аналіз небезпечного впливу хвостосховищ утворених від промислового виробництва на навколишнє середовище методами машинного навчання. При різних розмірах хвостосховищ, їх токсичності та технологічних параметрів визначено їх положення в кластерах, що дозволило класифікувати стан рівня їх небезпеки внаслідок аварії. Зроблено порівняльний аналіз підходів методів машинного навчання для визначення потенційної небезпеки хвостосховищ. За допомогою методу DBSCAN було визначено хвостосховища, які за своїми параметрами потрапляють до кореневого, граничного та шумового кластерів. Далі, всі хвостосховища, що потрапили до кореневого та граничного кластерів, було класифіковано наступними методами кластеризації – *k*-середніх та агломеративним методом. Кластеризація методом *k*-середніх будовалася на вибірках даних хвостосховищ шляхом порівняння метрики відстані і було розподілено на чотири кластери. Агломеративний метод дозволив визначити три групи кластерів. За цими результатами виконано порівняння цих методів та було зроблено висновок, що моделювання та пошук центрів кластеризації за допомогою *k*-середніх є більш комплексним рішенням задачі в порівнянні з методом агломеративної кластеризації. Точність використаного підходу кластеризації методом *k*-середніх виявився вищою, ніж у агломеративному методі кластеризації.

Таким чином, вдалося визначити спосіб кластерного аналізу хвостосховищ для при їх небезпеці для навколишнього середовища внаслідок аварії та використати за рахунок машинного навчання кращий спосіб їх класифікації.

ЛІТЕРАТУРА:

1. Nikolaieva I. O., Rudakov D. V. Development of a checklist for improvement of tailings safety. *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*. 2015. 2, 97–103

2. Proceedings of 7th International Conference Innovation Management, Entrepreneurship and Sustainability (IMES), 2019. Pp. 519–533

3. Kravchenko S., Hryshkun Ye. and Vlasenko O., Metody klasyfikatsii mashynnoho navchannia z vykorystanniam biblioteky scikit-learn, URL: http://tech.vernadskyjournals.in.ua/journals/2020/3_2020/part_1/21.pdf

4. Winkelmann-Oei G., Rudakov D., Shmatkov G., Nikolaieva I. A method for evaluation of tailings hazard. New Developments in Mining Engineering: Theoretical and Practical Solutions of Mineral Resources Mining / editors Bondarenko V., Kovalevska I., Pivnyak G. London: Taylor & Francis Group, 2015. pp. 33–38.

5. URL: https://www.sendaiplatform.org

6. Rico, M., Benito, G., Díez-Herrero, A. Floods from tailings dam failures. J. Hazard. Mater. 2008a. 154, 79–87.

7. Rico, M., Benito, G., Salgueiro, A.R., Díez -Herrero, A., Pereira, H.G. Reported tailings dam failures: a review of the European incidents in the worldwide context. J. Hazard. Mater. 2008b.152, 846–852.

8. Coduto D. P. Geotechnical Engineering: Principles and Practices. 1998.

9. Fredlund D. G., Rahardjo H., Fredlund M. D. Unsaturated Soil Mechanics in Engineering Practice. 2012.

10. Wozniak M. Editorial: Applying Machine Learning for Combating Fake News and Internet/Media Content Manipulation. 2021 URL: https://www.researchgate.net/publication/354462936_Editorial_Applying_Machine_Learning_for_Combating_Fake_News_and_InternetMedia_Content_Manipulation.

11. Astakhova N. N., Demidova L. A., Nikulchev E. V. Forecasting Method for Grouped Time Series with the Use of K-Means Algorithm. *Applied Mathematical Sciences*. 2015. Vol. 9. No. 97. P. 4813–4830.

REFERENCES:

1. Nikolaieva, I. O., Rudakov, D. V. (2015). Development of a checklist for improvement of tailings safety. *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu.* 2, 97–103

2. Proceedings of 7th International Conference Innovation Management, Entrepreneurship and Sustainability (IMES), 2019. Pp. 519–533

3. Kravchenko, S., Hryshkun, Ye. and Vlasenko, O. Metody klasyfikatsii mashynnoho navchannia z vykorystanniam biblioteky scikit-learn, Retrieved from: http://tech.vernadskyjournals.in.ua/journals/2020/3_2020/part_1/21.pdf

4. Winkelmann-Oei, G., Rudakov, D., Shmatkov, G., Nikolaieva, I. (2015). A method for evaluation of tailings hazard. New Developments in Mining Engineering: Theoretical and Practical Solutions of Mineral Resources Mining / editors Bondarenko V., Kovalevska I., Pivnyak G. London: Taylor & Francis Group, pp. 33–38.

5. Retrieved from: https://www.sendaiplatform.org

6. Rico, M., Benito, G., Díez-Herrero, A., 2008a. Floods from tailings dam failures. J. Hazard. Mater. 154, 79–87.

7. Rico, M., Benito, G., Salgueiro, A.R., Díez -Herrero, A., Pereira, H.G., 2008b. Reported tailings dam failures: a review of the European incidents in the worldwide context. J. Hazard. Mater. 152, 846–852.

8. Coduto, D. P. (1998). Geotechnical Engineering: Principles and Practices.

9. Fredlund, D. G., Rahardjo, H., Fredlund, M. D. (2012). Unsaturated Soil Mechanics in Engineering Practice.

10. Wozniak M. (2021). Editorial: Applying Machine Learning for Combating Fake News and Internet/Media Content Manipulation. Retrieved from: https://www.researchgate.net/publication/354462936_Editorial_Applying_Machine_Learning_for_Combating_Fake_News_and_InternetMedia_Content_Manipulation.

11. Astakhova, N. N., Demidova, L. A., Nikulchev, E. V. (2015). Forecasting Method for Grouped Time Series with the Use of K-Means Algorithm. *Applied Mathematical Sciences*. Vol. 9. No. 97. P. 4813–4830.

УДК 811.161.2’373.45:004.056=111
DOI https://doi.org/10.32782/IT/2024-4-12

Уляна ЖОРНОКУЙ
кандидат філологічних наук, завідувач кафедри іноземних мов, філології та журналістики, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе 210, Київ, Україна, 02121
ORCID: *0000-0003-3413-785X*
Scopus Author ID: *57260086300;*

Наталія КОВАЛЕНКО
кандидат філологічних наук, доцент кафедри іноземних мов, філології та журналістики, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе 210, Київ, Україна, 02121
ORCID: *0000-0001-9881-4835*

Бібліографічний опис статті: Жорнокуй, У., Коваленко, Н. (2024). Англомовні запозичення у сфері кібербезпеки на сучасному етапі розвитку української мови. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 92–98, doi: https://doi.org/10.32782/IT/2024-4-12

АНГЛОМОВНІ ЗАПОЗИЧЕННЯ У СФЕРІ КІБЕРБЕЗПЕКИ НА СУЧАСНОМУ ЕТАПІ РОЗВИТКУ УКРАЇНСЬКОЇ МОВИ

У статті здійснено комплексний аналіз англомовних запозичень у термінології кібербезпеки в українській мові.

Метою *дослідження стало визначення особливостей адаптації та функціонування англомовних запозичень у сфері кібербезпеки в українській мові. Задля реалізації завдань в контексті поставленої мети було застосовано* ***методи*** *структурного та лексико-семантичного аналізу, компаративний метод, метод систематизації та класифікації термінологічних одиниць.*

Наукова новизна *здійсненого дослідження полягає в тому, що вперше проведено системний аналіз структурних та функціональних особливостей англомовних термінів кібербезпеки в українській мові з урахуванням неологізмів та сленгових виразів.*

Досліджено основні способи адаптації термінів, включаючи транслітерацію, калькування та створення семантичних еквівалентів. Проаналізовано проблематику перекладу та стандартизації термінів кібербезпеки. Обґрунтовано необхідність систематизації термінологічного апарату та розробки відповідних навчальних матеріалів для забезпечення ефективної міжнародної співпраці, що дало змогу зробити такі ***висновки:*** *англомовні терміни у сфері кібербезпеки, будучи невід’ємною частиною сучасної української мови, забезпечують зв’язок із глобальною спільнотою. Проте їх адаптація та стандартизація потребують подальших досліджень і координації зусиль лінгвістів, перекладачів та фахівців у сфері кібербезпеки.*

Ключові слова: *англомовні запозичення, кібербезпека, термінологія, транслітерація, адаптація, неологізми, переклад.*

Uliana ZHORNOKUI
Candidate of Philological Sciences, Head of the Department of Foreign Languages, Philology, and Journalism, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121, ulianazhornokui@gmail.com
ORCID: *0000-0003-3413-785X*
Scopus Author ID: *57260086300;*

Nataliia KOVALENKO
Candidate of Philological Sciences, Associate Professor at the Department of Foreign Languages, Philology, and Journalism, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121, reseacher.ds@gmail.com
ORCID: *0000-0001-9881-4835*

To cite this article: Zhornokui, U., Kovalenko, N. (2024). Anhlomovni zapozychennia u sferi kiberbezpeky na suchasnomu etapi rozvytku ukrainskoi movy [English borrowings in the cybersecurity sector on the current stage of the Ukrainian language development]. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 4, 92–98, doi: https://doi.org/10.32782/IT/2024-4-12

ENGLISH BORROWINGS IN THE CYBERSECURITY SECTOR ON THE CURRENT STAGE OF THE UKRAINIAN LANGUAGE DEVELOPMENT

*The article deals with a comprehensive analysis of English loanwords in Ukrainian cybersecurity terminology. The **purpose** of the research was to determine the adaptation patterns and functioning of English borrowings in Ukrainian cybersecurity terminology. To achieve the research objectives, the study employed **methods** of structural and lexical-semantic analysis, comparative method, and systematic classification of terminological units.*

*The **scientific novelty** lies in conducting the first systematic analysis of structural and functional features of English cybersecurity terms in Ukrainian, including neologisms and slang expressions.*

*The research examines primary methods of term adaptation, including transliteration, calquing, and semantic equivalents creation. It analyzes challenges in translation and standardization of cybersecurity terms. The study substantiates the need for terminology systematization and development of educational materials to ensure effective international cooperation, leading to the following **conclusions:** English cybersecurity terms, being integral to modern Ukrainian language, facilitate connection with the global community. However, their adaptation and standardization require further research and coordinated efforts of linguists, translators, and cybersecurity specialists.*

***Key words:** English borrowings, cybersecurity, terminology, transliteration, adaptation, neologisms, translation.*

Постановка проблеми. Сфера кібербезпеки продовжує стрімко розвиватись, що, насамперед, зумовлюється швидким збільшенням технологічних загроз та вимог до захисту інформації. Вивчення кібербезпеки набуло особливої важливості в умовах широкомасштабного російського вторгнення в Україну, яке призвело до появи нових форм шахрайства в усіх сферах онлайн-послуг, що підтверджуютть і самі науковці (Жовтяк, 2023; Мосієвич, 2014; Шванова, 2023). Кібербезпека як нова міждисциплінарна галузь знань має на меті захист комп'ютерних систем та додатків, що вимагає ґрунтовного дослідження її термінологічного апарату. Стрімкий розвиток цієї сфери робить таке дослідження особливо **актуальним**. Зважаючи на свою стратегічну важливість, кібербезпека стала не лише пріоритетним напрямком державної політики, але й об'єктом лінгвістичних досліджень. Це пов'язано з тим, що фахівці галузі повинні досконало володіти англійською професійною мовою, яка є основною у глобальному кіберпросторі.

Постановка завдання. У цьому контексті англомовні терміни займають значуще місце, адже галузь залишається практично глобалізованою. Спостерігається, що превалююче місце у лексиці інформаційних технологій належить кібертермінам, більшість яких англомовного походження, зважаючи на міжнародний характер самої мови та й її домінуючу популярність серед опцій мови у додатках чи програмах, якими користуються люди в усьому світі. Таким чином, **метою** нашого дослідження є вивчення лексики цієї сфери задля визначення шляхів потрапляння термінів аналізованої галузі знань з англійської в українську мову, виділення продуктивних і непродуктивних способів творення серед лексики цього типу й окреслення перспектив подальшого розвитку термінології сфери кібербезпеки, що уможливить окреслення проблематики та її інтерпретації іншими зацікавленими обраною нами темою дослідниками.

Аналіз останніх досліджень. Якщо питанням перекладу ІТ-термінології займалися багато як і вітчизняних, так і зарубіжних дослідників (Борщ, 2011; Гаврилова, 2017; Мирошниченко & Шишкова, 2019), то проблемі комплексного аналізу англомовних термінів кібербезпеки ще не приділено достатньо уваги, за винятком окремих розвідок.

По-перше, роботи, присвячені цій тематиці часто висвітлюють окремий аспект аналізованого пласту термінології. Для прикладу, Жовтяк В. А. (Жовтяк, 2023) своє дослідження акцентує на вивченні та аналізі структурних особливостей англомовної термінології кібербезпеки, опрацьованої на основі термінологічних одиниць, дібраних із сучасних англомовних тлумачних словників. У ході здійсненого дослідження науковець, виділивши, що терміни кібербезпеки представлено однослівними, двослівними та багатослівними термінами, резюмує, що «…аналізована термінологія в більшості має іменникове підґрунтя, оскільки однією з основних функцій терміна є номінативна. Структура англомовних термінів кібербезпеки за частинами мови представлено герундієм та його поєднанням із іншими частинами мови, переважно іменником» (Жовтяк, 2023).

По-друге, чимало праць аналізують методично-дидактичні особливості вивчення термінології сфери кібербезпеки здобувачами вищої освіти. Наприклад, Ботвин Т.І. у своїй розвідці (Ботвин, 2022) проаналізовала особливості засвоєння та вживання англомовних термінів сектору кібербезпеки України через укладання глосаріїв та здійснила огляд основних проблемних моментів дотичних до теми статті в умовах воєнного стану. Перш за все, вважає автор, необхідно зробити «вичерпний аналіз всього дидактичного матеріалу освітнього компоненту згідно робочої програми, витягти термінологію, яка використовується мовою оригіналу, а потім надати пояснення термінів рідною мовою» (Ботвин, 2022). Проте, переконані, цей метод веде до розпорошення термінології сфери кібербезпеки за певними курсами, передбаченими освітніми програмами спеціальності «Кібербезпека та захист даних». На нашу думку, навпаки слушно уніфіковувати кібертермінологію задля налагодження продуктивної міжнародної співпраці, що особливо важливо в час динамічного розвитку сектору кібербезпеки та збільшення частотності кібератак як методів тиску та впливу у світовій політиці. В умовах сьогодення галузь кібербезпеки набуває критичного значення, оскільки українські комп'ютерні системи, як державні, так і приватні, зазнають постійних кібератак з боку російського агресора. З огляду на це, надзвичайно важливо, щоб усі користувачі комп'ютерів та початкуючі спеціалісти були обізнані з різними видами інтернет-шахрайства, вміли захищати себе в мережі та вживали превентивних заходів для запобігання подібним загрозам.

По-третє, чимало науковців зупинялися на дослідженні специфіки перекладу англійських лексичних одиниць цієї галузі знань засобами української мови, зосереджуючи свою увагу як на семантичних, фонетичних, так і на словотвірних особливостях. Серед дослідників варто відзначити роботи А. Б. Сарієвої (Сарієва, 2012), яка зосередилася на особливостях перекладу комп'ютерної термінології, виділивши чотири ключових види:

1) «Повне копіювання англомовного написання терміну»[1];

2) «Запозичення основи через транскрипцію, тобто, відтворення звукової форми початкового терміну за допомогою букв українського алфавіту»[2];

3) «Використання семантичних еквівалентів, тобто, реально існуючих українських слів, що повністю або частково відображають значення англомовного терміну»[3];

4) «Морфемне або лексичне калькування, тобто, переклад лексичної одиниці оригіналу через заміну її складових частин на лексичні відповідності в українській мові»[4] (Сарієва, 2012). Погоджуємося з авторкою вище зазначеної праці, що серед способів перекладу термінів лише копіювання та калькування мають чітко визначені сфери використання (Сарієва, 2012). Останній метод найбільш доречний для складних термінів, що містять більше однієї основи. Решта термінів перекладаються за допомогою другої або третьої моделі без чітко встановлених правил, спираючись на усталену практику перекладу. Особливо цікавим є досить розмите розмежування між термінами, що перекладаються за допомогою транскрипції, та тими, що передаються через семантичні еквіваленти.

Серед дослідників, котрих цікавила проблема кібертермінології в перекладацькому аспекті, виділимо розвідку Шванової О. В. Як зазначила сама авторка статті, що у аналізованій роботі доведено на прикладі термінолоічного матеріалу кіберсфери, що «…відтворення термінів, що є усталеними і зафіксовані в лексикографічних джерелах переважно здійснено способом транслітерації, але такий переклад не завжди у повній мірі передає внутрішню конотацію терміносполучення. Пропонуючи власні варіанти перекладу, ми в основному застосовували описовий спосіб, адже вважаємо, що таким чином можна передати всі відтінки внутрішнього значення терміносполучення» (Шванова, 2023).

По-четверте, варто зазначити, що частина розвідок з питань кібертермінології здійснені в контексті діахронічного та синхронічного зрізів. О. В. Гаврилова (Гаврилова, 2017) присвятила

[1] Звернемо вашу увагу на те, що до цього методу перекладу варто звертатися, коли іде мова про назви технологічних стандартів, програмного забезпечення, корпорацій, як-от: Microsoft, IOS, Android, Chat GPT, Claude тощо – У.Ж., Н.К.

[2] Для прикладу, скрінити, фішинг, фейк тощо.

[3] Серед слів цієї групи виділимо такі зразки: «firewall» – «захисна стіна» (система захисту мережі), «backdoor» – «чорний хід» (прихований спосіб доступу до системи), «honeypot» – «пастка» (система-приманка для виявлення зловмисників), «spoofing» – «підміна» (маскування під іншого користувача чи пристрій), «sniffer» – «перехоплювач» (програма для перехоплення мережевого трафіку), «keylogger» – «клавіатурний шпигун» (програма для відстеження натискань клавіш), «sandbox» – «пісочниця» (ізольоване середовище для тестування програм), «payload» – «навантаження» (шкідливий код у складі вірусу), «vulnerability» – «вразливість» (слабке місце в системі захисту).

[4] Наприклад, «cyber-attack» – «кібератака» (cyber – кібер, attack – атака), «malware» – «зловмисне програмне забезпечення» (mal – зло, ware – програмне забезпечення), «password» – «паролеслово» (pass – пропуск, word – слово), «cryptocurrency» – «криптовалюта» (crypto – крипто, currency – валюта), «cyberspace» – «кіберпростір» (cyber – кібер, space – простір), «cybersecurity» – «кібербезпека» (cyber – кібер, security – безпека), «anti-virus» – «антивірус» (anti – анти, virus – вірус), «cybercrime» – «кіберзлочин» (cyber – кібер, crime – злочин), «multi-factor» – «багатофакторний» (multi – багато, factor – фактор), «self-protection» – «самозахист» (self – само, protection – захист).

свої дослідження аналізу сучасного стану української комп'ютерної термінології, а І. Б. Ментинська (Ментинська, 2020) здійснила ґрунтовний аналіз тематичної та лексико-семантичної класифікації українських комп'ютерних термінів, зосередившись на історично-порівняльному методі.

Характерно, що багато із проаналізованих розвідок порушують проблему кібертермінології, наголошуючи на її архіактуальності в час повномасштабного вторгнення, що особливо важливо з кількох причин. Насамперед, російська агресія супроводжується масштабними кібератаками на критичну інфраструктуру України, що вимагає чіткого розуміння та правильного використання відповідної термінології для ефективної протидії цим загрозам. По-друге, активна міжнародна співпраця у сфері кібербезпеки потребує уніфікованого термінологічного апарату для безперешкодного обміну досвідом та інформацією з іноземними партнерами. Важливим аспектом є також необхідність швидкого та точного перекладу документації, інструкцій та рекомендацій щодо кіберзахисту, які надходять від міжнародних організацій та партнерів. Крім того, поява нових видів кіберзагроз та методів захисту вимагає постійного оновлення та систематизації термінологічної бази, що допомагає фахівцям ефективніше реагувати на сучасні виклики. Розвиток та стандартизація української термінології кібербезпеки також сприяє зміцненню національної безпеки та незалежності від російськомовних джерел інформації у цій критично важливій галузі.

Виклад основного матеріалу. Англомовні терміни стали фундаментальною основою для розвитку української термінології кібербезпеки, що зумовлено історичними та технологічними факторами. Оскільки більшість інновацій у сфері інформаційних технологій та кібербезпеки походить з англомовних країн, саме англійська мова стала першоджерелом для формування відповідної термінологічної бази. Це призвело до того, що значна частина української термінології кібербезпеки базується на англомовних термінах, які адаптуються до норм української мови.

У процесі формування української термінології кібербезпеки використовуються різні способи адаптації англомовних термінів. Серед них найпоширенішими є транслітерація та створення семантичних еквівалентів. Такий підхід дає можливість зберегти міжнародну впізнаваність термінів, одночасно забезпечуючи їх природне входження в українську мову.

Важливим аспектом впливу англомовних термінів є їхня роль у міжнародній комунікації та співпраці у сфері кібербезпеки. Знання оригінальних англомовних термінів та їх українських відповідників дозволяє фахівцям ефективно взаємодіяти з міжнародними партнерами, вивчати світовий досвід та швидко реагувати на нові кіберзагрози. Це особливо актуально в контексті інтеграції України до глобальної системи кібербезпеки.

Розвиток української термінології кібербезпеки на основі англомовних термінів також сприяє стандартизації професійної мови та покращенню якості технічної документації. При цьому важливо зберігати баланс між використанням міжнародної термінології та розвитком власне українських термінів, що відповідають нормам і традиціям української мови. Це дозволяє створити ефективну та зрозумілу систему термінів, яка відповідає як міжнародним стандартам, так і потребам українських фахівців.

Згадані особливості виділяла також В. А. Жовтяк Притаманні для термінології саме сфери кібербезпеки є риси на зразок: «1) інтеграція, що трапляється завдяки тому, що кібербезпека є ключовою ланкою в таких сферах, як бізнес (personal information breach), банківська справа, освіта (identity theft), тощо; 2) інтернаціональний характер, що відбувається завдяки термінам зі схожим написанням і значенням у принаймні трьох неспоріднених мовах (наприклад, cache, domain, host, hub, spam); 3) економія мовних ресурсів, реалізована за допомогою скорочення: URL – Uniform Resource Locator, POP 3 – Post Office Protocol, Version 3» (Жовтяк, 2023).

За даними досліджень (Шванова, 2023), три чверті термінів з галузі кібербезпеки запозичені без значних змін. Прикладами поширених термінів є «firewall», «phishing», «malware», «spoofing» тощо. Особливо цікаво, що багато з них запозичені з сленгу хакерів, що додає їм цікавих соціокультурних конотацій.

Англомовні терміни відіграють ключову роль у формуванні української термінології кібербезпеки, відображаючи глобальний характер цієї галузі та необхідність інтеграції в міжнародний інформаційний простір. Галузь інформаційних технологій у наш час є домінуючою за швидкістю розвитку, а, отже, цілком логічно, що «ця термінологія уособлює собою епіцентр найактивнішої частини словника, що постійно оновлюється та розвивається» (Мосієвич, 2014). Дослідження засвідчило, що утворення англомовної комп'ютерної лексики нерозривно пов'язане з процесом словотворення похідних

лексем та введенням нових слів у словниковий запас певної мови.

Для термінології кібербезпеки притаманна низка неологізмів, сленгізмів. Багато нових технічних термінів і концепцій з'являються у кібербезпеці через розробку нових технологій та методів атак, і їхнє введення часто відбувається англійською мовою. Ще однією особливістю термінології з кібербезпеки є постійне поповнення неологізмами (Мосієвич, 2014). Неологізм – це слово або зворот, що створено для позначення нового предмета або для вираження нового поняття. Згідно з М. І. Мостовим, чітких критеріїв визначення неологізму як мовного явища немає, а спроби відмежувати його за певними правилами поки що позитивних наслідків не дають (Мостовий, 1993).

Дослідження свідчать, що у сфері кібербезпеки значна частина англомовних термінів увійшла до української мови майже без змін або з мінімальною адаптацією, що зумовлено їхнім походженням та міжнародним характером. Зокрема, аналіз етимології цих термінів показує, що три чверті з них мають латинське або грецьке походження, що полегшує їх інтеграцію в українську мову (Гладун & Пучков, 2018). Яскравими прикладами є терміни «authentication» (автентифікація), «verification» (верифікація), «configuration» (конфігурація), «modification» (модифікація), які походять з латинської мови. Також широко використовуються терміни грецького походження, такі як «cryptography» (криптографія), «protocol» (протокол), «analysis» (аналіз), «algorithm» (алгоритм). Ця особливість значно спрощує процес адаптації термінів, оскільки подібні слова вже давно функціонують в українській науковій термінології різних галузей і легко адаптуються до української мови завдяки спільним словотворчим моделям та морфологічним елементам. Особливо це стосується термінів з префіксами «cyber- » (кібер-), «multi-» (мульти-), «inter-» (інтер-), «trans-» (транс-) та суфіксами «-tion» (-ція), «-ment» (-мент), «-ity» (-ість), які мають прямі відповідники в українській мові та дозволяють зберегти міжнародну впізнаваність термінів при їх перекладі.

Структурні особливості англомовних термінів також впливають на їх адаптацію. Багатокомпонентні терміни, такі як «denial of service attack» або «advanced persistent threat», часто перекладаються дослівно або калькуються, що може призводити до складнощів у розумінні та використанні (Жовтяк, 2023).

Зауважимо, що неологізми та сленгові вирази серед кібертермінів активно формуються та швидко поширюються в англомовному середовищі. Ці лексичні одиниці, що виникають як відповідь на нові виклики цифрової епохи, відіграють ключову роль у професійній комунікації, водночас створюючи значні труднощі для перекладу. Їхня унікальність полягає в поєднанні технічних понять і метафоричних значень, що вимагає від перекладача не лише володіння фаховою термінологією, але й глибокого розуміння контексту їх використання.

Терміни на кшталт «phishing», «spoofing» чи «ransomware» потребують не лише точного перекладу, але й адекватного тлумачення з урахуванням культурно-лінгвістичних особливостей цільової аудиторії. Помилки у передачі таких понять можуть спричинити серйозні комунікативні непорозуміння, що особливо критично в умовах міжнародної співпраці у сфері кіберзахисту. Наприклад, «phishing» у прямому перекладі як «фішинг» потребує супровідного пояснення щодо шахрайського характеру цього явища, що спрямоване на отримання конфіденційних даних. Аналогічно, переклад «spoofing» як «підробка» вимагає уточнення способів маніпулювання імітацією довірених джерел. Таким чином, ефективний переклад таких термінів забезпечує точність комунікації та сприяє професійному обміну знаннями. (Ніколіч, 2021).

Висновки і перспективи подальших досліджень. Англомовні терміни у сфері кібербезпеки є невід'ємною частиною сучасної української мови, забезпечуючи зв'язок із глобальною спільнотою. Проте їх адаптація та стандартизація потребують подальших досліджень і координації зусиль лінгвістів, перекладачів та фахівців у сфері кібербезпеки.

Подальші дослідження можуть зосередитися на проблемах уніфікації термінології та розробки стандартів для перекладу та використання англомовних термінів у сфері кібербезпеки з метою забезпечення єдності та зрозумілості у міжмовній комунікації; аналізу впливу англіцизмів на структуру та розвиток української мови; розробки навчальних матеріалів: створення посібників та словників, які допоможуть фахівцям та студентам краще розуміти та використовувати термінологію кібербезпеки; дослідження динаміки термінотворення з особливим акцентом на особливостях адаптації нових термінів у відповідь на швидкий розвиток технологій та загроз у кіберпросторі.

Таким чином, англомовні терміни є невід'ємною частиною української термінології кібербезпеки, проте їх адаптація та використання потребують подальшого вивчення та уніфікації для забезпечення ефективної комунікації та розвитку галузі.

ЛІТЕРАТУРА:

1. Борщ І. В. Особливості перекладу сучасних термінів комп'ютерної сфери з англійської українською мовою. Нова філологія. Запоріжжя : Запорізький національний університет. 2011. № 45. С. 175–177.

2. Ботвин Т. Щодо англомовної термінології системи сектору кібербезпеки в умовах воєнного стану: аналіз дефініцій. *Вчені записки ТНУ імені В. І. Вернадського. Серія: Філологія. Журналістика,* 2022. Том 33 (72), № 6, Ч. 1. С. 90–94. DOI: https://doi.org/10.32782/2710-4656/2022.6.1/15.

3. Гаврилова О. В. Місце комп'ютерної термінології в українській мові. *Лінгвістичні дослідження: Зб. наук. праць ХНПУ ім. Г.С. Сковороди,* 2017. Вип. 45. С. 189–193.

4. Гладун А. Я., Пучков О. О. Етимологічний аналіз англомовних термінів кібербезпеки. *Науковий вісник Національного університету біоресурсів і природокористування України. Серія: Філологія*, 2018. (292), 45–50. URL: http://publications.lnu.edu.ua/collections/index.php/foreighnphilology/article/view/4485/4971 (дата звернення – 23.12.2024).

5. Гладун А. Я., Пучков, О. О., Субач І. Ю., Хала К. О. *Англо-український словник термінів з інформаційних технологій та кібербезпеки*. ІСЗЗІ КПІ ім. Ігоря Сікорського. 2018. URL: https://ela.kpi.ua/server/api/core/bitstreams/034d3344-9f8a-4697-a679-944144bd95de/content (дата звернення – 20.12.2024).

6. Жовтяк В. А. Структурні особливості англомовних термінів кібербезпеки. *Закарпатські філологічні студії*, 2023. Випуск 34, Том 1. С. 79–84. DOI: https://doi.org/10.32782/tps2663-4880/2024.34.1.13.

7. Кальник О. П., Воробйова О. С., Симоненко А. В., Олешко О. П. Термінологічні проблеми перекладу наукових текстів в сфері ІТ-технологій. *Молодий вчений*. 2019. № 5.1 (69.1). С. 178–190.

8. Ментинська І. Б. Тематична та лексико-семантична класифікація українських українських комп'ютерних термінів. *Вчені записки ТНУ імені В.І. Вернадського. Серія: Філологія. Соціальні комунікації,* 2020. Том 31 (70), № 2, Ч. 1. С. 26–31.

9. Мирошниченко В. М., Шишкова І. С. Англійські лексичні новоутворення у сфері комп'ютерних технологій та особливості їх перекладу українською мовою. *Вісник Національного технічного університету «ХПІ». Серія: Актуальні проблеми розвитку українського суспільства,* Харків. № 1, 2019. С. 87–92.

10. Мосієвич І. Лінгвальні особливості англомовних термінів кібербезпеки та їх переклад українською мовою. *Вчені записки ТНУ імені В. І. Вернадського. Серія: Філологія. Журналістика, 2014.* Том 35 (74), № 5. Частина 1. С. 243–250. URL: https://www.philol.vernadskyjournals.in.ua/journals/2024/5_2024/part_1/42.pdf (дата звернення – 23.12.2024).

11. Мостовий М. І. Лексикологія англійської мови. Харків : Основа, 1993. 256 с.

12. Ніколіч І. Номінація термінів кібербезпеки в сучасній українській мові. *Науковий вісник Міжнародного гуманітарного університету. Серія: Філологія*, 2021. (59), частина 3, 373–377. URL: http://fch.udpu.edu.ua/article/view/4.pdf/4.pdf (дата звернення – 23.12.2024).

13. Сарієва А. Б. Особливості перекладу комп'ютерних термінів. *Наукові записки [Національного університету «Острозька академія»]. Сер.: Філологічна,* 2012. Вип. 25. С. 97–98. URL: http://nbuv.gov.ua/UJRN/Nznuoaf_2012_25_35 (дата звернення – 23.12.2024).

14. Субач І. Ю. Лінгвістичні маркери англомовних термінів кібербезпеки. *Науковий журнал «Грааль науки»*, 2022. № 12. С. 299–305. URL: https://archives.journal-grail.science/index.php/2710-3056/article/view/299(дата звернення – 23.12.2024).

15. Шванова О. В. Особливості перекладу термінів із кібербезпеки. *Науковий вісник Міжнародного гуманітарного університету. Сер.: Філологія*, 2023. № 59, Том 3. С. 180–183.

REFERENCES:

1. Borshch, I. V. (2011). Osoblyvosti perekladu suchasnykh terminiv kompyuternoi sfery z anhliiskoi ukrainskoiu movoiu [Features of translating modern computer terms from English into Ukrainian]. *Nova Filolohiia*, (45), 175–177. Zaporizhzhia: Zaporizhzhia National University [in Ukrainian].

2. Botvyn, T. (2022). Shchodo anhlomovnoi terminolohii systemy sektoru kiberbezpeky v umovakh voiennoho stanu: analiz definitsii [On English terminology in the cybersecurity sector under martial law: Definition analysis]. *Vcheni Zapisky TNU Imeni V. I. Vernadskoho. Seriia: Filolohiia. Zhurnalistyka*, 33(72) (6, Part 1), 90–94. https://doi.org/10.32782/2710-4656/2022.6.1/15 [in Ukrainian].

3. Havrylova, O. V. (2017). Mistse komputernoi terminolohii v ukrainskii movi [The place of computer terminology in the Ukrainian language]. *Linhvistychni Doslidzhennia: Zbirnyk Nauk. Prats KHNPU im. H. S. Skovorody*, (45), 189–193 [in Ukrainian].

4. Hladun, A. Ya., & Puchkov, O. O. (2018). Etymolohichnyi analiz anhlomovnykh terminiv kiberbezpeky [Etymological analysis of English cybersecurity terms]. *Naukovyi Visnyk Natsionalnoho Universytetu Bioresursiv*

i Pryrodokorystuvannia Ukrainy. Seriia: Filolohiia, (292), 45–50. Retrieved from: http://publications.lnu.edu.ua/collections/index.php/foreighnphilology/article/view/4485/4971 [in Ukrainian].

5. Hladun, A. Ya., Puchkov, O. O., Subach, I. Yu., & Khala, K. O. (2018). Anhlo-ukrainskyi slovnyk terminiv z informatsiinykh tekhnolohii ta kiberbezpeky [English-Ukrainian dictionary of terms in information technology and cybersecurity]. ISZZI KPI im. Ihoria Sikorskoho. Retrieved from: https://ela.kpi.ua/server/api/core/bitstreams/034d3344-9f8a-4697-a679-944144bd95de/content [in Ukrainian].

6. Zhovtiak, V. A. (2023). Strukturni osoblyvosti anhlomovnykh terminiv kiberbezpeky [Structural features of English cybersecurity terms]. *Zakarpatski Filolohichni Studii*, 34 (1), 79–84. https://doi.org/10.32782/tps2663-4880/2024.34.1.13 [in Ukrainian].

7. Kalnyk, O. P., Vorobiova, O. S., Symonenko, A. V., & Oleshko, O. P. (2019). Terminolohichni problemy perekladu naukovykh tekstiv v sferi IT-tekhnolohii [Terminological problems of translating scientific texts in the IT field]. *Molodyi Vchenyi*, 5(69.1), 178–190 [in Ukrainian].

8. Mentyntska, I. B. (2020). Tematychna ta leksyko-semantychna klasyfikatsiia ukrainskykh komputerovykh terminiv [Thematic and lexical-semantic classification of Ukrainian computer terms]. *Vcheni Zapysky TNU Imeni V. I. Vernadskoho. Seriia: Filolohiia. Sotsialni Komunikatsii,* 31 (70) (2, Part 1), 26–31 [in Ukrainian].

9. Myroshnychenko, V. M., & Shyshkova, I. S. (2019). Anhliiski leksychni novoformatsii u sferi komputerovykh tekhnolohii ta osoblyvosti yikh perekladu ukrainskoiu movoiu [English lexical innovations in computer technology and their translation into Ukrainian]. *Visnyk Natsionalnoho Tekhnichnoho Universytetu "KhPI". Seriia: Aktualni Problemy Rozvytku Ukrainskoho Suspilstva*, (1), 87–92 [in Ukrainian].

10. Mosiievych, I. (2024). Linhvalni osoblyvosti anhlomovnykh terminiv kiberbezpeky ta yikh pereklad ukrainskoiu movoiu [Linguistic features of English cybersecurity terms and their translation into Ukrainian]. *Vcheni Zapysky TNU Imeni V. I. Vernadskoho. Seriia: Filolohiia. Zhurnalistyka*, 35(74) (5, Part 1), 243–250. Retrieved from: https://www.philol.vernadskyjournals.in.ua/journals/2024/5_2024/part_1/42.pdf [in Ukrainian].

11. Mostovyi, M. I. (1993). Leksikolohiia anhliskoi movy [Lexicology of the English language]. Kharkiv: Osnova [in Ukrainian].

12. Nikolych, I. (2021). Nominatsiia terminiv kiberbezpeky v suchasnii ukrainskii movi [Nomination of cybersecurity terms in modern Ukrainian]. *Naukovyi Visnyk Mizhnarodnoho Humanitarnoho Universytetu. Seriia: Filolohiia*, (59, Part 3), 373–377. Retrieved from: http://fch.udpu.edu.ua/article/view/4.pdf/4.pdf [in Ukrainian].

13. Sariyeva, A. B. (2012). Osoblyvosti perekladu komputerovykh terminiv [Features of translating computer terms]. *Naukovi Zapysky [Natsionalnoho Universytetu "Ostrohka Akademiia"]. Seriia: Filolohichna*, (25), 97–98. Retrieved from: http://nbuv.gov.ua/UJRN/Nznuoaf_2012_25_35 [in Ukrainian].

14. Subach, I. Yu. (2022). Linhvistychni markery anhlomovnykh terminiv kiberbezpeky [Linguistic markers of English cybersecurity terms]. *Naukovyi Zhurnal "Hraal Nauky"*, (12), 299–305. Retrieved from: https://archives.journal-grail.science/index.php/2710-3056/article/view/299 [in Ukrainian].

15. Shvanova, O. V. (2023). Osoblyvosti perekladu terminiv iz kiberbezpeky [Features of translating cybersecurity terms]. *Naukovyi Visnyk Mizhnarodnoho Humanitarnoho Universytetu. Seriia: Filolohiia*, (59, Vol. 3), 180–183 [in Ukrainian].

УДК 004.9
DOI https://doi.org/10.32782/IT/2024-4-13

Денис ІВАНОВ
асистент кафедри інформаційних технологій та комп'ютерної інженерії, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького 19, Дніпро, Україна, 49005
ORCID: *0000-0001-8660-0928*
Scopus-Author ID: *58047075900*

Віта КАШТАН
кандидат технічних наук, доцент, доцент кафедри інформаційних технологій та комп'ютерної інженерії, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького 19, Дніпро, Україна, 49005
ORCID: *0000–0002–0395–5895*
Scopus-Author ID: *57201902879*

Бібліографічний опис статті: Іванов, Д., Каштан, В. (2024). Геоінформаційна технологія виявлення та моделювання затоплених зон міста Дубай на основі згорткової нейронної мережі. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 99–113, doi: https://doi.org/10.32782/IT/2024-4-13

ГЕОІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ВИЯВЛЕННЯ ТА МОДЕЛЮВАННЯ ЗАТОПЛЕНИХ ЗОН МІСТА ДУБАЙ НА ОСНОВІ ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ

Виявлення затоплених зон є важливим для ефективного реагування на стихійні лиха та їх управління, оскільки це дозволяє створювати системи раннього попередження та організовувати цілеспрямовану допомогу.

Мета роботи *полягає у розробці геоінформаційної технології для виявлення та моделювання затоплених зон у Дубаї на основі згорткових нейронних мереж.*

Методологія. *Геоінформаційна технологія дослідження складається з декількох етапів, що забезпечують комплексний підхід до аналізу та моделювання затоплення територій. Спершу здійснюється завантаження супутникових знімків та їх попередня обробка. Це включає корекцію атмосферних спотворень, вирівнювання кольорів та корекцію геометричних викривлень. Після цього знімки перетворюються у векторний формат для полегшення їх подальшої обробки та аналізу. Векторизація дозволяє точніше виділити ключові елементи на знімках, такі як річки, дороги, будівлі тощо. На наступному етапі інтегруються кліматичні та гідрографічні дані, а також цифрова модель рельєфу (ЦМР). Ці дані використовуються для побудови бази геоданих, яка включає інформацію про опади, гідрологічні умови, типи ґрунтів та рослинність на досліджуваній території. Застосування цифрової моделі рельєфу дозволяє враховувати висотні характеристики місцевості, що є критично важливим для точного моделювання водних потоків та зон затоплення. Для обробки та аналізу цих даних застосовуються методи машинного навчання, зокрема трансферне навчання, що дає змогу підвищити точність моделювання зон затоплення. За допомогою гідрологічної моделі розраховуються екстремальні гідрологічні параметри на основі кліматичних та гідрографічних даних, таких як інтенсивність опадів, тип ґрунту, умови землекористування та рослинний покрив. Ці параметри дозволяють створити криву ґрунтового номера та гідрограф, які відображають поведінку водного стоку на території. Наступним кроком є застосування гідравлічної моделі для визначення параметрів потоку, зокрема швидкості течії, глибини води та прогнозування зон потенційного затоплення. Моделювання затоплення базується на імплементації цифрової моделі рельєфу у форматі ASCII-сітки з розміром 30 м. Ця сітка імпортується в геоінформаційну систему (ГІС) QGIS для проведення розрахунків та моделювання зон затоплення. Завершальним етапом є створення карти зон небезпеки затоплення. Ця карта дозволяє ефективно візуалізувати потенційно небезпечні території, що є важливим інструментом для прийняття рішень в управлінні ризиками повеней та розробці заходів щодо захисту територій від затоплень.*

Наукова новизна *отриманих у роботі результатів полягає у застосуванні трансферного навчання для виявлення затоплених територій на основі супутникових знімків. Модифіковано архітектуру нейронної мережі шляхом додавання конволюційних шарів для підвищення точності виділення суттєвих ознак та використання максимального підсумовування для зменшення обчислювальної складності. Крім того, впроваджено щільний і згладжувальний шари для покращення стабільності навчання та запобігання перенавчанню. Запропонований підхід дозволив покращити ефективність прогнозування зон затоплення, підвищити точність класифікації та забезпечити вищий рівень автоматизації аналізу супутникових даних для управління природними ризиками.*

Висновки. *Дослідження було проведено для оцінки впливу потужної системи повільних штормів, що обрушилася на Об'єднані Арабські Емірати у квітні 2024 року, спричинивши значні опади і раптові повені. Для аналізу використано супутникові знімки Landsat-8, отримані до і після повені. Ці знімки були оброблені для отримання бінарних масок, а потім класифіковані на два основні класи: водні об'єкти та земна поверхня. Результати класифікації продемонстрували значне розширення затоплених територій після повені, що дозволило наочно відобразити масштаби та розподіл води. Метрики ефективності моделі, такі як точність, повнота та F1-міра, свідчать про високу точність у виявленні затоплених зон. Зокрема, точність моделі на навчальній вибірці склала 0.88, а на валідаційній – 0.87. Клас затоплених територій (FL) продемонстрував точність 0.91, що перевищує показники для класу інших територій (FV). Аналіз розподілу затоплення з використанням кроку 0,25 метра виявив, що затоплення є нерівномірним і значною мірою залежить від рельєфу місцевості. Найбільш вразливими до затоплення виявилися низько розташовані частини району, де рівень води досягав до 2 метрів.*

Ключові слова: *зона затоплення, згорткові нейронні мережі, класифікація зображень, моделювання, супутникові дані.*

Denys IVANOV
Assistant at the Department of Information Technology and Computer Engineering, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, ivanov.d.v@nmu.one
ORCID: *0000-0001-8660-0928*
Scopus-Author ID: *58047075900*

Vita KASHTAN
Candidate of Technical Science, Associate Professor, Associate Professor at the Department of Information Technology and Computer Engineering, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, kashtan.v.yu@nmu.one
ORCID: *0000–0002–0395–5895*
Scopus-Author ID: *57201902879*

To cite this article: Ivanov, D., Kashtan, V. (2024). Heoinformatsiina tekhnolohiia vyiavlennia ta modeliuvannia zatoplenykh zon mista Dubai na osnovi zghortkovoi neironnoi merezhi [Geoinformation technology for detecting and modeling flooded areas of Dubai based on convolutional neural network]. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 99–113, doi: *https://doi.org/10.32782/IT/2024-4-13*

GEOINFORMATION TECHNOLOGY FOR DETECTING AND MODELING FLOODED AREAS OF DUBAI BASED ON CONVOLUTIONAL NEURAL NETWORK

Detecting flooded areas is essential for effective disaster response and management, as it allows for the creation of early warning systems and the organization of targeted assistance.

The purpose of the work *is to develop geographic information technology for detecting and modeling flooded areas in Dubai based on convolutional neural networks.*

Methodology. *The geoinformation technology of the study consists of several stages that provide a comprehensive approach to the analysis and modeling of flooding. First, satellite images are downloaded and preprocessed. This includes correction of atmospheric distortions, color alignment, and geometric distortions. After that, the images are converted into a vector format to facilitate their further processing and analysis. Vectorization allows you to more accurately identify key elements in the images, such as rivers, roads, buildings, etc. The next step is to integrate climatic and hydrographic data, as well as a digital elevation model (DEM). This data is used to build a geodatabase that includes information on precipitation, hydrological conditions, soil types, and vegetation in the study area. Using a digital elevation model allows us to take into account the altitude characteristics of the terrain, which is critical for accurate modeling of water flows and flood zones. Machine learning methods, such as transfer learning, are used to process and analyze this data, which makes it possible to improve the accuracy of flood zone modeling. The hydrological model calculates extreme hydrological parameters based on climatic and hydrographic data, such as precipitation, soil type, land use, and vegetation cover. These parameters allow for the creation of a soil number curve and hydrograph that reflect the behavior of water runoff in the area. The next step is to apply a hydraulic model to determine flow parameters, such as flow velocity, and water depth, and predict areas of potential flooding. Flooding modeling is based on the implementation of a digital elevation model in ASCII grid format with a size of 30 meters. This grid is imported into the geographic information system (GIS) QGIS to perform calculations and modeling of flood zones. The final step is to create a map of flood hazard zones. This map allows for effective visualization of potentially hazardous areas, which is*

an important tool for decision-making in flood risk management and the development of measures to protect areas from flooding.

The scientific novelty *of the results obtained in this work consists of the application of transfer learning to detect flooded areas based on satellite images. The architecture of the neural network was modified by adding convolutional layers to improve the accuracy of extracting significant features and using maximum summation to reduce computational complexity. In addition, dense and smoothing layers were introduced to improve training stability and prevent overfitting. The proposed approach has improved the efficiency of flood zone forecasting, increased classification accuracy, and provided a higher level of automation of satellite data analysis for natural risk management.*

Conclusions. *The study assessed the impact of slow-moving storms that hit the United Arab Emirates in April 2024, causing significant rainfall and flash flooding. The analysis utilized Landsat-8 satellite images acquired before and after the flood. These images were processed to obtain binary masks and then categorized into two main classes: water bodies and land surfaces. The classification results demonstrated a significant expansion of the flooded areas after the flood, which allowed for a visual representation of the extent and distribution of water. The model performance metrics, such as accuracy, completeness, and F1-measure, indicate high accuracy in detecting flooded areas. In particular, the model's accuracy on the training sample was 0.88, and on the validation sample – 0.87. The flooded areas (FL) class demonstrated an accuracy of 0.91, which is higher than the class of other areas (FV). Analysis of the flood distribution using a 0.25-meter step revealed that flooding is uneven and largely depends on the terrain. The most vulnerable to flooding were low-lying parts of the district, where the water level reached up to 2 meters.*

Key words: *flood zone, convolutional neural networks, image classification, modeling, satellite data.*

Вступ. Повені представляють серйозну глобальну проблему, що становить значну загрозу для громад у всьому світі, викликаючи суттєві соціально-економічні та екологічні наслідки (Deolfa Jose Moises, 2023). Вчасне і точне виявлення та моніторинг повеней мають першорядне значення для ефективного реагування на стихійні лиха. Це дозволяє створювати системи раннього попередження, планувати евакуацію, розподіляти ресурси та цілеспрямовано надавати допомогу постраждалим. Тому, впровадження технологій автоматичного дистанційного зондування є важливим для забезпечення безпеки і добробуту суспільства.

Супутникові знімки стали цінним ресурсом для моніторингу повеней завдяки широкому діапазону зйомки, високій просторовій роздільній здатності, тощо. Проте, величезний обсяг даних і складна природа паводків вимагають передових обчислювальних методів для точного та ефективного аналізу. Картування міських повеней є складним завданням (Hsiao S.C., 2021). По-перше, міське середовище є дуже складним, з великою кількістю водних шляхів, які можуть бути надзвичайно вузькими, інколи менше метра в ширину. По-друге, затоплення у міських районах часто буває неглибоким і тимчасовим, що ускладнює його виявлення і моніторинг. По-третє, наявність ставків та інших водних об'єктів у містах означає, що масштаби затоплення можуть бути непередбачуваними і постійно змінюватися. Традиційні гідрологічні моделі, що використовуються для прогнозування повеней, не завжди можуть ефективно впоратися з цими факторами. Було зроблено багато спроб скласти карту міських повеней і ризиків затоплення за допомогою традиційних методів. Гідрологічні моделі високої роздільної здатності є ефективними в невеликих масштабах, таких як кілька міських кварталів, але обчислювальні ресурси та високоточні вхідні дані, необхідні для належного моделювання міських повеней у масштабі громади, не є широко доступними з поточною технологією. Вищеописані обмеження демонструють необхідність пошуку методів картографування або прогнозування повеней, які є менш обчислювально інтенсивними.

Перевагою дистанційного зондування є можливість виявлення повеней для великомасштабного картографування без потреби у високоточних вхідних даних і інтенсивних обчислювальних процесах, що покращує управління ризиками повеней. У той час як екстремальні повені, особливо ті, що підпадають під категорію 100-річних подій, добре вивчені та детально нанесені на карту в літературі, незначні повені важко нанести на карту та передбачити. Це менш серйозне затоплення, відоме як неприємне затоплення (nuisance flooding, NF), становить меншу небезпеку для життя та майна, але все ще може бути незручним або навіть небезпечним, особливо для водіїв. У цій статті представлено новий підхід на основі глибокого навчання для виявлення повеней на супутникових знімках. Використовуючи сильні сторони цих технологій, запропонована методологія спрямована на розробку точної та надійної моделі виявлення повеней.

Літературний огляд. Контрольовані методи класифікації часто використовуються для задач бінарної або багатовимірної класифікації. Ці

методи, такі як класифікатор K-Nearst Neighbor (K-NN) (Shahabi H., 2020), класифікатор Random Forest (RF) і опорна векторна машина (SVM) (Tavus, B., 2021), були застосовані до зображень Sentinel-1 SAR для виявлення повеней. Штучні нейронні мережі (ШНМ), наприклад, є популярною технікою машинного навчання, яка використовується в супутниковому дистанційному зондуванні та обробці зображень і демонструє великий потенціал для виявлення повеней із супутникових зображень дистанційного зондування. Методи ШНМ успішно застосовувалися для процесів виявлення затоплення, проте ще існує багато невирішених проблем (A. Emily Jenifer, 2022).

Крім того, розроблено системи на основі машинного навчання для виявлення змін між сухими зображеннями та зображеннями повеней, що дозволяє маскувати водойми та нормальний рівень води в озерах і річках. Для виявлення змін (Change Detection) (Li Y., 2018) необхідні принаймні два зображення: контрольне зображення (перед подією) та цільове зображення (під час події) з того самого супутника, орбітальної траєкторії, поляризації та покриття. Визначення порогу між водяними та неводними пікселями за зображеннями зворотного розсіювання Sentinel-1 SAR є критично важливим кроком для виявлення затоплених місць. Для оцінки порогового значення та необхідності відрізнити затоплені території від інших наземних покривів зазвичай використовується метод порогового визначення Оцу (Chawla I., 2020).

У підходах до виявлення змін, де необхідна система нечіткої класифікації, класичні показники текстури спільного входження, змішані з інформацією про амплітуду, можуть використовуватися для визначення порогового значення. Однак у реальних додатках і через дефіцит наземних правдивих даних для швидкого картографування повеней краще використовувати неконтрольовані методи виявлення змін. Алгоритми неконтрольованого машинного навчання є більш надійними завдяки вищій швидкості, меншим вимогам до навчальних даних і часу виконання обчислень, що забезпечує кращу ефективність обчислень (Cao H., 2019).

Неконтрольовані методи виявлення змін використовуються для автоматичного визначення змін у зображеннях без попереднього навчання на мічених даних. Це робить їх особливо цінними в умовах, коли швидкість та ефективність обробки є критичними, а доступ до наземних даних обмежений. Такі методи дозволяють оперативно реагувати на ситуації надзвичайного стану, такі як повені, і забезпечують точне картографування затоплених територій.

В роботі (Wu X., 2023) підкреслюється критична важливість виявлення повеней в режимі, близькому до реального часу, з акцентом на використанні методів глибокого навчання для збору даних з 16 паводків у басейні річки Янцзи. Це дослідження представляє ефективні методи створення наборів даних, пристосованих для навчання, тестування та застосування. Зокрема, дослідження показує, що згорткові нейронні мережі (CNN) перевершують традиційні методи виявлення повеней, підкреслюючи їх стійкість в умовах, близьких до реального часу.

У роботі (Abdirahman H., 2021) представлено нову та надійну систему виявлення повеней у реальному часі на основі методів машинного навчання та глибокого навчання, включаючи Random Forest, Naive Bayes, J48 та згорткові нейронні мережі. Ця система спрямована на виявлення підвищення рівня води та прогнозування потенційних повеней з гуманітарними наслідками до того, як вони відбудуться. Це дослідження робить внесок у галузі штучного інтелекту, інтелектуального аналізу даних та глибинного навчання, впроваджуючи інноваційний підхід до запобігання повеней за допомогою Arduino з GSM-модемами.

Робота (A. Emily J., 2022) зосереджена на точній ідентифікації затоплених територій за допомогою використання повністю згорткової мережі (FCN) на основі подвійних патчів, яка використовує злиття ознак на основі глибокого навчання. FCN незалежно навчаються на радарах із синтезованою апертурою (SAR) і мультиспектральних (MS) зображеннях, що дозволяє їм фіксувати відмінні риси, які згодом комбінуються для покращення можливостей виявлення повеней.

У роботі (Kashif A., 2019) представлено виявлення повеней на основі часових послідовностей супутникових знімків. Вони використовують комбінацію підходів класичного комп'ютерного зору і машинного навчання. Їхні результати підкреслюють ефективність їхніх методів у вирішенні проблем, пов'язаних із завданням MediaEval, і роблять внесок у ширшу сферу аналізу катастрофічних подій і використання супутникових даних.

Існуючі методи виявлення та картографування повеней мають ряд обмежень. Контрольовані методи, такі як K-NN, RF та SVM, залежать від наявності великої кількості мічених даних для навчання, що може бути проблематичним у реальних умовах. Методи глибокого навчання,

такі як CNN та FCN, хоч і демонструють високу точність, потребують значних обчислювальних ресурсів і можуть бути повільними при обробці великих обсягів даних у режимі реального часу. Неконтрольовані методи виявлення змін, хоча і є більш надійними та швидкими, можуть бути менш точними через відсутність попереднього навчання на мічених даних. Це може призводити до високої частоти хибнопозитивних або хибнонегативних виявлень. Крім того, велика різноманітність міських середовищ, різні типи затоплень та неоднорідність даних ускладнюють розробку універсальних моделей для виявлення повеней. Обмежена доступність даних з високою просторовою та часовою роздільною здатністю також є важливою перешкодою для точного і своєчасного моніторингу повеней.

Мета дослідження: розробка геоінформаційної технології для виявлення та моделювання затоплених зон у Дубаї на основі згорткових нейронних мереж (CNN). Це дозволить підвищити точність класифікації затоплених територій і покращити управління ризиками та відновлювальними роботами.

Виклад основного матеріалу. Загальна структура запропонованої геоінформаційної технології представлена на рисунку 1 і включає наступні етапи: завантаження супутникових знімків, векторизація первинних даних, інтеграція кліматичних та гідрографічних даних, розробка бази геоданих, створення карти досліджуваної території, застосування методів машинного навчання, моделювання затоплення та отримання карти зон небезпеки затоплення.

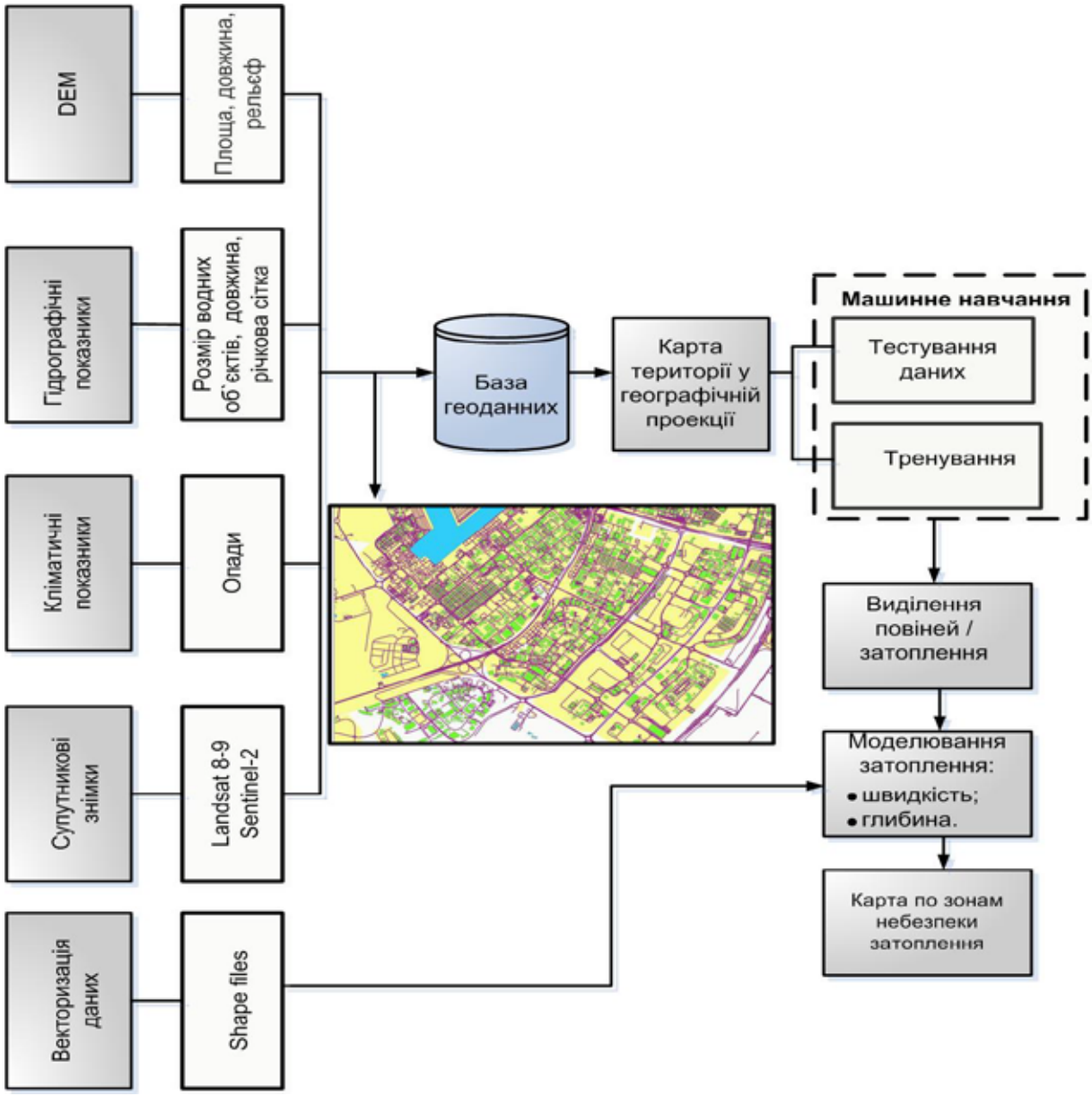


Рис. 1. Схема запропонованої геоінформаційної технології

З платформи Copernicus Open Access Hub завантажуються оптичні знімки з супутників Sentinel-2 або Landsat-8,9. Після завантаження зображень виконується ряд попередніх обробок, включаючи радіометричну та атмосферну корекції. Ці корекції спрямовані на усунення різноманітних спотворень, таких як вплив атмосфери, сонячне освітлення та інші фактори, що можуть виникати під час зйомки та передачі зображень з космосу. Радіометрична корекція забезпечує відповідність даних реальним фізичним величинам, що досягається шляхом перетворення цифрових значень (DN) пікселів у значення відбивної здатності. Це включає корекцію впливу сенсора та калібрування інструменту:

$$\text{Відбивна здатність} = \frac{L_{TOA}}{E_{sun} \cdot \cos(\theta) \cdot d^2} \qquad (1)$$

де L_{TOA} – радіометрична яскравість на вершині атмосфери; E_{sun} – спектральна сонячна константа; θ – зенітний кут сонця; d – відстань між Землею і Сонцем в астрономічних одиницях.

Атмосферна корекція усуває ефекти розсіювання та поглинання світла в атмосфері, що значно покращує якість зображень для подальшого аналізу.

В роботі запропоновано використати метод Dark Object Subtraction (DOS), що базується на припущенні, що найтемніші об'єкти в зображенні (зазвичай вода або тінь) мають нульову відбивну здатність, і будь-яке значення, відмінне від нуля, є результатом атмосферного розсіювання:

$$\text{Відкориговане значення} = DN - DN_{\min}, \qquad (2)$$

де $DN_{\min}$ – мінімальне значення DN у зображенні.

Після завантаження та попередньої обробки супутникових знімків здійснюється їх перетворення у векторний формат для полегшення подальшої обробки. Векторизація – це процес перетворення растрових зображень (пікселів) у векторні дані (геометричні форми, такі як точки, лінії та полігони). Цей етап дозволяє більш ефективно працювати з даними, забезпечуючи кращу інтеграцію з іншими геопросторовими даними та зменшуючи обсяг даних. Векторизація супутникових знімків включає кілька ключових етапів. Спочатку зображення сегментуються для визначення різних об'єктів на основі їхніх спектральних характеристик. Після сегментації виділяються контури об'єктів, які згодом будуть перетворені у векторні форми. Контури об'єктів перетворюються у точки, лінії та полігони, що формують векторні шари. Для цього використовуються алгоритми, такі як алгоритм Дугласа-Пекера для спрощення ліній та полігонів.

Наступним етапом є завантаження кліматичних показників, включають дані про опади, температуру, вологість, швидкість вітру та інші кліматичні фактори, які можуть впливати на ймовірність та масштаб повеней. Ці дані зазвичай отримуються з метеорологічних станцій або кліматичних моделей. Для дослідження в регіоні Дубай обрані дані з чотирьох метеорологічних станцій, які значною мірою сприяють аналізу затоплень у цьому регіоні. Вони включають станції Дубай, Шарджа, Аль-Айн і Джебель-Алі, оскільки дані про кількість опадів з цих станцій є критичними для моделювання повеней у регіоні.

Гідрографічні показники включають інформацію про річки, озера, водосховища та інші водні об'єкти. Вони містять дані про рівень води, швидкість течії, об'єм води та інші характеристики, що впливають на динаміку затоплень. Цифрова модель рельєфу (DEM) є ключовим компонентом для аналізу та моделювання повеней. DEM представляє собою тривимірну цифрову модель поверхні Землі, яка включає висоти на певній площі. Ця модель використовується для розрахунків потоків води та визначення зон можливого затоплення. DEM завантажується з відповідних джерел, таких як USGS або платформи, що надають дані LIDAR. Обробка DEM включає згладжування даних, виправлення помилок та заповнення прогалин для забезпечення точного представлення рельєфу. Аналіз рельєфу використовується для розрахунку нахилів, напрямків стоку та інших характеристик, важливих для моделювання водного потоку та зон затоплення. Усі значення висоти в цій моделі зберігаються в метрах відносно геоїда WGS84 EGM96. Цей геоїд базується на еліпсоїді WGS84, а коефіцієнти обчислюються на основі глобальної бази даних 30-хвилинних середніх вільних гравітаційних аномалій, а також на даних, отриманих з супутників і безпосередньої альтиметрії (вимірювання висоти).

Після виконання векторизації супутникових знімків та інтеграції кліматичних, гідрографічних даних і DEM, система готова до подальшого аналізу та моделювання. Ці етапи потрібні для точного визначення зон затоплення та розробки карт небезпеки, що є важливими для ефективного управління ризиками повеней.

Створення бази геоданих. Спочатку підготовлюються векторні шейп-файли та растрові

зображення, які є основою для подальшого аналізу. Визначаються таблиці та поля для зберігання інформації, встановлюється структура бази геоданих (БГД), яка забезпечить цілісність та ефективність зберігання даних. База геоданих створюється в середовищі QGIS, де імпортуються вихідні дані з відповідних джерел. Цей етап включає імпорт векторних та растрових даних у визначену структуру бази. Для забезпечення цілісності даних встановлюються зв'язки між таблицями, що дозволяє зберігати логічні відносини між різними наборами даних. Використовуючи інструменти QGIS, проводиться редагування та внесення змін у дані. На цьому етапі застосовується алгоритм дерева рішень, частина процесу машинного навчання, для аналізу зв'язків між описовими ознаками та цільовими мітками для безперервних даних. Алгоритм дерева рішень використовує методи визначення змінних та значень поділу за допомогою домішок, що генеруються в кожному вузлі. Після створення БГД дані оптимізуються та індексуються для підвищення продуктивності та швидкості доступу до них.

Класифікація під контролем є поширеним методом для вилучення інформації із зображень, що включає два основних етапи: навчання та класифікацію. На етапі навчання для кожного класу вибирається набір репрезентативних зразків, який служить основою для навчання моделі класифікації. На етапі класифікації використовуються навчені моделі для оцінки ймовірності того, що кожен піксель зображення належить до одного з класів. Пікселі класифікуються відповідно до класу, що має найбільшу ймовірність, визначену класифікатором. Класифікація застосовується до зображень з використанням трьох класів: водні об'єкти, затоплені території та інші. Клас водних об'єктів представляє постійні водні площі в обох наборах даних. Для всіх трьох методів класифікації використовується однакова навчальна вибірка. У численних дослідженнях були застосовані різні моделі згорткових нейронних мереж (CNN), такі як ResNet50, VGG16, InceptionV3 тощо (Wu X., Abdirahman H., A. Emily J., Kashif A.), які показали значні результати у різних сферах. У цій роботі використовується підхід трансферного навчання, де попередньо навчені моделі застосовуються для виявлення повеней. Модифіковано архітектуру моделей, додавши кілька конволюційних шарів для виділення суттєвих ознак місцевості та максимальне підсумовування для зменшення розмірності просторових даних. У другій гілці мережі застосовується щільний шар для інтеграції ознак та згладжувальний шар для підвищення стабільності навчання. Останній шар використовує максимальне підсумовування для збереження ключових ознак з подальшим переходом до щільного шару. Процес навчання моделі виконується з використанням оптимізатора Adam та бінарної втрати перехресної ентропії для мінімізації похибок. Ефективність оцінюється шляхом вимірювання показників точності на валідаційних даних, як відображено на Рисунку 2. Процес навчання моделі включає використання скомпільованого набору даних з оптимізатором Adam та бінарною втратою перехресної ентропії як цільовою функцією. Навчання проводиться шляхом ітерацій через кілька епох, де відбувається безперервне оновлення вагових коефіцієнтів моделі з метою мінімізації втрат і підвищення точності. Ефективність моделі оцінюється на валідаційному наборі даних шляхом вимірювання таких показників, як точність і втрати.

Вхідний шар приймає зображення розміру:

$$H \times W \times C, \tag{3}$$

де H – висота, W – ширина, C – кількість каналів.

Конволюційний шар застосовує фільтри для виділення ознак:

$$Y_{i,j,k} = \sum_{m,n} X_{i+m,j+n} \cdot W_{m,n,k} + b_k, \tag{4}$$

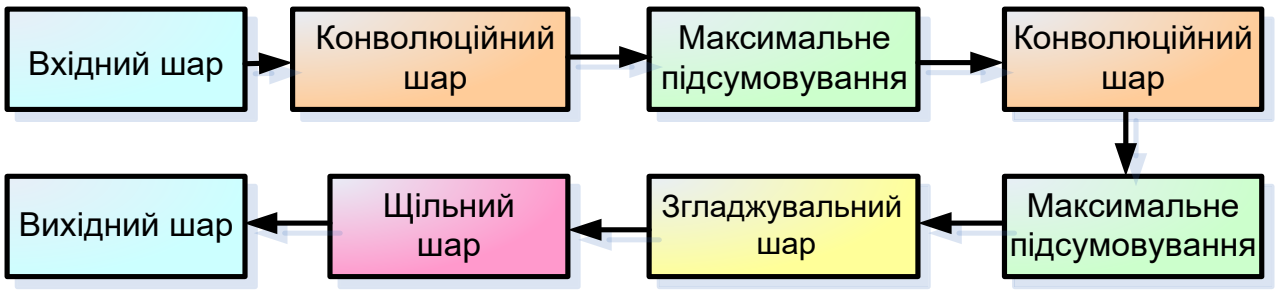


Рис. 2. CNN архітектура

де X – вхідні дані, W – ваги фільтрів, b – зсув, Y – вихідні ознаки.

Максимальне підсумовування зменшує розмірність карт ознак

$$Y_{i,j,k} = \max_{m,n} X_{s(i,m),s(j,n),k}, \quad (5)$$

де s – розмір вікна підсумовування.

Згладжувальний шар перетворює багатовимірні дані в одновимірний вектор.

Щільний шар використовує функцію активації для навчання:

$$Y_i = f(\sum_j W_{ij} \cdot X_j + b_i), \quad (6)$$

де f – функція активації, наприклад, ReLU або сигмоїдна функція.

Моделювання затоплення є ключовим етапом в оцінці та прогнозуванні потенційних зон затоплення. Воно включає використання як гідрологічних, так і гідравлічних моделей для детального аналізу обсягів водного стоку та умов течії потоку, що допомагає у визначенні зон потенційного затоплення. На першому етапі використовуються гідрологічні моделі для оцінки обсягів річкового стоку, що генерується під впливом екстремальних погодних подій або комбінацій різних серій подій у часі. Гідрологічна модель обчислює екстремальні значення, враховуючи дані про опади, топографію, тип ґрунту та умови землекористування, такі як рослинний покрив. Це дозволяє створити криву ґрунтового номера і гідрограф, які відображають поведінку водного стоку на території. Наступний етап передбачає використання гідравлічної моделі для розрахунку умов течії потоку, зокрема швидкості потоку, глибини води та розміру потенційного затоплення. Моделювання затоплення починається з імплементації цифрової моделі рельєфу (ЦМР) у форматі файлу сітки ASCII. Цей файл імплементується у геоінформаційну систему QGIS для подальшої обробки. Обчислення сітки виконуються на основі ASCII файлу, при цьому розмір сітки визначається на рівні 30 м. Розмір сітки обирається за формулою:

$$0.3 м/с \le \frac{Q_{max}}{A_{elem}} \le 3 \text{ м/с} \quad (7)$$

де Q_{max} – максимальний стік, розрахований для елемента сітки, A_{elem} – площа елемента сітки. Інтерполяція сітки забезпечує визначення репрезентативної висоти кожного елемента.

Для моделювання небезпеки затоплення використовуються кумулятивні значення періодів повернення (10, 30, 50, 100 і 300 років) кліматичного та гідрологічного моделювання. Результати моделювання, такі як максимальна глибина потоку та максимальна швидкість потоку, обробляються у програмному забезпеченні QGIS.

Для оцінки небезпеки затоплення карти класифікуються на зони високого, середнього та низького ризику, використовуючи критерії Гарсії та Лопеса. Зони визначаються за глибиною води h:

$$v \cdot h, \quad (8)$$

де v – швидкість потоку.

Наприклад, зона h>1,5 м і v×h>1,5 м/с класифікується як зона високої небезпеки. Зони з 0,5<h<1,5 м і v×h<1,5м/с класифікуються як зона середньої небезпеки. Зони з 0,1<h<0,5 м і v×h<0,5 м/с вважаються зонами низького рівня небезпеки. Результати класифікації відображаються на картах із кольоровою індикацією рівня небезпеки: червоний для високого рівня, жовтий для середнього та зелений для низького рівня.

Експерименти. Об'єднані Арабські Емірати (ОАЕ) є державою, розташованою на Аравійському півострові, відомою своїм сухим пустельним кліматом. У квітні 2024 року ця країна зазнала впливу потужної системи повільних штормів, що призвела до значних опадів, які перевищили річну норму за кілька днів. Це викликало раптові повені в східних регіонах країни, що спричинило затоплення доріг та порушення транспортної інфраструктури. Для аналізу впливу повені на регіон було використано супутникові знімки Landsat-8, отримані до (рис. 3 а) та після (рис. 3 б) повені. Знімки до повені, зроблені на початку квітня 2024 року, використовувались для визначення базового стану території. Після завершення штормів, знімки, отримані 19 квітня 2024 року, дозволили оцінити масштаби затоплення та його вплив на інфраструктуру.

Для подальшого аналізу впливу повені на регіон було застосовано модель згорткових нейронних мереж (CNN). Спочатку для отримання бінарних масок (рис. 4) було проведено класифікацію пікселів, де водні об'єкти були позначені як 1, а суша – як 0. Це дозволило створити бінарні маски, які наочно відображають присутність водних об'єктів на супутникових знімках.

Після отримання бінарних масок було створено знімки з чіткою класифікацією територій як суші або води. На цих знімках водні об'єкти були виділені червоним кольором, що дозволяє візуально ідентифікувати контури затоплених зон. Це дає змогу легко порівняти зміни між зображеннями до і після повені, особливо в областях,

а)

б)

Рис. 3. Супутникові знімки космічного апарату Landsat 8-9 синтезовані у R-G-B канали: а) 18.03.2024 року; б) 19.04.2024 року

що зазнали затоплення. Зміни між зображеннями до і після повені стали очевидними, особливо в областях, що зазнали затоплення. Це наочно демонструє масштаби і розподіл води після повені, що є важливим для планування відновлювальних заходів та управління територіями. На зображеннях після обробки машинного навчання (рис. 5) чітко видно зміну класів у затоплених районах, що підтверджує ефективність застосованої методики класифікації.

Для оцінки продуктивності моделі класифікації у даному дослідженні були використані комбінація категоріальної перехресної ентропії та індексу Dice як функція втрат, а також метод Адама як оптимізатор. Метрика точності, яка була застосована для оцінки результатів моделювання, є F1-score, що дозволяє комплексно оцінити ефективність моделі в умовах дисбалансу класів. Точність визначається як відношення кількості правильно класифікованих позитивних зразків до загальної кількості класифікованих як позитивні, а повнота – як відношення кількості правильно класифікованих позитивних зразків до загальної кількості позитивних зразків у даних.

Точність:

$$\text{Точність} = \frac{TP}{TP + FP} \tag{8}$$

де *TP* – кількість правильних позитивних прогнозів, *FP* – кількість хибних позитивних прогнозів.

Повнота:

$$\text{Повнота} = \frac{TN}{TN + FP} \tag{9}$$

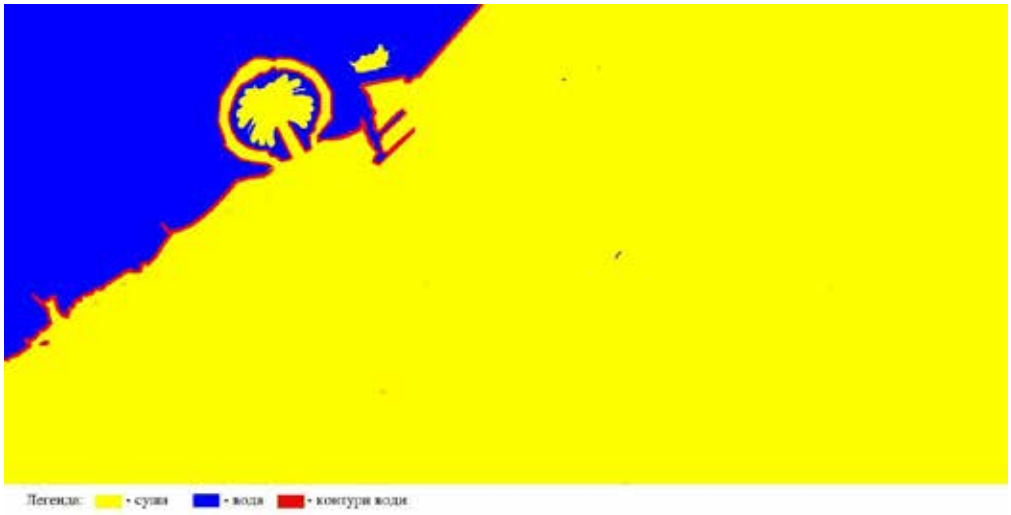
де *FN* – кількість хибних негативних прогнозів.

F1 Score:

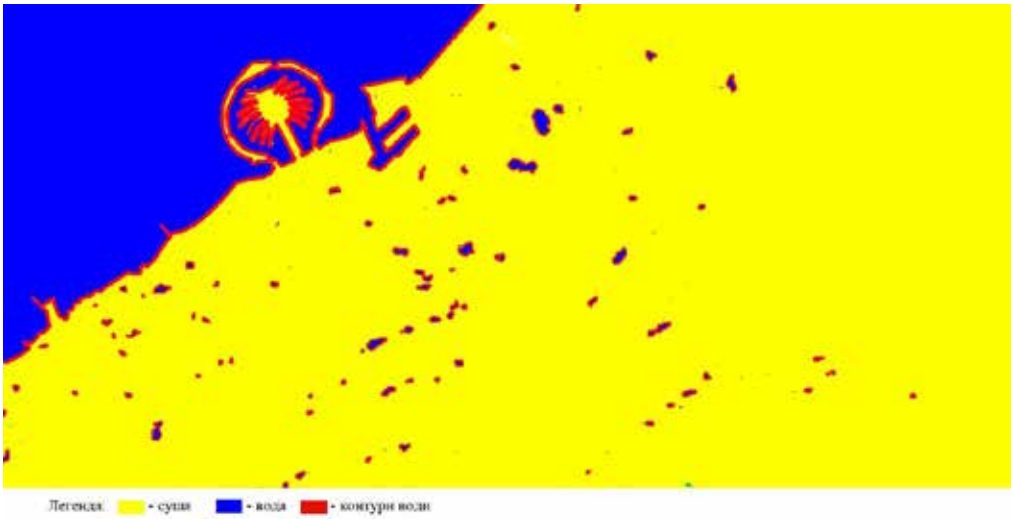
$$\text{F1}. \frac{2 \cdot \textit{Точність} \cdot \textit{Повнота}}{\textit{Точність} + \textit{Повнота}} \tag{10}$$

Результати. У таблиці 1 представлені результати роботи моделі, отримані на навчальній та валідаційній вибірках. Дані показують, що

Рис. 4. Бінарна маска супутникового знімка 19.04.2024 року



а)



б)

Рис. 5. Нейромережева класифікація супутникового знімку Landsat 8-9: а) до повені 18.03.2024 року; б) після повені 19.04.2024 року

результати тестування на валідаційних даних та на тестових даних є подібними, що свідчить про стабільність моделі.

Таблиця 1

Результати роботи моделі на навчальній та валідаційній вибірках

Вибірка	Точність	Повнота	Повнота
Навчальна	0.88	0.85	0.86
Валідаційна	0.87	0.84	0.85

Таблиця 2 демонструє оцінку F1 для кожного класу, розраховану на основі прогнозів моделі, отриманих з тестових даних. Результати вказують на те, що модель демонструє високу точність у класифікації класу водних об'єктів (FL) порівняно з класом земної поверхні (FV). Клас FL, що представляє затоплені території, передбачений з вищою точністю, ніж клас FV.

Таблиця 2

Оцінка F1 для кожного класу на основі тестових даних

Клас	Точність	Повнота	F1
FL	0.91	0.89	0.90
FV	0.84	0.78	0.81

У процесі дослідження було побудовано кілька графіків, які наочно демонструють інтенсивність опадів у Дубаї під час шторму. На рис. 6 представлено результат у вигляді графіка з використанням сітки даних з координатами, що охоплюють Дубай та прилеглі території. Інтенсивність опадів була відображена методом контурної заливки, що дозволяє детально показати зміни в кількості опадів. Кольорова шкала показує величину опадів у кожній точці, де більш інтенсивні опади позначені темнішими відтінками синього. Графік дозволяє візуально оцінити розподіл опадів по всій території Дубаю та визначити області з найбільш інтенсивними опадами.

Рисунок 7а показує площі затоплення за період з 15 квітня по 19 квітня 2024 року. З графіка видно, що площа затоплення поступово збільшувалась з 10.2 гектарів 15 квітня до 42.61 гектарів 19 квітня. Це відображає наростаючий ефект шторму, що супроводжувався інтенсивними опадами, які викликали значні затоплення. Кругова діаграма на рис. 7 б допомагає зрозуміти масштаб затоплення в порівнянні з іншими територіями, вказуючи на суттєву частку затоплених земель.

На рис. 8 значна частина території Дубаї (райони: Hasyan 1, Hasyan 2, Jebel Ali, Palm та Jebel Ali) затоплена водою глибиною від 1.75 до 2 метрів, що позначено чорними крапками. Це вказує на серйозний вплив затоплення на ці райони. Інші кольори вказують на поступове зменшення рівня води в різних частинах території. Зображення також показує інші рівні затоплення, розподілені по різних ділянках: рожеві ділянки (1.5 – 1.75 метра) займають меншу площу порівняно з червоними; жовті, світло-зелені, зелені, бірюзові та блакитні ділянки позначають райони, в яких поступово зменшується рівень затоплення.

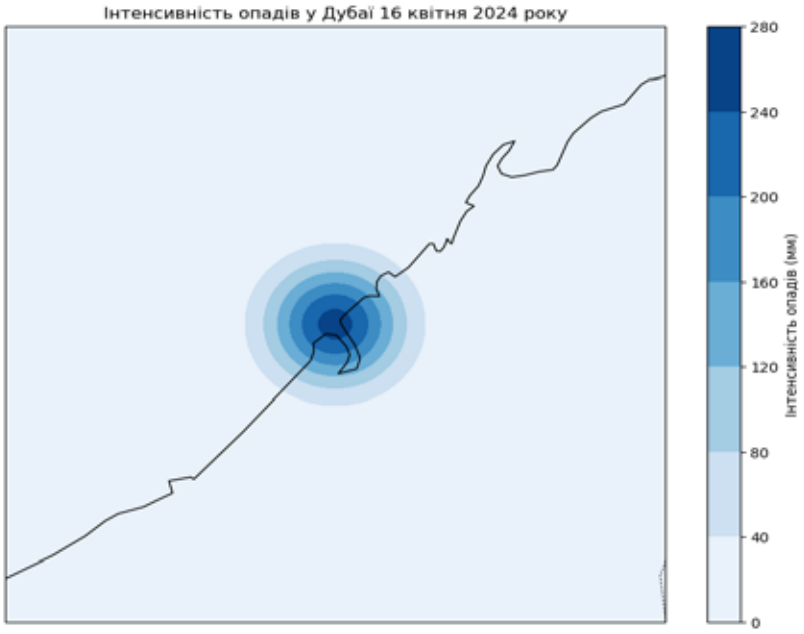


Рис. 6. Графік інтенсивності опадів в Дубаї

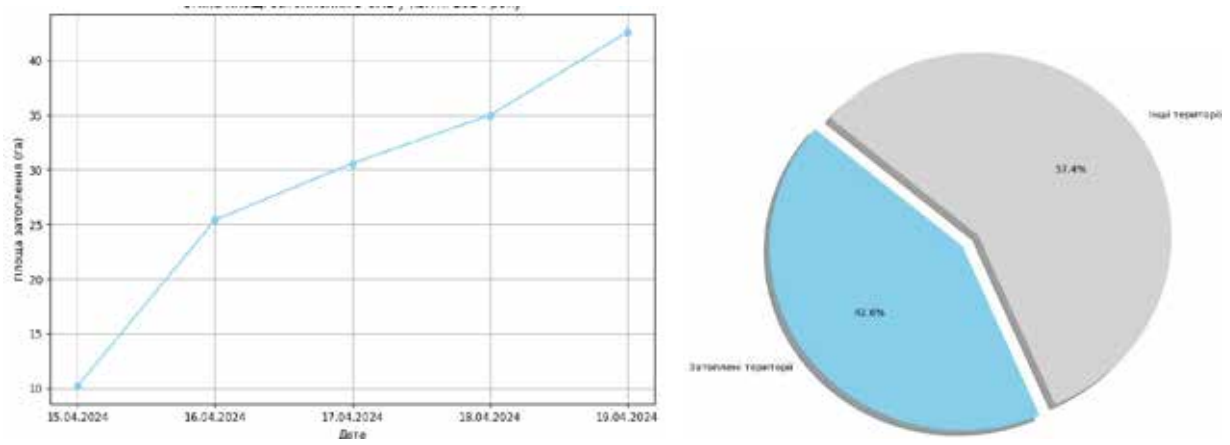


Рис. 7. Зміни площі затоплення в Дубаї: а) графік в період з 15.04.2024 року по 20.04.2024 року; б) діаграмма співвідношення площі затоплених та інших територій

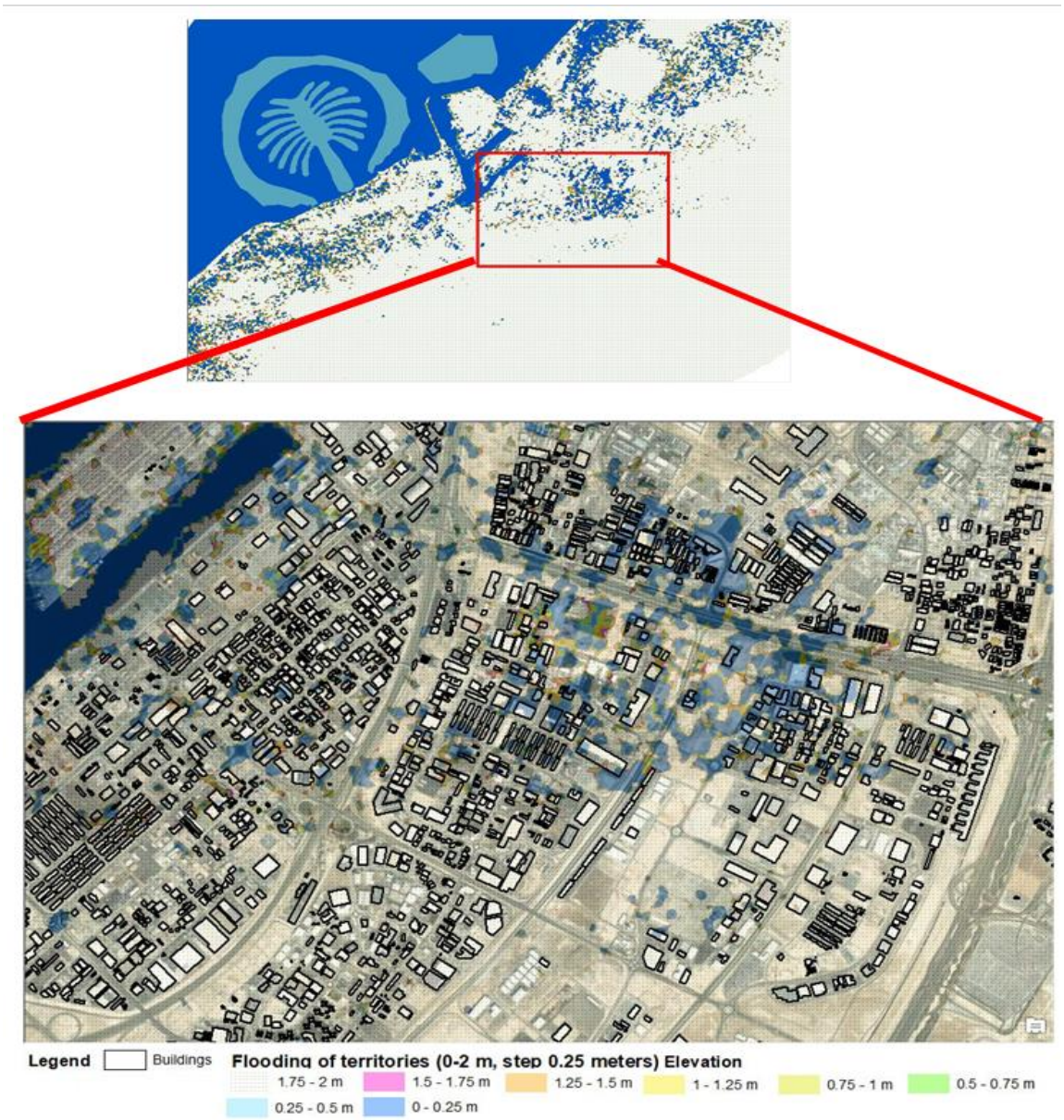


Рис. 8. Карта зон затоплення Дубаї

На рисунку 9 представлено вісім карт, що відображають зони затоплення району Jebel Ali Industrial 2. Карти демонструють глибину затоплення в діапазоні від 0 до 2 метрів з кроком 0,25 метра. Кожна карта відображає території, які будуть затоплені на певну глибину, що дозволяє здійснити детальний аналіз впливу поступового підвищення рівня води на цей район.

Загальний розподіл затоплення вказує на його нерівномірність, з більшою глибиною затоплення в північній і центральній частині району. Деякі території залишаються незатопленими або затопленими на незначну глибину, що вказує на різницю у висоті та особливостях рельєфу. Крок 0,25 метра, який використовується в даних картах, дозволяє детально відобразити динаміку затоплення та зрозуміти, як поступове підвищення рівня води впливатиме на територію. Початковий етап затоплення (0 – 0,25 метра) охоплює обмежені території, переважно в низько розташованих частинах району. Зі збільшенням рівня води до 0,5 метра затоплені території розширюються, переважно вздовж річок і каналів. При підвищенні рівня води до 0,75 метра спостерігається значне розширення зон затоплення, особливо в низько розташованих районах. Коли глибина затоплення досягає 1 метра, затоплення охоплює ще більше територій, включаючи деякі забудовані райони. Подальше підвищення рівня води до 1,25 метра призводить до затоплення ще більших частин промислових зон.

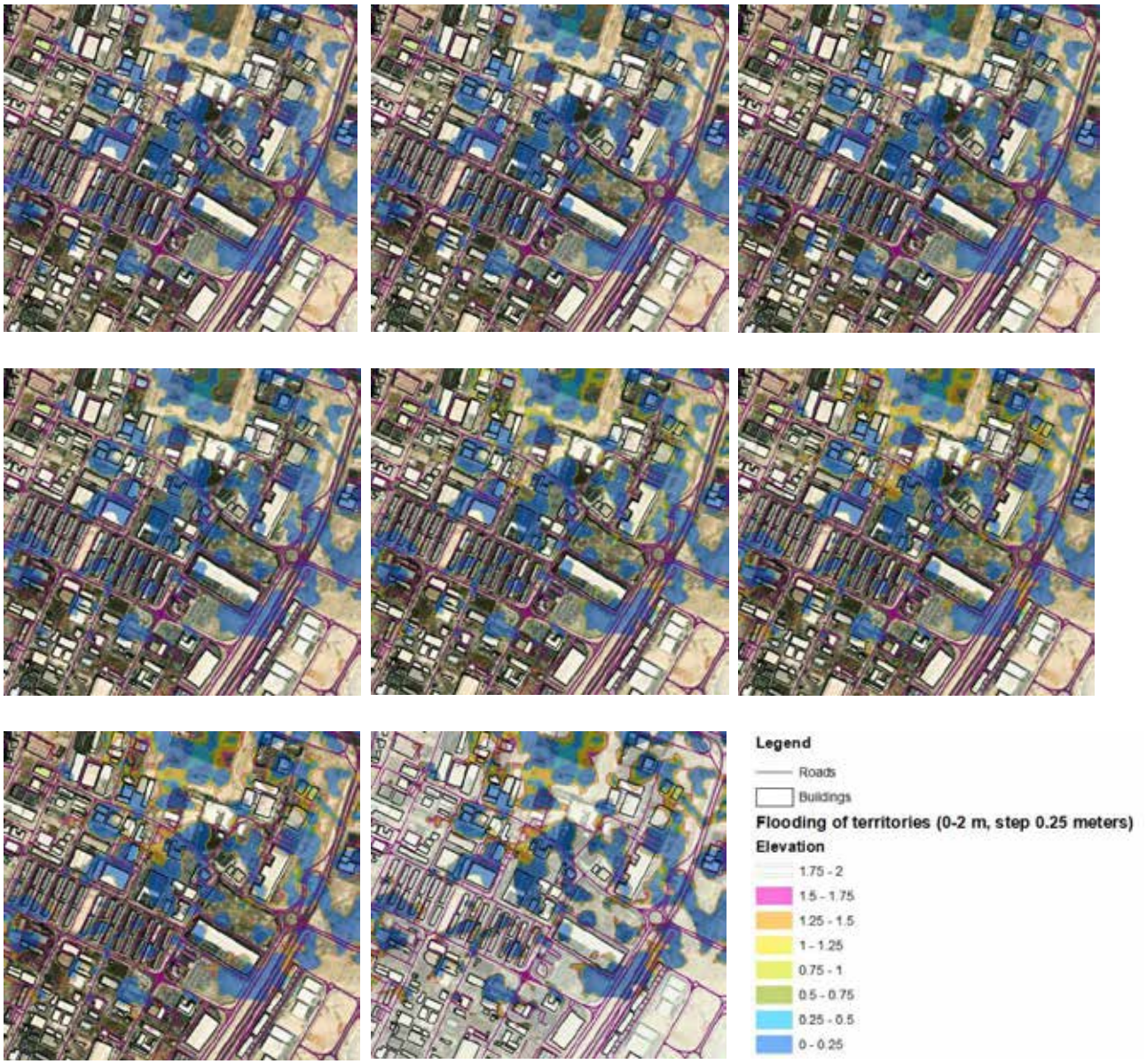


Рис. 9. Карта зон затоплення району Jebel Ali Industrial 2

При досягненні рівня води 1,5 метра затоплення поширюється на ще більші території. На глибині 1,75 метра спостерігається майже повне затоплення низькорозташованих частин району, а при максимальній глибині затоплення до 2 метрів охоплюються найнижчі території, з мінімальними залишками незатоплених зон. Виходячи з отриманих даних, райони з низькою абсолютною висотою є найбільш вразливими до затоплення, з максимальним затопленням до 2 метрів. Прогресивне затоплення з кроком 0,25 метра демонструє, як поступове підвищення рівня води впливає на територію, надаючи цінну інформацію для планування заходів щодо захисту та адаптації. Для районів з глибиною затоплення понад 1 метр слід розглянути додаткові захисні заходи, такі як будівництво дамб або підвищення рівня будівель. Отримані результати можуть бути використані для подальшого аналізу та розробки стратегій управління ризиками затоплення, а також для інформування громадськості та планування розвитку інфраструктури в районі Jebel Ali Industrial 2.

Висновки. Розроблено геоінформаційну технологію виявлення та моделювання затоплених зон Дубаю на основі згорткової нейронної мережі з використанням супутникових даних. Дослідження показало, що запропонований підхід ефективно виявляє та класифікує затоплені території, що підтверджується отриманими метриками точності та повноти.

Результати класифікації на навчальній та валідаційній вибірках продемонстрували високий рівень точності, повноти та F1-міри. На навчальній вибірці модель досягла точності 0.88, повноти 0.85 та F1-міри 0.86. На валідаційній вибірці точність склала 0.87, повнота 0.84, а F1-міра – 0.85. Це свідчить про стабільність та надійність моделі при застосуванні до нових даних. Крім того, аналіз результатів по окремих класах показав, що модель краще справляється з виявленням класу FL (затоплені території), ніж класу FV (інші території). Точність для класу FL становить 0.91, повнота – 0.89, а F1-міра – 0.90. Для класу FV точність складає 0.84, повнота – 0.78, а F1-міра – 0.81. Це вказує на високу ефективність моделі для виявлення затоплених зон. Загальний розподіл затоплення показав нерівномірність затоплених зон, з більшою глибиною затоплення у північній і центральній частинах району. Аналіз з використанням кроку 0,25 метра дозволив детально відобразити динаміку затоплення, показуючи, як поступове підвищення рівня води впливатиме на територію. Початкові етапи затоплення охоплюють обмежені території, переважно в районах з низькою абсолютною висотою. Отримані результати надають цінну інформацію для планування заходів щодо захисту та адаптації, особливо для районів з глибиною затоплення понад 1 метр, де можуть знадобитися додаткові захисні заходи, такі як будівництво дамб або підвищення рівня будівель. Запропонована методика виявлення та моделювання затоплених зон може бути використана для подальшого аналізу, розробки стратегій управління ризиками затоплення, а також для інформування громадськості та планування розвитку інфраструктури.

ЛІТЕРАТУРА:

1. Deolfa Jose Moises, Olivia Kungum. Strengthening Namibia's Flood Early Warning System through a Critical Gap Analysis. *Sustainability*. 2023.15. 524.

2. Hsiao S. C., Chiang W. S., Jang J. H., Wu H. L. Lu, W. S. Chen, W. B. Wu Y. T. Flood risk influenced by the compound effect of storm surge and rainfall under climate change for low-lying coastal areas. *Sci. Total Environ*. 2021. 764. 144439.

3. Shahabi H., Shirzadi A., Ghaderi K., Omidvar E., Al-Ansari N., Clague J. J. Geertsema M., Khosravi K., Amini A., Bahrami S. et al. Flood Detection and Susceptibility Mapping Using Sentinel-1 Remote Sensing Data and a Machine Learning Approach: Hybrid Intelligence of Bagging Ensemble Based on K-Nearest Neighbor Classifier. *Remote Sens*. 2020. 12. 266.

4. Tavus B., Kocaman S., Gokceoglu C. Flood damage assessment with Sentinel-1 and Sentinel-2 data after Sardoba dam break with GLCM features and Random Forest method. *Sci. Total Environ*. 2021. 816. 151585

5. Emily Jenifer A, Sudha N. DeepFlood: A deep learning based flood detection framework using feature-level fusion of multi-sensor remote sensing images. *Journal of Universal Computer Science*. 2022. Vol. 28. P. 329–343.

6. Li Y., Martinis S., Plank S., Ludwig R. An automatic change detection approach for rapid flood mapping in Sentinel-1 SAR data. *Int. J. Appl. Earth Obs. Geoinf*. 2018. 73. P.123–135.

7. Chawla I., Karthikeyan L., Mishra A.K. A review of remote sensing applications for water security: Quantity, quality, and extremes. *J. Hydrol*. 2020. 585. 124826.

8. Cao H., Zhang H., Wang C., Zhang B. Operational Flood Detection Using Sentinel-1 SAR Data over Large Areas. *Water*. 2019. 11. 786.

9. Wu X., Zhang Z., Xiong S., Zhang W., Tang J., Li Z., An B., Li R. A Near-Real-Time Flood Detection Method Based on Deep Learning and SAR Images. *Remote Sensing*. 2023. 15(8):2046.

10. Abdirahman H., Abdullahi A., Mohamed E., Siti H., Octavio R. R. A Real-Time Flood Detection System Based on Machine Learning Algorithms with Emphasis on Deep Learning. *International Journal of Engineering Trends and Technology*. 2021. 69. P.249–256.

11. A. Emily J., Sudha N. DeepFlood: A deep learning based flood detection framework using feature-level fusion of multi-sensor remote sensing images. *JUCS -Journal of Universal Computer Science*. 2022. 28(3). P. 329-343

12. Kashif A., Konstantin P., Mohib U., Michael R., Nicola Conci, Johannes L., Ala Al-F. Multi-Modal Machine Learning for Flood Detection in News, Social Media and Satellite Sequences. *Computer Vision and Pattern Recognition*, MediaEval 2019.

REFERENCES:

1. Deolfa Jose Moises and Olivia Kungum. (2023). Strengthening Namibia's Flood Early Warning System through a Critical Gap Analysis. *Sustainability*. 15. 524. [in English].

2. Hsiao, S. C., Chiang, W. S., Jang, J. H., Wu, H. L. Lu, W. S. Chen, W. B. Wu, Y. T. (2021). Flood risk influenced by the compound effect of storm surge and rainfall under climate change for low-lying coastal areas. *Sci. Total Environ*. 764. 144439. [in English].

3. Shahabi, H. Shirzadi, A., Ghaderi, K., Omidvar, E., Al-Ansari, N., Clague, J. J. Geertsema, M., Khosravi, K., Amini, A., Bahrami, S. et al. (2020). Flood Detection and Susceptibility Mapping Using Sentinel-1 Remote Sensing Data and a Machine Learning Approach: Hybrid Intelligence of Bagging Ensemble Based on K-Nearest Neighbor Classifier. *Remote Sens*. 12. 266. [in English].

4. Tavus, B., Kocaman, S., Gokceoglu, C. (2021). Flood damage assessment with Sentinel-1 and Sentinel-2 data after Sardoba dam break with GLCM features and Random Forest method. *Sci. Total Environ*. 816. 151585. [in English].

5. A. Emily Jenifer and Sudha Natarajan. (2022). DeepFlood: A deep learning based flood detection framework using feature-level fusion of multi-sensor remote sensing images. *Journal of Universal Computer Science*. Vol. 28. P. 329–343. [in English].

6. Li Y., Martinis S., Plank S., Ludwig R. (2018). An automatic change detection approach for rapid flood mapping in Sentinel-1 SAR data. *Int. J. Appl. Earth Obs. Geoinf.* 73. P.123–135. [in English].

7. Chawla, I., Karthikeyan, L., Mishra, A. K. (2020). A review of remote sensing applications for water security: Quantity, quality, and extremes. *J. Hydrol*. 585. 124826. [in English].

8. Cao, H., Zhang, H., Wang, C., Zhang, B. (2019). Operational Flood Detection Using Sentinel-1 SAR Data over Large Areas. *Water.* 11. 786. [in English].

9. Wu, X., Zhang, Z., Xiong, S., Zhang, W., Tang, J., Li, Z., An, B., Li, R. (2023). A Near-Real-Time Flood Detection Method Based on Deep Learning and SAR Images. *Remote Sensing*. 15(8):2046. [in English].

10. Abdirahman, H., Abdullahi, A., Mohamed, E., Siti, H., Octavio, R. R. (2021). A Real-Time Flood Detection System Based on Machine Learning Algorithms with Emphasis on Deep Learning. *International Journal of Engineering Trends and Technology.* 69. P.249–256. [in English].

11. A. Emily, J., Sudha, N. (2022). DeepFlood: A deep learning based flood detection framework using feature-level fusion of multi-sensor remote sensing images. *JUCS -Journal of Universal Computer Science*. 28(3). P. 329–343. [in English].

12. Kashif, A., Konstantin, P., Mohib, U., Michael, R., Nicola Conci, Johannes, L., Ala, Al-F. (2019). Multi-Modal Machine Learning for Flood Detection in News, Social Media and Satellite Sequences. *Computer Vision and Pattern Recognition, MediaEval* [in English].

УДК 004.75:004.49
DOI https://doi.org/10.32782/IT/2024-4-14

Антоніна КАШТАЛЬЯН
кандидат технічних наук, доцент кафедри фізики та електротехніки, докторанка, Хмельницький національний університет, 11, вул. Інститутська, м. Хмельницький, Україна, 29016
ORCID: *0000-0002-4925-9713*
Scopus Author ID: *57218242499*

Бібліографічний опис статті: Каштальян, А. (2024). Характерні властивості централізації в архітектурі мультикомп'ютерних систем антивірусних комбінованих приманок і пасток. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 114–124, doi: https://doi.org/10.32782/IT/2024-4-14

ХАРАКТЕРНІ ВЛАСТИВОСТІ ЦЕНТРАЛІЗАЦІЇ В АРХІТЕКТУРІ МУЛЬТИКОМП'ЮТЕРНИХ СИСТЕМ АНТИВІРУСНИХ КОМБІНОВАНИХ ПРИМАНОК І ПАСТОК

В роботі здійснено аналіз систем попередження, виявлення і протидії зловмисному програмному забезпеченню та комп'ютерним атакам, зокрема обманних систем та систем з приманками і пастками. При розробленні таких систем в їх архітектуру часто закладають механізми, які забезпечують їх адаптивність, тобто пристосування до оточуючого операційного середовища та внутрішніх і зовнішніх впливів. Під цими механізмами першочергово розглядають механізми, які забезпечують перебудову архітектури систем в процесі їх функціонування з метою відповіді на зловмисні загрози. Одним з основних елементів при перебудові архітектури систем без втручання користувача є центри систем. Деталізації механізмів перебудови саме центрів систем приділено недостатньо уваги, зокрема про це не заявляють і розробники комерційних систем.

В роботі визначено характерні властивості централізації в мультикомп'ютерних системах та здійснено деталізацію характерних властивостей, їх поєднання та подання. Така деталізація характерних властивостей є основою для встановлення зв'язків між ними, синтезу систем з такими центрами, а також оцінювання ефективності варіантів централізації. Згідно розробленого подання варіантів централізації в архітектурі систем було оцінено потенційну кількість варіантів централізації, до яких може перейти система при виборі наступного варіанту централізації.

Метою статті *забезпечення деталізації характерних властивостей централізації для визначення наступного варіанту центру в архітектурі мультикомп'ютерних систем попередження, виявлення та протидії зловмисному програмному забезпеченню і комп'ютерним атакам таким чином, щоб системи самостійно приховували свій центр, а також щоб забезпечувались ефективна взаємодія між їх вузлами і швидке прийняття рішень та підтримувалась цілісність системи в процесі експлуатації.*

Методологія *полягає у застосуванні наукових методів: системного аналізу синтезу, порівняння. В роботі представлений аналіз характерних властивостей централізації для визначення наступного варіанту центру в архітектурі мультикомп'ютерних систем попередження, виявлення та протидії зловмисному програмному забезпеченню і комп'ютерним атакам. Згідно аналізу подано формальне представлення характерних властивостей згідно яких будуть синтезовані варіанти централізації в архітектурі систем.*

Наукова новизна *полягає у розроблених моделях характерних властивостей централізації для використання їх в методі визначення наступного варіанту центру в архітектурі мультикомп'ютерних систем попередження, виявлення та протидії зловмисному програмному забезпеченню та комп'ютерним атакам, які на відмінну від відомих моделей згруповано за характерними властивостями і надають змогу формувати з них наступний варіант централізації без залучення користувача та уникнення повного або значного часткового перебору варіантів.*

Висновки*. Розроблено характерні властивості централізації в архітектурі мультикомп'ютерних систем приманок та пасток для виявлення та протидії ЗПЗ та КА. Характерні властивості згруповано в множини характерних властивостей. Згідно такого визначення отримано показники для використання їх при визначенні наступного варіанту централізації в архітектурі мультикомп'ютерних систем, яке вони повинні здійснювати самостійно без залучення адміністратора. Аналіз запропонованого рішення підтвердив перспективність напряму досліджень. В роботі здійснено постановку експерименту для розробленої системи згідно поданої централізації. Результати проведеного експерименту підтверджують перспективність досліджень в напрямі перебудови центру систем.*

Ключові слова: *централізація; захист інформації; обманні системи; мультикомп'ютерні системи; зловмисне програмне забезпечення; комп'ютерні атаки.*

Antonina KASHTALIAN
PhD, Associate Professor at the Department of Physics and Electrical Engineering, Doctoral Student, Khmelnytskyi National University, 11, Instytuts'ka Str., Khmelnytskyi, Ukraine, 29016, yantonina@ukr.net
ORCID: *0000-0002-4925-9713*
Scopus Author ID: *57218242499*

To cite this article: Kashtalian, A. (2024). Kharakterni vlastyvosti tsentralizatsii v arkhitekturi multykompiuternykh system antyvirusnykh kombinovanykh prymanok i pastok [Characteristic properties of centralization in the architecture of multi-computer systems of antiviral combined baits and traps]. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 4, 114–124, doi: https://doi.org/10.32782/IT/2024-4-14

CHARACTERISTIC PROPERTIES OF CENTRALIZATION IN THE ARCHITECTURE OF MULTI -COMPUTER SYSTEMS OF ANTIVIRAL COMBINED BAITS AND TRAPS

The work analyzes the prevention, detection and counteraction systems of malicious software and computer attacks, including fraudulent systems and systems with baits and traps. When developing such systems, mechanisms that provide their adaptability, that is, adaptation to the surrounding operating environment and internal and external influences, are often laid in their architecture. These mechanisms first consider mechanisms that provide restructuring of systems architecture in the process of their functioning in order to respond to malicious threats. One of the main elements in the restructuring of the architecture of systems without user intervention is the centers of systems. The detail of the mechanisms of restructuring of the centers of systems is not given enough attention, in particular, the developers of commercial systems do not declare.

The work identifies the characteristic properties of centralization in multicomputer systems and detail the characteristic properties, their combination and presentation. Such detail of characteristic properties is the basis for establishing relationships between them, the synthesis of systems with such centers, as well as evaluating the effectiveness of centralization options. According to the developed submission of centralization options in the architecture of systems, the potential number of centralization options, which can be transferred when choosing the next centralization option, was evaluated.

The purpose of the article *is to detail the characteristic properties of centralization to identify the next version of the center in the architecture of systems of prevention, detection and counteraction to malicious software and computer attacks so that the systems are concealed their center independently, as well as to ensure effective interaction between their nodes And the rapid decision -making and the integrity of the system was maintained during operation.*

The methodology *is to apply scientific methods: systematic analysis of synthesis, comparison. The work presents an analysis of the characteristic properties of centralization to determine the next version of the center in the architecture of the Milty -Computer Systems Preventing, Detection and Combating Mallen software and computer attacks. According to the analysis, a formal representation of the characteristic properties according to which will synthesize the variants of centralization in the architecture of systems.*

Scientific novelty *lies in the designed models of characteristic properties of centralization to use them in the method of determining the next variant of the center in the architecture of multi-computer systems of prevention, detection and counteraction Able to form the following centralization option without involving the user and avoiding a complete or significant partial interview of options.*

Conclusions. *The characteristic properties of centralization in the architecture of multi-computer baits and traps to detect and counteract the RFS and Ka. The characteristic properties are grouped into the set of characteristic properties. According to this definition, the indicators were obtained to use them in determining the following centralization in the multi -computer architecture, which they must be carried out independently without involving the administrator. The analysis of the proposed solution confirmed the prospect of the research. The experiment was staged for the developed system according to the submitted centralization. The results of the experiment confirm the prospect of research in the direction of restructuring of the center of systems.*

Key words: *centralization, protection of information, fraudulent systems, multicomputer systems, malicious software; Computer attacks.*

Вступ. При розробленні мультикомп'ютерних систем (A. Kashtalian et al., 2023) для виявлення та протидії зловмисному програмному забезпеченню (ЗПЗ) і комп'ютерним атакам (КА), зокрема і систем з приманками та пастками, а також обманних систем, потрібно забезпечити такі системи механізмами перебудови (B. Savenko et al., 2023, O. Savenko et al., 2020) їх архітектури таким чином, щоб ускладнити зловмисникам процес виявлення центру систем. При здійсненні синтезу мультикомп'ютерних систем антивірусних комбінованих приманок і пасток згідно різних характеристичних властивостей та принципів (A. Kashtalian et

al., 2023, A. Kashtalian et al., 2024) потрібно враховувати особливості централізації, тобто архітектуру та функційні можливості центрів таких систем. Ефективність функціонування муьтикомп'ютерних систем антивірусних комбінованих приманок і пасток залежить від організації та особливостей функціонування центру в архітектурі таких систем, оскільки це впливає на комунікацію між вузлами. Вузли системи розподілені, тоді час на прийняття рішень та надсилання відповідних повідомлень є важливими характеристиками. Крім того, приховування центру мультикомп'ютерних систем антивірусних комбінованих приманок і пасток для уникнення його виявлення зловмисниками, які діють ззовні або з середини периметру корпоративної мережі, є важливою характеристикою та спроможністю систем класу $\mathfrak{S}$ (A. Kashtalian et al., 2023, A. Kashtalian et al., 2024). Тому, потрібно на архітектурному рівні при синтезі систем класу $\mathfrak{S}$ (A. Kashtalian, 2023) забезпечити їх централізацію таким чином, щоб системи самостійно приховували свій центр, а також щоб забезпечувались ефективна взаємодія між їх вузлами і швидке прийняття рішень та підтримувалась цілісність системи в процесі експлуатації.

Аналіз останніх досліджень і публікацій. Обманні системи із приманками є сучасними засобами запобігання та виявлення зловмисних дій (S. Suratkar et al., 2022), які можуть виконувати глибокий аналіз даних атак (S. Lysenko et al., 2022). Існуючі системи приманок піддаються постійним спробам детектування зловмисниками та виявлення їх вразливості. Приманки та їх системи, які мають постійні характеристики, не можуть протистояти виявленню. Ще одним недоліком статичних схем розгортання приманок є неможливість адаптації мережі після розгортання. Подолати ці недоліки дозволяють мережі приманок, які мають динамічні властивості (S. Leyi et al., 2019). В роботі (W. Fan et al., 2015) запропоновану гнучку систему керування віртуальною мережею приманок. Також при використанні систем приманок має досягатися компроміс щодо функціональних можливостей та супутніх витрат, що неможливо забезпечити маючи незмінну конфігурацію приманок. В роботі (J.C. Acosta et al., 2021) запропонований підхід, що дозволяє приманкам динамічно вводити ресурси відповідно до виявлених дій зловмисника. Приманки з динамічними характеристиками забезпечують кращий обман зловмисників, в тому числі завдяки динамічному розташуванню. В роботі (Y.Li et al., 2019) запропонована розподілена мережа з динамічно розташованими приманками, яка виконує періодичну зміну сервісів, заплутуючи зловмисника та розпізнаючи його дії. Для забезпечення динамічних властивостей обманних систем використовують можливості програмно керованих мережа, зокрема в роботі (W. Han et al., 2016) запропоновану мережу приманок HoneyMix, що дозволяє уникати виявлення зловмисником, в роботі (M. Baykara et al., 2019) рознлянуто централізований підхід на основі приманок із програмно-визначеним перемиканням для зменшення хибно позитивних спрацювань.

Для оптимізації роботи обманних систем використовують комбіновані системи приманок різного рівня взаємодії Комбіновані системи приманок динамічно конфігуруються, збалансовують використання двох рівнів складності приманок, приманок з високим рівнем взаємодії для більш досвідчених зловмисників, та низького рівня взаємодії для менш досвідчених (A. Anwar et al., 2022) та забезпечують передачу трафіку між різними рівнями (S. Schindler et al., 2015). В роботі (M. Wegerer et al., 2016) пропонується архітектура на основі програмно керованої мережі, застосованої до комбінованої приманки для моделювання топології мережі та пере направлення трафіку атаки, використовується розширюваність та керованість контролера програмно керованої мережі для імітації великої та реалістичної мережі для залучення зловмисників та пере направлення атак на приманку з високим рівнем взаємодії.

Важливою тактикою для обманних систем є їх невизначеність для зловмисників, що збільшує витрати зловмисників на атаки і зменшує їх успішність (S. Achleitner et al., 2016). В обманних системах має забезпечуватися розгортання обманних об'єктів, яке дозволяє інтегрувати обманні вузли в робочі, що дозволяє виявляти нові атаки з мінімальними втратами для засобу захисту (M. Zaman et al., 2023). Невизначеність системи досягається динамічними змінами, в роботі (C.J. Chiang et al., 2016) пропонується обманна система, що динамічно змінює вигляд мережі хостів в реальному часі. В роботі (A.S. Ehab, 2016) запропоновано методи кібермутації, які дозволяють змінювати динамічно конфігурацію системи, що забезпечує відволікання зловмисників та збільшення їх витрат на проведення атак.

Таким чином, проведений аналіз досліджень підтверджує необхідність розроблення нових підходів для забезпечення систем механізмами перебудови їх центру в процесі функціонування системи без втручання адміністратора. Це дасть змогу створювати системи, які буде

складно зрозуміти зловмисникам в процесі їх функціонування в частині знаходження та будови їх центру.

Методологія дослідження. Серед визначальних властивостей для систем класу $\mathfrak{S}$ було виділено в [4] таку властивість як $\mathfrak{V}_2$, яка характеризує типи та кількість центрів в архітектурі системи. В загальному випадку ця властивість щодо центру системи може буде реалізована так: цілісний центр в одній компоненті; центр поділений на рівнозначні частини в різних компонентах; центр поділений ієрархічно і міститься в різних компонентах; центр є цілісним, ієрархічним та міститься в різних компонентах. З використанням такої стратегії щодо синтезу центру в архітектурі систем класу $\mathfrak{S}$ можуть бути різні варіанти та варіації, які будуть основою при функціонуванні центру системи в поточний момент часу. Наприклад, центр систем класу $\mathfrak{S}$ в поточний момент часу може бути в одному вузлі, а далі протягом певного інтервалу часу вже бути розподілений в декількох компонентах системи у різних вузлах корпоративної мережі. Також, зловмисник з середини мережі може відімкнути обладнання і корпоративна розділиться на два або більше сегментів. Таку модель зловмисної поведінки потрібно враховувати при організації функціонування центру системи.

Для реалізації в архітектурі систем класу $\mathfrak{S}$ таких властивостей центру розподілимо на підмножини всі можливі варіанти та їх варіації, а також здійснимо дослідження центру на предмет відповідності його очікуваним результатам з ефективності його приховування, прийняття рішень, організації комунікації між компонентами системи в різних вузлах корпоративної мережі та забезпечення підтримки цілісності системи, враховуючи гнучкі переходи центру системи до різних варіантів.

Виділимо спочатку потенційні варіанти організації центру системи, а потім їх варіації. У системах з більше ніж однією компонентою, які розміщені у різних вузлах корпоративної мережі, можна синтезувати в архітектурі системи центр чотирма різними варіантами: централізована архітектура; частково централізована архітектура; частково децентралізована архітектура; децентралізована архітектура. Позначимо ці варіанти в множині варіантів централізації так:

$$M_{\mathfrak{V}_2,\text{centr},v_1} = \left\{ m_{\mathfrak{V}_2,\text{centr},v_1,1}, m_{\mathfrak{V}_2,\text{centr},v_1,2}, m_{\mathfrak{V}_2,\text{centr},v_1,3}, m_{\mathfrak{V}_2,\text{centr},v_1,4} \right\}, \quad (1)$$

де елемент $m_{\mathfrak{V}_2,\text{centr},v_1,1}$ відображає централізовану архітектуру систем класу $\mathfrak{S}$; елемент $m_{\mathfrak{V}_2,\text{centr},v_1,2}$ – частково централізовану архітектуру системи; $m_{\mathfrak{V}_2,\text{centr},v_1,3}$ – частково децентралізовану архітектуру системи; $m_{\mathfrak{V}_2,\text{centr},v_1,4}$ – децентралізовану архітектуру системи.

Наступним параметром, який потребує врахування, буде варіант з розподіленням центру системи між її компонентами в різних вузлах корпоративної мережі. Задамо його множиною $M_{\mathfrak{V}_2,\text{centr},v_2}$ так:

$$M_{\mathfrak{V}_2,\text{centr},v_2} = \left\{ m_{\mathfrak{V}_2,\text{centr},v_2,1}, m_{\mathfrak{V}_2,\text{centr},v_2,2} \right\}, \quad (2)$$

де елемент $m_{\mathfrak{V}_2,\text{centr},v_2,1}$ відображає поділ центру системи між компонентами в різних вузлах для систем класу $\mathfrak{S}$; елемент $m_{\mathfrak{V}_2,\text{centr},v_2,2}$ відображає неподільність центру системи між компонентами в різних вузлах для систем класу $\mathfrak{S}$, тобто перебування його в одній компоненті.

Для варіанту, який задано множиною за формулою (2), потрібно врахувати можливість тимчасового розділення системи на дві або декілька частин. Причини розділення можуть бути різними. Але їх потрібно врахувати. Бо при наявності центру в одній із частин після розділення системи, решта частин системи повинні підтримувати своє функціонування. Інакше після повернення до повноцінного функціонування системи в цілісному поданні, довіра до частин, які функціонували без центру системи буде знижена і потребуватимуть повного дослідження її компоненти та вузли корпоративної мережі. Також, можуть бути і інші причини, але всі вони за відсутності центру системи в її частинах, які відокремились, знижують довіру до результатів її роботи. Тому, параметром, який потребує врахування, буде варіант з розподіленням центру системи між її компонентами в різних вузлах корпоративної мережі при розподіленні її на незв'язні частини. Навіть для централізованої архітектури систем класу $\mathfrak{S}$ при її розділенні на незв'язні частини вважатимемо, що центр системи буде в кожній з таких частин. Задамо такий варіант з наявністю центру системи в незв'язних частинах систем класу $\mathfrak{S}$ множиною $M_{\mathfrak{V}_2,\text{centr},v_3}$ так:

$$M_{\mathfrak{V}_2,\text{centr},v_3} = \left\{ m_{\mathfrak{V}_2,\text{centr},v_3,1}, m_{\mathfrak{V}_2,\text{centr},v_3,2} \right\}, \quad (3)$$

де елемент $m_{\mathfrak{V}_2,\text{centr},v_3,1}$ відображає наявність центру системи в незв'язних частинах систем класу $\mathfrak{S}$ після їх розділення; елемент $m_{\mathfrak{V}_2,\text{centr},v_3,2}$ відображає наявність центру системи лише в одній з незв'язних частин систем класу $\mathfrak{S}$ після їх розділення, тобто в решті частин системи центр відсутній.

У варіантах з розділенням центру системи між компонентами у різних вузлах корпоративної мережі потрібно враховувати задані зв'язки між компонентами центру системи. Вони

впливатимуть на швидкість прийняття рішень, бо будуть залежати від часу витраченого на обмін повідомленнями між собою. Ці зв'язки можуть бути встановлені між всіма компонентами, в яких буде міститись центр системи, а можуть бути задані лише частково, тобто не повністю і, відповідно, не формуватимуть відношення «всі до всіх». Між компонентами, в яких міститься центр системи, може бути організована ієрархічна підпорядкованість. Такі особливості в архітектурі центрів систем класу $\mathfrak{S}$ потрібно врахувати. Тому, задамо множиною $M_{\mathfrak{A}_2,centr,v_4}$ особливості зв'язків між компонентами при його розподілені між компонентами в різних вузлах корпоративної мережі так:

$$M_{\mathfrak{A}_2,centr,v_4} = \left\{ m_{\mathfrak{A}_2,centr,v_4,1}, m_{\mathfrak{A}_2,centr,v_4,2}, \ldots, m_{\mathfrak{A}_2,centr,v_4,n_{centr,v_4}} \right\}, \quad (4)$$

де елемент $m_{\mathfrak{A}_2,centr,v_4,1}$ відображає зв'язки між компонентами з центром системи архітектури систем класу $\mathfrak{S}$ згідно відношення «всі до всіх»; елемент $m_{\mathfrak{A}_2,centr,v_4,2}$ відображає зв'язки між компонентами з центром системи архітектури систем класу $\mathfrak{S}$ згідно відношення «один до всіх» архітектуру систем класу $\mathfrak{S}$; $m_{\mathfrak{A}_2,centr,v_4,3}$ відображає зв'язки між компонентами з центром системи архітектури систем класу $\mathfrak{S}$ згідно послідовної топології; $m_{\mathfrak{A}_2,centr,v_4,4}$ відображає зв'язки між компонентами з центром системи архітектури систем класу $\mathfrak{S}$ згідно топології «кільце»; $m_{\mathfrak{A}_2,centr,v_4,5}$ відображає зв'язки між компонентами з центром системи архітектури систем класу $\mathfrak{S}$ згідно топології «решітка»; $m_{\mathfrak{A}_2,centr,v_4,6}$ відображає зв'язки між компонентами з центром системи архітектури систем класу $\mathfrak{S}$ згідно топології «тор»; $m_{\mathfrak{A}_2,centr,v_4,7}$ відображає зв'язки між компонентами з центром системи архітектури систем класу $\mathfrak{S}$ згідно топології «куб»; n_{centr,v_4} – кількість варіантів зв'язків між компонентами з центром системи архітектури систем класу $\mathfrak{S}$.

Введемо, також, множину $M_{\mathfrak{A}_2,centr,v_5}$, елементи якої будуть характеризувати ієрархію компонентів, що містять центр системи, або будуть вказувати на те, що всі компоненти з центром системи є рівнозначними. Рівнів ієрархії компонентів з центром системи може бути декілька і не обов'язково два, як їх мінімальна кількість. Тобто, частина елементів може бути на вищому рівні ієрархії, а частина на нижчому рівні ієрархії. Більша кількість рівнів ієрархії може передбачати встановлення певної топології для таких компонентів, а також визначення для компонентів центру системи з певних рівнів ієрархії заданої функційної спроможності в ієрархії прийняття рішення. Ієрархія для компонентів центру системи може і не стосуватись обмеженої функційної спроможності, а може бути введена лише для оптимізації зв'язків між компонентами. Задамо *множину* $M_{\mathfrak{A}_2,centr,v_5}$ переліком таких елементів:

$$M_{\mathfrak{A}_2,centr,v_5} = \left\{ m_{\mathfrak{A}_2,centr,v_5,1}, m_{\mathfrak{A}_2,centr,v_5,2}, m_{\mathfrak{A}_2,centr,v_5,3}, \ldots, m_{\mathfrak{A}_2,centr,v_5,n_{centr,v_5}} \right\}, (5)$$

де елемент $m_{\mathfrak{A}_2,centr,v_5,1}$ відображає один рівень ієрархії, тобто фактично відображає відсутність ієрархічних рівнів; елемент $m_{\mathfrak{A}_2,centr,v_5,2}$ відображає два рівні ієрархії; елемент $m_{\mathfrak{A}_2,centr,v_5,3}$ відображає три рівні ієрархії між компонентами з центром системи архітектури систем класу $\mathfrak{S}$; $m_{\mathfrak{A}_2,centr,v_5,n_{centr,v_5}}$ відображає $n_{centr,v_5} - 1$ рівнів ієрархії між компонентами з центром системи архітектури систем класу $\mathfrak{S}$; n_{centr,v_5} – кількість елементів множини $M_{\mathfrak{A}_2,centr,v_5}$.

Напрям передачі повідомлень між компонентами систем згідно встановлених ***зв'язків*** може бути одностороннім, двостороннім або комбінованим. Тобто, передача повідомлень при організації комунікації між частинами центру системи може мати обмеження. Без обмежень буде двостороння комунікація. З обмеженнями буде одностороння комунікація, при якій повідомлення можуть переміщуватись за встановленим маршрутом. Наприклад, якщо є перехід від однієї компоненти до другої, то зворотного переходу може не бути, а для того, щоб реалізувати передачу в зворотньому напрямі потрібно надсилати повідомлення за маршрутом, який преведе у вказану компоненту проходячи через певні компоненти. Аналогічно, для змішаної комунікації частина компонент буде забезпечена двосторонньою комунікацією, а частина односторонньою. Задамо множиною $M_{\mathfrak{A}_2,centr,v_6}$ варіанти передачі повідомлень з різними типами комунікації так:

$$M_{\mathfrak{A}_2,centr,v_6} = \left\{ m_{\mathfrak{A}_2,centr,v_6,1}, m_{\mathfrak{A}_2,centr,v_6,2}, m_{\mathfrak{A}_2,centr,v_6,3} \right\}, \quad (6)$$

де елемент $m_{\mathfrak{A}_2,centr,v_6,1}$ відображає двосторонню комунікацію між компонентами; елемент $m_{\mathfrak{A}_2,centr,v_6,2}$ відображає односторонню комунікацію між компонентами; елемент $m_{\mathfrak{A}_2,centr,v_6,3}$ відображає змішану комунікацію між компонентами.

Дійсно, при різних варіантах в архітектурі систем з частковою централізацією переважатиме змішана комунікація між *компонентами*. У варіантах централізованої або децентралізованої архітектури, якщо не встановлено додаткових обмежень, переважатиме двостороння комунікація. Встановлення обмежень на здійснення комунікації між компонентами центру системи може бути необхідним, враховуючи специфіку завдань системи, потребу в заплутуванні зловмисників та приховуванні компонент центру системи.

Центр системи може бути в частині компонент, а не точно в усіх компонентах системи. Тобто, центр системи може бути розподіленим між певною частиною компонент, а в частині компонент його може не бути взагалі. Також, можливі варіанти, коли центр системи розподілено в усі компоненти. *Задамо множиною* $M_{\mathfrak{Y}_2,\text{centr},v_7}$ варіанти передачі повідомлень з різними типами комунікації так:

$$M_{\mathfrak{Y}_2,\text{centr},v_7} = \left\{ m_{\mathfrak{Y}_2,\text{centr},v_7.1}, m_{\mathfrak{Y}_2,\text{centr},v_7.2} \right\}, \quad (7)$$

де елемент $m_{\mathfrak{Y}_2,\text{centr},v_7,1}$ відображає наявність розподілених частин центру системи в усіх компонентах систем класу $\mathfrak{S}$; елемент $m_{\mathfrak{Y}_2,\text{centr},v_7.2}$ відображає наявність розподілених частин центру системи в певних компонентах систем класу $\mathfrak{S}$, тобто не в усіх компонентах.

Згідно множини, яку задано формулою (7), можуть бути варіації щодо перебування розподілених частин центру системи в компонентах в активному та неактивному станах. Введемо для визначення цих варіацій множину $M_{\mathfrak{Y}_2,\text{centr},v_8}$ *так:*

$$M_{\mathfrak{Y}_2,\text{centr},8} = \left\{ m_{\mathfrak{Y}_2,\text{centr},v_8,1}, m_{\mathfrak{Y}_2,\text{centr},v_8.2}, m_{\mathfrak{Y}_2,\text{centr},v_8.3} \right\}, \quad (8)$$

де елемент $m_{\mathfrak{Y}_2,\text{centr},v_8,1}$ відображає наявність розподілених частин центру системи в компонентах, які для нього задані, винятково в активному стані; елемент $m_{\mathfrak{Y}_2,\text{centr},v_8.2}$ відображає наявність розподілених частин центру системи в компонентах, які для нього задані, в двох станах (активний стан, пасивний стан); елемент $m_{\mathfrak{Y}_2,\text{centr},v_8,3}$ відображає наявність всіх розподілених частин центру системи в компонентах, які для нього задані, винятково в пасивному стані.

Активний стан розподіленої частини центру системи в певній компоненті означає його функціонування в поточний момент часу. Пасивний стан розподіленої частини центру системи в певній компоненті означає його перебування в поточний момент часу в очікуванні вказівки про перехід до активного стану.

Центр системи може бути розподілений так, що в усіх компонентах системи, в яких передбачено його наявність, буде однаковий функціонал. Тобто, всі компоненти, в яких буде наявний функціонал центру системи, будуть мати однакові функціонали. Якщо центр системи біде характеризуватись елементом $m_{\mathfrak{Y}_2,\text{centr},v_5,1}$ з формули (5), то всі компоненти з центром системи будуть фактично формувати децентралізовану архітектуру центру системи. Але центр системи може не бути в контексті ієрархії згідно формули (5), а може мати розділення таким чином, що декілька основних функцій будуть в окремих компонентах і матимуть вищий пріоритет порівняно з рештою функцій, які будуть перебувати в решті компонент, в яких задано функціонал з центром системи. В цьому варіанті центр системи складається з функцій, які розподілено між різними компонентами, і частина з них або одна є основними чи основною. Таким чином, центр розподілений за його частинами. Також, може бути варіант розподілу центру системи таким чином, що частина компонент міститиме основні функції центру системи, а решта компонент, в яких задано центр системи, другорядні. При цьому варіанті частина компонент з основними функціями центру системи буде рівнозначною між собою. Це важливо у випадку розпаду системи на декілька частин через розділення вузлів корпоративної мережі на певний інтервал часу. Тоді, буде збережено стійкість системи і вона зможе продовжити свою роботу. Введемо для визначення цих варіантів з розподіленням функцій центру системи множину $M_{\mathfrak{Y}_2,\text{centr},v_9}$ так:

$$M_{\mathfrak{Y}_2,\text{centr},9} = \left\{ m_{\mathfrak{Y}_2,\text{centr},v_9,1}, m_{\mathfrak{Y}_2,\text{centr},v_9.2}, m_{\mathfrak{Y}_2,\text{centr},v_9.3} \right\}, \quad (9)$$

де елемент $m_{\mathfrak{Y}_2,\text{centr},v_9,1}$ відображає випадок, коли всі компоненти, в яких наявний функціонал центру системи, будуть мати однакові функціонали; елемент $m_{\mathfrak{Y}_2,\text{centr},v_9.2}$ відображає випадок, коли центр системи має розділення таким чином, що декілька основних функцій будуть в окремих компонентах і мають вищий пріоритет порівняно з рештою функцій, які перебувають в решті компонент, в яких задано функціонал з центром системи; елемент $m_{\mathfrak{Y}_2,\text{centr},v_9.3}$ відображає випадок, коли варіант розподілу центру системи такий, що частина компонент містить основні функції центру системи, а решта компонент, в яких задано центр системи, другорядні, і, при цьому варіанті частина компонент з основними функціями центру системи буде рівнозначною між собою.

Функціонал центру системи може бути сформовано таким чином, що він буде складатись з незалежних підсистем. Для їх запуску в різних вузлах корпоративної мережі можна використовувати процедуру віддаленого запуску. Тоді, такий функціонал розподіленої частини центру системи може міститись в різних компонентах, в який задано функціонал центру системи, і буде формуватись з різних незалежних підсистем різних компонент. Головною незалежною підсистемою буде та підсистема, яка при встановлені функціоналу центру системи визначена і закріплена за відповідною компонентою. Таку організацію функціоналу компоненти центру системи

вважатимемо динамічною. Якщо ж функціонал компоненти центру системи складається з незалежних підсистем і всі вони встановлені в одну компоненту та запускаються в ній при виклику з головної підсистеми, то таку організацію функціоналу компоненти центру системи вважатимемо статичною. У випадку наявності одночасно двох варіантів з динамічною та статичною організацією функціоналу компоненти центру системи, тоді вважатимемо таку організацію змішаною. Для систем *класу* $\mathfrak{S}$ змішана організація може виникнути після повернення системи до цілісної, якщо перед цим було розділення. Під час її розділення в одній частині могла бути задіяна модель із динамічною організацією, а в решті частин – модель з статичною організацією. Після об'єднання фактично стає змішана організація. Задамо ці випадки множиною $M_{\mathfrak{V}_2,\text{centr},v_{10}}$ *так:*

$$M_{\mathfrak{V}_2,\text{centr},10} = \left\{ m_{\mathfrak{V}_2,\text{centr},v_{10},1}, m_{\mathfrak{V}_2,\text{centr},v_{10}.2}, m_{\mathfrak{V}_2,\text{centr},v_{10}.3} \right\}, \quad (10)$$

де елемент $m_{\mathfrak{V}_2,\text{centr},v_{10},1}$ відображає динамічну організацію функціоналу розподіленої частини центру системи*; елемент* $m_{\mathfrak{V}_2,\text{centr},v_{10}.2}$ відображає статичну організацію функціоналу розподіленої частини центру системи; елемент $m_{\mathfrak{V}_2,\text{centr},v_{10}.3}$ відображає змішану організацію функціоналу розподіленої частини центру системи.

Розглянемо введені множини та їх елементи згідно формул (1)-(10) в поєднанні для забезпечення централізації в системах класу $\mathfrak{S}$. Таке поєднання множин та їх елементів між собою забезпечить відображення особливостей центру системи за різновидами його архітектури, які необхідно використовувати при функціонуванні систем класу $\mathfrak{S}$ з метою приховування центру системи та заплутування зловмисників. Фактично, перебуваючи в одному із визначених центром системи можливих варіантів централізації, система перебуває в заданому стані, який характеризує її архітектуру в поточний момент часу. Протягом часу функціонування системи можуть самостійно змінювати варіанти централізації, переходячи при цьому до іншого стану.

Подамо в табл. 1 фрагмент варіантів централізації з особливостями, які відображені елементами в формулах (1)-(10).

Таким чином, здійснено подання характерних властивостей варіантів централізації в системах класу $\mathfrak{S}$ множинами, які задано за формулами (1)-(10), та встановлено кількість можливих варіантів. Отриманий результат щодо кількості варіантів відображає потенційну складність при здійсненні вибору наступного варіанту централізації. Розроблене подання через характерні властивості необхідне для формування правил визначення наступного варіанту централізації в архітектурі систем класу $\mathfrak{S}$, враховуючи потенційно велику кількість можливих варіантів.

Дослідження ефективності застосування теоретико-множинного подання характерних властивостей для подання централізації в архітектурі мультикомп'ютерних систем. Розглянемо ефективність розроблених характерних властивостей та їх поділ на множини в контексті досягнення поставленої мети щодо

Таблиця 1

Фрагмент варіантів централізації в архітектурі систем

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
	$M_{\mathfrak{V}_2,\text{centr},v_1}$				$M_{\mathfrak{V}_2,\text{centr},v_3}$		$M_{\mathfrak{V}_2,\text{centr},v_5}$		$M_{\mathfrak{V}_2,\text{centr},v_4}$				$M_{\mathfrak{V}_2,\text{centr},v_6}$					$M_{\mathfrak{V}_2,\text{centr},v_8}$			$M_{\mathfrak{V}_2,\text{centr},v_7}$		$M_{\mathfrak{V}_2,\text{centr},8}$			$M_{\mathfrak{V}_2,\text{centr},9}$			$M_{\mathfrak{V}_2,\text{centr},10}$		
	$m_{\mathfrak{V}_2,\text{centr},v_1,1}$	$m_{\mathfrak{V}_2,\text{centr},v_1,2}$	$m_{\mathfrak{V}_2,\text{centr},v_1,3}$	$m_{\mathfrak{V}_2,\text{centr},v_1,4}$	$m_{\mathfrak{V}_2,\text{centr},v_3,1}$	$m_{\mathfrak{V}_2,\text{centr},v_3,2}$	$m_{\mathfrak{V}_2,\text{centr},v_5,1}$	$m_{\mathfrak{V}_2,\text{centr},v_5,2}$	$m_{\mathfrak{V}_2,\text{centr},v_4,1}$	$m_{\mathfrak{V}_2,\text{centr},v_4,2}$	...	$m_{\mathfrak{V}_2,\text{centr},v_4,n,\text{centr},v_4}$	$m_{\mathfrak{V}_2,\text{centr},v_6,1}$	$m_{\mathfrak{V}_2,\text{centr},v_6,2}$	$m_{\mathfrak{V}_2,\text{centr},v_6,3}$	...	$m_{\mathfrak{V}_2,\text{centr},v_6,n,\text{centr},v_6}$	$m_{\mathfrak{V}_2,\text{centr},v_8,1}$	$m_{\mathfrak{V}_2,\text{centr},v_8,2}$	$m_{\mathfrak{V}_2,\text{centr},v_8,3}$	$m_{\mathfrak{V}_2,\text{centr},v_7,1}$	$m_{\mathfrak{V}_2,\text{centr},v_7,2}$	$m_{\mathfrak{V}_2,\text{centr},v_8,1}$	$m_{\mathfrak{V}_2,\text{centr},v_8,2}$	$m_{\mathfrak{V}_2,\text{centr},v_8,3}$	$m_{\mathfrak{V}_2,\text{centr},v_9,1}$	$m_{\mathfrak{V}_2,\text{centr},v_9,2}$	$m_{\mathfrak{V}_2,\text{centr},v_9,3}$	$m_{\mathfrak{V}_2,\text{centr},v_{10},1}$	$m_{\mathfrak{V}_2,\text{centr},v_{10},2}$	$m_{\mathfrak{V}_2,\text{centr},v_{10},3}$
1	1	0	0	0	0	1	1	0	1	0	0	0	1	0	0	0	0	1	0	0	1	0	1	0	0	1	0	0	1	0	0
2	1	0	0	0	0	1	1	0	1	0	0	0	0	1	0	0	0	1	0	0	1	0	1	0	0	1	0	0	1	0	0
3	1	0	0	0	0	1	1	0	1	0	0	0	0	0	1	0	0	1	0	0	1	0	1	0	0	1	0	0	1	0	0
4	1	0	0	0	0	1	1	0	1	0	0	0	0	0	0	1	0	1	0	0	1	0	1	0	0	1	0	0	1	0	0
5	1	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	1	1	0	0	1	0	1	0	0	1	0	0	1	0	0
6	1	0	0	0	0	1	1	0	1	0	0	0	1	0	0	0	0	0	1	0	1	0	1	0	0	1	0	0	1	0	0
7	1	0	0	0	0	1	1	0	1	0	0	0	0	1	0	0	0	0	1	0	1	0	1	0	0	1	0	0	1	0	0
8	1	0	0	0	0	1	1	0	1	0	0	0	0	0	1	0	0	0	1	0	1	0	1	0	0	1	0	0	1	0	0
9	1	0	0	0	0	1	1	0	1	0	0	0	0	0	0	1	0	0	1	0	1	0	1	0	0	1	0	0	1	0	0
10	1	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	0	1	0	0	1	0	0
11	1	0	0	0	0	1	1	0	1	0	0	0	1	0	0	0	0	0	0	1	1	0	1	0	0	1	0	0	1	0	0
12	1	0	0	0	0	1	1	0	1	0	0	0	0	1	0	0	0	0	0	1	1	0	1	0	0	1	0	0	1	0	0

синтезу систем, які можуть в процесі свого функціонування перебудовувати свою архітектуру без залучення адміністратора. Основною частиною мультикомп'ютерних систем, яка потребує дослідження щодо перебудови є центр системи. Задані характерні властивості централізації в архітектурі таких систем за формулами (1)-(10) дають змогу визначити не тільки потенційну кількість варіантів централізації, але й визначити стійкість та ймовірність виявлення зловмисником центру системи.

Потенційну кількість варіантів централізації визначимо згідно врахування характерних властивостей кожної множини, які визначені формулами (1)-(10). Оскільки, з кожної множини може бути взято тільки по одній характерній властивості, тоді кількість таких характерних властивостей буде дорівнювати 10 з 31 можливої характерної властивості. Тоді, 21 характерна властивість буде відсутня в системі. Взагалі допустимо, що система може вибирати довільних варіант централізації з різними характерними властивостями, як подано фрагментом потенційних варіантів в табл. 1. Наявність певної характерної властивості позначено в табл. 1 одиницею, а відсутність – нулем. Кількість виборів однієї характерної властивості з першої множини буде дорівнювати кількості перестановок без повторень з чотирьох елементів по одному елементу. Аналогічно визначаємо для решти дев'яти множин. Отримані значення для кожної характерної властивості з кожної з десяти множин за правилом добутку перемножуємо і отримуємо потенційно можливу кількість варіантів централізації, тобто

$$K_{centr} = \prod_{i=1}^{10} C_1^{N^{centr}_{M_{\mathfrak{A}_2,\,centr,i}}}, \qquad (11)$$

де $N^{centr}_{M_{\mathfrak{A}_2,\,centr,i}}$ – кількість елементів в множині характерних властивостей $M_{\mathfrak{A}_2,\,centr,i}$; $i = 1,2,\ldots,10$.

Подані за формулами (1)-(10) множини характерних властивостей дають змогу, також, встановити варіанти централізації, в які неможливо перейти з попереднього поточного варіанту централізації. Кількість таких варіантів є невеликою і пов'язана переважно з станом системи, коли вона розподіляється на дві непорожні незв'язні частини, а потім при поверненні до повноцінного варіанту пробує вибрати наступний варіант централізації.

Ймовірність виявлення центру мультикомп'ютерної системи є оберненою величиною до кількості компонент в системі. Для такого визначення можуть обиратись тільки активні в поточний момент часу компоненти. Щодо компонент, в яких міститься функціонал з центром системи можливі варіанти. Наприклад, система в поточний момент часу повністю централізована у варіанті без розподілення центру між декількома компонентами. Також, може бути варіант децентралізованої архітектури, тоді всі компоненти системи містять функціонал центру системи. Але у варіантах децентралізованої архітектури чи частково децентралізованої, частково централізованої або централізованої з певною кількістю компонент з функціоналом центру вважатимемо, що вихід з ладу однієї з компонент системи не призведе до виходу з ладу центру системи. Тому, ймовірність пошуку центру системи буде визначатись так:

$$P_{zl} = \frac{1}{K_{komp}}, \qquad (12)$$

де K_{komp} – кількість компонент в системі.

Якщо припустити, що всі процеси в корпоративній мережі будуть спрощеними щодо дій зловмисника і пошук центру системи для виведення його з ладу буде відповідати моделі, яка описана з припущення за формулою (12), тоді розроблення різної кількості варіантів щодо централізації в системі недоцільне. Але поведінка зловмисника переважно визначається його цілеспрямованістю та набором певних інструментів досягнення мети. Наприклад, після пошуку однієї компоненти з центром системи, він як правило переходить до іншої компоненти, потім до наступної компоненти, продовжуючи захоплювати або виводити з ладу компоненти системи чи комп'ютерні станції. В процесі таких дій зловмисник як правило постійно виконує шаблонні дії з використанням наявних інструментів. Система проти якої спрямовані дії зловмисника повинна мати набір інструментів для відповіді на прояви зловмисних дій. Одним з варіантів на повторювані дії зловмисника є варіант з перебудовою архітектури системи і її центру. В такому випадку ймовірність виявлення центру системи зловмисником буде суттєво менше значення, яке отримується за формулою (12). Тому, запропоноване рішення щодо зміни централізації в архітектурі мультикомп'ютерних систем згідно характерних властивостей, які задано множинами за формулами (1)-(10), покращує стійкість таких систем щодо зловмисних впливів на їх центр.

Висновки. Розроблено характерні властивості централізації в архітектурі мультикомп'ютерних систем приманок та пасток для виявлення та протидії ЗПЗ та КА. Характерні властивості згруповано в множини характерних властивостей. Згідно такого визначення отримано

показники для використання їх при визначенні наступного варіанту централізації в архітектурі мультикомп'ютерних систем, яке вони повинні здійснювати самостійно без залучення адміністратора. Аналіз запропонованого рішення підтвердив перспективність напряму досліджень.

Напрямом подальших досліджень буде деталізація механізмів централізації в архітектурі мультикомп'ютерних систем для формування зв'язків між ними та на їх основі правил, які будуть забезпечувати перехід систем до наступного варіанту централізації.

ЛІТЕРАТУРА:

1. Kashtalian A., Lysenko S., Savenko B., Sochor T., Kysil, T. Principle and method of deception systems synthesizing for malware and computer attacks detection. *Radioelectronic and Computer Systems,* (023. 0(4), 112–151. doi:https://doi.org/10.32620/reks.2023.4.10

2. B. Savenko, A. Kashtalian, S. Lysenko and O. Savenko, "Malware Detection By Distributed Systems with Partial Centralization," *2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, Dortmund, Germany, 2023, pp. 265–270, doi: 10.1109/IDAACS58523.2023.10348773.

3. Savenko O., Sachenko A., Lysenko S., Markowsky G., Vasylkiv N. BOTNET DETECTION APPROACH BASED ON THE DISTRIBUTED SYSTEMS. *International Journal of Computing*, 2020. 19(2), 190–198. https://doi.org/10.47839/ijc.19.2.1761

4. Kashtalian A., Lysenko S., Savenko O., Nicheporuk A., Sochor T., Avsiyevych V. Multi-computer malware detection systems with metamorphic functionality. *Radioelectronic and Computer Systems*, 2024. (1), 152–175. doi:https://doi.org/10.32620/reks.2024.1.13

5. Каштальян А. С. Концептуальна модель архітектури мультикомп'ютерних систем із приманками та пастками для виявлення та протидії зловмисному програмному забезпеченнюю *Information Technology: Computer Science, Software Engineering and Cyber Security*, 2023, № 3, С. 22

6. Suratkar S., Shah K., Sood A. An adaptive honeypot using Q-Learning with severity analyzer. *J Ambient Intell Human Comput*, 2022, 13, P.4865–4876.

7. Lysenko S, Bobrovnikova K, Kharchenko V, Savenko O. IoT Multi-Vector Cyberattack Detection Based on Machine Learning Algorithms: Traffic Features Analysis, Experiments, and Efficiency. *Algorithms*, 2022. 15(7), 239.

8. Leyi S., Yang L., Liu T., Liu J., Shan B., Chen H. Dynamic Distributed Honeypot Based on Blockchain. *IEEE Access*, 2019. P. 1–1. 10.1109/ACCESS.2019.2920239.

9. Fan W., Fernández D., Du Z. Adaptive and Flexible Virtual Honeynet. 2015. 10.1007/978-3-319-25744-0_1.

10. Acosta J. C., Basak A., Kiekintveld C., Kamhoua C. Lightweight On-Demand Honeypot Deployment for Cyber Deception. In: Gladyshev, P., Goel, S., James, J., Markowsky, G., Johnson, D. (eds) Digital Forensics and Cyber Crime. ICDF2C 2021. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Springer, Cham*, 2021, vol. 441.

11. Li Y, Shi L, Feng H. A Game-Theoretic Analysis for Distributed Honeypots. *Future Internet*, 2019. 11(3), 65.

12. Han W., Zhao Z., Doupé A., Ahn G.-J. HoneyMix: Toward SDN-based Intelligent Honeynet. 2016. 10.1145/2876019.2876022.

13. Baykara M., Das R. oftSwitch: a centralized honeypot-based security approach usingsoftware-defined switching for secure management of VLAN networks. *Turkish journal of electrical engineering & computer sciences*, 2019, 27, P.3309–3325

14. Anwar A. H., Zhu M., Wan Z., Cho J. -H., Kamhoua C. A., Singh M. P. Honeypot-Based Cyber Deception Against Malicious Reconnaissance via Hypergame Theory. *GLOBECOM 2022 – 2022 IEEE Global Communications Conference, Rio de Janeiro, Brazil*, 2022, P. 3393–3398.

15. Schindler S., Schnor B., Scheffler T. Sven Schindler. Hyhoneydv6: A hybrid Honeypot Architecture for Ipv6 Networks. *International Journal of Intelligent Computing Research (IJICR)*, 2015, June, Volume 6, Issue 2, P. 562–570

16. Wegerer M., Tjoa S. Defeating the Database Adversary Using Deception – A MySQL Database Honeypot. 2016 *International Conference on Software Security and Assurance (ICSSA), Saint Pölten, Austria*, 2016, P. 6–10.

17. Achleitner S., Porta T., McDaniel P., Sugrim S., Krishnamurthy S., Chadha R. Cyber Deception: Virtual Networks to Defend Insider Reconnaissance. 2016. P.57–68. 10.1145/2995959.2995962.

18. Zaman M., Tao L., Maldonado M., Liu C., Sunny A., Xu S., Chen L. Optimally Blending Honeypots into Production Networks: Hardness and Algorithms. Science of Cyber Security : 5th International Conference,

SciSec 2023, Melbourne, VIC, Australia, July 11–14, Proceedings. Springer-Verlag, Berlin, Heidelberg, 2023, P.285–304.

19. Chiang C. J., Gottlieb Y. M., Sugrim S., Chadha R., Serban C., Poylisher A., Marvel L. M., Santos J. ACyDS: *An adaptive cyber deception system. MILCOM 2016 – 2016 IEEE Military Communications Conference*, P. 800–805.

20. Ehab Al-Shaer. A Cyber Mutation: Metrics, Techniques and Future Directions. *In Proceedings of the 2016 ACM Workshop on Moving Target Defense (MTD '16). Association for Computing Machinery, New York, NY, USA,* 2016, 1.

REFERENCES:

1. Kashtalian, A., Lysenko, S., Savenko, B., Sochor, T., & Kysil, T. (2023). Principle and method of deception systems synthesizing for malware and computer attacks detection. *Radioelectronic and Computer Systems,* 0(4), P.112–151.

2. Savenko, B., Kashtalian, A., Lysenko, S. & Savenko, O. (2023). "Malware Detection By Distributed Systems with Partial Centralization," *2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, *Dortmund, Germany*, P. 265–270.

3. Savenko, O., Sachenko, A., Lysenko, S., Markowsky, G., & Vasylkiv, N. (2022). Botnet detection approach based on distributed systems. *International Journal of Computing*, P.190–198

4. Kashtalian, A., Lysenko, S., Savenko, O., Nicheporuk, A., Sochor, T., & Avsiyevych, V. (2024). Multi-computer malware detection systems with metamorphic functionality. *Radioelectronic and Computer Systems*, (1), P.152–175. doi:https://doi.org/10.32620/reks.2024.1.13

5. Kashtalian, A. (2023). A conceptual model of the architecture of multi-computer systems with decoys and traps for detecting and countering malware and computer attacks. *Information Technology: Computer Science, Software Engineering and Cyber Security*, № 3, P. 22–31.

6. Suratkar, S., Shah, K., Sood, A. (2022). An adaptive honeypot using Q-Learning with severity analyzer. *J Ambient Intell Human Comput*, 13, P. 4865–4876.

7. Lysenko, S., Bobrovnikova, K., Kharchenko, V., Savenko, O. (2022). IoT Multi-Vector Cyberattack Detection Based on Machine Learning Algorithms: Traffic Features Analysis, Experiments, and Efficiency. *Algorithms*, 15(7), 239.

8. Leyi, S., Yang, L., Liu, T., Liu, J., Shan, B., Chen, H. (2019). Dynamic Distributed Honeypot Based on Blockchain. *IEEE Access*, P. 1–1. 10.1109/ACCESS.2019.2920239.

9. Fan, W., Fernández, D., Du, Z. (2015). Adaptive and Flexible Virtual Honeynet. 10.1007/978-3-319-25744-0_1.

10. Acosta, J. C., Basak, A., Kiekintveld, C., Kamhoua, C. (2021). Lightweight On-Demand Honeypot Deployment for Cyber Deception. In: Gladyshev, P., Goel, S., James, J., Markowsky, G., Johnson, D. (eds) Digital Forensics and Cyber Crime. ICDF2C 2021. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Springer, Cham*, vol. 441.

11. Li, Y, Shi, L, Feng, H. (2019). A Game-Theoretic Analysis for Distributed Honeypots. *Future Internet*, 11(3), 65.

12. Han, W., Zhao, Z., Doupé, A., Ahn, G.-J. (2016). HoneyMix: Toward SDN-based Intelligent Honeynet. 10.1145/2876019.2876022.

13. Baykara, M., Das, R. (2019). oftSwitch: a centralized honeypot-based security approach usingsoftware-defined switching for secure management of VLAN networks. *Turkish journal of electrical engineering & computer sciences*, 27, P.3309–3325

14. Anwar, A. H., Zhu, M., Wan, Z., Cho, J. -H., Kamhoua, C. A., Singh, M. P. (2022). Honeypot-Based Cyber Deception Against Malicious Reconnaissance via Hypergame Theory. *GLOBECOM 2022 – 2022 IEEE Global Communications Conference, Rio de Janeiro, Brazil*, P. 3393–3398.

15. Schindler, S., Schnor, B., Scheffler, T. (2015). Sven Schindler. Hyhoneydv6: A hybrid Honeypot Architecture for Ipv6 Networks. *International Journal of Intelligent Computing Research (IJICR)*, June, Volume 6, Issue 2, P. 562–570

16. Wegerer, M., Tjoa, S. (2016). Defeating the Database Adversary Using Deception – A MySQL Database Honeypot. 2016 *International Conference on Software Security and Assurance (ICSSA), Saint Pölten, Austria*, P. 6–10.

17. Achleitner, S., Porta, T., McDaniel, P., Sugrim, S., Krishnamurthy, S., Chadha, R. (2016). Cyber Deception: Virtual Networks to Defend Insider Reconnaissance. P.57–68. 10.1145/2995959.2995962.

18. Zaman, M., Tao, L., Maldonado, M., Liu, C., Sunny, A., Xu, S., Chen, L. (2023). Optimally Blending Honeypots into Production Networks: Hardness and Algorithms. Science of Cyber Security: 5th International Conference, SciSec 2023, Melbourne, VIC, Australia, July 11–14, Proceedings. Springer-Verlag, Berlin, Heidelberg, P.285–304.

19. Chiang, C. J., Gottlieb, Y. M., Sugrim, S., Chadha, R., Serban, C., Poylisher, A., Marvel, L. M., & Santos, J. ACyDS: *An adaptive cyber deception system. MILCOM 2016 – 2016 IEEE Military Communications Conference*, P. 800–805.

20. Ehab, Al-Shaer. (2016). A Cyber Mutation: Metrics, Techniques and Future Directions. *In Proceedings of the 2016 ACM Workshop on Moving Target Defense (MTD '16). Association for Computing Machinery, New York, NY, USA,* 1.

УДК 177:316.4]:004-049.5
DOI https://doi.org/10.32782/IT/2024-4-15

Ольга КИВЛЮК
доктор філософських наук, професор, завідувачка кафедри філософії та психології, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе 210, м. Київ, Україна, 02121
ORCID: *0000-0002-7900-9299*
Scopus Author ID: *57196318811*

Бібліографічний опис статті: Кивлюк, О. (2024). Теоретико-методологічні засади інформаційної та кібернетичної безпеки в умовах цифровізації суспільних процесів. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 125–131, doi: https://doi.org/10.32782/IT/2024-4-15

ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ЗАСАДИ ІНФОРМАЦІЙНОЇ ТА КІБЕРНЕТИЧНОЇ БЕЗПЕКИ В УМОВАХ ЦИФРОВІЗАЦІЇ СУСПІЛЬНИХ ПРОЦЕСІВ

Мета роботи. *Метою дослідження є теоретичний аналіз сутності інформаційної та кібернетичної безпеки в умовах цифровізації суспільних процесів, що передбачає концептуалізацію філософських, соціальних, політично-правових, етичних та технологічних аспектів взаємодії людини, суспільства та цифрових технологій в контексті загроз та ризиків.*

Методологія дослідження: *діалектичний метод дозволив розглянути соціальні процеси в контексті цифровізації суспільства; міждисциплінарний підхід створив можливості поєднання філософських, соціальних, технічних та політико-правових аспектів аналізу інформаційної та кібернетичної безпеки; системний підхід забезпечив розгляд інформаційної та кібернетичної безпеки як багаторівневої системи, яка охоплює технічні, організаційні та соціальні компоненти в контексті ідентифікацій взаємозв'язків між загрозами, ризиками та заходами безпеки; феноменологічний метод скорегував необхідність вивчення процесів цифровізації суспільства в контексті ризиків та потенційних загроз; прогностичний підхід змоделював можливі сценарії розвитку загроз в умовах цифровізації щодо підвищення рівня інформаційної та кібернетичної безпеки; оціночний та порівняльний методи забезпечення механізмів регулювання інформаційної та кібернетичної безпеки у відповідності до сучасних викликів цифровізації.*

Наукова новизна *полягає у інтегрованому підході дослідження впливу цифровізації на безпекові аспекти суспільних процесів, прогнозування кіберзагроз та негативних впливів інформаційних викликів, удосконалення механізмів забезпечення інформаційної та кібернетичної безпеки в цифровому вимірі, адаптації методологічних підходів щодо цифрової трансформацій, розвитку інформаційної інфраструктури, заходів захисту від цифрових маніпуляцій та дезінформації в межах соціально-філософських, політико-правових та етичних аспектів.*

Висновки. *В результаті теоретичного дослідження було здійснено концептуалізацію філософських, соціальних, політично-правових, етичних та технологічних аспектів інформаційної та кібернетичної безпеки в умовах цифровізації суспільних процесів, а також спрогнозовано механізми захисту інформаційного простору від існуючих та можливих загроз у контексті сучасних викликів техногенної цивілізації.*

Ключові слова: *інформаційна безпека, кібербезпека, цифровізація, людина, цифрові технології, суспільні процеси, техногенна цивілізація.*

Olga KYVLIUK
Doctor of Philosophical Sciences, Professor, Head of the Department of Philosophy and Psychology, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121,
o.p.kyvliuk@gmail.com
ORCID: *0000-0002-7900-9299*

To cite this article: Kyvliuk, O. (2024). Teoretyko-metodolohichni zasady informatsiynoyi ta kibernetychnoyi bezpeky v umovakh tsyfrovizatsiyi suspil nykh protsesiv [Theoretical and methodological foundations of information and cyber security in the context of digitalisation of social processes] *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 125–131, doi: https://doi.org/10.32782/IT/2024-4-15

THEORETICAL AND METHODOLOGICAL FOUNDATIONS OF INFORMATION AND CYBER SECURITY IN THE CONTEXT OF DIGITALISATION OF SOCIAL PROCESSES

Aim of the study. *The purpose of the study is to theoretically analyse the essence of information and cyber security in the context of the digitalisation of social processes, which involves conceptualising the philosophical, social, political, legal, ethical and technological aspects of the interaction between man, society and digital technologies in the context of threats and risks.*

Methodology. *The dialectical method allowed us to consider social processes in the context of the digitalisation of society; the interdisciplinary approach created opportunities to combine philosophical, social, technical, political and legal aspects of the analysis of information and cyber security; the systemic approach ensured consideration of data and cyber security as a multilevel system that covers technical, organisational and social components in the context of identifying the relationships between threats, risks and security measures; the phenomenological method adjusted the need to study the processes of digitalisation of society in the context of risks and potential threats; the systemic approach ensured consideration of information and cyber security as a multilevel system that covers technical, organisational and social components in the context of identifying the relationships between threats, risks and security measures; the phenomenological method adjusted the need to study the processes of digitalisation of society in the context of risks and potential threats;*

The scientific novelty *lies in an integrated approach to studying the impact of digitalisation on the security aspects of social processes, forecasting cyber threats and the negative effects of information challenges, improving mechanisms for ensuring information and cyber security in the digital dimension, adapting methodological approaches to digital transformations, developing information infrastructure, measures to protect against digital manipulation and disinformation within the socio-philosophical, political, legal and ethical aspects.*

Conclusions. *As a result of the theoretical study, the author conceptualised the philosophical, social, political, legal, ethical and technological aspects of information and cyber security in the context of the digitalisation of social processes, and also predicted the mechanisms for protecting the information space from existing and possible threats in the context of modern challenges of technogenic civilisation.*

Key words: *information security, cybersecurity, digitalisation, human, digital technologies, social processes, and technogenic civilisation.*

Актуальність проблеми. Цифровізація суспільних процесів є не просто феноменом, а й незаперечним реальним фактом. Даний процес, як і технологізація, комп'ютеризація, чи то інформатизація – є неспинним і незворотнім, а отже, має як свої переваги так і недоліки. Зважаючи на умови протидії воєнній агресії Російської Федерації, завдання забезпечення та запобігання загрозам кіберзлочинності, інформаційного маніпулювання, протиправного розповсюдження персональних даних громадян, державних реєстрів, національних та комерційних таємниць тощо, стає вкрай складним і більш ніж актуальним. Тому важливо, щоб проблематика інформаційної та кібернетичної безпека носила не тільки технократичний характер, а й набула антропологічного, етично-ціннісного значення, не зважаючи на її цифровий, а подекуди віртуальний вимір.

Теоретико-методологічні засади інформаційної та кібернетичної безпеки як концептуальний підхід, фокусується на осмисленні принципів, цінностей, і стратегій захисту інформації та цифрових інфраструктур від реальних та можливих загроз, використовуючи як технічнологічні, так і етичні підходи.

Захист інформації від несанкціонованого доступу; гарантії точності, цілісності та незмінності (за умови необхідності) даних; забезпечення безперебійної доступності до інформаційних систем, авторизації користувача, здійснення пошукових завдань; прозорість та відповідальність оператора/організації щодо збереження, обробки, представлення та передачі даних; дотримання етичних стандартів, враховуючи права людини, приватність та інші соціальні аспекти формують основні принципи філософії інформаційної та кібернетичної безпеки.

Філософські питання, що лежать в основі концептуального підходу до цифровізації суспільних процесів, які потребують заходів безпеки в інформаційно-комунікаційному просторі можна визначити наступним чином:

- Де закінчується свобода і починається контроль?
- Як знайти баланс між правом на приватність і потребою в здійсненні заходів для забезпечення безпеки?
- Як гарантувати, що технології та штучний інтелект сприятимуть захисту людської гідності, особистої приватності та дотриманню прав людини, уникатимуть дискримінації, а також будуть використовуватися для загального блага, а не для маніпуляцій чи необґрунтованого контролю?
- Чи повинна людина обмежувати розвиток технологій та штучного інтелекту через ризики, які вони створюють у сфері безпеки?

- Як забезпечити інклюзивність та рівний доступ до цифрових технологій для всіх верств населення в контексті запобігання цифрової нерівності?
- Яке місце відводиться людині у системі кібербезпеки?
- Як гармонізувати технократичні та гуманістичні цінності у цифровому суспільстві?

Це, безумовно, лише частина питань, які вимагають детального розгляду та пошуку відповідей.

Аналіз останніх досліджень і публікацій. Акцентуючи на міждисциплінарності та багаторівневості процесу цифровізації, соціально-філософський аспект даної проблематики досліджувався в роботах як вітчизняних науковців: Воронокової В. та Нікітенко В., які в монографії «Філософія цифрової людини і цифрового суспільства: теорія і практика» аналізуються взаємодія цифрового суспільства та цифрової людини як соціального феномену; Олексенко Р. зосереджується на ціннісному вимірі цифрової освіти та цифрової людини у діджиталізованому суспільстві, Бєлов Д. розглядає доктринальні засади цифрового права; так і відомих закордонних вчених серед яких відмітимо: Андрюкайтене Р. описує цифрові суспільні трансформації, Бріньолфссон Е., Макафі Е. досліджують вплив цифрових технологій на економіку та суспільство, Скінер К., Штаб К., Даґоґо О., Пентленд А. також вивчають різні аспекти цифровізації, включаючи економічні, соціальні та технологічні зміни в сучасному суспільстві та інші.

Сутність поняття інформаційної безпеки розглядалася в роботах: Золотар О., яка досліджує права та свободу людини в умовах інформаційного суспільства його перспектив та загроз; Бикова О., який аналізує правовий та культурологічний виміри інформаційної безпеки, підкреслюючи її важливість у сучасному суспільстві; Валлерстейн І. розглядає інформаційну безпеку з урахуванням трансдисциплінарної стратегії досліджень; Петрик В. відмічає, що поняття «безпека» і «небезпека» є діалектично взаємозалежними, тобто не існують у природі окремо.

Незважаючи на дотичність і синтезованість інформаційної та кібернетичної безпеки, більшу увагу другому було сконцентровано в дослідженнях: Деннінг Д. підкреслює, що кібербезпека спрямована на захист комп'ютерних систем, мереж та цифрової інформації від несанкціонованого доступу, модифікації чи знищення; Столлінгс В. концентрується на протидії кіберзагрозам та мережевій безпеці; Бурячок В. досліджує концептуальні основи створення ефективної системи кібернетичної безпеки на державному рівні та наголошує на важливість міжнародного співробітництва у сфері кібербезпеки.

Мета дослідження полягає у теоретичному аналіз сутності інформаційної та кібернетичної безпеки в умовах цифровізації суспільних процесів, що передбачає концептуалізацію філософських, соціальних, технічних аспектів взаємодії людини, суспільства та цифрових технологій в контексті загроз та ризиків.

Цифровізація суспільних процесів. Трансформація суспільних процесів пов'язана з цифровізацією, а саме: зростання кількості цифрових даних та електронних транзакцій, розвиток технологій штучного інтелекту та Інтернету речей, трансформація соціальних комунікацій та бізнес-процесів, розмивання кордонів між цифровою та фізичною реальністю, формування нових моделей соціальної комунікації як в локальному так і глобальному форматах – створює нові виклики для суспільства в цілому та конкретної людини зокрема, концентруючи увагу на гармонійному балансі між безпекою та правами людини (Pentland, 2014).

Цифрове суспільство виникло як унікальний і багатогранний соціальний феномен, який інтегрував сучасні технології у всі сфери життя людини. Воно проходить етапи еволюції, трансформуючись із простих двовимірних взаємодій у складні тривимірні системи, де реальне і віртуальне поступово зливаються в єдине ціле. Цей процес супроводжується «розмиванням» меж між фізичною та цифровою реальністю, що надає нових можливостей для соціальної взаємодії, водночас викликаючи низку викликів та змін у людській діяльності.

Цифрова «ерозія» стала помітною на різних етапах розвитку суспільства. Вона проявляється через перехід соціальної активності людини в площину, де центральними елементами стають інформація, дані та цифрові коди. У результаті цього формується новий шар ідентичності людини – її «цифрові атрибути». Ці атрибути, включаючи онлайн-профілі, віртуальну репутацію, цифрову компетенцію та інші характеристики, стали невід'ємною частиною особистості та вплинули на те, як людина взаємодіє з суспільством. Процес розвитку цифрового суспільства можна розглядати як взаємну трансформацію: з одного боку, цифровізація змінює соціальну діяльність людини, з іншого боку – «цифрові атрибути» людей формують характер і темпи цієї трансформації. Взаємодія цих аспектів сприяє формуванню нових

механізмів управління, комунікації та інтеграції у цифровому просторі (Кивлюк, Воронкова, Нікітенко, 2023).

Цифровізація реалізує накопичення інформаційних ресурсів і даних, їхню циркуляцію та обробку. Водночас інтелектуалізація додає цьому процесу зміст, інтегруючи дані в системи прийняття рішень та оптимізації діяльності. Такий симбіоз дозволяє суспільству адаптуватися до викликів цифрової епохи та максимально використовувати потенціал технологій.

Проте цифровізація соціальної діяльності людини не лише створює умови для функціонування цифрового суспільства, але й глибоко впливає на саму людину. Вона змінює її поведінку, цінності та взаємодію зі світом, формуючи нові підходи до комунікації, роботи та навіть мислення. Люди поступово набувають нових навичок, адаптуючись до швидких технологічних змін, що водночас відкриває нові горизонти та ставить перед суспільством складні моральні, етичні та соціальні питання. Цифрове суспільство можна розглядати як динамічну систему, що формується завдяки взаємодії технологій, даних і людської діяльності. Це суспільство, де цифрові інструменти не лише полегшують життя, а й створюють нові вимоги та загрози, змінюючи уявлення про реальність та соціокультурну взаємодію.

Досліджуючи права людини через призму держаної політики цифровізації суспільних процесів у сфері нацбезпеки, Д. Чижов розглядає два напрямки наукового пошуку: забезпечення інформаційної безпеки як складової національної безпеки в умовах цифровізації та упровадження цифрової трансформації Збройних сил і забезпечення інформаційної безпеки. Він вважає, що інформаційна і кібернетична системи мають розглядатися синергетично та системно з врахуванням природи технократичних, міждисциплінарних, темпоральних підходів. Обґрунтовуючи значимість інформаційної та кібернетичної безпеки особистості, звертає увагу на симбіоз безпеки «держава – суспільство – людина» та акцентує на переході держаної безпеки в оборонну, у зв'язку з воєнним станом в країні (Чижов, 2021).

Інформаційна безпека. Інформаційна безпека являє собою стан захищеності інформаційного простору і охоплює захист інформації в усіх її формах та проявах, функціонує в інтересах громадян, організацій, держави, що окрім того включає заходи захисту від негативних інформаційних маніпуляцій, пропаганди та інформаційно-психологічну безпеку. Інформаційна безпека – це не тільки засіб, чи то механізм запобігання, або подолання негативних проявів цифровізації, – це засіб для досягнення інформаційно-технічного прогресу з врахуванням гуманістичних, етичних, соціальних, політично-правових тощо аспектів.

З точки зору філософсько-антропологічного підходу варто наголосити на цінності та захищеності психічного, ментального, навіть фізичного здоров'я людини від деструктивних інформаційних впливів, що призводить до негативних наслідків.

О. Золотар у своїй монографії «Інформаційна безпека людини: теорія і практика» зазначає, що у філософському контексті важливо зосередитися на трьох ключових аспектах дослідження даної проблематики: онтологічному, де природа інформаційної безпеки є складною багатовимірною категорією, що включає аналіз сутності інформаційних загроз, їх впливу на людину, суспільство та ноосферу; гносеологічному, що аналізує пізнавальний аспект, який здійснює людина як суб'єкт пізнання, інформаційної безпеки, включаючи вивчення процесів отримання, обробки, зберігання та передачі знань, а також методів їх захисту; логічному, де досліджуються основи побудови системи аргументів, стратегій та структур інформаційної безпеки, а також розглядаються принципи логічної послідовності в процесі захисту, прийнятті рішень і розробці політик безпеки, що враховують причинно-наслідкові зв'язки між загрозами, ризиками та заходами захисту (Золотар, 2018).

Тобто, автор пропонує системний підхід до аналізу інформаційної безпеки, який охоплює не лише технічні аспекти, але й філософські, що є важливим заданням для забезпечення гармонійного функціонування інформаційного простору.

Аналіз сучасних публікацій дозволив узагальнити механізми забезпечення інформаційної безпеки особистості – це:

по-перше, загальна освіченість кожної людини, тобто рівень її теоретичної освіти та практичних компетентностей щодо інформаційних можливостей та загроз в контексті успішної життєдіяльності та гармонійного саморозвитку;

по-друге, здійснення загальнодержавних заходів задля створення безпекового рівня як в межах задоволення потреб особистості в інформаційних продуктах, так і можливих загроз;

по-третє, створення дієвого правового поля в контексті державного контролю функціонування інформаційного простору/середовища, щодо запобіганню інформаційних небезпек та відповідних процедур покарання;

по-четверте, гарантування захищеності інтересів людини від можливих гіпотетичних інформаційних загроз різного роду;

по-п'яте, забезпечення інформаційна рівності.

Тобто, з огляду на зазначене вище, людина, при всьому її бажанні та освіченості, не зможе створити свій особистий безпековий інформаційний вимір, а для цього має існувати суспільний інформаційно-безпековий рівень. От наприклад, у Філософському енциклопедичному словнику, «безпека національна – поняття політології та соціальної філософії, яке визначає певний ступінь захищеності особистості, суспільства, держави від внутрішніх та зовнішніх загроз, що дозволяє їм нормально існувати та розвиватися» (Шинкарук, 2002: с. 49). Національна безпека – це і соціальна система, це і державний пріоритет, це і національно-культурний процес, це законодавча доктрина, це політична інституція, це економічний розвиток тощо.

Отже, інформаційна безпека суспільства залежить від сукупності інтегрованих політичних, правових, економічних, соціальних, технологічних процесів, а також і окремих особистостей, дії яких спрямовані на попередження, запобігання, виявлення, подолання негативних подій, або вже їх наслідків.

Зрозуміло, що взаємозалежність високого рівня інформаційної безпеки суспільства і окремої людини в контексті інтелектуального, етичного, духовного, критичного та креативного розвитку очевидна. А провідником, так би мовити посередником, між людиною і суспільством, що гармонізує дану взаємозалежність є – держава, яка має створювати умови для розвитку та продуктивного функціонування, стабільності та захищеності інформаційної інфраструктури від негативних впливів: інформаційні війни, маніпулювання, інформаційний тероризм, незаконне оприлюднення конфіденційної інформації, кіберзлочинність і т.д., тобто все те, що загрожує та шкодить національним інтересам, суверенітету держави, економічній та політичній стабільності, міжнародному співробітництву.

Кібернетична безпека. Кібернетична безпека в епоху цифровізації сфокусована на забезпеченні конфіденційності, цілісності та доступності інформації в кіберпросторі, що охоплює захист критичної інформаційної інфраструктури та протидію кіберзагрозам.

Для ефективного захисту інформації та активів у сфері кібернетичної безпеки недостатньо застосовувати єдиного векторного підходу. Необхідно враховувати різноманітні види засобів та заходів контролю кібербезпеки та оптимально інтегрувати їх у діяльнісну сферу. Усі технологічні процеси стикаються з ризиками безпеки та конфіденційності. Хоча сучасні цифрові технології здатні ефективно протидіяти кіберзагрозам, цього замало. Важливо мати гарантії, що державні, політичні, комерційні процеси та етично-правова поведінка окремих особистостей також спрямовані на зниження або усунення цих ризиків. Оскільки забезпечення кібербезпеки є складним і багатогранним завданням, як на державному, комерційному так і особистісному рівнях, дедалі частіше впроваджують стандартизовані підходи для управління інформаційною безпекою. У цьому контексті ключову роль відіграють системи управління інформаційною безпекою, які: виявляють загрози в сфері ІТ-безпеки швидко і на ранніх стадіях; своєчасно впроваджують відповідні заходи для нейтралізації ІТ-ризиків; гарантують відповідність внутрішнім вимогам та міжнародним стандартам; зменшують навантаження на ІТ-персонал за допомогою автоматизації виявлення вразливостей; забезпечують надання повної доказової бази щодо інцидентів комп'ютерної безпеки. Таким чином, системний підхід до кібербезпеки дозволяє створити ефективніші механізми захисту, що враховують як технологічні, так і організаційні аспекти (Кивлюк; Воронкова, 2022).

Превентивність, комплексність, адаптованість, багаторівневість, стандартизованість, безперервність, економічна доцільність, організаційна узгодженість, правова захищеність – це основні принципи функціонування кібернетичної безпеки.

Аналіз існуючих науково-практичних розвідок дозволив узагальнити напрямки та завдання розвитку кібернетичної безпеки задля підвищення її рівня:

- освітня діяльність в контексті «кібергігієни» (підвищення обізнаності щодо основ кібербезпеки користувачів ІТ, створення навчальних програм, курсів та тренінгів для моделювання можливих загроз та пропедевтики їх подолання);
- розвиток культури кібербезпеки, як наслідку високого рівня кіберосвіченості;
- автоматизація процесів реагування на кіберзагрози (наприклад, SOAR-платформи дозволяють автоматизувати рутинні завдання кіберзахисту та прискорюють реагування на інциденти);
- міжнародна співпраця в контексті глобальних ініціатив та обміну інформацією задля боротьби з кіберзлочинністю та аналізу кіберзагроз;

- розробка стандартів безпеки (наприклад, GDPR або CCPA), їх моніторинг та управління;
- удосконалення механізмів для захисту конфіденційних даних, тобто вдосконалення методів біометричної автентифікації;
- використання штучного інтелекту для безперервного аналізу та алгоритму передбаченню кіберзагроз у режимі реального часу, але при умові що загроза йде не від самого ШІ;
- розвиток квантової криптографії, тобто сучасні еволюційні технології створюють нові загрози, а з їх розвитком треба бути «озброєним» новим рівнем безпеки;
- використання блокчейн-технологій для забезпечення захисту (наприклад, R3 Corda? Quorum, Dragonchain, Hyperledger Besu) та децентралізації управління даними;
- політика «нульової довіри» з елементами мікросегментація для мінімізації ризиків;

Висновки і перспективи подальших досліджень. Формування нових парадигм інформаційної та кібернетичної безпеки вимагає створення інтегрованих теоретико-методологічних підходів щодо дослідження викликів цифровізації суспільства. Етичні питання в контексті соціальної відповідальності цифровізації потребують комплексного вивчення та розробки стандартів етичного використання цифрових даних у цифровому середовищі. Серед перспективних завдань подальших наукових розвідок також відзначаємо: дослідження впливу цифровізації на соціальні інституції, освітні ініціативи щодо підвищення рівня обізнаності серед громадян та кібернетичної культури, правове регулювання та міжнародна стандартизація протистоянню кіберзагрозам, розробку інтелектуальних систем безпеки в контексті застосування штучного інтелекту. Взагалі, подальші дослідження мають бути зорієнтовані на створення інтегрованих, адаптивних і стійких систем інформаційної та кібернетичної безпеки, здатних ефективно протидіяти існуючим і гіпотетично можливим викликам цифрового світу.

Отже, в результаті теоретичного дослідження було здійснено концептуалізацію філософських, соціальних, політично-правових, етичних та технологічних аспектів інформаційної та кібернетичної безпеки в умовах цифровізації суспільних процесів, а також спрогнозовано механізми захисту інформаційного простору від існуючих та можливих загроз у контексті сучасних викликів техногенної цивілізації.

ЛІТЕРАТУРА:

1. Бріньолфссон Е., Макафі Е. Друга епоха машин: робота, прогрес та процвітання в часи надзвичайних технологій. Київ: K. FUND, 2016. 236 с.

2. Бурячок В. Л. Основи формування державної системи кібернетичної безпеки : монографія. К. : НАУ, 2013. 431 с.

3. Даґоґо О. Нове мислення. Від Айнштейна до штучного інтелекту: наука і технології, що змінили світ / пер. з англ. І. Возняка. Харків : Віват, 2021. 368 с.

4. Золотар О. О. Інформаційна безпека людини : теорія і практика : монографія. Київ : ТОВ «Видавничий дім «АртЕк», 2018. 446 с.

5. Кивлюк О., Воронкова В., Нікітенко В. Цифрові права як вираження цифрових атрибутів людини: соціально-філософське обґрунтування. Освітній дискурс : збірник наукових праць. Київ : ТОВ«НІА «НТІ». 2023. Випуск 44 (4-6). С. 7–26. https://doi.org/10.33930/ed.2019.5007.44(4-6)-1.

6. Кивлюк О. П., Воронкова В. Г. Філософська рефлексія інформаційної безпеки у цифровому середовищі: проблеми, ризики, правове забезпечення. Innovative resources of modern science : collective monograph. Compiled by V. Shpak; Chairman of the Editorial Board S. Tabachnikov. Sherman Oaks, California : GS Publishing Services, 2022. С. 160–172. (186 p.) DOI: 10.51587/9798-9866-95907-2022-009-160-172.

7. Петрик В., Канарський Ю. Методи гібридної війни Росії проти України. Напрями протидії. Information Technology and Security. 2015. Vol. 3, Iss. 1 (4). P. 30–37.

8. Скінер К. Людина цифрова. Фабула. 2020. 272 с.

9. Філософський енциклопедичний словник : энциклопедия / НАН України, Ін-т філософії ім. Г. С. Сковороди ; голов. ред. В. І. Шинкарук. Київ : Абрис, 2002. 742 с.

10. Чижов Д. А., Державна політика забезпечення прав людини у сфері національної безпеки в умовах цифровізації. Науковий вісник Національної академії внутрішніх справ, 2021. № 4 (121). с. 46–52.

11. Alex Pentland, Social Physics: How Social Networks Can Make Us Smarter, Scribe Publications, 2014, p. 300.

12. Denning D., Information Warfare and Security, ACM Press, 1999, p. 522.

REFERENCES:

1. Brynjolfsson, E., & McAfee, A. (2016). Druha epokha mashyn: robota, prohres i protsvitannya v chas nadzvychaynykh tekhnolohiy [The Second Machine Age: Work, Progress, and Prosperity in a Time of Extraordinary Technology]. Kyiv. FUND. [in Ukrainian].

2. Buryachok, V. L. (2013). Osnovy formuvannya systemy kiberbezpeky derzhavy: monohrafiya [Fundamentals of the formation of a state cyber security system: monograph] Kyiv: NAU. [in Ukrainian].

3. Dagogo, O. (2021). Nove myslennya. Vid Eynshteyna do shtuchnoho intelektu: nauka i tekhnolohiyi, yaki zminyly svit [New Thinking. From Einstein to Artificial Intelligence: Science and Technologies That Changed the World] (trans. from English by I. Wozniak). Kharkiv: Vivat. [in Ukrainian].

4. Zolotar, O. O. (2018). Informatsiyna bezpeka lyudyny: teoriya i praktyka [Human Information Security: Theory and Practice: Monograph]. Kyiv: LLC “ArtEk Publishing House”. [in Ukrainian].

5. Kyvlyuk, O., Voronkova, V., & Nikitenko, V. (2023). Tsyfrovi prava yak vyrazhennya tsyfrovykh atrybutiv lyudyny: sotsial'no-filosofs'ke obgruntuvannya [Digital rights as an expression of human digital attributes: a socio-philosophical justification]. *Educational discourse: collection of scientific works*, 44(4-6), P. 7–26. [in Ukrainian]. https://doi.org/10.33930/ed.2019.5007.44(4-6)-1

6. Kyvlyuk, O. P., & Voronkova, V. G. (2022). Filosofs'ka refleksiya informatsiynoyi bezpeky v tsyfrovomu seredovyshchi: problemy, ryzyky, pravove zabezpechennya [Philosophical reflection on information security in the digital environment: problems, risks, legal support]. *Innovative resources of modern science: collective monograph* (pp. 160–172). Sherman Oaks, California: GS Publishing Services. https://doi.org/10.51587/9798-9866-95907-2022-009-160-172

7. Petryk, V., & Kanarsky, Y. (2015). Methods of Russia’s hybrid war against Ukraine [Methods of Russia’s hybrid war against Ukraine]. Directions of counteraction. *Information Technology and Security*, 3(1), 30–37. [in Ukrainian].

8. Skinner, K. (2020). Tsyfrova lyudyna [Digital Man]. Fabula. [in Ukrainian].

9. hynkaruk, V. I. (Ed.). (2002). Filosofs'kyy entsyklopedychnyy slovnyk: Entsyklopediya [Philosophical Encyclopedic Dictionary: Encyclopedia]. Kyiv: Abrys. [in Ukrainian].

10. Chyzhov, D. A. (2021). Derzhavna polityka zabezpechennya prav lyudyny u sferi natsional'noyi bezpeky v umovakh tsyfrovizatsiyi [State policy of ensuring human rights in the sphere of national security in the context of digitalization]. *Naukovyy visnyk Natsional'noyi akademiyi vnutrishnikh spravScientific* (4), 46–52. [in Ukrainian].

11. Pentland, A. (2014). *Social Physics: How Social Networks Can Make Us Smarter*. Scribe Publications.

12. Denning, D. (1999). *Information Warfare and Security*. ACM Press.

УДК 004.93
DOI https://doi.org/10.32782/IT/2024-4-16

Павло КУНДЕНКО
аспірант кафедри інформаційних технологій і систем, Український державний університет науки і технологій, вул. Лазаряна, 2, Дніпро, Україна, 49010
ORCID: *0009-0000-9388-2783*

Олександр СОКОЛ
аспірант кафедри інформаційних технологій і систем, Український державний університет науки і технологій, вул. Лазаряна, 2, Дніпро, Україна, 49010
ORCID: *0009-0002-8851-2045*

Наталія КАЛІНІНА
кандидат технічних наук, доцент, доцент кафедри інформаційних технологій і систем, Український державний університет науки і технологій, вул. Лазаряна, 2, Дніпро, Україна, 49010
ORCID: *0009-0001-6416-4907*

Вікторія ГНАТУШЕНКО
доктор технічних наук, професор, завідувач кафедри інформаційних технологій і систем, Український державний університет науки і технологій, вул. Лазаряна, 2, Дніпро, Україна, 49010
ORCHID: *0000-0001-5304-4144*
Scopus-Author ID: *57218682553*

Бібліографічний опис статті: Кунденко, П., Сокол, О., Калініна, Н., Гнатушенко, В. (2024). Нейромережевий підхід виявлення водних об'єктів на мультиспектральних зображеннях. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 132–143, doi: https://doi.org/10.32782/IT/2024-4-16

НЕЙРОМЕРЕЖЕВИЙ ПІДХІД ВИЯВЛЕННЯ ВОДНИХ ОБ'ЄКТІВ НА МУЛЬТИСПЕКТРАЛЬНИХ ЗОБРАЖЕННЯХ

Мета роботи. *Метою данного дослідження є створення та впровадження ефективного нейромережевого підходу на основі моделі U-Net для детектування структурних елементів водного середовища України шляхом аналізу мультиспектральних зображень із супутника Sentinel-2. Це передбачає підвищення точності та швидкості виявлення змін у водних масах та елементах екомережі, що сприятиме оперативному реагуванню на екологічні виклики.*

Методологія. *Для підготовки даних застосовується автоматичний підхід до формування навчального набору з використанням нормалізованого різницевого водного індексу (NDWI). На супутникових знімках Sentinel-2 (зелений та ближній інфрачервоний канали) генеруються бінарні маски води на основі порогового значення, що дає змогу створювати великий набір даних для навчання без ручної анотації. Модель базується на глибинній нейронній мережі U-Net: енкодер виокремлює високорівневі ознаки, а декодер відновлює просторову роздільну здатність і формує сегментаційну карту. Під час навчання застосовується комбінована функція втрат, яка поєднує бінарну крос-ентропію та коефіцієнт Дайса. Оцінка моделі здійснюється за метриками F1-міри, точності, повноти й індексу Жаккара (IoU). Модель тестується на повнорозмірних знімках різних ділянок, демонструючи високий рівень узагальнення.*

Наукова новизна. *Запропоновано підхід, що полягає в об'єднанні автоматичного формування масок водних об'єктів (за допомогою NDWI) та нейромережі U-Net, адаптованої до задач бінарної сегментації великих супутникових зображень. Така інтеграція уникає ручної розмітки та забезпечує точні результати виявлення води. Використання комбінованої функції втрат покращує чутливість моделі до тонких водних структур, а стійкість до варіативності спектральних характеристик і атмосферних перешкод підтверджується експериментами в різних умовах зйомки.*

Висновки. *Розроблено ефективну модель для сегментації водних об'єктів на супутникових знімках Sentinel-2 з використанням глибокої нейронної мережі архітектури U-Net. Процес включав підготовку даних, де автоматично генерувалися маски водних об'єктів на основі індексу NDWI, та детальне навчання моделі з комбінованою функцією втрат, що поєднує бінарну крос-ентропію та коефіцієнт Дайса.*

Модель досягла високих показників точності, підтверджених метриками F1-міри (0.8897), точності (0.8721), повноти (0.9080) mIoU (0.8013), що свідчить про її здатність до точного виявлення водних

об'єктів різних розмірів і форм. Перевагами використаних підходів є поєднання глибинного навчання з попередньою підготовкою даних, що дозволило автоматизувати процес виявлення водних об'єктів та забезпечити високу точність сегментації. Модель демонструє гнучкість, масштабованість та обчислювальну ефективність, що робить її придатною для практичного застосування в моніторингу водних ресурсів і екологічних дослідженнях. Майбутні напрямки дослідження направлені на експерименти з іншими архітектурами нейронних мереж, використання альтернативних методів створення масок та врахування просторово-часової інформації, що дасть змогу підвищити ефективність і універсальність моделі.

Ключові слова: *згорткові нейронні мережі, семантична сегментація, супутникові знімки, водні ресурси, екологічний моніторинг.*

Pavlo KUNDENKO
Postgraduate Student at the Department of Information Technology and Systems, Ukrainian State University of Science and Technologies, 2, Lazariana Str., Dnipro, Ukraine, 49010, pavel.kundenko@gmail.com
ORCHID: *0009-0000-9388-2783*

Oleksander SOKOL
Postgraduate Student at the Department of Information Technology and Systems, Ukrainian State University of Science and Technologies, 2, Lazariana Str., Dnipro, Ukraine, 49010, wawjiu4ok@gmail.com
ORCID: *0009-0002-8851-2045*

Nataliia KALININA
Candidate of Technical Science, Associate Professor, Associate Professor at the Department of Information Technology and Systems, Ukrainian State University of Science and Technologies, 2, Lazariana Str., Dnipro, Ukraine, 49010, natalyku13@gmail.com
ORCID: *0009-0001-6416-4907*

Viktoriia HNATUSHENKO
Doctor of Technical Science, Professor, Head of Department of Information Technology and Systems, Ukrainian State University of Science and Technologies, 2, Lazariana Str., Dnipro, Ukraine, 49010, vik.v.hnatushenko@ust.edu.ua
ORCID: *0000–0003–3140–3788*
Scopus-AuthorID: *57218682553*

To cite this article: Kundenko, P., Sokol, O., Kalinina. N., Hnatushenko, V. (2024). Neiromerezhevyi pidkhid vyiavlennia vodnykh obiektiv na multyspektralnykh zobrazhenniakh [Neural network approach to detecting water objects in multispectral images]. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 132–143, doi: *https://doi.org/10.32782/IT/2024-4-16*

NEURAL NETWORK APPROACH TO DETECTING WATER OBJECTS IN MULTISPECTRAL IMAGES

Aim of the Study. *The goal of this research is to develop and implement an effective neural network approach based on the U-Net model for detecting structural elements of Ukraine's aquatic environment through the analysis of multispectral images from the Sentinel-2 satellite. This entails enhancing both accuracy and speed in identifying changes in water bodies and ecological network components, thereby enabling prompt responses to environmental challenges.*

Methodology. *An automatic approach to forming the training set is utilized, relying on the Normalized Difference Water Index (NDWI). Binary water masks are generated based on a threshold value applied to the Sentinel-2 green and near-infrared channels, enabling the creation of a large training dataset without manual labeling. The model is built around the U-Net architecture: the encoder extracts high-level features, while the decoder restores spatial resolution and produces the segmentation map. During training, a combined loss function that merges binary cross-entropy with the Dice coefficient is applied. Model evaluation uses F1 score, precision, recall, and Intersection over Union (IoU). Testing the model on full-resolution images of different areas demonstrates a high degree of generalization.*

Scientific Novelty. *The proposed approach integrates automatic generation of water body masks (via NDWI) with a U-Net-based architecture tailored for binary segmentation of large satellite images. This integration eliminates the need for manual labeling and ensures accurate water detection results. The use of a combined loss function improves the model's sensitivity to fine water structures, while experiments under various imaging conditions confirm its robustness against spectral and atmospheric variability.*

Conclusions. *An effective model has been developed for the segmentation of water bodies in Sentinel-2 satellite images using a deep neural network with a U-Net architecture. The process involved data preparation, wherein*

water masks were automatically generated based on the NDWI index, followed by thorough training of the model with a combined loss function that merges binary cross-entropy and the Dice coefficient. The model achieved high accuracy metrics, confirmed by an F1 score of 0.8897, precision of 0.8721, recall of 0.9080, and an IoU of 0.8013, demonstrating its capability for accurately detecting water bodies of various sizes and shapes.The advantages of the chosen approaches stem from combining deep learning with prior data processing, which automated the detection of water bodies and ensured high segmentation accuracy. The model exhibits flexibility, scalability, and computational efficiency, making it suitable for practical applications in water resource monitoring and environmental studies. Future research directions include experimenting with different neural network architectures, employing alternative methods for mask generation, and incorporating spatiotemporal information to further enhance the model's effectiveness and versatility.

Key words: *convolutional neural networks, semantic segmentation, satellite imagery, water resources, environmental monitoring.*

Постановка проблеми. Моніторинг змін у структурних елементах водного середовища є критично важливим для збереження біорізноманіття та стійкого розвитку екосистем. Актуальність цього питання зростає в контексті глобальної зміни клімату (Climatechange, 2021), яка призводить до підвищення температури води, зміни рівня моря та збільшення частоти екстремальних погодних явищ. Висхідні екологічні виклики, спричинені також антропогенними факторами та військовими конфліктами, погіршують якість водних ресурсів через забруднення та руйнування екосистем (UNEP, 2021). Постійний розвиток сфери нейронних мереж відкриває нові можливості для аналізу мультиспектральних зображень, що використовуються для моніторингу водного середовища (LiW., 2019). Сучасні нейронні мережі, зокрема глибокі згорткові нейронні мережі, здатні обробляти великі обсяги даних з високою точністю, дозволяючи виявляти навіть незначні зміни у структурі водних об'єктів (MaL., 2019). Проте, попри значний прогрес, існує простір для покращення методів обробки та аналізу даних, зокрема в контексті підвищення точності, швидкості та інтерпретованості результатів (ZhangC., 2020). З огляду на зазначене, дослідження, спрямовані на розробку ефективних нейромережевих підходів до моніторингу водного середовища та виявленню його елементів на мультиспектральних зображеннях, є актуальними та мають значний науковий і практичний інтерес.

Аналіз останніх досліджень і публікацій. Останніми роками спостерігається значний прогрес у застосуванні штучного інтелекту (ШІ) та глибокого навчання для моніторингу водних мас та елементів екомережі. Це зумовлено необхідністю оперативного реагування на екологічні виклики та можливістю обробки великих обсягів даних дистанційного зондування(Dobrovolska, 2024; Hnatushenko V.V., 2019). Одним із ключових напрямків є використання згорткових нейронних мереж (CNN) для класифікації та сегментації водних об'єктів на мультиспектральних зображеннях. Наприклад, Li та інші розробили метод для виявлення нафтових розливів на основі поляриметричних SAR-даних, використовуючи глибоке навчання. Їхня модель досягла високої точності виявлення, що сприяє швидкому реагуванню на екологічні аварії (Li, 2019).

Інші дослідження зосереджуються на виявленні цвітіння водоростей, яке може мати негативний вплив на водні екосистеми. Gao та співавтори успішно застосували CNN для виявлення цвітіння водоростей у прибережних зонах, аналізуючи дані з супутникових зображень (Gao, 2020).У сфері аналізу наземних екосистем глибоке навчання також демонструє високий потенціал. Zhang та інші використали глибокі нейронні мережі для класифікації типів наземного покриву, що дозволяє відстежувати зміни в біорізноманітті та стані екосистем (Zhang, 2021). Це сприяє більш ефективному управлінню природними ресурсами та збереженню екологічного балансу.Комбінація мультиспектральних зображень із можливостями глибокого навчання дозволяє отримувати більш детальну інформацію про стан водних та наземних екосистем. Chenta інші досліджували застосування глибоких нейронних мереж для покращення роздільної здатності супутникових зображень, що забезпечує більш точний аналіз структурних елементів екомережі (Chen, 2021, p.5899-5921).Попри значний прогрес, існують виклики, пов'язані з необхідністю великих обсягів якісних навчальних даних та забезпеченням інтерпретованості моделей ШІ. Подальші дослідження спрямовані на розвиток методів трансферного навчання та удосконалення алгоритмів, щоб підвищити точність і надійність моделей для практичного застосування в екологічному моніторингу.

Мета дослідження. Метою даного дослідження є створення та впровадження ефективного нейромережевого підходу на основі моделі U-Net для детектування структурних елементів водного середовища України шляхом аналізу мультиспектральних зображень із супутника

Sentinel-2. Це передбачає підвищення точності та швидкості виявлення змін у водних масах та елементах екомережі, що сприятиме оперативному реагуванню на екологічні виклики.

Основна частина. В дослідженні використовувалися супутникові знімки Sentinel-2, надані Європейським космічним агентством (ЄКА). Sentinel-2 надає мультиспектральні зображення високої роздільної здатності з просторовим розрізненням 10 м у видимому та ближньому інфрачервоному (NIR) діапазонах, що робить їх придатними для моніторингу водного середовища. Для тренування і валідації був обраний знімок з ідентифікатором «S2B_MSIL2A_20240501T085559_N0510_R007_T36UUA_20240501T114536.SAFE».

Для аналізу були обрані три основні спектральні канали: канал B3 (зелений, 10 м), який є чутливим до рослинності та використовується в обчисленні індексів, пов'язаних з водою; канал B8 (ближній інфрачервоний, NIR, 10 м), що допомагає виявляти водні об'єкти завдяки своїй реакції на вологість; та канал TCI (TrueColorImage), істинно-кольорове зображення для візуалізації та загального аналізу.

Зображення завантажувалися та нормалізувалися шляхом масштабування значень пікселів до діапазону [0, 1], що сприяло стабільності та ефективності навчання нейронної мережі (SanturkarS., 2018, p. 2483–2493). Це забезпечувало стандартизацію даних і покращувало збіжність алгоритму під час тренування.

Для автоматичного генерування масок водних об'єктів використовувався нормалізований різницевий водний індекс (NDWI), який ефективно виявляє водні поверхні, базуючись на різниці відбиття між зеленим та ближнім інфрачервоним каналами (DuanW., 2019). Формула NDWI має вигляд:

$$NDWI = \frac{B3 - B8}{B3 + B8},$$

де *B3*– відбиття у зеленому каналі, а *B8* – відбиття у ближньому інфрачервоному каналі.

Процес створення масок складався з кількох етапів. Спочатку завантажувалися спектральні канали B3 та B8 і перетворювалися до відповідного числового формату. Якщо роздільна здатність каналів не співпадала, здійснювалося масштабування одного з них для вирівнювання розмірів зображень. Це було важливим для коректного обчислення NDWI та забезпечення точності подальшого аналізу.Далі для кожного пікселя зображення розраховувався NDWI. Щоб уникнути ділення на нуль та некоректних значень, застосовувалися відповідні методи обробки даних, такі як додавання малих констант або фільтрація пікселів з нульовою сумою каналів. Після обчислення NDWI встановлювалося порогове значення -0.015; пікселі з NDWI вище цього порогу вважалися водними

Рис. 1. Зображення, використане в навчанні моделі

об'єктами. Це значення було визначене емпірично на основі аналізу гістограм NDWI та може бути адаптоване залежно від специфічних умов зйомки та характеристик регіону дослідження.

На основі порогового значення генерувалася бінарна маска, де водні пікселі позначалися як 1, а всі інші – як 0. Маска зберігалася для подальшого використання в процесі навчання моделі. Це дозволило автоматизувати процес створення навчальних даних та зменшити потребу в ручній розмітці, що є трудомістким завданням.

Через високу роздільну здатність супутникових зображень Sentinel-2 було прийнято рішення розбивати їх на менші фрагменти (патчі) розміром 512×512 пікселів. Це дозволило зменшити обчислювальні вимоги, оскільки робота з меншими зображеннями знижує потреби в оперативній пам'яті та підвищує швидкість обробки даних. Крім того, такий підхід покращив навчання моделі, оскільки вона ефективніше навчається на локальних особливостях зображення, що сприяє точнішій сегментації водних об'єктів.Розбиття здійснювалося рівномірно по всьому зображенню, забезпечуючи відповідність між патчами зображень та відповідними масками. Патчі зберігалися у зручному для подальшої обробки форматі, що дозволяло швидко завантажувати їх під час тренування та забезпечувало ефективний потік даних.

Для тренування моделі були відібрані найбільш інформативні патчі, які містили значну кількість водних пікселів. Такий підхід забезпечив збалансованість даних: модель отримувала достатню кількість прикладів з водними об'єктами, що покращувало її здатність до їх коректного виявлення. Це також підвищило ефективність тренування, оскільки фокусування на інформативних патчах зменшувало час навчання без компромісу щодо якості моделі. Валідаційний набір даних формувався з патчів, які не використовувалися під час тренування. Це забезпечувало об'єктивну оцінку продуктивності моделі та її здатності до узагальнення на невідомих даних. Такий підхід є стандартною практикою в машинному навчанні і дозволяє уникнути перенавчання моделі на тренувальних даних.

Архітектура нейронної мережі. Для задачі сегментації водних об'єктів була обрана модель U-Net, яка є одним із провідних підходів для семантичної сегментації зображень, особливо в галузі дистанційного зондування та медичної візуалізації (ZhouZ., 2019, p. 1856-1867). U-Net поєднує в собі симетричну архітектуру, що складається з шляху стиснення (енкодера) та шляху розширення (декодера). Така структура дозволяє моделі враховувати як локальні, так і глобальні контексти зображення, що є критичним для точного виявлення водних об'єктів.

Енкодер моделі складається з послідовності згорткових шарів, які поступово зменшують просторові розміри вхідного зображення та збільшують кількість каналів. Це дозволяє моделі захоплювати високорівневі особливості та контекстну інформацію. Зокрема, енкодер містить чотири блоки, кожен з яких має два послідовних згорткових шари з активаційною функцією та нормалізацією. Після кожного блоку застосовується операція максимального пулінгу,

Рис. 2. Розмітка цільових даних знімку

яка зменшує розміри зображення та підвищує ефективність навчання.

Перший блок енкодера приймає вхідне зображення розміром *H×W×3* (три колірні канали) і містить згорткові шари з кількістю каналів *64*. Кожен наступний блок подвоює кількість каналів (*128*, *256*, *512*), що дозволяє моделі вловлювати більш складні та абстрактні особливості зображення. Максимальний пулінг з розміром ядра *2×2* зменшує просторові розміри, забезпечуючи збереження важливої інформації.У центрі моделі знаходиться бот-тлнек, який відповідає за максимально глибоке представлення вхідного зображення. Він складається з двох згорткових шарів з кількістю каналів *1024*, активаційними функціями та нормалізацією. Це дозволяє моделі навчитися розпізнавати найбільш високорівневі особливості, що є важливими для сегментації.

Декодер моделі відповідає за відновлення просторових розмірів зображення та зменшення кількості каналів. Він також складається з чотирьох блоків, симетричних до енкодера. Кожен блок декодера включає транспоновану згортку (апсемплінг), яка збільшує розміри зображення, та два згорткових шари з активацією та нормалізацією. Кількість каналів на кожному рівні декодера зменшується відповідно до енкодера (*512*, *256*, *128*, *64*).

Ключовою особливістю U-Net є використання пропускних з'єднань між відповідними рівнями енкодера та декодера. Це дозволяє моделі поєднувати контекстну інформацію з глибоких шарів з просторовою інформацією високої роздільної здатності з нижчих рівнів. Математично це можна виразити формулою:

$$y_i = Conv\left(Up\left(y_{i+1}\right) \oplus x_i\right), \qquad (1)$$

де x_i – вихід енкодера на рівні i, y_i – вихід декодера на рівні i, $Up\left(y_{i+1}\right)$ – результат апсемплінгу з наступного рівня декодера, $\oplus$ – операція конкатенації по каналах, а $Conv$ – згортковий блок.

Для активації після кожного згорткового шару використовується функція ReLU (RectifiedLinearUnit), яка вводить нелінійність та запобігає проблемам зникаючих градієнтів (RamachandranP., 2018). Нормалізація після згорткових шарів стабілізує процес навчання та сприяє швидшій збіжності моделі (IoffeS., 2015, p. 448-456). Ці елементи критично важливі для ефективного навчання глибоких нейронних мереж.

Вихідний шар моделі складається зі згорткового шару з фільтром розміру *1×1*та кількістю каналів *1*, що підходить для задачі бінарної сегментації. На цьому шарі використовується сигмоїдна активаційна функція, яка переводить

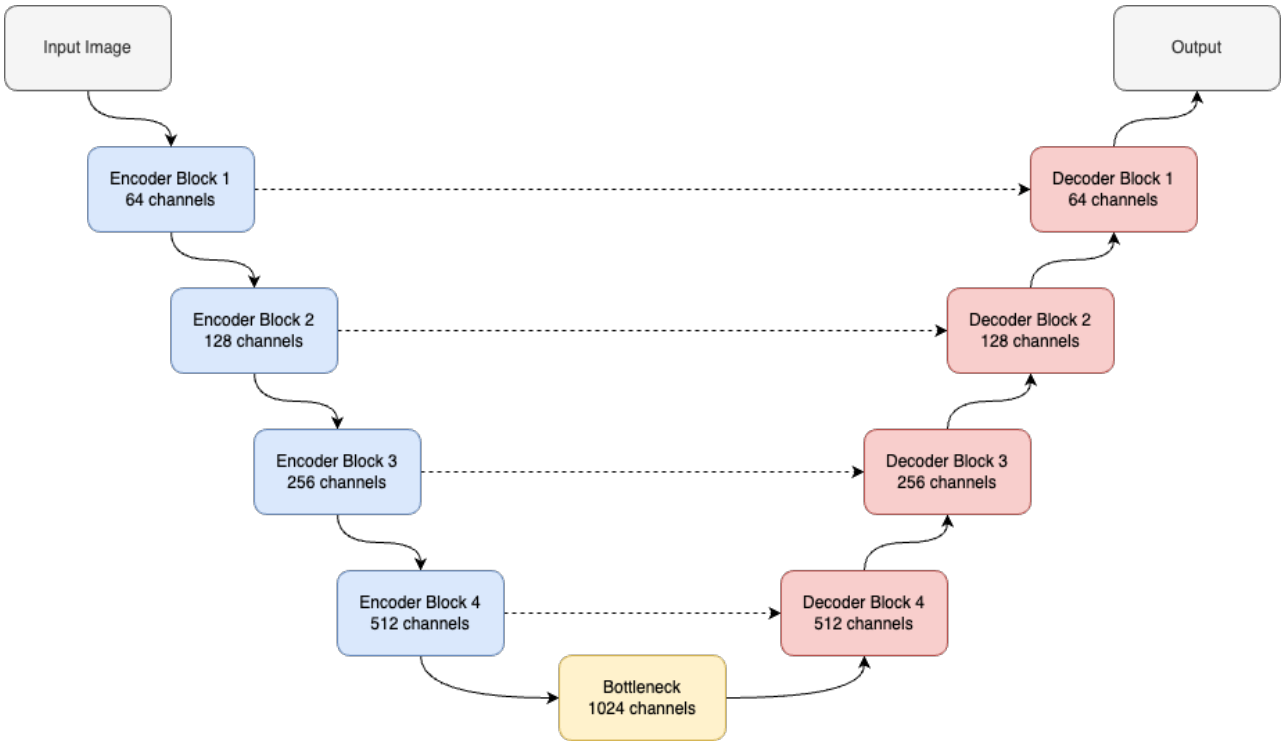


Рис. 3. Архітектура моделі U-net

вихідні значення в діапазон [0, 1], інтерпретуючи їх як ймовірність належності пікселя до класу «вода».

Параметри моделі були обрані з метою забезпечення балансу між точністю та обчислювальною ефективністю. Розмір фільтрів встановлено на *3×3*для всіх згорткових шарів, що дозволяє моделі вловлювати локальні патерни та текстури. Кількість фільтрів подвоюється на кожному рівні енкодера, досягаючи максимуму в боттлнеку, і відповідно зменшується в декодері. Використання паддінгу гарантує збереження розмірів зображення після згортки, а страйд встановлений на *1* для згорткових шарів та *2* для операцій максимального пулінгу та апсемплінгу.

Операція згортки для вхідного тензора *I* та ядра *K* визначається як:

$$(I * K)(i,j) = \sum_{m=-k}^{k} \sum_{n=-k}^{k} I(i+m, j+n) \cdot K(m,n), \quad (2)$$

де k – радіус ядра згортки. Транспонована згортка (апсемплінг) використовується для збільшення просторових розмірів вхідного тензора, що дозволяє відновити роздільну здатність зображення:

$$Up(I)(i,j) = \sum_{m=-k}^{k} \sum_{n=-k}^{k} I\left(\frac{i-m}{s}, \frac{j-n}{s}\right) \cdot K(m,n), \quad (3)$$

де s – страйд апсемплінгу.

Обґрунтування вибору архітектури полягає в її здатності ефективно вирішувати задачі семантичної сегментації з високою точністю навіть при обмеженій кількості навчальних даних (ZhouZ., 2019, p. 1856-1867). Пропускні з'єднання дозволяють моделі поєднувати високорівневі контекстуальні ознаки з низькорівневими просторовими деталями, що є особливо важливим для точного виявлення водних об'єктів різних розмірів та форм. Симетрична структура U-Net полегшує процес навчання та забезпечує збалансоване представлення характеристик на різних масштабах, що сприяє покращенню якості сегментації.Використання цієї архітектури в нашому дослідженні дозволило досягти високої точності виявлення водних об'єктів на супутникових знімках. Модель показала високу ефективність та гнучкість, що підтверджується отриманими результатами та метриками оцінки. Це свідчить про доцільність застосування U-Net у задачах моніторингу навколишнього середовища та підкреслює її потенціал для подальшого розвитку та вдосконалення.

Навчання та оцінка моделі. Після підготовки даних та побудови архітектури моделі було проведено навчання глибокої нейронної мережі для задачі семантичної сегментації водних об'єктів на супутникових знімках. Основні аспекти процесу навчання включали вибір функції втрат, оптимізатора, гіперпараметрів та стратегії навчання.

Для задачі бінарної сегментації використовувалася комбінована функція втрат, яка поєднує бінарну крос-ентропію (BinaryCross-Entropy, BCE) та коефіцієнт Дайса (DiceLoss).Формула бінарної крос-ентропії:

$$L_{BCE} = -\frac{1}{N} \sum_{i=1}^{N} \left[y_i \log(p_i) + (1-y_i) \log(1-p_i) \right]. \quad (4)$$

Формула коефіцієнта Дайса має вигляд:

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N} y_i p_i + \epsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} p_i + \epsilon}, \quad)$$

де y_i – істинне значення пікселя (0 або 1), p_i – передбачена ймовірність належності пікселя до класу «вода», ϵ – мале число для уникнення ділення на нуль.

Комбінована функція втрат визначається як:

$$L_{total} = \alpha \cdot L_{BCE} + \beta \cdot L_{Dice}, \quad (6)$$

де α та β – вагові коефіцієнти, що визначають внесок кожної складової функції втрат.

Комбінування цих двох функцій втрат дозволяє враховувати як піксельну точність (через BCE), так і загальну форму та перекриття передбаченої та істинної масок (через DiceLoss) (YeungM., 2021; LiX., 2018, p. 2663-2674).

Для оптимізації ваг моделі використовувався метод адаптивної оптимізації зі швидкістю навчання $\eta = 1 \times 10^{-4}$, що забезпечує стабільну збіжністьта поступове оновлення ваг(LuoL., 2019). Розмір батчу був встановлений на рівні 16, що дозволило збалансувати використання пам'яті та статистичну стабільність градієнтів. Навчання моделі проводилося протягом 50 епох, після чого спостерігалася стабілізація функції втрат на валідаційному наборі даних.

Стратегія навчання передбачала розбиття даних на навчальний та валідаційний набори в пропорції 80% та 20%, що відповідає рекомендаціям сучасних досліджень (BudaM., 2018, p. 249-259). Патчі даних перемішувалися перед кожною епохою, щоб уникнути впливу порядку даних на процес навчання.

Для запобігання перенавчанню та покращення збіжності використовувалися наступні методи:

- Рання зупинка (EarlyStopping): навчання припинялося, якщо функція втрат на валідаційному наборі переставала зменшуватися протягом 5 епох (PrecheltL., 2012, p. 53-67).

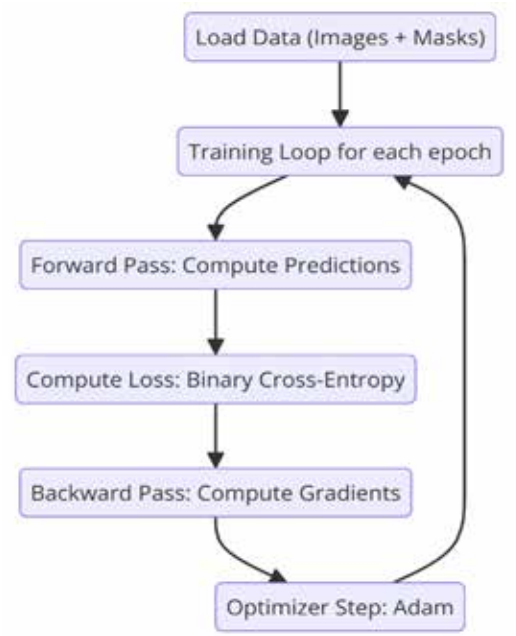


Рис. 4. Схема навчання моделі

- Зменшення швидкості навчання на плато: Швидкість навчання зменшувалася на фактор 0.1, якщо функція втрат на валідаційному наборі не покращувалася протягом 3 епох (SmithL.N., 2018).

Для оцінки продуктивності моделі були обрані наступні метрики:

1. F1-міра (F1 Score): враховує як точність, так і повноту, що особливо важливо в умовах незбалансованих класів, характерних для даної задачі (YangP., 2020).

$$F1\,Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}, \quad (7)$$

2. Точність (Precision): відсоток правильно передбачених позитивних прикладів серед всіх передбачених позитивних.

$$Precision = \frac{TP}{TP + FP}, \quad (8)$$

3. Повнота (Recall): відсоток правильно передбачених позитивних прикладів серед всіх реальних позитивних.

$$Recall = \frac{TP}{TP + FN}, \quad (9)$$

4. Індекс Жаккара (IntersectionoverUnion, IoU): відношення площі перетину передбаченої та істинної масок до площі їх об'єднання.Широко використовується в задачах сегментації та дає інтуїтивне розуміння перекриття між передбаченою та істинною масками (RahmanM.A., 2016, p. 234-244).

$$IoU = \frac{TP}{TP + FP + FN}, \quad (10)$$

деTP (TruePositives) – кількість правильно передбачених пікселів води,FP (FalsePositives) – кількість пікселів, неправильно передбачених як вода, FN (FalseNegatives) – кількість пікселів води, які не були виявлені.

Оцінка моделі проводилася на валідаційному наборі даних після кожної епохи навчання. Передбачені ймовірності переводилися в бінарні значення з використанням порогу 0.5, що є стандартною практикою в задачах бінарної класифікації (LiuX., 2019).На основі передбачених та істинних масок обчислювалися зазначені вище метрики.Результати метрик використовувалися для моніторингу продуктивності моделі та прийняття рішень щодо коригування гіперпараметрів або стратегії навчання.

Після завершення навчання модель показала високі результати на валідаційному наборі даних. Отримані метрики:

Метрика	Значення
F1-міра (F1 Score)	0.8897
Точність (Precision)	0.8721
Повнота (Recall)	0.9080
Індекс Жаккара (IoU)	0.8013

Ці показники свідчать про те, що модель досягла високої точності та повноти у виявленні водних об'єктів. Високе значення F1-міри вказує на збалансованість між точністю та повнотою, а значення IoU понад 0.8 свідчить про значне перекриття між передбаченими та істинними масками.Модель демонструвала стійкість до перенавчання завдяки застосуванню методів регуляризації та контролю за функцією втрат на валідаційному наборі. Використання комбінованої функції втрат сприяло більш точному виявленню як дрібних деталей, так і загальних контурів водних об'єктів.

Застосування моделі та її переваги. Після успішного навчання та оцінки модель була протестована на нових супутникових знімках для перевірки її здатності до узагальнення та практичного застосування. Для ефективного прогнозування на повнорозмірних зображеннях використовувався метод розбиття зображення на менші фрагменти (патчі). Це було необхідно, оскільки безпосередня обробка великих зображень є обчислювально складною та вимагає значних ресурсів пам'яті.Вихідне зображення розміром $H \times W$ пікселів ділилося на патчі розміром $P \times P$ пікселів із певним кроком S. Використання перекриття між патчами ($S<P$) дозволяло зменшити артефакти на межах патчів і покращити якість сегментації. Кожен патч передавався через модель для отримання

передбаченої маски. Оскільки модель була навчена на патчах того ж розміру, прогнозування було ефективним та точним.Після обробки всі передбачені маски патчів об'єднувалися в одне повнорозмірне зображення з урахуванням можливих перекриттів. При перекритті пікселів використовувалося середнє або максимальне значення для остаточного рішення, що забезпечувало цілісність передбаченої маски та підвищувало загальну якість сегментації.Приклад передбачень моделі на повнорозмірних знімках Sentinel-2 (ідентифікаторзнімку-«S2A_MSIL2A_20240430T083601_N0510_R064_T36UXU_20240430T120153.SAFE»):

Обраний підхід має кілька важливих переваг. По-перше, він забезпечує обчислювальну ефективність. Розбиття на патчі зменшує вимоги до оперативної пам'яті, дозволяючи виконувати прогнозування на звичайному обладнанні без необхідності використання високопродуктивних обчислювальних систем. Крім того, можлива паралельна обробка патчів, що прискорює процес прогнозування.

По-друге, метод характеризується гнучкістю та масштабованістю. Модель легко застосовується до зображень будь-якого розміру, оскільки процес розбиття та збирання не залежить від розмірів вхідного зображення. Це особливо важливо для обробки зображень з високою роздільною здатністю, що необхідно для детального аналізу територій.

По-третє, відбувається покращення якості прогнозування. Обробка патчів дозволяє моделі фокусуватися на локальних особливостях зображення, що сприяє більш точному виявленню дрібних водних об'єктів. Перекриття патчів і об'єднання прогнозів зменшують артефакти на межах, підвищуючи загальну якість сегментації.

Потенційні покращення. Для підвищення ефективності моделі можна розглянути альтернативні підходи до створення масок водних об'єктів. Замість використання фіксованого порогу на основі NDWI, доцільно застосувати методи кластеризації або глибокого навчання для автоматичного виявлення води з більшою точністю (LiB., 2020). Об'єднання декількох спектральних індексів, таких як MNDWI чи AWEI, також може покращити результати в умовах змінної освітленості та атмосферних впливів (DuanW., 2019).

Експериментування з іншими архітектурами нейронних мереж, такими як U-Net++ (ZhouZ., 2018), Attention U-Net(OktayO., 2018) або DeepLabv3+ (ChenL.C., 2018), може сприяти покращенню сегментації завдяки кращому врахуванню багаторівневих особливостей та механізмів уваги. Впровадження цих підходів дозволить моделі збільшити точність ідентифікування складних структур водних об'єктів та підвищити загальну продуктивність системи.

Висновок. У даній роботі було розроблено ефективну модель для сегментації водних об'єктів на супутникових знімках Sentinel-2 з використанням глибокої нейронної мережі архітектури U-Net. Процес включав підготовку даних, де автоматично генерувалися маски водних об'єктів на основі індексу NDWI, та детальне навчання моделі з комбінованою функцією втрат, що поєднує бінарну крос-ентропію

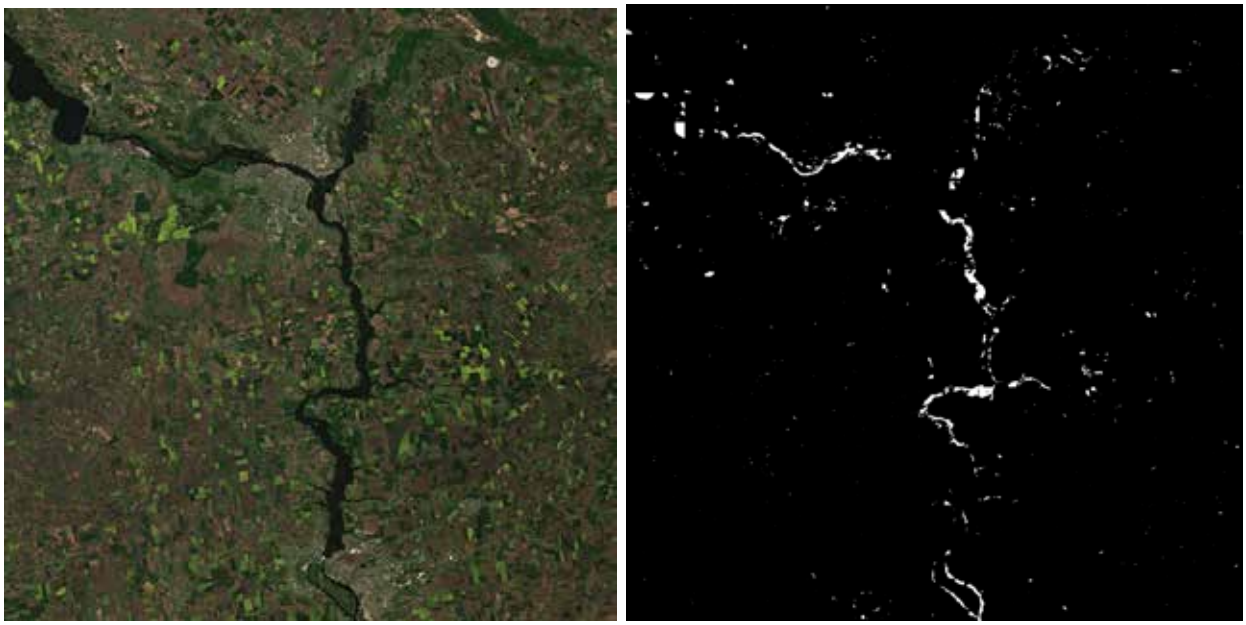


Рис. 5. Результат передбачення моделі

та коефіцієнт Дайса. Модель досягла високих показників точності, підтверджених метриками F1-міри, точності, повноти та індексу Жаккара, що свідчить про її здатність до точного виявлення водних об'єктів різних розмірів і форм. Перевагами використаних підходів є поєднання глибинного навчання з попередньою підготовкою даних, що дозволило автоматизувати процес виявлення водних об'єктів та забезпечити високу точність сегментації. Модель демонструє гнучкість, масштабованість та обчислювальну ефективність, що робить її придатною для практичного застосування в моніторингу водних ресурсів і екологічних дослідженнях. Майбутні напрямки дослідження направлені на експерименти з іншими архітектурами нейронних мереж, використання альтернативних методів створення масок та врахування просторово-часової інформації, що дасть змогу підвищити ефективність і універсальність моделі.

ЛІТЕРАТУРА:

1. IPCC. *Climate Change 2021: The Physical Science Basis*. URL: https://www.ipcc.ch/report/ar6/wg1/
2. United Nations Environment Programme (UNEP). *Environmental Issues Emerging from Conflict and Disaster*. 2021. URL: https://www.unep.org/explore-topics/disasters-conflicts
3. Li W., Fu H., Yu L., Cracknell A. Deep learning based oil spill detection from polarimetric SAR data. *Remote Sensing of Environment*, 2019. 233, 111368. URL: https://www.sciencedirect.com/science/article/pii/S0034425719304563
4. Ma L., Liu Y., Zhang X., Ye Y., Yin G., Johnson B. A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2019. 152, 166–177. URL: https://www.sciencedirect.com/science/article/pii/S0924271618303333
5. Zhang C., Liu F., & Li W. A review of remote sensing image classification techniques: the role of spatio-contextual information. *European Journal of Remote Sensing*, (2020). 53(1), 176–197. URL: https://www.tandfonline.com/doi/full/10.1080/22797254.2020.1727854
6. Li W., Fu H., Yu L., Cracknell A. Deep learning based oil spill detection from polarimetric SAR data. *Remote Sensing of Environment*, 2019. 233, 111368. https://doi.org/10.1016/j.rse.2019.111368
7. Gao B., Huang J., Li J., Chen Z. Detecting algal blooms in coastal waters using deep learning and satellite imagery. *Remote Sensing*, 2020. 12(7), 1135. https://doi.org/10.3390/rs12071135
8. Zhang C., Li W., Travis D. Gaps fill of land cover classification with deep learning. *Remote Sensing of Environment*, 2021. 267, 112716. https://doi.org/10.1016/j.rse.2021.112716
9. Chen Y., Jiang H., Li C., Jia X., Ghamisi P. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 2021. 59(7), 5899–5921. https://doi.org/10.1109/TGRS.2020.3028704
10. Santurkar S., Tsipras D., Ilyas A., Madry A. How Does Batch Normalization Help Optimization?*Advances in Neural Information Processing Systems*, 2018. 31, 2483–2493. URL: https://proceedings.neurips.cc/paper/2018/file/905056c1f1abf51a8521faa9756a31d8-Paper.pdf
11. Duan W., He B., Nover D., Yang G., Chen W., Meng H., Liu C. Water bodies' mapping from Sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the SWIR band. *Remote Sensing*, 2019. 11(23), 2779. https://doi.org/10.3390/rs11232779
12. Zhou Z., Siddiquee M.M.R., Tajbakhsh N., Liang J. UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 2019. 39(6), 1856–1867. https://doi.org/10.1109/TMI.2019.2959609
13. Ramachandran P., Zoph B., Le Q. V. Searching for activation functions. *arXiv preprint arXiv:1710.05941*. 2018. URL: https://arxiv.org/abs/1710.05941
14. Ioffe S., Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *Proceedings of the 32nd International Conference on Machine Learning*, 2015. 37, 448–456. http://proceedings.mlr.press/v37/ioffe15.html
15. Yeung M., Sala E., Schönlieb C. B., Rundo, L. Unified focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Computerized Medical Imaging and Graphics*, 2021. 89, 101885. https://doi.org/10.1016/j.compmedimag.2021.101885
16. Li X., Chen H., Qi X., Dou Q., Fu C. W., Heng P. A. H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes. *IEEE Transactions on Medical Imaging*, 2018. 37(12), 2663–2674. https://doi.org/10.1109/TMI.2018.2845918
17. Luo L., Xiong Y., Liu Y., Sun X. Adaptive gradient methods with dynamic bound of learning rate. *Proceedings of the 7th International Conference on Learning Representations (ICLR)*. 2019. URL: https://arxiv.org/abs/1902.09843

18. Buda M., Maki A., Mazurowski M. A. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 2018. 106, 249–259. https://doi.org/10.1016/j.neunet.2018.07.011

19. Prechelt L. Early stopping – but when?. In *Neural Networks: Tricks of the Trade* (2012). (pp. 53–67). Springer. https://doi.org/10.1007/978-3-642-35289-8_5

20. Smith L. N. A disciplined approach to neural network hyper-parameters: Part 1 – learning rate, batch size, momentum, and weight decay. *arXiv preprint arXiv:1803.09820*. 2018. URL: https://arxiv.org/abs/1803.09820

21. Yang P., Zhang L., Zhao W., Xie X. Retinal vessel segmentation using a U-Net based convolutional neural network. *Symmetry*, 2020. 12(4), 611. https://doi.org/10.3390/sym12040611

22. Rahman M. A., Wang Y. Optimizing intersection-over-union in deep neural networks for image segmentation. In *International Symposium on Visual Computing* (2016). (pp. 234–244). Springer. https://doi.org/10.1007/978-3-319-50835-1_22

23. Liu X., Song L., Liu S., Zhang Y. A review of deep-learning-based medical image segmentation methods. *Sustainability*, 2019. 11(7), 1897. https://doi.org/10.3390/su11071897

24. Li B., Liu Y., Wang Y., Li M., Liu Y. A comparison of clustering algorithms for water body extraction from remote sensing images. *Remote Sensing*, 2020. 12(10), 1590. https://doi.org/10.3390/rs12101590

25. Duan W., He B., Nover D., Yang G., Chen W., Meng H., Liu C. Water bodies' mapping from Sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the SWIR band. *Remote Sensing*, 2019. 11(23), 2779. https://doi.org/10.3390/rs11232779

26. Zhou Z., Siddiquee M.M.R., Tajbakhsh N., Liang J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* 2018. (pp. 3–11). Springer. https://doi.org/10.1007/978-3-030-00889-5_1

27. Oktay O., Schlemper J., Le Folgoc L., Lee M., Heinrich M., Misawa K., … Rueckert D. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv preprint arXiv:1804.03999*. 2018. URL: https://arxiv.org/abs/1804.03999

28. Chen L. C., Zhu Y., Papandreou G., Schroff F., Adam H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)* 2018. (pp. 801–818). https://doi.org/10.1007/978-3-030-01234-2_49

29. Olena Dobrovolska, Knut Schmidtke, Viktoriia Hnatushenko, Svitlana Sytnyk and Iryna Dmytriieva Relationship between economic development, forest resources, and forest fires: European context. *Environmental Economics*, 2024. 15(2), 77–92. doi:10.21511/ee.15(2).2024.06.

30. Hnatushenko V. V, Mozgovoy D. K., HnatushenkoVik. V., Spirintsev V. V., Udovyk I. M. All-weather monitoring of oil and gas production areasusingsatellite data//Scientific bulletin of National Mining University. *State Higher Educational Institution "National Mining University"*, Dnipro, 2019. № 6 (158). С. 99–105.

REFERENCES:

1. IPCC. *Climate Change 2021: The Physical Science Basis*. Retrieved from: https://www.ipcc.ch/report/ar6/wg1/

2. United Nations Environment Programme (UNEP). *Environmental Issues Emerging from Conflict and Disaster*. 2021. Retrieved from: https://www.unep.org/explore-topics/disasters-conflicts

3. Li, W., Fu, H., Yu, L., & Cracknell, A. (2019). Deep learning based oil spill detection from polarimetric SAR data. *Remote Sensing of Environment*, 233, 111368. Retrieved from: https://www.sciencedirect.com/science/article/pii/S0034425719304563

4. Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152, 166–177. Retrieved from: https://www.sciencedirect.com/science/article/pii/S0924271618303333

5. Zhang, C., Liu, F., & Li, W. (2020). A review of remote sensing image classification techniques: the role of spatio-contextual information. *European Journal of Remote Sensing*, 53(1), 176–197. Retrieved from: https://www.tandfonline.com/doi/full/10.1080/22797254.2020.1727854

6. Li, W., Fu, H., Yu, L., & Cracknell, A. (2019). Deep learning based oil spill detection from polarimetric SAR data. *Remote Sensing of Environment*, 233, 111368. https://doi.org/10.1016/j.rse.2019.111368

7. Gao, B., Huang, J., Li, J., & Chen, Z. (2020). Detecting algal blooms in coastal waters using deep learning and satellite imagery. *Remote Sensing*, 12(7), 1135. https://doi.org/10.3390/rs12071135

8. Zhang, C., Li, W., & Travis, D. (2021). Gaps fill of land cover classification with deep learning. *Remote Sensing of Environment*, 267, 112716. https://doi.org/10.1016/j.rse.2021.112716

9. Chen, Y., Jiang, H., Li, C., Jia, X., &Ghamisi, P. (2021). Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 59(7), 5899–5921. https://doi.org/10.1109/TGRS.2020.3028704

10. Santurkar, S., Tsipras, D., Ilyas, A., & Madry, A. (2018). How Does Batch Normalization Help Optimization?*Advances in Neural Information Processing Systems*, 31, 2483–2493. https://proceedings.neurips.cc/paper/2018/file/905056c1f1abf51a8521faa9756a31d8-Paper.pdf

11. Duan, W., He, B., Nover, D., Yang, G., Chen, W., Meng, H., & Liu, C. (2019). Water bodies' mapping from Sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the SWIR band. *Remote Sensing*, 11(23), 2779. https://doi.org/10.3390/rs11232779

12. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., & Liang, J. (2019). UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 39(6), 1856–1867. https://doi.org/10.1109/TMI.2019.2959609

13. Ramachandran, P., Zoph, B., & Le, Q.V. (2018). Searching for activation functions. *arXiv preprint arXiv:1710.05941*. Retrieved from: https://arxiv.org/abs/1710.05941

14. Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *Proceedings of the 32nd International Conference on Machine Learning*, 37, 448–456. http://proceedings.mlr.press/v37/ioffe15.html

15. Yeung, M., Sala, E., Schönlieb, C.B., & Rundo, L. (2021). Unified focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Computerized Medical Imaging and Graphics*, 89, 101885. https://doi.org/10.1016/j.compmedimag.2021.101885

16. Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.W., & Heng, P.A. (2018). H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes. *IEEE Transactions on Medical Imaging*, 37(12), 2663–2674. https://doi.org/10.1109/TMI.2018.2845918

17. Luo, L., Xiong, Y., Liu, Y., & Sun, X. (2019). Adaptive gradient methods with dynamic bound of learning rate. *Proceedings of the 7th International Conference on Learning Representations (ICLR)*. Retrieved from: https://arxiv.org/abs/1902.09843

18. Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106, 249–259. https://doi.org/10.1016/j.neunet.2018.07.011

19. Prechelt, L. (2012). Early stopping – but when?. In *Neural Networks: Tricks of the Trade* (pp. 5–-67). Springer. https://doi.org/10.1007/978-3-642-35289-8_5

20. Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: Part 1 – learning rate, batch size, momentum, and weight decay. *arXiv preprint arXiv:1803.09820*. Retrieved from: https://arxiv.org/abs/1803.09820

21. Yang, P., Zhang, L., Zhao, W., & Xie, X. (2020). Retinal vessel segmentation using a U-Net based convolutional neural network. *Symmetry*, 12(4), 611. https://doi.org/10.3390/sym12040611

22. Rahman, M. A., & Wang, Y. (2016). Optimizing intersection-over-union in deep neural networks for image segmentation. In *International Symposium on Visual Computing* (pp. 234–244). Springer. https://doi.org/10.1007/978-3-319-50835-1_22

23. Liu, X., Song, L., Liu, S., & Zhang, Y. (2019). A review of deep-learning-based medical image segmentation methods. *Sustainability*, 11(7), 1897. https://doi.org/10.3390/su11071897

24. Li, B., Liu, Y., Wang, Y., Li, M., &Liu, Y. (2020). A comparison of clustering algorithms for water body extraction from remote sensing images. *Remote Sensing*, 12(10), 1590. https://doi.org/10.3390/rs12101590

25. Duan, W., He, B., Nover, D., Yang, G., Chen, W., Meng, H., & Liu, C. (2019). Water bodies' mapping from Sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the SWIR band. *Remote Sensing*, 11(23), 2779. https://doi.org/10.3390/rs11232779

26. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., & Liang, J. (2018). UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (pp. 3-11). Springer. https://doi.org/10.1007/978-3-030-00889-5_1

27. Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M., Heinrich, M., Misawa, K., … & Rueckert, D. (2018). Attention U-Net: Learning Where to Look for the Pancreas. *arXiv preprint arXiv:1804.03999*. Retrieved from: https://arxiv.org/abs/1804.03999

28. Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 801–818). https://doi.org/10.1007/978-3-030-01234-2_49

29. Olena Dobrovolska, Knut Schmidtke, Viktoriia Hnatushenko, Svitlana Sytnyk & Iryna Dmytriieva (2024). Relationship between economic development, forest resources, and forest fires: European context. *Environmental Economics*, 15(2), 77–92. doi:10.21511/ee.15(2).2024.06.

30. Hnatushenko, V. V, Mozgovoy, D. K., Hnatushenko,V. V., Spirintsev, V. V., Udovyk, I. M. (2019). All-weather monitoring of oil and gas production areasusingsatellite data//Scientific bulletin of National Mining University. *State Higher Educational Institution "National Mining University"*, Dnipro, № 6 (158). С. 99–105.

УДК 004.8
DOI https://doi.org/10.32782/IT/2024-4-17

Марина МОЛЧАНОВА
викладач кафедри комп'ютерних наук, Хмельницький національний університет, вул. Інститутська, 11, Хмельницький, Україна, 29016
ORCID: *0000-0001-9810-936X*
Scopus Author ID: *57456998700*

Бібліографічний опис статті: Молчанова, М. (2024). Нейромережеве визначення цілей пропаганди у текстовому контенті з візуальною аналітикою результатів. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 144–150, doi: https://doi.org/10.32782/IT/2024-4-17

НЕЙРОМЕРЕЖЕВЕ ВИЗНАЧЕННЯ ЦІЛЕЙ ПРОПАГАНДИ У ТЕКСТОВОМУ КОНТЕНТІ З ВІЗУАЛЬНОЮ АНАЛІТИКОЮ РЕЗУЛЬТАТІВ

Мета робити *полягає у створенні методу нейромережевого визначення цілей пропаганди у текстовому контенті з візуальною аналітикою результатів, який повинен включати розширення множини цілей пропаганди та використання контекстних вікон для аналізу взаємозв'язків між пропагандистськими прийомами та цілями.*

Методологія. *Розроблений метод передбачає аналіз текстового контенту за допомогою нейромереж, що дозволяє виявляти цілі та прийоми пропаганди. Використовуються контекстні вікна для оцінки взаємозв'язків між прийомами та цілями пропаганди. Результати представляються у вигляді візуалізації семантичної моделі пропаганди, що включає оцінку семантичної значущості та важливості взаємозв'язків.*

Наукова новизна. *Основною відмінністю запропонованого методу є розширення множини цілей пропаганди шляхом доповнення множини їх словесних варіантів. Впровадження контекстних вікон для аналізу взаємозв'язків між пропагандистськими прийомами та цілями дозволяє підвищити точність виявлення пропаганди та забезпечити наочне відображення результатів.*

Висновки. *Експериментальні результати підтвердили ефективність розробленого методу, який демонструє покращення виявлення цілей пропаганди у порівнянні з існуючими підходами. Висока кореляція з експертними оцінками підтверджує надійність і точність підходу. Візуальна аналітика забезпечує прозорість та пояснюваність результатів, що сприяє глибшому розумінню механізмів пропагандистського впливу. Метод має значний потенціал для використання в автоматичному аналізі текстового контенту та може бути адаптований для різних текстових джерел.*

Ключові слова: *пропаганда, цілі пропаганди, текстовий контент, візуальна аналітика.*

Maryna MOLCHANOVA
Teacher at the Department of Computer Science, Khmelnytskyi National University, 11 Instytutska Str., Khmelnytskyi, Ukraine, 29016, m.o.molchanova@gmail.com
ORCID: *0000-0001-9810-936X*
Scopus Author ID: *57456998700*

To cite this article: Molchanova, M. (2024). Neiromerezheve vyznachennia tsilei propahandy u tekstovomu kontenti z vizualnoiu analitykoiu rezultativ [Neural network identification of propaganda goals in text content with visual analytics of results]. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 144–150, doi: https://doi.org/10.32782/IT/2024-4-17

NEURAL NETWORK IDENTIFICATION OF PROPAGANDA GOALS IN TEXT CONTENT WITH VISUAL ANALYTICS OF RESULTS

The work aim *is to create a method for neural network identification of propaganda goals in text content with visual analytics of the results, which should include expanding the set of propaganda goals and using context windows to analyze the relationships between propaganda techniques and goals.*

Methodology. *The developed method involves analyzing text content using neural networks, which allows identifying propaganda goals and techniques. Context windows are used to assess the relationships between propaganda techniques and goals. The results are presented in the form of visualization of the semantic model of propaganda, which includes assessing the semantic significance and importance of relationships.*

Scientific novelty. *The main difference of the proposed method is the expansion of the set of propaganda goals by supplementing the set of their verbal variants. The introduction of context windows to analyze the relationships between propaganda techniques and goals allows to increase the accuracy of propaganda detection and provide a visual display of the results.*

Conclusions. *Experimental results confirmed the effectiveness of the developed method, which demonstrates improved detection of propaganda targets compared to existing approaches. High correlation with expert assessments confirms the reliability and accuracy of the approach. Visual analytics provides transparency and explainability of the results, which contributes to a deeper understanding of the mechanisms of propaganda influence. The method has significant potential for use in automatic analysis of text content and can be adapted for various text sources.*

Key words: *propaganda, propaganda goals, text content, visual analytics.*

Актуальність проблеми. У сучасному світі пропаганда є одним із найефективніших засобів маніпулювання масовою свідомістю, використовуючи текстовий контент для формування потрібних наративів. Її цілі можуть варіюватися від політичного впливу до соціальних чи економічних трансформацій. Вона спрямована на різні цілі, серед яких – окремі особи, соціальні групи, організації (Nazarov V., 2023; Молчанова М.О., 2023).

Пропаганда діє через широкий спектр прийомів, що впливають на емоційний стан, когнітивні процеси та поведінкові реакції аудиторії. Важливість вивчення цього явища обумовлена не лише необхідністю ідентифікації цих прийомів, але й розумінням того, на кого вони спрямовані та які мають наслідки, тобто виникає потреба визначати цілі пропаганди (Молчанова М.О, 2024, С. 101-116).

Аналіз останніх досліджень. На сучасному етапі завдання ідентифікації пропаганди в текстовому контенті, як і інші аспекти семантичного аналізу, зокрема визначення токсичного змісту або аналіз емоційної складової текстів (Krak I., 2024, С. 16-28), вирішуються за допомогою сучасних методів обробки природної мови (Zalutska O., 2023, С. 344–356). У контексті виявлення пропагандистського впливу важливо не лише розпізнати самі пропагандистські прийоми, але й встановити, на які цілі або аудиторії вони спрямовані. Ідентифікація цілей пропаганди має вагоме значення для розкриття механізмів маніпуляції свідомістю, що робить це питання особливо актуальним для наукових досліджень.

У статті (Ahmad P. N., 2024, С. 1–15.) розглядається проблема зростання кількості пропагандистських текстів та активностей у цифровому просторі, що стало наслідком стрімкого розвитку соціальних медіа. Для вирішення цього питання запропоновано автоматизований підхід до визначення семантичних характеристик пропаганди у текстах новин. Метод ґрунтується на використанні мета-навчання для класифікації текстів на рівні речень і включає поєднання багатозадачного навчання, моделі CRF, BiLSTM та попередньо натренованих мовних моделей. Ефективність підходу підтверджено на основі аналізу корпусу новинних текстів, де модель досягла F1-оцінки 0.61 для багатомовних даних і 0.688 для одномовних.

У статті (Rodríguez-García R., 2024) представлено дослідження визначення пропаганди в текстовому контенті. У двох із п’яти сценаріїв MTL перевершило інші методи, досягнувши F1-оцінки 0.78, що значно перевищує базовий результат трансформерів (0.68). Автори також рекомендують досліджувати вплив моральних і логічних аспектів для подальшого вдосконалення моделей.

У статті (Szwoch J., 2024, С. 4330) досліджується застосування LLM, таких як GPT-4, для автоматичного визначення та класифікації пропагандистських прийомів у текстах електронної преси. Результати демонструють, що моделі здатні досягати високої точності в окремих аспектах, наприклад, 74% точності у бінарному виявленні пропаганди в польських текстах і 69% у класифікації прийомів. Однак загальні показники F1-оцінки були нижчими за базові рівні, коливаючись у межах 20-50%. Експерименти також показали, що результати моделей є нестабільними та важко відтворюваними.

Описані підходи до визначення пропаганди із застосуванням великих мовних моделей мають низку обмежень, які впливають на їх ефективність і практичну придатність. Однією з ключових проблем є нестабільність результатів, яка ускладнює відтворення отриманих даних. Багато експериментів показали, що F1-оцінка моделей суттєво поступається базовим підходам, що свідчить про їх обмежену точність у виявленні та класифікації пропаганди. Також описані підходи не містять візуальної аналітики результатів, що не викликає високий рівень довіри до отриманих результатів

Таким чином виникає потреба в застосуванні не тільки підходу на основі класифікації текстів, а й на основі визначення цілей пропаганди у текстовому контенті з візуальною аналітикою результатів.

Метою роботи є розробка методу нейромережевого визначення цілей пропаганди у текстовому контенті з візуальною аналітикою результатів, що дозволить визначати, на кого та на що орієнтовані конкретні пропагандистські прийоми, а також надавати візуальну інтерпретацію результатів.

Виклад основного матеріалу дослідження. Використання прийомів пропаганди в текстовому контенті може бути виявлене за допомогою сучасних нейромережевих підходів (Krak I., 2024, С. 158–170), що водночас дозволяє створювати моделі, спеціалізовані на детекції конкретних пропагандистських прийомів. Запропонований метод забезпечує цілісний аналіз взаємозв'язків між прийомами пропаганди та їхніми цілями, а також спроможний узагальнювати альтернативні згадування цілей у текстах. Його унікальність полягає у поєднанні двох завдань: ідентифікації цілей пропаганди та кількісного співвіднесення цих цілей із виявленими пропагандистськими прийомами.

Запропонований метод передбачає використання 17 попередньо навчених нейромережевих моделей архітектури трансформер, кожна з яких спеціалізується на ідентифікації одного з основних пропагандистських прийомів, таких як «Appeal to fear-prejudice», «Causal Oversimplification», «Doubt», «Exaggeration», «Flag-Waving», «Labeling», «Loaded Language», «Minimisation», «Name Calling», «Repetition», «Appeal to Authority», «Black and White Fallacy», «Reductio ad Hitlerum», «Red Herring», «Slogans», «Thought Terminating Clichés» та «Whataboutism» (Yoosuf S., 2019, С. 87–91).

Навчання моделей здійснювалося на основі розмічених даних, підготовлених командою «Analysis Project» (Propaganda Analysis Project, 2024), яка провела детальний аналіз текстів, виявивши фрагменти, що містять зазначені прийоми, та анотувавши їх відповідно до типу. Для цієї мети створено корпус із 550 новинних статей (Zenodo, 2024), розмічених вручну на рівні кожного фрагмента для точного визначення застосованих прийомів пропаганди.

Метод нейромережевого визначення цілей пропаганди з візуальною аналітикою результатів базується на використанні моделей глибокого навчання та складається з кількох ключових етапів. Спочатку здійснюється розпізнавання іменованих сутностей для формування первинної множини цілей пропаганди. Далі проводиться попередня обробка тексту та розширення цієї множини шляхом додавання альтернативних словесних подань цілей. Наступним кроком є створення контекстних вікон для кожного словесного подання, з урахуванням мінімального порогу розміру вікна. На основі цих контекстних вікон виконується оцінка рівня застосування пропагандистських прийомів із використанням нейромережевих моделей. Завершальним етапом є визначення значущих зв'язків між прийомами та цілями пропаганди, враховуючи мінімальний рівень прояву пропаганди, та подальше представлення результатів у візуальній формі. Схема та кроки методу нейромережевого визначення цілей пропаганди у текстовому контенті з візуальною аналітикою результатів наведено на рисунку 1.

Запропонований метод нейромережевого визначення цілей пропаганди базується на аналізі тестового тексту, множини прийомів пропаганди та набору спеціалізованих моделей, натренованих на визначення цих прийомів.

Перший крок передбачає пошук іменованих сутностей (NER) із видаленням повторів на рівні лем, що дозволяє отримати множину унікальних цілей пропаганди.

На другому кроці здійснюється розширення цієї множини за допомогою моделі «FastText», яка виявляє семантично схожі цілі, розширюючи аналіз за межі класичних NER. Використання «FastText», донавченої на пропагандистських текстах. У результаті цього кроку формується ширша множина цілей пропаганди завдяки включенню різних варіантів їх словесних подань. Параметр мінімальної семантичної близькості встановлюється залежно від конкретної задачі та визначається експериментально. У рамках цього дослідження порогове значення не використовувалося.

Третій крок зосереджений на формуванні контекстних вікон, які представляють речення, що містять цілі пропаганди. Якщо у межах одного контекстного вікна трапляється кілька різних цілей пропаганди, то таке вікно не дублюється. Однак, якщо в контекстному вікні присутні різні словесні варіанти однієї й тої ж самої цілі пропаганди, то таке вікно буде дублюватися.

На четвертому кроці виконується аналіз контекстних вікон нейромережевими моделями для визначення рівня застосування пропагандистських прийомів. Контент вікон векторизується, а моделі оцінюють його відповідність конкретним прийомам, виявленим у тексті.

Завершальним, п'ятим кроком, є побудова множини значущих зв'язків між прийомами та цілями пропаганди. При цьому враховується порогове значення, нижче якого прийом не вважається застосованим до відповідної групи цілей.

Результатом методу є візуальна аналітика семантичної моделі пропаганди для тексту,

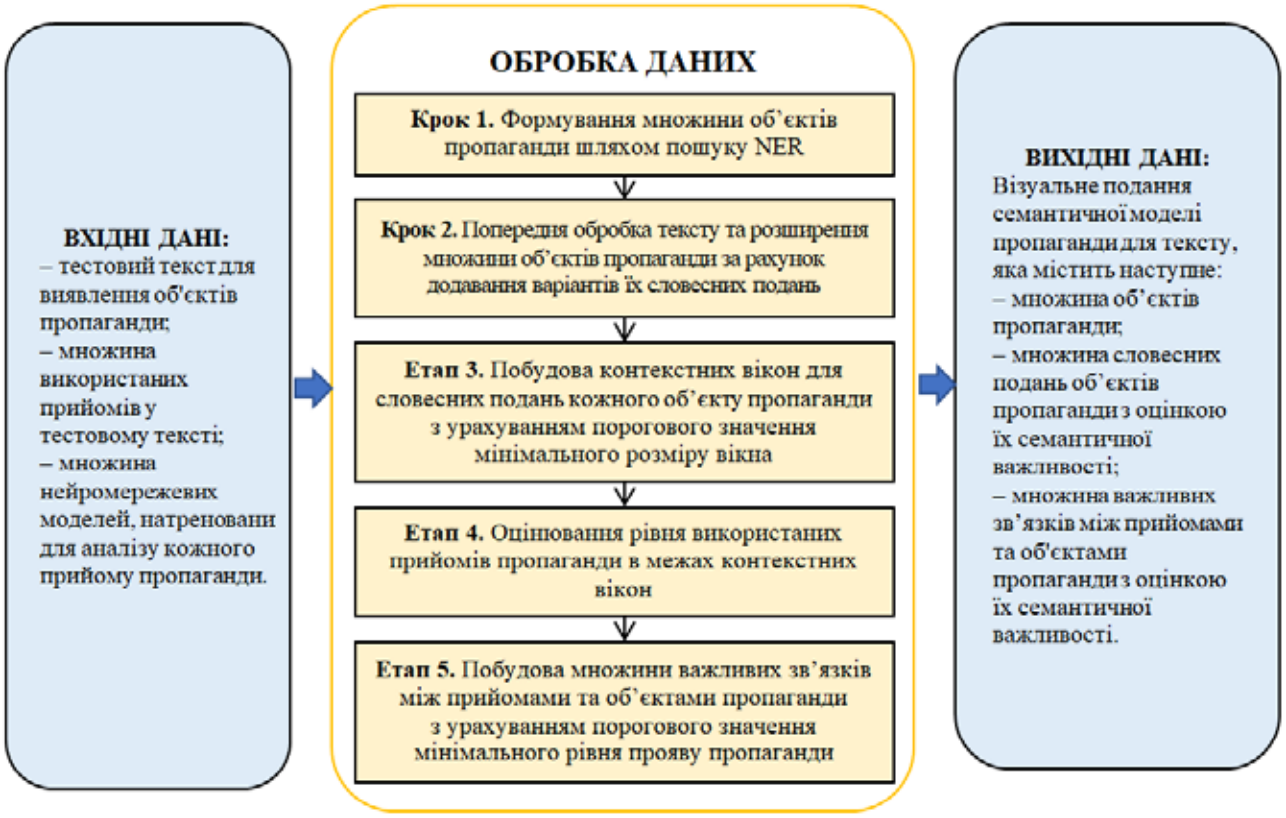


Рис. 1. Схема та кроки методу нейромережевого визначення цілей пропаганди у текстовому контенті з візуальною аналітикою

яка демонструє сукупність цілей пропаганди, їхні словесні варіанти з відповідними оцінками семантичної значущості, а також ключові зв'язки між пропагандистськими прийомами і цілями пропаганди із зазначенням рівня їхньої семантичної важливості.

У результаті було запропоновано метод, який забезпечує нейромережеве визначення цілей пропаганди у текстовому контенті, визначення їхньої належності до застосованих у тексті пропагандистських прийомів, а також розширення множини цілей пропаганди за рахунок додавання їхніх словесних варіантів. Використання контекстних вікон дозволило встановити взаємозв'язки між пропагандистськими прийомами та цілями, що сприяло підвищенню ефективності ідентифікації цілей пропаганди та забезпечило можливість візуальної аналітики результатів.

Експеримент, результати та дискусія. З метою оцінки ефективності запропонованого методу нейромережевого визначення цілей пропаганди у текстовому контенті з візуальною аналітикою було створено спеціалізоване програмне забезпечення. Розроблене програмне забезпечення дозволяло ідентифікувати цілі пропаганди та аналізувати їх у контексті застосованих прийомів пропаганди. Результати, представлені у вигляді візуальної аналітики, порівнювалися з висновками авторитетних джерел та експертів у сфері визначення і протидії пропаганді.

Для аналізу було використано пости із соціальних мереж, які попередньо опрацював «Центр стратегічних комунікацій» (https://spravdi.gov.ua/). Дані матеріали вже містили експертні оцінки та мітки, що дозволило зіставити їх із даними, отриманими завдяки запропонованому методу. Створене програмне забезпечення реалізує запропонований метод нейромережевого визначення цілей пропаганди у текстовому контенті з візуальною аналітикою. На рисунку 2 наведено покроковий опис запропонованого методу.

З огляду на вміст тестового тексту, прийоми пропаганди майже рівномірно розподілені по контекстним вікнам. Проте, для першого контекстного вікна прийом «Loaded Language» має менш виражений характер, тоді як прийом «Appeal to fear-prejudice» виявляється більш вираженою порівняно з другим контекстним вікном.

При навчанні нейромережевих моделей архітектури трансформер, заснованих на переднавчених моделях BERT, було досягнуто

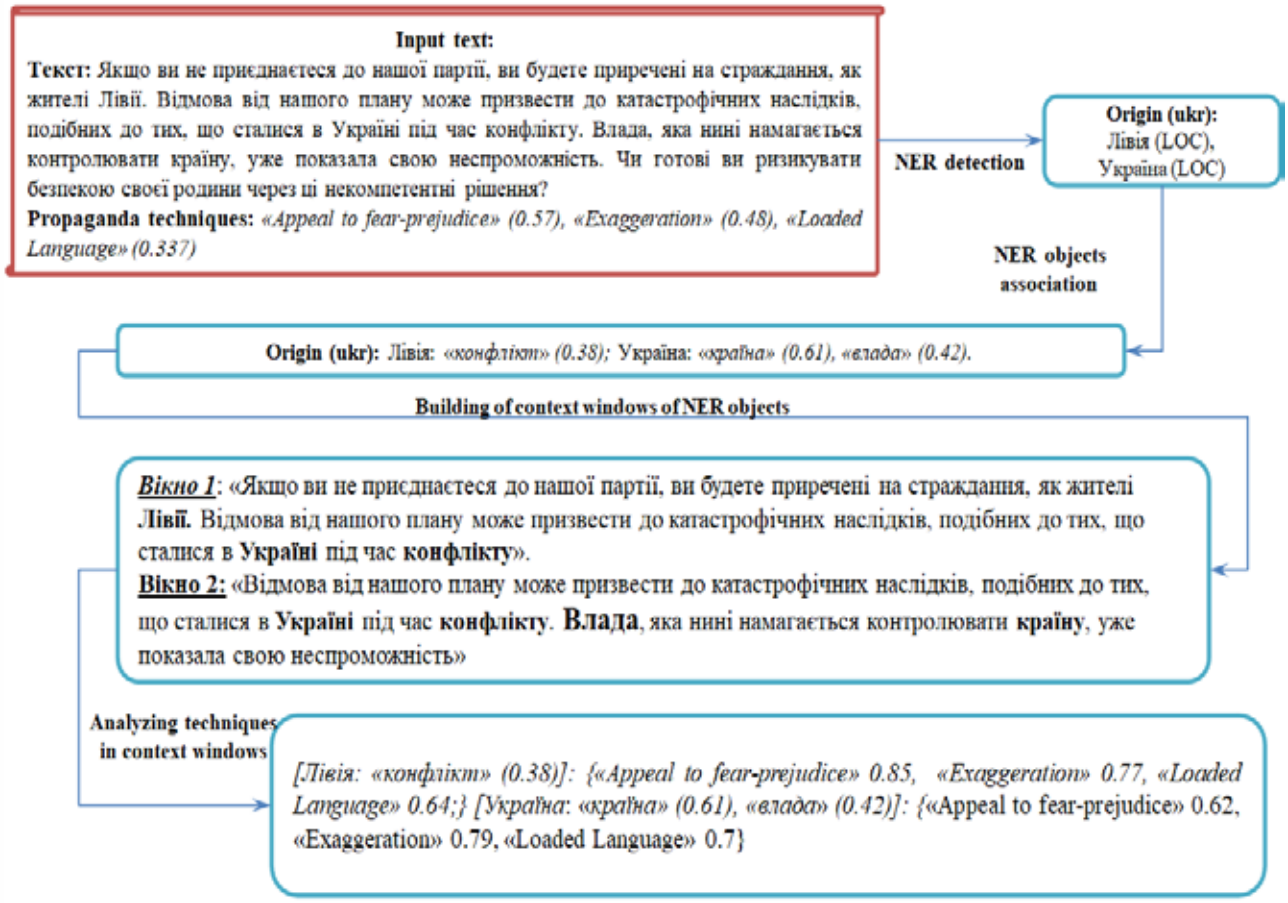


Рис. 2. Приклад перетворення даних цілей пропаганди за методом нейромережевого визначення цілей пропаганди у текстовому контенті з візуальною аналітикою

точності понад 80% для визначення прийомів пропаганди. Результати точності моделей BERT представлені в таблиці 1.

Таблиця 1

Точність визначення прийомів пропаганди моделями нейронних мереж

Прийом пропаганди	Accuracy
«Appeal to fear-prejudice»	0.87
«Causal Oversimplification»	0.82
«Doubt»	0.93
«Exaggeration»	0.8
«Flag-Waving»	0.92
«Labeling»	0.96
«Loaded Language»	0.97
«Minimisation»	0.9
«Name Calling»	0.92
«Repetition»	0.94
«Appeal to Authority»	0.89
«Black and White Fallacy»	0.91
«Reductio ad hitlerum»	0.87
«Red Herring»	0.8
«Slogans»	0.86
«Thought terminating Cliches»	0.8
«Whataboutism»	0.79

Приклад виконаного аналізу за допомогою розробленого програмного забезпечення показано на рисунку 3.

Як видно з рисунку 3, було виявлено не тільки цілі пропаганди та використані прийоми, а й визначено зв'язок між знайденими цілями та прийомами, до яких вони належать.

Отже, в результаті дослідження ефективності запропонованого методу було встановлено, що метод нейромережевого визначення цілей пропаганди у текстовому контенті з візуальною аналітикою забезпечує результати, які мають високу кореляцію з висновками експертів. Крім того, завдяки розгляду пропаганди як цілісної моделі та використанню візуальної аналітики отриманих результатів, вдалося здійснити комплексний аналіз взаємозв'язків між прийомами та цілями пропаганди, а також забезпечити узагальнення для цілей пропаганди та їх альтернативних згадок у текстах.

Висновки. Розроблено метод нейромережевого визначення цілей пропаганди у текстовому контенті з візуальною аналітикою результатів. Основною відмінністю цього методу є розширення множини цілей пропаганди через

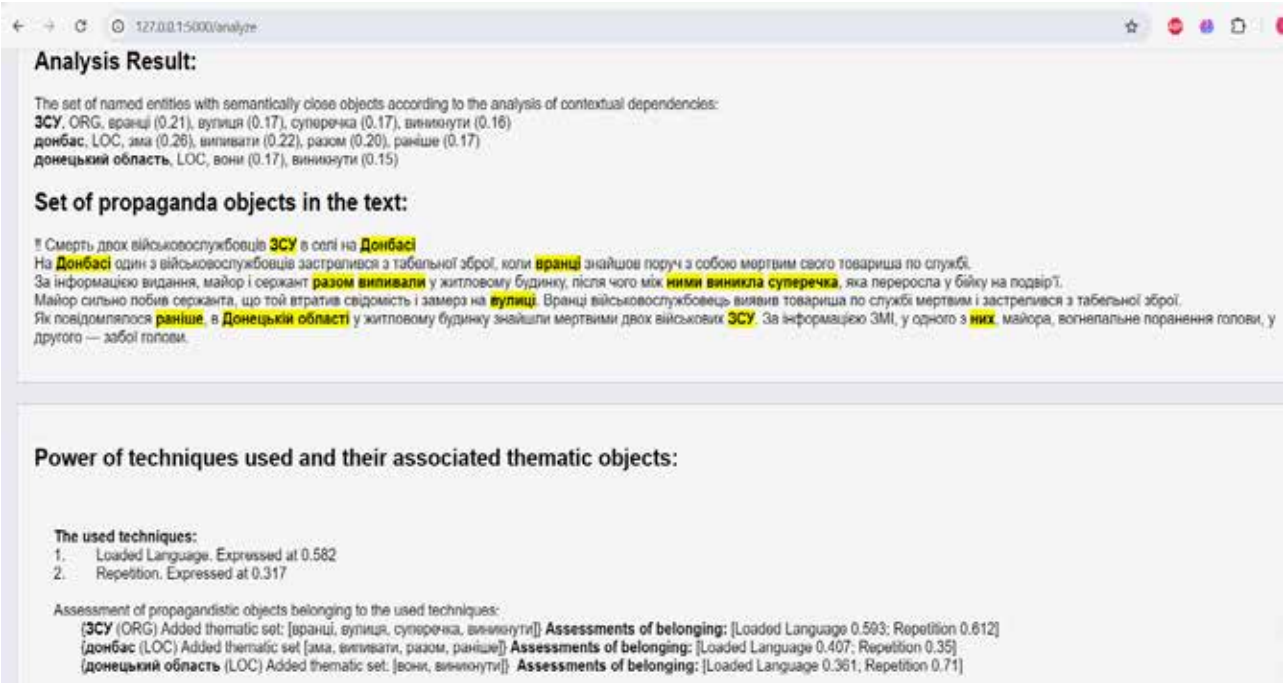


Рис. 3. Аналіз тестового тексту з пропагандою розробленим програмним забезпеченням

додавання варіантів їх словесного вираження та використання контекстних вікон для аналізу зв'язків між прийомами й цілями пропаганди, що дозволило значно покращити якість визначення та забезпечити наочне представлення результатів.

У межах основної мети підвищення точності та якості визначення прийомів і цілей пропаганди за семантичними маркерами у текстовому контенті з використанням нейронних мереж та візуальною аналітикою результатів було розроблено метод, що дозволяє визначати цілі пропаганди та встановлювати, на кого або на що спрямовані пропагандистські прийоми. Метод вирішує ключові проблеми, зокрема відсутність комплексного аналізу взаємозв'язків між прийомами та цілями пропаганди, а також нестачу узагальнень і альтернативних згадок цілей у текстах.

Експериментально підтверджено ефективність запропонованого методу, який, на відміну від існуючих аналогів, поєднує пошук іменованих сутностей за допомогою бібліотеки «STANZA» із розширенням переліку цілей пропаганди завдяки бібліотеці машинного навчання «FastText». Крім того, метод оцінює, як знайдені цілі пов'язані з пропагандистськими прийомами. Дослідження показали, що результати, отримані за допомогою цього методу, узгоджуються з експертними оцінками. Для наочного представлення результатів надається візуальна аналітика, яка дозволяє визначати цілі впливу у контексті відповідних пропагандистських прийомів.

ЛІТЕРАТУРА:

1. Nazarov V., Molchanova M. Information System for Detecting Abusive Speech in Audio Content by Means of Natural Language. *Proceedings of V International Scientific and Practical Conference «Modern strategies of global scientific solutions». Stockholm, Sweden: International Scientific Unit*y, 2023. С. 132–135

2. Молчанова М. О., Мазурець О. В., Собко О. В. Віт Р. В., Назаров, В. В. Алгоритм виявлення аб'юзивного вмісту в україномовному аудіоконтенті для імплементації в об'єктно-орієнтовану інформаційну систему. *Вісник Хмельницького національного університету. Технічні науки*. 2024. № 1 (331). С. 101–106.

3. Молчанова М. О., Залуцька О. О., Бармак О. В. Метод інтелектуального аналізу тональності текстів. *Матеріали XII Всеукраїнської науково-практичної конференції «Глушковські читання». Київ, 2023*. С. 113–116

4. Krak I., Zalutska O., Molchanova M., Mazurets O., Bahrii R., Sobko O., Barmak O. Abusive Speech Detection Method for Ukrainian Language Used Recurrent Neural Network. *CEUR Workshop Proceedings*. 2024. Vol. 3688. С. 16–28.

5. Zalutska O., Molchanova M., Sobko O., Mazurets O., Pasichnyk O., Barmak O., Krak I. Method for Sentiment Analysis of Ukrainian-Language Reviews in E-Commerce Using RoBERTa Neural Network. *CEUR Workshop Proceedings. 2023. Vol. 3387*. С. 344–356

6. Ahmad P. N., Yuanchao L., Aurangzeb K., Anwar M. S., Haq Q. M. U. Semantic web-based propaganda text detection from social media using meta-learning. *Service Oriented Computing and Applications. 2024.* С. 1–15.

7. Rodríguez-García R., Centeno R., Rodrigo Á. Together we can do it! A roadmap to effectively tackle propaganda-related tasks. Internet Research. 2024

8. Szwoch J., Staszkow M., Rzepka R., Araki K. Limitations of Large Language Models in Propaganda Detection Task. *Applied Sciences*. 2024. Т. 14, № 10. С. 4330

9. Krak I., Molchanova M., Mazurets O., Sobko O., Zalutska O., Barmak O. Method for Neural Network Detecting Propaganda Techniques by Markers With Visual Analytic. *CEUR Workshop Proceedings.* 2024. Vol. 3790. С. 158–170.

10. Yoosuf S., Yang Y. Fine-Grained Propaganda Detection with Fine-Tuned BERT. Proceedings of the Second Workshop on Natural Language *Processing for Internet Freedom: Censorship, Disinformation, and Propaganda. Hong Kong, China*, 2019. С. 87–91.

11. Propaganda Analysis Project. URL: https://propaganda.math.unipd.it/index.html. Дата звернення (20.11.2024)/

12. Zenodo. Propaganda. Proppy Corpus 1.0. URL: https://zenodo.org/records/3271522#.XS6qRUUzau4. Дата звернення (20.11.2024)

REFERENCES:

1. Nazarov, V., & Molchanova, M. (2023). Information system for detecting abusive speech in audio content by means of natural language. *Proceedings of the V International Scientific and Practical Conference «Modern strategies of global scientific solutions», Stockholm, Sweden: International Scientific Unity,* 132–135.

2. Molchanova, M. O., Mazurets, O. V., Sobko, O. V., Vit, R. V., & Nazarov, V. V. (2024). Alhorytm vyiavlennia abiuzyvnoho vmistu v ukrainomovnomu audiokontenti dlia implementatsii v obiektno-oriientovanu informatsiinu systemu [Algorithm for detecting abusive content in Ukrainian-language audio content for implementation in an object-oriented information system]. *Visnyk Khmelnytskoho natsionalnoho universytetu. Tekhnichni nauky,* 1(331), 101–106 [in Ukrainian].

3. Molchanova, M. O., Zalutska, O. O., & Barmak, O. V. (2023). Metod intelektualnoho analizu tonalnosti tekstiv [Method of intellectual analysis of text tone]. *Materialy XII Vseukrainskoi naukovo-praktychnoi konferentsii «Hlushkovski chytannia»,* 113–116 [in Ukrainian].

4. Krak, I., Zalutska, O., Molchanova, M., Mazurets, O., Bahrii, R., Sobko, O., & Barmak, O. (2024). Abusive speech detection method for Ukrainian language using recurrent neural network. *CEUR Workshop Proceedings,* 3688, 16–28.

5. Zalutska, O., Molchanova, M., Sobko, O., Mazurets, O., Pasichnyk, O., Barmak, O., & Krak, I. (2023). Method for sentiment analysis of Ukrainian-language reviews in e-commerce using RoBERTa neural network. *CEUR Workshop Proceedings,* 3387, 344–356.

6. Ahmad, P. N., Yuanchao, L., Aurangzeb, K., Anwar, M. S., & Haq, Q. M. U. (2024). Semantic web-based propaganda text detection from social media using meta-learning. *Service Oriented Computing and Applications*, 1–15.

7. Rodríguez-García, R., Centeno, R., & Rodrigo, Á. (2024). Together we can do it! A roadmap to effectively tackle propaganda-related tasks. *Internet Research*.

8. Szwoch, J., Staszkow, M., Rzepka, R., & Araki, K. (2024). Limitations of large language models in propaganda detection task. *Applied Sciences,* 14(10), 4330.

9. Krak, I., Molchanova, M., Mazurets, O., Sobko, O., Zalutska, O., & Barmak, O. (2024). Method for neural network detecting propaganda techniques by markers with visual analytic. *CEUR Workshop Proceedings,* 3790, 158–170.

10. Yoosuf, S., & Yang, Y. (2019). Fine-grained propaganda detection with fine-tuned BERT. Proceedings of the Second Workshop on Natural Language *Processing for Internet Freedom: Censorship, Disinformation, and Propaganda, Hong Kong, China*, 87–91.

11. Propaganda Analysis Project. (2024, November 20). Retrieved from: https://propaganda.math.unipd.it/index.html.

12. Zenodo. (2019, November 20). Propaganda. Proppy Corpus 1.0. Retrieved from: https://zenodo.org/records/3271522#.XS6qRUUzau4.

UDC 621.926.5-52
DOI https://doi.org/10.32782/IT/2024-4-18

Igor NOVYTSKYI
Doctor of Technical Sciences, Professor, Professor at the Department of Cyberphysical and Information-Measuring Systems, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave, Dnipro, Ukraine, 49005, novytskyi.i.v@nmu.one
ORCID: *0000-0002-8780-6589*
Scopus Author ID: *59312716800*

Yuliia SHEVCHENKO
Assistant Lecturer at the Department of System Analysis and Control, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave, Dnipro, Ukraine, 49005, shevchenko.yu.o@nmu.one
ORCID: *0000-0002-3895-3937*
Researcher ID: *JVN-9643-2024*
Scopus Author ID: *59312616400*

Maksym TRYPUTEN
Student, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave, Dnipro, Ukraine, 49005, tryputen.m.m@nmu.one

To cite this article: Novytskyi, I., Shevchenko, Yu., Tryputen, M. (2024). Adaptyvna systema upravlinnia protsesom samopodribnennia rud u barabannykh mlynakh [Adaptive control system of the self-grinding process of ores in autogenous mills]. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 4, 151–158, doi: https://doi.org/10.32782/IT/2024-4-18

ADAPTIVE CONTROL SYSTEM OF THE SELF-GRINDING PROCESS OF ORES IN AUTOGENOUS MILLS

The main goal *is development of an adaptive self-regulating system for controlling the loading process of drum mills for autogenous ore grinding, providing specified indicators of control quality under conditions of uncontrolled disturbances.*

Methodology. *To achieve this goal, methods of system analysis and synthesis of automatic control theory systems and mathematical modeling methods were used to assess the quality of processes for regulating the parameters of the adaptive system.*

Scientific novelty. *A method is proposed for solving the problem of synthesizing an adaptive loading control system for autogenous mills, which takes into account in its formulation restrictions on the level of control action and a possible change in the structure of the mathematical model of the object.*

Conclusions. *The feasibility of using adaptive self-regulating systems for controlling the loading of autogenous fuel mills is substantiated. The problem of synthesizing the main loop of the control system is formulated and solved, taking into account real restrictions on the control action and possible changes in the structure of the mathematical model of the object. The dependence of the control time of the main loop on the restrictions on the control action has been studied. As a result of modeling the processes in the self-tuning loop of the adaptive system, it was established that the time for setting the model parameters is significantly less than the decay time of the correlation function of the processes that cause drift of the object parameters. Thus, the fulfillment of the conditions of quasi-stationarity indicates the operability of the adaptive system.*

Practical significance. *The use of the proposed adaptive system for controlling the loading of autogenous mills provides the necessary indicators of control quality under conditions of restrictions on the control action and the non-stationary nature of the facility.*

Key words: *autogenous mill, modeling, adaptive system, mathematical model, loading control, self-tuning, quality indicators, uncontrolled disturbances, uncontrolled disturbances.*

Ігор НОВИЦЬКИЙ
доктор технічних наук, професор, професор кафедри кіберфізичних та інформаційно-вимірювальних систем, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, м. Дніпро, Україна, 49005
ORCID: *0000-0002-8780-6589*
Scopus Author ID: *59312716800*

Юлія ШЕВЧЕНКО
асистент кафедри системного аналізу та управління, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, м. Дніпро, Україна, 49005
ORCID: *0000-0002-3895-3937*
Researcher ID: *JVN-9643-2024*
Scopus Author ID: *59312716800*

Максим ТРИПУТЕНЬ
студент, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, м. Дніпро, Україна, 49005

Бібліографічний опис статті: Новицький, І., Шевченко, Ю., Трипутень, М. (2024). Адаптивна система управління процесом самоподрібнення руд у барабанних млинах. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 151–158, doi: https://doi.org/10.32782/IT/2024-4-18

АДАПТИВНА СИСТЕМА УПРАВЛІННЯ ПРОЦЕСОМ САМОПОДРІБНЕННЯ РУД У БАРАБАННИХ МЛИНАХ

Метою цієї роботи *є розробка адаптивної самонастроювальної системи управління процесом завантаження барабанних млинів самоподрібнення руд, що забезпечує заданих показників якості управління в умовах дії неконтрольованих збурень.*

Методологія передбачає *використання методів системного аналізу і методів синтезу з теорії автоматичного управління, а також методи математичного моделювання для оцінки якості процесів налаштування параметрів адаптивної системи.*

Наукова новизна. *Запропоновано метод вирішення задачі синтезу адаптивної системи управління завантаженням млинів самоподрібнення, які враховують у своїй умові обмеження на рівень управляючого впливу і можливість зміни структури математичної моделі цього об'єкту.*

Висновки. *Обґрунтована доцільність використання адаптивних самонастроювальних систем управління завантаженням млинів самоподрібнення. Вирішена задача синтезу основного контуру системи управління з урахуванням реальних обмежень на управляючий вплив і можливе змінення структури математичної моделі об'єкта. Досліджена залежність часу регулювання основного контуру від обмежень на управляючий вплив. У результаті моделювання процесів у контурі самоналаштування адаптивної системи встановлено, що час настройки параметрів моделі значно менше часу спадання кореляційної функції процесів, що визивають дрейф параметрів об'єкту. Виконання умови квазістаціонарності свідчить про працездатність розробленої адаптивної системи управління млином.*

Практична значимість. *Використання запропонованої адаптивної системи управління завантаженням млинів самоподрібнення забезпечує необхідні показники якості управління в умовах обмеження на управляючий вплив та не стаціонарність цього об'єкту управління.*

Ключові слова: *млин самоподрібнення, моделювання, адаптивна система, математична модель, управління завантаженням, самоналаштування, показники якості, неконтрольовані збурення.*

Introduction. The emergence and development of modern means of information processing makes it possible to implement effective multi-level systems of automatic management of technological processes of enrichment technology, in which the object of management is the entire technological line, and the economic efficiency of its work is taken as a quality criterion **(**Sokur et al., 2020)**.**

Such generalized economic criteria are the most sensitive to the modes of operation of the resource-intensive devices of the two-stage technological lines (Novytskyi et al., 2022). Therefore, the development of effective local control systems for aggregates such as drum mills, which account for more than half of all beneficiation technology costs (Sokur et al., 2016), is an urgent task.

Setting the problem of research. Fig. 1 presents a functional scheme of the control system with a classifier for the degree of filling of a self-grinding mill working.

It is known that for self-grinding mills, the degree of drum filling is a critical parameter (Novytskyi et al., 2021), the control and regulation of which ensures trouble-free (absence of jamming) and efficient operation of the mill.

Therefore, the purpose of managing the local system of Fig. 1 – to ensure, by adjusting the frequency of rotation of the feeder blade V, the stabilization of the degree of filling of the mill φ at the level of the task φ_{val}, the value of which is determined at the top level of the hierarchy.

As an object of control, a self-crushing mill with a classifier along the channel «feeder blade speed V – degree of drum filling φ» is described with sufficient accuracy by the transfer function:

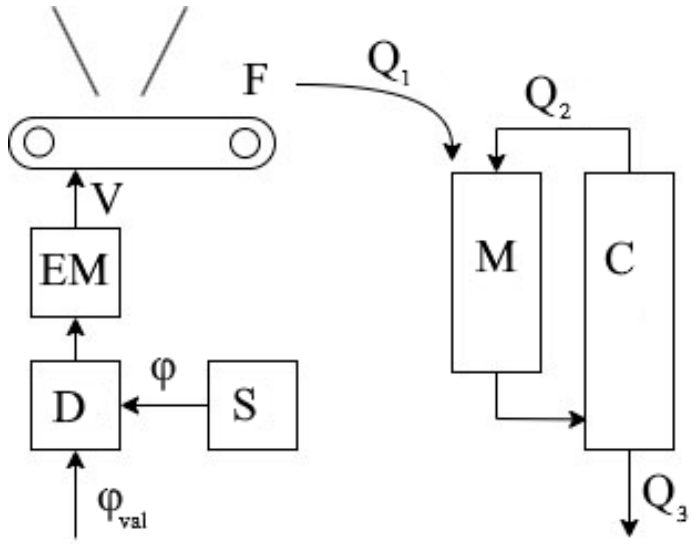


Fig. 1. Control system for the degree of filling of a self-grinding mill working: M – autogenous mill, C – classifier, S – mill filling sensor, F – feeder, D – computing device, EM – executive mechanism, Q_1, Q_2, Q_3 – respectively, flows of initial ore, reversible loading and finished product

$$W_{OC} = \frac{K}{T_2^2 p^2 + T_1 p + 1}. \quad (1)$$

The main task of managing the degree of filling φ is complicated by the fact that the self-grinding drum mill is essentially an inertial and non-stationary object – the parameters of the transfer function (1), such as K, T_1, T_2, as change in a fairly wide range (up to 15% of nominal values). The main reason for the drift of the model parameters is a change in the physical and mechanical properties of the original ore sent for grinding (Khotskina, 2014). In addition to affecting the transmission coefficient K, these disturbances can qualitatively change the structure of the object (1), which acquires the properties of value $T_1 < 2T_2$ an oscillating link.

Thus, to ensure the effective operation of the control system of such a non-stationary object, it is necessary to provide for adaptive regulation of the parameters of the main circuit.

The **main goal of the research** is development of an adaptive self-regulating system for controlling the loading process of drum mills for autogenous ore grinding, using methods of analysis and synthesis of automatic control theory systems and mathematical modeling methods and providing specified indicators of control quality under conditions of uncontrolled disturbances.

Literature review. A large number of works by domestic and foreign authors are devoted to solving the problems of increasing the efficiency of the grinding mill as due to the improvement of structural elements of mills (Arvidsson, 2006) and classifying devices (Morkun et al., 2018) and to optimization the management process.

Most of them involve a traditional approach based on the use of information about the characteristics of the input flow of material (Pistun et al., 2002) and the finished product of the grinding process (Fedoryshyn et al., 2012). The emergence of modern means of information processing and management contributed to the development of the direction based on the use of adaptive systems (Feng et al., 2019). Conventionally, they can be divided into two types: systems with adjustment of model parameters (Novytskyi & Shevchenko, 2014), which is considered in this research paper, and systems based on non-parametric identification methods. The second type working with non-parametric models (mainly neural networks) have a number of advantages.

However, they have a long adjustment time and their application in production requires a strict justification of quasi-stationarity conditions.

Solving the problem. Synthesis of the system must be performed in two important main stages:

1) determination of the structure and calculation of the nominal values of the parameters of the main control scheme;

2) determination of the structure of the self-regulation scheme.

In the process of synthesizing the main scheme of the system, two factors must be taken into account:

– the roots of the characteristic equation of the object model (1) can be complex;

– the speed of the feed web is limited by size, i.e.

$$V \le V_{max} \ . \qquad (2)$$

Taking into account these conditions, it is advisable to use the synthesis method (Novytskyi & Us, 2017), which provides an aperiodic response of a closed control system to a typical input influence with a given adjustment time T_{adj} and re-adjustment of no more than 2%. In Fig. 2 shows the structural diagram of the main circuit where W_{d1} and W_{d2} are the transfer functions of correction devices:

$$W_{d1} = \frac{C_1 T_2^2}{p + C_2}, \qquad (3)$$

$$W_{d2} = \frac{C_3 (p + C_2)}{p + C_4}.$$

Calculation of the values of the parameters C_1, C_2, C_3, C_4 that ensure the aperiodic response of the closed system Fig. 2 with a given adjustment time T_{adj} is performed as follows. The transmission function of the system on the channel « $\varphi_{val} - \varphi$ » is as follows:

$$W_c(p) = \frac{W_{d1} \cdot W_{d2} \cdot W_{oc}}{1 + W_{d2} \cdot W_{oc}} =$$

$$= \frac{KC_1C_3}{p^3 + \frac{T_1 + C_4T_2^2}{T_2^2} \cdot p^2 + \frac{KC_3 + T_1C_4 + 1}{T_2^2} \cdot p + \frac{C_4 + C_2C_3K}{T_2^2}}.$$

We denote:

$$\omega^3 = KC_1C_3 = \frac{C_4 + C_2C_3K}{T_2^2}, \qquad (4)$$

and next reduce the transfer function $W_c(p)$ to the normalised form:

$$W_c(\bar{p}) = \frac{1}{\bar{p}^3 + \frac{T_1 + C_4T_2^2}{T_2^2\omega} \cdot \bar{p}^2 + \frac{KC_3 + T_1C_4 + 1}{T_2^2\omega^2} \cdot \bar{p} + 1}, \qquad (5)$$

where $\bar{p} = p / \omega$ is a normalised argument.

To ensure the aperiodicity of the response, the parameters of the transfer function (5) for the third-order system should have the following values (Dorf & Bishop, 2022):

$$\omega = \frac{4{,}04}{T_{adj}}.$$

$$\alpha = \frac{T_1 + C_4T_2^2}{T_2^2 \cdot \omega} = 1{,}9.$$

$$\beta = \frac{KC_3 + T_1C_4 + 1}{T_2^2 \cdot \omega^2} = 2{,}2.$$

Thus, the sequence of calculating the values of the parameters C_1, C_2, C_3, C_4 of the correcting devices will be as follows:

$$\omega = \frac{4{,}04}{T_{adj}}.$$

$$C_4 = \frac{\alpha T_2^2 \omega - T_1}{T_2^2}.$$

$$C_3 = \frac{\beta T_2^2 \omega^2 - T_1C_4 - 1}{K}. \qquad (6)$$

$$C_1 = \frac{\omega^3}{KC_3}.$$

$$C_2 = \frac{C_1T_2^2KC_3 - C_4}{KC_3}.$$

The output data are the parameters of the object model K, T_1, T_2 and the specified adjustment time T_{adj}.

It should be noted that the minimum possible setting time of T_{adj} is determined by restrictions on the control influence, in this case it is the maximum speed of the feeder blade V according (2).

In (Fig. 3, *a)* presents the transient process of changing the degree of filling $\varphi(t)$ in the control system for $\varphi_{val} = 40\%$, defined nominal values of the object's parameters: $K = 60\,sec \cdot m^{-1}; T_1 = 16\,sec; T_2 = 12\,sec; V_{max} = 1\,\frac{m}{sec}$ and the parameters of the correction devices (6) calculated according to $C_1 = 0{,}001$; $C_2 = 0{,}0852$; $C_3 = 0{,}0405$; $C_4 = 0{,}1448$. In (Fig. 3, *b*) shows the process of changing the control influence $V(t)$ – the speed of the feeder.

From (Fig. 3, *a, b*), it follows that condition (2) for limiting the speed of the feeder web in the control

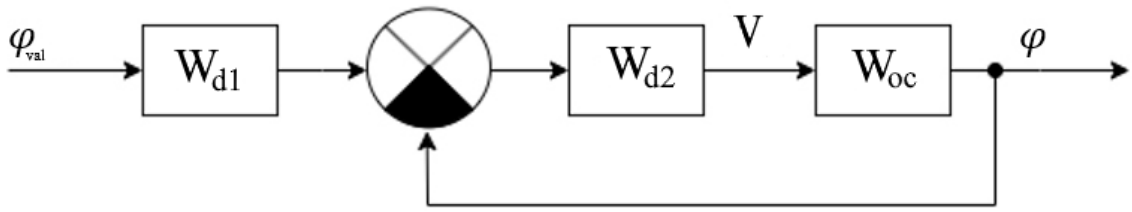
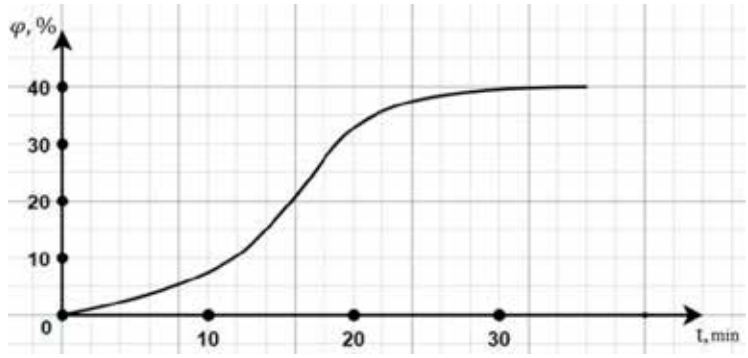
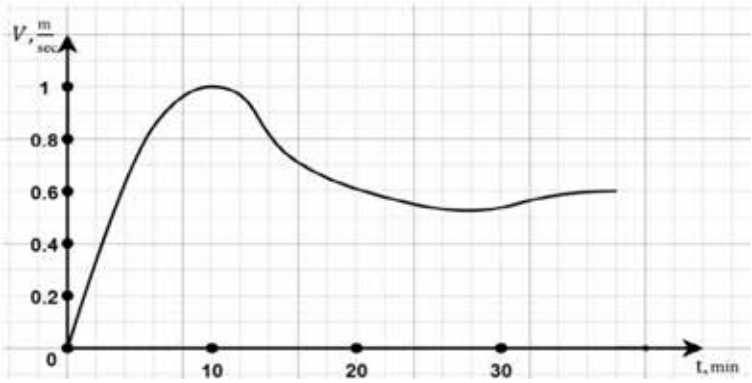


Fig. 2. Structural diagram of the main circuit of the adaptive control system



a)



b)

Fig. 3. Processes of changes in the degree of φ(t) filling (a) and the speed of the feeder blade $V(t)$ in the management system (b)

process will be fulfilled if the specified adjustment time T_{adj} is at least 30 minutes. In the opposite case, condition (2) is violated, overregulation increases and oscillations occur in the system.

To solve the problem of adapting the system to the physical and mechanical properties of the original ore, which change and cause changes in the values of the model parameters of the control object K, T_1, T_2 it is advisable to use a self-adjusting system with an adjustable model (Fig. 4).

The control influence $V(t)$ simultaneously enters the inputs of the control object and the model. The mismatch between object and model outputs ε (mismatch function) is used to adjust model parameters $P = \{K_M, T_{1M}, T_{2M}\}$ according to the gradient law:

$$\frac{dP}{dt} = -\alpha grad\varepsilon^2, \qquad (7)$$

where P – is the adjustable element, α – is the coefficient that determines the adjustment speed.

In the computing device D (Fig. 4) according to the adjusted parameters of the model K_M, T_{1M}, T_{2M} in accordance with expressions (6), the values of the parameters of the correction devices – C_1, C_2, C_3, C_4, are calculated, in which an aperiodic response is ensured in the main circuit and the processes correspond (Fig. 3).

Calculation of processes in the adaptive system (Fig. 4) was performed by the methods Δt. In picture (Fig. 5) we can see the processes of setting parameters of the control object model K_M, T_{1M}, T_{2M} with simultaneous jump-like application of them on the object by 10 %.

From the graphs (Fig. 5) obtained, it follows that the process of setting model parameters ends within 40 minutes. This is significantly less than the period of quasi-stationarity, which is determined by the rate of change in the properties of the original ore and, in the presence of storage bunkers, is several hours.

After setting the parameters of the model, the processes in the main circuit along the channel « $\varphi_{val} - \varphi$ » correspond to the curves in (Fig. 3) and have the desired quality indicators.

Conclusions and prospects for further development. In this work, an adaptive system is proposed (Fig. 2) to control the loading process of self-grinding mills, which provides the specified characteristics of transient processes in the main control circuit in conditions of non-stationarity of the object.

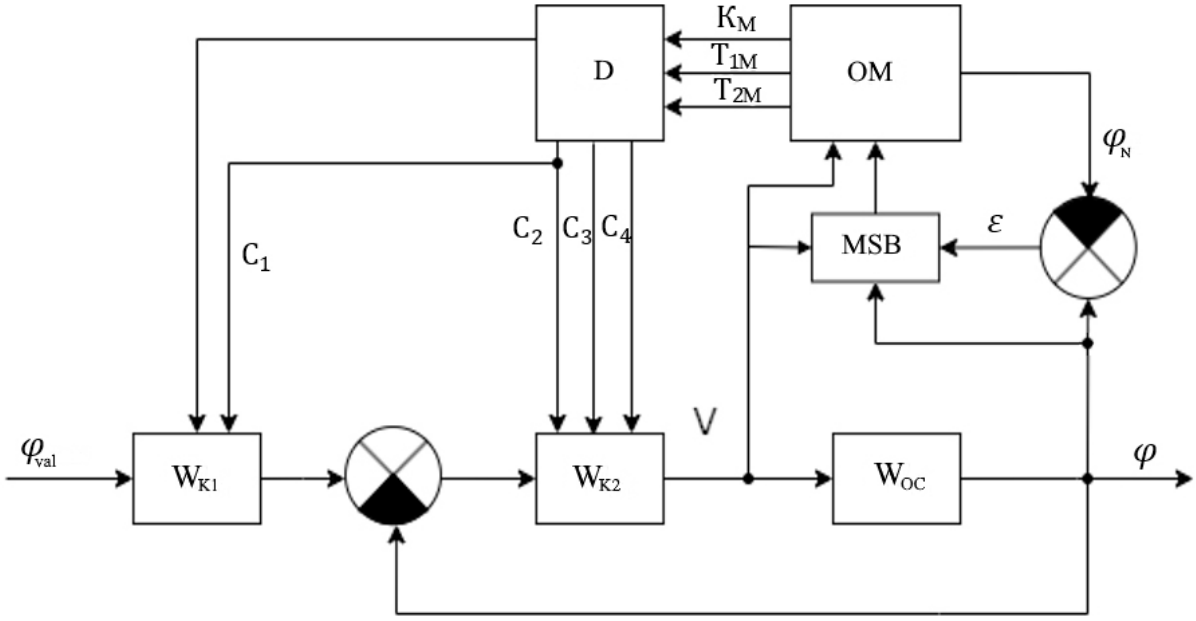


Fig. 4. Structural diagram of the adaptive control system: MSB – model setting block, OM – object model, D – computing device

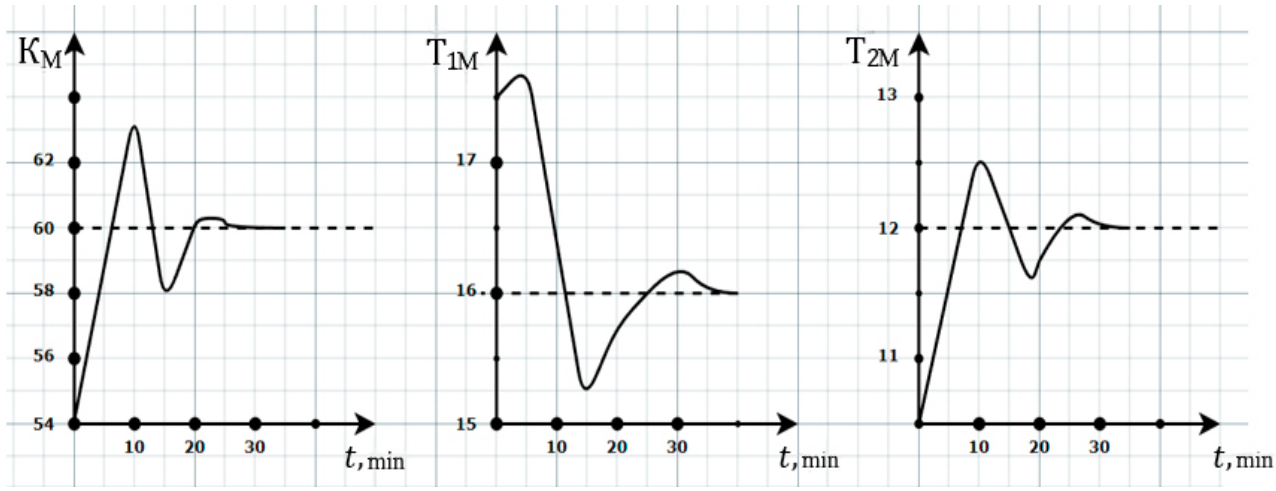


Fig. 5. Processes of setting parameters of the adaptive system object model

At the same time, the specified (minimum) adjustment time in the main circuit is determined by the limitation on the level of influence of control – speed of the feeder V (2). Using of an adaptive self-adjusting control system (Fig. 4) for the filling of self-grinding mills allows ensuring the necessary indicators of the quality of management in the conditions of variable parameters of the object of control (6). As a result of modeling the processes in the adaptive system, it is established that setting its parameters ends much faster (Fig. 5) than changing the parameters of the control object.

So, the following is done in this research paper:

1) use of an adaptive self-regulating control system for the filling of self-grinding mills allows you to ensure the necessary indicators of the quality of management in the conditions of changes in the parameters of the object;

2) the adjustment time of the main loop of the filling control system of the self-grinding mill is determined by restrictions on the level of control influence.

In the future, it is proposed to carry out research on the development of a method for determining the parameter α (7), which affects the speed of setting the parameters of our adaptive system model.

BIBLIOGRAPHY:

1. Sokur V. I. Investigation of the process of crushing solid materials in the centrifugal disintegrators / V. I. Sokur, V. S. Biletskyi, I. O. Vidmid, E.M. Robota. *Eastern-European Journal of Enterprise Technologies*. 2016. № 81. P. 34–40.

2. Novytskyi I. V. A systematic approach to solving the problems of optimal management of two-stage technological lines / I.V. Novytskyi, V.V. Sliesariev, Y.O. Shevchenko. *Collection of research papers of the National Mining University*. 2022. № 71. P. 211–216. doi: https://doi.org/10.33271/crpnmu/71.222

3. Sokur V. I. Ore preparation: crushing, grinding, classification. Kremenchuk. V.I. Sokur, V.S. Bileskiy., I.O. Vidmid., E.M. Robota. Kremenchug 2020. ISBN 978-617-639-272-9.

4. Novytskyi I. V. Основні принципи організації пошукових процедур при управлінні процесом самоздрібнювання в барабанних млинах / I.V. Novytskyi, V.V. Sliesariev, A.V. Malienko. *Collection of research papers of the National Mining University*. 2021. № 66. P. 245–253. doi: 10.33271/crpnmu/66.245

5. Khotskina V. B. Automatization of control processes at the first stage of ore enrichment using quickly acting regressive models. *Naukovyi Visnyk Natsionalnogo Hirnychoho Universytetu*. 2014. № 3. P. 145–151.

6. Novytskyi I. V. Обґрунтування критерія оптимального керування процесом самоподрібнення руд у барабанних млинах / I.V. Novytskyi, Y.O. Shevchenko. *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu.* 2024. № 4. P. 211–216. doi: https://doi.org/10.33271/nvngu/2024-4/061

7. Arvidsson S. Remote control and automation in Swedish iron ore mining. Metallurgy – An integral part of Innovative Sweden: The Ministry of Industry. Employment and Communications, Sweden. 2006.

8. Morkun V., Morkun N. Estimation of the crushed ore particles density in the pulp flow based on the dynamic effects of high energy ultrasound. *Archives of Acoustics*. 2018. № 43(1). P. 61–67. doi: 10.24425/118080

9. Pistun Y. Automatic control and optimization of ball mills / Y. Pistun, V. Zagraj, A. Skobalo /// Proc. of VIII Forum of Power Engineers, Techn. Univ. of Opole, May 29–31. 2002. ISBN 83-88492-04-7. PP. 575–581, Publ. House of Tech. Univ. of Opole. Opole. Poland.

10. Fedoryshyn R. The improved system for automation and optimization of solid material grinding by means of ball mills / R. Fedoryshyn, H. Nykolyn, V. Zagraj, Y. Pistun. Annals of DAAAM for 2012 & Proceedings of the 23rd International DAAAM Symposium. Austria. 2012. P. 053–056. ISBN 978-3-901509-91-9.

11. Feng L., Yang F., Zhang W., Tian H. Model Predictive Control of Duplex Inlet and Outlet Ball Mill System Based on Parameter Adaptive Particle Swarm Optimization. Mathematical Problems in Engineering. 2019.

12. Novytskyi I. V. Adaptyvna systema upravlinnya zavantazenyam barabanyh mlyniv somopodribnennya [Adaptive loading control system for autogenous drum mills] / I. V. Novytskyi, Y. O. Shevchenko. Collection of research papers of the National Mining University. 2014. № 44. P. 103–109. url: https://ir.nmu.org.ua/handle/123456789/152406

13. Bai Y., He F. Modeling on the effect of coal loads on kinetic energy of balls for ball mills. *Energies*. 2015. № 8. PP. 6859–6880.

14. Novytskyi I. V Modern theory of healing: textbook for universities / I. V. Novytskyi, S. A. US. *National Mining University*. Dnipro. 2017. ISBN 978-966-350-661-6

15. Dorf R. Modern Control Systems / R. C. Dorf, R. H. Bishop. Prentice Hall. 2008. ISBN 0132270285.

REFERENCES:

1. Sokur, V., Biletskyy, L., Sokur, D., & Bozyk, I. (2016). Investigation of the process of crushing solid materials in the centrifugal disintegrators. *Eastern-European Journal of Enterprise Technologies,* 3/7(81), 34–40.

2. Novytskyi, I., Sliesariev, V., & Shevchenko, Y. (2022). A systematic approach to solving the problems of optimal management of two-stage technological lines. Collection of research papers of the National Mining University, 71, 2011- 216. https://doi.org/10.33271/crpnmu/71.222

3. Sokur, M. I., Biletskyi, V. S., Vidmid, I. O., & Robota, E. M. (2020). Ore preparation (crushing, grinding, classification): monograph. ISBN 9786176392729.

4. Novytskyi, I., Sliesariev, V., & Maliienko, A. (2021). The basic principles of organizing prospecting procedures for managing the process of autogenous grinding of ores in drum mills. *Collection of Research Papers of the National Mining University,* 66, 245–253. https://doi.org/10.33271/crpnmu/66.245

5. Khotskina, V. (2014). Automatization of control processes at the first stage of ore enrichment using quickly acting regressive models. *Naukovyi Visnyk Natsionalnogo Hirnychoho Universytetu,* 3, 145–151.

6. Novytskyi, I., & Shevchenko, Y. (2024). Обґрунтування критерія оптимального керування процесом самоподрібнення руд у барабанних млинах. *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu,* 4, 61–66. https://doi.org/10.33271/nvngu/2024-4/061

7. Arvidsson, S. (2006). Remote control and automation in Swedish iron ore mining. Metallurgy – An integral part of Innovative Sweden: The Ministry of Industry. *Employment and Communications,* Sweden.

8. Morkun, V., & Morkun, N. (2018). Estimation of the crushed ore particles density in the pulp flow based on the dynamic effects of high energy ultrasound. *Archives of Acoustics,* 43(1), 61–67. doi: 10.24425/118080

9. Pistun, Y., Zagraj, V., & Skobalo, A. (2002). Automatic control and optimization of ball mills, *Proc. of VIII Forum of Power Engineers, Techn. Univ. of Opole*, 575–581, ISBN 83-88492-04-7.

10. Fedoryshyn, R., Nykolyn, H., Zagraj, V., & Pistun, Y. (2012). The improved system for automation and optimization of solid material grinding by means of ball mills. Annals of DAAAM for 2012 & Proceedings of the 23rd International DAAAM Symposium, 053–056, ISBN 978-3-901509-91-9.

11. Feng, L., Yang, F., Zhang, W., & Tian, H. (2019). Model Predictive Control of Duplex Inlet and Outlet Ball Mill System Based on Parameter Adaptive Particle Swarm Optimization. *Mathematical Problems in Engineering*.

12. Novytskyi, I., & Shevchenko, Y. (2014). Adaptive loading control system for autogenous drum mills. *Collection of research papers of the National Mining University,* 44, 103–109. url: https://ir.nmu.org.ua/handle/123456789/152406

13. Bai, Y., He, F. (2015). Modeling on the effect of coal loads on kinetic energy of balls for ball mills. *Energies,* 8, 6859–6880.

14. Novytskyi, I., & Us, S. (2017). Modern theory of healing: textbook for universities. *National Mining University*. ISBN 978-966-350-661-6.

15. Dorf, R. & Bishop, R. (2022). Modern Control Systems. *14th Edition*. ISBN 978-0-13-730725-8.

УДК 004.896
DOI https://doi.org/10.32782/IT/2024-4-19

Віктор ОЛЕВСЬКИЙ
доктор технічних наук, професор, професор кафедри інформаційних технологій та комп'ютерної інженерії, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, м. Дніпро, Україна, 49005
ORCID: *0000-0003-3824-1013*
Scopus Author ID: *56419822400*

Ірина УДОВИК
кандидат технічних наук, доцент, деканка факультету інформаційних технологій, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького 19, Дніпро, Україна, 49005
ORCHID: *0000-0002-5190-841X*
Scopus-Author ID: *55998874400*

Дмитро ГРИЩАК
кандидат фізико-математичних наук, докторант кафедри інформаційних технологій та комп'ютерної інженерії, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, м. Дніпро, Україна, 49005
ORCID: *0000-0001-8956-8468*
Scopus Author ID: *57220062399*

Юлія ОЛЕВСЬКА
кандидат фізико-математичних наук, доцент, доцент кафедри прикладної математики, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, м. Дніпро, Україна, 49005
ORCID: *0000-0002-0235-1360*
Scopus Author ID: *57191851171*

Юлія ЖУРАВЛЬОВА
асистент кафедри інформаційних технологій та комп'ютерної інженерії, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, м. Дніпро, Україна, 49005, zhuravlova.yu.s@nmu.one
ORCID: *0000-0003-3146-6826*
Scopus Author ID: *57904162900*

Бібліографічний опис статті: Олевський, В., Удовик, І., Грищак, Д., Олевська, Ю., Журавльова, Ю. (2024). Методологічні основи інформаційних технологій аналізу багатоканальних аерокосмічних зображень. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 159–168, doi: https://doi.org/10.32782/IT/2024-4-19

МЕТОДОЛОГІЧНІ ОСНОВИ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ АНАЛІЗУ БАГАТОКАНАЛЬНИХ АЕРОКОСМІЧНИХ ЗОБРАЖЕНЬ

У статті аналізуються методологічні основи інформаційних технологій, що використовуються для обробки та аналізу багатоканальних аерокосмічних зображень. Сучасні космічні супутники забезпечують високоякісні зображення в різних спектральних діапазонах, що відкриває нові можливості для моніторингу земних процесів і дослідження навколишнього середовища.

Мета роботи *– виявити ключові етапи та технології, що забезпечують ефективний аналіз багатоканальних аерокосмічних зображень.*

Методологія *дослідження полягає у інтеграції сучасних інформаційних технологій для всебічного аналізу багатоканальних аерокосмічних зображень. Початковий етап передбачає попередню обробку даних, яка включає корекцію геометрії, атмосферну корекцію та нормалізацію яскравості. Ці процеси необхідні для усунення спотворень, які можуть вплинути на точність аналізу. Подальший спектральний аналіз*

є критично важливим для класифікації об'єктів на зображеннях. Застосування алгоритмів, таких як підтримуючі вектори (SVM) та випадковий ліс (Random Forest), дозволяє автоматизувати процес класифікації і підвищити його точність. У статті також розглядається просторовий аналіз, що передбачає вивчення просторових зв'язків між об'єктами. Геостатистичні методи, такі як інтерполяція, використовуються для оцінки просторового розподілу різних явищ, що відбуваються на поверхні Землі. Цей підхід особливо важливий для вивчення змін у ландшафтах та їх структурних характеристик. Часовий аналіз на основі часових рядів є ще одним важливим аспектом, що дозволяє дослідникам моніторити зміни в покритті території, розвиток інфраструктури та екологічні процеси. Використання таких даних є важливим для вивчення тенденцій та прогнозування майбутніх змін. Стаття акцентує увагу на застосуванні сучасних інформаційних технологій, зокрема геоінформаційних систем (ГІС) та методів машинного навчання, що суттєво підвищують ефективність аналізу багатоканальних космічних зображень. Інтеграція цих технологій дозволяє забезпечити точність і оперативність у дослідженнях, що стосуються навколишнього середовища, та сприяє кращому розумінню глобальних змін на планеті.

***Наукова новизна** дослідження полягає у створенні методологічних основ та розвитку сучасних інформаційних технологій аналізу багатоканальних аерокосмічних зображень. В роботі запропонована послідовність кроків обробки зображень на основі розробленої в НТУ «ДП» нової технологія геометричної та спектральної корекцій космічних зображень високого просторового розрізнення з урахуванням фізичних механізмів фіксації інформації та розробленого авторами алгоритму глибокого навчання з попередньою обробкою зображень для системи підтримки прийняття рішень.*

***Висновки** статті підкреслюють важливість подальшого розвитку методів і технологій в цій галузі, а також необхідність міждисциплінарного підходу для забезпечення всебічного аналізу та моніторингу. Розроблені в роботі методологічні основи обробки багатоканальних аерокосмічних зображень та удосконалена методика їх аналізу дозволяє створити на базі первинних зображень інформаційну систему підтримки прийняття рішень з керування літальних апаратів з урахуванням розташованих на земній поверхні об'єктів.*

***Ключові слова:** методи машинного навчання, багатоканальні космічні зображення, інформаційні технології, нейронна мережа, геоінформаційні системи.*

Viktor OLEVSKYI
Doctor of Technical Sciences, Professor, Professor at the Department of Information Technology and Computer Engineering, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, olevskyi.v.i@nmu.one
***ORCID:** 0000-0003-3824-1013*
***Scopus Author ID:** 56419822400*

Iryna UDOVYK
Candidate of Technical Science, Associate Professor, Dean of Information Technologies Department, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, udovyk.i.m@nmu.one
***ORCID:** 0000-0002-5190-841X*
***Scopus-Author ID:** 55998874400*

Dmytro HRYSHCHAK
Candidate of Technical Sciences, Doctorant at the Department of Information Technology and Computer Engineering, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, hryshchak.d.d@nmu.one
***ORCID:** 0000-0001-8956-8468*
***Scopus Author ID:** 57220062399*

Yuliia OLEVSKA
Candidate of Physical-Mathematical Sciences, Associate Professor, Associate Professor at the Department of Applied Mathematics, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, olevska.yu.b@nmu.one
***ORCID:** 0000-0002-0235-1360*
***Scopus Author ID:** 56419822400*

Yuliia ZHURAVLOVA
Assistant at the Department of Information Technology and Computer Engineering, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, zhuravlova.yu.s@nmu.one
***ORCID:** 0000-0003-3146-6826*
***Scopus Author ID:** 57904162900*

To cite this article: Olevskyi, V., Udovyk I., Hryshchak, D., Olevska, Yu., Zhuravlova, Yu. (2024). Metodolohichni osnovy informatsiinykh tekhnolohii analizu bahatokanalnykh aerokosmichnykh zobrazhen [Methodological foundations of information technologies for the analysis of multi-channel aerospace images]. Information Technology: Computer Science, Software Engineering and Cyber Security, 4, 159–168, doi: https://doi.org/10.32782/IT/2024-4-19

METHODOLOGICAL FOUNDATIONS OF INFORMATION TECHNOLOGIES FOR THE ANALYSIS OF MULTI-CHANNEL AEROSPACE IMAGES

The article analyzes the methodological foundations of information technologies used for processing and analysis of multi-channel aerospace images. Modern space satellites provide high-quality images in various spectral ranges, which opens up new opportunities for monitoring Earth processes and studying the environment.

***The purpose** of the work is to identify the key stages and technologies that provide effective analysis of multi-channel aerospace images.*

***The research methodology** consists in the integration of modern information technologies for comprehensive analysis of multi-channel aerospace images. The initial stage involves data preprocessing, which includes geometry correction, atmospheric correction, and brightness normalization. These processes are necessary to eliminate distortions that can affect the accuracy of the analysis. Further spectral analysis is critical for object classification in images. The use of algorithms such as support vectors (SVM) and random forest (Random Forest) allows you to automate the classification process and increase its accuracy. The article also considers spatial analysis, which involves the study of spatial relationships between objects. Geostatistical methods, such as interpolation, are used to estimate the spatial distribution of various phenomena occurring on the Earth's surface. This approach is especially important for studying changes in landscapes and their structural characteristics. Temporal analysis based on time series is another important aspect that allows researchers to monitor changes in land cover, infrastructure development, and ecological processes. Using such data is essential for studying trends and predicting future changes. The article focuses on the application of modern information technologies, in particular, geographic information systems (GIS) and machine learning methods, which significantly increase the efficiency of the analysis of multi-channel space images. The integration of these technologies allows for accuracy and efficiency in research related to the environment and contributes to a better understanding of global changes on the planet.*

***The scientific novelty** of the research lies in the creation of methodological foundations and the development of modern information technologies for the analysis of multi-channel aerospace images. The paper proposes a sequence of image processing steps based on the new technology of geometric and spectral correction of space images of high spatial resolution developed at NTU «DP», taking into account the physical mechanisms of information fixation and the deep learning algorithm developed by the authors with pre-processing of images for the decision support system.*

***The conclusions** of the article emphasize the importance of further development of methods and technologies in this field, as well as the need for an interdisciplinary approach to ensure comprehensive analysis and monitoring. The methodological foundations of processing multi-channel aerospace images developed in the work and the improved technique of their analysis allow creating an information system for decision-making support for aircraft control based on primary images, taking into account objects located on the earth's surface.*

***Key words:** machine learning methods, multi-channel space images, information technologies, neural networks, geo-information systems.*

Актуальність проблеми. Багатоканальні аерокосмічні зображення – це дані, отримані від дронів чи супутників, які фіксують електромагнітне випромінювання в різних спектральних діапазонах. Кожен канал може відображати різні фізичні властивості об'єктів, що дозволяє отримувати більш детальну інформацію про них. Використання багатьох каналів значно покращує можливості для класифікації, детекції змін та моніторингу. У сучасному світі, де інформація стає критично важливим ресурсом, аналіз багатоканальних космічних зображень відіграє ключову роль у ряді галузей, таких як екологія, сільське господарство, урбаністика, моніторинг кліматичних змін та управління природними ресурсами. Космічні супутники надають можливість отримувати детальну інформацію про стан Землі в різних спектральних діапазонах, що дозволяє виявляти та аналізувати численні природні та антропогенні явища. Основними причинами актуальності дослідження методологічних основ інформаційних технологій для аналізу таких зображень є наступні.

1. Зростаюча потреба в моніторингу навколишнього середовища. Зміни клімату, забруднення довкілля, руйнування природних екосистем вимагають постійного спостереження та аналізу. Багатоканальні зображення забезпечують цінні дані для оцінки стану природних ресурсів та екосистем.

2. Розвиток технологій дистанційного зондування. Інновації у сфері супутникових технологій дозволяють отримувати зображення з високою роздільною здатністю та різними

спектральними характеристиками. Це створює нові можливості для детального аналізу об'єктів і явищ на Землі.

3. Необхідність інтеграції даних. З метою отримання комплексного розуміння складних процесів, які відбуваються на Землі, важливо поєднувати дані з різних джерел. Методологічні основи інформаційних технологій дозволяють інтегрувати інформацію з космічних зображень з іншими формами даних, такими як наземні спостереження, демографічні дані та інше.

4. Зростання обсягу даних. Завдяки постійному вдосконаленню технологій збору даних, обсяги інформації, що генеруються космічними супутниками, зростають експотенційно. Це вимагає розробки нових методів та алгоритмів для обробки, аналізу та візуалізації цих даних.

5. Економічні та соціальні виклики. Ефективне управління ресурсами та сталий розвиток територій вимагають високоякісного аналізу даних. Використання інформаційних технологій для аналізу космічних зображень може сприяти оптимізації прийняття рішень у сфері екології, агрономії, урбаністики та інших галузях.

Таким чином, дослідження методологічних основ інформаційних технологій для аналізу багатоканальних космічних зображень є актуальним і необхідним кроком для розвитку сучасних методів моніторингу та управління природними ресурсами. Аналіз таких зображень є складним процесом, що потребує використання спеціалізованих інформаційних технологій. У цій статті розглянемо методологічні основи аналізу багатоканальних космічних зображень.

Аналіз останніх досліджень і публікацій. Аналіз багатоканальних космічних зображень є галуззю, яка динамічно розвивається та активно досліджується в наукових колах. Методологічні основи інформаційних технологій у цій сфері об'єднують різні підходи, від обробки зображень до статистичних методів і машинного навчання. Нижче наведено огляд основних джерел, що висвітлюють ці теми.

Класичний підручник (Gonzalez R. C., Woods R. E., 2018) охоплює основи обробки цифрових зображень, включаючи методи корекції, фільтрації та виявлення ознак. Він є основою для розуміння подальшого аналізу космічних зображень. У виданні (Jensen J. R., 2015) описуються принципи дистанційного зондування та методи аналізу даних, отриманих з космічних супутників. Книга також акцентує увагу на екологічних застосуваннях та управлінні ресурсами. Ще одне авторитетне джерело (Lillesand T., Kiefer R. W., Chipman J., 2015) детально описує технології та методи обробки космічних зображень, а також різноманітні аспекти їх інтерпретації. У книзі (Kogan F., 2019) обговорюються різні концепції та застосування дистанційного зондування, зокрема методи, що використовуються для аналізу змін у покритті території та екологічних змін. Оглядове дослідження (Hansen, M., Loveland, T., 2012) аналізує різні методи картографування земельного покриття на основі супутникових зображень, включаючи машинне навчання та статистичні методи. У статті (Adegun, A., Viriri, S., Tapamo, JR., 2023) розглядаються новітні підходи глибокого навчання для аналізу космічних зображень, а також їх переваги та недоліки в різних застосуваннях. Книга (Mather, P.M. & Magaly, K., 2011) пропонує всебічний огляд комп'ютерної обробки супутникових зображень, з акцентом на алгоритми обробки та методи візуалізації даних.

Аналіз багатоканальних космічних зображень є складним процесом, що супроводжується низкою проблем, які можуть вплинути на точність і достовірність отриманих результатів. Нижче наведено основні проблеми та можливі шляхи їх вирішення.

Атмосферні умови можуть значно спотворювати сигнал, що отримується від земних об'єктів. Це призводить до зміни спектральних характеристик зображень. Подолання цих спотворень можливо шляхом використання алгоритмів атмосферної корекції (Gonzalez, R. C., Woods, R. E., 2018), таких як 6S (Second Simulation of a Satellite Signal in the Solar Spectrum) або Dark Object Subtraction (DOS).

Рух супутника, нахил сенсора та інші фактори можуть призвести до геометричних спотворень зображень. Для виправлення геометричних спотворень використовують геометричну корекцію та ресемплінг. Можуть бути корисними такі алгоритми як cubic convolution (Hansen, M. C., Loveland, T. R., 2012).

Багатоканальні зображення генерують великі обсяги даних, що ускладнює їх обробку та аналіз. Для подолання цієї перешкоди використовують методи зменшення розмірності, такі як PCA (Principal Component Analysis) та тензорний аналіз для спрощення даних перед аналізом (Hansen, M. C., Loveland, T. R., 2012).

Неправильна класифікація або виявлення об'єктів через подібність спектральних характеристик долається шляхом використання алгоритмів машинного навчання, таких як Random Forest або SVM (Support Vector Machine), для підвищення точності класифікації (Han W., Zhang X., Wang Y. et al, 2023).

Недостатня кількість даних або їх нерівномірний розподіл можуть ускладнити аналіз (Kogan F., 2022). Шляхи подолання недоліку: застосування методу інтерполяції та екстраполяції для заповнення прогалин у даних; використання даних з інших джерел, таких як наземні спостереження.

Таким чином, аналіз багатоканальних космічних зображень стикається з численними викликами, але сучасні методи та технології можуть суттєво поліпшити точність і ефективність цього процесу. Використання алгоритмів обробки даних, машинного навчання та інтеграція різних джерел інформації дозволяє подолати багато з цих проблем, відкриваючи нові можливості для досліджень і практичних застосувань.

Мета дослідження полягає у виявленні ключових етапів та технологій, що забезпечують ефективний аналіз багатоканальних космічних зображень.

Виклад основного матеріалу дослідження. Дослідження в області аналізу багатоканальних космічних зображень є багатогранними і активно розвиваються. Важливою є інтеграція у них сучасних інформаційних технологій, що дозволяють покращити точність і ефективність аналізу. Це надає необхідну теоретичну базу та практичні приклади, що роблять їх корисними для науковців та інженерів у цій галузі. Методологічні основи аналізу полягають у проведенні наступних етапів дослідження.

Перед початком аналізу необхідно провести попередню обробку зображень, наприклад за методикою (Гнатушенко В.В., Каштан В.Ю., 2017), яка включає:

– корекцію геометрії: усунення спотворень, викликаних рухом супутника або атмосферними умовами.

– атмосферну корекцію: видалення впливу атмосфери на спектральні дані.

– нормалізацію яскравості: приведення зображень до єдиного масштабу для подальшого аналізу.

Найбільш перспективною є розроблена у роботах (Shevchenko V.Yu., Hnatushenko V.V., та ін., 2015; Kashtan V.Yu., Shedlovska Y.I., Hnatushenko V.V., 2017) нова технологія геометричної та спектральної корекцій космічних зображень високого просторового розрізнення з урахуванням фізичних механізмів фіксації інформації. Для визначення ефективності розробленої інформаційної технології отримані кількісні оцінки якості синтезованих багатоканальних зображень, зокрема ентропія по Шеннону, сигнальна ентропія та інші. В роботі пропонується метод об’єднання на основі використання пакетної побудови вейвлет-базисів (Каштан В.Ю., Гнатушенко В.В, 2016) з декореляцією первинних видових даних, бікубічної інтерполяції, методу HSV та гіперсферичного перетворення (HCT) (Шевченко В.Ю., Гнатушенко В.В., Кавац О.О., 2015). Основними етапами перетворення первинних багатоканальних зображень є завантаження фотограмметричних панхроматичного (ПЗ) і багатоканальних (БКЗ) зображень, отриманих зі супутника; потім передискретизація БКЗ і приведення його розмірності до розмірності ПЗ на основі бікубічної інтерполяції. Далі проводиться геометрична, радіометрична та контрастна корекція БКЗ та ПЗ і перетворення зображення з формату RGB в кольорову систему HSV. Після заміни компоненти яскравості багатоканального зображення такою ж компонентною панхромного зображення, проводиться обернене перетворення отриманого на попередньому етапі зображення з формату HSV в кольорову систему RGB. Отримане зображення вже буде мати підвищене просторове розрізнення у порівнянні з первинним знімком у натуральних кольорах. Наступним етапом є застосування розробленого методу геометричної та спектральної корекції за допомогою пакетного вейвлет-перетворення, також представленого в (Каштан В., Гнатушенко В., 2015). Зворотний пакетний вейвлет-розклад та одержання зображення БКС у форматі RGB дозволяє зробити перетворення 8-канального зображення, яке після інтерполяції в гіперсферичний колірний простір HCT і зворотного HCT-перетворення призводить до отримання результату Fusion.

Аналіз отриманих результатів дозволяє відзначити, що використанні запропонованої методики підвищує якість первинних космічних знімків та якість розпізнавання об’єктів на 10–12%. В подальших дослідженнях ми будемо користуватися саме цією технологією, лише замінивши бікубічну інтерполяцію на розроблену нами багатовимірну інтерполяцію типу Паде, яка є більш точною та компактною (Olevskyi V.I., Olevska Yu.B., Olevskyi O.V., Hnatushenko V.V., 2023).

Подальший спектральний аналіз є ключовим етапом обробки зображень. Він включає класифікацію зображень шляхом використання алгоритмів для визначення типів покриття, таких як SVM (Support Vector Machine) або Random Forest. SVM є потужним інструментом для обробки та аналізу багатоканальних космічних зображень. Він відмінно підходить для завдань класифікації та регресії завдяки своїй здатності знаходити оптимальну гіперплощину,

що розділяє дані. SVM можна використовувати для класифікації різних ознак на зображеннях, таких як ліси, водойми, урбанізовані території тощо. Спочатку можна витягнути важливі характеристики з багатосмугових зображень, такі як текстури, кольори та форми, а потім використовувати SVM для класифікації. SVM може працювати з великими обсягами даних, що часто зустрічається на космічних знімках. Використовуючи основні функції (наприклад, функцію радіального базису), SVM можуть ефективно обробляти нелінійні дані. Метод можна також адаптувати до конкретних завдань шляхом вибору коригування гіперпараметрів. SVM передбачає наступні кроки:

– Збір та попередня обробка зображень, включаючи корекцію атмосфери та нормалізацію даних.

– Визначення та виділення тих ознак, які будуть використовуватися для класифікації.

– Розбиття даних на навчальні та тестові вибірки, навчання моделі SVM на навчальній вибірці.

– Тестування та валідація, під час яких оцінюються продуктивність моделі на тестовому зразку, використовуючи такі показники, як точність та повнота.

– Використання навченої моделі для класифікації нових космічних зображень.

SVM добре працює навіть з невеликою кількістю прикладів навчання. Метод демонструє високу точність класифікації завдяки своїй здатності знаходити оптимальну межу між класами. Можливість використовувати різні основні функції дозволяє адаптувати SVM до різних типів даних. Таким чином, SVM є важливим інструментом при аналізі багатоканальних аерокосмічних знімків, що дозволяє ефективно класифікувати і витягувати корисну інформацію з великих обсягів даних.

Ще одним із найпопулярніших алгоритмів машинного навчання, який успішно застосовується в обробці та аналізі багатоканальних космічних зображень, є метод випадкового лісу (Random Forest, RF). Цей метод заснований на ансамблевому підході, що використовує множину вирішальних дерев для підвищення точності класифікації та зменшення перенавчання. RF може класифікувати різні типи об’єктів, такі як сільськогосподарські угіддя, водоймища, ліси та урбанізовані зони, що корисно для моніторингу навколишнього середовища та управління природними ресурсами. Метод випадкового лісу ефективно працює з великими обсягами даних, що притаманно космічних зображень, забезпечуючи хорошу масштабованість. RF може враховувати інформацію з різних спектрів (наприклад, видимий, інфрачервоний) та отримувати корисні ознаки з багатоканальних зображень, він стійкий до шуму в даних і може ефективно обробляти пропуски, що особливо важливо для супутникових знімків. Етапи використання Random Forest наступні:

– Збір, попередня обробка та нормалізація космічних зображень. Це може включати корекцію атмосферних ефектів і виділення областей, що цікавлять.

– Визначення ключових ознак, таких як текстури, спектральні індекси (наприклад, NDVI для рослинності), які можуть покращити класифікацію.

– Поділ даних на навчальну та тестову вибірки. Модель випадкового лісу навчається на вибірці з використанням великої кількості дерев (зазвичай від 100 до 1000).

– Оцінка продуктивності моделі на тестовій вибірці з використанням метрик, таких як точність та повнота.

– Використання навченої моделі для класифікації нових космічних зображень та отримання карток класифікації.

RF часто демонструє відмінні результати у завданнях класифікації завдяки агрегуванню передбачень кількох дерев. RF дозволяє оцінювати важливість ознак, що допомагає зрозуміти, які характеристики найбільш значущі для класифікації. Метод може бути налаштований для різних типів завдань, включаючи класифікацію та регресію. Таким чином, випадковий ліс є потужним інструментом для аналізу багатоканальних аерокосмічних зображень, дозволяючи отримувати точні класифікації та отримувати важливу інформацію з великих обсягів даних.

Всі ці дії необхідні для виділення певних ознак, аналізу спектральних характеристик для виявлення певних об’єктів або явищ. Саме це є практичною ціллю аерокосмічних досліджень земною поверхні. Виявляються цікаві для дослідника об’єкти и явища і далі проводиться їх просторовий аналіз. Він передбачає вивчення просторових зв’язків між об’єктами на зображеннях, геостатистичний аналіз з використанням методів інтерполяції для оцінки просторового розподілу явищ, а також аналіз фрагментації зображення і вивчення структури ландшафтів та їх зміни з часом. Аналіз змін у часі є важливим для моніторингу екологічних та соціальних процесів. Використовуються часові ряди (Гнатушенко В., Олевська Ю., Олевський В., 2024), що дозволяють відстежувати переміщення досліджуваних об’єктів, зміни в покритті території, розвитку інфраструктури та ін.

Використання інформаційних технологій та геоінформаційних систем (ГІС) є основою для аналізу багатоканальних аерокосмічних зображень з метою візуалізації даних та подальшої їх інтеграції до систем автоматичного керування літальними пристроями або систем підтримки прийняття рішень. На цьому етапі доцільно створювати нейронні мережі алгоритмами машинного навчання і застосувати їх для автоматизації, класифікації та виявлення аномалій. Нами розроблена система машинного навчання нейронної мережі для виявлення раку молочної залози (Aziukovskyi O., Hnatushenko V., Zavizion V., та інші, 2024), яка може бути застосована також для аналізу аерокосмічних знімків (рис. 1).

Початковий рівень – це згорткова нейронна мережа (VGGNet), яка ідентифікує підозрілі на наявність небезпеки зони на зображеннях. Основною метою цього першого рівня є побудова глибокої архітектури з численними згортковими шарами для ефективного вилучення функцій із зображень. Це забезпечує високу точність розпізнавання завдяки спеціалізації мережі для певного типу зображення, тим самим зменшуючи кількість помилкових

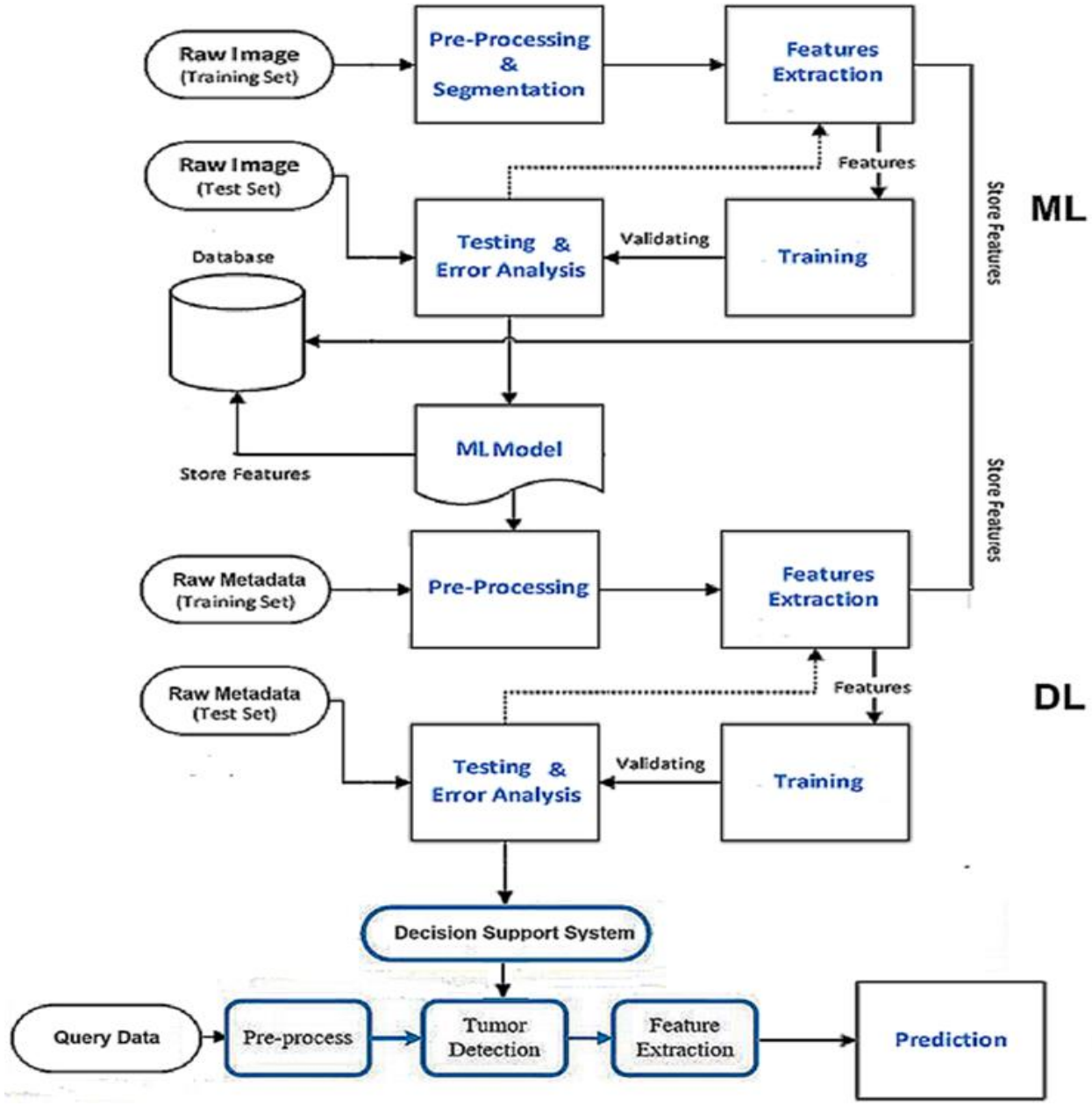


Рис. 1. Алгоритм глибокого навчання з попередньою обробкою зображень для системи підтримки прийняття рішень (Aziukovskyi O., Hnatushenko V., ZavizionV., та інші, 2024)

спрацьовувань. Модель глибокого навчання для системи підтримки прийняття рішень (DSS) навчається за допомогою бази даних, як показано на рис. 1.

Другий рівень призначений для створення DSS. Він виконує класифікацію проблемних об'єктів, виявлених на першому шарі, і генерує рекомендації для подальших досліджень і протоколів дій. У процесі глибокого навчання другий рівень використовує дані об'єктів з першого рівня разом із метаданими з бази даних.

Існуючі DSS стикаються зі значними проблемами у розпізнаванні підозрілих об'єктів на основі монохроматичних даних. Наша архітектура нейронної мережі для DSS вирішує це, об'єднуючи кілька зображень (2 або 3) на першому рівні, щоб підвищити точність виявлення об'єктів. Ці зображення масштабуються до єдиного розміру, який відповідає їхньому фізичному розташуванню на поверхні, а статистичний критерій узгодження застосовується до точок контуру об'єкту із заданим рівнем достовірності.

Ця модель для глибокого навчання може проводити поглиблений аналіз великих наборів даних і відкривати нові відомості, для пошуку яких вона, можливо, не була спеціально навчена.

При інтерпретації та візуалізації даних створюються псевдокольорові зображення, що дозволяють використання кольорових схем для візуалізації різних спектральних каналів. Створюється також 3D-візуалізація та тривимірні моделі для більш глибокого аналізу даних. Обробка зображень може виконуватися програмним забезпеченням для аналізу, таке як ENVI, ERDAS Imagine та інші.

Останнім часом з'явилося декілька новітніх напрямків в обробці багатоканальних супутникових знімків (Макс Поляков, 2024). Зовсім скоро у Національному управлінні океанічних і атмосферних досліджень (NOAA) мають з'явитися нові типи гіперспектральних сенсорів для аналізу складу атмосфери (ACX), які дозволяють отримувати дані в кількох спектральних діапазонах. Анонсовано використання радарів з синтезованою апертурою (SAR). Ці радари дозволяють отримувати високоякісні зображення незалежно від погодних умов і часу доби. Вони активно використовуються для моніторингу лісових пожеж, розливів нафти та інших екологічних катастроф.

Глобальні сервіси з космічної обізнаності (SSA) використовують супутникові дані для моніторингу космічного простору, включаючи відстеження космічного сміття та прогнозування зіткнень. Сервіси космічної обізнаності розширюють та пришвидшують систематизацію інформації, отриманої з різних супутників, підключених до єдиної мережі. Дані типи сервісів пропонують своїм клієнтам послуги з відстеження, моніторингу земної поверхні, орбітального контролю тощо. Як частина єдиної архітектури SSA-супутники за запитом можуть швидко надати необхідний перелік даних щодо ситуації на Землі. Це виявляється особливо корисним у випадку екстреного реагування на надзвичайні ситуації природного та техногенного характеру, або за необхідності швидкої реакції щодо вирішення питань національної і глобальної безпеки.

Висновки і перспективи подальших досліджень. Аналіз багатоканальних аерокосмічних зображень є складним, але водночас надзвичайно важливим процесом для отримання інформації про наземні об'єкти. Використання методологічних основ та сучасних інформаційних технологій дозволяє досягти високих результатів у вивченні різноманітних технічних, екологічних, соціальних та економічних процесів. Подальший розвиток технологій та методів аналізу обіцяє ще більші можливості для дослідників у цій галузі. Розроблені в роботі методологічні основи обробки багатоканальних аерокосмічних зображень та удосконалена методика їх аналізу дозволяє створити на базі первинних зображень інформаційну систему підтримки прийняття рішень з керування літальних апаратів з урахуванням розташованих на земній поверхні об'єктів.

ЛІТЕРАТУРА:

1. Adegun A. A., Viriri S., Tapamo J. R. Review of deep learning methods for remote sensing satellite images classification: Experimental survey and comparative analysis. *Journal of Big Data.* 2023. Vol. 10. № 93. P. 1–24.

2. Aziukovskyi O., Hnatushenko V., Zavizion V., Olevskyi V., Bulana T., Ivanov D., Gadiatskyi V. Decision Support System Oncoassist for Breast Cancer Computer-Aided Diagnostic. *Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту (ISDMCI'2024):* матеріали XX міжнародної наукової інтернет-конференції. Херсон: Книжкове видавництво ФОП Вишемирський В., 2024. С. 21–22.

3. Gonzalez R. C., Woods R. E. Digital image processing (4th ed.): Pearson Education, 2018. 1019 p.

4. Han W., Zhang X., Wang Y., Wang L., Huang X., Li J., Wang Sh., Chen W., Li X., Feng R., Fan R., Zhang X., Wang Y. A survey of machine learning and deep learning in remote sensing of geological environment: Challenges, advances, and opportunities. *ISPRS Journal of Photogrammetry and Remote Sensing.* 2023. № 202. P. 87–113.

5. Hansen M., Loveland T. A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment.* 2012. № *122*. P. 66–74.

6. Jensen J.R. Remote sensing of the environment: An earth resource perspective (2nd ed.): Pearson Education, 2009. 592 p.

7. Kashtan V. Yu., Shedlovska Y. I., Hnatushenko V. V. Processing technology of multispectral remote sensing images. *International Young Scientists Forum on Applied Physics,* Lviv, October 16–20 2017. Lviv, 2017. P. 355–358.

8. Kogan F. Remote sensing land surface changes. *Springer, Cham.* 2022. 462 p.

9. Lillesand T., Kiefer R. W., Chipman J. Remote sensing and image interpretation (7th ed.): John Wiley & Sons, 2015. 768 p.

10. Mather P. M., Magaly K. Computer processing of remotely-sensed images: An introduction (4th ed.). John Wiley & Sons, 2011. 464 p.

11. Olevskyi V. I., Olevska Yu. B., Olevskyi O. V., Hnatushenko V. V. Raster image processing using 2D Padé-type approximations. *Journal of Physics: Conference Series.* 2023. Vol. 2675. № 1. P. 012015.

12. Shevchenko V. Yu., Hnatushenko V. V., Kavats O. O. Pansharpening technology of high-resolution multispectral and panchromatic satellite images. *Naukovyi Visnyk NGU.* 2015. Vol. 4. № 148. P. 91–98.

13. Гнатушенко В. В., Олевська Ю. Б., Олевський В. І. Методи апроксимації рядами та їх застосування в біологічних і технічних задачах. Кременчук : Видавництво «НОВАБУК», 2024. 202 с.

14. Каштан В. Ю., Гнатушенко В. В. Інформаційна технологія підвищення інформативності багатоканальних даних на основі пакетних вейвлет-перетворень. *Комп'ютерно-інтегровані технології: освіта, наука, виробництво*. 2016. С. 77–83.

15. Каштан В. Ю., Гнатушенко В. В. Технологія геометричної та спектральної корекцій оптико-електронних космічних знімків. *Вісник ХНТУ.* 2017. Вип. 3. № 62. С. 286–291.

16. Поляков М. Нові можливості супутникових сенсорів: як змінилася EOS-індустрія за минулі роки. URL: https://maxpolyakov.com/ua/novi-mozhlivosti-suputnikovikh-sensoriv-yak-zminilasya-eos-industriya-za-minuli-roki/ (дата звернення: 24.09.2024).

17. Шевченко В. Ю., Гнатушенко В. В., Кавац О. О. Підвищення просторового розрізнення багатоканальних аерокосмічних зображень високого просторового розрізнення на основі гіперсферичного перетворення. *Науковий журнал Запорізького національного технічного університету, радіоелектроніка, інформатика, управління.* 2015. Вип. 1. № 32. С. 73–79.

REFERENCES:

1. Adegun, A. A., Viriri, S., & Tapamo, J. R. (2023). Review of deep learning methods for remote sensing satellite images classification: Experimental survey and comparative analysis. *Journal of Big Data,* 10(93), 1–24.

2. Aziukovskyi, O., Hnatushenko, V., Zavizion, V., Olevskyi, V., Bulana, T., Ivanov, D., & Gadiatskyi, V. (2024). Decision Support System Oncoassist for Breast Cancer Computer-Aided Diagnostic. *Intelectualni systemy pryiniattia rishen i problemy obchysliuvalnoho intelektu (ISDMCI'2024): Materialy XX Mizhnarodnoi naukovoi internet-konferentsii* [Intelligent Decision-Making Systems and Computational Intelligence Issues (ISDMCI'2024): Proceedings of the 20th International Scientific Internet Conference] (pp. 21–22). Kherson: Knyzhkove vydavnytstvo FOP Vishemirskyi V. S.

3. Gonzalez, R.C., & Woods, R.E. (2018). *Digital image processing* (4th ed.). Pearson Education.

4. Han, W., Zhang, X., Wang, Y., Wang, L., Huang, X., Li, J., … & Wang, Y. (2023). A survey of machine learning and deep learning in remote sensing of geological environment: Challenges, advances, and opportunities. *ISPRS Journal of Photogrammetry and Remote Sensing,* 202, 87–113.

5. Hansen, M., & Loveland, T. (2012). A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment,* 122, 66–74.

6. Jensen, J.R. (2009). *Remote sensing of the environment: An earth resource perspective* (2nd ed.). Pearson Education.

7. Kashtan, V. Yu., Shedlovska, Y. I., & Hnatushenko, V. V. (2017). Processing technology of multispectral remote sensing images [Conference paper]. *International Young Scientists Forum on Applied Physics,* October 16–20, Lviv, Ukraine, 355–358.

8. Kogan, F. (2022). *Remote sensing land surface changes*. Springer, Cham.

9. Lillesand, T., Kiefer, R.W., & Chipman, J. (2015). *Remote sensing and image interpretation* (7th ed.). John Wiley & Sons.

10. Mather, P. M. & Magaly, K. (2011). *Computer processing of remotely-sensed images: An introduction* (4th ed.). John Wiley & Sons.

11. Olevskyi, V. I., Olevska, Yu. B., Olevskyi, O. V., & Hnatushenko, V. V. (2023). Raster image processing using 2D Padé-type approximations. *Journal of Physics: Conference Series, 2675*(1), 012015.

12. Shevchenko, V. Yu., Hnatushenko, V. V., & Kavats, O. O. (2015). Pansharpening technology of high-resolution multispectral and panchromatic satellite images. *Naukovyi Visnyk NGU,* 4(148), 91–98.

13. Hnatushenko, V. V., Olevska, Yu. B., & Olevskyi, V. I. (2024). Metody aproksymatsii riadamy ta yikh zastosuvannia v biolohichnykh i tekhnichnykh zadachakh [Methods of series approximations and their application in biological and technical tasks]. Kremenchuk: Vydavnytstvo «NOVABOOK» [in Ukrainian].

14. Kashtan, V. Yu., & Hnatushenko, V. V. (2016). Informatsiina tekhnolohiia pidvyshchennia informatyvnosti bahatokanalnykh danykh na osnovi paketnykh veivlet-peretvoren [Information technology for enhancing the informativeness of multichannel data based on packet wavelet transforms]. *Kompiuterno-intehrovani tekhnolohii: osvita, nauka, vyrobnytstvo* [Computer-Integrated Technologies: Education, Science, Production], 77–83 [in Ukrainian].

15. Kashtan, V. Yu., & Hnatushenko, V. V. (2017). Tekhnolohiia heometrychnoi ta spektralnoi korektsii optyko-elektronnykh kosmichnykh znimkiv [Technology of geometric and spectral correction of optical-electronic satellite images]. *Visnyk KhNTU,* 3(62), 286–291 [in Ukrainian].

16. Polyakov, M. (2024). Novi mozhlyvosti suputnykovykh sensoriv: yak zminylasia EOS-industriia za mynuli roky [New capabilities of satellite sensors: How the EOS industry has changed over the past years]. Retrieved from: https://maxpolyakov.com/ua/novi-mozhlivosti-suputnikovikh-sensoriv-yak-zminilasya-eos-industriya-za-minuli-roki/ [in Ukrainian].

17. Shevchenko, V. Yu., Hnatushenko, V. V., & Kavats, O. O. (2015). Pidvyshchennia prostorovoho rozriznennia bahatokanalnykh aerokosmichnykh zobrazhen vysokoi prostorovoi rozriznennia na osnovi hipersferychnoho peretvorennia [Enhancing spatial resolution of high-resolution multichannel aerospace images using hyperspherical transformation]. *Naukovyi zhurnal Zaporizkoho natsionalnoho tekhnichnoho universytetu, radioelektronika, informatyka, upravlinnia* [Scientific Journal of Zaporizhzhia National Technical University, Radioelectronics, Informatics, Control], *1*(32), 73–79 [in Ukrainian].

УДК 534.2:004.94
DOI https://doi.org/10.32782/IT/2024-4-20

Сергій ПРИХОДЧЕНКО
кандидат технічних наук, доцент, доцент кафедри програмного забезпечення комп'ютерних систем, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького 19, Дніпро, Україна, 49005
ORCID: *0000-0002-6562-0601*
Scopus-Author ID: *57201251400*

Лариса КОРЯШКІНА
доктор технічних наук, доцент, професор кафедри системного аналізу та управління, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького 19, Дніпро, Україна, 49005
ORCID: *0000-0001-6423-092X*
Scopus-Author ID: *55844269100*

Ольга ШЕВЦОВА
PhD, доцент кафедри програмного забезпечення комп'ютерних систем, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, м. Дніпро, Україна, 49005
ORCID: *0000-0002-0148-5877*
Scopus Author ID: *57220267804*

Євгеній ЗАГИНАЙЛО
аспірант кафедри програмного забезпечення комп'ютерних систем, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, м. Дніпро, Україна, 49005
ORCID: *0009-0004-5109-2053*

Бібліографічний опис статті: Приходченко, С., Коряшкіна, Л., Шевцова, О., Загинайло, Є. (2024). Концептуальна модель архітектури програмного забезпечення обробки багатоджерельних акустичних сигналів. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 169–177, doi: https://doi.org/10.32782/IT/2024-4-20

КОНЦЕПТУАЛЬНА МОДЕЛЬ АРХІТЕКТУРИ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ОБРОБКИ БАГАТОДЖЕРЕЛЬНИХ АКУСТИЧНИХ СИГНАЛІВ

Робота присвячена розробці програмного забезпечення для моделювання поширення звукових хвиль у просторах з багатьма аудіо джерелами.

Метою *є створення гнучкої та масштабованої програмної системи для моделювання акустичних хвиль, що поєднує чисельні методи, оптимізацію обчислень і сучасну візуалізацію, надаючи ефективний інструмент для аналізу складних акустичних процесів у різних середовищах.*

Методологія. *Запропоновано архітектуру програмного забезпечення симуляції звукових хвиль, яка базується на модульному підході та використовує методи скінченних і граничних елементів для розв'язання рівнянь акустики, механізми підтримки паралельних обчислень, візуалізацію результатів симуляції.*

Наукова новизна *полягає у розробці комплексного підходу щодо створення програмного забезпечення для симуляції звукових хвиль, використовуючи обчислювальні методи і алгоритми, які сумісно розв'язують рівняння акустики і рівняння, GPU-прискорення та багатопотокову обробку, що дозволяє досягти високої точності і ступені достовірності моделювання та оптимізації обчислювальних ресурсів.*

Висновки. *В розробленій моделі програмного забезпечення передбачені модуль візуалізації, який забезпечує тривимірну інтерпретацію результатів, і модуль збереження даних, що слугує для обробки великих обсягів інформації. Запропонована модель дозволяє досліджувати звукові коливання, та моделювати звукові хвилі з декількох джерел, що робить її корисною для вирішення широкого спектра задач, зокрема проектування акустичних середовищ і аудіотехнологій.*

Ключові слова: *комп'ютерна симуляція, акустичне моделювання, багатоджерельний звук, архітектура програмного забезпечення.*

Serhii PRYKHODCHENKO
PhD, Associate Professor, Associate Professor at the Department of Computer Systems Software, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, prykhodchenko.s.d@nmu.one
ORCID: *0000-0002-6562-0601*
Scopus-Author ID: *57201251400*

Larysa KORIASHKINA
Doctor of Technical Sciences, Associate Professor, Professor at the Department of System Analysis and Control, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, koriashkina.l.s@nmu.one
ORCID: *0000-0001-6423-092X*
Scopus-Author ID: *55844269100*

Olha SHEVTSOVA
PhD, Associate Professor at the Department of Computer Systems Software, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, shevtsova.o.s@nmu.one
ORCID: *0000-0002-0148-5877*
Scopus Author ID: *57220267804*

Yevhenii ZAHYNAILO
Postgraduate Student at the Department of Computer Systems Software, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, Zahynailo.Yev.O@nmu.one
ORCID: *0009-0004-5109-2053*

To cite this article: Prykhodchenko, S., Koriashkina, L., Shevtsova, O., Zahynailo, Ye. (2024). Kontseptualna model arkhitektury prohramnoho zabezpechennia obrobky bahatodzherelnykh akustychnykh syhnaliv [Conceptual model of the software architecture for processing multisource acoustic signals]. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 169–177, doi: https://doi.org/10.32782/IT/2024-4-20

CONCEPTUAL MODEL OF THE SOFTWARE ARCHITECTURE FOR PROCESSING MULTI-SOURCE ACOUSTIC SIGNALS

The paper is devoted to the development of software for modeling the propagation of sound waves in spaces with multiple audio sources.

The objective *is to create a flexible and scalable software system for acoustic wave modeling that combines numerical methods, computational optimization, and modern visualization, providing an efficient tool for analyzing complex acoustic processes in various environments.*

Methodology. *The proposed software architecture for sound wave simulation is based on a modular approach and employs finite element and boundary element methods to solve acoustic equations, mechanisms for supporting parallel computations, and simulation result visualization.*

Scientific Novelty. *The scientific novelty lies in developing a comprehensive approach to creating software for sound wave simulation using computational methods and algorithms that jointly solve acoustic equations and related equations, GPU acceleration, and multi-threaded processing. This allows achieving high accuracy and reliability in modeling and optimizing computational resources.*

Conclusions. *The developed software model includes a visualization module that ensures three-dimensional interpretation of results and a data storage module for processing large volumes of information. The proposed model enables the study of sound oscillations and the simulation of sound waves from multiple sources, making it useful for solving a wide range of tasks, including the design of acoustic environments and audio technologies.*

Key words: *Computer simulation, Acoustic modelling, Multi-source sound, Software architecture.*

Вступ. Комп’ютерна симуляція звукових хвиль є надзвичайно важливою галуззю, яка поєднує фізику, математику та інформатику для вирішення складних задач моделювання звуку. Дана робота присвячена проектуванню програмного забезпечення (ПЗ) для симуляції процесів поширення, відбивання, заломлювання та поглинанняя звукових хвиль у матеріалах з різними фізичними властивостями. Відомо, що розповсюдження звуку у рідкому та газоподібному середовищах характеризується наявністю конвективних та диссипативних ефектів. Тому особлива увага під час створення ПЗ приділена розробці обчислювальних методів і алгоритмів,

які розв'язують сумісно рівняння акустики і рівняння, що описують динамику середовища. Ключовими аспектами є висока точність і ступінь достовірності моделювання та оптимізація обчислювальних ресурсів. Сучасні методи, такі як кінцеві елементи (FEM), методи граничних елементів (BEM) та дискретно-часові методи, дозволяють вирішувати проблеми в тривимірному просторі з високою деталізацією. Зокрема, розвиток GPU-прискорених обчислень значно покращив можливості симуляцій у реальному часі.

Обчислювальна симуляція звуку охоплює широкий спектр практичних застосувань. В акустиці приміщень симуляції використовують для проектування залів із ідеальною акустикою або шумовими бар'єрами. У медичній галузі на основі комп'ютерного моделювання вдосконалюються ультразвукові технології діагностики та терапевтичні методи. У сфері аудіотехнологій обчислювальна симуляція допомагає створювати реалістичні звукові ефекти у віртуальних середовищах, наприклад у відеоіграх чи кіно.

Метою роботи є створення гнучкої та масштабованої програмної системи для моделювання акустичних хвиль, що поєднує чисельні методи, оптимізацію обчислень і сучасну візуалізацію, надаючи ефективний інструмент для аналізу складних акустичних процесів у різних середовищах.

Об'єктом дослідження є архітектура програмного забезпечення для моделювання акустичних хвиль, зокрема методи оптимізації обчислювальних ресурсів, алгоритми чисельного аналізу та інтеграція інструментів візуалізації. Особлива увага приділяється розподіленим і паралельним обчисленням для складних задач.

Предметом дослідження є методи і алгоритми чисельного моделювання акустичних хвиль у багатоджерельних середовищах, програмна архітектура, що забезпечує оптимізацію обчислювальних процесів і ефективну обробку великих обсягів даних під час розв'язання фізичних рівнянь.

Постановка проблеми

Моделювання звукових хвиль на комп'ютері стикається з кількома суттєвими викликами, які потребують уваги дослідників та інженерів. По-перше, чисельна стабільність і точність. Звукові хвилі описуються складними математичними рівняннями, такими як рівняння хвиль, які можуть бути чутливими до вибору чисельних методів та параметрів моделювання. Неправильний вибір дискретизації або методу розв'язання може призвести до чисельної нестабільності, що в свою чергу вплине на точність результатів.

Відтак, потрібно розробити комплексний підхід щодо успішного моделювання звукових хвиль, включаючи вдосконалення чисельних методів, оптимізацію обчислювальних процесів і інтеграцію нових технологій для підвищення точності та ефективності симуляцій.

Літературний огляд

Математичні моделі та методи, що використовуються в системі обчислювальної симуляції звукових хвиль, ґрунтуються на фізичних законах поширення звуку в середовищі.

Основою є хвильове рівняння, яке описує просторово-часову динаміку звукових хвиль у середовищах із різними фізичними властивостями. У випадках, коли необхідно враховувати більше деталей, використовуються узагальнені рівняння, такі як рівняння Гельмгольца для гармонічних хвиль або рівняння Нав'є–Стокса.

Рівняння Гельмгольца може допомогти у вирішенні питань акустики, сейсмології, електромагнітної радіації (Juraev et al., 2024). Рівняння Гельмгольца є рівнозначним хвильовому рівнянню за умови незмінної частоти.

Поширення звукових хвиль в рідкому або газоподібному середовищі уже включене в рівняння Нав'є–Стокса, яке описує динаміку руху рідини або газу (Sacasa-Cespedes, 2024) і дозволяє визначити основні параметри течії.

Вибір чисельного методу інтегрування вказаних рівнянь залежить від вимог щодо точності та продуктивності. Приміром, розглянута в (Rosen, Godin & Raghuvanshi, 2020) модель хвильового акустичного поширення, використовує метод низькочастотних скінченних різниць у симуляції зі статичною сценою задля моделювання динамічного поширення акустичної хвилі у просторі. Такий метод заснований на дискретизації рівнянь у просторі та часі та має можливість покращення завдяки паралельним обчисленням.

У випадках складнішої геометрії чи в неоднорідних середовищах, як, наприклад, в комплексному підході до віброакустичного аналізу (Citarella et al., 2007), перевагу надають методу скінченних елементів (FEM). Цей метод дозволяє враховувати нерівномірний розподіл властивостей матеріалів задля аналізу структурної динаміки елементів, що в свою чергу дає можливість передбачити і проконтролювати потенційні проблеми в конструкції на етапі створення дизайну.

Метод граничних елементів (BEM) ефективний для задач, де потрібно моделювати звукові

хвилі в необмежених середовищах, або в задачах моделювання аероакустики, біоакустики або, навіть, звукової візуалізації в комп'ютерній анімації (Xun et al., 2008). Для прискорення обчислень застосовуються адаптивні методи сіток, які динамічно змінюють розподіл точок дискретизації залежно від локальної складності звукового поля. Це дозволяє зосередити ресурси в областях із високою інтенсивністю хвиль або складною взаємодією, наприклад, поблизу перешкод чи в зонах заломлення. До прикладу наведемо блоко-структурний адаптивний метод сіток. Ціллю даного методу є створення паралельного алгоритму адаптивних сіток, що слугує підвищенню обчислювальної ефективності та зберіганню в рамках задачі прогнозування шуму аеродвигуна. Натомість одним з висновків даного методу було те, що при збільшенні кількості процесорів обчислювальна продуктивність падала через необхідність балансування обчислювального навантаження всієї адаптивної сітки. А, отже, покращення існуючих методів шляхом паралельних обчислень мають бути протестовані на різних конфігураціях системи задля агрументації оцінки ефективності.

Розглядаючи інноваційні підходи до локалізації кількох джерел звуку у двовимірному просторі, робота (Le Moing et al., 2020) використовує глибокі нейронні мережі. Автори пропонують вдосконалену архітектуру «енкодер-декодер», яка включає роздільну обробку даних з різних масивів мікрофонів і використання пар мікрофонів для покращення характеристик моделі. Впроваджено два нових методи представлення результатів: теплові карти і вдосконалену сіткову модель, що забезпечують вищу точність локалізації навіть за наявності обмежених навчальних даних. Результати експериментів, проведених на синтетичних і реальних даних, демонструють перевагу цього підходу порівняно з існуючими методами, хоча адаптація до реального світу все ще вимагає більшого обсягу навчальних даних.

Методика Competitive K-means Clustering (Lee & Choi, 2010) використовується для локалізації звуків з декількох джерел, забезпечуючи поділ акустичних сигналів у складних звукових сценах. Алгоритм ефективно групує звукові дані, покращуючи точність визначення місця розташування звукових джерел завдяки конкурентному підходу. Експерименти показують високу стійкість моделі навіть в умовах шумових перешкод. Це робить підхід придатним для застосування в робототехніці, аудіосистемах та інших завданнях із багатозадачною обробкою звуку.

Метод локалізації джерела звуку за статистичним зрівнюванням джерел (Gao et al., 2019) використовує статистичну врівноваженість компонентів джерел для локалізації звуку в багатокомпонентних акустичних середовищах. Підхід ґрунтується на поділі сигналів з використанням статистичних характеристик, таких як енергія і кореляція, щоб поліпшити точність локалізації. Модель показує високу ефективність у складних звукових сценах, особливо за наявності звуків, що перекриваються. Це робить метод перспективним для систем аудіомоніторингу та робототехніки в шумних середовищах.

Метод локалізації багатоджерельного звукозапису (Jia, 2018) ґрунтується на покроковому видаленні компонентів домінуючого джерела для аналізу залишкових сигналів. Цей підхід дає змогу поетапно виділяти менш виражені джерела звуку, забезпечуючи більш точне їх визначення в зашумлених і багатокомпонентних середовищах. Використовуючи статистичний аналіз домінуючих компонентів, модель ефективно зменшує взаємні перешкоди між джерелами. Цей підхід перспективний для аудіомоніторингу, систем розпізнавання та роботизованих додатків. Одним з прикладів такого програмного забезпечення є застосунок, що використовувався в роботі (Wang, 2006).

Матеріали та методи. Архітектурна модель програмного забезпечення для системи обчислювальної симуляції звукових хвиль побудована за принципами багаторівневої модульної організації, що забезпечує високу гнучкість, ефективність та простоту в обслуговуванні. UML-діаграма компонентів розробленої системи комп'ютерного моделювання звукових хвиль представлена на рис. 1. У центрі системи знаходиться обчислювальний модуль, який відповідає за чисельне моделювання фізичних процесів, пов'язаних із поширенням звукових хвиль. Цей компонент реалізує алгоритми, такі як метод скінченних або граничних елементів, і включає механізми для дискретизації геометрії середовища, моделювання часової динаміки та врахування крайових умов.

Для взаємодії з користувачами передбачено модуль управління параметрами симуляції, який дозволяє налаштовувати фізичні властивості середовища, задавати початкові умови та конфігурувати джерела звуку. Інтерфейс цього модуля підтримує як текстові конфігураційні файли, так і інтеграцію з графічними інтерфейсами, якщо такі є частиною системи.

Значна увага приділена модулю зберігання даних, який оптимізований для роботи з великими обсягами інформації, що утворюються під

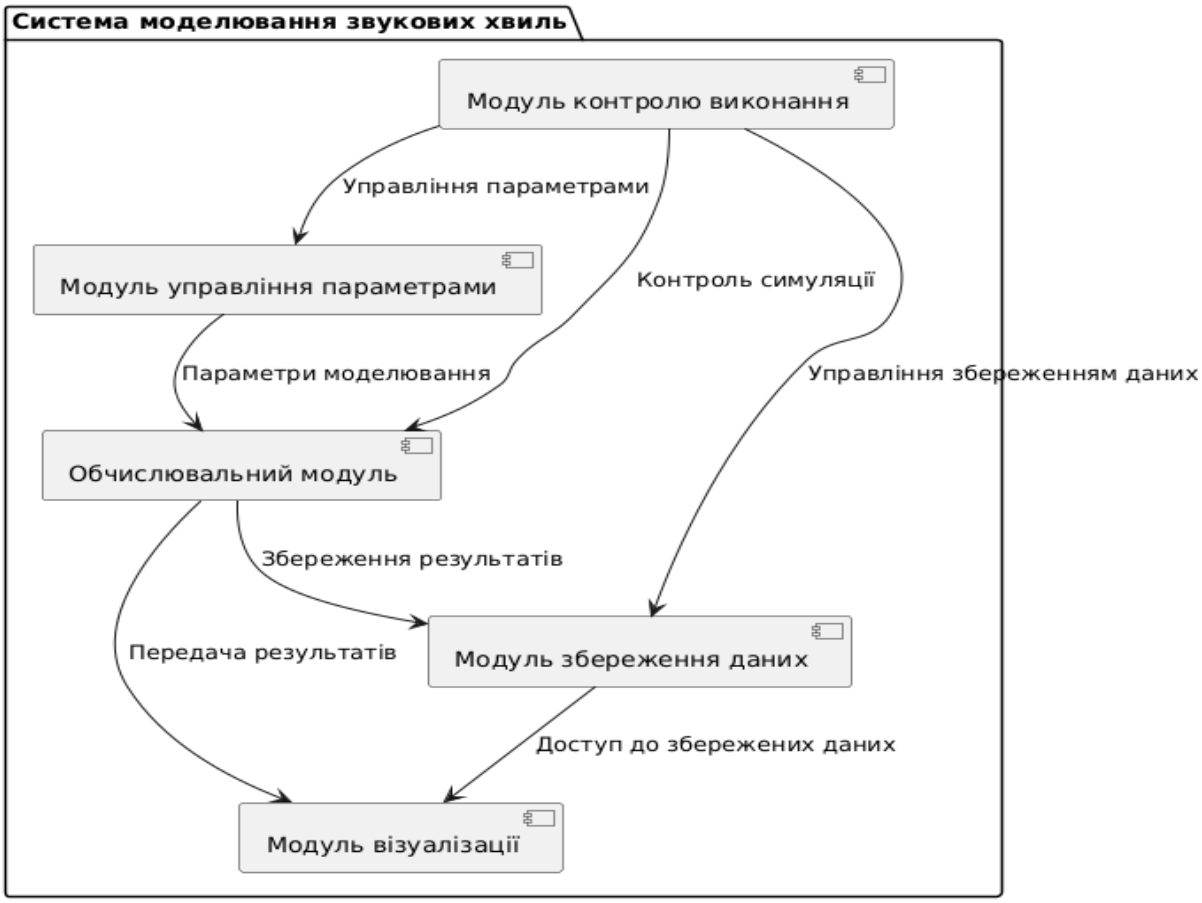


Рис. 1. UML-діаграма компонентів

час моделювання. Цей модуль забезпечує ефективне збереження результатів у форматах, придатних для багатопотокового доступу, таких як HDF5 або NetCDF, і надає засоби для фільтрації та попередньої обробки даних, які згодом можуть бути використані для аналізу чи візуалізації.

Візуалізація результатів симуляції реалізована через модуль, що дозволяє створювати статичні зображення, анімації та інтерактивні тривимірні сцени. Основою для цього є графічні бібліотеки, наприклад, VTK або OpenGL, які забезпечують масштабованість залежно від обсягу оброблюваних даних і необхідного рівня деталізації.

Система також включає механізми оптимізації продуктивності через підтримку паралельних обчислень, що реалізуються за допомогою багатопотокової обробки CPU або GPU-прискорення. Ця функціональність інтегрована в обчислювальний модуль, що дозволяє адаптуватися до різних обчислювальних середовищ – від суперкомп'ютерів до стандартних робочих станцій.

UML Sequence diagram розробленої системи обчислювальної симуляції звуку представлена на рис. 2.

На верхньому рівні архітектури знаходиться модуль контролю виконання, який координує роботу всіх компонентів, розподіляє ресурси і забезпечує обробку виняткових ситуацій. Він відповідає за стабільність системи, перевірку введених даних на коректність і ведення журналів подій для діагностики. Завдяки модульній структурі ця система легко розширюється і адаптується до нових задач без значних змін у базовій архітектурі, що робить її зручною для розробників і дослідників.

Обчислювальний модуль для моделювання поширення звукових хвиль організований як система класів, кожен з яких відповідає за певний аспект симуляції. Центральним елементом є клас Solver, який реалізує основні чисельні методи, такі як метод скінченних елементів, скінченних різниць або граничних елементів. Він забезпечує дискретизацію рівнянь, розрахунок часових кроків і контроль стабільності обчислень.

Геометрія моделюваного середовища представлена через клас Mesh, що зберігає інформацію про вузли, елементи та їхні зв'язки. Цей підхід дозволяє працювати з нерівномірними чи адаптивними сітками. Фізичні

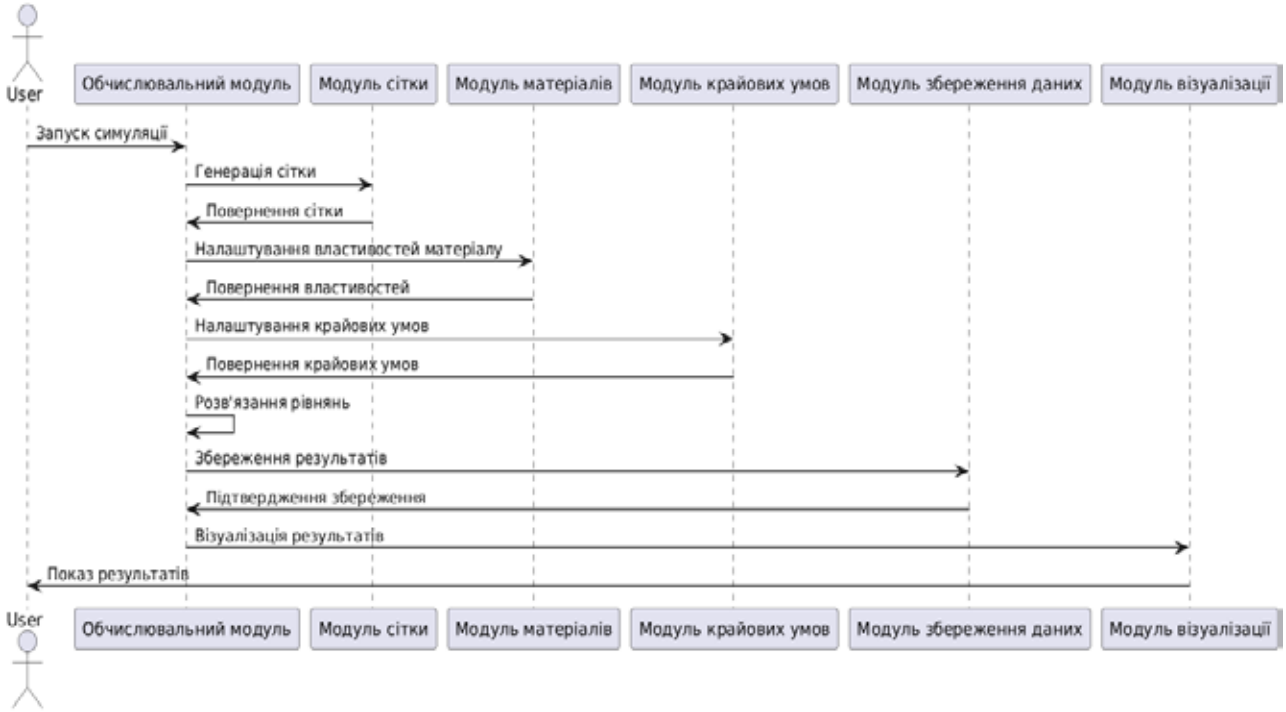


Рис. 2. UML Sequence diagram

характеристики матеріалів, зокрема густина, швидкість звуку та коефіцієнти поглинання, управляються класом MaterialProperties, який забезпечує доступ до цих даних для інших компонентів системи.

Крайові умови моделі визначаються класом BoundaryConditions, який інтегрується з Solver для встановлення фіксованих значень, відбиття чи поглинання хвиль на межах середовища. Для підвищення продуктивності система використовує клас ParallelComputingManager, що забезпечує паралельні обчислення через багатопоточну обробку або GPU-прискорення. Цей клас розподіляє задачі між обчислювальними ресурсами й оптимізує обмін даними між процесами.

Результати симуляції структуруються класом SimulationData, який забезпечує зручний доступ до вихідних даних, їхнє збереження і візуалізацію. Крім того, цей клас підтримує інтерфейси для проміжного аналізу, що дозволяє здійснювати моніторинг процесу обчислень.

Для перевірки коректності моделі використовується клас ValidationTools, який тестує результати на основі контрольних задач і забезпечує їх порівняння з аналітичними чи експериментальними даними. Цей підхід дозволяє гарантувати точність і надійність чисельних розв'язків у системі.

Програмне забезпечення для обчислювальної симуляції звукових хвиль є складною багаторівневою системою, структура якої базується на принципах модульності, масштабованості та чіткого розподілу відповідальності між компонентами. Компонентна структура цього програмного забезпечення складається з взаємопов'язаних модулів, кожен із яких має чітко визначену роль у загальному процесі моделювання. Ключовою метою такої архітектури є забезпечення ефективності, гнучкості та надійності системи.

Центральне місце займає обчислювальний модуль, який реалізує чисельне розв'язання рівнянь, що описують поширення звукових хвиль. Цей модуль є ядром системи, яке безпосередньо виконує складні математичні операції з урахуванням заданих фізичних параметрів. Він складається з кількох підкомпонентів, серед яких розрахунковий блок для обчислення значень на кожному етапі симуляції, блок інтеграції часових і просторових параметрів, а також механізми оптимізації для ефективного використання апаратних ресурсів.

Модуль управління параметрами забезпечує інтерфейс для взаємодії з обчислювальним ядром, приймаючи від користувача початкові дані та передаючи їх у форматі, який адаптований для чисельних методів. Він виконує перевірку вхідних параметрів, їхню нормалізацію та, за необхідності, автоматичну оптимізацію. Цей модуль також включає підсистему для збереження наборів параметрів, що дозволяє

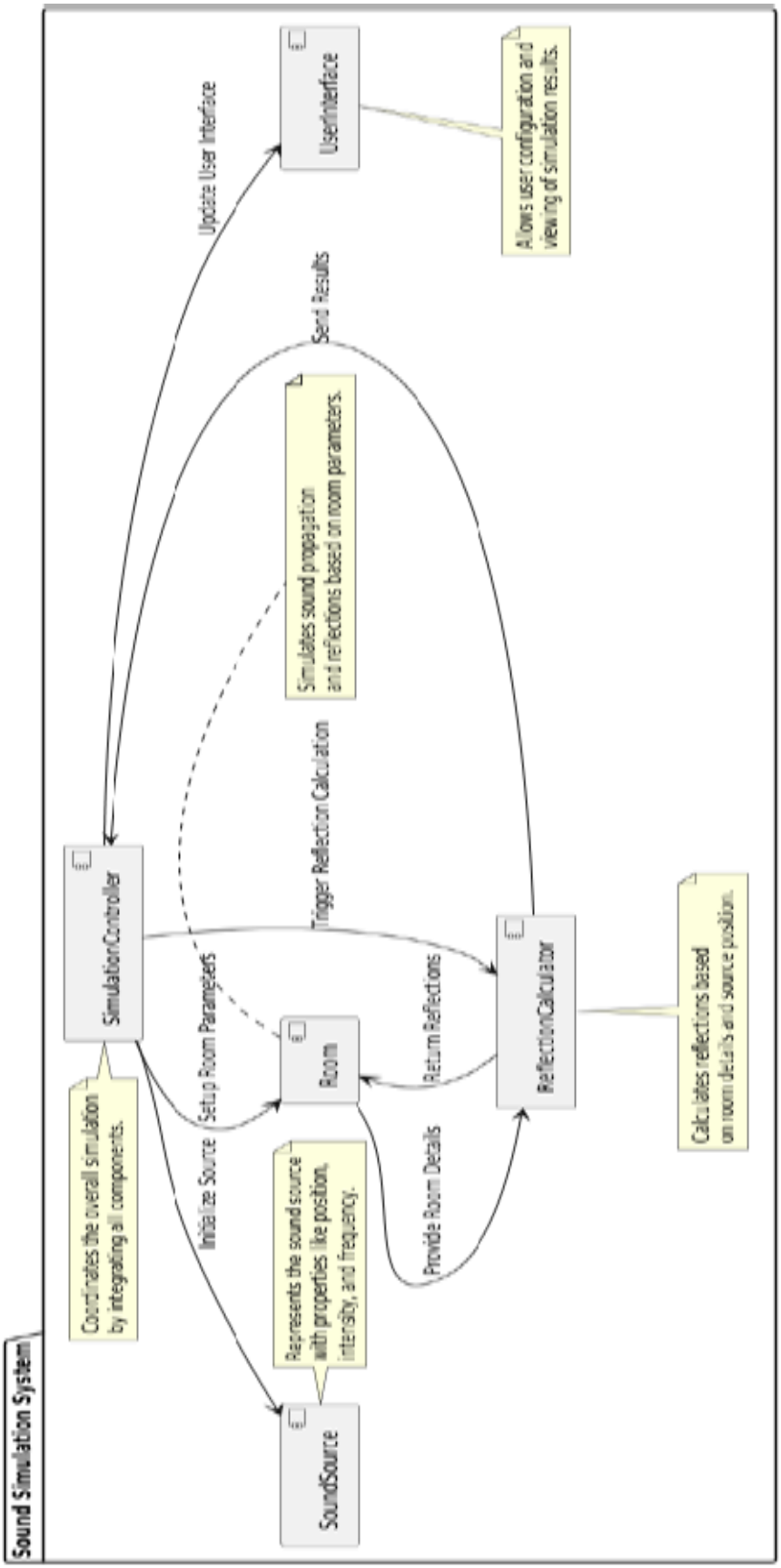


Рис. 3. Діаграма компонентів UML для розробленого інформаційної системи

повторно використовувати типові конфігурації для швидкого запуску симуляцій.

Модуль візуалізації призначений для представлення результатів обчислень у зручній для користувача формі. Цей компонент приймає дані, отримані від обчислювального модуля, і трансформує їх у графічні чи текстові формати. Він підтримує різні типи візуалізації, такі як двовимірні графіки, тривимірні моделі та анімації. Особлива увага приділяється інтерактивності, що дозволяє користувачеві змінювати параметри відображення в реальному часі для детальнішого аналізу результатів.

Модуль збереження даних забезпечує довготривале зберігання результатів симуляції, параметрів, а також журналів виконання завдань. Він базується на реляційній базі даних, де інформація організована у вигляді взаємопов'язаних таблиць. Цей модуль також відповідає за обробку запитів на доступ до даних, забезпечуючи ефективний пошук і експорт результатів у потрібному форматі. Окрім того, він інтегрується з механізмами резервного копіювання для захисту від втрати даних.

Модуль контролю виконання відіграє роль координатора роботи системи. Він управляє чергою завдань, забезпечуючи послідовність їх виконання відповідно до заданих пріоритетів. Цей модуль реалізує механізми моніторингу стану системи в реальному часі, що дозволяє виявляти та виправляти можливі збої. Окрім цього, він відповідає за логування процесів, що є критично важливим для відтворення сценаріїв у разі виникнення помилок.

Система має інтеграційний шар, який забезпечує взаємодію між її компонентами. Комунікація між модулями реалізована за допомогою внутрішнього API, що дозволяє досягти високої швидкості обміну даними та мінімізувати ризики конфліктів. Інтеграційний шар також відповідає за управління потоками даних між модулями, забезпечуючи їхню синхронізацію та узгодженість. Загалом, така компонентна структура дозволяє гнучко адаптувати програмне забезпечення до потреб різних користувачів, розширювати функціональність шляхом додавання нових модулів і гарантувати стабільну роботу навіть за умов значного навантаження.

Схема модулів розробленої програми представлена на рис. 3.

Програмне забезпечення, створене для симуляції звукових хвиль, є результатом комплексного поєднання математичного моделювання, архітектурного планування та передових програмних технологій. Цей інструмент розроблено для точного розрахунку, відображення та аналізу процесів поширення звуку в замкнених або відкритих середовищах. Робота програмного забезпечення базується на тісній інтеграції кількох ключових модулів, кожен з яких виконує конкретну роль, забезпечуючи надійну та послідовну роботу всієї системи.

Висновки. Представлена концепуальна модель програмного забезпечення для обчислювальної симуляції звукових хвиль. Створене програмне забезпечення є високотехнологічною системою, що поєднує точні математичні моделі, ефективні алгоритми та сучасні методи візуалізації. Воно надає дослідникам і інженерам інструмент для аналізу та оптимізації акустичних середовищ, дозволяючи моделювати ситуації, які складно або неможливо відтворити у реальних умовах. Завдяки модульному підходу система є гнучкою, масштабованою та придатною для використання у широкому спектрі прикладних задач, від архітектурного акустичного проєктування до досліджень в області фізики звуку.

ЛІТЕРАТУРА:

1. Juraev D., Agarwal P., Elsayed E., Nauryz T. Helmholtz equations and their applications in solving physical problems. 2024. 4. 54–64.

2. Sacasa-Cespedes S. A Geometric Approach to the Navier-Stokes Equations. 2024. 10.48550/arXiv.2411.18724.

3. Rosen M., Godin K. W., Raghuvanshi N. Interactive sound propagation for dynamic scenes using 2D wave simulation. *Computer Graphics Forum*, 2020. 39, 39–46. https://doi.org/10.1111/cgf.14099.

4. Citarella R., Federico L., Cicatiello A. Modal acoustic transfer vector approach in a FEM–BEM vibro-acoustic analysis, *Engineering Analysis with Boundary Elements*, Volume 31, Issue 3, 2007. Pages 248–258, ISSN 0955-7997, https://doi.org/10.1016/j.enganabound.2006.09.004.

5. Yijun L. On the BEM for acoustic wave problems, *Engineering Analysis with Boundary Elements,* 2019. Volume 107, Pages 53–62, ISSN 0955-7997, https://doi.org/10.1016/j.enganabound.2019.07.002.

6. Xun H., Xin Z., Simon K. R. Adaptive mesh refinement computation of acoustic radiation from an engine intake, *Aerospace Science and Technology*, 2008. Vol. 12, Issue 5, 418–426, ISSN 1270-9638, https://doi.org/10.1016/j.ast.2007.09.004.

7. Le Moing G., Vinayavekhin P., Inoue T., Vongkulbhisal J., Munawar A., Tachibana R. Agravante D. Learning Multiple Sound Source 2D Localization. 2020. 10.48550/arXiv.2012.05515.

8. Lee B. Choi J. Multi-source sound localization using the competitive k-means clustering. 2010. 1–7. 10.1109/ETFA.2010.5641169.

9. Gao Sh. Jia M., Wu Yu. Jia Yi., Mingchen W. Multiple Sound Sources Localization by using Statistical Source Component Equalization. 2019. 320–325. 10.1145/3373509.3373582.

10. Jia M. Wu Yu., Bao Ch., Wang, J. Multiple Sound Sources Localization with Frame-by-Frame Component Removal of Statistically Dominant Source. *Sensors.* 2018. 18. 3613. 10.3390/s18113613.

11. Wang M., Freund J. B., Lele, S. K. COMPUTATIONAL PREDICTION OF FLOW-GENERATED SOUND. *Annual Review of Fluid Mechanics,* 2006. 38, 483–512.

REFERENCES:

1. Juraev, D., Agarwal, P., Elsayed, E. & Nauryz, T. (2024). Helmholtz equations and their applications in solving physical problems. 4. 54–64.

2. Sacasa-Cespedes, S. (2024). A Geometric Approach to the Navier-Stokes Equations. 10.48550/arXiv.2411.18724.

3. Rosen, M., Godin, K. W. & Raghuvanshi, N. (2020). Interactive sound propagation for dynamic scenes using 2D wave simulation. Computer Graphics Forum, 39: 39-46. https://doi.org/10.1111/cgf.14099.

4. Citarella, R., Federico, L. & Cicatiello, A. (2007). Modal acoustic transfer vector approach in a FEM–BEM vibro-acoustic analysis, *Engineering Analysis with Boundary Elements*, Volume 31, Issue 3, Pages 248–258, ISSN 0955-7997, https://doi.org/10.1016/j.enganabound.2006.09.004.

5. Yijun, L. (2019). On the BEM for acoustic wave problems, Engineering *Analysis with Boundary Elements*, Volume 107, Pages 53–62, ISSN 0955-7997, https://doi.org/10.1016/j.enganabound.2019.07.002.

6. Xun, H., Xin, Z. & Simon K. R. (2008). Adaptive mesh refinement computation of acoustic radiation from an engine intake, *Aerospace Science and Technology*, Vol. 12, Issue 5, 418-426, ISSN 1270-9638, https://doi.org/10.1016/j.ast.2007.09.004.

7. Le Moing, G., Vinayavekhin, P., Inoue, T., Vongkulbhisal, J., Munawar, A., Tachibana, R. & Agravante, D. (2020). Learning Multiple Sound Source 2D Localization. 10.48550/arXiv.2012.05515.

8. Lee, B. & Choi, J. (2010). Multi-source sound localization using the competitive k-means clustering. 1–7. 10.1109/ETFA.2010.5641169.

9. Gao, Sh. Jia, M., Wu, Yu. Jia, Yi. & Mingchen, W. (2019). Multiple Sound Sources Localization by using Statistical Source Component Equalization. 320–325. 10.1145/3373509.3373582.

10. Jia, M. Wu, Yu., Bao, Ch. & Wang, J. (2018). Multiple Sound Sources Localization with Frame-by-Frame Component Removal of Statistically Dominant Source. Sensors. 18. 3613. 10.3390/s18113613.

11. Wang, M., Freund, J. B., & Lele, S. K. (2006). COMPUTATIONAL PREDICTION OF FLOW-GENERATED SOUND. Annual Review of Fluid Mechanics, 38, 483–512.

УДК 004.49:004.75
DOI https://doi.org/10.32782/IT/2024-4-21

Павло РЕГІДА
старший викладач кафедри комп'ютерної інженерії та інформаційних систем, Хмельницький національний університет, вул. Інститутська 11, м. Хмельницький, Україна, 29016
ORCID: *0000-0002-6591-7069*
Scopus Author ID: *57202317133*

Олег САВЕНКО
доктор технічних наук, професор, професор кафедри комп'ютерної інженерії та інформаційних систем, Хмельницький національний університет, вул. Інститутська 11, м. Хмельницький, Україна, 29016
ORCID: *0000-0002-4104-745X*
Scopus Author ID: *54421023400*

Бібліографічний опис статті: Регіда, П., Савенко, О. (2024). Метод виявлення зловмисної активності в інфікованих програмах. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 4, 178–186, doi: https://doi.org/10.32782/IT/2024-4-21

МЕТОД ВИЯВЛЕННЯ ЗЛОВМИСНОЇ АКТИВНОСТІ В ІНФІКОВАНИХ ПРОГРАМАХ

Розробники зловмисного програмного забезпечення (ЗПЗ) застосовують різноманітні техніки маскування, що дозволяють уникнути виявленню під час їх виконання на хості користувача. Серед ключових класів таких технік визначають зокрема: анти-емуляційні та поліморфні, через їх можливість протидіяти гібридному методу виявлення, який часто застосовується в сучасних АПЗ. Саме тому, дослідження методів маскування, їх комбінування та загальний вплив на хост корчувача дозволить сформувати метод їх виявлення. В зв'язку із цим, буде доречно застосувати технології емуляції, пісочниці та розподілених систем.

Метою статті *є виявлення зловмисної поведінки в інфікованих програмах із техніками уникнення виявлення на основі аналізу зміни поведінки їх виконання в модифікованих ізольованих середовищах.*

Методологія *полягає у застосуванні наукових методів: синтез, аналіз та порівняння. В роботі представлений аналіз частини сучасних методів уникнення виявлення, що застосовуються в інфікованих програмах. Розглянуто засіб виконання інфікованої програми, та архітектуру системи для організації розподіленого виявлення зловмисної поведінки. Представлено алгоритм формування поведінки виконання програми в ізольованому середовищі.*

Наукова новизна *полягає у формуванні методу виявлення зловмисного прояву у інфікованих програмах, що застосовують захисні техніки протидії емуляції та поліморфізму. Приділено увагу стратегії виконання таких інфікованих програм в різних середовищах, що позначається на їх поведінці виконання. Представлені результати досліджень, що демонструють ефективність запропонованого методу в рамках досліджених технік уникнення виявлення.*

Висновки. *Запропонований метод дозволяє покращити виявлення інфікованих програм, які використовують поліморфні техніки перетворення, для уникнення виявлення в ситуації коли емульоване середовище нею було визначено. Представлено архітектуру системи, що забезпечує множину модифікованих ізольованих середовищ для проведення аналізу виконання програм.*

Ключові слова: *виявлення вірусів, емуляція, пісочниця, маскувальні техніки.*

Pavlo REHIDA
Senior Lecturer at the Department of Computer Engineering and Information Systems, Khmelnytskyi National University, 11, Instytuts'ka Str., Khmelnytskyi, Ukraine 29016, pavlo.rehida@gmail.com
ORCID: *0000-0002-6591-7069*
Scopus Author ID: *57202317133*

Oleg SAVENKO
Doctor of Technical Sciences, Professor, Professor at the Department of Computer Engineering and Information Systems, Khmelnytskyi National University, 11, Instytuts'ka Str., Khmelnytskyi, Ukraine, 29016, savenko_oleg_st@ukr.net
ORCID: *0000-0002-4104-745X*
Scopus Author ID: *54421023400*

To site this article: Rehida, P., Savenko, O. (2024). Metod vyiavlennia zlovmysnoi aktyvnosti v infikovanykh prohramakh. [Method for detecting malicious activity in infected programs]. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 178–186, doi: https://doi.org/10.32782/IT/2024-4-21

METHOD FOR DETECTING MALICIOUS ACTIVITY IN INFECTED PROGRAMS

Developers of malwares employ various avoid techniques to evade detection during execution on user host. Key classes of such techniques include anti-emulation and polymorphic techniques, due to their ability to counteract hybrid detection methods commonly used in modern antivirus programs. Therefore, studying avoid techniques, their combinations, and their impact on the host will help in developing detection methods. In this context, it is appropriate to use emulation technologies, sandboxes, and distributing systems.

The purpose *is to detect malicious activity in infected programs that use evasion techniques by analysing changes in their execution behaviour in modified isolated environments.*

The methodology *involves the use of scientific methods: synthesis, analysis, and comparison. The paper presents an analysis of modern evasion techniques in infected programs. It discusses the execution tool for infected programs and the system architecture for organizing distributed detection of malicious activity. An algorithm to form program execution behaviour in an isolated environment is also presented.*

The scientific novelty *lies in developing a method for detecting, malicious activity in infected programs that use anti-emulation and polymorphic techniques. Special attention is given to the strategy of executing such programs in different environments, which impacts their execution behaviour. The presented research results demonstrate the effectiveness of proposed method within the scope of studied evasion techniques.*

Conclusions. *The proposed method improves the detection of infected programs that use polymorphic transformation techniques to evade detection, particularly when the emulated environment has been identified. The architecture of the presented system provides a set of modified isolated environments for analysing program execution.*

Key words: *malware detection, emulation, sandbox, avoid techniques.*

Актуальність проблеми. Для вирішення поставної задачі виявлення наявності зловмисного прояву в інфікованій програмі запропоновано використати розподілений підхід із використанням різних модифікованих ізольованих середовищ. Основною відмінністю середовищ є застосування досліджених технік анти-емуляції в якості параметрів їх модифікації. Сучасне зловмисне програмне забезпечення використовує велику кількість технік уникнення від виявлення антивірусними програмними засобами. Тому, пропонується провести аналіз виконання програм в різних умовах, що дозволить виявити відмінність у виконанні, і тим самим підтвердити наявність зловмисного прояву.

Аналіз останніх досліджень і публікацій. Розробники сучасних екземплярів зловмисного програмного забезпечення досліджують техніки їх виявлення підходами, які використовуються в антивірусних програмних засобах. Основною метою таких досліджень, є формування нових стратегій для приховання зловмисного прояву. Аналіз звітів наукових лабораторій та наукових публікацій тематика яких є дослідження ЗПЗ, а також їх методів впливу на хост користувача показують, що як і кількість так і методи впливу із кожним роком все збільшуються (Razaulla, 2023; Savenko 2021). Метою створення таких ЗПЗ може бути і модифікація існуючого програмного забезпечення, яке застосовується отримання як фінансової вигоди, так і для отримання несанкціонованого доступу до хосту користувач (Savenko, 2020), та навіть отримання доступу до елементів кіберфізичних систем (Lysenko, 2022). Для того щоб вирішити цю проблему, фахівці кібербезпеки почали розробляти та застосовувати антивірусні програмні засоби (АПЗ), як основний засіб захисту хосту користувача. Враховуючи динаміку появи нових ЗПЗ та методів їх впливу, сучасні АПЗ постійно вдосконалюються шляхом аналізу нових загроз від ЗПЗ та впровадженням стратегій їх протидії.

В більшості сучасних настільних та мобільних операційних системах по замовчуванню використовуються вбудоване АПЗ (Numminen, 2023), яке використовує велику кількість підходів для виявлення ЗПЗ. В літературі, такі підходи узагальнюють у три основні методи виявлення: статичний (Omar, 2022) динамічний (Leon, 2021) та гібридний (Yunus, 2020). Окрім вбудованих методів, широко застосовуються і сторонні рішення, які відрізняються розширеним функціоналом, що представлені додатковими методами виявлення.

Методи виявлення сучасних АПЗ базуються на використанні гібридного методу виявлення. Динамічний аналіз здійснюється в ізольованому середовищі з емуляцією апаратних компонентів хоста, що дозволяє безпечно виконувати аналіз програм на наявність ЗПЗ та його

поведінки. Крім того, АПЗ активно використовують бази вже відомих загроз при застосуванні статичного аналізу. Щоб протидіяти таким методам виявлення, розробники ЗПЗ наділяють їх властивостями, що дозволяють підвищити шанс уникнення. Такі властивості представлені у вигляді технік, які базуються на досліджених методах виявлення. В літературі такі техніки, розділяються на категорії за методом аналізу або впливу на хост користувача. (Faruki, 2023)

Серед основних технік уникнення виявлення визначають техніки протидії емуляції, через те, що сучасні засоби виявлення базуються на використанні ізольованих середовищ із емуляцією. Саме тому, для ЗПЗ важливо виявити наявність емуляції в середовищі виконання, адже його наявність із високою імовірністю означає використання АПЗ. Такі техніки спрямовані для пошуку невідповідності у характеристиках виконуваного середовища у порівнянні із реальним середовищем користувача, які базуються на аналізі: швидкості виконання інструкцій процесора (D'Elia, 2020), ідентифікації апаратних засобів (Apostolopoulos, 2021), пошук артефактів емуляції (Nappa, 2021).

Ще однією важливою категорією таких технік визначають як поліморфні. За допомогою такого підходу, в окремих випадках, ЗПЗ може змінювати код виконання інфікованої програми. Такий підхід дозволяє оминути статичний метод (Nicheporuk, 2020) виявлення, який базується на порівнянні із уже відомими ЗПЗ, які знаходяться у відповідній базі АПЗ. Це відбувається завдяки використанню відповідних технік, які можуть змінювати код виконання програми, при цьому зберігаючи її функціональні властивості (Gorment, 2023). Зазвичай, для такої модифікації використовуються декілька технік одночасно, до яких входять: використання додаткового коду (Gibert, 2022), перепризначення регістру, заміни інструкцій (Han, 2022) та розділення коду.

Для використання ізольованого середовища з метою аналізу програм на наявність зловмисної активності, використовують технологію емуляції апаратних засобів хоста (Fedák, 2022). Це також дозволяє отримати більше можливостей в плані аналізу процесу виконання програми (Liu, 2022) та сформувати поведінку виконання програми. Дослідження поведінки, в свою чергу, дозволить визначити наявність зловмисної поведінки (Arabo, 2020) у програмі, що була проаналізована.

Аналіз виконання програми стає можливим із використанням за допомогою зчитування вмісту основних регістрів процесору (Schaik, 2021) за допомогою використання програмних переривань (Amit, 2015).

Успішне застосування розподілених систем у наукових та практичних завданнях (Peng, 2020) сприяло їх широкому розвитку. Особливу категорію таких систем становлять такі, що базуються на грід-технологіях (Hamdan, 2020). Хоча організація цих систем вимагає вирішення додаткових завдань, пов'язаних з гарантуванням безпеки (Kong, 2022; Rehida 2023) та коректності їхньої роботи, вони знаходять широке застосування у сучасних дослідженнях та розробках.

Мета дослідження. Метою дослідження є розробка методу виявлення інфікованих програм, що використовують техніки уникнення виявлення на основі аналізу виконання програми в модифікованих ізольованих середовищах.

Виклад основного матеріалу дослідження. Розробники ЗПЗ застосовують для виконання зловмисної дії використовують різні підходи, і одним і ключових напрямків є інфікування файлів. Інфіковане програмне забезпечення отримує додаткові функціональні властивості, як наприклад: виконати зловмисну дію. Розглянуті техніки уникнення виявлення, також стають частиною інфікованих програм. Тому, якщо представити набір технік уникнення емуляції як $M_{ae} = \{ae_1, ae_2 \ldots ae_n\}$, а техніки поліморфізму як $M_{pt} = \{pt_1, pt_1 \ldots pt_m,\}$. При цьому, кількість технік, які використовуються в рамках одного вірусу не може налічувати усі їх варіації, так як така суттєва модифікація оригінального програмного забезпечення (ПЗ) буде викликати підозру навіть зі сторони користувача. Нехай оригінальне ПЗ та ЗПЗ це *sw* та *mw* відповідно. Тоді інфіковану програму можна представити у наступному вигляді:

$$Isw_i = \{sw_i, mw_i, M_{ae_i}, M_{pt_i}\} \quad (1)$$

Так як мета будь-якого вірусу виконати, який є частиною інфікованої програми виконати свою зловмисну дію, він може змінювати власну стратегію виконання на основі результатів виконання цих стратегій. Тому, у випадку коли емуляція виявлена, інфіковане програмне може приховати свою зловмисну поведінку. І для того щоб виявити наявність зловмисної поведінки, пропонується дослідити поведінку виконання програми у різних модифікованих середовищах. Розглянуті раніше техніки модифікації коду з метою обходу виявлення, часто застосовують саме цю зміну на рівні інструкцій, що подаються на виконання в Центральний процесор (ЦП).

Для пошуку такої відмінності було розроблено базовий емулятор, який може опрацьовувати декілька десятків інструкцій низького

рівня різного типу, такі як: арифметичні, логічні, із плаваючої комою, строкові, потоків керування, роботи із даними та системні. З точки зору виконання таких інструкцій, компоненти ЦП можна розділити на активні та пасивні. Активними компонентами будемо визначати такі, стан яких змінюється при виконанні інструкцій. При виконання набору інструкцій, які собою представляють певну програму, поступово змінюється вміст регістрів. Тому, визначимо стан ЦП як множину набору значень в поточний момент виконання програми. І враховуючи, що в сучасних процесорах здебільшого використовуються наступні групи регістрів: основні регістри – $R_b = \{r_{RAX}, r_{RBX}, r_{RCX}, r_{RDX}\}$, регістри вказівники – $R_p = \{r_{RBP}, r_{RSP}\}$, строкові регістри – $R_s = \{r_{RSI}, r_{RDI}\}$ а також загальні регістри – $R_r = \{r_8 \ldots r_{15}\}$, то поточний стан ЦП можемо представити як:

$$CPU_{state} = \{R_b, R_p, R_s, R_r\} \quad (2)$$

Тоді, враховуючи що кожна програма містить велику кількість інструкції, то її поведінку можна представити у вигляді набору таких станів. Дослідження станів дозволить виявити власне чи відбулась зміна виконання програми у різних модифікованих середовищах. А, для модифікації середовища використовуються знання про існуючі техніки виявлення емуляції. В таких середовищах, вони будуть використовуватись в якості приманок, а цю складову частину можемо представити як $EE_{ae} = \{ae_1, ae_2 \ldots ae_n\}$. Тому, слід очікувати, що інфікована програма буде виконуватись по різному в середовищах які мають і не мають визначені модифікації. Тому при використанні деякої кількості таких середовищ із різними параметрами, формування поведінки для кожного із них буде виглядати наступним чином

$$B_{Isw_i} = \begin{cases} EE_1\big|_{EE_{ae_1}} \xrightarrow{Isw_i} B_1 \,\big|\, EE_{ae_1} \subset M_{ae_i} \\ EE_2\big|_{EE_{ae_2}} \xrightarrow{Isw_i} B_2 \,\big|\, EE_{ae_2} \cap M_{ae_i} = \varnothing \end{cases} \quad (3)$$

В результаті аналізу завдяки такому підходу, буде сформована множина поведінок, і для подальшого виявлення потрібно буде порівняти усі отримані поведінки.

Для забезпечення такого підходу, буде доцільно сформувати розподілену систему. Запропонована система має бути централізованою, а обчислювальні елементи (ОЕ) автономними та гетерогенними. Це дозволить об'єднати велику кількість ОЕ для вирішення поставленого завдання. Основним завданням ОЕ буде проводити аналіз надісланих зразків потенційних інфікованих програм, та формувати поведінку його виконання. Саме тому, можемо визначити ряд основних функцій CE_f, до яких будуть входити: $ce_{f_{gt}}$ – отримання завдання від ЦС, $ce_{f_{ie}}$ – ініціалізація модифікованого емульованого середовища, $ce_{f_{pe}}$ – запуск на виконання отриманого зразка ІП, $ce_{f_{fs}}$ – формування стану ЦП в визначений момент часу, $ce_{f_{cb}}$ – формування поведінки програми на основі накопичених станів ЦП, $ce_{f_{sr}}$ – відправка результату аналізу на центральний сервер. Тоді, список функцій такого спеціалізованого ПЗ, буде виглядати наступним чином:

$$CE_f = \{ce_{f_{gt}}, ce_{f_{ie}}, ce_{f_{pe}}, ce_{f_{fs}}, ce_{f_{cb}}, ce_{f_{sr}}\}. \quad (4)$$

Основними завданнями які визначені перед центральним сервером є: надавати інтерфейсу для користувачів, що хочуть проаналізувати підозрілі програми, аналізувати стан системи на кількість ОЕ, формувати завдання для активних ОЕ, проводи аналіз отриманих результатів. Тому, можемо визначити набір основних функцій центрального серверу CS_f, а саме: $cs_{f_{ut}}$ – отримання зразка інфікованої програми від користувача, $cs_{f_{fa}}$ – формування активних ОЕ в поточний момент часу, $cs_{f_{tf}}$ – формування завдання на основі активних елементів, $cs_{f_{ar}}$ – збір результатів аналізу програми, $cs_{f_{dr}}$ – визначення наявності зловмисного прояву, $cs_{f_{rb}}$ – сповіщення користувача про результат розподіленого аналізу програми. Отже, список функцій центрального серверу буде так:

$$CS_f = \{cs_{f_{ut}}, cs_{f_{fa}}, cs_{f_{tf}}, cs_{f_{ar}}, cs_{f_{dr}}, cs_{f_{rb}}\}. \quad (5)$$

Запропонована система зображена на рисунку 1, який також представляє які основні модулі приймають участь для забезпечення проведення розподіленого аналізу.

Метод розподіленого виявлення інфікованих програм на основі аналізу поведінки виконання. Представимо запропонований метод більш детально, у вигляді кроків.

1. *Отримання системою ІП на аналіз.* Для користувача, при роботі із системою, передбачений відповідний інтерфейс, який дозволяє завантажувати підозрілий файл IS на аналіз. Завантажені файли у систему записуються у чергу CS_q, звідти вибираються коли система вільна. В разі чого, формується множина:

$$CS_q = \{Isw_1, Isw_2, \ldots, Isw_n\}. \quad (6)$$

2. *Формування списку активних ОЕ та генерація параметрів для ізольованих середовищ.* Так як розглянута система працює із набором автономних ОЕ, центральний сервер

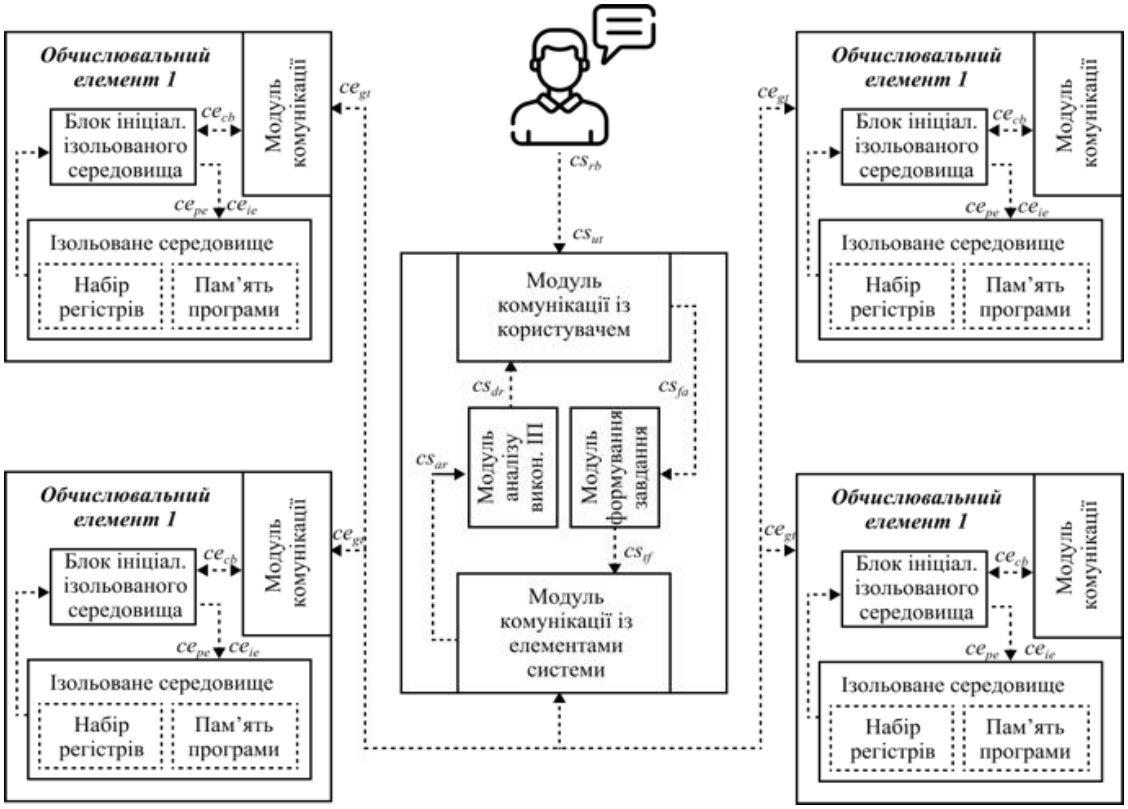


Рис. 1. Архітектура системи розподіленого виявлення зловмисного прояву

перевіряє ті хто долучені в систему, і відповідно до цього формує набір параметрів для ізольованих середовищ. Таким чином, для розподіленого аналізу ІП, використовується наступний набір параметрів:

$$T_{Isw\,i} = \left\{ Isw_i, EE^1_{ae_1} \ldots EE^1_{ae_k}, EE^2_{ae_1} \ldots EE^2_{ae_l}, \ldots, EE^n_{ae_1} \ldots EE^n_{ae_m} \right\}. \quad (7)$$

де, n – кількість активних ОЕ, на яких можна розгорнути ізольоване середовище, а k, l, m – це кількість пасток яка застосовується для кожного середовища відповідно.

3. *Генерація ізольованих середовищ на ОЕ*. Кожен ОЕ від центрального сервера отримає наступний множину, для виконання поставленого завдання $T^n_{IS_i} = \left\{ Isw_i, EE^n : EE^n = \left\{ EE^n_{ae_1}, \ldots, EE^n_{ae_m} \right\} \right\}$. На основі отриманого набору ОЕ формує модифіковане ізольоване середовище, та ініціює емулятор для виконання програми. На цьому етапі, система працює із n набором середовищ, що будуть аналізувати ІП що знаходиться у черзі під номером i, що буде мати наступний вигляд.

$$Isw_i = \left\{ EE^1, EE^2, \ldots, EE^n \right\}. \quad (8)$$

4. *Аналізу програми під час виконання її на ОЕ.* Для проведення аналізу виконання програми запропоновано використати програмні переривання, для зчитування даних про стан регістрів. Враховуючи те, що інформація про зміну стану процесора буде накопичуватись, то визначений стан (2) варто подати наступним чином:

$$CPU_{state} = \left\{ \vec{R}_b, \vec{R}_p, \vec{R}_s, \vec{R}_r \right\} \quad (9)$$

У виразі (9) усі групи регістрів представлені у вигляді векторів, так як кожен вектор містить інформацію про зміну стану відповідного до його регістру. Але такий запис не є зручним для проведення порівняння, тому пропонується виконати наступні перетворення, для формування єдиного значення, що буде характеризувати поведінку в цілому:

$$B = FH\left(\sum_{i=0}^{n} FH\left(\sum_{j=0}^{m} r_i \right) \right) \quad (10)$$

де, i – кількість виконаних програмних переривань, j – кількість регістрів ЦП, та FH функція хешування.

5. *Аналіз отриманих результаті центральним сервером.* Після того, як усі ОЕ повернуть результати виконання програм в різних модифікованих ізольованих середовищах, центральний сервер формує набір поведінок $SB_{Isw_i} = \left\{ B^1_{Isw_i}, B^2_{Isw_i}, \ldots, B^n_{Isw_i} \right\}$. Оскільки сучасні віруси завдяки своїм методам захисту можуть змінювати свою поведінку, центральний сервер буде керуватись правилом (11) для виявлення зловмисного прояву.

$$\neg\left(\forall x, y \in \left\{B^1_{Isw_j}, B^2_{Isw_j}, \ldots, B^n_{Isw_j}\right\}, x = y\right) \Rightarrow Isw_i \text{ is malware} \quad (11)$$

Після чого, про результат аналізу буде сповіщений користувач. На цьому етапі усі ОЕ знищують свої ізольовані середовища, та переходять у режим очікування. Запропонований метод дозволяє підвищити виявлення для ІП які застосовують техніки виявлення емуляції та поліморфні техніки, а також стратегії що базуються на поєднанні їх особливостей. Результативність використання такого методу залежить від кількості досліджених технік уникнення та правильному формуванню наборів параметрів для ізольованих середовищ в розподіленій системі з автономними ОЕ. Запропонований метод графічно представлено на рисунку 2.

Для проведення експериментів, було згенеровано 4 моделі ІП, які використовують визначену кількість технік виявлення емуляції середовища, а також 2 набори модифікованих ізольованих середовищ по 8 у кожному. Таблиця 1 демонструє визначені техніки ІП, а також модифікації середовищ. Таблиця 2 показує зафіксовану кількість різних поведінок в результаті проведення аналізу.

Висновки. У дослідженні запропоновано модель системи для проведення аналізу поведінки виконання ІП в ізольованому середовищі. Даний підхід передбачає формування модифікованих середовищ, на основі знань про техніки виявлення емульованого середовища в якості пасток. Це дозволить створити умови, при яких ІП буде змушена використати одну із технік поліморфізму, що дозволить уникнути виявленню. Для проведення експерименту було створено 4 моделі ІП, які реагують на модифікації в ізольованих середовищах згідно дослідженнями в проаналізованих наукових публікаціях. Кожна модель була проаналізована в двох наборах середовищ із відмінним

Рис. 2. Графічне представлення методу виявлення зловмисного прояву

Таблиця 1

Вхідні дані для проведення експерименту з дослідження поведінки ІП

ІП	*Набір м. середовищ 1 (GEE_1)*		*Набір м. середовищ 2 (GEE_2)*	
$Isw_1(ae_1, ae_2)$	$EE^1_1(ae_1, ae_2)$	$EE^1_5(ae_1, ae_5)$	$EE^2_1(ae_4, ae_5)$	$EE^2_5(ae_5)$
$Isw_2(ae_3)$	$EE^1_2(ae_4)$	$EE^1_6(ae_5)$	$EE^2_2(ae_1)$	$EE^2_6(ae_1, ae_6)$
$Isw_3(ae_3, ae_5)$	$EE^1_3(ae_4, ae_5)$	$EE^1_7(ae_2)$	$EE^2_3(ae_5, ae_6)$	$EE^2_7(ae_2, ae_6)$
$Isw_4(ae_2, ae_6)$	$EE^1_4(ae_1, ae_4)$	$EE^1_8(ae_2, ae_6)$	$EE^2_4(ae_2, ae_4)$	$EE^2_8(ae_4)$

Таблиця 2

Результати експерименту виявлення зловмисної поведінки

ІП	*Поведінки у GEE_1*	*Поведінки у GEE_2*	*Результат*
$Isw_1(ae_1, ae_2)$	5 x $B^1_{Isw_1}$, 3 x $B^2_{Isw_1}$	4 x $B^1_{Isw_1}$, 4 x $B^2_{Isw_1}$	*виявлено*
$Isw_2(ae_3)$	8 x $B^1_{Isw_2}$	8 x $B^1_{Isw_2}$	*не виявлено*
$Isw_3(ae_3, ae_5)$	3 x $B^1_{Isw_3}$, 5 x $B^2_{Isw_3}$	3 x $B^1_{Isw_3}$, 5 x $B^2_{Isw_3}$	*виявлено*
$Isw_4(ae_2, ae_6)$	3 x $B^1_{Isw_4}$, 5 x $B^2_{Isw_4}$	4 x $B^1_{Isw_4}$, 4 x $B^2_{Isw_4}$	*виявлено*

набором пасток, за допомогою централізованої системи. Відмінності поведінки зафіксовані для 1-ї, 3-ї та 4-ї моделей. Щодо 2-ї моделі, так як жодне з середовищ не використовувало модифікацію згідно із її технікою виявлення емуляції, поведінка залишилась однакова. Кількість досліджених технік уникнення виявлення, що застосовується ІП, та їх використання для модифікації середовища виконання, напряму підвищує шанс на виявлення нових. Тому, предметом подальших досліджень є вивчення нових технік захисту ІП від виявлення, їх комбінація в рамках одного екземпляру ІП, та стратегії їх застосування.

ЛІТЕРАТУРА:

1. Razaulla S., Fachkha C., Markarian C., Gawanmeh A., Mansoor W., Fung B. C., Assi C. The age of ransomware: A survey on the evolution, taxonomy, and research directions. *IEEE Access.* 2023. Vol. 11. P. 40698–40723. DOI: 0.1109/ACCESS.2023.3268535

2. Savenko B., Lysenko S., Bobrovnikova K., Savenko O., Markowsky G. Detection DNS tunneling botnets. *11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, 22 September, 2021. P. 64-69. DOI: 10.1109/IDAACS53288.2021.9661022

3. Savenko O., Sachenko A., Lysenko S., Markowsky G.,Vasylkiv N. Botnet detection approach based on the distributed systems. *International Journal of Computing.* 2020. Vol. 19. № 2. P. 190–198. DOI: 10.47839/ijc.19.2.1761

4. Lysenko S., Bobrovnikova K., Kharchenko V., Savenko O. IoT multi-vector cyberattack detection based on machine learning algorithms: traffic features analysis, experiments, and efficiency. *Algorithms.* 2022. Vol 15. № 7. P. 239. DOI: 10.3390/a15070239

5. Numminen A. Windows technical hardening against the most prevalent threats: Master Thesis : 2023, 61 p.

6. Omar M. Static Analysis of Malware. *Defending Cyber Systems through Reverse Engineering of Criminal Malware.* 2022. P. 9–17. DOI: 10.1007/978-3-031-11626-1_2

7. Leon, R. S., Kiperberg M., Leon Zabag A. A., Zaidenberg N. J. Hypervisor-assisted dynamic malware analysis. *Cybersecurity.* 2021. Vol. 4. P. 1–14. DOI: 10.1186/s42400-021-00083-9

8. Yunus Y. K. B. M., Ngah S. B. Review of hybrid analysis technique for malware detection. *The 6th International Conference on Software Engineering & Computer Systems*, Vol. 769, 25–27 September, 2019, Pahang, Malaysia. P. 012075. DOI: 10.1088/1757-899X/769/1/012075

9. Faruki P., Bhan R., Jain V., Bhatia S., El Madhoun N., Pamula R. A survey and evaluation of android-based malware evasion techniques and detection frameworks. *Information.* 2023. Vol. 14. № 7. P.374. DOI: 10.3390/info14070374

10. D'Elia D. C., Coppa E., Palmaro F., Cavallaro L. On the dissection of evasive malware. *IEEE Transactions on Information Forensics and Security*. 2020. Vol. 15. P. 2750–2765. DOI: 10.1109/TIFS.2020.2976559

11. Apostolopoulos T., Katos V., Choo K. K. R., Patsakis C. Resurrecting anti-virtualization and anti-debugging: Unhooking your hooks. *Future Generation Computer Systems.* 2021. Vol. 116. P. 393–405. DOI: 10.1016/j.future.2020.11.004

12. Nappa A., Papadopoulos P., Varvello M., Gomez D.A., Tapiador J., Lanzi A. Pow-how: An enduring timing side-channel to evade online malware sandboxes. *26th European Symposium on Research in Computer Security,* 4-8 October, 2021, Darmstadt, Germany. P. 86–109. DOI: 10.1007/978-3-030-88418-5_5

13. Nicheporuk A., Savenko O., Nicheporuk A., Nicheporuk Y. An Android Malware Detection Method Based on CNN Mixed-Data Model. *16th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer,* Vol. 2732, October 06-10, 2020, Kharkiv, Ukraine. P. 198–213.

14. Gorment N. Z., Selamat A., Krejcar O. Obfuscated malware detection: impacts on detection methods. *Asian Conference on Intelligent Information and Database Systems*, Vol. 1863, 2023. P. 55–66. DOI: 10.1007/978-3-031-42430-4_5

15. Gibert D., Fredrikson M., Mateu C., Planes J., Le Q. Enhancing the insertion of NOP instructions to obfuscate malware via deep reinforcement learning. *Computers & Security.* 2022. Vol. 113. P. 102543. DOI: 10.1016/j.cose.2021.102543

16. Han S. H., Lee D. Kernel-based real-time file access monitoring structure for detecting malware activity. *Electronics.* 2022. Vol. 11, № 12, 1871. DOI: 10.3390/electronics11121871

17. Fedák A., Štulrajter J. Evasion of antivirus with the help of packers. *Science & Military Journal.* 2022. Vol. 17, № 1. P 14–22. DOI: 10.52651/sam.a.2022.1.14-22

18. Liu S., Feng P., Wang S., Sun K., Cao J. Enhancing malware analysis sandboxes with emulated user behavior. *Computers & Security.* 2022. Vol. 115, 102613. DOI: 10.1016/j.cose.2022.102613

19. Arabo A., Dijoux R., Poulain T., Chevalier G. Detecting ransomware using process behavior analysis. *Procedia Computer Science.* 2020. Vol. 168. P 289–296. DOI: 10.1016/j.procs.2020.02.249

20. Van Schaik S., Minkin M., Kwong A., Genkin D., Yarom Y. CacheOut: Leaking data on Intel CPUs via cache evictions. *IEEE Symposium on Security and Privacy (SP).* 24-27 May, 2021, San Francisco, CA, USA. P. 339–354. DOI: 10.1109/SP40001.2021.00064

21. Amit N., Tsafrir D., Schuster A., Ayoub A., Shlomo, E. Virtual CPU validation. *Proceedings of the 25th Symposium on Operating Systems Principles.* 04 October, 2015. P. 311–327. DOI: 10.1145/2815400.2815420

22. Peng P., Soljanin E., Whiting P. Diversity vs. parallelism in distributed computing with redundancy. *IEEE International Symposium on Information Theory (ISIT).* 21-26 June, 2020, Los Angeles, CA, USA. P. 257–262. DOI: 10.1109/ISIT44484.2020.9174030

23. Hamdan S., Ayyash M., Almajali, S. Edge-computing architectures for internet of things applications: A survey. *Sensors.* 2020. Vol. 20, № 22, P. 6441. DOI: 10.3390/s20226441

24. Kong X., Wu Y., Wang H., Xia F. Edge computing for internet of everything: A survey. *IEEE Internet of Things Journal.* 2022. Vol 9, № 23, P. 23472–23485. DOI: 10.1109/JIOT.2022.3200431

25. Rehida P., Sochor T., Martynyuk V., Tarasova O., Orlenko V. A distributed malware detection model based on sandbox technology. *Intelligent Information Technologies & Systems of Information Security (IntelITSIS)*, Vol. 3373, 22-24 March, 2023, Khmelnytskyi, Ukraine. P. 475–485.

REFERENCES:

1. Razaulla, S., Fachkha, C., Markarian, C., Gawanmeh, A., Mansoor, W., Fung, B. C., & Assi, C. (2023). The age of ransomware: A survey on the evolution, taxonomy, and research directions. *IEEE* Access, 11, 40698–40723. https://doi.org/10.1109/ACCESS.2023.3268535

2. Savenko, B., Lysenko, S., Bobrovnikova, K., Savenko, O., & Markowsky, G. (2021, September). Detection DNS tunneling botnets. *11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, 1, 64–69, IEEE. https://doi.org/10.1109/IDAACS53288.2021.9661022

3. Savenko, O., Sachenko, A., Lysenko, S., Markowsky, G., & Vasylkiv, N. (2020). Botnet detection approach based on the distributed systems. *International Journal of Computing*, *19*(2), 190–198. https://doi.org/10.47839/ijc.19.2.1761

4. Lysenko, S., Bobrovnikova, K., Kharchenko, V., & Savenko, O. (2022). IoT multi-vector cyberattack detection based on machine learning algorithms: traffic features analysis, experiments, and efficiency. *Algorithms*, *15*(7), 239. https://doi.org/10.3390/a15070239

5. Numminen, A. (2023). *Windows technical hardening against the most prevalent threats* [Master thesis, University of Jyväskylä]. JYX Digital Repository. Retrieved from: https://jyx.jyu.fi/handle/123456789/87270

6. Omar, M. (2022). Static Analysis of Malware. *Defending Cyber Systems through Reverse Engineering of Criminal Malware* (pp. 9–17). Springer, Cham. https://doi.org/10.1007/978-3-031-11626-1_2

7. Leon, R. S., Kiperberg, M., Leon Zabag, A. A., & Zaidenberg, N. J. (2021). Hypervisor-assisted dynamic malware analysis. *Cybersecurity*, 4, 1–14. https://doi.org/10.1186/s42400-021-00083-9

8. Yunus, Y. K. B. M., & Ngah, S. B. (2020, February). Review of hybrid analysis technique for malware detection. *IOP conference series: materials science and engineering* 769(1), 012075). IOP Publishing. https://doi.org/10.1088/1757-899X/769/1/012075

9. Faruki, P., Bhan, R., Jain, V., Bhatia, S., El Madhoun, N., & Pamula, R. (2023). A survey and evaluation of android-based malware evasion techniques and detection frameworks. *Information*, 14(7), 374. https://doi.org/10.3390/info14070374

10. D'Elia, D. C., Coppa, E., Palmaro, F., & Cavallaro, L. (2020). On the dissection of evasive malware. *IEEE Transactions on Information Forensics and Security*, *15*, 2750–2765. https://doi.org/10.1109/TIFS.2020.2976559

11. Apostolopoulos, T., Katos, V., Choo, K. K. R., & Patsakis, C. (2021). Resurrecting anti-virtualization and anti-debugging: Unhooking your hooks. *Future Generation Computer Systems*, 116, 393–405. https://doi.org/10.1016/j.future.2020.11.004

12. Nappa, A., Papadopoulos, P., Varvello, M., Gomez, D. A., Tapiador, J., & Lanzi, A. (2021). Pow-how: An enduring timing side-channel to evade online malware sandboxes. *Computer Security–ESORICS 2021: 26th European Symposium on Research in Computer Security,* 26, 86–109. Springer International Publishing. https://doi.org/10.1007/978-3-030-88418-5_5

13. Nicheporuk, A., Savenko, O., Nicheporuk, A., & Nicheporuk, Y. (2020, October). An Android Malware Detection Method Based on CNN Mixed-Data Model. *16th International Conference on ICT in Education,*

Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer (ICTERI), 2732, pp. 198–213. Retrieved from: https://ceur-ws.org/Vol-2732/20200198.pdf

14. Gorment, N. Z., Selamat, A., & Krejcar, O. (2023, July). Obfuscated malware detection: impacts on detection methods. *Asian Conference on Intelligent Information and Database Systems*, 55-66. Springer, Cham. https://doi.org/10.1007/978-3-031-42430-4_5

15. Gibert, D., Fredrikson, M., Mateu, C., Planes, J., & Le, Q. (2022). Enhancing the insertion of NOP instructions to obfuscate malware via deep reinforcement learning. *Computers & Security*, 113, 102543. https://doi.org/10.1016/j.cose.2021.102543

16. Han, S. H., & Lee, D. (2022). Kernel-based real-time file access monitoring structure for detecting malware activity. *Electronics*, 11(12), 1871. https://doi.org/10.3390/electronics11121871

17. Fedák, A., & Štulrajter, J. (2022). Evasion of antivirus with the help of packers. *Science & Military Journal*, *17*(1), 14–22. https://doi.org/10.52651/sam.a.2022.1.14-22

18. Liu, S., Feng, P., Wang, S., Sun, K., & Cao, J. (2022). Enhancing malware analysis sandboxes with emulated user behavior. *Computers & Security*, 115, 102613. https://doi.org/10.1016/j.cose.2022.102613

19. Arabo, A., Dijoux, R., Poulain, T., & Chevalier, G. (2020). Detecting ransomware using process behavior analysis. *Procedia Computer Science*, 168, 289–296. https://doi.org/10.1016/j.procs.2020.02.249

20. Van Schaik, S., Minkin, M., Kwong, A., Genkin, D., & Yarom, Y. (2021, May). CacheOut: Leaking data on Intel CPUs via cache evictions. *2021 IEEE Symposium on Security and Privacy (SP),* 339–354. IEEE. https://doi.org/10.1109/SP40001.2021.00064

21. Amit, N., Tsafrir, D., Schuster, A., Ayoub, A., & Shlomo, E. (2015, October). Virtual CPU validation. *Proceedings of the 25th Symposium on Operating Systems Principles*, 311–327. https://doi.org/10.1145/2815400.2815420

22. Peng, P., Soljanin, E., & Whiting, P. (2020, June). Diversity vs. parallelism in distributed computing with redundancy. *2020 IEEE International Symposium on Information Theory*, 257–262. IEEE. https://doi.org/10.1109/ISIT44484.2020.9174030

23. Hamdan, S., Ayyash, M., & Almajali, S. (2020). Edge-computing architectures for internet of things applications: A survey. *Sensors*, 20(22), 6441. https://doi.org/10.3390/s20226441

24. Kong, X., Wu, Y., Wang, H., & Xia, F. (2022). Edge computing for internet of everything: A survey. *IEEE Internet of Things Journal*, 9(23), 23472–23485. https://doi.org/10.1109/JIOT.2022.3200431

25. Rehida, P., Sochor, T., Martynyuk, V., Tarasova, O., & Orlenko, V. (2023). A distributed malware detection model based on sandbox technology. 4th International Workshop on Intelligent Information Technologies & Systems of Information Security *IntelITSIS*, 3373, 475–485. Retrieved from: https://ceur-ws.org/Vol-3373/paper32.pdf

UDC [005.092:614.2]-024.84
DOI https://doi.org/10.32782/IT/2024-4-22

Oleksii SERHIEIEV
Postgraduate Student at the Department of System Analysis and Control, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, serhieiev.o.s@nmu.one
ORCHID: *0000-0001-5781-5540*
Scopus-Author ID: *57220190743*

To cite this article: Serhieiev, O. (2024). Zastosuvannia systemnoho pidkhodu do analizu medychnoi lohistyky rehionalnoho rivnia [Applying the systems approach to the analysis of regional-level medical logistics]. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 187–196, doi: https://doi.org/10.32782/IT/2024-4-22

APPLYING THE SYSTEMS APPROACH TO THE ANALYSIS OF REGIONAL-LEVEL MEDICAL LOGISTICS

The paper presents an in-depth analysis of the medical logistics system with a focus on improving decision-making processes for infrastructure development. The relevance of the work lies in the study of the system's robustness in crisis situations. The crisis revealed the inability of the system to meet the unprecedented demand for medicines and medical devices, emphasizing the need to create a more sustainable and adaptive logistics infrastructure.

The purpose of work. *To analyse and build a medical logistics system at the regional level. For this purpose, the key components and functions, including transport networks, regional and subregional centers and distribution hubs, should be additionally described. To analyse the legislative acts on the implementation of medical logistics, highlighting the general algorithm of actions for infrastructure development.*

Methodology. *A systematic approach was used to analyse the medical logistics system at the regional level. Using analysis and synthesis, the main components of the system are identified, and its organizational chart in times of crisis is presented. The process of providing assistance and organizing the relevant logistics infrastructure is considered using the IDEF0 notation, where inputs, outputs, controls, and mechanisms are identified. The process of opening regional, subregional and distribution centers is analysed separately and a methodology for its organization was proposed.*

Scientific novelty. *A methodology for opening regional, subregional and distribution centers is developed, which involves making appropriate decisions based on the results of solving the two-stage location-activation problem using software. The article details specific steps for organizing some processes of the regional medical logistics system that are potentially at risk of human factor influence in decision-making. The algorithm provides for the location of centers of different levels and determination of the transportation strategy. The author proposes to improve this process by automating some of the steps. Most of the steps can be optimized by applying the developed methodology and software for solving multi-stage location-activation problems. The application of the proposed modifications will allow taking into account more qualitative and quantitative indicators when making decisions on scaling the infrastructure in the field of medical logistics.*

Conclusions*. The proposed methodology and software make it possible to optimize some elements of the regional medical logistics system, improving their ability to adapt during emergencies, make more effective management decisions, achieve effective indicators by reducing costs and meeting the needs of the population.*

Key words: *system analysis, medical logistics, optimal solutions, business process, methodology.*

Олексій СЕРГЄЄВ
аспірант кафедри системного аналізу та управління, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького 19, Дніпро, Україна, 49005
ORCHID: *0000-0001-5781-5540*
Scopus-Author ID: *57220190743*

Бібліографічний опис статті: Сергєєв, О. (2024). Застосування системного підходу до аналізу медичної логістики регіонального рівня. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 4, 187–196, doi: https://doi.org/10.32782/IT/2024-4-22

ЗАСТОСУВАННЯ СИСТЕМНОГО ПІДХОДУ ДО АНАЛІЗУ МЕДИЧНОЇ ЛОГІСТИКИ РЕГІОНАЛЬНОГО РІВНЯ

У роботі представлено поглиблений аналіз системи медичної логістики з основним акцентом на вдосконалення процесів ухвалення рішень щодо розвитку інфраструктури. Актуальність роботи полягає у дослідженні робастності системи в кризових ситуаціях. Під час яких була виявлена нездатність системи задовольнити безпрецедентний попит на ліки та вироби медичного призначення, підкресливши необхідність створення більш стійкої та адаптивної логістичної інфраструктури.

Мета роботи. *Проаналізувати та побудувати систему медичної логістики регіонального рівня. Для неї додатково навести ключові компоненти та функції, що включають транспортні мережі, регіональні та субрегіональні центри та розподільчі хаби. Проаналізувати законодавчі акти щодо впровадження медичної логістики із висвітленням загального алгоритму дій для розвинення інфраструктури.*

Методологія. *Застосовано системний підхід для аналізу системи медичної логістики регіонального рівня. За допомогою аналізу та синтезу виділено основні компоненти системи, наведено її організаційна діаграма в кризові періоди. Процес надання допомоги та організації відповідної логістичної інфраструктури розглянуто з використанням нотації IDEF0, де виділено входи, виходи, керування та механізми. Окремо було проаналізовано процес відкриття регіональних, субрегіональних та дистрибуційних центрів та запропоновано методику щодо його організації.*

Наукова новизна. *Розроблена методика відкриття регіональних, субрегіональних та центрів дистрибуції, що передбачає ухвалення відповідних рішень базуючись на результатах розв'язання двоетапної задачі розміщення-активації із застосуванням її у вигляді програмного забезпечення. Деталізовано конкретні кроки з організації деяких процесів регіональної системи медичної логістики, що потенційно перебувають під ризиком впливу людського фактору під час ухвалення рішень. Алгоритм передбачає розміщення центрів різних рівнів та визначення стратегії транспортування. Автором запропоновано вдосконалити цей процес шляхом автоматизації деяких процесів. Більшість кроків може бути оптимізована за допомогою застосування розробленої методики та програмного забезпечення для розв'язання багатоетапних задач розміщення-активації. Застосування запропонованих модифікацій дозволить враховувати більше якісних та кількісних показників при ухваленні рішень щодо масштабування інфраструктури у сфері медичної логістики.*

Висновки. *Запропоновані методологія та програмне забезпечення дозволяють оптимізувати деякі елементи регіональної системи медичної логістики, покращуючи їх здатність до адаптації під час надзвичайних ситуацій, ухвалення більш ефективних управлінських рішень, досягнення результативних показників шляхом зменшення витрат та задоволення потреб населення.*

Ключові слова: *системний аналіз, медична логістика, оптимальні рішення, бізнес-процес, методологія.*

Introduction. The last few years have shown us the weaknesses and areas for improvement of the medical logistics system. The global pandemic of 2020 unreadiness to meet the demand for medicines and immunobiological products in large quantities. This also highlights the strain on the system when we had to move quickly and efficiently during a crisis. The situation got worse with the full-scale invasion in 2022 which created an additional demand for moving large quantities of medicines as part of humanitarian aid. This unprecedented demand put a lot of pressure on the existing infrastructure, we need a robust and resilient medical logistics system that can adapt to crisis. The above-mentioned circumstances also introduced new challenges that require modernization and optimization of medical logistics transportation.

At the state level, one of the major challenges is to provide medicines to remote and underserved areas. To address this need we need to establish new regional, sub-regional and distribution centers in unserved areas. The continuous strain on the system also showed us the importance of developing adaptive strategies and leveraging technology to improve the efficiency and responsiveness of medical logistics.

The logistics industry is crucial for economic and social development. The paper (Lu, 2022) explores the challenges and strategies for optimizing logistics management from a supply chain perspective, emphasizing the importance of logistics optimization for sustainable industry growth. The work (Umoren, 2021) proposes a computational framework using fuzzy logic to optimize healthcare logistics, addressing resource constraints and improving the quality of care and quality of experience in Nigerian hospitals. The research (Mykhalchuk, 2020) introduces a new logistic model for Ukrainian healthcare, focusing on optimizing financial, material, human, and informational flows. It discusses the implementation of logistics concepts like enterprise resource planning, total quality management and others to enhance medical service quality and efficiency. The paper (Yaspal, 2023) addresses the increased production of infectious medical waste due to COVID-19, proposing a data-driven framework for optimizing reverse logistics in medical waste management. The article (Gadiiak, 2021) highlights the significance of

logistics in healthcare management, focusing on optimizing resource flow processes. It outlines the benefits of applying logistic approaches in medical institutions for effective resource control and quality improvement in medical services. From mathematical side, the problem of logistics is wildly researched as well. The paper (Koriashkina, 2023) presents mathematical models for the distribution of material resources related to preventive measures in case of technogenic emergencies. It examines optimal zoning tasks, assigning areas to social service objects with the possibility of overlapping zones when the nearest center cannot fully provide services. The study also considers the optimal location of new emergency logistics centers and the redistribution of loads across all structural elements, as well as the selection of locations for structural subdivisions based on existing objects. The criteria for optimality include minimizing the service time to the farthest object or the total distance to the nearest centers from consumers densely distributed in the given territory. Organizational costs for establishing new centers are also considered. The study (Koriashkina, 2024) focuses on enhancing the efficiency of evacuation logistics in regions affected by emergencies through a systematic approach. It highlights the importance of mathematical modeling and optimization theories in ensuring the rational distribution of transport and material resources. The research delves into two-stage evacuation logistics, emphasizing the role of first-stage centers (initial reception) and second-stage centers (further aid). The proposed models address optimal partitioning and resource allocation, aiming to minimize evacuation time and maximize coverage. The practical application of these models provides valuable insights for developing preventive measures and planning effective rescue operations.

The goal of this research is to conduct an in-depth examination of the existing state of the medical logistics system, evaluate the current procedures in place, and identify potential areas for improvement. Through this analysis, the author aims to offer recommendations for enhancing the logistics system, thereby increasing its quality and resilience. Ultimately, the objective is to ensure that the medical logistics system can effectively manage both routine operations and unforeseen emergencies, thereby safeguarding public health and welfare.

Key components of the medical logistics system. Healthcare systems need to be adaptable and strong in the face of uncertain crises like pandemics and floods. One key aspect of a strong healthcare system is the transportation network that helps distribute drugs effectively across emergencies. We see the effects of insufficient transportation infrastructure in distributing drugs across crises like pandemics and wars. On the other hand, we see the lives saved and cures contained across crises when healthcare systems have robustly structured transportation networks. While there are logistical and economic issues in setting up a transportation network in healthcare systems, the lives saved and the cures contained never cease to outweigh the costs. The key components of the Medical Logistics System are:

– Verkhovna Rada of Ukraine;
– Ministry of Health of Ukraine;
– suppliers designated by the MOH of Ukraine;
– regional centers;
– subregional centers;
– distribution centers.

The key components are illustrated in Figure 1.

Therefore, it is critical to focus on and invest in these health systems to ensure that they are resilient in times of crisis and can do their job effectively to protect the public. Lives saved are the goal and public health protection is the immediate effect across crises. To ensure this effect and goal, we shall first focus on the processes and current state of the medical logistics system across crises. We describe the decision-making process and the organization of the medical logistics system across a crisis. The process of medical logistics system organization in a crisis is shown in Figure 2. The Ukraine governments laws and orders are used for compiling the diagram: resolutions of the Cabinet of Ministers of Ukraine (No. 228 from 2024, No. 1805-p from 2021, No. 333-p from 2014); order of Ministry of Health (No. 232 from 2015); letter of the National Bank of Ukraine (No. 12-13780 111/1922 from 2004).

We will use the elements of the medical logistics system in Figure 2 to build a model for business processes. A context diagram in IDEF0 notation for medical logistics will be created depicting high-level functions and interactions only. The diagram has four types of arrows:

– Inputs: emergency; requests for medical assistance.
– Outputs: meeting the end-customers' needs.
– Elements of control: legislative acts; regulatory instructions for utilization and storage; internal regulations and orders.
– Mechanisms: Verkhovna Rada of Ukraine; Ministry of Health of Ukraine; regional health departments; regional centers.

Let's decompose Diagram A0 into the following elements:

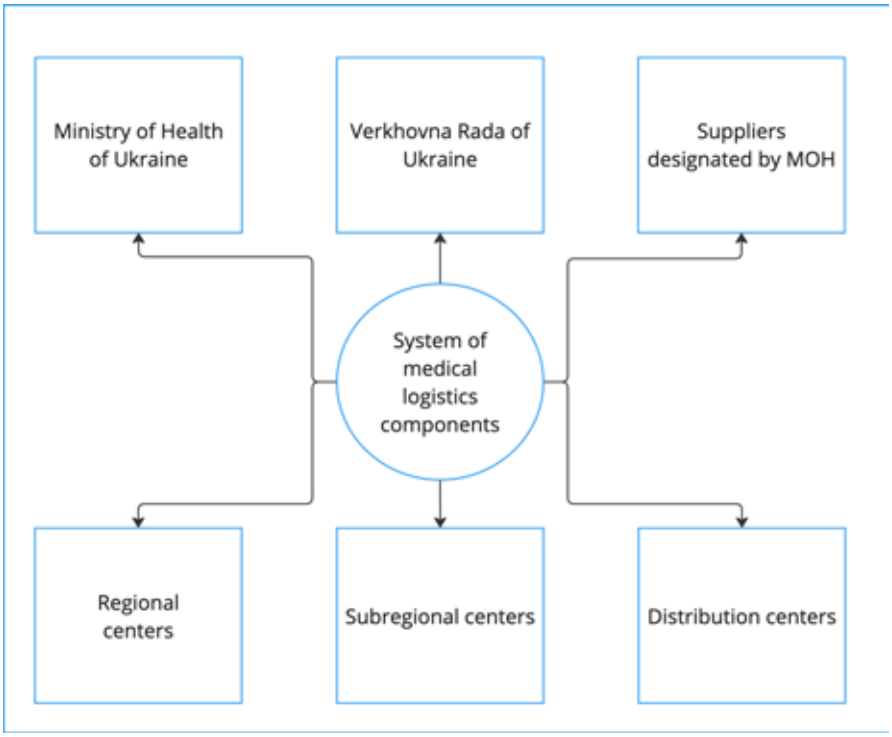


Fig. 1. Key components of the system of medical logistics

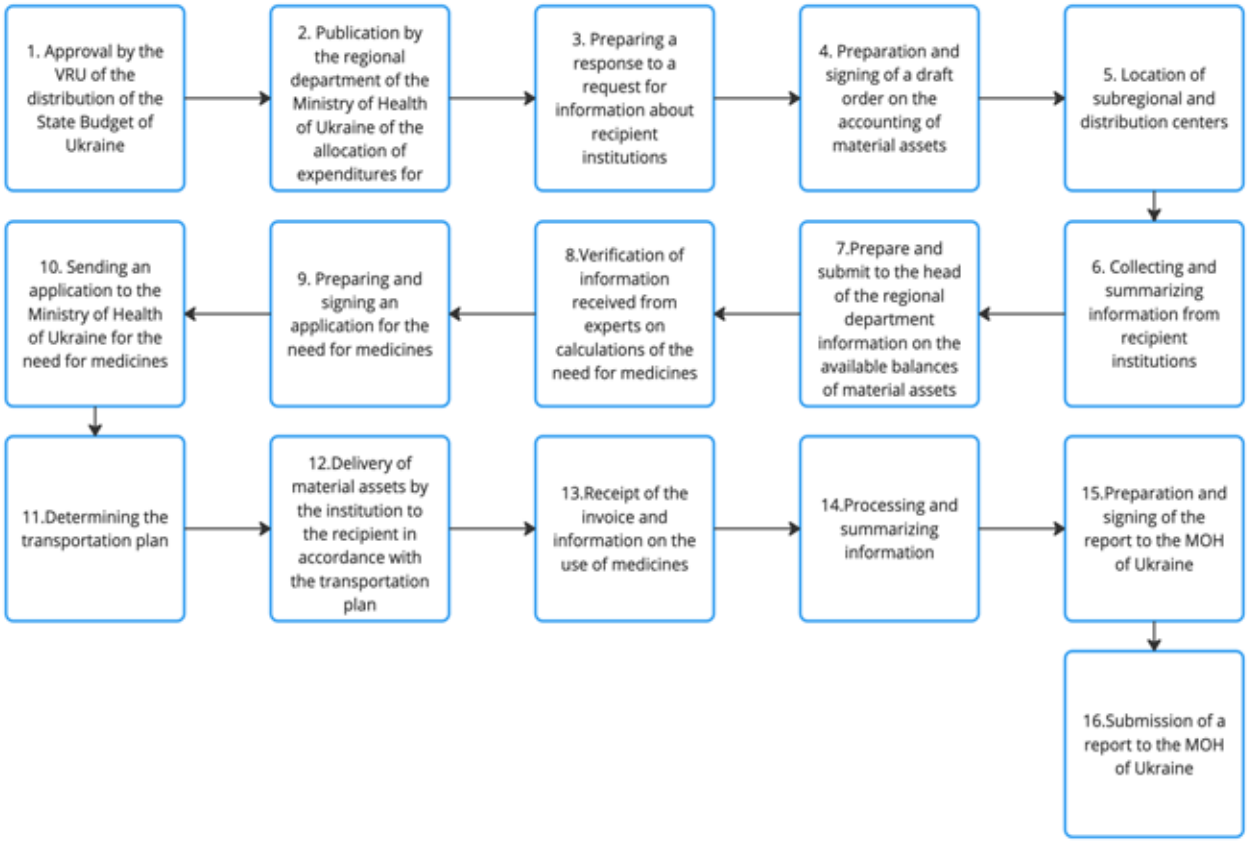


Fig. 2. Organization of the medical logistics system in a crisis

– Adoption of the legislative initiative of the Verkhovna Rada and preparation of orders of the Ministry of Health;

– identification of subregional centers and distribution centers;

– collecting and summarizing information from recipient institutions. Expert research;

– working with experts to calculate the need for medicines and medical supplies;

– determining the transportation strategy;

– receiving of material assets by the recipient institution in accordance with the transportation strategy;

– reporting by regional departments to the Ministry of Health.

The decomposed context diagram, that visualizes IDEF0 notation is shown in Figure 3.

The medical logistics system at the regional level performs the following functions through subsystems:

– Basic: ordering; supply; distribution.

– Key: purchasing management; transportation; inventory management; physical distribution.

– Additional: storage; cargo handling; IT support; accounting.

Moreover, the medical logistics system can be represented by several subsystems: legal, medical, IT, supply chain management and finance. When it comes to the legal system, it covers laws, regulations, and ethics, including contractual agreements and healthcare regulations. The medical one focuses on the clinical aspects of ensuring that drugs and medical supplies are appropriate, safe, and meet the needs of healthcare professionals and patients. The IT subsystem uses technology to manage information, automate processes and communications to improve the efficiency, accuracy, and reliability of logistics operations. The next one is supply chain management – it procures, stores, and distributes pharmaceuticals and supplies to ensure they are available on time and in good condition. The final one is the finance subsystem that manages the accounting aspects, including budgeting, cost tracking and optimization, invoicing and financial reporting. The mentioned functions and classifications for subsystems can be visualized in the form of the diagram in Figure 4.

It is worth highlighting some specific steps from the medical logistics organization steps from Figure 2 that potentially might contain lots of manual decision-making processes. These steps are the following:

– Step 5: Location of subregional and distribution centers.

– Step 6: Collecting and summarizing information from recipient institutions.

– Step 11: Determining the transportation plan.

These steps can be represented as certain algorithm of how a decision-making person runs the process of determining geographical locations for centers of different stages. An algorithm

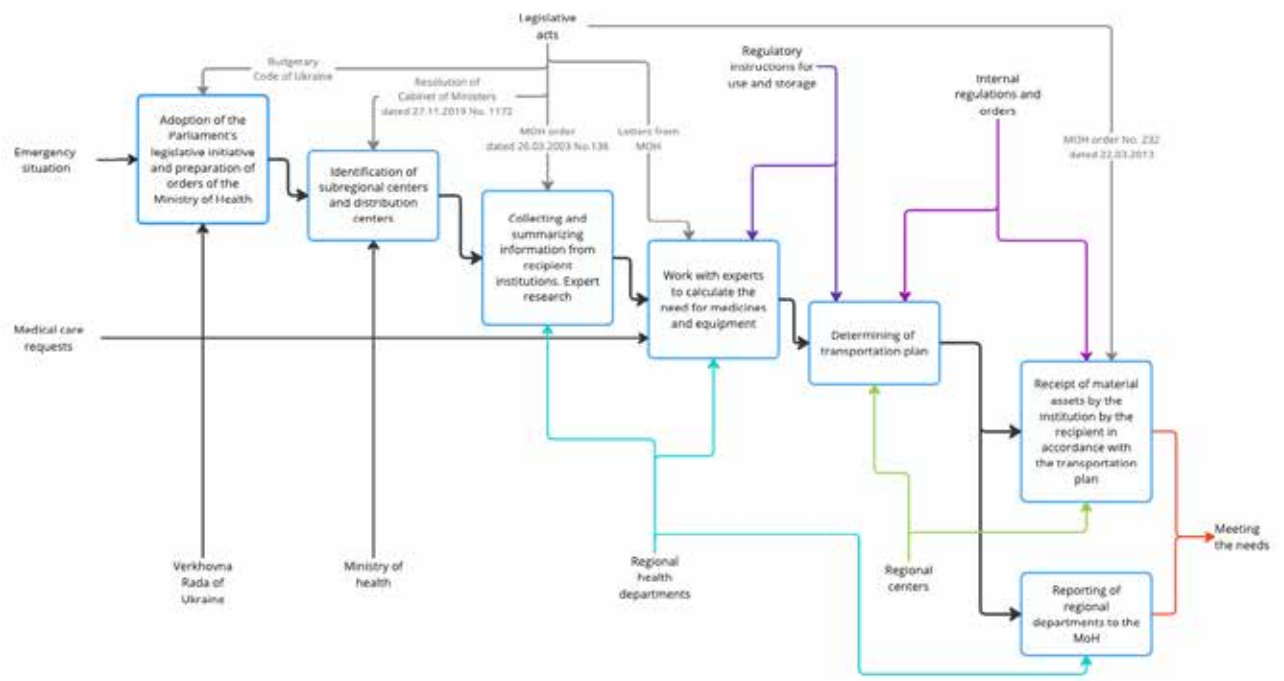


Fig. 3. Decomposition for diagram A0 for medical logistics

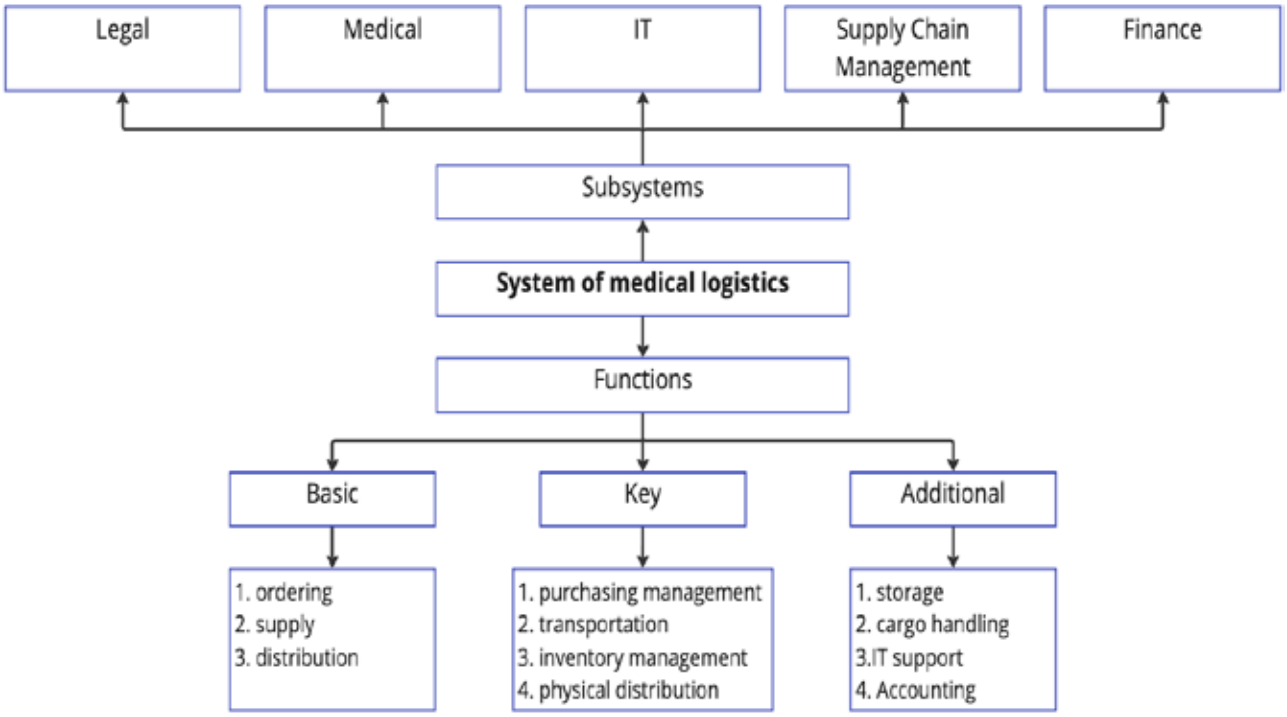


Fig. 4. Structure of medical logistics system

for opening regional, sub-regional and distribution centers and organizing transportation can be formulated as follows.

Step 1. The Ministry of Health submits a request for medicines and medical supplies from the regions to the relevant regional departments.

Step 2. The regional departments collect data on health needs, population and infrastructure capacity of different regions.

Step 3. Based on the information provided by the regional departments, the Ministry of Health decides which regional centers to open. This decision is made at the government level and considers factors such as geographic coverage, population density and specific medical needs. As a result, a report is prepared for the Ministry of Health.

Step 4. The Ministry of Health identifies potential sub-regional centers from the set of existing oblast medical institutions.

Step 5. The regional department evaluates the potential sub-regional centers proposed by the MoH to determine their suitability on the basis of the local healthcare demands and logistical needs.

Step 6. Detailed analysis of potential demand for medicines and planning for potential demand is carried out by the planning service of the regional department.

Step 7. Based on the demand planning, the regional department selects a certain number of sub-regional centers to be activated from the MoH list.

Step 8. Following the demand requests and the projected demand plan, medicines and SMEs are categorized:

– Immunobiological medicines;
– humanitarian aid;
– Medical products
– medical equipment;
– consumable materials.

Step 9. For each category of medical supplies, the regional department will determine whether the approach to aiding will be stationary or mobile:

– The need for stationary and mobile immunization sites is determined.

– Determine the need for stationary and mobile humanitarian distribution points.

– Medical supplies may be sent to stationary medical facilities or mobile points.

Step 10. Activation or location of distribution centers:

– For inpatient care, the department either activates existing facilities or establishes new ones as distribution centers (DCs) if existing facilities are insufficient.

– For mobile healthcare, the department determines the locations of mobile DCs.

Step 11. The regional department continuously monitors the efficiency of work and demand in the centers and makes the necessary adjustments to meet the changing needs of the population for medical care.

Step 12. The regional department assigns responsibility to the regional centers for developing a transportation plan in accordance with existing and planned needs.

The described algorithm is depicted in Figure 5.

It is proposed to improve this process through the implementation of certain automation measures. Most stages will be either eliminated or optimized, and new stages will be introduced. The Ministry of Health will organize the operation of a unified information system called MedData. The regional departments will use software to automate decision-making regarding the location and activation of centers. As a result, sub-regional centers, distribution centers, and transport schemes will be activated and placed. Regional centers will utilize the provided logistics software, which includes recommended transport schemes. The visualization of the improved process diagram is shown in Figure 6.

A more detailed description and formulation of the problem including mathematical models, solution approaches, and software implementation proposed by the author are detailed in (Us, Serhieiev, 2023; Serhieiev, Us, 2023). These elements are designed to integrate seamlessly to address the identified challenges.

Conclusions. This study analyzed the medical logistics system to improve decision-making processes for infrastructure development. The analysis highlighted several critical areas that require optimization and modernization to meet the increasing demands and challenges of both routine operations and unforeseen emergencies, such as the global pandemic and humanitarian crises. The key functions and components of the systems are described to emphasize their importance. The organization of the medical logistics system in a crisis situation is detailed according to steps derived from legislative documents. In addition, the vision of a medical logistics system with the use of a systematic approach is proposed. The medical process is presented in the form of IDEF0 notation to highlight the main steps during this process.

The main steps that could potentially contain a lot of manual decision-making processes are selected. For these steps, a more detailed algorithm is provided, where the reader can observe the set of actions for opening and locating centers of different stages. Improvements to the aforementioned algorithm have been proposed to eliminate manual decisions and replace them with the use of methodology and software implementation: it is proposed to make better use of the unified information system, MedData, managed by the Ministry of Health, which will centralize data management and improve coordination between health departments and centers. By using advanced

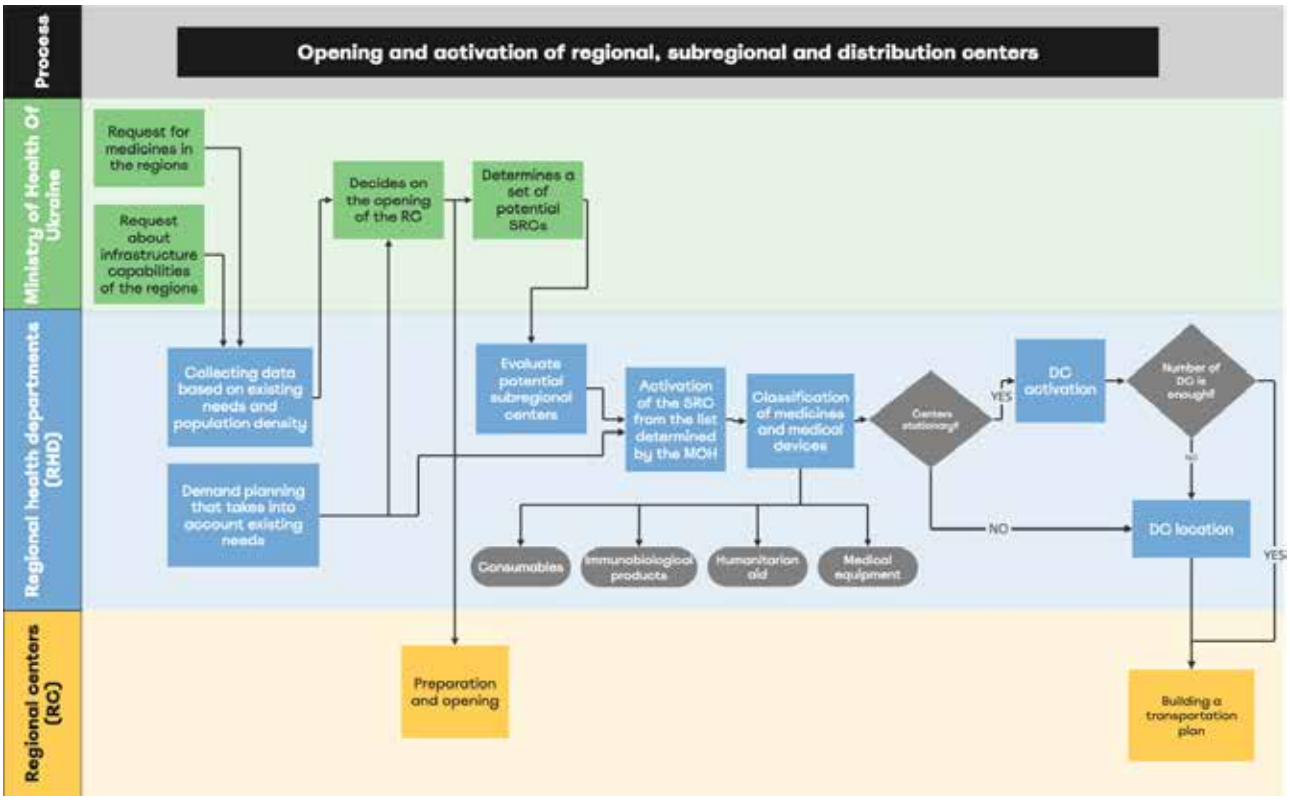


Fig. 5. Opening and activation of regional, subregional and distribution centers

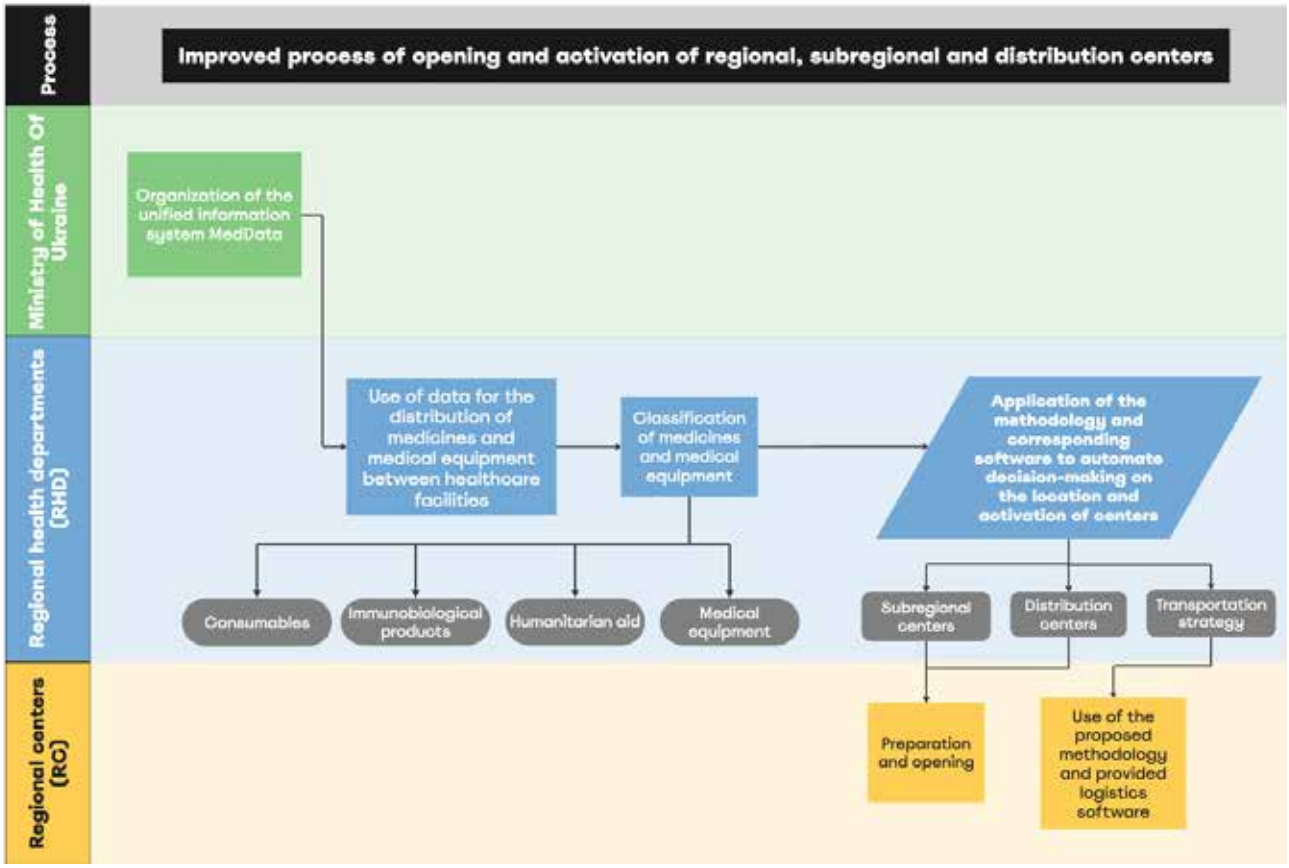


Fig. 6. Improved process of opening and activation of regional, subregional and distribution centers

software to automate decision-making processes, regional departments can streamline the location and activation of sub-regional and distribution centers. This will not only reduce human error and increase efficiency, but also ensure a rapid response in times of crisis. Ultimately, the provided improvements aim to build a more efficient, reliable, and responsive medical logistics system that can safeguard public health by effectively managing both routine and emergency medical logistics operations.

This publication was prepared as a part of the scientific theme 0123U100011 «Problems of analysis, modeling, and optimization of technological processes in complex systems of different nature», which is implemented on the System Analysis and Control Department at Dnipro University of Technology.

BIBLIOGRAPHY:

1. Research on Optimization of Logistics Management Process from the Perspective of Supply Chain Management / J. Lu et al. *Frontiers in Business, Economics and Management*. 2022. Vol. 6, no. 3. P. 122–125. https://doi.org/10.54097/fbem.v6i3.3328 (date of access: 28.06.2024).

2. Healthcare Logistics Optimization Framework for Efficient Supply Chain Management in Niger Delta Region of Nigeria / I. J. Umoren et al. *International Journal of Advanced Computer Science and Applications*. 2021. Vol. 12, no. 4. https://doi.org/10.14569/ijacsa.2021.0120475 (date of access: 28.06.2024).

3. Концептуальні засади логістики нового часу в сучасній українській медицині. V. M. Mykhalchuk та ін. *Вісник соціальної гігієни та організації охорони здоров'я України*. 2020. № 3. С. 67–73. https://doi.org/10.11603/1681-2786.2020.3.11625 (дата звернення: 28.06.2024).

4. A data-driven digital transformation approach for reverse logistics optimization in a medical waste management system / B. Yaspal et al. *Journal of Cleaner Production*. 2023. P. 139703. https://doi.org/10.1016/j.jclepro.2023.139703 (date of access: 28.06.2024).

5. Gadiiak L., Pysarenko V. LOGISTICS APPROACH IN HEALTHCARE MANAGEMENT. *Investytsiyi: praktyka ta dosvid*. 2021. No. 7. P. 113. https://doi.org/10.32702/2306-6814.2021.7.113 (date of access: 28.06.2024).

6. Larysa K., Serhii D. МАТЕМАТИЧНІ МОДЕЛІ ТА МЕТОДИ РОЗМІЩЕННЯ ОБ'ЄКТІВ І ЗОНУВАННЯ ТЕРИТОРІЙ В СИСТЕМАХ ЕКСТРЕНОЇ ЛОГІСТИКИ. *System technologies*. 2024. Т. 6, № 149. С. 107–122. https://doi.org/10.34185/1562-9945-6-149-2023-09 (дата звернення: 10.07.2024).

7. Two-stage problems of optimal location and distribution of the humanitarian logistics system's structural subdivisions / L. S. Koriashkina et al. *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*. 2024. No. 1. P. 130–139. https://doi.org/10.33271/nvngu/2024-1/130 (date of access: 10.07.2024).

8. Про затвердження Порядку використання коштів, передбачених у державному бюджеті для виконання програм та здійснення централізованих заходів з охорони здоров'я: Постанова Каб. Міністрів України від 17.03.2011 р. № 298: станом на 25 черв. 2024 р. URL: https://zakon.rada.gov.ua/laws/show/298-2011-п#Text (дата звернення: 28.06.2024).

9. Про організацію контролю за цільовим та раціональним використанням матеріальних цінностей, отриманих централізовано за бюджетними програмами та заходами: Наказ від 22.03.2013 р. № 232: станом на 31 берез. 2015 р.

10. Про порядок відображення в бухгалтерському обліку інформації про рух запасів матеріальних цінностей та ювілейних і пам'ятних монет карбованцевого номіналу: Лист Нац. банку України від 30.12.2004 р. № 12-111/1922-13780. URL: https://zakon.rada.gov.ua/laws/show/v3780500-04#Text (дата звернення: 28.06.2024).

11. Про схвалення Концепції реформування місцевого самоврядування та територіальної організації влади в Україні: Розпорядж. Каб. Міністрів України від 01.04.2014 р. № 333-р. URL: https://zakon.rada.gov.ua/laws/show/333-2014-р#Text (дата звернення: 28.06.2024).

12. Про схвалення Стратегії реформування системи управління державними фінансами на 2022-2025 роки та плану заходів з її реалізації: Розпорядж. Каб. Міністрів України від 29.12.2021 р. № 1805-р. URL: https://zakon.rada.gov.ua/laws/show/1805-2021-р#Text (дата звернення: 28.06.2024).

13. Інформаційно-аналітичний ресурс, для обміну інформацією щодо залишків лікарських засобів та медичних виробів. *Головна – Med Data*. URL: https://meddata.com.ua/ (дата звернення: 28.06.2024).

14. Serhieiev O. S., Us S. A. Modified genetic algorithm approach for solving the two-stage location problem. *Radio Electronics, Computer Science, Control*. 2023. No. 3. P. 159. https://doi.org/10.15588/1607-3274-2023-3-16 (date of access: 28.06.2024).

15. Svitlana U., Oleksii S. An algorithm for solving a two-stage continuous-discrete location problem for medical logistics optimization. *System technologies*. 2024. Vol. 5, no. 148. P. 71–85. https://doi.org/10.34185/1562-9945-5-148-2023-07 (date of access: 28.06.2024).

REFERENCES:

1. Lu, J., Xu, H., Kan, X., & Lei, Y. (2022). Research on optimization of logistics management process from the perspective of supply chain management. *Frontiers in Business, Economics and Management*, *6*(3), 122–125. https://doi.org/10.54097/fbem.v6i3.3328

2. Umoren, I. J., E., U., P., A., & C., K. (2021). Healthcare logistics optimization framework for efficient supply chain management in niger delta region of nigeria. *International Journal of Advanced Computer Science and Applications*, 12(4). https://doi.org/10.14569/ijacsa.2021.0120475

3. Mykhalchuk, V. M., Kolomoyets, A. V., Tolstanov, O. K., & Hbur, Z. V. (2020). Kontseptualni zasady lohistyky novoho chasu v suchasnii ukrainskii medytsyni [Conceptual principles of modern logistics in modern Ukrainian medicine]. *Visnyk sotsialnoi hihiieny ta orhanizatsii okhorony zdorovia Ukrainy*, (3), 67–73. https://doi.org/10.11603/1681-2786.2020.3.11625 [in Ukraine].

4. Yaspal, B., Jauhar, S. K., Kamble, S., Belhadi, A., & Tiwari, S. (2023). A data-driven digital transformation approach for reverse logistics optimization in a medical waste management system. *Journal of Cleaner Production*, 139703. https://doi.org/10.1016/j.jclepro.2023.139703

5. Gadiiak, L., & Pysarenko, V. (2021). Logistics approach in healthcare management. *Investytsiyi: Praktyka ta dosvid*, (7), 113. https://doi.org/10.32702/2306-6814.2021.7.113

6. Larysa, K., & Serhii, D. (2024). Matematychni modeli ta metody rozmishchennia obiektiv i zonuvannia terytorii v systemakh ekstrenoi lohistyky. *System Technologies*, *6*(149), 107–122. https://doi.org/10.34185/1562-9945-6-149-2023-09

7. Koriashkina, L. S., Dziuba, S. V., Us, S. A., Stanina, O. D., & Odnovol, M. M. (2024). Two-stage problems of optimal location and distribution of the humanitarian logistics system's structural subdivisions. *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*, (1), 130–139. https://doi.org/10.33271/nvngu/2024-1/130

8. Pro zatverdzhennia Poriadku vykorystannia koshtiv, peredbachenykh u derzhavnomu biudzheti dlia vykonannia prohram ta zdiisnennia tsentralizovanykh zakhodiv z okhorony zdorovia [On the approval of the

Procedure for the use of funds provided for in the state budget for the implementation of programs and the implementation of centralized measures for health protection], Postanova Kabinetu Ministriv Ukrainy № 298 (2024). Retrieved from: https://zakon.rada.gov.ua/laws/show/298-2011-п#Text [in Ukraine].

9. Pro orhanizatsiiu kontroliu za tsilovym ta ratsionalnym vykorystanniam materialnykh tsinnostei, otrymanykh tsentralizovano za biudzhetnymy prohramamy ta zakhodamy [On the organization of control over the targeted and rational use of material assets received centrally under budget programs and measures], Nakaz № 232 (2015) [in Ukraine].

10. Pro poriadok vidobrazhennia v bukhhalterskomu obliku informatsii pro rukh zapasiv materialnykh tsinnostei ta yuvileinykh i pamiatnykh monet karbovantsevoho nominalu [On the procedure for displaying in accounting information on the movement of stocks of tangible assets and jubilee and commemorative coins of the ruble denomination], Lyst Natsionalnoho banku Ukrainy № 12-13780 111/1922(2004). Retrieved from: https://zakon.rada.gov.ua/laws/show/v3780500-04#Text [in Ukraine].

11. Pro skhvalennia Kontseptsii reformuvannia mistsevoho samovriaduvannia ta terytorialnoi orhanizatsii vlady v Ukraini [On the approval of the Concept of reforming local self-government and territorial organization of power in Ukraine], Rozporiadzhennia Kabinetu Ministriv Ukrainy № 333-r (2014). Retrieved from: https://zakon.rada.gov.ua/laws/show/333-2014-p#Text [in Ukraine].

12. Pro skhvalennia Stratehii reformuvannia systemy upravlinnia derzhavnymy finansamy na 2022-2025 roky ta planu zakhodiv z yii realizatsii [On the approval of the Strategy for Reforming the State Finance Management System for 2022-2025 and the Action Plan for its Implementation], Rozporiadzhennia Kabinetu Ministriv Ukrainy № 1805-r (2021). Retrieved from: https://zakon.rada.gov.ua/laws/show/1805-2021-p#Text [in Ukraine].

13. Informatsiino-analitychnyi resurs, dlia obminu informatsiieiu shchodo zalyshkiv likarskykh zasobiv ta medychnykh vyrobiv [Informational and analytical resource for the exchange of information on the remains of medicines and medical devices]. (b. d.). *Holovna – Med Data*. Retrieved from: https://meddata.com.ua/ [in Ukraine].

14. Serhieiev, O. S., & Us, S. A. (2023). Modified genetic algorithm approach for solving the two-stage location problem. *Radio Electronics, Computer Science, Control*, (3), 159. https://doi.org/10.15588/1607-3274-2023-3-16

15. Svitlana, U., & Oleksii, S. (2024). An algorithm for solving a two-stage continuous-discrete location problem for medical logistics optimization. *System Technologies*, *5*(148), 71–85. https://doi.org/10.34185/1562-9945-5-148-2023-07

УДК 004.8
DOI https://doi.org/10.32782/IT/2024-4-23

Олена СОБКО
викладач кафедри комп'ютерних наук, Хмельницький національний університет, вул. Інститутська, 11, Хмельницький, Україна, 29016
ORCID: *0000-0001-5371-5788*
Scopus Author ID: *57456803900*

Бібліографічний опис статті: Собко, О. (2024). Нейромережевий пошук і класифікація кіберзалякувань у текстових повідомленнях. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 197–205, doi: https://doi.org/10.32782/IT/2024-4-23

НЕЙРОМЕРЕЖЕВИЙ ПОШУК І КЛАСИФІКАЦІЯ КІБЕРЗАЛЯКУВАНЬ У ТЕКСТОВИХ ПОВІДОМЛЕННЯХ

У статті висвітлено проблему пошуку і класифікації кіберзалякувань у текстових повідомленнях, що є одним із ключових викликів сучасного інформаційного суспільства. Актуальність дослідження зумовлена необхідністю створення ефективних систем, здатних забезпечувати точне, етичне та прозоре нейромережеве виявлення кіберзалякувань. Особливе значення приділяється адаптації таких систем до чутливих тем, як дискримінація за віковими, етнічними, гендерними та релігійними ознаками.

Мета роботи *полягає у створенні комплексного методу до нейромережевих пошуку і класифікації кіберзалякувань у текстових повідомленнях, який передбачає забезпечення репрезентативності даних у датасеті, що використовується для навчання моделі, дотримання етичного принципу справедливості в розробці моделі та можливість інтерпретації результатів моделі щодо типів виявлених кіберзалякувань.*

Новизна *запропонованого підходу полягає у створенні нового методу, який дозволяє не тільки оцінювати наявність кібербулінгу в текстових повідомленнях, але й визначати з високою точністю прояв кожного з видів кібербулінгу, забезпечуючи формування репрезентативно збалансованих датасетів для навчання нейромережевих моделей, що виконується у три етапи. На першому етапі проводиться оцінка репрезентативності датасету для навчання нейромережевих моделей для задачі виявлення та класифікації кіберзалякувань. Зокрема, метод дозволяє мінімізувати відхилення у розподілі даних за класами, що досягає максимуму лише 0,04%. На другому етапі використовуються моделі нейромережевої класифікації: BiLSTM для бінарної класифікації кіберзалякувань, яка демонструє точність 96%, та BERT для мультилейблової класифікації за різними типами кіберзалякувань з точністю 94%. Третій етап передбачає застосування моделі LIME, яка забезпечує візуальну інтерпретацію рішень нейромережі, дозволяючи користувачам отримати пояснення кожного виявленого типу кіберзалякувань.*

Методологія *дослідження базується на поєднанні сучасних підходів до машинного навчання, якісного аналізу репрезентативності даних та використання інтерпретаційних моделей. Інтеграція цих підходів спрямована на створення прозорих і довірених систем виявлення кіберзалякувань, що можуть бути застосовані у реальних умовах.*

Результати *демонструють ефективність запропонованого методу, який не лише підвищує точність і прозорість процесу виявлення та класифікації кіберзалякувань, але й відповідає Цілям Сталого Розвитку № 5, № 10 та № 16, що дозволяє запропонованому комплексному методу бути релевантним для використання в системах, де етичність і точність є важливими.*

Ключові слова: *кіберзалякування, репрезентативність, інтерпретація результатів, BERT, LIME.*

Olena SOBKO
Teacher at the Department of Computer Science, Khmelnytskyi National University, 11, Instytutska Str., Khmelnytskyi, Ukraine, 29016, olenasobko.ua@gmail.com
ORCID: *0000-0001-5371-5788*
Scopus Author ID: *57456803900*

To cite this article: Sobko, O. (2024). Neiromerezhevyi poshuk i klasyfikatsiia kuberzaliakuvan u tekstovykh povidomlenniakh [Neural network search and classification of cyberbullying in text messages]. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 197–205, doi: https://doi.org/10.32782/IT/2024-4-23

NEURAL NETWORK SEARCH AND CLASSIFICATION OF CYBERBULLYING IN TEXT MESSAGES

The article highlights the problem of searching and classifying cyberbullying in text messages, which is one of the key challenges of the modern information society. The relevance of the study is due to the need to create effective systems capable of ensuring accurate, ethical and transparent neural network detection of cyberbullying. Particular importance is given to adapting such systems to sensitive topics such as discrimination on the basis of age, ethnicity, gender and religion.

***The purpose** of the work is to create a comprehensive method for neural network search and classification of cyberbullying in text messages, which involves ensuring the representativeness of the data in the dataset used to train the model, adhering to the ethical principle of fairness in the development of the model and the ability to interpret the results of the model regarding the types of detected cyberbullying.*

***The novelty** of the proposed approach lies in the creation of a new method that allows not only to assess the presence of cyberbullying in text messages, but also to determine with high accuracy the manifestation of each type of cyberbullying, ensuring the formation of representatively balanced datasets for training neural network models, which is performed in three stages. At the first stage, the representativeness of the dataset for training neural network models for the task of detecting and classifying cyberbullying is assessed. In particular, the method allows minimizing deviations in the distribution of data by classes, which reaches a maximum of only 0.04%. At the second stage, neural network classification models are used: BiLSTM for binary classification of cyberbullying, which demonstrates an accuracy of 96%, and BERT for multi-label classification by different types of cyberbullying with an accuracy of 94%. The third stage involves the application of the LIME model, which provides a visual interpretation of the neural network solutions, allowing users to obtain an explanation for each detected type of cyberbullying.*

***The research** methodology is based on a combination of modern approaches to machine learning, qualitative analysis of data representativeness and the use of interpretive models. The integration of these approaches is aimed at creating transparent and trusted cyberbullying detection systems that can be applied in real-world conditions.*

***The results** demonstrate the effectiveness of the proposed method, which not only increases the accuracy and transparency of the cyberbullying detection and classification process, but also meets Sustainable Development Goals No. 5, No. 10 and No. 16, which allows the proposed comprehensive method to be relevant for use in systems where ethics and accuracy are important.*

***Key words:** cyberbullying, neural networks, interpretation of results, BERT, LIME.*

Актуальність проблеми. Кіберзалякування, або булінг у інтернет-середовищі, є серйозною соціальною проблемою сучасності, яка набуває дедалі більшої актуальності з розвитком інтернет-комунікацій. Явище кіберзалякування представляє собою агресивну, образливу або маніпулятивну поведінку в онлайн-просторі, що завдає значної психологічної шкоди жертві (Teng T. H., 2023, C. 55533-55560). Особливо вразливими до кіберзалякувань є підлітки та молодь, які активно користуються соціальними мережами та месенджерами для спілкування, що може призвести до втрати впевненості в собі, депресії чи навіть суїцидальних думок.

Складність пошуку кіберзалякування обумовлена його багатогранністю та контекстуальною природою. Агресивні дії часто приховані за сарказмом, жартами чи культурними особливостями спілкування, що робить їх важкими для однозначної ідентифікації (Unnava S., 2024, C. 200–206). До того ж, великий обсяг текстової інформації, яка щодня створюється в мережі, значно ускладнює процес аналізу. Нерепрезентативні дані можуть призводити до упередженості моделі (систематичної помилки в прогнозах, яка призводить до несправедливих або некоректних результатів), що ставить під сумнів її ефективність та етичність. Використання традиційних методів є малоефективним, тому виникає потреба у застосуванні нейромережевих технологій, здатних аналізувати тексти з урахуванням їхнього змісту, стилю і контексту, допомагаючи у вирішенні цієї проблеми (Pagano T.P., 2024, C. 15).

Аналіз останніх досліджень і публікацій. Проблема пошуку кіберзалякувань у текстових повідомленнях ускладнюється через різноманіття його форм і проявів: від прямих образ до прихованих дій, таких як сарказм або маніпуляції (Собко О.В., 2024, C. 262–265). Сучасні дослідження часто стикаються з обмеженнями, пов’язаними з нерівномірністю даних і недостатньою увагою до локальних особливостей кіберзалякувань, що підкреслює потребу в нових підходах до їх пошуку та класифікації (Krak I., 2024, C. 16–28).

Наприклад, у (Harish D., 2023, C. 1-6) автори розглядаються алгоритми машинного навчання для автоматизації процесу пошуку кіберзалякувань у соціальних мережах. У дослідженні використано різні моделі машинного навчання, такі як згорткові нейронні мережі, підтримуючі векторні машини та логістична регресія для пошуку кіберзалякувань у текстах. Результати

показали, що модель логістичної регресії з оптимізацією за допомогою Stochastic Average Gradient перевершує інші методи в точності визначення кіберзалякувань.

Стаття (Orrù G., 2023, С. 430) описує розробку технологій для пошуку кіберзалякувань у рамках проєкту BullyBuster. Зокрема, вона зосереджується на застосуванні машинного навчання та обробки природної мови для пошуку та класифікації агресивної поведінки в онлайн-спілкуванні. Підхід поєднує методи аналізу настроїв і контексту для підвищення точності пошуку кіберзалякувань, враховуючи лінгвістичні та поведінкові аспекти.

У статті (Samee, N. A., 2023) автори пропонують комбінувати векторні уявлення слів, емоційні характеристики та федеративне навчання для покращення пошуку кіберзалякувань. Автори стверджують, що векторні уявлення слів покращують розуміння контексту, а емоційні характеристики дозволяють виявляти афективні аспекти тексту. Федеративне навчання дозволяє зберігати конфіденційність даних, навчаючи моделі на розподілених пристроях. Використання моделей BERT, CNN, DNN та LSTM показує, що цей підхід перевищує традиційні методи в точності пошуку кіберзалякувань.

Попри значний прогрес у дослідженнях кіберзалякувань, проведений аналіз показує, що існуючі підходи часто мають суттєві недоліки: нерепрезентативні дані, недостатнє врахування контексту повідомлень і складнощі з ідентифікацією типів кіберзалякувань, що створює потребу у розробці нового, відмінного від існуючих, підходу, який враховуватиме репрезентативне формування датасету з метою навчання нейромережевих моделей для пошуку і класифікації кіберзалякувань у текстових повідомленнях з подальшою інтерпретацією отриманих результатів.

Мета роботи полягає у створенні комплексного методу до нейромережевих пошуку і класифікації кіберзалякувань у текстових повідомленнях, який передбачає забезпечення репрезентативності даних у датасеті, що використовується для навчання моделі, дотримання етичного принципу справедливості в розробці моделі та можливість інтерпретації результатів моделі щодо типів виявлених кіберзалякувань.

Виклад основного матеріалу дослідження. Запропонований у статті комплексний метод до нейромережевих пошуку і класифікації кіберзалякувань у текстових повідомленнях складається з трьох етапів, що наведені на рисунку 1. Наведені етапи забезпечують не лише точний пошук та класифікацію кіберзалякувань, а й враховує етичні принципи справедливості для формування датасету, що призначений для навчання нейромереж з метою пошуку кіберзалякувань. Також важливою є інтерпретація результатів, яка надається для поясненності прийнятих нейромережевою моделюю рішень.

На першому етапі відбувається оцінка та коригування репрезентативності датасету для пошуку та класифікації кіберзалякувань, схема та кроки якого наведені на рисунку 2.

Вхідними даними цього етапу є вибірка текстових даних, яка містить певну кількість елементів, що відповідають визначеним цільовим пропорціям за етичними аспектами. Для

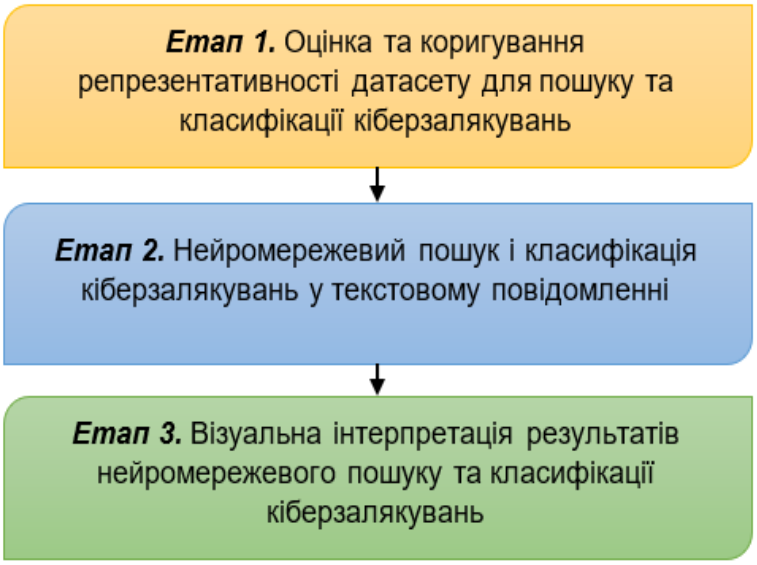


Рис. 1. Схема етапів нейромережевих пошуку і класифікації кіберзалякувань у текстових повідомленнях

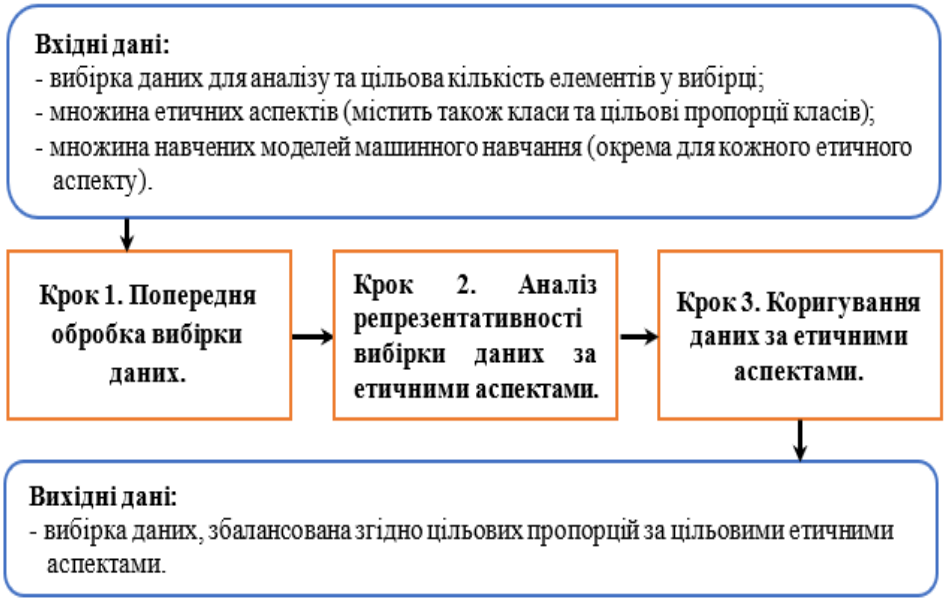


Рис. 2. Схема оцінювання та коригування репрезентативності датасету

кожного аспекту використовується попередньо навчена модель машинного навчання, що виконує аналіз конкретного етичного критерію.

На першому кроці здійснюється попередня обробка даних, яка включає видалення непотрібних елементів (знаків пунктуації, цифр) та некоректних записів (порожніх або неінформативних текстів).

Другий крок полягає в аналізі репрезентативності вибірки, що включає векторизацію елементів, класифікацію за етичними аспектами, оцінку пропорцій класів та пошуку відхилень від заданих пропорцій. Також перевіряється, чи є достатнім обсяг даних для представлення кожного класу.

На третьому кроці здійснюється коригування вибірки, яке включає видалення надлишкових елементів, аугментацію менших за кількість зразків у класах та формування збалансованої вибірки згідно з етичними вимогами. Результатом етапу є вибірка, що відповідає встановленим пропорціям та етичним критеріям.

На другому етапі відбувається нейромережевий пошук і класифікація кіберзалякувань у текстовому повідомленні, схема наведена на рисунку 3.

Перший крок етапу нейромережевого пошуку та класифікації кіберзалякувань у текстовому повідомленні включає попередню обробку вхідного тексту. На цьому кроці текст очищується від зайвих символів, після чого перетворюється у векторне представлення, що підготовлене для обробки нейромережею.

На другому кроці нейромережева модель, яка пройшла навчання для бінарної класифікації кіберзалякувань, аналізує текст на наявність таких ознак, визначаючи інтенсивність проявів кіберзалякування. Якщо рівень прояву кіберзалякування перевищує 50%, текстове повідомлення вважається таким, що містить кіберзалякування, і передається на наступний етап для визначення типу.

Третій крок етапу передбачає використання нейромережевої моделі, яка спеціалізується на визначенні типів кіберзалякувань. Ця модель проводить аналіз текстового повідомлення і визначає відсоткове співвідношення кожного з типів кіберзалякувань.

Результатом цього етапу є загальна оцінка рівня кіберзалякувань у текстовому повідомленні, а також окремі оцінки для кожного типу кіберзалякувань.

На третьому етапі відбувається візуальна інтерпретація результатів нейромережевого пошуку та класифікації кіберзалякувань, схема кроків якого подана на рисунку 4.

Вхідними даними цього етапу є модель трансформерної архітектури, попередньо навчена для мультилейблової класифікації кіберзалякувань, що здатна ідентифікувати різні типи кіберзалякувань, включаючи вікові, гендерні, етнічні, релігійні та загальні категорії. Крім того, використовується інтерпретаційна модель, яка дозволяє пояснити, як окремі слова чи фрази впливають на результати класифікації.

На першому кроці текстове повідомлення проходить через процес токенізації.

Після чого на другому кроці здійснюється прогнозування ймовірностей, що вказують на належність текстового повідомлення до кожного із знайдених типів кіберзалякування.

Рис. 3. Схема нейромережевого пошуку і класифікації кіберзалякувань у текстовому повідомленні

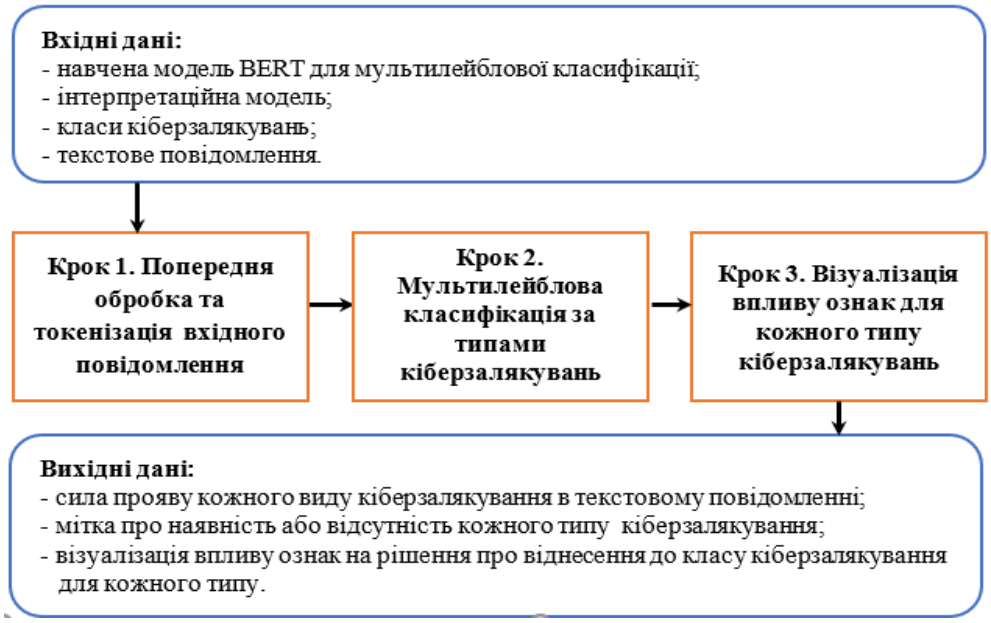


Рис. 4. Схема візуальної інтерпретації результатів нейромережевого пошуку та класифікації кіберзалякувань

Третій крок полягає в інтерпретації результатів, що проводиться за допомогою спеціальної інтерпретаційної моделі, яка візуалізує вплив конкретних слів чи фраз на класифікацію тексту.

Кінцевими результатами цього етапу є ймовірності для кожного типу кіберзалякування та візуальне представлення, що підкреслюють слова, які мали найбільший вплив на класифікацію текстового повідомлення за типами кіберзалякувань.

Таким чином кожен з описаних етапів є частиною запропонованого комплексного методу нейромережевих пошуку і класифікації кіберзалякувань у текстових повідомленнях, що враховує різні групи людей, як за віковими, так і за етнічними чи релігійними ознаками для завдання пошуку типів кіберзалякувань у текстовому повідомленні, а також надає пояснення рішень моделі щодо визначених у текстовому повідомленні типів кіберзалякувань шляхом візуальної інтерпретації.

Експеримент, результати та дискусія. З метою оцінки ефективності запропонованого підходу до нейромережевого визначення цілей пропаганди у текстовому контенті з візуальною аналітикою було розроблено програмне забезпечення. Для апробації методу були використанні набори «Cyberbullying Classification» (Kaggle. Cyberbullying Classification) та «Cyberbullying Detection Dataset» (Kaggle. Cyberbullying Detection Dataset). Однак ці датасети не мають міток щодо статі, віку, релігії та етнічності авторів повідомлень, тому для навчання моделей машинного навчання, що розмічають вхідні дані, використовувались додаткові датасети, які враховують три етичні аспекти принципу справедливості: гендер, вік та релігію (Kaggle.com. Tweet Files for Gender Guessing, Live.european-language-grid.eu. TAG-it Dataset Distribution, Kaggle.com. Cyberbullying Tweets). Оскільки класи в цих датасетах були нерівномірно представлені, вони були збалансовані за чисельністю, щоб уникнути негативного впливу на якість навчання моделей. Кількість зразків для кожного класу в навчальних вибірках для EML за етичними аспектами показана на рисунку 5.

У якості моделей машинного навчання, які б здійснювали розмітку вхідного датасету за етичними аспектами обрано різні архітектури: класифікатор SVM та моделі глибокого навчання BERT, LSTM. Результати обчислення статичних метрик кращих моделей машинного навчання для вікового, гендерного та релігійно аспектів наведено в таблиці 1

Для різних класів було досягнуто різних рівнів роздільної здатності: для релігійної ознаки, при використанні класифікатора BERT, який показав найкращі результати серед навчених моделей для класифікації текстів за релігійним аспектом, дані виявились добре роздільними. За гендерною ознакою, де застосовувався класифікатор LSTM, що продемонстрував кращу ефективність у порівнянні з іншими моделями, дані показали середню роздільність. Водночас, за віковою ознакою, використовуючи класифікатор SVM, дані виявились погано роздільними.

У процесі формування репрезентативного датасету з урахуванням вікового та гендерного етичних аспектів, на основі демографічних підгруп населення України (Idss. Національні демографічні прогнози), було застосовано

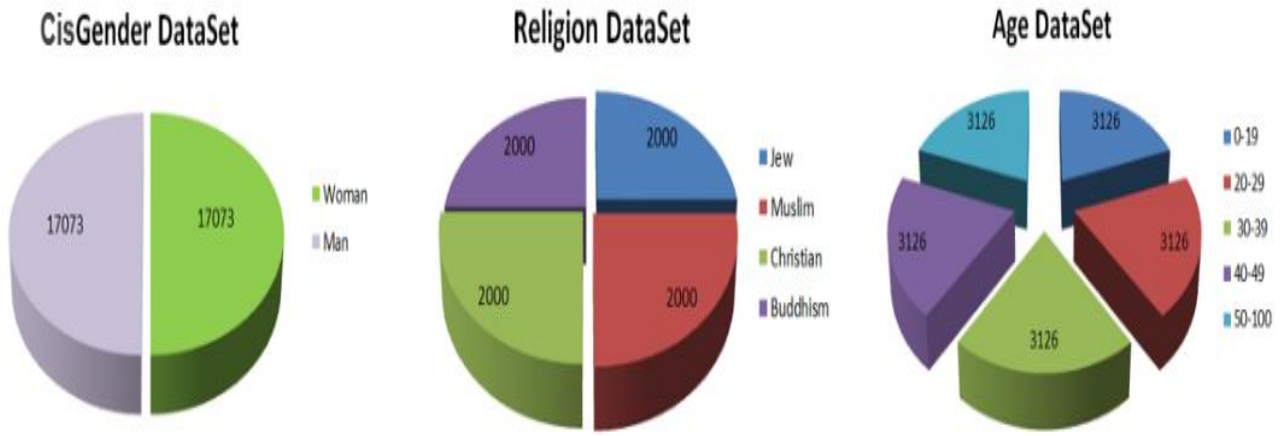


Рис. 5. Класи та кількість зразків у датасетах для навчання моделей машинного навчання

Таблиця 1

Статистичні метрики Accuracy, Precision, Recall та F1-score моделей машинного навчання

Модель машинного навчання	*Accuracy*	*Precision*	*Recall*	*F1-score*
Релігійний етичний аспект				
BERT	0.910	0.980	0.74 0	0.840
Гендерний етичний аспект				
LSTM	0.70	0.770	0.670	0.720
Віковий етичний аспект				
SVM	0.815	0.770	0.779	0.770

метод аугментації для створення збалансованої вибірки текстових даних (таблиця 2).

Відхилення розподілу зразків за віковими та гендерними класами в коригованому датасеті були мінімальними. Зокрема, мінімальне відхилення становило 0.00%, максимальне – 0.04%, а середнє – 0.02%. Це свідчить про високу точність коригування вибірки, що дозволяє забезпечити достатнє представлення кожної категорії без значних диспропорцій.

Для нейромережевих пошуку та класифікації кіберзалякувань у текстових повідомленнях сформований датасет використовується в кілька етапів. Спочатку застосовується модель для бінарної класифікації, яка визначає загальний рівень кіберзалякувань у текстовому повідомленні. Для цієї задачі використано нейромережу BiLSTM, яка досягла високих показників ефективності: Accuracy 0.96, Precision 0.91, Recall 0.93 та F1 Score 0.92. Після цього здійснюється мультилейблова класифікація для визначення конкретних типів кіберзалякувань. Для цього обрано модель BERT, що показала також відмінні результати: Accuracy 0.94, Precision 0.93, Recall 0.93 та F1 Score 0.93. Модель була навчена для пошуку таких типів кіберзалякувань, як вікове, гендерне, релігійне, етнічне та інші.

У дослідженні запропонованого методу було використано текстовий зразок «Your God has no place here. Stick to your country and stop dragging your outdated traditions and religions into ours». Модель BERT визначила ймовірності наявності вікових, етнічних, гендерних, релігійних та інших типів кіберзалякувань у текстовому зразку, зокрема релігійне кіберзалякування мало ймовірність 99.86%, тоді як інші види були значно менш представлені.

Результати мультилейблової класифікації типів кіберзалякувань, отримані за допомогою моделі BERT, використовуються на етапі візуальної інтерпретації за допомогою LIME. Це дозволяє отримати візуальне представлення, де слова виділяються кольорами в залежності від їх впливу на класифікацію, з більш насиченими кольорами для слів із більшим впливом. Додатково, яскравість слів коригується для відображення позитивного або негативного впливу, що дозволяє точніше пояснити прийняті моделі рішення (рисунок 6).

Запропонована інтерпретація результатів нейромережевих пошуку і класифікації кіберзалякувань є важливим інструментом візуальної аналітики, який сприяє забезпеченню прозорості та етичності в рішеннях штучного інтелекту. Вона дозволяє краще зрозуміти, чому і як були прийняті ті чи інші рішення моделлю, що є важливим кроком для запобігання упередженості та неетичних впливів у таких системах.

Висновки. Розроблено комплексний метод нейромережевих пошуку і класифікації кіберзалякувань у текстових повідомленнях, який складається з трьох етапів, спрямованих на аналіз текстового повідомлення з урахуванням етичних аспектів, таких як вікові та етнічні ознаки.

Перший етап передбачає оцінку та коригування репрезентативності датасету, що забезпечує відповідність вибірки етичним вимогам, з мінімальними відхиленнями від ідеального розподілу (0.00%-0.04%). Це дозволяє адаптувати датасет до репрезентативного для подальшого аналізу.

Другий етап передбачає нейромережевий пошук і класифікацію кіберзалякувань у текстовому повідомленні, що забезпечує не лише загальну оцінку рівня кіберзалякувань, але й мультилейблову класифікацію за різними типами, такими як вікові, релігійні та етнічні. Моделі BiLSTM для бінарної класифікації та BERT для мультилейблової класифікації показали високі результати: точність 96% та 94% відповідно.

Таблиця 2

Розподіл зразків у створеній репрезентативній вибірці після проведення аугментації

Вікові демографічні підгрупи	***0-19 років***	***20-29 років***	***30-39 років***	***40-49 років***	***50-100 років***
Відсоткове відношення демографічних груп за гендером та віком у популяції України					
Чоловіки	9.67%	5.64%	8.96%	7.79%	15.56%
Жінки	9.04%	4.53%	7.96%	7.47%	23.38%
Відсоткове відношення демографічних груп за гендером та віком у текстовій вибірці					
Чоловіки	9.65%	5.62%	8.94%	7.80%	15.57%
Жінки	9.05%	4.57%	7.97%	7.45%	23.38%
Одержане відхилення від репрезентативного розподілу					
Чоловіки	0.02%	0.02%	0.02%	0.01%	0.02%
Жінки	0.01%	0.04%	0.01%	0.02%	0.00%

Рис. 6. Інтерпретація результатів нейромережевих пошуку і класифікації типів кіберзалякувань у текстових повідомленнях

Третій етап надає візуальну інтерпретацію результатів пошуку, що забезпечує пояснення рішень моделі через візуальні інструменти. Такий підхід сприяє підвищенню прозорості та довіри до систем штучного інтелекту, особливо для чутливих тем, як кіберзалякування.

Отже, запропонований комплексний метод нейромережевих пошуку і класифікації кіберзалякувань у текстових повідомленнях враховує соціальні фактори, що дозволяє ефективно здійснювати пошук та класифікацію кіберзалякувань та забезпечує поясненність рішень моделі.

ЛІТЕРАТУРА:

1. Teng T. H., Varathan, K. D. Cyberbullying detection in social networks: A comparison between machine learning and transfer learning approaches. *IEEE Access,* vol. 11, 2023. С. 55533–55560.

2. Unnava S., Parasana S. R. A Study of Cyberbullying Detection and Classification Techniques: A Machine Learning Approach. *Engineering, Technology & Applied Science Research,* 14(4), 2024. P. 15607–15613.

3. Pagano T. P., Loureiro R. B., Lisboa F.V.N., Peixoto R. M., Guimarães G.A.S., Cruz G.O.R., Araujo M. M., Santos L. L., Cruz M.A.S., Oliveira E.L.S. Bias and Unfairness in Machine Learning Models: A Systematic Review on Datasets, Tools, Fairness Metrics, and Identification and Mitigation Methods. *Big Data Cogn. Comput.,* 7(1), 2023. P. 15.

4. Собко О. В. Метод інтелектуального пошуку кіберзалякувань у текстовому контенті. *Розвитки інформаційно-керуючих систем та технологій: монографія. Львів-Торунь: Lina-Pres,* 2024. С. 267–287.

5. Krak I., Zalutska O., Molchanova M., Mazurets O., Bahrii R., Sobko O., Barmak O. Abusive Speech Detection Method for Ukrainian Language Used Recurrent Neural Network. *CEUR Workshop Proceedings.* Vol. 3688, 2024. С. 16–28.

6. Harish D., Alamelu M., Manimaran M. Automatic Detection of Cyberbullying on Social Media Using Machine Learning. *In 2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, 2023. С. 1–6.

7. Orrù G., Galli A., Gattulli V., Gravina M., Micheletto M., Marrone S., Sansone C. Development of Technologies for the Detection of (Cyber) Bullying Actions: The BullyBuster Project. *Information,* 14(8), 430, 2023.

8. Samee N. A., Khan U., Khan S., Jamjoom M. M., Sharif M., Kim D. H. Safeguarding Online Spaces: A Powerful Fusion of Federated Learning, *Word Embeddings, and Emotional Features for Cyberbullying Detection. IEEE Access, vol. 11,* 2023. С. 124524–124541.

9. Kaggle.com. Cyberbullying Classification, 2021. URL: https://www.kaggle.com/datasets/andrewmvd/cyberbullying-classification?resource=download. Дата останнього звернення: 2024/12/02.

10. Kaggle.com. CyberBullying Detection Dataset, 2024. URL: https://www.kaggle.com/datasets/sayankr007/cyber-bullying-data-for-multi-label-classification. Дата останнього звернення: 2024/12/02.

11. Kaggle.com. Tweet Files for Gender Guessing, 2019. URL: https://www.kaggle.com/datasets/aharless/tweet-files-for-gender-guessing. Дата останнього звернення: 2024/12/02.

12. Live.european-language-grid.eu. TAG-it Dataset Distribution, 2024. URL: https://live.european-language-grid.eu/catalogue/corpus/8112/download. Дата останнього звернення: 2024/12/02.

13. Cyberbullying Tweets. URL: https://www.kaggle.com/datasets/soorajtomar/cyberbullying-tweets. Last accessed: 2024/10/27. Дата останнього звернення: 2024/12/02.

14. Idss.org.ua. Національні демографічні прогнози 2023. URL: https://idss.org.ua/forecasts/nation_pop_proj. Дата останнього звернення: 2024/12/02.

REFERENCES:

1. Teng, T. H., Varathan, K. D. (2023). Cyberbullying detection in social networks: A comparison between machine learning and transfer learning approaches. *IEEE Access*, vol. 11, C. 55533–55560.

2. Unnava, S., Parasana, S. R. (2024). A Study of Cyberbullying Detection and Classification Techniques: A Machine Learning Approach. Engineering*, Technology & Applied Science Research,* 14(4), P. 15607–15613.

3. Pagano, T. P., Loureiro, R. B., Lisboa, F.V.N., Peixoto, R. M., Guimarães, G.A.S., Cruz, G.O.R., Araujo, M. M., Santos, L. L., Cruz, M.A.S., Oliveira, E.L.S. (2023). Bias and Unfairness in Machine Learning Models: A Systematic Review on Datasets, Tools, Fairness Metrics, and Identification and Mitigation Methods. *Big Data Cogn. Comput.,* 7(1), P. 15.

4. Sobko, O. V. (2024). Doslidzhenia efektyvnosti metodu otsiniuvannia ta koryhulyzatsii reprezentatyvnosti datasetu za FATE-pryntsypom spravedlyvosti [Research on the effectiveness of the method for assessing and adjusting the representativeness of a dataset according to the FATE principle of fairness]. *Perspektyvy suchasnoi nauky: teoriia i praktyka: materialy VIII Mizhnarodnoyi naukovo-praktychnoi konferentsii,* 217–221 [in Ukrainian].

5. Sobko, O. V. (2024). Metod intelektualnoho vyiavlennia kyberzaliakuvan v tekstovomu kontenti [A method for intelligently detecting cyberbullying in text content]. *Rozvytok informatsiino-kervuiuchykh system ta tekhnolohii: monohrafiia. L'viv-Torun: Lina-Pres,* 267–287 [in Ukrainian].

6. Krak, I., Zalutska, O., Molchanova, M., Mazurets, O., Bahrii, R., Sobko, O., & Barmak, O. (2024). Abusive speech detection method for Ukrainian language using recurrent neural network. *CEUR Workshop Proceedings, 3688*, 16–28.

7. Harish, D., Alamelu, M., & Manimaran, M. (2023). Automatic detection of cyberbullying on social media using machine learning. *In 2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA),* 1–6.

8. Orrù, G., Galli, A., Gattulli, V., Gravina, M., Micheletto, M., Marrone, S. & Sansone, C. (2023). Development of technologies for the detection of (cyber) bullying actions: The BullyBuster project. *Information,* 14(8), 430.

9. Samee, N. A., Khan, U., Khan, S., Jamjoom, M. M., Sharif, M., & Kim, D. H. (2023). Safeguarding online spaces: A powerful fusion of federated learning, word embeddings, and emotional features for cyberbullying detection. *IEEE Access,* 11, 124524–124541.

10. Kaggle.com. Cyberbullying Classification, 2021. Retrieved from: https://www.kaggle.com/datasets/andrewmvd/cyberbullying-classification?resource=download. Last accessed: 2024/12/02.

11. Kaggle.com. CyberBullying Detection Dataset, 2024. Retrieved from: https://www.kaggle.com/datasets/sayankr007/cyber-bullying-data-for-multi-label-classification. Last accessed: 2024/12/02.

12. Kaggle.com. Tweet Files for Gender Guessing, 2019. Retrieved from: https://www.kaggle.com/datasets/aharless/tweet-files-for-gender-guessing.

13. Live.european-language-grid.eu. TAG-it Dataset Distribution, 2024. Retrieved from: https://live.european-language-grid.eu/catalogue/corpus/8112/download. Last accessed: 2024/12/02.

14. Cyberbullying Tweets. Retrieved from: https://www.kaggle.com/datasets/soorajtomar/cyberbullying-tweets. Last accessed: 2024/10/27. Last accessed: 2024/12/02.

15. Idss.org.ua. Natsionalni demohrafichni prohnozy 2023. [National demographic projections 2023]. Retrieved from: https://idss.org.ua/forecasts/nation_pop_proj. Last accessed 2024/10/27 [in Ukrainian].

УДК 658.012.4:621 DOI
DOI https://doi.org/10.32782/IT/2024-4-24

Віталій ТУПКАЛО
доктор технічних наук, професор, професор кафедри кібербезпеки, інформаційних технологій та економіки, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе 210, м. Київ, Україна, 02121
ORCID: *0000-0002-6594-530X*

Бібліографічний опис статті: Тупкало, В. (2024). Блоково-композиційний метод формування контрольних сигнатур функціонування цифрових систем на основі властивостей сигнатурного поля TSF[2^n,P^m(x)]. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 206–214, doi: https://doi.org/10.32782/IT/2024-4-24

БЛОКОВО-КОМПОЗИЦІЙНИЙ МЕТОД ФОРМУВАННЯ КОНТРОЛЬНИХ СИГНАТУР ФУНКЦІОНУВАННЯ ЦИФРОВИХ СИСТЕМ НА ОСНОВІ ВЛАСТИВОСТЕЙ СИГНАТУРНОГО ПОЛЯ TSF[2^N,P^M(X)]

В контексті посилення актуальності проблеми забезпечення надійності функціонування різноманітних цифрових інформаційно-вимірювальних систем комплексів управління об'єктами критичної інфраструктури в умовах зовнішнього шкідливого впливу (кібератак), актуальною стає задача розробка цих комплексів управління із властивістю контролепридатності структурних вузлів (пристроїв) у реальному масштабі часу їх функціонування. При цьому визначальними умовами рішення задачі синтезу повинні бути дві: досягнення високої здатності апаратного функціонального контролю виявляти багатократні помилки у двійкових кодах при їх передачі, обробці та зберіганні та апаратна надлишковість функціонального контролю повинна бути мінімальною. Тому на базовому схемотехнічному рівні побудови цифрових систем виникає, в першу чергу, необхідність подальшого розвитку (нового етапу) схемотехніки мікросхем середнього та високого ступеня інтеграції.

Мета роботи. *На основі критичного аналізу ряду звісних методів функціонального апаратного контролю запропоновано схемотехнічний сигнатурно-функціональної підхід щодо створення нового покоління мікросхем середнього та високого ступеня інтеграції із вбудованими вузлами функціонального сигнатурного функціонального контролю, які б забезпечували високу достовірність контролю за мінімальної схемотехнічної (апаратної) надлишковості.*

Метод. *Рішення задачі синтезу ґрунтується на використанні математичного апарату авторської сигнатурної алгебри сигнатурного поля В. Тупкало TSF[2^n,P^m(x)].*

Наукова новизна. *Розроблено нове методичне рішення задачі синтезу формувачів контрольних сигнатур паралельного комбінаційного типу для багато розрядних операндів шляхом їх агрегації з окремих уніфікованих типових блоків малої розрядності. Доказана можливість двох способів такої агрегації при побудові формувачів сигнатур комбінаційного типу.*

Висновки. *Виходячи з двох запропонованих варіантів використання моделей паралельних формувачів сигнатур комбінаційного типу, можна запропонувати два можливих напрямків подальшого розвитку схемотехніки мікросхем середнього та високого ступеня інтеграції для побудови цифрових інформаційно-вимірювальних систем комплексів управління широкого призначення. По-перше, запропоновані моделі можуть бути реалізовані як окремі серії мікросхем вузлів сигнатурного контролю для побудова функціонально контролепридатних (самоконтролюючих у реальному часі функціонування) цифрових інформаційно-вимірювальних систем комплексів управління, по-друге – як функціональна інтегрована складова низки базових мікросхем схемотехніки цифрових систем. Подальшим напрямом розвитку теми статті є розробка методів схемотехніки сигнатурного функціонального контролю цифрових інформаційно-вимірювальних систем комплексів управління на основі множини сигнатурних полів В. Тупкало TSF[2^{kn},P^m(x)].*

Ключові слова: *функціональний сигнатурний контроль, методи контролю цифрових систем, функціональна контролепридатність, кіберстійкість.*

Vitaliy TUPKALO
Doctor of Technical Sciences, Professor, Professor at the Department of Cybersecurity, Information Technologies and Economics, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121
ORCID: *0000-0002-6594-530X*

To cite this article: Tupkalo, V. (2024). Blokovo-kompozytsiinyi metod formuvannia kontrolnykh syhnatur funktsionuvannia tsyfrovykh system na osnovi vlastyvostei syhnaturnoho polia TSF[2^n,P^m(x)] [Block-composition method for forming control signatures of the functioning of digital systems based on the properties of the signature field TSF[2^n,P^m(x)]]. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 206–214, doi: https://doi.org/10.32782/IT/2024-4-24

BLOCK-COMPOSITION METHOD OF FORMING CONTROL SIGNATURES OF DIGITAL SYSTEMS OPERATION BASED ON THE PROPERTIES OF THE SIGNATURE FIELD TSF[2^N,P^M(X)]

In the context of increasing relevance of the problem of ensuring the reliability of the functioning of various digital information and measuring systems of control complexes by critical infrastructure objects in conditions of external harmful influence (for example, cyberattacks), the task of developing these control complexes with the property of controllability of structural units (devices) in the real time scale of their functioning becomes relevant. At the same time, the determining conditions for solving the synthesis problem should be two: achieving a high ability of hardware functional control to detect multiple errors in binary codes during their transmission, processing and storage, and hardware redundancy of functional control should be minimal. Therefore, at the basic circuit level of building digital systems, there is, first of all, the need for further development (a new stage) of the circuitry of microcircuits of medium and high integration.

Purpose of the work. *Based on a critical analysis of a number of well-known methods of functional hardware control, a circuit-technical signature-functional approach is proposed to create a new generation of microcircuits of medium and high integration with built-in functional signature functional control nodes, which would provide high reliability of control with minimal circuit-technical (hardware) redundancy.*

Method. *The solution to the synthesis problem is based on the use of the mathematical apparatus of the author's signature algebra of the field V. Tupkalo TSF[2^n,P^m(x)].*

Scientific novelty. *A new methodological solution to the problem of synthesizing control signature generators of the parallel combinational type for multi-bit operands by aggregating them from separate unified typical blocks of small bit size has been developed.*

Conclusions. *Based on the two proposed options for using models of parallel combinational signature generators, it is possible to propose two possible directions for further development of circuit design of microcircuits of medium and high degree of integration for building digital information and measuring systems of control complexes of wide purpose. Firstly, the proposed models can be implemented as separate series of microcircuits of signature control nodes for building functionally controllable (self-controlled in real time functioning) digital information and measuring systems of control complexes, secondly – as a functional integrated component of a number of basic microcircuits of circuit design of digital systems. A further direction of development of the topic of the article is the development of circuit design methods of signature functional control of the digital information and measuring systems based on the set of signature fields V. Tupkalo TSF[2^{kn},P^m(x)].*

Key words: *functional signature control, digital systems control methods, functional controllability, cyber resilience.*

Аналіз останніх досліджень і публікацій. З аналізу стану сучасної бази мікросхем для побудови цифрових інформаційно-вимірювальних пристроїв та систем (Прищепа, Погребняк, 2006; Бондаренко, Бородін, Карнаушенко, 2020; Піддубний, Товкач, 2021) у контексті забезпечення їх ефективної контролепридатності при експлуатації можна зробити висновок, що існуючі серії мікросхем є тільки функціонально-орієнтованими для синтезу множини цифрових інформаційно-вимірювальних систем комплексів управління (ЦІВСКУ) за різним функціонально-цільовим призначенням. Тому в умовах можливого зовнішнього шкідливого впливу (наприклад, кібератак) на ЦІВСКУ актуальною стає проблема подальшого розвитку (нового етапу) схемотехніки мікросхем середнього та високого ступеня інтеграції. Одним з перспективних напрямів рішення цієї проблеми є розробка нового покоління серій мікросхем з вбудованими елементами функціонального контролю (ФК) щодо можливості на їх основі розробляти вузли апаратного ФК для об'єктно-орієнтованих цифрових пристроїв та систем. При цьому визначальними умовами рішення задачі синтезу повинні бути дві: досягнення високої здатності апаратного ФК виявляти багатократні помилки у двійкових кодах при їх передачі, обробці та зберіганні, а також апаратна надлишковість ФК повинна бути мінімальною. Тому, на нашу думку, є необхідність у проведенні обґрунтування та формування пропозиції щодо подальшого напряму розвитку системотехніки мікросхем середнього та високого ступеня інтеграції із вбудованими апаратними елементами функціонального контролю з можливістю на їх основі розробляти цифрові інформаційно-вимірювальні пристрої та системи комплексів

управління критичної інфраструктури з високою функціональною контролепридатністю у реальному часі функціонування.

Виклад основного матеріалу дослідження. В контексті визначеної проблеми забезпечення функціональної контролепридатності цифрових пристроїв та систем відомий ряд методів функціонального контролю, в яких формуються двійкові контрольні характеристики (сигнатури), а саме:

– контроль цифрових кодів на парність / непарність (контроль однократних помилок у блоках n-розрядних двійкових послідовностей шляхом сумування їх одиниць по модулю два) (Бондаренко, Глушко, Меньков, 2003);

– числовий (двійковий) контроль (Локазюк, 1996; Чегренець, Руденко, 2016);

– програмно-логічні методи контролю ходу виконання програм (Говорущенко, Гнатчук,2010; Романкевич, 2016);

– сигнатурний контроль виконання логічних та арифметичних операцій на основі апарату синтезу сигнатурної алгебри (Тупкало,2021; Тупкало, 2024).

Результат аналізу цих методів представлений в таблиці 1.

З аналізу даних Таблиці 1 можна зробити наступні висновки:

1) загальний принцип методів ФК – відстеження правильності виконання операційних дій

Таблиця 1

Аналіз відомих методів ФК, в яких формуються двійкові контрольні характеристики (сигнатури)

№	Метод	Сутність методу	Контрольна характеристика
1.	Контроль цифрових кодів на парність / непарність (Бондаренко, Глушко, Меньков, 2003)	Сумування по модулю два одиниць у старших n – розрядах поточних контрольованих двійкових послідовностей з подальшим порівнянням результату з еталонним значенням заздалегідь сформованого (n + 1)-го контрольного розряду.	Один додатковий розряд (1 або 0) для супроводу n- розрядних двійкових послідовностей. Недолік – можливість виявити тільки однократні помилки у блоках n-розрядних двійкових послідовностей.
2.	Числовий (двійковий) контроль (Локазюк, 1996; Чегренець, Руденко, 2016)	Здійснюється за допомогою контрольних k-розрядних кодів, що представляють собою остачу від ділення двійкових n-розрядних чисел на деякий модуль p (двійкове число) з подальшим порівнянням результату остачі з еталонним значенням заздалегідь сформованої k-розрядної остачі. (n + 1) – го контрольного розряду.	Двійковий k-розрядний код, який формується відповідним вузлом ділення (апаратна надлишковість). Величина модуля p, від якої залежить величина кратності виявлених операційних помилок, повинна бути по можливості невелика, з тим щоб код залишку від ділення на число p будь-яких чисел не вимагали великого обсягу обладнання для реалізації числового контролю.
3.	Програмно-логічні методи контролю ходу виконання програм (Говорущенко, Гнатчук,2010; Романкевич, 2016)	Кожній ділянці (фрагменту адресації кортежу команд) перед виконанням присвоюється свій m-розрядний код, який попередньо запам'ятовується. В кінці виконання кожної ділянки здійснюється контроль на «свій код». Якщо коди не співпадають, то це свідчить про збій правильного виконання ділянки.	Контрольний m- розрядний двійковий код, як остача (сигнатура) ділення поліному контрольованого n -розрядного коду (n > m) на незвідний примітивний поліном $P^m(x)$ ступеню m. При цьому вибір величини ступеню поліному m визначає досяжну кратність виявлення помилки контролю на «свій код». Формування сигнатур відбувається за допомогою m – розрядного регістру зсуву зі зворотними зв'язками відповідно поліному $P^m(x)$. Недолік – затримка у часі формування коду сигнатур контрольованого кортежу ходу виконання програм.
4.	Сигнатурний контроль виконання логічних та арифметичних операцій на основі апарату логічного синтезу сигнатурної алгебри (Тупкало, 2021; Тупкало, 2024)	*Базовою операцією сигнатурного контролю є логічне співвідношення сигнатур вхідних поточних операндів з сигнатурою кінцевого результату виконання певних цільових операцій.* Реалізація методу зводиться до побудови контрольного вузла комбінаційного типу шляхом відповідної комутації функціонально закінчених комбінаційних елементів з кінцевої множини (стандартного набору) [9].	Формування m – розрядних сигнатур, як остачі ділення поліному контрольованого n -розрядного коду числа (n > m) на незвідний примітивний поліном $P^m(x)$ ступеню m: $A(x) \equiv sigA(x) = A(x) mod\ P(x)$. Синтез формувача сигнатур (sig) відбувається вузлом комбінаційної логіки паралельного типу (Tupkalo, Cherepkov, 2023)

цифрових пристроїв та систем, заснований на формуванні та порівнянні еталонних та поточних контрольних характеристик (сигнатур) операційних двійкових n- розрядних послідовностей (кодів);

2) в контексті визначеної вище проблема подальшого розвитку (нового етапу) схемотехніки мікросхем середнього та високого ступеня інтеграції з вбудованими елементами ФК в першу чергу слід звернути увагу на 3 і 4 методи ФК з точки зору принципової можливості забезпечення ефективної контролепридатності (кратність виявлення помилок у двійкових операційних кодах) цифрових пристроїв та систем, а саме: m- розрядна сигнатура формується, як остача ділення поліному контрольованого n-розрядного коду числа на обраний (твірний) незвідний примітивний поліном $P^m(x)$ ступеню m (n > m). При цьому в рамках сигнатурного поля Тупкало (Tupkalo signature field) $TSF[2^n, P^m(x)]$ кратність виявлення помилок у двійкових кодах визначається звісним вибором (Тупкало, Черепков, 2024) величини цього ступеню при співвідношенні $n = (2^m – 1)$;

3) метод сигнатурного контролю 4 може одночасно задовольнити, як вимогу виявлення багатократних помилок у двійкових операційних кодах, так і вимогу мінімальної апаратної надлишковості контролюючого вузла, а саме: базовою операцією формування сигнатур є остачі ділення поліномів контрольованих n – розрядних кодів чисел (n > m) на незвідний примітивний поліном $P^m(x)$ ступеню m. При цьому синтез апаратної надлишковості ФК може звестися до побудови контрольного вузла шляхом відповідної комутації функціонально закінчених комбінаційних елементів з кінцевої множини (стандартного набору). Тому цей синтез запропоновано здійснювати на основі використання апарату логічного синтезу авторської сигнатурної алгебри поля $TSF[2^n, P^m(x)]$ (зокрема, методики синтезу формувача сигнатур паралельного типу на основі використання сигнатурної утворюючої матриці незвідного примітивного полінома (Тупкало, Черепков,2023).

Згідно вищезазначеної мети дослідження пропонуються наступні визначення.

Визначення 1. Сигнатурний апаратний функціональний контроль – це вбудовані в структурні функціональні частини об'єктів контролю цифрових систем вузли сигнатурного контролю комбінаційного типу з необхідними вимогами до ефективності контролю.

Визначення 2. Сигнатурна функціональна контролепридатність ЦІВСКУ – це придатність системи до виявлення помилок у реальному масштабі часу при виконанні її цільових вимірювальних (обчислювальних) функцій $f_i \in F(X_{[L]})$ за допомогою певних вбудованих засобів сигнатурного функціонального контролю.

В контексті сутності визначення 2 модель сигнатурного функціонального контролю ЦІВСКУ представлена на рис. 1

Виходячи з вищезазначеного, сутність задачі синтезу визначимо так: нехай для заданого об'єкту функціонального контролю G доступними для контролю є тільки його входи та виходи. Треба визначити метод синтезу апаратної надлишковості комбінаційного типу згідно моделі рис. 1 при умові, щоб необхідна ефективність сигнатурного функціонального контролю (СФК) забезпечувалося би і при довільному співвідношенні k = L / n при обраному твірному незвідному примітивному поліномі $P^m(x)$ ступеню m. Тобто, не було б обмежень при реалізації СФК на величину розрядності L двійкових операндів об'єкту контроля.

З метою реалізації зазначеної задачі синтезу пропонується доказ наступної теореми (сигнатурна композиційна теорема Віталія Тупкало).

Теорема 1. Сигнатура L – розрядного двійкового числа U_L, що є послідовною конкатенацією з k окремих n- розрядних блокових частин U_{in} (i = 1,2, …) при обраному незвідному примітивному твірному поліномі сигнатур $P^m(x)$, дорівнює сумі за модулем два сигнатур його k окремих n- розрядних блокових частин U_{in} при умові, що L = kn, $n = (2^m – 1)$:

$$\mathrm{sig}\, U_L = \mathrm{sig}\,(U_{1n} \prec U_{2n} \prec \ldots \prec U_{kn}) =$$

$$= \mathrm{sig}\, U_{1n} \oplus \mathrm{sig}\, U_{2n} \oplus \ldots \oplus \mathrm{sig}\, U_{kn}, \qquad (1)$$

де $\prec$ – операція конкатенації та «склеювання» окремих n-розрядних блокових частин Uin у порядку їхньої послідовності (відношення передування) від молодших до старших розрядів числа UL.

Доказ Теореми 1. Введемо наступне визначення.

Визначення 3. α-кратна сигнатура n-розрядного числа $\mathrm{sig}^{\alpha}A_n$ – це сигнатура перетвореного числа виду $(A_n \prec W_{kn})$ – де W_{kn} – число нульової

kn – розрядності (k = 1,2, …).

Наприклад, для 7-ми розрядного числа A_7 = = 0111101 його двократною сигнатурою є його перше перетворення sig (0111101.0000000) = = sig^2(0111101), трикратною сигнатурою є друге перетворення sig (0111101.0000000.0000000) = = sig^3(0111101) і т.д.

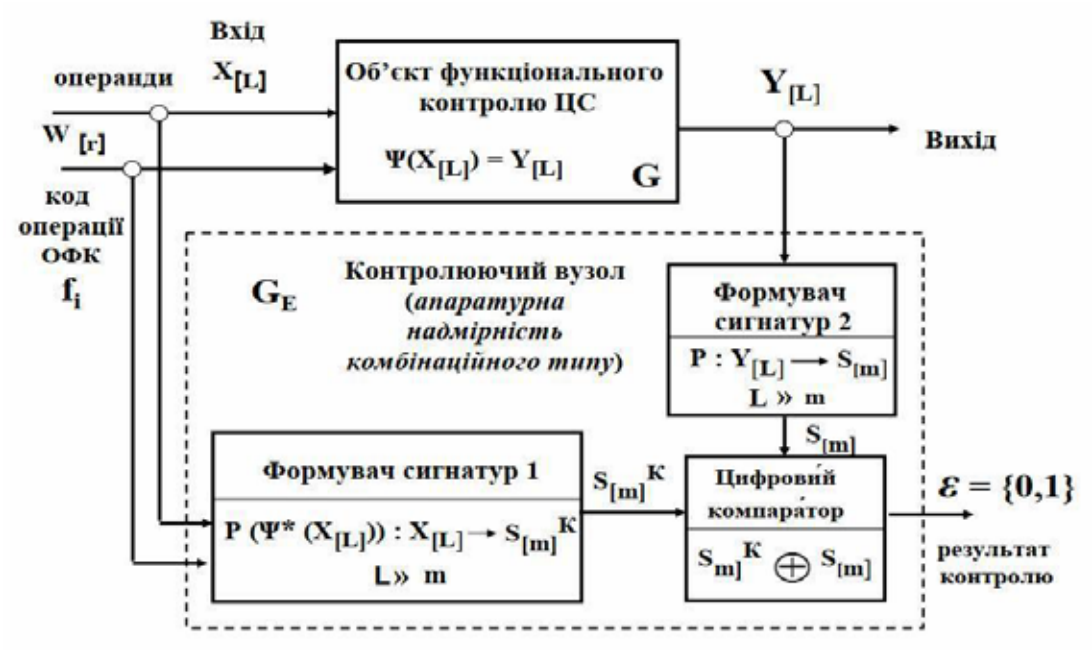


Рис. 1. Модель сигнатурного функціонального контролю (авторська модель)

G_E – вузол сигнатурного контролю, що складається з двох m-розрядних формувачів сигнатур (ФС); m-розрядного компаратора (КК) двох контрольних сигнатур; f_i – управлінські коди вибору виду контрольованих операцій з множини F.

Тоді, якщо двійкове kn-розрядне число U_L є результатом конкатенації з k окремих n- розрядних чисел

$$U_L = (U_{1n} \prec U_{2n} \prec \ldots \prec U_{kn}), \qquad (2)$$

то на основі вище наведеного визначення, сигнатура результату конкатенації цих чисел є сумою за модулем два (в рамках розрядності L = kn) їх відповідних перетворень ($U_{in} \prec W_{in}$) у низці (послідовності) формування конкатенації. Тобто,

$$\mathrm{sig}U_L = \mathrm{sig}\, U_{1n} \oplus \mathrm{sig}^2 U_{2n} \oplus \mathrm{sig}^3 U_{3n} \oplus \ldots \oplus \mathrm{sig}^{(k-1)} U_{kn}, \qquad (3)$$

що і треба було доказати.

Перевіримо справедливість рівняння (3) по наступним вхідним даним:

для формування сигнатур використано регістр зсуву зі зворотними зв'язками за видом твірного поліному сигнатур $P^3(x) = x^3 + x + 1$; n = 7;

$$U_L = 0000111.0010101.1000110.$$

Згідно (3) маємо: U_{17} **=** 1000110; U_{27} **=** 0010101; U_{37} = 0000111.

Тоді при використанні Додатку А маємо:

$$\mathrm{sig}\, U_L = (00001110010101100011 0)_{21} = \mathbf{110} =$$

$$= \mathrm{sig}^3 (0000111) \oplus \mathrm{sig}^2 (0010101) \oplus \mathrm{sig}1000110 =$$

$$= \mathrm{sig}\, (0000111.0000000.0000000) \oplus$$

$$\mathrm{sig}\, (0010101.0000000) \oplus \mathrm{sig}\, (1000110\ ) =$$

$$= 101 \oplus 011 \oplus 000 = \mathbf{110}.$$

Ліва і права частина рівняння співпали.

Відносно рівнянню (3) слід зауважити, що оскільки операція формування m-розрядних сигнатур n-розрядних чисел на основі використання незвідного примітивного поліному $P^m(x)$ ступеню m при умові співвідношення n = (2^m – 1) є лінійною (Тупкало, Черепков, 2024) і сигнатури створюють кільце сигнатурного поля TSF [2^n, $P^m(x)$], то для рівняння (3) справедливим є застосування принципу суперпозиції за модулем два:

$$\mathrm{sig}\, U_L = \mathrm{sig}\, (U_{1n} \oplus U_{2n} \oplus \ldots \oplus U_{kn}). \qquad (4)$$

Перевіримо справедливість рівняння (4) при попередніх вхідних даних і використовуємо таблицю поля TSF[7,$P^3(x)$] у Додатку:

$$\mathrm{sig}\, U_L = 110 =$$

$$= \mathrm{sig}\, (1000110 \oplus 0010101 \oplus 0000111) =$$

$$= \mathrm{sig}\, (1010100) = \mathbf{110}.$$

Результати перевірок обох рівнянь (3) і (4) співпали.

Приклад рішення задачі синтезу та пропозиції

Зауваження до теореми 1. Відносно обов'язковості умови L = nk Теореми 1 слід

звернути увагу на наступне. Якщо після конкатенації окремих двійкових чисел у моноблок U_L загальна кількість розрядів моноблока не дорівнює величині L = kn, то необхідно провести доповнення недостатньої кількості розрядів нульовими розрядами зі сторони старших розрядів моноблоку U_L. Пояснення цього зауваження зробимо на прикладі блок-схеми шістнадцяти розрядного арифметичного двійкового суматора, представленого на рис. 2.

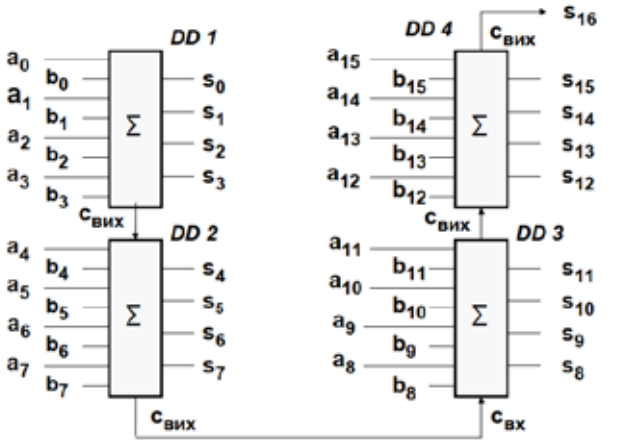


Рис. 2. Блок-схема шістнадцяти розрядного арифметичного двійкового суматора

Приклад 1. Вхідні дані: A_{16} = 1 1 0 1 0 0 1 1 0 1 1 1 0 1 0 1;

B_{16} = 1 1 1 1 0 1 1 1 0 1 0 1 0 0 0 1;

$D_{17} = (A + B)_{17}$ = 1 1 1 0 0 1 0 1 0 1 1 0 0 0 1 1 0 ;

$P^3(x) = x^3 + x + 1$.

Тоді маємо: $n = 2^m – 1 = 2^3 – 1 = 7$; L_n / n = 17: 7 = 2 і остача 3.

Цей результату в контексті вищевикладеного зауваження до теореми 1 відносно двох альтернативних рівнянь (1) і (4) дає підстави стверджувати, що побудова вузла сигнатурного контролю може мати два варіанти.

Варіант № 1. Реалізація вузла сигнатурного ФК ґрунтується на рівнянні (1). Тобто, для виконання умови теореми 1 щодо організації сигнатурного контролю операції складання двійкового 16- ти розрядного суматора рис. 2 за mod $P^3(x)$ необхідно мати вузол контролю, який складається з трьох 7-ми розрядних по входу і трьома виходами (коду сигнатур при m = 3) формувачів сигнатур комбінаційного паралельного типу. При цьому цей варіант враховує сутність вищезазначене зауваження до теореми 1. Графічне пояснення сутності формування коду нульового доповнення при вхідних даних прикладу 1 представлено на рис. 3. Таким чином, на входи старших розрядів третього формувача сигнатури блоку U_{3n} необхідно подати код нульового доповнення $Q^4 = q_3q_2q_1q_0$.

Згідно вхідних даних прикладу 1 доказ коректності визначення варіанту № 1 синтезу апаратної надлишковості ФК для 16- ти розрядного арифметичного двійкового суматора на основі рівняння (1) зводиться до наступної послідовності обчислень:

$$\mathrm{sig}D_{17} = \mathrm{sig}\, U_{17} = \mathrm{sig}\,(11100101011000110)_{17} =$$

$$= \mathrm{sig}(\mathbf{0000}111.0010101.1000110)_{21} =$$

$$= \mathrm{sig}\,(0000111)_7 \oplus \mathrm{sig}\,(0010101)_7$$

$$\oplus\ \mathrm{sig}\,(1000110)_7.$$

Якщо для формування сигнатур було б використано регістр зсуву зі зворотними зв'язками за видом твірного поліному сигнатур $P^3(x) = x^3 + x + 1$ (див. Рис. 4,а), то на 17-му такті зсуву було б отримано результат:

$$\mathrm{sig}\,(11100101011000110)_{17} = 110.$$

У нашому випадку для формування сигнатур паралельним способом використовується формувач сигнатур (7x3) поля TSF[7,$P^3(x)$]

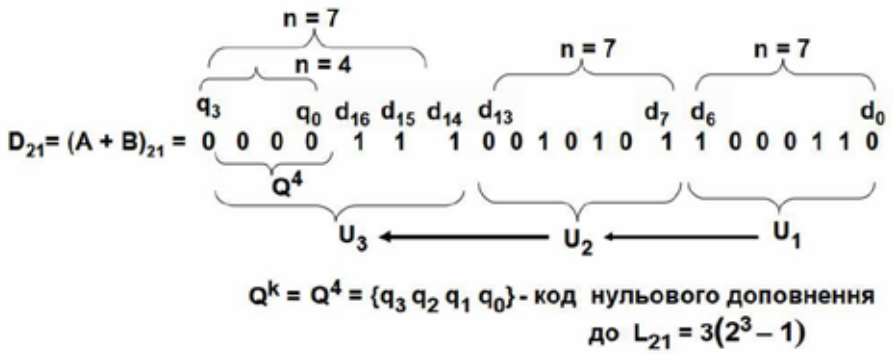


Рис. 3. Сутність формування коду нульового доповнення

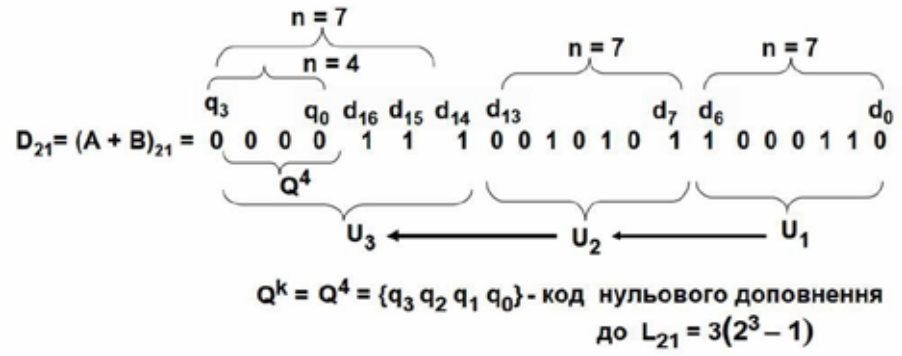


Рис. 4. Формування сигнатур:

а) у послідовний спосіб з використанням регістру зсуву зі зворотніми зв'язками;
б) у паралельний спосіб з використанням формувачів комбінаційного типу.

комбінаційного паралельного типу, який представлений на рис. 5. Тому, повертаючись до рівняння (3) і використовуючи таблицю В. Тупкало поля TSF[7,P^3(x)] у Додатку А, отримуємо:

$$\text{sig}\,(0000111)_7 \oplus \text{sig}\,(0010101)_7 \oplus \text{sig}\,(1000110)_7 =$$
$$= 101 \oplus 011 \oplus 000 = 110,$$

тобто є повний збіг із попереднім отриманим результатом.

Технічна реалізація паралельного формувача сигнатур для 21-го розрядного моноблока U_{21} зводиться до побудови комбінаційного вузла 3(7x3) згортки трьох сигнатур від трьох 7-ми розрядних блоків 21-го розрядного моноблока U_{21} за модулем два пірамідального типу з числом функціональних ступеней згортки, рівного двом. При цьому на входи чотирьох старших розрядів формувача сигнатури третього блоку U_{3n} необхідно подати код нульового доповнення $Q^4 = q_3q_2q_1q_0$. При цьому час виконання операції формування 3-х розрядної сигнатури моноблока U_{21} становить вчетверенну величину затримки двовходового елемента сума за модулем два М2.

Варіант № 2. Реалізація вузла сигнатурного ФК грунтується на основі рівняння (4). Згідно вхідних даних прикладу 1 доказ коректності визначення варіанту № 2 синтезу апаратної надлишковості ФК для 16-ти розрядного арифметичного двійкового суматора на основі рівняння (4) зводиться до наступної послідовності обчислень з використанням таблиці поля TSF[7,P^3(x)] у Додатку А, отримуємо:

$$\text{sig}_{[3]}A_{[7]} = \begin{cases} S_3 = a_7 \oplus a_5 \oplus a_4 \oplus a_3 \\ S_2 = a_6 \oplus a_4 \oplus a_3 \oplus a_2 \\ S_1 = a_5 \oplus a_3 \oplus a_2 \oplus a_1 \end{cases}$$

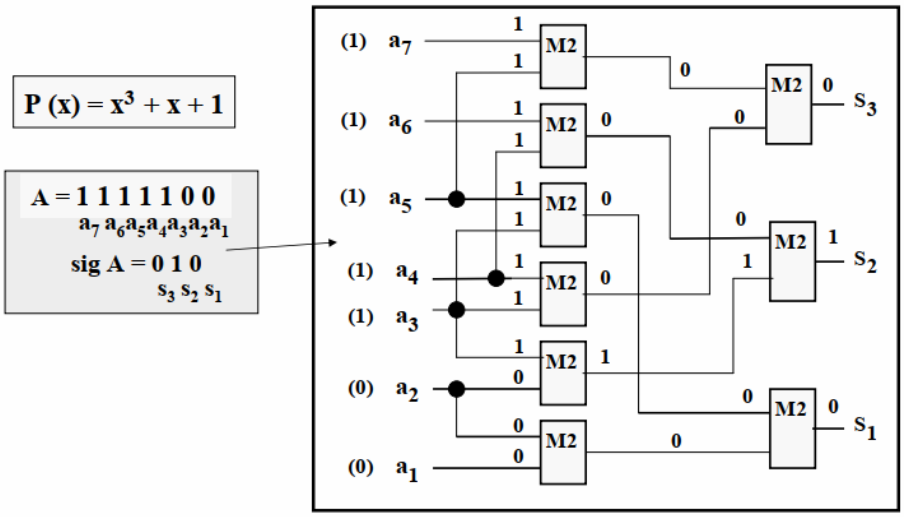


Рис. 5. Формувач сигнатур (7x3) поля TSF[7,P^3(x)] комбінаторного паралельного типу (авторська модель)

$$\text{sig } U_{17} = \text{sig } (11100101011000110)_{17} =$$
$$= \text{sig } (0000111.0010101.1000110)_{21} = \mathbf{110} =$$
$$= \text{sig } (0000111 \oplus 0010101 \oplus 1000110) =$$
$$\text{sig } (1010100)_7 = 110, \qquad (4)$$

тобто є повний збіг із отриманим результатом варіанта № 1.

Технічна реалізація паралельного формувача сигнатур (7х3) комбінаційного типу за цим варіантом зводиться до побудови комбінаційного вузла згортки за модулем два трьох 7-ми розрядних блоків 21-го розрядного моноблока U_{21} з урахуванням нульове доповнення $Q^4 = q_3q_2q_1q_0$ блоку U_{3n} шляхом подачі на відповідні входи суматора сигнал логічної одиниці. З отриманої на першому функціональному ступені згортки за модулем два 7-ми розрядного результату W_7 формується контрольна 3-х розрядна сигнатура $\text{sig}W_7$, що дорівнює сигнатурі 21-го розрядного моноблока U_{21}. Таким чином, контрольний вузол пірамідального комбінаційного типу за варіантом № 2 має, як і варіант № 1, два функціональні ступеня згортки. При цьому час виконання операції формування сигнатури моноблока $(A+B)_{21}$ становить також вчетверенну величину затримки двовходового елемента згортки за модулем два М2. Вузол сигнатурного контролю за варіантом № 2 представлений на рис. 6.

Виходячи з вищезазначеного в контексті теореми 1 слід зауважити, що поле $TSF[2^n, P^m(x)]$ при співвідношенні $n = (2^m – 1)$ є окремим випадком сімейства полів В. Тупкало (Tupkalo Signature Field) $TSF[2^{kn}, P^m(x)]$. При варіації змінних поля n, m, k в загальному сенсі $TSF[2^{kn}, P^m(x)]$ – це

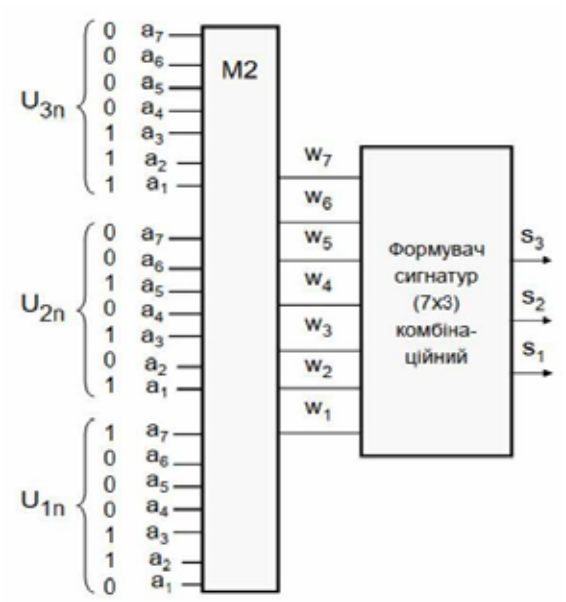


Рис. 6. Вузол сигнатурного контролю за варіантом № 2

множина (сімейство) сигнатурних полів В. Тупкало, кожне з яких є скінченими. Згідно даного твердження пропонується наступне визначення.

Визначення 3. Сигнатурне поле В. Тупкало (Tupkalo Signature Field) TSF $[2^{kn}, P^m(x)]$ є кінцевим полем (множиною) сюрективного відображення двох взаємно пов'язаних підможин: підмножини L = kn – розрядних (k = 1,2,…) двійкових чисел U_L та підмножини значень (класів) m – розрядних сигнатур $\text{sig}U_L(x) = U_L(x) \bmod P^m(x)$ при умові L > m, $n = (2^m – 1)$; $P^m(x)$ – незвідний примітивний поліном поля (породжувальний поліном m – розрядних сигнатур поля).

ЛІТЕРАТУРА:

1. Прищепа М. М., Погребняк В. П. Мікроелектроніка. У 3-х частинах. Ч. 2. Елементи мікросхемотехніки: Навч. посіб / М. М. Прищепа. Київ: Вища школа, 2006. 503 с.

2. Бондаренко І. М., Бородін О. В., Карнаушенко В. П. Сучасна компонентна база електронних систем: навч. посібник для студентів ЗВО. Харків: ХНУРЕ, 2020. 268 с.

3. Піддубний В. О., Товкач І. О. Елементна база радіоелектронної апаратури: В 4 ч. Ч. 4. Основи мікроелектроніки: навч. посіб. для студ. спеціальності 172 «Телекомунікації та радіотехніка». Київ :КПІ ім. Ігоря Сікорського, 2021. 98 с.

4. Бондаренко І. М., Глушко А. П., Меньков О. М. Коди та кодування. Навч. посібник. Харків. ХІ ВПС, 2003. 116 с.

5. Локазюк В. М. Контроль і діагностування обчислювальних пристроїв та систем: Навч. посібник для вузів. Хмельницький: ТУП, 1996. 175 с.

6. Чегренець В. М., Руденко Н. В. Комп'ютер та комп'ютерна арифметика. Київ: Державний Університет Телекомунікацій, Навчально-науковий Інститут Телекомунікацій та Інформатизації, 2016. 120 с

7. Говорущенко Т. О., Гнатчук Є. Г. Особливості відмовостійких комп'ютерних систем з програмованою логікою як об'єктів діагностування. *Вісник Хмельницького національного університету*. № 5, 2010. С. 222–226.

8. Авраменко А. С., Авраменко В. С., Косенюк Г. В. Тестування програмного забезпечення. Навчальний посібник. Черкаси: ЧНУ імені Богдана Хмельницького, 2017. 284 с.

9. Тупкало В. М. Розробка моделей кіберстійких інформаційних систем управління на основі використання математичного апарату сигнатурної алгебри. V Міжнарод. наук.-практ. конф. «Управління якістю в освіті та промисловості: досвід, проблеми та перспективи», 20-21 травня 2021 р. зб. тез доп. Львів: ЛА «Піраміда», 2021. С. 186–187.

10. Тупкало В. М. Розробка моделей кіберстійких інформаційних систем управління на основі використання математичного апарату сигнатурної алгебри: авторське свідоцтво № 131020; заяв. 31.10.2024; опубл. 29.11.2024, Бюл. № 84.

11. Tupkalo V., Cherepkov S. Systems engineering of cybersecured digital and information measuring systems based on the signature Boolean-polynomial algebra synthesis apparatus. *Measurements infrastructure*. № 5. 2023. pp.1–14.

12. Tupkalo V., Cherepkov S. Signature functional control of digital automatic device register operations. *Measurements infrastructure*. № 7. 2024. pp.1–13.

REFERENCES:

1. Pryshchepa, M. M., Pogrebnyak, V. P. (2006). Mikroelektronika [Microelectronics]. In 3 parts. Part 2. Elements of microcircuitry: Textbook / M. M. Pryshchepa. Kyiv: Higher School.

2. Bondarenko, I. M., Borodin, O. V., Karnaushenko, V. P. (2020). Suchasna komponentna baza elektronnykh system [Modern component base of electronic systems]: textbook for students of higher educational institutions. Kharkiv: KhNURE.

3. Piddubny, V. O., Tovkach, I. O. (2021). Elementna baza radioelektronnoi aparatury: V 4 ch. Ch. 4. Osnovy mikroelektroniky [Elemental base of radioelectronic equipment: In 4 parts. Part 4. Fundamentals of microelectronics]: teaching aids for students of specialty 172 «Telecommunications and radio engineering». Kyiv: Igor Sikorsky Kyiv Polytechnic Institute.

4. Bondarenko, I. M., Glushko, A. P., Menkov, O. M. (2003). Kody ta koduvannia [Codes and coding]. Textbook. Kharkiv. XI Air Force.

5. Lokaziuk, V. M. (1996). Kontrol i diahnostuvannia obchysliuvalnykh prystroiv ta system: Navch. posibnyk dlia vuziv [Control and diagnostics of computing devices and systems: Textbook for universities]. Khmelnytsky. 175 p.

6. Chegrenets, V. M., Rudenko, N. V. (2016). Kompiuter ta kompiuterna aryfmetyka [Computer and computer arithmetic]. Kyiv: State University of Telecommunications, Educational and Scientific Institute of Telecommunications and Informatization. 120 p.

7. Govorushchenko, T. O., Gnatchuk, E. G. (2010). Osoblyvosti vidmovostiikykh kompiuternykh system z prohramovanoiu lohikoiu yak obiektiv diahnostuvannia [Features of fault-tolerant computer systems with programmable logic as objects of diagnostics]. *Bulletin of Khmelnytsky National University*. № 5, P. 222–226

8. Avramenko, A. S., Avramenko, V. S., Kosenyuk, G. V. (2017). Testuvannia prohramnoho zabezpechennia [Software testing.] Textbook. Cherkasy: Bohdan Khmelnytskyi National University. 284 p.

9. Tupkalo, V. M. (2021). Rozrobka modelei kiberstiikykh informatsiinykh system upravlinnia na osnovi vykorystannia matematychnoho aparatu syhnaturnoi alhebry [Development of models of cyber-resistant information management systems based on the use of the mathematical apparatus of signature algebra]. V International Scientific and Practical Conference «Quality Management in Education and Industry: Experience, Problems and Prospects». P. 186–187.

10. Tupkalo, V. M. (2024). Rozrobka modelei kiberstiikykh informatsiinykh system upravlinnia na osnovi vykorystannia matematychnoho aparatu syhnaturnoi alhebry [Development of models of cyber-resistant information management systems based on the use of the mathematical apparatus of signature algebra]: author's certificate No. 131020; application. 10/31/2024; publ. 11/29/2024, Bull. No. 84.

11. Tupkalo, V., Cherepkov, S. (2023). Systems engineering of cybersecured digital and information measuring systems based on the signature Boolean-polynomial algebra synthesis apparatus. *Measurements infrastructure*. № 5. pp.1–14.

12. Tupkalo, V., Cherepkov, S. (2024). Signature functional control of digital automatic device register operations. *Measurements infrastructure*. № 7. pp.1–13.

Додаток А. Сигнатурне поле TSF[7, $P^3(x)$], $P^3(x) = x^3 + x + 1$
(Tupkalo, Cherepkov, 2023**)**

Addition A. Signature field TSF[7, $P^3(x)$], $P^3(x) = x^3 + x + 1$
(Tupkalo, Cherepkov, 2023**)**

№	A	sigA	№	A	sigA	№	A	sigA	№	A	sigA	№	A	sigA
1	0000000	000	31	0011110	111	61	0111100	110	91	1011010	100	121	1111000	101
2	0000001	001	32	0011111	110	62	0111101	111	92	1011011	101	122	1111001	100
3	0000010	011	33	0100000	010	63	0111110	101	93	1011100	000	123	1111010	110
4	0000011	010	34	0100001	011	64	0111111	100	94	1011101	001	124	1111011	111
5	0000100	111	35	0100010	001	65	1000000	100	95	1011110	011	125	1111100	010
6	0000101	110	36	0100011	000	66	1000001	101	96	1011111	010	126	1111101	011
7	0000110	100	37	0100100	101	67	1000010	111	97	1100000	110	127	1111110	001
8	0000111	101	38	0100101	100	68	1000011	110	98	1100001	111	128	1111111	000
9	0001000	110	39	0100110	110	69	1000100	011	99	1100010	101			
10	0001001	111	40	0100111	111	70	1000101	010	100	1100011	100			
11	0001010	101	41	0101000	100	71	1000110	000	101	1100100	001			
12	0001011	100	42	0101001	101	72	1000111	001	102	1100101	000			
13	0001100	001	43	0101010	111	73	1001000	010	103	1100110	010			
14	0001101	000	44	0101011	110	74	1001001	011	104	1100111	011			
15	0001110	010	45	0101100	011	75	1001010	001	105	1101000	000			
16	0001111	011	46	0101101	010	76	1001011	000	106	1101001	001			
17	0010000	101	47	0101110	000	77	1001100	101	107	1101010	011			
18	0010001	100	48	0101111	001	78	1001101	100	108	1101011	010			
19	0010010	110	49	0110000	111	79	1001110	110	109	1101100	111			
20	0010011	111	50	0110001	110	80	1001111	111	110	1101101	110			
21	0010100	010	51	0110010	100	81	1010000	010	111	1101110	100			
22	0010101	011	52	0110011	101	82	1010001	000	112	1101111	101			
23	0010110	001	53	0110100	000	83	1010010	010	113	1110000	011			
24	0010111	000	54	0110101	001	84	1010011	011	114	1110001	010			
25	0011000	011	55	0110110	011	85	1010100	110	115	1110010	000			
26	0011001	010	56	0110111	010	86	1010101	111	116	1110011	001			
27	0011010	000	57	0111000	001	87	1010110	101	117	1110100	100			
28	0011011	001	58	0111001	000	88	1010111	100	118	1110101	101			
29	0011100	100	59	0111010	010	89	1011000	111	119	1110110	111			
30	0011101	101	60	0111011	011	90	1011001	110	120	1110111	110			

Ідеал поля TSF(7,3) при $P(x)=x^3+x+1$		
1	0000000	000
14	0001101	000
24	0010111	000
27	0011010	000
36	0100011	000
47	0101110	000
53	0110100	000
58	0111001	000
71	1000110	000
76	1001011	000
82	1010001	000
93	1011100	000
102	1100101	000
105	1101000	000
115	1110010	000
128	1111111	000
модель Тупкало В.М.		

УДК 004.056/331.4
DOI https://doi.org/10.32782/IT/2024-4-25

Любов ЧЕРЕМІСІНА
викладач кафедри кібербезпеки, інформаційних технологій та економіки, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе 210, м. Київ, Україна, 02121
ORCID: *0009-0005-0719-0745*

Наталія МАКСИМОВИЧ
старший викладач кафедри філософії та психології, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе 210, м. Київ, Україна, 02121
ORCID: *0000-0002-7405-9801*

Бібліографічний опис статті: Черемісіна, Л., Максимович, Н. (2024). Управління системою охорони праці та безпекою життєдіяльності персоналу в умовах виробничих компаній, що пов'язані з захистом даних. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 216–223, doi: https://doi.org/10.32782/IT/2024-4-25

УПРАВЛІННЯ СИСТЕМОЮ ОХОРОНИ ПРАЦІ ТА БЕЗПЕКОЮ ЖИТТЄДІЯЛЬНОСТІ ПЕРСОНАЛУ В УМОВАХ ВИРОБНИЧИХ КОМПАНІЙ, ЩО ПОВ'ЯЗАНІ З ЗАХИСТОМ ДАНИХ

У виробничих компаніях, які працюють з обробкою та зберіганням великих обсягів чутливої інформації, питання охорони праці та безпеки життєдіяльності персоналу набувають особливої актуальності.

Мета роботи. *Обґрунтування умов формування вимог до системи управління охороною праці та безпекою життєдіяльності персоналу в виробничих компаніях, що займаються захистом даних. Ця система є частиною загальної системи управління підприємства, спрямованою на ідентифікацію небезпек, оцінку та управління ризиками в галузі безпеки та гігієни праці, пов'язаними з діяльністю щодо захисту інформації.*

Методологія дослідження. *Був використаний регресійний аналіз. Методи статистичної обробки інформації дають змогу побудувати математичну модель планування заходів щодо розподілу ресурсів на потреби охорони праці та безпеки життєдіяльності в галузі захисту даних. За побудованою моделлю можна не тільки розрахувати значення показника, але й виявити внесок кожного із факторів.*

Наукова новизна. *Запропонована система, побудована на основі дворівневої інформаційно-вимірювальної системи, може бути застосована на підприємствах з різноманітними технологічними процесами та територіально розподіленою виробничою структурою. Також проаналізована математична модель для ефективного застосування в оптимізації розподілу ресурсів на потреби охорони праці, формуванні плану заходів на майбутнє та оцінюванні значущості статистичної інформації, що використовується для прийняття управлінських рішень. Комплекс математичних моделей, що описують різні показники та процеси управління охороною праці, є частиною системи підтримки прийняття рішень.*

Висновки. *Автоматизація системою управління охорони праці та безпекою життєдіяльності персоналу в виробничих компаніях, що займаються захистом даних дозволить підвищити ефективність забезпечення здоров'я та безпеки персоналу, забезпечення дотримання норм охорони праці, а також розширити увагу до екологічних питань та своєчасно адаптуватися до змін у технологічних процесах.*

Ключові слова: *охорона праці, безпека життєдіяльності, кібербезпека, захист даних, виробниче здоров'я, система управління.*

Liubov CHEREMISINA
Lecturer at the Department of Cybersecurity, Information Technology and Economics, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121, cheremisina1112@gmail.com
ORCID: *0009-0005-0719-0745*

Natalia MAKSYMOVYCH
Senior Lecturer at the Department of Philosophy and Psychology, Kyiv University of Intellectual Property and Law of the National university of Odessa law academy, 210, Kharkivske highway, Kyiv, Ukraine, 02121, maksimovich@gmail.com
ORCID: *0000-0002-7405-9801*

To cite this article: Cheremisina, L., Maksymovych, N. (2024). Upravlinnya systemoyu okhorony pratsi ta bezpekoyu zhyttiediyal'nosti personalu v umovakh vyrobnychykh kompaniy, shcho pov'yazani z zakhystom danykh [Management of the occupational health and safety system and personnel life safety in production companies related to data protection]. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 216–223, doi: https://doi.org/10.32782/IT/2024-4-25

MANAGEMENT OF THE OCCUPATIONAL HEALTH AND SAFETY SYSTEM AND PERSONNEL LIFE SAFETY IN PRODUCTION COMPANIES RELATED TO DATA PROTECTION

In manufacturing companies that work with the processing and storage of large amounts of sensitive information, the issues of labor protection and life safety of personnel become especially relevant.

Aim of the study. *Substantiation of the conditions for the formation of requirements for the occupational health and safety management system for personnel in manufacturing companies engaged in data protection. This system is part of the overall management system of the enterprise, aimed at identifying hazards, assessing and managing occupational safety and health risks related to information protection activities.*

Research methodology. *Regression analysis was used. Methods of statistical information processing make it possible to build a mathematical model for planning measures for the allocation of resources for the needs of labor protection and life safety in the field of data protection. According to the constructed model, it is possible not only to calculate the value of the indicator, but also to identify the contribution of each of the factors.*

Scientific novelty. *The proposed system, built on the basis of a two-level information and measurement system, can be applied at enterprises with a variety of technological processes and a geographically distributed production structure. A mathematical model for effective application in optimizing the allocation of resources for labor protection needs, forming an action plan for the future and assessing the significance of statistical information used for managerial decision-making is also analyzed. A set of mathematical models describing various indicators and processes of occupational health and safety management is part of the decision support system.*

Conclusions. *Automation of the occupational health and safety management system of personnel in manufacturing companies engaged in data protection will increase the efficiency of ensuring the health and safety of personnel, ensuring compliance with labor protection standards, as well as expanding attention to environmental issues and timely adapting to changes in technological processes.*

Key words: *occupational health and safety, life safety, cybersecurity, data protection, occupational health, management system.*

У сучасному світі інформаційні технології займають важливе місце в діяльності кожної компанії і захист даних стає ключовим елементом забезпечення безпеки. Важливою складовою цього є забезпечення фізичної та інформаційної безпеки, а також створення ефективних систем управління для запобігання можливим ризикам.

Система охорони праці в компаніях, що пов'язані з захистом даних, має враховувати не лише фізичні аспекти безпеки, але й аспекти, пов'язані з використанням технічних засобів для захисту інформації. Це включає в себе наявність спеціалізованих систем для захисту від несанкціонованого доступу, захист від вірусів і шкідливих програм, а також заходи, що сприяють збереженню конфіденційності і цілісності даних.

Головна мета управління охороною праці та безпекою життєдіяльності персоналу – забезпечення безпеки, зберігання здоров'я і високої працездатності людини у процесі праці. Виконання поставленої задачі на практиці вимагає удосконалення методів контролю, аналізу, моделювання і прогнозування стану охорони праці. Традиційні форми управління умовами праці на сучасному етапі виявляються недостатніми, оскільки при плануванні заходів щодо розподілу ресурсів на потреби охорони праці в галузі кібербезпеки та захисту даних не впроваджуються спеціальні алгоритми моделювання, які дозволяють оцінити наслідки виконання цих заходів. Тому на практиці управління переважають емпіризм і приватні рішення, навіть у компаніях які спеціалізуються на захисті даних. Оцінити наслідки прийнятих управлінських рішень і на цій основі раціонально сформувати план заходів можна за допомогою спеціальних математичних моделей.

Математична модель – це система співвідношень, яка визначає характеристики станів об'єкта управління, а через них і керуючі впливи, при яких досягається оптимальне управління (Гулевець, 2016). Основною використовуваною інформацією для управління охороною праці підприємства є: інформація про виробничий травматизм, форми звітності, витрати засобів на потреби охорони праці, виконання заходів з охорони праці та інша статистична інформація. Якісне управління охороною праці

вимагає точності вихідних даних. Точність статистичної інформації – це, насамперед, рівень відповідності документально Зареєстрованого значення розглядуваного фактора дійсному його значенню. Відхилення значень зареєстрованих документально факторів від дійсних їхніх значень є помилками статистичного спостереження. Точність статистичних матеріалів забезпечується детальною розробкою бланків і інструкцій до їх заповнення та чіткою організацією роботи і систематичним контролем матеріалу. Для максимально можливого зменшення числа випадків появи помилок необхідно розрізняти помилки різних видів і причини їх виникнення.

Сьогодні виробничі компанії прагнуть, з одного боку, зменшити витрати, пов'язані з охороною здоров'я та безпекою праці, з іншого боку − підвищити безпеку виробництва, ефективно керуючи пов'язаними з ним ризиками для людини, та одночасно поліпшити корпоративний імідж. З цією метою, підприємства всього світу впроваджують у себе системи управління виробничим здоров'ям і безпекою (СУ ВЗБ), орієнтуючись на вимоги міжнародного стандарту OHSAS 18000 «Система менеджменту здоров'я та безпеки персоналу» (Гулевець, Ревук, Кружило, 2000).

Аналіз закордонного досвіду в галузі менеджменту виробничого здоров'я та безпеки свідчить, що регулювання його необхідно здійснювати на підставі оцінки ризиків виникнення нещасних випадків та аварій на виробництві.

Національний стандарт ДСТУ ISO 45001:2019 (ISO 45001:2018, IDT) «Системи управління охороною здоров'я та безпекою праці. Вимоги та настанови щодо застосування», прийнятий методом перекладу, – ідентичний щодо ISO 45001:2018 (ISO 45001:2010, IDT) (версія en) «Occupational health and safety management systems – Requirements with guidance for use».

Відповідальне за цей стандарт в Україні – Державне підприємство «Український науково-дослідний і навчальний центр проблем стандартизації, сертифікації та якості» (ДП «УкрНДНЦ»), яке виконує функції НОС.

Цей стандарт прийнято на заміну ДСТУ OHSAS 18001:2010 (OHSAS 18001:2007, IDT) «Системи управління гігієною та безпекою праці. Вимоги», прийнятого методом перекладу.

СУ ВЗБ – це частина загальної системи управління виробничої компанії, яка спрямована на ідентифікацію небезпек, оцінку та управління ризиками в сфері безпеки та гігієни праці, що пов'язані з діяльністю організації. Система охоплює політику та цілі в галузі виробничого здоров'я та безпеки, організаційну структуру, процедури та ресурси для розробки, впровадження, досягнення, аналізу та підтримки в робочому стані політики організації в сфері безпеки та гігієни праці. Запроваджуючи СУ ВЗБ, кожна організація має визначити законодавчі та інші нормативні вимоги стосовно безпеки та гігієни праці, які вона зобов'язана виконати, та які стосуються її видів діяльності, продукції та послуг, обладнання та приміщень.

Відповідність організації законодавчим і нормативним актам країни є одним з найважливіших критеріїв аудиту під час сертифікації СУ ВЗБ. На кожному більш-менш великому підприємстві яке спеціалізується на захисті інформації існує, як мінімум, два рівня управління − на рівні виробничої компанії, як об'єкту управління в цілому, та на рівні окремого структурного підрозділу. Органом управління на першому рівні є роботодавець. Координація дій здійснюється через службу забезпечення виробничим здоров'ям і безпекою (СЗВЗБ). Органом управління на другому рівні є керівник структурного підрозділу. У оперативному вирішенні питань забезпечення виробничим здоров'ям і безпекою та здійсненні контролю роль органу управління здійснюють керівники дільниць, майстри, виконроби, енергетики, механіки.

Підприємства, які спеціалізуються на захисті інформації, можуть бути дуже різними за розміром та напрямками діяльності. Ось кілька типів таких підприємств:

1. IT-компанії з кібербезпеки – компанії надають послуги з кіберзахисту, таких як розробка програмного забезпечення для захисту від вірусів, зловмисних програм, мережевих атак та інших загроз. Приклад: Palo Alto Networks, Fortinet, Check Point Software Technologies.

2. Консалтингові компанії з кібербезпеки – спеціалізуються на наданні консультаційних послуг у сфері захисту даних, кіберзахисту, а також розробці політик безпеки та навчанні персоналу. Приклад: Deloitte, KPMG, PwC (завдяки їхньому кібербезпековому консалтингу).

3. Компанії з захисту персональних даних (Data Protection) – це підприємства допомагають організаціям забезпечити відповідність вимогам щодо захисту персональних даних, таким як GDPR. Приклад: Vormetric, Varonis, OneTrust.

4. Постачальники антивірусного програмного забезпечення – розробляють програмне забезпечення для захисту від вірусів, шкідливих програм, хакерських атак та інших кіберзагроз. Приклад: McAfee, Symantec (Norton), Kaspersky Lab.

5. Постачальники рішень для безпечного

зберігання даних- компанії, що займаються розробкою технологій для збереження, шифрування та захисту даних на серверах і в хмарних системах. Приклад: Box, Dropbox, Cloudflare.

6. Фірми з управління інцидентами кібербезпеки – компанії, що спеціалізуються на виявленні, реагуванні та вирішенні інцидентів, що пов'язані з кіберзагрозами. Приклад: FireEye, CrowdStrike.

Управлінська інформація складається з даних ідентифікації небезпек для кожного структурного підрозділу, визначених ступенів ризику особливо небезпечних та шкідливих технологічних процесів, стану виконання планів робіт, наказів, розпоряджень, приписів. Управлінські рішення доводяться до об'єкту управління у вигляді наказів, розпоряджень, приписів, планів робіт тощо. Сучасні проблеми забезпечення виробничого здоров'я та безпеки зачіпають різні аспекти діяльності всіх колективів підприємств і окремих фахівців. Складність цих проблем полягає в тому, що рішення їх має забезпечуватися в кожній виробничій ланці та на кожному робочому місці, на всіх етапах розвитку виробничих процесів, на основі аналізу технології, системного та комплексного методів.

В результаті аналізу ситуацій, що склалися на об'єктах виробничого здоров'я та безпеки (ОВЗБ), формуються висновки щодо стабілізації умов праці. У керуючому органі (КО) на підставі цих висновків виробляються команди, які приводять в дію виконавчі органи (ВО). Останні впливають на умови праці, з метою приведення їх у відповідність з нормативними вимогами. Так само як і ланка аналізу керуючий і виконавчий органи в існуючих СУ ВЗБ представляють собою певні інженерно-технічні служби. Пропонована система управління легко може бути автоматизована при використанні сучасної обчислювальної техніки. Автоматизація системи дозволить підвищити ефективність забезпечення здоров'я та безпеки, розширити питання екології, своєчасно адаптуватися до різних змін технологічних процесів.

Автоматизована СУ ВЗБ синтезована на базі дворівневої інформаційно вимірювальні системи (ІВС), може бути застосована для підприємств з різнорідними технологічними процесами та територіально-роз'єднаною виробничою структурою. ОВЗБ в цьому випадку обладнуються контрольно – вимірювальною апаратурою (КВА), яка має галузеві особливості та включає в собі діагностичні комплекси для об'єктивної оцінки стану працівників. Пульт керування (ПК) виконує збір інформації про стан умов на робочих місцях, перетворення цієї інформації для зберігання та передачі на вищий рівень ІВС, а також виробляє ряд логічних операцій в порівнянні виміряних значень з допустимими нормами та в необхідних випадках видає інформацію на пристрій відображення.

На підставі об'єктивних інформаційних даних керівники приймають рішення про зміну умов праці за допомогою деякого ВО. При досить глибокій автоматизації ВО може управлятися сигналами, що виробляються на верхньому рівні ІВС. Інформація з ПК надходить в центральний пристрій ІВС, який за допомогою електронно – обчислювальної машини (ЕОМ) проводить обробку даних відповідно до вирішуваних завдань за приватними алгоритмами або стандартними підпрограмами. При цьому центральний пристрій (ЦП) ІВС користується нормативними матеріалами, даними щодо умов праці в попередній період, характеристиками технологічних процесів, які зберігаються в пристрої довготривалої пам'яті. Крім того, в ЦП ІВС надходить зовнішня інформація, що міститься в законах і указах, положеннях, постановах уряду, рішеннях профспілкових органів, галузевих наказах і інструкціях. Зовнішня інформація включає в себе також і відомості про заходи щодо забезпечення безпечних методів праці, методики атестації робочих місць, пропаганди безпечних способів організації трудових процесів, планування заходів з гігієни та безпеки праці, досягнень та передовий досвід управління охороною здоров'я та безпекою. Результати обробки інформації надходять на пристрій відображення інформації (ПВІ) СЗВЗБ, яка на основі цих результатів розробляє пропозиції для керівництва підприємства щодо змін на ОВЗБ з метою поліпшення умов праці. Зв'язок від ланки управління підприємством до ВО може реалізуватися або через ЦП ІВС (в тому випадку, коли необхідна висока оперативність), або через службові (виробничі) канали та керівників служб, або через канали зв'язку з ПВІ ЦП ІВС та керівників служб.

Як показує практика, статистичні дані з охорони праці в ряді випадків містять помилки і неточності внаслідок об'єктивних і суб'єктивних причин. Помилки через неуважність реєстратора в значній мірі можуть бути усунуті при використанні засобів автоматизації і спеціалізованого програмного забезпечення для впровадження і збереження інформації. У цьому випадку виключаються неточності через нерозбірливий почерк, помилки форматів і кодів введених даних, а отже, й помилки, пов'язані з виконанням математичних розрахунків при формуванні документів. Проте існує

ряд помилок (у більшості випадків допущених свідомо), які спотворюють реальні дані. Стандартні засоби логічного контролю ці помилки виявити не можуть.

Існує ряд підходів до визначення можливості події НВ у конкретних виробничих умовах. Зокрема, прогнозування кількості НВ і матеріальних, наслідків від РІВ може бути виконане за допомогою теоретико-ймовірного методу (Ткачук, Ткачук, Ревук, Кружилко, 1996). На жаль, цей метод не дозволяє оцінити ступінь впливу факторів на досліджуваний показник.

Методи статистичної обробки інформації дають змогу побудувати математичну модель досліджуваного показника. За побудованою моделлю можна не тільки розрахувати значення показника, але й виявити внесок кожного із факторів. Для вирішення поставленої задачі був використаний регресійний: аналіз. Як досліджувані показники прийнято число НВ, число непрацездатних днів, розмір матеріальних збитків у результаті НВ. Для побудови залежності між показниками і факторами, які характеризують умови праці, необхідно мати базу даних (БД) про НВ. Розглянемо процедуру побудови моделі на прикладі показника – числа днів непрацездатності К На підставі побудованої моделі визначимо ступінь впливу кожного з реєстрованих факторів й оцінимо точність вихідних даних.

Визначимо залежність Y = F(X), де X – множина факторів, яка характеризує НВ. Будь-яка функціональна залежність може бути подана у вигляді полінома Колмогорова-Габора (Гулевець, 2016):

$$Y=a_0+\Sigma\, a_i x_i+\Sigma\, \Sigma\, a_{ij}\, x_i\, x_j+...$$

x_i та x_j – незалежні перемінні; a_0 та a_i- постійні коефіцієнти.

Важливою складовою управління охороною праці є виявлення небезпек і оцінка ризиків, пов'язаних як з фізичними факторами, так і з технологічними аспектами захисту інформації. Для цього застосовуються методи ризик-менеджменту, аналізу безпеки інформаційних систем і розробка стандартів безпеки на всіх етапах виробничого процесу.

У компаніях, що працюють з інформаційними технологіями та захистом даних, система управління охороною праці повинна бути інтегрована з іншими системами управління, такими як система інформаційної безпеки та управління якістю. Вона повинна включати чітко визначені процедури для запобігання аваріям, навчання персоналу та реагування на надзвичайні ситуації.

Отже, загальний вигляд залежності буде поліномний. Ця властивість визначення залежностей знайшла широке застосування в моделюванні різних показників охорони праці.

Інформацію щодо СУ ВЗБ слід виділити в групи: – щодо стан гігієни та безпеки праці, – щодо наслідки незадовільного стану гігієни та безпеки праці, – щодо функціонування СУ ВЗБ.

Перша група інформації дає можливість оцінювати фактичний стан всього комплексу чинників, що впливають на безпеку праці. Інформація другої групи відображає відомості стосовно різних форм шкоди, завданої здоров'ю працюючих (травматизм, профзахворювання) через незадовільні умови гігієни та безпеки праці. Третя група інформації дає можливість судити стосовно працездатності СУ ВЗБ взагалі, її структурних ланок і органів.

Другий клас інформації – адресна керуюча, ієрархічна. Наповнення її здійснюється ВП і призначене для ВО при диференційованому забезпеченні, що передбачає адресність і необхідний ступінь деталізації. Щоб визначити яка інформація повинна надходити кожному керівнику служби, цеху, відділу потрібно суворо керуватися їх функціональними обов'язками. Інформаційні потоки організовуються нормативними актами підприємства:

–розпоряджувальний: накази, розпорядження, вказівки; -стверджувальними: акти розслідування нещасних випадків, комплексних перевірок контрольних комісій;

–положеннями, що встановлюють завдання, функції, права, обов'язки та відповідальність підрозділів, служб, відділів підприємства та їх керівників щодо забезпечення виробничого здоров'я та безпеки;

–посадовими інструкціями, які встановлюють права та функціональні обов'язки осіб адміністративного та інженерно-технічного персоналу; -поточними планами поліпшення гігієни та безпеки праці, санітарно – оздоровчих заходів;

–вказівками матеріально-технічного забезпечення та засобами індивідуального захисту.

Дії, що управляють (рішення) в напрямку від ВО до ВП. Її джерела:

–план роботи начальників служб, цехів, відділів; -наказ, вказівки керівника виробничої компанії;

–преміювання (депреміювання) та інші заходи притягнення до відповідальності; -припис інспекції;

–матеріали відомчих комісій.

Інформація факторна збурена – чинники навколишнього середовища та технологічних процесів виробництва, які формують умови праці

на робочих місцях кожного фахівця. З робочих зон нижнього рівня на верхньому рівні також приймається інформація наступного класу – звана ергономічною, що є інструментом реального відображення умов і результату трудового процесу. Інформаційними характеристиками є:

–якість і ефективність виконаної роботи;

–продуктивність праці;

–стан здоров'я;

–психофізіологічні показники – втома, помилки та їх ціна, пам'ять, реакція людини й т.і. З периферійних пристроїв факторна збурена та паралельно ергономічна інформація надходять в пристрій прийому ІРЦ ПЗБ. Перша порівнюється з нормативною базою терміналів безпеки праці (зберігання в базі даних).

При відхиленні отриманих показників від вимог зони комфортності та нормального самопочуття включаються звукова та світлова сигналізації. Одночасно відбувається обробка даних на ЕОМ, пов'язаної з ІРЦ ПЗБ лініями сполучення, або персональної ЕОМ служби ЗВЗБ. За результатами обробки отримуємо команду щодо прийняття рішення. Право прийняття тактичного рішення залишається за людиною (керуючим органом).

Для ефективного управління охороною праці в таких компаніях варто використовувати математичні моделі, які дозволяють оптимізувати розподіл ресурсів і визначити пріоритети для забезпечення безпеки. Це включає в себе розробку планів з урахуванням необхідних матеріальних і людських ресурсів для виконання заходів, спрямованих на забезпечення здоров'я та безпеки персоналу.

Апарат математичного моделювання може ефективно застосовуватися в області охорони праці не тільки для оптимізації розподілу ресурсів на потреби охорони праці й безпеки життєдіяльності, а й у формуваннях плану заходів на майбутній період, й для оцінки вірогідності й значимості інформації, використовуваної для прийняття управлінських рішень. Комплекс математичних моделей, які описують різноманітні показники і процеси управління охороною праці, являють собою систему підтримки прийняття рішень.

Автоматизація процесів охорони праці і безпеки життєдіяльності персоналу дозволяє підвищити ефективність цих систем. Вона включає в себе використання сучасних технологій для моніторингу стану безпеки, своєчасного виявлення інцидентів та автоматичне реагування на потенційні загрози. Це забезпечує швидке реагування на зміни в технологічному процесі або в умовах роботи.

Управління системою охорони праці та безпеки життєдіяльності персоналу є важливим аспектом діяльності виробничих компаній, що займаються захистом даних. Для забезпечення ефективної роботи таких компаній необхідно інтегрувати системи охорони праці з іншими аспектами управління, такими як кібербезпека, захист даних та управління ризиками. Використання математичних моделей, автоматизація процесів та постійне вдосконалення системи безпеки дозволяють мінімізувати ризики для здоров'я персоналу та забезпечити належний захист інформації.

ЛІТЕРАТУРА:

1. Гулевець В. Д. Система управління охороною здоров'я та безпеки персоналу в умовах виробничих компаній. *Проблеми охорони праці в Україні.* 2016. № 32, С. 84–93.

2. Гулевець В. Д., Ревук О. Г., Кружило О. Є. Алгоритм оцінки значимості і достовірності статистичної інформації про стан охорони праці підприємства. 2000. *Вісник НАУ.* № 1-2. С. 257–260.

3. Ткачук С. П., Ткачук К. Н., Ревук А. Г., Кружилко О. Є. Методологічні основи визначення критеріїв оцінки умов праці. *Охорона праці.* 1996. № 8. С. 34–39.

REFERENCES:

1. Ohronevets, V. D. (2016). Systema upravlinnia okhoronoi zdorovia ta bezpeky personalu v umovakh vyrobnychnykh kompanii [System of health and safety management and personnel safety in the conditions of manufacturing companies]. *Problemy okhorony pratsi v Ukraini.* 32, 84–93 [in Ukrainian].

2. Ohronevets, V. D., Revuk, O. H., Kruzhylo, O. Ye. (2000). Alhorytm otsinky znachymosti i dostovirnosti statystychnoyi informatsiyi pro stan okhorony pratsi pidpryiemstva [Algorithm for assessing the significance and reliability of statistical information about the state of labor protection at an enterprise]. *Visnyk NAU,* № 1-2, 257–260 [in Ukrainian].

3. Tkachuk, S. P., Tkachuk, K. N., Revuk, A. H., Kruzhylko, O. Ye. (1996). Metodolohichni osnovy vyznachennia kryteriyiv otsinky umov pratsi [Methodological foundations for determining the criteria for assessing working conditions]. *Okhorona pratsi,* № 8, 34–39 [in Ukrainian].

УДК 159.9:005.95/.96:004.056
DOI https://doi.org/10.32782/IT/2024-4-26

Олеся ЧЕРНЯКОВА
кандидат психологічних наук, доцент, доцент кафедри філософії та психології, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе, 210, Київ, Україна, 02121
ORCID: *0000-0002-0384-4829*

Бібліографічний опис статті: Чернякова, О. (2024). Психотехнології вирішення конфліктів розробників систем захисту інформації в умовах кіберзагроз. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 4, 222–226, doi: https://doi.org/10.32782/IT/2024-4-26

ПСИХОТЕХНОЛОГІЇ ВИРІШЕННЯ КОНФЛІКТІВ РОЗРОБНИКІВ СИСТЕМ ЗАХИСТУ ІНФОРМАЦІЇ В УМОВАХ КІБЕРЗАГРОЗ

Мінливі умови сучасності, спричинені цифрою трансформацією та інтенсифікацією інформаційного середовища значно підвищують складність роботи фахівців з інформаційної безпеки, що, у свою чергу, провокує виникнення міжособистісних і професійних конфліктів. Високий рівень стресу, необхідність оперативного реагування на загрози та багатозадачність створюють сприятливе підґрунтя для конфліктних ситуацій у командах, які розробляють та впроваджують системи захисту інформації.

Мета. *Важливим викликом для сучасної науки є визначення ефективних психотехнологій для діагностики, попередження та вирішення конфліктів серед розробників систем інформаційної безпеки, з урахуванням специфіки їхньої професійної діяльності. Основна увага, на нашу думку, має бути приділена аналізу впливу когнітивних, емоційних і соціальних чинників на процеси комунікації та співпраці в команді.*

Стаття присвячена теоретичному аналізу психотехнологій вирішення конфліктів серед фахівців з інформаційної безпеки в умовах зростаючих кіберзагроз. В роботі розглянуто сутність ключових понять: «кіберзагроза», «конфлікт», а також визначено специфіку виникнення та протікання конфліктів у професійних колективах розробників систем захисту інформації. Особливий акцент зроблено на впливі стресогенних чинників, спричинених кіберзагрозами, на психологічний клімат у командах, що займаються захистом інформації.

Методологічною основою *дослідження є міждисциплінарний підхід, який поєднує положення психології конфлікту, кібербезпеки та організаційної психології. Теоретичний аналіз дозволив виявити основні джерела конфліктів серед фахівців у сфері інформаційної безпеки, зокрема конкуренцію за ресурси, напруження через високий рівень відповідальності та ризиків, а також проблеми міжособистісної взаємодії в умовах інтенсивної роботи.*

На основі проведеного аналізу розроблено низку практичних рекомендацій, спрямованих на профілактику конфліктів. Зокрема, запропоновано методики підвищення стресостійкості фахівців, розвиток навичок конструктивної комунікації, впровадження програм тімбілдінгу та створення сприятливих умов для ефективної роботи команд. Особлива увага приділена використанню психотехнологій, що сприяють своєчасному розпізнаванню та вирішенню конфліктів, запобіганню їхньому загостренню та зниженню негативного впливу на професійну діяльність.

Результати *дослідження можуть бути використані як основа для подальших прикладних розробок у сфері організації діяльності команд фахівців з інформаційної безпеки, а також для розробки навчальних програм, спрямованих на формування стресостійкості та підвищення ефективності командної взаємодії в умовах сучасних кіберзагроз.*

Ключові слова: *психотехнології, вирішення конфліктів, інформаційна безпека, кіберзагрози, розробники систем захисту, емоційний інтелект, стрес-менеджмент.*

Olesia CHERNIAKOVA
PhD in Psychology, Associate Professor, Associate Professor at the Department of Philosophy and Psychology, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121, cherniakovaolesia@gmail.com
ORCID: *0000-0002-0384-4829*

To cite this article: Cherniakova, O. (2024). Psykhotekhnolohii vyrishennia konfliktiv rozrobnykiv system zakhystu informatsii v umovakh kiberzahroz [Psychotechnologies for resolving conflicts between developers of information protection systems in the face of cyber threats]. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 4, 222–226, doi: https://doi.org/10.32782/IT/2024-4-26

PSYCHOTECHNOLOGIES FOR CONFLICT RESOLUTION OF DEVELOPERS OF INFORMATION PROTECTION SYSTEMS IN THE CONDITIONS OF CYBER THREATS

The changing conditions of modernity, caused by digital transformation and intensification of the information environment, significantly increase the complexity of the work of information security specialists, which, in turn, provokes the emergence of interpersonal and professional conflicts. High levels of stress, the need for prompt response to threats, and multitasking create a favorable environment for conflict situations in teams that develop and implement information protection systems.

An important challenge for modern science is to identify effective psychotechnologies for diagnosing, preventing, and resolving conflicts among information security system developers, taking into account the specifics of their professional activities. In our opinion, the main attention should be paid to the analysis of the influence of cognitive, emotional, and social factors on the processes of communication and cooperation in the team.

The article is devoted to the theoretical analysis of psychotechnologies for resolving conflicts among information security specialists in the face of growing cyber threats. The paper examines the essence of the key concepts: «cyber threat», «conflict», and also identifies the specifics of the emergence and course of conflicts in professional teams of information protection system developers. Special emphasis is placed on the influence of stress-producing factors caused by cyber threats on the psychological climate in teams involved in information protection.

***The methodological** basis of the study is an interdisciplinary approach that combines the principles of conflict psychology, cybersecurity and organizational psychology. Theoretical analysis allowed us to identify the main sources of conflicts among specialists in the field of information security, in particular, competition for resources, tension due to a high level of responsibility and risks, as well as problems of interpersonal interaction in conditions of intensive work.*

Based on the analysis, a number of practical recommendations aimed at conflict prevention were developed. In particular, methods were proposed to increase the stress resistance of specialists, develop constructive communication skills, implement team building programs and create favorable conditions for effective team work. Particular attention is paid to the use of psychotechnologies that contribute to the timely recognition and resolution of conflicts, preventing their escalation and reducing the negative impact on professional activity.

***The results** of the study can be used as a basis for further applied developments in the field of organizing the activities of teams of information security specialists, as well as for the development of training programs aimed at developing stress resistance and increasing the effectiveness of team interaction in the face of modern cyber threats.*

***Key words:** psychotechnologies, conflict resolution, information security, cyber threats, developers of protection systems, emotional intelligence, stress management.*

Актуальність проблеми вирішення конфліктів серед розробників систем захисту інформації в умовах кіберзагроз обумовлена зростаючою роллю інформаційної безпеки в сучасному цифровому світі. В умовах швидкого розвитку технологій кіберзагроз зростають вимоги до кваліфікації та ефективності роботи фахівців, що призводить до підвищення рівня стресу, напруження та професійних конфліктів у команді розробників. Це негативно впливає на продуктивність і безпеку інформаційних систем. Розробка психотехнологій для ефективного вирішення таких конфліктів є необхідною умовою для підтримання стабільності в командах і забезпечення належного рівня захисту інформації в умовах постійних загроз.

Аналіз останніх досліджень в рамках окресленої проблематики вказує на значний інтерес до цього питання як в Україні, так і за її межами. Вітчизняні науковці, зокрема, зосереджуються на дослідженні психологічних аспектів взаємодії в команді, розвитку емоційного інтелекту та механізмів управління стресом у високотехнологічних професіях. Відзначено важливість впливу організаційних факторів на виникнення конфліктів, а також розглядаються методи попередження та вирішення конфліктних ситуацій за допомогою психотехнологій, медіації та тренінгів.

Зарубіжні дослідження, зокрема в Європі та США, акцентують увагу на інтеграції психотерапевтичних і управлінських підходів для оптимізації міжособистісних відносин в умовах інтенсивних кіберзагроз. Зокрема, розглядаються специфічні методи роботи з командами високої напруги, зокрема в IT-секторі, де конфлікти можуть мати критичні наслідки для безпеки систем.

Перспективи подальших досліджень, на нашу думку, мають бути зосереджені на глибшому вивченні психосоціальних аспектів професійних конфліктів у кібербезпеці, зокрема на визначенні ефективних методів профілактики та нейтралізації стресу в умовах високого навантаження. Недостатньо вивченими залишаються питання автоматизації процесів вирішення конфліктів за допомогою новітніх технологій, таких як штучний інтелект або алгоритми для моніторингу емоційного стану працівників. Складнощі досліджень зумовлені складністю інтерпретації психосоціальних факторів, які впливають на виникнення та розвиток конфліктів в умовах

швидких змін у кіберпросторі та новітніх загроз.

Метою нашого дослідження є систематизація та формулювання практичних рекомендацій щодо комплексного підходу до профілактики, запобіганню та вирішенню конфліктів розробників інформаційних систем в умовах кіберзагроз.

В рамках дослідження шляхів вирішення конфліктів розробників систем захисту інформації в умовах кіберзагроз, в першу чергу, на нашу думку, варто визначити сутність та зміст поняття кібезагрози.

Кіберзагроза – це сукупність реальних або потенційних факторів і явищ, що становлять загрозу для інтересів особистості, суспільства та держави шляхом порушення доступності, цілісності, достовірності, автентичності та повноти інформації, яка обробляється чи передається в межах критично важливих об'єктів національної інформаційної інфраструктури (Сахав А. І., 2020, с. 267).

Наступним важливим етапом нашого дослідження є визначення психологічного змісту поняття конфліктів. Конфлікт є складною системою, що характеризується адаптивними структурами та еволюційними механізмами. Його глибоке вивчення потребує, з одного боку, застосування системного підходу, а з іншого – інтеграції знань із різних соціальних і наукових дисциплін. У зв'язку з цим у сучасній науці накопичено значний обсяг теоретичних і практичних розробок, спрямованих на дослідження цього явища. Перехід від індустріальної до інформаційної епохи зумовив виникнення нових типів конфліктів, відомих як інформаційні конфлікти (Шевченко С. М., 2022, с. 151).

Термін «конфлікт» походить від латинського слова, що означає «зіткнення». Різні наукові школи та течії пропонують власні теорії та концепції, які пояснюють природу конфлікту з різних точок зору, що ускладнює як його тлумачення, так і практичне використання в науковій діяльності. У науковій літературі нараховується велика кількість визначень цього поняття. Український дослідник Ю. Мацієвський виділяє три основні аспекти розуміння конфлікту: структурна несумісність інтересів, яка проявляється у взаємному виключенні групових цілей через обмежену кількість дефіцитних ресурсів; дії або взаємовідносини, спрямовані на завдання шкоди або знищення суперника; стан ворожості між окремими особами або групами (Жекало Г., 2015, с. 62).

Будь-який колектив формується з працівників, які істотно відрізняються один від одного за віком, рівнем освіти, стажем роботи, професійним і життєвим досвідом. Також спостерігаються відмінності у цілях і завданнях, які люди ставлять перед собою, та у способах їх досягнення. Важливо враховувати розбіжності у функціональних обов'язках, рольових позиціях і статусах. Крім того, значущими є особистісні особливості, різниця у ціннісних орієнтаціях, життєвих установках та підходах до вирішення проблем, які виникають у процесі роботи. Ці фактори створюють передумови для появи конфліктних ситуацій та виникнення конфліктів у колективах. Основною причиною розвитку конфліктів часто стає суперечність інтересів або прагнення відстояти свої позиції, що проявляється через конфліктну поведінку. У межах організації це може проявлятися у формі вилучення ресурсів, приховування важливої інформації, запровадження обмежень чи дискредитації. Такі дії, як правило, провокують негативні емоції, включно з гнівом, бажанням помсти та взаємною ворожістю. Ці емоції, у свою чергу, підсилюють конфліктну поведінку, що загострює суперечності інтересів. У результаті конфлікт має тенденцію до ескалації, перетворюючись на динамічний і швидкозростаючий процес (Щербакова І. М, Терещеня І. О., 2014, с. 186).

І. М. Щербакова та І. О. Терещеня у ході емпіричного дослідження дійшли висновку, що в трудових організаціях професійне вирішення конфліктів за участі психологів практично не здійснюється. Проте саме психолог має можливість і повинен виконувати функцію медіатора. У цій ролі він залишається нейтральним, не підтримуючи жодну зі сторін, що дозволяє об'єктивно та беземоційно вибудовувати конструктивний діалог між учасниками професійного конфлікту (Щербакова І. М, Терещеня І. О., 2014, с. 187). Пропонуємо більш детально зупинитися на тих наукових працях, які присвячені вивченню конфліктів саме в рамках інформаційної безпеки та кібербезпеки.

При вивченні стадій конфлікту в системах безпеки можна виокремити декілька ключових етапів, що характеризують розвиток конфліктної ситуації. Першим етапом є виникнення конфліктної ситуації, коли відбувається початкове загострення суперечностей. Другий етап–латентна стадія–характеризується прихованим існуванням конфлікту, що не демонструється явно. Третя стадія–активна–означає безпосередню реалізацію конфліктних дій, де виявляються відкриті зіткнення інтересів. Завершення конфлікту відбувається на останньому етапі, коли конфлікт досягає свого завершення або переходить на нову стадію. Глибина конфліктів в контексті інформаційної безпеки визначається складністю суперечностей, що є основою

конфлікту в інформаційній системі. Для оцінки глибини конфлікту можна виділити кілька рівнів. Початковий рівень характеризується виявленням одного інциденту, при якому відхилення оцінки множини ознак від еталонних значень не перевищує встановленої межі. На середньому рівні виявляються кілька інцидентів, що відбуваються одночасно, і відхилення оцінки ознак все ще знаходяться в межах допустимих значень. Високий рівень вказує на одночасне наявність декількох інцидентів, де відхилення від еталонних ознак перевищує певну межу, що свідчить про значну інтенсивність конфлікту в системі (Шевченко С. М., 2022, с. 156).

На нашу думку, вирішення проблеми конфліктів розробників систем захисту інформації в умовах кіберзагроз вимагає комплексного підходу. Зокрема, окремої уваги вимагає проблематика профілактики виникненню конфліктів, що включає розвиток емоційної компетентності, емоційного інтелекту, емпатії тощо серед студентів, що здобувають освіту на спеціальністю кібербезпека та інформаційні технології (Капітон А. М., 2022). В результаті аналізу наукової літератури, присвяченої окресленій проблематиці, нами було зроблено спробу розробити практичні рекомендації, спрямовані на профілактику та запобігання професійним конфліктам в контексті інформаційних технологій, отже:

1. Профілактика конфліктів: основним елементом профілактики є створення сприятливого психологічного клімату в команді розробників. Для цього важливо:

– Формування комунікаційних навичок: навчання фахівців ефективним технікам комунікації, включаючи активне слухання, відкрите обговорення проблем і питань без критики, що дозволяє знизити рівень непорозумінь та напруження.

– Чітке визначення ролей і відповідальності: упорядкування організаційної структури команди, визначення чітких функціональних обов'язків кожного члена колективу сприяє зменшенню територіальних конфліктів та дублювання обов'язків.

– Створення середовища для конструктивної критики: заохочення відкритого обговорення проблемних питань, де кожен розробник може висловити свою думку без страху бути осудженим. Це дозволяє знизити рівень конфліктів, пов'язаних з незадоволенням відсутністю можливості висловити свою позицію.

2. Запобігання конфліктам на етапах розробки системи захисту інформації полягає в:

– Забезпеченні інформаційної прозорості: створення відкритих каналів комунікації між різними підрозділами, що працюють над проектом, дозволяє своєчасно виявляти потенційні джерела напруження і конфліктних ситуацій.

– Планування та координація робіт: ретельне планування проекту, узгодження термінів та обсягів робіт між усіма учасниками забезпечує уникнення непорозумінь, пов'язаних з перевантаженням або, навпаки, недостатнім завантаженням працівників.

– Інтеграція системи управління стресом: враховуючи високий рівень стресу в умовах кіберзагроз, важливо застосовувати стратегії зниження останнього на організаційному рівні, такі як регулярні перерви, тренінги з управління стресом і підтримка добробуту працівників.

3. Вирішення конфліктів: у разі виникнення конфліктів серед розробників, необхідно використовувати такі психотехнології для їх вирішення:

– Психологічне втручання: проведення індивідуальних або групових консультацій із залученням психологів або спеціалістів з управління конфліктами, які допоможуть розкрити причини конфлікту та знайти конструктивне рішення.

– Медіація: призначення нейтральної особи для проведення медіації, яка допоможе сторонам конфлікту обговорити свої претензії і досягти взаємопогодженого рішення.

– Техніка «win-win»: пошук рішень, що задовольняють усі сторони конфлікту, враховуючи інтереси кожного учасника, з метою досягнення гармонії в роботі команди без втрат для результативності проекту.

4. Розвиток навичок емоційного інтелекту є важливим інструментом у профілактиці і вирішенні конфліктів. Це включає:

– Самоусвідомлення: розвиток здатності фахівців усвідомлювати власні емоції та їх вплив на робочі процеси, що допомагає знижувати рівень емоційних реакцій у напружених ситуаціях.

– Саморегуляція: навчання ефективним стратегіям самоконтролю, що дозволяє уникнути емоційних спалахів і вирішувати конфлікти в конструктивному руслі.

– Емпатія: розвиток здатності до емпатії допомагає зрозуміти точки зору колег, що сприяє взаєморозумінню і покращенню міжособистісних взаємин у команді.

5. Використання технологій для запобігання конфліктам: інтеграція технологічних рішень, таких як автоматизовані системи моніторингу роботи команди, платформи для спільної роботи та аналізу продуктивності, може значно знизити ймовірність виникнення конфліктів. Використання цих інструментів дозволяє

здійснювати контроль над процесами розробки та забезпечує своєчасне виявлення потенційних проблем.

Загалом, ефективне вирішення конфліктів серед розробників систем захисту інформації в умовах кіберзагроз вимагає поєднання організаційних заходів, психологічних підходів та використання інноваційних технологій для запобігання і вирішення проблем. Застосування таких методів дозволяє не лише покращити командну атмосферу, але й забезпечити безпеку інформаційних систем на високому рівні.

Перспективою подальшого дослідження є впровадження розроблених нами рекомендацій з метою визначення їх ефективності в контексті профілактики та запобіганню професійних конфліктів розробників інформаційних систем в умовах кіберзагроз.

ЛІТЕРАТУРА:

1. Жекало Г. Основні підходи до визначення поняття «конфлікт». *Східноукраїнський конфлікт: типологія, особливості, шляхи деескалації. Сер. Філософія*. 2015. № 2 (134). С. 62–65.

2. Сахав А. І. Поняття кіберзагроз на сучасному етапі. *Математичні методи, моделі та інформаційні технології у науці, освіті, економіці, виробництві* : матеріал доп. учасн. II Всеукраїнської науково-практичної Інтернет-конференції з проблем вищої освіти і науки, 29 квітня 2020 р. Маріуполь, 2020. С. 266–270.

3. Шевченко С. М., Складанний П. М., Негоденко О. В. Дослідження прикладних аспектів теорії конфліктів у системах безпеки. *Кібербезпека: освіта, наука, техніка*. 2022. № 2 (18). С. 150–162.

4. Щербакова І. М., Терещеня І. О. Психотехнології вирішення професійних конфліктів. *Наукові пошуки*. 2014. Вип. 10. С. 185–187.

5. Капітон А. М. Педагогічні умови формування професійної компетентності майбутніх фахівців з кібербезпеки. *Стратегічні пріоритети інформаційної безпеки держави у сфері оборони в умовах воєнного стану* : зб. матеріалів II Міжвід. наук.-практ. конф., 29 листоп. 2022 р. Київ: НУОУ, 2022. С. 115–116.

REFERENCES:

1. Zhekalo, H. (2015). Osnovni pidkhody do vyznachennia poniattia «konfdikt». *Skhidnoukrainskyi konflikt: typolohiia, osoblyvosti, shchliakhy deeskalatsii. Ser. Filosofiia*, 2 (134), 62–65 [in Ukrainian].

2. Sakhav, A. I. (2020). Poniattia kiberzahroz na suchasnomu etapi. *Matematychni metody, modeli ta informatsiini tekhnolohii u nautsi, osviti, ekonomitsi, vyrobnytstvi* : material dop. uchasn. II Vseukrainskoi naukovo-praktychnoi Internet-konferentsii z problem vyshchoi osvity i nauky, 29 kvitnia 2020 r. Mariupol, 266–270 [in Ukrainian].

3. Shevchenko, S. M., Skladannyi, P. M. & Nehodenko O. V. (2022). Doslidzhennia prykladnykh aspektiv teorii konfliktiv u systemakh bezpeky. *Kiberbezpeka: osvita, nauka, tekhnika*, 2 (18), 150–162 [in Ukrainian].

4. Shcherbakova, I. M., Tereshchenia, I. O. (2014). Psykhotekhnolohii vyrishennia profesiinykh konfliktiv. *Naukovi poshuky*, 10, 185–187 [in Ukrainian].

5. Kapiton, A. M. (2022). Pedahohichni umovy formuvannia profesiinoi kompetentnosti maibutnikh fakhivtsiv z kiberbezpeky. Stratehichni priorytety informatsiinoi bezpeky derzhavy u sferi oborony v umovakh voiennoho stanu : zb. materialiv II Mizhvid. nauk.-prakt. konf., 29 lystop. 2022 r. Kyiv: NUOU, 115–116 [in Ukrainian].

УДК 004.056:004.75
DOI https://doi.org/10.32782/IT/2024-4-27

Максим ШАБАН
кандидат технічних наук, викладач кафедри кібербезпеки, інформаційних технологій та економіки, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе, 210, м. Київ, Україна, 02121
ORCID: *0000-0003-2706-8235*

Бібліографічний опис статті: Шабан, М. (2024). Апробація апарату регулярних виразів при аналізі функціонального профілю захисту. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 227–229, doi: https://doi.org/10.32782/IT/2024-4-27

АПРОБАЦІЯ АПАРАТУ РЕГУЛЯРНИХ ВИРАЗІВ ПРИ АНАЛІЗІ ФУНКЦІОНАЛЬНОГО ПРОФІЛЮ ЗАХИСТУ

У статті проведено аналіз можливостей використання апарату регулярних виразів для побудови функціонального профілю захисту (ФПЗ) в інформаційних системах. Методологія полягає в застосуванні регулярних виразів, такі як побудова синтаксичних дерев, визначення часткових та внутрішніх патернів, а також розробка алгоритмів пошуку підстрок у тексті. Функціональний профіль захисту розглядається як строкова послідовність, де кожен символ відображає певний елемент захисту. За допомогою регулярних виразів можливо автоматизувати процес перевірки на повноту, узгодженість та ідентифікацію функцій безпеки в інформаційних системах.

Метою *роботи є демонстрація регулярних виразів, що є ефективним інструментом для аналізу ФПЗ, оскільки дозволяють стандартизувати та оптимізувати процес оцінки безпеки інформаційних систем. Було розглянуто практичні приклади використання синтаксичних дерев, які забезпечують ефективне виявлення префіксів, підстрок та інших строкових структур. Описано методологію пошуку часткових патернів у текстових даних, що є основою для автоматизованих систем прийняття рішень під час проведення державної експертизи інформаційних систем.*

Висновки. *Наукова новизна полягає у використанні отриманих результатів для створення програмного забезпечення, яке підтримує прийняття рішень щодо оцінки функціонального профілю захисту, що значно скорочує час аналізу та підвищує ефективність процесу. Робота підкреслює важливість використання регулярних виразів у сучасних методах забезпечення безпеки інформації.*

Ключові слова: *регулярні вирази, функціональний профіль захисту, синтаксичне дерево, часткові патерни, державна експертиза.*

Maksym SHABAN
Candidate of Technical Sciences, Lecturer at the Department of Cybersecurity, Information Technologies and Economics, Kyiv University of Intellectual Property and Law of the National University «Odesa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121, sabanmaxim@gmail.com
ORCID: *0000-0003-2706-8235*

To cite this article: Shaban, M. (2024). Aprobatsiia aparatu rehuliarnykh vyraziv pry analizi funktsionalnoho profiliu zakhystu [Testing of regular expression apparatus in analysis of functional profile of protection]. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 4, 227–229, doi: https://doi.org/10.32782/IT/2024-4-27

TESTING OF REGULAR EXPRESSION APPARATUS IN ANALYSIS OF FUNCTIONAL PROFILE OF PROTECTION

The article analyzes the possibilities of using regular expressions to build a functional security profile (FSP) in information systems. The methodology consists in the use of regular expressions, such as the construction of syntax trees, the definition of partial and internal patterns, as well as the development of algorithms for searching for substrings in the text. A functional security profile is considered as a string sequence where each character represents a specific security element. Using regular expressions, it is possible to automate the process of checking for completeness, consistency, and identification of security features in information systems.

The purpose *of this paper is to demonstrate regular expressions as an effective tool for analyzing security features, as they allow standardizing and optimizing the process of assessing the security of information systems. Practical examples of the use of syntactic trees, which provide effective detection of prefixes,*

substrings and other string structures, were considered. The methodology for searching for partial patterns in text data is described, which is the basis for automated decision-making systems during the state examination of information systems.

Conclusions. *The scientific novelty lies in the use of the obtained results to create software that supports decision-making on the assessment of the functional security profile, which significantly reduces the analysis time and increases the efficiency of the process. The work emphasizes the importance of using regular expressions in modern methods of ensuring information security.*

Key words: *regular expressions, functional protection profile, syntax tree, partial patterns, state expertise.*

Вступ. Регулярний вираз – це послідовність символів, що визначає шаблон пошуку в тексті. Він використовується для операцій типу «знайти» або «знайти та замінити» в текстових даних. У сучасних інформаційних системах регулярні вирази відіграють важливу роль при аналізі функціонального профілю захисту (ФПЗ), що є основою для оцінки безпеки автоматизованих систем.

Актуальність. Під час державних експертиз інформаційних систем виникає необхідність в оцінці рівня захищеності даних (Щербина і Скопа, 2008, с. 25). Для цього застосовується функціональний профіль захисту (Давиденко і Шабан, 2014, с. 116), який містить список необхідних функцій безпеки та їх рівнів. Проблеми оцінки захищеності автоматизованих систем часто пов'язані з класифікацією ресурсів, що потребують захисту, і встановленням відповідності між функціями безпеки та реальними загрозами. Регулярні вирази дозволяють стандартизувати процес аналізу, полегшуючи виявлення вразливостей і протиріч.

ФПЗ, як строкова послідовність. Функціональний профіль захисту можна уявити як строкову послідовність (Шабан, 2016, с. 93). Її елементи (символи) відображають окремі функції або атрибути безпеки. Строки можуть бути:

1. **Лінійними** – послідовність із чітко визначеним початком і кінцем.
2. **Циклічними** – без чіткого початку чи кінця.
3. **Нескінченними** – строка, що має лише один кінець (лівий або правий) чи не має їх взагалі.

Регулярні вирази допомагають автоматизувати аналіз цих строкових послідовностей (Фридл, 2008).

Внутрішні патерни. Патерни (шаблони) є ключовим елементом регулярних виразів. Вони діляться на:

- **Внутрішні** (характерні для будь-якої строки).
- **Часткові** (конкретні підстроки, які можна виділити за допомогою пошуку).

Для аналізу використовується **синтаксичне дерево**, яке відображає всі можливі підстроки множини рядків. Листки дерева відповідають конкретним строкам, а ребра позначені символами рядків. Це дерево дозволяє ефективно знаходити префікси, часткові патерни та інші елементи строкової структури.

Часткові патерни. Часткові патерни знаходять застосування в задачах пошуку підстрок. Наприклад, алгоритм пошуку підстрок у тексті працює так:

```
for i=1 to n – m + 1 do
j= порівняти(i,m)
if j = m + 1 then
output i
```

Цей алгоритм простий, але ефективний. Він порівнює кожну позицію тексту з заданим шаблоном і виводить усі входження шаблону в текст.

Використання регулярних виразів для аналізу ФПЗ. Регулярні вирази полегшують такі завдання:

1. Пошук та ідентифікація функцій безпеки.
2. Визначення кореляцій між функціями та ризиками.
3. Автоматизація перевірки на повноту та узгодженість профілю.

Це дозволяє стандартизувати підхід до аналізу ФПЗ, скорочуючи час та підвищуючи точність оцінки.

Висновки. У статті проаналізовано використання регулярних виразів для роботи з функціональним профілем захисту. Було розглянуто ключові методи, такі як побудова синтаксичного дерева, пошук часткових патернів і автоматизація строкових операцій. Ці підходи можуть бути інтегровані у системи підтримки прийняття рішень під час проведення державних експертиз інформаційних систем, що значно підвищує їх ефективність.

ЛІТЕРАТУРА:

1. Давиденко А. Н., Шабан М. Р. Разработка методики проведения экспертиз комплексных систем защиты информации. Зб. наук. праць ІПМЕ НАН України. Київ, 2014. Вип. 73. С. 114–121.

2. Щербина В., Скопа А. А. Проблемы оценки защищенности автоматизированных систем. *Науково-технічний журнал «Захист інформації»*. Київ, 2008. № 4. С. 23–28.

3. Шабан М. Р. Формалізація правил перевірки повноти та несуперечності функціонального профілю захисту. Зб. наук. праць ІПМЕ НАН України. Київ, 2016. Вип. 76. С. 89–94.

4. Фридл Дж Регулярные выражения, 3-е издание. Фридл Дж – СПб.: Символ_Плюс, 2008. С. 188–190.

REFERENCES:

1. Davidenko, A. N., & Shaban, M. R. (2014). Development of a methodology for conducting expertise of comprehensive information protection systems. *Collection of scientific works of the Institute of Modeling Problems in Power Engineering of the NAS of Ukraine*, (73), 114–121. [in Ukrainian].

2. Shcherbyna, V., & Skopa, A. A. (2008). Problems of assessing the security of automated systems. *Scientific and Technical Journal "Information Security"*, (4), 23–28. [in Ukrainian].

3. Shaban, M. R. (2016). Formalization of rules for checking the completeness and consistency of the functional protection profile. *Collection of scientific works of the Institute of Modeling Problems in Power Engineering of the NAS of Ukraine*, (76), 89–94. [in Ukrainian].

4. Friedl, J. (2008). *Mastering Regular Expressions* (3rd ed.). St. Petersburg: Symbol_Plus. 188–190. [in Ukrainian].

УДК 004.056:004.75
DOI https://doi.org/10.32782/IT/2024-4-28

Максим ШАБАН
кандидат технічних наук, викладач кафедри кібербезпеки, інформаційних технологій та економіки, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе, 210, м. Київ, Україна, 02121
ORCID: *0000-0003-2706-8235*

Наталія ДЯЧЕНКО
кандидат наук з державного управління, доцент кафедри кібербезпеки, інформаційних технологій та економіки, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе, 210, м. Київ, Україна, 02121
ORCID: *0000-0002-4306-7665*
Scopus Author ID: *57216565101*

Бібліографічний опис статті: Шабан, М., Дяченко, Н. (2024). Організація правил перевірки функціонального профілю захисту. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 230–234, doi: https://doi.org/10.32782/IT/2024-4-28

ОРГАНІЗАЦІЯ ПРАВИЛ ПЕРЕВІРКИ ФУНКЦІОНАЛЬНОГО ПРОФІЛЮ ЗАХИСТУ

У статті розглянуто процес формалізації правил перевірки функціонального профілю захисту (ФПЗ), який є основою для проведення державних експертиз комплексних систем захисту інформації (КСЗІ). ***Науковою новизною*** *роботи є використання функціонального профілю захисту, що використовується для оцінки повноти та несуперечності реалізації функціональних послуг безпеки (ФПБ) у грід-системах, а також для ідентифікації загроз інформаційній безпеці. Особлива увага приділяється формалізації правил, визначених нормативним документом НД ТЗІ 2.5.004-99 «Критерії оцінки захищеності інформації в комп'ютерних системах від несанкціонованого доступу», які дозволяють автоматизувати процес перевірки ФПЗ.*

Метою статті *є розгляд та вдосконалення трьох основних завдань, пов'язаних з ФПЗ: визначення рівнів реалізації ФПБ, оцінка їхньої повноти та несуперечності, а також ідентифікація опису функціональних послуг у вхідній документації. Описано математичну модель ФПЗ, яка враховує рівні ФПБ, набір критеріїв (конфіденційність, цілісність, доступність тощо) та умови реалізації послуг. Формалізація правил дозволяє ефективно перевіряти відповідність профілю нормативним вимогам та автоматизувати цей процес.*

Методологія *статті полягає в формулюванні умов виконання функціональних послуг безпеки, таких як «Довірча конфіденційність», із використанням матриць та рівнянь. Аналіз таких послуг демонструє, що їхня взаємозалежність і винятковість можуть бути враховані в автоматизованих системах перевірки ФПЗ.*

Висновки. *Запропоновані підходи є основою для створення систем підтримки прийняття рішень, що підвищують точність та швидкість проведення державної експертизи КСЗІ. Результати дослідження спрямовані на вдосконалення методологій забезпечення інформаційної безпеки у грід-системах.*

Ключові слова: *функціональний профіль захисту, функціональні послуги безпеки, автоматизація перевірки, грід-системи, державна експертиза.*

Maksym SHABAN
Candidate of Technical Sciences, Lecturer at the Department of Cybersecurity, Information Technologies and Economics, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121, sabanmaxim@gmail.com
ORCID: *0000-0003-2706-8235*

Natalia DIACHENKO
Candidate of Sciences in Public Administration, Associate Professor at the Department of Cybersecurity, IT and Economics, Kyiv University of Intellectual Property and Law, National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121, n.diachenko@ukr.net
ORCID: *0000-0002-4306-7665*
Scopus Author ID: *57216565101*

To cite this article: Shaban, M., Diachenko, N. (2024). Orhanizatsiia pravyl perevirky funktsionalnoho profiliu zakhystu [Organization of rules for verification of the functional profile of protection]. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 4, 230–234, doi: https://doi.org/10.32782/IT/2024-4-28

ORGANIZATION OF RULES FOR VERIFICATION OF THE FUNCTIONAL PROFILE OF PROTECTION

The article considers the process of formalizing the rules for verifying the functional security profile (FS), which is the basis for conducting state examinations of integrated information security systems (IISS). The scientific novelty of the work is the use of a functional security profile used to assess the completeness and consistency of the implementation of functional security services (FSS) in grid systems, as well as to identify threats to information security. Particular attention is paid to the formalization of the rules defined by the normative document ND TZI 2.5.004-99 «Criteria for assessing the security of information in computer systems against unauthorized access», which allow automating the process of checking the FSIs.

***The purpose** of the article is to consider and improve three main tasks related to the FSI: determining the levels of implementation of the FSI, assessing their completeness and consistency, and identifying the description of functional services in the incoming documentation. The article describes a mathematical model of the FSF that takes into account the levels of FSF, a set of criteria (confidentiality, integrity, availability, etc.) and the conditions for the implementation of services. The formalization of rules allows to effectively check the compliance of the profile with regulatory requirements and automate this process.*

***The methodology** of the article is to formulate the conditions for the implementation of functional security services, such as «Trust Confidentiality», using matrices and equations. The analysis of such services demonstrates that their interdependence and exclusivity can be taken into account in automated systems for verifying FSS.*

***Conclusions.** The proposed approaches are the basis for the creation of decision support systems that increase the accuracy and speed of the state examination of IPSS. The results of the study are aimed at improving the methodologies for ensuring information security in grid systems.*

***Key words:** functional protection profile, functional security services, automation of verification, grid systems, state expertise.*

Одним з ключових завдань при проведенні державної експертизи є ідентифікація функціонального профілю захисту (ФПЗ). Основною метою створення ФПЗ є нейтралізація загроз несанкціонованого доступу до інформації (Давиденко і Шабан, 2014, с. 114), яка забезпечується реалізацією комплексів засобів захисту (КЗЗ) політики функціональних послуг (Новіков і Тимошенко, 2002, с. 40). У завдання ідентифікації (Зегжда і Калінін, 2002, с. 7) входять такі підзадачі: визначення рівнів функціональних послуг безпеки (ФПБ) реалізованих в комплексній системі захисту інформації (КСЗІ) об'єкта експертизи; визначення повноти та несуперечності профілю; ідентифікація опису (Лукацький, 2016) ФПБ у вхідних документах. При проведенні другої підзадачі необхідно враховувати правила побудови функціонального профілю захисту визначених НД ТЗІ 2.5.004-99 «Критерії оцінки захищеності інформації в комп'ютерних системах від несанкціонованого доступу». Формалізація цих правил повинна створити сприятливі умови для автоматизації перевірки повноти та несуперечності ФПЗ.

Отже, перед нами стоять три завдання: визначення рівнів ФПБ реалізованих КСЗІ об'єкта експертизи; визначення повноти та несуперечності профілю; ідентифікація опису ФПБ у вхідних документах.

Визначення рівнів ФПБ реалізованих КСЗІ об'єкта експертизи включає попередній аналіз об'єкта експертизи в ході якого, шляхом пошагового опрацювання кожної ФПБ, формується ФПЗ.

На етапі визначення повноти та несуперечності ФПЗ експерт повинен перевірити відповідність отриманого на попередньому етапі ФПЗ вимогам НД ТЗІ 2.5.004-99 (Яловець, 2011, с. 25).

Останнє завдання, яке повинен вирішити експерт – опрацювати вхідні документи на предмет пошуку опису реалізації умов, що пред'являються до кожної ФПБ.

Розглянемо друге завдання більш детально. Вимоги до ФУБ задані у вигляді таблиць показують відповідність рівня реалізованої послуги переліком необхідних вимог для її коректної роботи (Гільгурт, Дурняк і Коростиль, 2014, с. 3).

Для автоматизації перевірки повноти та несуперечності профілю формалізуємо правила, які визначені у НД ТЗІ 2.5.004-99 «Критерії оцінки захищеності інформації в комп'ютерних системах від несанкціонованого доступу».

Перше правило: будь-який профіль зобов'язаний включати в себе контроль цілісності

КСЗІ. Позначимо профіль через множіство Fp тоді: $F_p = \{ФПБ0, ФПБ1, ..., ФПБN\} \cap НЦ_i \neq \varnothing$,

де ФПБNN – функціональна послуга безпеки, а

$НЦ_i$ – «Цілісність комплексу засобів захист» і-ого рівня.

Більш детально розберемо необхідність цієї умови. Розберемо ФПБ «Цілісність комплексу засобів захисту» більш детально. В умовах виконання НЦ рівня 1 сказано: «Політика цілісності КЗЗ повинна визначати склад КЗЗ і механізми контролю цілісності компонентів, що входять до складу КЗЗ»; «В разі виявлення порушення цілісності будь-якого із своїх компонентів КЗЗ повинен повідомити адміністратора та автоматично відновити відповідність компонента еталону, або перевести КС до стану з якого повернути її до нормального функціонування може тільки адміністратор або користувачі, яким надані відповідні повноваження»; «Повинні бути описані обмеження, дотримання яких дозволяє гарантувати, що послуги безпеки доступні тільки через інтерфейс КЗЗ і всі запити на доступ до захищених об'єктів контролюються КЗЗ». Тобто, невиконання цих правил автоматично робить неможливим гарантування безпеки експлуатації об'єкта експертизи і унеможливлює проведення державної експертизи.

Друге правило: відповідно до НД ТЗІ 2.5.004-99 «Критерії оцінки захищеності інформації в комп'ютерних системах від несанкціонованого доступу» деякі з ФПБ є умовно пов'язаними з іншими ФПБ. Іншими словами в НД ТЗІ 2.5.004-99 «Критерії оцінки захищеності інформації в комп'ютерних системах від несанкціонованого доступу» при наявності послуг потрібна наявність інших ФПБ, наприклад: для виконання умов ФПБ КД рівня 1 необхідною умовою є виконання умов ФПБ НИ рівня 1. Формалізуємо дану умову:

ФПБ $КД1 = \{УКД1, УКД2, ..., УКДN\} \cap УКД7.1 \neq \varnothing$.

Третє правило: якщо послуга має усі 4-ри рівня, то в функціональному профілі захисту може бути тільки одна. Третє правило говорить про поглинання старшим ФПБ молодших. Тобто, в одному профілі не можуть бути ФПБ однієї спрямованості різного рівня. Наприклад, якщо в ФПБ присутня послуга КД рівня 3 – це автоматично означає, що в даному ФПБ не можуть бути послуги КД рівня 1 або 2. Функціональні послуги одного типу мають властивості винятковості, причому пріоритет віддається послугам вищого рівня.

У подальшому постало питання необхідності якось формалізувати, з математичної точки зору, функціональний профіль послуг безпеки (ФППБ). Функціональний профіль послуг безпеки Fu – це функціональний профіль, що включає в себе множину актуальних послуг безпеки реалізованої КСЗІ об'єкта експертизи. ФППБ залежить від 3 параметрів: Кр – група критеріїв; Ri – рівень реалізованої послуги безпеки; Ui – кон'юнкція умов реалізації виконання і-тої ФПБ.

$$Fu(K_p R_i U_i);$$

$$Kp = \{C, I, A, O, G\};$$

де C – критерій конфіденційності; I – критерій цілісності; A – критерій доступності; O – критерій спостережності; G = критерій гарантій. Ri – це кількість рівнів кожної ФПБ, яка персоніфіковано залежить від виду послуги. Максимальний рівень, згідно з нормативним документом НД ТЗІ 2.5.004-99 «Критерії оцінки захищеності інформації в комп'ютерних системах від несанкціонованого доступу», четвертий, але в залежності від ФПБ зустрічаються ФПБ з одним рівнем тощо.

$$Ri = \{1, 2, 3, 4\};$$

де i = 1-23 – це кількість ФПБ згідно з даним ФПЗ, де 23 – це максимальна кількість ФПБ отриманих з нормативного документу НД ТЗІ 2.5.004-99 «Критерії оцінки захищеності інформації в комп'ютерних системах від несанкціонованого доступу».

Кон'юнкція умов реалізації виконання і-тої ФПБ (Ui) – це сума усіх включених ФПБ до профілю захисту, де Ni – кількість умов до і-послуги; $Y_{i,j}$ – це j умова реалізації і-ої послуги.

$$Ui, R_i = \bigcup_{j=1}^{Ni} Y_j \underline{F} (i, R_i, j) \quad (1.1);$$

Розглянемо приклад аналізу КД.

$$КД1 = \{УКД1.1, УКД2.1, УКД3.1, УКД4.1, УКД6.1, УКД7.1\} \quad (1.2)$$

$$КД2 = \{УКД1.1, УКД2.2, УКД3.1, УКД4.2, УКД5.2, УКД6.1, УКД7.1\}$$

$$КД3 = \{УКД1.2, УКД2.2, УКД3.1, УКД4.3, УКД5.2, УКД6.1, УКД7.2\}$$

$$КД4 = \{УКД1.2, УКД2.3, УКД3.1, УКД4.4, УКД5.2, УКД6.1, УКД7.2\},$$

де УКД – умова виконання ФПБ «Довірча конфіденційність».

На основі цього побудуємо матрицю значень для кожного рівня ФПБ КД використовуючи рівняння:

$$Ui, R_i = \bigcup_{j=1}^{Ni} Y_j \underline{F} (i, R_i, j) \quad (1.1),$$

де U = КД, Y = УКД, $\underline{КД} = \underline{УКД}$;

$$КД = \{КД1\, КД2\, КД3\, КД4\} \quad (1.3);$$

Виходячи з цього представимо вектор УКД у вигляді добутку матриць:

$$\overline{УКД} = \{УКД1 УКД2 УКД3 УКД4 УКД5 УКД6 УКД7\} \begin{Bmatrix} 1122 \\ 1223 \\ 1111 \\ 1224 \\ 0112 \\ 1111 \\ 1113 \end{Bmatrix} (1.4);$$

Для КД 1 це буде мати такий вигляд:

$$\overline{УКД} = \{1234567\} \begin{Bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 1 \end{Bmatrix} (1.5);$$

Для КД 2:

$$\overline{УКД} = \{1234567\} \begin{Bmatrix} 1 \\ 2 \\ 1 \\ 2 \\ 1 \\ 1 \\ 1 \end{Bmatrix} (1.6);$$

Для КД 3:

$$\overline{УКД} = \{1234567\} \begin{Bmatrix} 2 \\ 2 \\ 1 \\ 3 \\ 2 \\ 1 \\ 2 \end{Bmatrix} (1.7);$$

Для КД 4:

$$\overline{УКД} = \{1234567\} \begin{Bmatrix} 2 \\ 3 \\ 1 \\ 4 \\ 2 \\ 1 \\ 2 \end{Bmatrix} (1.8);$$

Опис 1.5 – 1.8 є окремими випадками формули 1.1 за умови і = 1. Проведемо обчислення за формулою 1.3. Порахуємо добуток матриць згідно з правилом множення матриць: перший елемент першої матриці множимо на перший елемент другої матриці. Далі множимо другий елемент першої матриці на другий елемент другої матриці тощо. Коли число стовпців в першому співмножнику рівне числу рядків у другому, то в цьому випадку говорять, що форма матриць узгоджена. Як наслідок, отримаємо КД 1:

$$КД1 = \{УКД1.1, УКД2.1, УКД3.1, УКД4.1, УКД6.1, УКД7.1\}$$

Підводячи підсумок слід зазначити, що ми розглянули окремий випадок кон'юнкції умов реалізації виконання ФПБ КД. Розглянутий випадок підтверджує відповідність 1.2 до рівняння 1.1. Рівняння 1.1 дозволяє формалізувати процедуру перевірки повноти та несуперечності функціонального профілю захисту.

Висновки. Таким чином, завдання визначення повноти та несуперечності ФПЗ може бути зведена до перевірки запропонованих вище правил, а саме: правило контролю цілісності КСЗІ; правило контролю взаємозв'язку одних ФПБ по відношенню до інших; правило контролю рівнів ФПБ, що дозволяє автоматизувати цей процес шляхом побудови відповідної комп'ютерної системи підтримки прийняття рішень під час проведення державної експертизи на відповідність інформації циркулюючої в грід-системі. Проведена робота дає теоретичну основу для практичної реалізації системи автоматизації перевірки повноти та несуперечності ФПЗ.

ЛІТЕРАТУРА:

1. Зегжда Д. П., Калінін М. О. Методика аналізу захищеності інформаційних систем. *Проблеми інформаційної безпеки. Комп'ютерні системи*, 2002. (3), 7–12.

2. Лукацький А. Упевненість фахівців з безпеки у своїх силах похитнулась. 2016. URL: http://gblogs.cisco.com/ru/asr2016-2/.

3. Давиденко А. Н., Шабан М. Р. Розробка методики проведення експертиз комплексних систем захисту інформації. *Збірник наукових праць Інституту проблем моделювання в енергетиці імені Г. Є. Пухова НАН України*, 2014. (73), 114–121.

4. Новіков О. М., Тимошенко А. О. (2002). Логіко-функціональні моделі безпеки інформації в інформаційно-обчислювальних системах з відкритою архітектурою. *Наукові вісті НТУУ «КПІ»*, (2), 40–46.

5. Гільгурт Я., Дурняк Б. В., Коростиль Ю. М. Протидія атакам алгоритмічної складності на системи виявлення вторгнень. *Моделювання та інформаційні технології*, 2014. (71), 3–12.

6. Яловець А. Л. Представлення та обробка знань з точки зору математичного моделювання. Київ: Наукова думка. 2011.

REFERENCES:

1. Zegzhda, D. P., & Kalinin, M. O. (2002). Methodology for analyzing the security of information systems. *Problems of Information Security. Computer Systems*, (3), 7–12.

2. Lukatskyi, A. (2016). Confidence of security specialists has been shaken [Electronic resource]. Retrieved from: http://gblogs.cisco.com/ru/asr2016-2/.

3. Davidenko, A. N., & Shaban, M. R. (2014). Development of a methodology for conducting expertise of comprehensive information protection systems. *Collection of scientific works of the Institute of Modeling Problems in Power Engineering of the NAS of Ukraine*, (73), 114–121.

4. Novikov, O. M., & Tymoshenko, A. O. (2002). Logical-functional models of information security in information-computing systems with open architecture. *Scientific News of NTUU "KPI"*, (2), 40–46.

5. Hilhurt, Y., Durniak, B. V., & Korostyl, Y. M. (2014). Counteracting algorithmic complexity attacks on intrusion detection systems. *Modeling and Information Technologies*, (71), 3–12.

6. Yalovets, A. L. (2011). Representation and processing of knowledge from the perspective of mathematical modeling. Kyiv: Naukova Dumka.

УДК 502.7:004.94
DOI https://doi.org/10.32782/IT/2024-4-29

Геннадій ШВАЧИЧ
доктор технічних наук, професор, професор кафедри програмного забезпечення комп'ютерних систем, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, м. Дніпро, Україна, 49005
ORCID: *0000-0002-9439-5511*
Scopus Author ID: *56509642500*

Павло ІЩУК
аспірант кафедри програмного забезпечення комп'ютерних систем, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, м. Дніпро, Україна, 49005
ORCID: *0000-0001-6399-6771*

Бібліографічний опис статті: Швачич, Г., Іщук, П. (2024). Ідентифікація динаміки довкілля: виклики та перспективи розподіленого моделювання. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 235–241, doi: https://doi.org/10.32782/IT/2024-4-29

ІДЕНТИФІКАЦІЯ ДИНАМІКИ ДОВКІЛЛЯ: ВИКЛИКИ ТА ПЕРСПЕКТИВИ РОЗПОДІЛЕНОГО МОДЕЛЮВАННЯ

Дослідження спрямовані на розв'язок коефіцієнтних задач екології методами розподіленого моделювання.

Мета роботи *полягає у моделюванні процесів забруднення атмосфери, які адекватні до тих, що описують реальні процеси. При цьому виникає можливість досліджувати такі питання як детальний аналіз стану атмосфери міста або промислового району, короткостроковий прогноз якості повітря в регіоні, оцінка довгострокових програм по очищенню повітря, оптимальне управління викидами, трансграничне перенесення і т. п.*

Методологія. *Розглядаються обернені задачі екології, постановка яких формулюється через аналіз взаємозв'язків типу «причина-наслідок». Причинні характеристики екологічного процесу, згідно з прийнятою математичною моделлю, визначаються коефіцієнтами рівняння перенесення шкідливих домішок. У такій інтерпретації встановлення причинно-наслідкових зв'язків є метою прямих задач екології. У протилежному випадку, якщо необхідно відновити причинні характеристики за інформацією про поле забруднення атмосфери, формулюється одна з постановок зворотних задач екології, які належать до класу некоректних задач за Адамаром.*

Наукова новизна *проведених досліджень полягає у розробці нового підходу до обробки даних на всіх етапах екологічного експерименту із застосуванням сучасних обчислювальних засобів. Обчислювальні експерименти виконувалися із застосуванням розподілених паралельних обчислювальних систем. Це зумовлено не лише принциповими обмеженнями максимальної швидкодії звичайних послідовних ЕОМ, але й постійною наявністю задач, для розв'язування яких існуючі обчислювальні ресурси виявляються недостатніми. До таких задач належать, зокрема, задачі моделювання клімату.*

Висновки. *У даних дослідженнях розглянуті постановка і методи розв'язку задач ідентифікації динаміки довкілля, суть яких полягає в оцінці вхідних параметрів за фактичною інформацією про модельовану систему, відому з експерименту. Для аналізу оцінки можливості реалізації математичних моделей на засобах обчислювальної техніки був розроблений комплекс програм для розрахунку перенесення шкідливих домішок. Наведено розв'язок тестової задачі щодо ідентифікації динаміки довкілля.*

Ключові слова: *екологія, розподілене моделювання, обернені задачі, пакет програм, експеримент, довкілля.*

Hennadii SHVACHYCH
Doctor of Technical Sciences, Professor, Professor at the Department of Software Engineering, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, sgg1@ukr.net
ORCID: *0000-0002-9439-5511*
Scopus Author ID: *56509642500*

Pavlo ISHCHUK
Postgraduate Student at the Department of Software Engineering, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, Vinceras@gmail.com
ORCID: *0000-0001-6399-6771*

To cite this article: Shvachych, H., Ishchuk, P. (2024). Identyfikatsiia dynamiky dovkillia: vyklyky ta perspektyvy rozpodilenoho modeliuvannia [Identification of environmental dynamics: challenges and prospects of distributed modeling]. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 235–241, doi: https://doi.org/10.32782/IT/2024-4-29

IDENTIFICATION OF ENVIRONMENTAL DYNAMICS: CHALLENGES AND PROSPECTS OF DISTRIBUTED MODELING

The research is of significant importance as it aims to effectively address and solve coefficient problems in ecology using distributed modeling methods.

The work aims *to model atmospheric pollution processes that are not only adequate to those describing real processes but also have practical applications. This provides an opportunity to investigate and address issues such as a detailed analysis of the atmosphere state of a city or industrial area, short-term forecast of air quality in the region, assessment of long-term air purification programs, optimal emission management, transboundary transfer, etc.*

Methodology. *Inverse ecological problems, which involve determining the cause of an observed effect, are considered and formulated through the analysis of cause-effect relationships. According to the adopted mathematical model, the ecological process's causal characteristics are determined by the coefficients of the equation for the transfer of harmful impurities. In this interpretation, establishing cause-effect relationships is the goal of direct ecological problems, which involve determining the effect of a known cause. In the opposite case, if it is necessary to restore the causal characteristics based on information about the atmospheric pollution field, one of the formulations of inverse ecological problems is formulated, which belongs to the class of Hadamard-incorrect problems.*

The scientific novelty *of the conducted research is a significant breakthrough in the field. It lies in developing a new approach to data processing at all stages of an ecological experiment using modern computing resources. Computational experiments were performed using distributed parallel computing systems. This is not only due to the fundamental limitations of the maximum speed of conventional sequential computers but also to the constant presence of problems for the solution of which existing computing resources are insufficient. Such problems include, in particular, climate modeling problems.*

Conclusions. *These studies consider the formulation and methods of solving problems of identifying environmental dynamics, which essence is to estimate input parameters based on actual information about the modeled system known from the experiment. To analyze the assessment of the possibility of implementing mathematical models using computer technology, a set of programs was developed to calculate the transfer of harmful impurities. The solution of the test problem of identifying environmental dynamics is presented.*

Key words: *ecology, distributed modeling, inverse problems, software package, experiment, environment.*

Актуальність проблеми. Бурхливий розвиток промисловості в усіх країнах світу поставив перед людством гостру проблему охорони довкілля з метою збереження екологічних систем, що історично сформувалися в різних регіонах нашої планети. У зв'язку з цим екологічна наука нині просто-таки вимушена займатися проблемами невідомої раніше складності. Нові завдання вимагають і нових способів їх рішення. Але вже сьогодні математичне моделювання, обчислювальний експеримент на паралельних кластерних системах з досить повним математичним наповненням тих явищ, які ми вивчаємо, утворює основу нової технології наукового пошуку, аналізу і прогнозу (Петрик, 2020;, Семеренко, 2018). Реальність сьогоднішнього дня полягає в тому, що у зв'язку з розвитком паралельних обчислень і, особливо з появою обчислювальних кластерів, зникли принципові проблеми в потенційно нескінченному збільшенні пікової продуктивності комп'ютерів (Shvachych, 2021; Shvachych,2022). Це відкриває шлях до широкого застосування математичного моделювання, що зробить результативнішими як фундаментальні, так і прикладні дослідження (Moroz, 2020; Shvachych, 2021). Отже, проблема ідентифікації динаміки довкілля є на сьогоднішній день актуальною та першочерговою.

Постановка задачі. Важливим завданням нині є прогнозування змін екологічних систем під впливом природних процесів і антропогенного впливу (Mamuzić, 2024). Теоретичне дослідження таких явищ базується на загальнодоступних моделях механіки суцільних середовищ і, зокрема, на рівняннях Навье-Стоксу, на рівняннях енергії і турбулентної дифузії, встановлених ще в позаминулому столітті і перевірених

в численних експериментах. У математичному вираженні це системи багатовимірних нелінійних рівнянь, що мають в лінеаризованому вигляді форму лінійних взаємозв'язаних задач з початковими даними в еволюційній векторній формі Коши-Ковалевської (Швачич, 2008):

$$\frac{Df}{Dt} = Af\ , \qquad (1)$$

$$f_{|t=t_0} = f_{\Delta}\ , \qquad (2)$$

де: *D/Dt* – субстанціональна похідна;

A – оператор, що включає просторові змінні і дисипативні функції;

f – шукана вектор-функція залежних змінних.

Суть проблем аеродинаміки довкілля полягає в тому, щоб описати поширення в природних середовищах різних забрудників, які викидаються в довкілля. При цьому перенесення забруднюючих речовин визначається двома процесами. Ці процеси включають конвективне перенесення через усереднений рух середовища та дифузію, викликану турбулентністю. Відповідно, математична модель (1), (2) повинна адекватно відображати як розподіл середніх швидкостей, так і особливості турбулентної дифузії. Зважаючи на це рівняння усередненого руху містять члени турбулентного перенесення (турбулентна напруга), то для замикання турбулентної задачі можна використовувати цілком певні моделі турбулентності (Vozna, 2021; Lienhart, 2002). У ідеальному випадку модель турбулентності має бути настільки загальною, щоб для усіх практично важливих задач її можна було застосувати без змін. Зауважимо, що моделі з двома рівняннями перенесення типу $\kappa - \varepsilon$ є найбільш зручними при моделюванні подібних течій.

Мета даної роботи полягає у моделюванні процесів забруднення атмосфери на основі моделей (1), (2), які адекватно тих, що описують реальні процеси. При цьому виникає можливість досліджувати такі питання як детальний аналіз стану атмосфери міста або промислового району, короткостроковий прогноз якості повітря в регіоні, оцінка довгострокових програм по очищенню повітря, оптимальне управління викидами, трансграничне перенесення і т. п.

Основні результати досліджень. Розглянемо деяку область Ω, обмежену підстилаючою поверхнею S_o з лінійними розмірами *L*, вільними бічними і верхньою S_n поверхнями в декартовій системі координат.

Поверхня S_o в прийнятій системі координат є функцією горизонтальних змінних x, y, що враховують рельєф місцевості, міську забудову та інші неоднорідності на поверхні Землі. Верхня межа S_n в загальному випадку може змінюватися в часі і в просторі. У окремому випадку, коли поверхні S_o і S_n площини, простір регіону представляється у вигляді паралелепіпеда.

У багатьох завданнях математичної фізики тимчасова змінна t в (1), (2) грає особливу роль. Обчислювальний алгоритм в таких задачах складається з актів переходу системи від одного стану до іншого. Для механіки суцільних середовищ система диференціальних рівнянь (СДР) (1) є СДР у частинних похідних другого порядку. Тому до умови (2) необхідно додати граничні умови. Необхідні граничні умови і належний їх вибір залежить від фізичної постановки задачі. Для даного класу задач на поверхні S_o задаються умови прилипання для поля швидкостей, а також граничні умови 1-3 або 4-го роду для температурних або концентраційних функцій. Верхня межа S_n розрахункової області задається або як межа атмосферного пограничного шару, або як межа інверсійного шару (Лаврик, 2010). Умови на вільних поверхнях (бічних поверхнях розрахункової області) зазвичай спирається на припущення про загасання мезо-масштабних обурень на межах розрахункової області. Проте, для при складному рельєфі місцевості обурення поширюються на дуже великі відстані. У зв'язку з цим розумніше на вільних межах задавати граничні умови 1-го роду, уточнюючи їх значення за конкретних умов шляхом порівняння з даними натурних випробувань. Окрім цього, необхідно відмітити, що місце розташування вільних меж не може бути фіксованим. Тому у ряді випадків доцільно задавати прозорі межі, т. т. умови 4-го роду, коли на розрахунковій межі області витримується умова безперервності шуканих функцій і її градієнтів, але при цьому при цьому враховується поведінка шуканих функцій на значних відстанях від вільних меж розрахункової області.

Таким чином, математична модель (1), (2) із заданою областю визначення (рис. 1) і заданими граничними умовами, визначена і достатня для реалізації у вигляді програмних засобів на скалярних або паралельних обчислювальних пристроях (Башков, 2011; Іващенко, 2011). Хід розв'язання, за якого при заданих умовах однозначності, що стосуються причинних характеристик досліджуваного процесу, буде визначати певний стан довкілля. За таких умов його можна розглядати як розв'язок прямої задачі. З точки зору співвідношень причина – наслідок

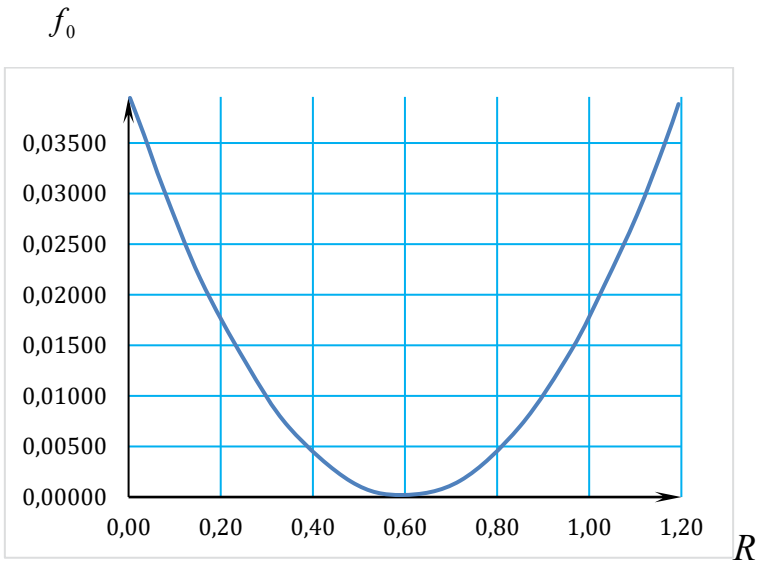


Рис. 1. Графічна інтерпритація розв'язку коефіцієнтної оберненої задачі для визначення потужності точкового джерела забруднення

кожній прямій задачі, в межах прийнятої структури (1), (2), може бути зіставлена деяка безліччю обернених задач. Отже, встановлення причинно – наслідкових зв'язків складає ланцюг прямих задач екології. І навпаки, якщо за певною інформацією про стан довкілля вимагається відновити причинні характеристики, то маємо ту або іншу постановку оберненої задачі. Таким чином, розв'язок будь-якої конкретної задачі з цього класу визначається не лише як функція просторових координат і часу, але і як функція вхідних параметрів R. Тоді розв'язок прямої задачі (1), (2) реалізує перетворення

$$f = f_R(x, y, z, t, R), \qquad (3)$$

де R – вектор – функція причинних параметрів.

Розглядаючи математичну модель, де параметрами, що управляють, є компоненти R, сформулюємо мету управління. Вибиратимемо управління R з умови певної узгодженості відомої з експерименту функції $f = f_e(t)$ з розрахунковим значенням з $f = f_p(t)$, обчисленим за допомогою моделі і такою, що відповідає цьому управлінню R. Нехай окрім керованого об'єкту і поставленої мети управління задано і критерій якості, представлений у вигляді функціонала

$$J(R) = \int_{t_0}^{t_k} Y(R, f_p(t), f_e(t))dt \qquad (4)$$

де $Y(R, f_p, f_e)$ – відома скалярна функція, безперервна разом з усіма частинними похідними за всією сукупності аргументів. Варіюючи параметрами, що управляють, набуватимемо різних значень функціонала в просторі керованих параметрів. Таким чином, розв'язок будь-якої оберненої задачі полягатиме в тому, щоб знайти таке управління, що реалізовує мету, для якої функціонал (4) на розв'язках керованої системи приймає можливе найменше значення. Для розв'язків сформульованих обернених задач застосовуються чисельні методи теорії оптимізації, що використовують принципи максимуму Понтрягіна, Белмана та інші. При цьому можливі два випадки:

1) оптимальне розв'язок відшукається в просторі параметрів;

2) задача оптимізації розв'язується у функціональному просторі.

Ці два різні підходи застосовані для розв'язку обернених задач теорії теплопровідності за допомогою одного класу ітераційних методів, відомих під назвою градієнтних. У точній екстремальній постановці визначення параметрів управління відповідає мінімізації середньоквадратичної нев'язки, що характеризує ухилення розрахункової функції від експериментальної в метриці простору вхідних даних. В цьому випадку як ухилення приймається значення функціонала (4), де $Y(t) = (f_p(t) - f_э(t))^2$. Величина функціонала з такою нев'язкою для цього випадку є функціоналом в просторі параметрів управління і його числовим значенням визначає відстань між $f_p(A,t)$ і $f_e(A,t)$ у функціональному просторі L_2. Для одновимірних просторових задач з одним, максимум двома параметрами управління, рішення зворотних задач в такій постановці досить просто зводиться до визначення мінімуму функції багатьох змінних.

Для згладжених експериментальних даних, допустимо, процедурою сплайн – згладжування або пропустивши їх через фільтр, отримаємо функцію I(R), що має достатній запас аналітичної. Тоді в околі її мінімуму допускається розкладання в ряд Тейлора:

$$J_{p+\varepsilon_{R,1}}(\varepsilon_R) = J_{p,1} + \varepsilon_R J_{p,2} + \varepsilon_R^2 J_{p,3} + \ldots \qquad (5)$$

де

$$\left.\begin{aligned} \varepsilon_R &= \frac{R - R_p}{R_{p+1} - r} \in [-1, 1], \\ J_{p,2} &= \frac{1}{2}(J_{p+1,1} - J_{p-1,1}) \\ J_{p,3} &= \frac{1}{2}(J_{p+1,1} + J_{p-1,1} - 2J_{p,1}) \end{aligned}\right\} \qquad (6)$$

безрозмірний нормований аргумент і значення тейлоровських компонент, представлених на вузлах рівномірної сітки $\overline{p = 1, 2m-1}$ центральними різницями. Прирівнявши частинну похідну від (5) по ε_r, отримаємо відому інтерполяційну формулу

$$R = R_p - \frac{(R_{p+1} - R_p)}{2} \frac{(J_{p+1,1} - J_{p-1,1})}{(J_{p+1,1} + J_{p-1,1} - 2J_{p,1})}, \qquad (7)$$

яку можна використовувати для уточнення параметра управління.

Рішення зворотної задачі в такій постановці складається з двох етапів. Перший етап передбачає процес відокремлення мінімуму функціонала на виконання умови

$$(J_{p,1} \times J_{p+1,1}) < 0, \;\; p = 1, 2, \ldots, \qquad (8)$$

коли $J_{p,1}$ *та* $J_{p+1,1}$ мають різні знаки. Тут значення R_p і R_{p+1} можуть бути вибране з апріорної інформації досить довільно. Зрозуміло, що при виконанні умови (8) Imin лежить десь між цими точками. Довизначивши значення функціонала в центральному вузлі, подальше його уточнення реалізується по формулі (7).

Для двох параметрів $R_1 = \alpha$ і $R = \beta$ будується таблиця сіткових функції $J_{p,1}^{\nu,1}$. Процедура відділення мінімуму досить просто вирішується методом простого сортування і далі уточнюється за інтерполяційними формулами типу (7).

Отже, розв'язок будь-якої оберненої задачі з цього класу полягатиме в тому, щоб знайти таке управління, що реалізовує мету, для якого відповідний функціонал на рішеннях керованої системи приймає можливе найменше значення. У такій постановці розв'язок тієї або іншої оберненої задачі може бути реалізоване методом, розробленим в цій роботі.

Результати моделювання процесу ідентифікації динаміки довкілля. Проілюструємо запропонований в роботі підхід для розв'язку деякої прикладної задачі. У цій роботі демонструється паралельна обробка інформації та паралельний тип мислення на прикладі використання багатопроцесорної обчислювальної системи (Башков, 2011; Іващенко, 2011). Завдяки цьому підходу проводилися обчислювальні експерименти.

Тестова задача. Розглянемо розв'язок оберненої задачі з метою визначення потужності джерела Q, що викидає в атмосферу шкідливі домішки. В якості додаткової вхідної інформації, відомої з експерименту, виступає функція:

$$\phi_{|x=x_*} = \phi_e,$$

де x_* – відома координата, а права частина відповідає значенню функції, відомої з експерименту.

На рис. 1 приведені результати розв'язку цієї задачі при точних значеннях функції ϕ_e, відомої з експерименту. По осі абсцис відкладалося значення параметра управління $R = Q$, по осі ординат – квадрата нев'язки $f_0 = (\phi_e - \phi_p)^2$.

Екстремум нев'язким відповідає точці мінімуму цієї функції, що має абсцису:

$$Q \approx 0.6\,[кг/с],$$

і ординату

$$f_0 \approx 0,$$

що відповідає точному рози'явку прямої задачі, представленої в цій редакції на рис. 1.

Висновки

1. У цій роботі розглянуті постановка і методи розв'язку задач ідентифікації динаміки довкілля, суть яких полягає в оцінці вхідних параметрів за фактичною інформацією про модельовану систему, відому з експерименту.

2. Для аналізу оцінки можливості реалізації математичних моделей на засобах обчислювальної техніки був розроблений комплекс програм для розрахунку перенесення шкідливих домішок. Наведено розв'язок тестової задачі щодо ідентифікації динаміки довкілля.

3. Ідентифікація процесів перенесення шкідливих домішок в атмосфері оберненими методами може бути використана для визначення за експериментальними даними коефіцієнтів рівняння, параметрів джерел забруднення, параметрів граничних функцій і т. п.

ЛІТЕРАТУРА:

1. Петрик М. Р., Бойко І. В. Високопродуктивні обчислення та математичне моделювання в наукових дослідженнях. Тернопіль : Тернопільський національний технічний університет імені Івана Пулюя, 2020. 108 с.

2. Семеренко В. П. Технології паралельних обчислень : навчальний посібник. Вінниця : ВНТУ, 2018. 104 с.

3. Shvachych G**.** Implementation of system software of multiprocessor computing systems for solving applied tasks. *Актуальні напрями розвитку технічного та виробничого потенціалу національної економіки* : монографія / G. G. Shvachych, D. M. Moroz. Дніпро: 2021. С. 111–136.

4. Shvachych G., Shcherbyna P., Moroz D. Aggregation of computing channels based on the Nvidia Cuda platform for control modes of components of technological systems. *Information Technology: Computer Science, Software Engineering and Cyber Security*. 2022. № 2. P. 85–92.

5. Shvachych G., Moroz B., Martynenko A., Hulina I., Busygin V., Moroz D. Model of Speed Spheroidization of Metals and Alloys Based on Multiprocessor Computing Complexes. *Machine Learning for Predictive Analysis.* Networks and Systems. Springer. 2020. P. 33–41.

6. Shvachych G., Vozna N., Ivashchenko O., Bilyi O., Moroz D. Efficient algorithms for parallelizing tridiagonal systems of equations. *System technologies.* Dnipro. 2021. № 5 (136). P. 110–119.

7. I. Mamuzić G. G. Shvachych P. O. Ishchuk Challenges of Parallel Numerical Integration of Pollution Transport Equations. *Book of abstracts of the 17th International Symposium of Croatian Metallurgical Society – SHNM '2004 Materials and metallurgy (published in Metalurgija* 63 (2024) 3-4), P. 477–492.

8. Швачич Г. Г., Шмукин А. А. Особенности конструирования параллельных вычислительных алгоритмов для ПЭВМ в задачах тепло- и массообмена. *Восточно-европейский журнал передовых технологий*. 2004. № 3. С. 42–47.

9. Shvachych G., Vozna N., Ivashchenko O., Bilyi O., Moroz D. Method of lines in distributed problems of experimental data processing. *International Academy Journal Web of Scholar.* 2021. № 2(52). P. 1–7.

10. Lienhart H., Stoots C., Becker S. Flow and Turbulence Structures in the Wake of a Simplified Car Model (Ahmed Model). New Results in Numerical and Experimental Fluid Mechanics III. 2002. Vol. 12, p. 323–330.

11. Лаврик В. І. Моделювання і прогнозування стану довкілля : [навч. посіб]. / В. І. Лаврик. К. : ВД КМ Академія, 2010. 400 с.

12. Башков Є. О., Іващенко В. П., Швачич Г. Г. Високопродуктивна багатопроцесорна система на базі персонального обчислювального кластера. *Проблеми моделювання та автоматизації проектування*. 2011. Вип. 9 (179). С. 312–324.

13. Пат. 61944 Україна, МПК C21D 1/26, G06F 15/16 (2011.01). Модуль високоефективної багатопроцесорної системи підвищеної готовності / Іващенко В.П., Башков Є.О., Швачич Г.Г., Ткач М.О. ; патентовласники Національна металургійна академія України, Донецький національний технічний університет. – № u 2010 09341; заявл. 26.07.2010; опубл. 10.03.2011, Бюл. № 5.

REFERENCES:

1. Petryk, M. R. & Boiko, I. V. (2020). Vysokoproduktyvni obchyslennia ta matematychne modeliuvannia v naukovykh doslidzhenniakh [High performance computing and math modeling in scientific research]. Ternopil : Ternopilskyi natsionalnyi tekhnichnyi universytet imeni Ivana Puliuia. 108 s.

2. Semerenko, V. P. (2018). Tekhnolohii paralelnykh obchyslen [Parallel computing technologies]. Vinnytsia : VNTU. 104 s.

3. Shvachych, G**.** & Moroz, D. (2021). Implementation of system software of multiprocessor computing systems for solving applied tasks. *Aktualni napriamy rozvytku tekhnichnoho ta vyrobnychoho potentsialu natsionalnoi ekonomiky : monohrafiia [Current trends in the development of the technical and production potential of the national economy: monograph]*. Dnipro. s. 111–136.

4. Shvachych, G., Shcherbyna, P. & Moroz D. (2022). Aggregation of computing channels based on the Nvidia Cuda platform for control modes of components of technological systems. *Information Technology: Computer Science, Software Engineering and Cyber Security*. № 2. s. 85–92.

5. Shvachych, G., Moroz, B., Martynenko, A., Hulina, I., Busygin, V. & Moroz, D. (2020). Model of Speed Spheroidization of Metals and Alloys Based on Multiprocessor Computing Complexes. *Machine Learning for Predictive Analysis.* Networks and Systems. Springer. P. 33–41.

6. Shvachych, G., Vozna, N., Ivashchenko, O., Bilyi, O. & Moroz D. (2021). Efficient algorithms for parallelizing tridiagonal systems of equations. *System technologies.* Dnipro. № 5 (136). P. 110–119.

7. I. Mamuzić, G. G. Shvachych & P. O. Ishchuk (2024). Challenges of Parallel Numerical Integration of Pollution Transport Equations. *Book of abstracts of the 17th International Symposium of Croatian Metallurgical Society – SHNM '2004 Materials and metallurgy (published in Metalurgija* 63 (2024) 3-4), P. 477–492.

8. Shvachych, H. H. & Shmukyn A. A. (2004). Osobennosty konstruyrovanyia parallelnыkh vыchyslytelnыkh alhorytmov dlia PЭVM v zadachakh teplo- y massoobmena [Peculiarities of designing parallel computational algorithms for PCs in heat and mass transfer problems]. *Vostochno-evropeiskyi zhurnal peredovыkh tekhnolohyi* [*Eastern European journal of advanced technologies*]. № 3. s. 42–47.

9. Shvachych, G., Vozna, N., Ivashchenko, O., Bilyi, O. & Moroz, D. (2021). Method of lines in distributed problems of experimental data processing. *International Academy Journal Web of Scholar.* № 2(52). P. 1–7.

10. Lienhart, H., Stoots, C. & Becker S. (2002). Flow and Turbulence Structures in the Wake of a Simplified Car Model (Ahmed Model). New Results in Numerical and Experimental Fluid Mechanics III. Vol. 12, p. 323–330.

11. Lavryk, V. I. (2010). Modeliuvannia i prohnozuvannia stanu dovkillia [Modeling and forecasting of the state of the environment]. K. : VD KM Akademiia. 400 s.

12. Bashkov, Ye. O., Ivashchenko, V. P. & Shvachych, H. H. (2011). Vysokoproduktyvna bahatoprotsesorna systema na bazi personalnoho obchysliuvalnoho klastera [A high-performance multiprocessor system based on a personal computing cluster]. *Problemy modeliuvannia ta avtomatyzatsii proektuvannia* [*Problems of modeling and design automation*]. 2. Vyp. 9 (179). S. 312–324.

13. Pat. 61944 Ukraina, MPK C21D 1/26, G06F 15/16 (2011.01). Modul vysokoefektyvnoi bahatoprotsesornoi systemy pidvyshchenoi hotovnosti [Module of a highly efficient multiprocessor system of high availability]/ Ivashchenko, V.P., Bashkov, Ye.O., Shvachych, H.H., Tkach, M.O.; patentovlasnyky Natsionalna metalurhiina akademiia Ukrainy, Donetskyi natsionalnyi tekhnichnyi universytet. № u 2010 09341; zaiavl. 26.07.2010; opubl. 10.03.2011, Biul. № 5.

УДК 004.42:004.6
DOI https://doi.org/10.32782/IT/2024-4-30

Андрій ЯРМОЛАТІЙ
викладач кафедри кібербезпеки, інформаційних технологій та економіки, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе 210, м. Київ, Україна, 02121
ORCID: *0009-0004-8655-9928*

Любов ЧЕРЕМІСІНА
викладач кафедри кібербезпеки, інформаційних технологій та економіки, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе 210, м. Київ, Україна, 02121
ORCID: *0009-0005-0719-0745*

Валентин ГАЛУНЬКО
доктор філософії з галузі права, доцент кафедри адміністративного права, інтелектуальної власності та цивільно-правових дисциплін, Київський університет інтелектуальної власності та права Національного університету «Одеська юридична академія», Харківське шосе, 210, Київ, Україна, 02121
ORCID: *0000-0002-8133-6766*

Бібліографічний опис статті: Ярмолатій, А., Черемісіна, Л., Галунько, В. (2024). Побудова веб-додатків із сучасними системами аутентифікації на базі OAuth 2.0 PKCE та біометрики для підвищення кіберстійкості цих додатків. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 242–251, doi: https://doi.org/10.32782/IT/2024-4-30

ПОБУДОВА ВЕБ-ДОДАТКІВ ІЗ СУЧАСНИМИ СИСТЕМАМИ АУТЕНТИФІКАЦІЇ НА БАЗІ OAUTH 2.0 PKCE ТА БІОМЕТРИКИ ДЛЯ ПІДВИЩЕННЯ КІБЕРСТІЙКОСТІ ЦИХ ДОДАТКІВ

Сучасний світ бізнесу в своїй технічній основі базується на множинній інтеграцій бізнес платформ різного призначення, зокрема веб-додатків та веб-платформ, що поєднує не лише передачу та синхронізацію даних, але і синхронізацію бізнес процесів і процесів прийняття рішень, які цими даними управляються. Відповідно цілісність і кіберзахист таких даних є ключовим аспектом убезпечення від ризиків і створення підґрунтя для прийняття правильних рішень. В умовах стрімкого розвитку цифрових технологій та значного зростання кіберзагроз, питання забезпечення безпеки веб-додатків бізнесу набуває особливої актуальності а подекуди є важливою частиною корпоративної стратегії при виборі або створенні нових веб-систем. Одним із основних інструментів забезпечення безпеки інтеграцій між веб-платформами Протокол OAuth 2.0 з механізмом Proof Key for Code Exchange (PKCE) є одним з основних інструментів для захисту таких інтеграцій.

Мета роботи. *Метою даної роботи є аналітичний розгляд технічних аспектів практичного впровадження протоколу OAuth 2.0 і створення пропозиції щодо його вдосконалення через застосування біометричних технологій алгоритму Proof Key for Code Exchange (PKCE) що особливо важливо розуміти під час створення веб-додатків (веб-програмування). Зокрема запропонований підхід комбінованої біометрії під час етапу делегування користувачем прав до клієнтського додатку.*

Методологія. *У дослідженні застосовано комплексний підхід, що включав аналіз принципів роботи протоколу OAuth 2.0, моделювання потенційних загроз під час делегування користувачем прав клієнтському додатку. Методологія дослідження охоплювала теоретичний аналіз літератури з кібербезпеки та технологій біометричної верифікації, теоретичне моделювання можливих сценаріїв атак та впровадження додаткового шару автентифікації.*

Наукова новизна. *У цій роботі вперше досліджено потенційну вразливість протоколу OAuth 2.0 PKCE до соціального інжинірингу на етапі делегації прав користувачем клієнтському додатку. Запропоновано інноваційний підхід до підсилення цього етапу шляхом впровадження біометричної верифікації, яка базується на унікальному алгоритмі, що забезпечує підвищений рівень захисту від атак з використанням методів соціальної маніпуляції.*

Висновки. *Забезпечення безпеки інтеграцій веб-платформ є критичним аспектом для підприємств у сучасному бізнес-середовищі. Особливо важливим є захист процесу делегування прав користувачем клієнтському додатку, оскільки цей етап є вразливим до атак, зокрема соціального інжинірингу. Запропоно-*

ване вдосконалення протоколу OAuth 2.0 з використанням механізму Proof Key for Code Exchange (PKCE) через інтеграцію біометричної верифікації на етапі делегування прав надає додатковий рівень захисту, що дозволяє знизити ризики, пов'язані з соціальними маніпуляціями. Таким чином, запропонований підхід може бути застосований у практиці веб-програмування для забезпечення більш високого рівня безпеки інтеграцій, де високий рівень кіберстійкості має надзвичайно важливе значення.

Ключові слова: *OAuth 2.0, PKCE, веб-програмування, кібербезпека, аутентифікація, біометричні системи, веб-платформи, кіберстійкість.*

Andrii YARMOLATII
Lecturer at the Department of Cybersecurity, Information Technology and Economics, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121, ayarmolatii@gmail.com
ORCID: *0009-0004-8655-9928*

Liubov CHEREMISINA
Lecturer at the Department of Cybersecurity, Information Technology and Economics, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121, cheremisina1112@gmail.com
ORCID: *0009-0005-0719-0745*

Valentyn HALUNKO
PhD in law, Associate Professor at the Department of Administrative Law, Intellectual Property and Civil-Law Disciplines, Kyiv University of Intellectual Property and Law of the National University «Odessa Law Academy», 210, Kharkivske highway, Kyiv, Ukraine, 02121, valentinvalentin0987@gmail.com
ORCID: *0000-0002-8133-6766*

To cite this article: Yarmolatii, A., Cheremisina, L., Halunko, V. (2024). Pobudova veb-dodatkiv iz suchasnymy systemamy autentyfikatsiyi na bazi OAuth 2.0 PKCE ta biometryky dlya pidvyshchennya kiberstijkosti tsikh dodatkiv [Building web applications with modern authentication systems based on oauth 2.0 pkce and biometrics to enhance the cyber resilience of these applications]. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 242–251, doi: https://doi.org/10.32782/IT/2024-4-30

BUILDING WEB APPLICATIONS WITH MODERN AUTHENTICATION SYSTEMS BASED ON OAUTH 2.0 PKCE AND BIOMETRICS TO ENHANCE THE CYBER RESILIENCE OF THESE APPLICATIONS

The modern business world is technically founded on the integration of various business platforms, including web applications and web platforms. This integration encompasses not only data transfer and synchronization but also the synchronization of business processes and decision-making processes governed by these data. Therefore, the integrity and cybersecurity of such data are key aspects in mitigating risks and laying the groundwork for making sound decisions. Amid the rapid development of digital technologies and the significant rise in cyber threats, ensuring the security of business web applications has become particularly relevant and is often an essential part of corporate strategy when selecting or creating new web systems. One of the primary tools for securing integrations between web platforms is the OAuth 2.0 protocol with the Proof Key for Code Exchange (PKCE) mechanism.

Objective. *The aim of this work is to provide an analytical review of the technical aspects of the practical implementation of the OAuth 2.0 protocol and to create a proposal for its improvement through the application of biometric technologies using the Proof Key for Code Exchange (PKCE) algorithm, which is particularly important to understand when developing web applications (web programming). Specifically, the proposed approach involves combining biometrics during the user delegation phase of granting rights to a client application.*

Methodology. *The study applies a comprehensive approach, which includes an analysis of the principles of OAuth 2.0 protocol operation, modeling potential threats during the user's delegation of rights to the client application. The research methodology encompassed a theoretical analysis of literature on cybersecurity and biometric verification technologies, theoretical modeling of possible attack scenarios, and the implementation of an additional authentication layer.*

Scientific novelty. *This work investigates for the first time the potential vulnerability of the OAuth 2.0 PKCE protocol to social engineering during the delegation of user rights to a client application. An innovative approach is proposed to strengthen this stage by implementing biometric verification based on a unique algorithm that provides an enhanced level of protection against attacks using social manipulation techniques.*

Conclusions. *Ensuring the security of web platform integrations is a critical aspect for businesses in the modern business environment. The protection of the user rights delegation process to the client application is particularly important, as this stage is vulnerable to attacks, including social engineering. The proposed improvement of the OAuth 2.0 protocol with the use of the Proof Key for Code Exchange (PKCE) mechanism through the integration of biometric verification at the rights delegation stage provides an additional layer of protection, reducing risks associated with social manipulation. Therefore, the proposed approach can be applied in web programming practices to ensure a higher level of integration security, where a high level of cyber resilience is of utmost importance.*

Key words: *OAuth 2.0, PKCE, web programming, cybersecurity, authentication, biometric systems, web platforms, cyber resilience.*

Актуальність проблеми. У сучасному бізнес-середовищі інтеграція різних веб-платформ і сторонніх спеціалізованих веб-додатків є необхідною умовою для досягнення ефективності та гнучкості в управлінні інформацією. Згідно дослідницької платформи Gartner Magic Quadrant серед сучасних лідерів і найпоширеніших веб-платформ на ентерпрайз рівні що створюють можливість розробляти веб-додатки і інтегруватися з ними є такі веб-платформи як ServiceNow і Salesforce. Обидва ці інструменти дозволяють організаціям швидко автоматизувати бізнес-процеси та забезпечити ефективну взаємодію між різними компонентами корпоративних IT-систем. Ці платформи активно використовуються для управління IT-сервісами, взаєминами з клієнтами, підтримки співробітників та іншими критичними завданнями організацій, включаючи навіть операції кібербезпеки.

Проте ефективний і стійкий до атак обмін даними через створення інтеграцій веб-додатків з такими платформами потребує надійного механізму аутентифікації та авторизації, що одночасно забезпечував би і захист від несанкціонованого доступу і зменшення ризику викрадення зловмисниками атрибутів аутентифікації (паролі, токени тощо) що зазвичай передаються незахищеними каналами зв'язку або зберігаються в незахищених додатках клієнтської сторони інтеграції. Для забезпечення такого захисту сьогодні широко використовується протокол OAuth 2.0, спеціально підсилений механізмом PKCE (Proof Key for Code Exchange), що значно підвищує рівень безпеки особливо при інтеграціях з відкритими API та незахищеними додатками але має потенційні слабкі місця в алгоритмі, які потребують підсилення в сферах критично важливого надвисокого рівня кіберстійкості.

Аналіз останніх досліджень і публікацій. Серед останніх досліджень (Бодак, Дорошенко, 2022) досить детально і послідовно розглядаються протокол OAuth 2.0 і алгоритм PKCE відносно відкритих клієнтів в контексті проблеми авторизації таких клієнтів, приводиться набір потенційних атак і пропонується рішення через використання моделі BFF (Backend For Frontend). З іншої сторони в цій та інших (Білодід, 2024, с. 76–90) роботах мало уваги приділяється передумовам і причинам створення підсиленого флоу OAuth 2.0 PKCE, адже розуміння необхідності застосування того чи іншого механізму аутентифікації при побудові захищеної інеграції ефективної передачі даних між системами(платформами) базується на порівнянні знань про попередні або альтернативно існуючі механізми. Детальний розгляд таких механізмів не є фокусом даної статті, але слід зауважити, що більшість попередньо існуючих механізмів (серед них Базова Аутентифікація – Basic Authentication, використання API Key в заголовку надісланих пакетів інформації тощо) виконувались в один крок і вимагали від користувача ділитися своїми паролями зі сторонніми додатками та веб-сервісами для отримання доступу до їх облікових даних, а оскільки можливості 100% гарантувати, що при передачі паролей інформація не буде отримана і використана зловмисником, виникла необхідність в новому підході до алгоритму аутентифікації, що і призвело до появи OAuth 2.0.

Метою даної статті є аналіз алгоритму роботи OAuth2.0 та PKCE і потенційно вразливих моментів та пропозиція по їх підсиленню при веб програмуванні додатків та платформ для отримання високого рівня їх кіберстійкості.

Архітектура веб-додатків і веб-додатків і платформ. На початку слід згадати загальну структуру більшості веб-додатків яка має клієнт-серверну архітектуру (рис. 1):

Як вказано на рис. 1. – весь додаток – це поєднання(інтеграція) окремих систем і механізмів як на клієнтській та і на серверній частині, але серверна частина завжди складніша і має бути потужніша, адже своїм основним призначенням має обробку великої кількості запитів від множини клієнтських додатків та процессінг даних з бази даних.

Якщо серверну частину умовно можна вважати більш захищеною по причині віддаленості від зловмисника і складності фізичного доступу, то клієнтська частина і сама взаємодія між сервером і клієнтом потребує створення додаткового захисту – кіберстійкої системи аутентифікації.

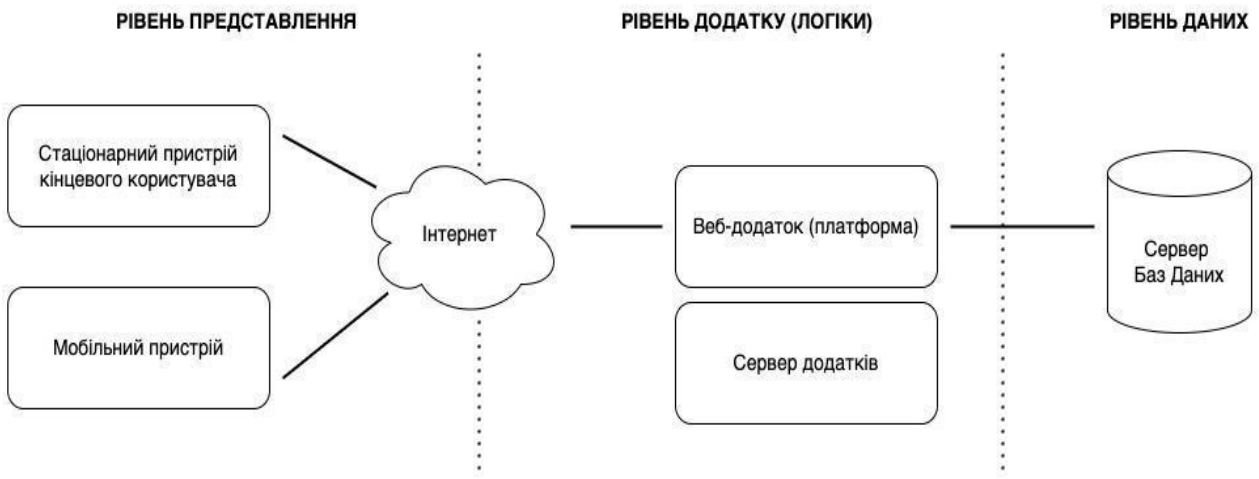


Рис. 1. Загальна схема веб-додатків

При цьому підхід до створення веб-додатків в веб-програмуванні може різнитися в залежності від багатьох факторів, в тому числі від середовища і предметної області для яких цей додаток розробляється. Веб-програмування додатків до вже існуючої платформи значно спрощений, але розробка веб-додатку так само часто може розпочинатися з нуля і вимагатиме більше часу на вирішення нетривіальної задачі правильного вибору компонент. Проте оскільки більшість веб-додатків пишуться розробниками на сьогоднішній день на основі JavaScript та Node.js, це автоматично звужує набір варіантів та задає шаблон розробки – використання бібліотек і модулів, які значною мірою скорочують час розробки.

На сьогоднішній день в веб-програмуванні серед модулів і бібліотек JavaScript/Node.js для реалізації системи аутентифікації OAuth 2.0 найчастіше використовуються: passport, passport-oauth2, simple-oauth2, express-oauth-server, express-oauth2-bearer, openid-client, oidc-provider, jsonwebtoken тощо. Серед них найчастіше використовують passport, passport-oauth2 та simple-oauth2. Їх широке застосування і популярність супроводжуються підтримкою глобального комьюніті, а серед переваг визначають простоту використання базових сценаріїв і широкий спектр інтеграційних можливостей з іншими системами.

Передумови OAuth 2.0 та Implicit Flow.

На рис. 2. продемонстрована класична та історично перша архітектура інтеграції двох систем з використанням паролю (або API Key) для аутентифікації. Літерами (а) та (б) вказано небезпечні елементи такої архітектури, де а) не захищений канал передачі аутентифікаційних даних – email, sms тощо та б) незахищений додаток, до якого може отримати доступ зловмисник і рано чи пізно отримати паролі що там зберігаються. Звісно, що в кожному окремому випадку величина такої ймовірності різна, але проблемою залишається наявність такої можливості для зловмисника, тому використовувати такий підхід в критично важливих процесах було б дуже ризиковано.

Враховуючи вказані вище ризики, OAuth 2.0 був розроблений як стандарт для так званої делегованої авторизації, що дозволяє стороннім додаткам отримувати обмежений доступ до ресурсів користувача без необхідності передавати свої облікові дані. Тепер, замість однокрокового алгоритму(очевидний флоу) аутентифікації маємо протокол, який передбачає і додатковий елемент безпеки (крок) і використання так званих «токенів доступу», що генеруються сервером авторизації і використовуються клієнтами для доступу до захищених ресурсів (неявний потік – Implicit Flow). Ключовими елементами протоколу OAuth 2.0 є наступні:

1. Сервер авторизації (Authorization Server) – сервер, який відповідає за аутентифікацію користувачів і видачу токенів доступу
2. Сервер ресурсів (Resource Server) – сервер, що зберігає захищені ресурси (наприклад, дані користувачів або іншу необхідну інформацію), і перевіряє токени доступу
3. Клієнт (Client) або клієнтський додаток – додаток, який звертається до ресурсів користувача за допомогою отриманого токена доступу
4. Тип Авторизації (Authorization Grant) – тип авторизації, який застосовує OAuth для видачі коду доступу.

Розглянемо класичний спенарій роботи OAuth 2.0, що зображено на Рис. 3

Як видно на рис. 3., клієнтський додаток перед запитом до серверу ресурсів спочатку

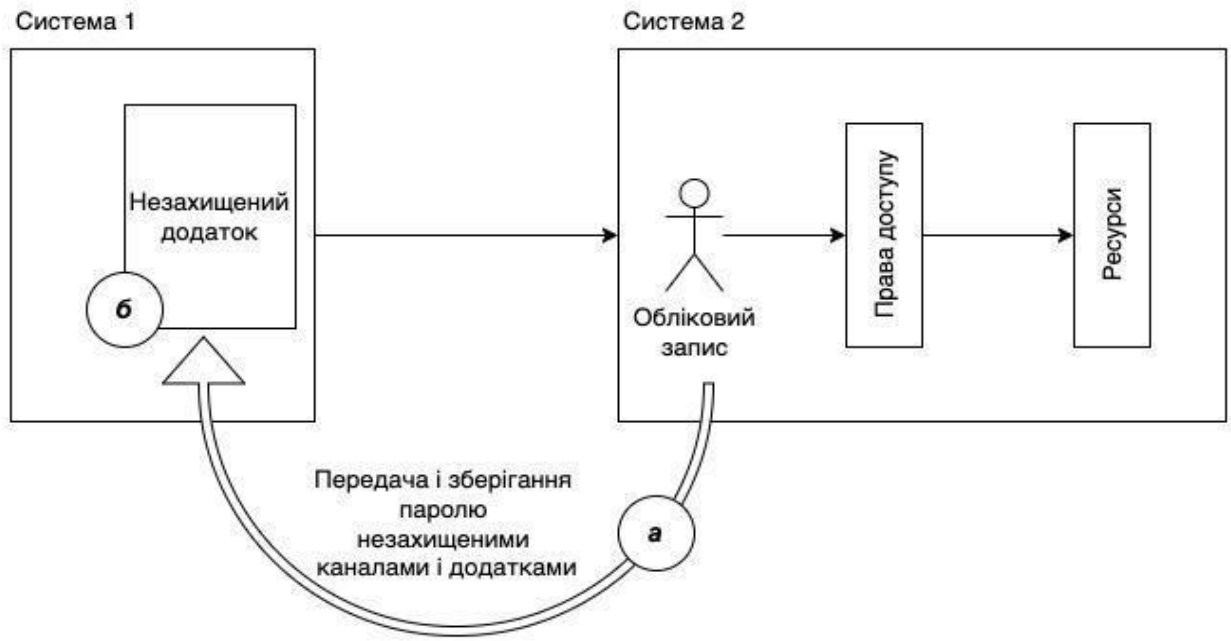


Рис. 2. Передача і зберігання паролей незахищеними каналами та додатками

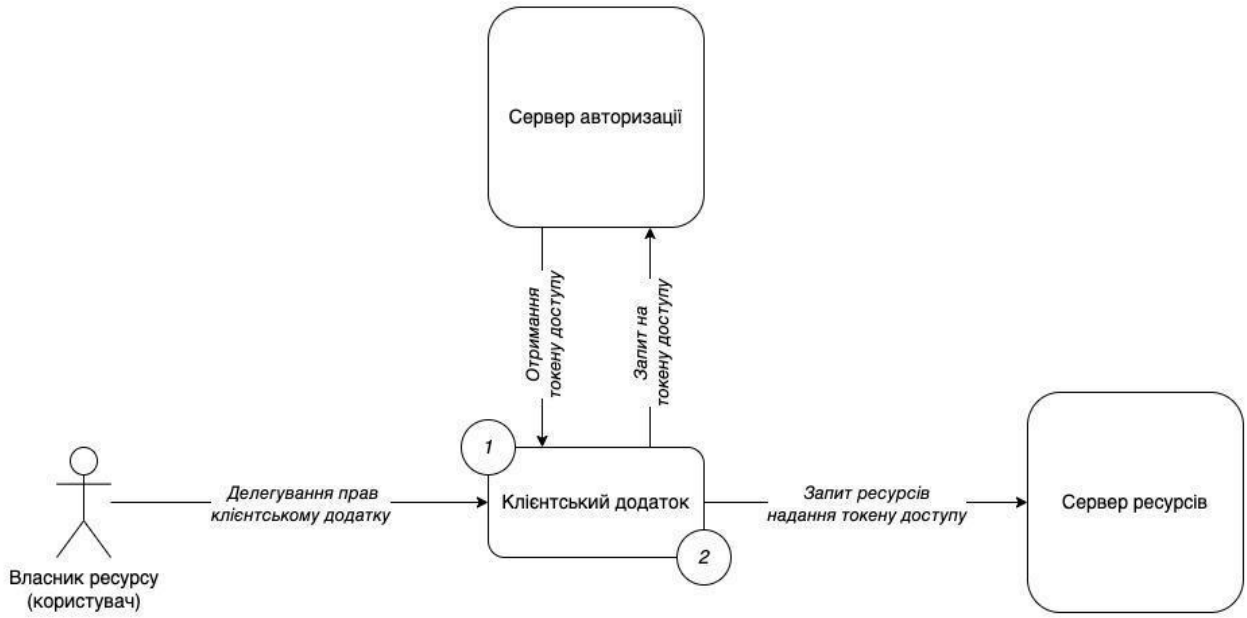


Рис. 3. Класичний сценарій OAuth 2.0

взаємодіє з авторизаційним сервером наступним чином:

1. Отримання коду авторизації: клієнт перенаправляється на авторизаційний сервер, де користувач надає дозвіл на доступ до своїх ресурсів.
2. Обмін коду на токен доступу: після отримання коду клієнт відправляє його на сервер для отримання токена доступу.

Варто зауважити, що все частіше сервер авторизації і сервер ресурсів є технічно двома компонентами однією і тієї ж системи (платформи) і навіть більше того – веб-додаток чи платформа мають універсальні механізми, що дозволяють їм вести себе і як сервер і як клієнтський додаток. Так, наприклад, в веб-платформі ServiceNow зберігаються і інформація як ресурс і налаштування аутентифікації та авторизації. Для цього достатньо зайти в аплікейшн меню System OAuth та обравши пункт Application Registry (реєстр додатків), перейти до конфігураційної форми (рис. 4.)

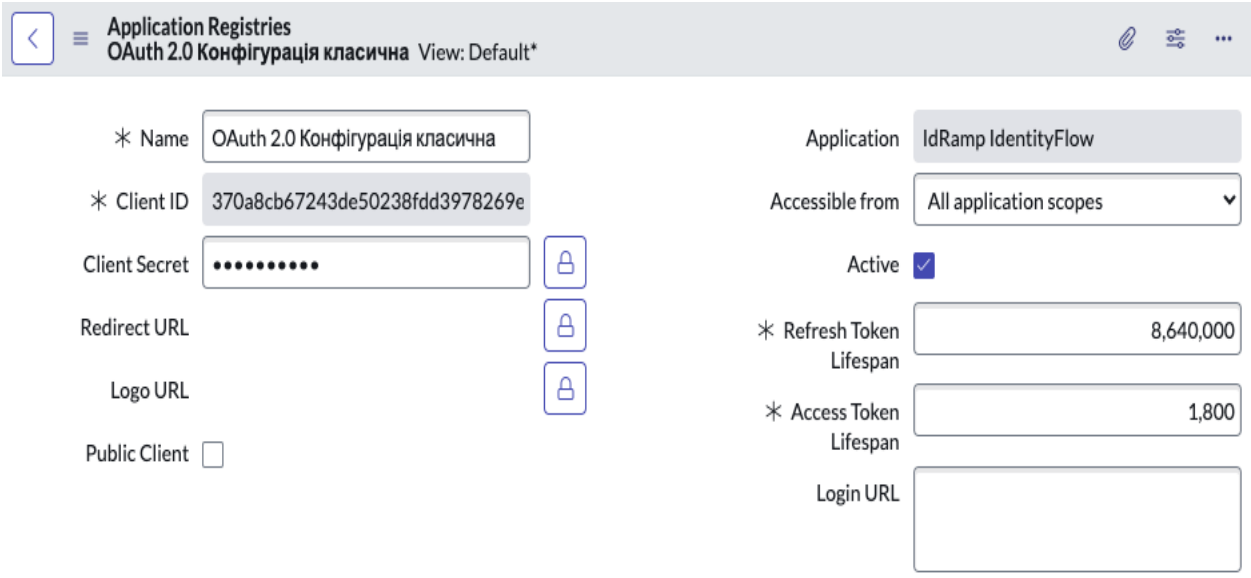


Рис. 4. Конфігурація OAuth 2.0 аутентифікації на веб-платформі ServiceNow

Все, що необхідно зробити для конфігурації серверу аутентифікації на даному етапі – просто додати Redirect URL, адже параметри Client ID та Client Secret генеруються автоматично системою. Серверна частина сконфігурована і можнавикористовувати клієнтський додаток (наприклад Postman) для тестування. За схожим принципом створюється конфігурація OAuth-серверу і в веб-платформі Salesforce, де спочатку необхідно перейти до менеджеру додатків (App Manager) а потім створити нову сутність підключаємого додатку (Connected App). В обох платформах відбувається збереження інформації про підключаємий веб-додаток.

Підсилення аутентифікації OAuth 2.0 механізмом PKCE. При аналізі форми конфігурації OAuth платформи ServiceNow (рис. 4) чекбокс Public Client було навмисне не використано адже розглядався класичний сценарій OAuth 2.0. І як вже було згадано раніше (Бодак, Дорошенко, 2022) – такий механізм має ряд вразливостей при роботі з публічними (відкритими, незахищеними) клієнтськими додаткам, що в свою чергу може понизити рівень кіберстійкості такої системи. На Рис. 5 продемонстрована схема атаки перехоплення коду авторизації:

Як видно на рис. 5., серед можливих вразливостей OAuth 2.0 існує можливість атаки на перехоплення коду авторизації додатком зловмисника на пристрої кінцевого користувача. Інші можливі атаки були розглянуті раніше (Бодак, Дорошенко, 2022). Така ситуація пояснюється архітектурою взаємодії з сервером авторизації, де не сам додаток, а операційна система чи браузер комунікують з сервером для отримання коду авторизації і далі передають його додатку, про існування якого сервер авторизації не знає, тобто немає безпосередньої привязки а значить і контролю до клієнтського додатку. Це в свою чергу призвело до необхідності покращення алгоритму захисту OAuth 2.0 що і призвело до створення механізму PKCE (Proof Key for Code Exchange), який був вперше представлений у жовтні 2014 року як частина RFC 7636, доповнення до OAuth 2.0. PKCE був створений для вирішення проблем безпеки, пов'язаних з аутентифікацією в мобільних додатках та інших клієнтах, які не можуть безпечно зберігати конфіденційні дані, як-от клієнтські секрети.

Серед причин виникнення PKCE виділимо наступні:

- ***Відсутність клієнтського секрету***: У мобільних додатках і браузерних клієнтах немає безпечного способу зберігати клієнтський секрет. PKCE усуває необхідність у клієнтському секреті для захисту процесу авторизації.
- ***Захист від атаки підміни коду (Code Injection)***: Без PKCE зловмисник може перехопити код авторизації та обміняти його на токен доступу, якщо у них є можливість перехопити трафік між клієнтом і сервером. PKCE додає додатковий рівень безпеки, використовуючи динамічно згенерований секрет для підтвердження того, що запит на отримання токена доступу походить від того ж клієнта, що й запит на авторизацію.

Рис. 5. Перехоплення Коду Авторизації

● ***Поліпшення безпеки***: PKCE покращує безпеку OAuth 2.0, забезпечуючи, що тільки клієнт, який ініціював авторизацію, може обміняти код на токен доступу. Це досягається шляхом додавання додаткового параметра code_verifier і його хеш-версії code_challenge до запиту авторизації.

Алгоритм роботи PKCE. На рис. 6 показана загальна схема роботи PKCE алгоритму в контексті OAuth 2.0 аутентифікації.

1. Клієнт(клієнтський додаток) генерує випадковий код перевірки (code_verifier) і обчислює хеш (code_challenge).

2. Клієнт направляє запит на авторизацію разом з code_challenge.

3. Користувач надає дозвіл на доступ до своїх даних.

4. Клієнт отримує код авторизації.

5. Клієнт відправляє код авторизації разом з code_verifier для отримання токена доступу.

6. Авторизаційний сервер перевіряє відповідність code_challenge і code_verifier, і якщо вони співпадають, видає токен доступу.

Таким чином, навіть якщо зловмисник отримає код авторизації, він не зможе використати його для отримання токена без доступу до первісного коду перевірки. Хоча цей механізм і підвищує кіберстійкість OAuth 2.0, особливо вразливим є третій крок, де користувач надає дозвіл на доступ до своїх даних. Адже на практиці надання доступу відбувається наступним чином – користувачу виводиться або діалогове вікно або сторінка, на якій в кращому випадку необхідно ввести логін і пароль користувача а не рідко може бути просто кнопка “Дозволити” без особливої верифікації персони, яка надає доступ, що робить алгоритм вразливим для соціального інжинірингу та потенційних маніпуляцій користувачем.

Удосконалення OAuth 2.0 за допомогою PKCE. Розглянуту вище слабку точку алгоритму PKCE можна підсилити різними методами, але розглянемо три найперспективніші – використання біометричної технології, додаткова верифікація власника ресурсів (користувача) та комбінований підхід.

1. Використання біометричної технології. Біометричні технології, такі як сканування відбитків пальців, розпізнавання обличчя або голосу, можуть бути інтегровані як додатковий рівень автентифікації перед наданням доступу до даних. Це значно ускладнить можливість несанкціонованого доступу, оскільки зловмиснику буде потрібно не лише код авторизації, але й фізичний доступ до біометричних даних користувача.

2. Додаткова верифікація власника ресурсів. Це може включати відправлення одноразового

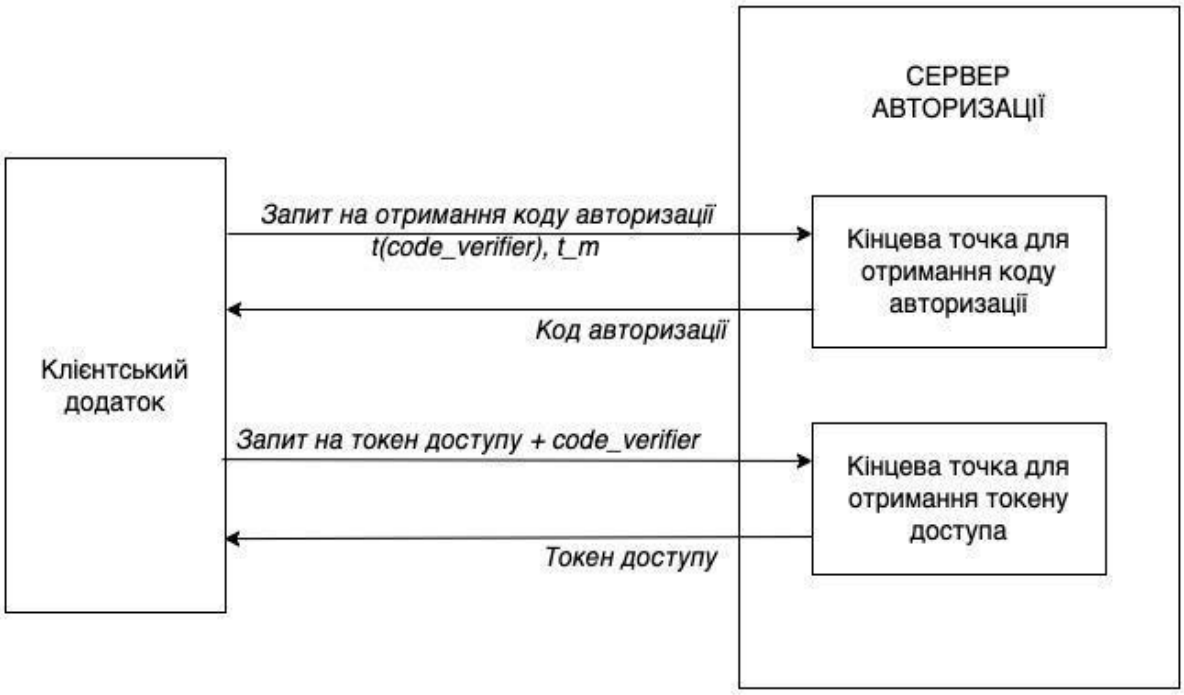


Рис. 6. Загальна схема роботи OAuth 2.0 PKCE

пароля (OTP) на зареєстрований номер телефону або електронну пошту користувача для підтвердження дії. Такий підхід забезпечує додатковий рівень захисту, оскільки користувач має підтвердити своє бажання надати доступ через окремий канал зв'язку.

3. Комбінований підхід. Комбінований підхід передбачає використання біометричної верифікації, яка здійснюється на сторонньому сервісі. Це забезпечує додатковий рівень безпеки, оскільки верифікація проходить поза основною системою, зменшуючи ризик компрометації локальних даних. Таким чином, навіть якщо зловмисник отримає доступ до локальних даних користувача, він не зможе пройти біометричну перевірку на сторонньому сервісі. Комбінований підхід рекомендовано до реалізації де потрібно створити особливо високий рівень кіберстійкості, адже комплексна система захисту завжди використовує поєднання засобів і заходів а не окремий інструмент.

Розглянемо схему комбінованої верифікації користувача OAuth 2.0 PKCE

Як видно на рис. 7, делегування доступу клієнтському додатку (клієнту) супроводжується додатковою верифікацією зі стороннім сервером верифікації та використанням біометричних технологій. В такій схемі, перед тим як делегується доступ, користувачу буде надіслане посилання (через sms або email) з посиланням для проходження верифікації. Користувач повинен довести, що це він делегує доступ а не хтось за його пристроєм. Тому перейшовши за посиланням, користувач ініціює процедуру верифікації і при цьому застосовує біометрику – або сканування відбитку пальця, або сканування обличчя тощо. Якщо через біометричні атрибути користувач довів що це він, то сервер верифікації дає сигнал дозволу на делегування. Окрім цього, слід застосувати комбінацію біометричних перевірок наступним алгоритмом: якщо користувач попередньо не пройшов, наприклад, біометричну верифікацію відбитком пальця, то наступна верифікація буде вже скануванням обличчя або комбінацією обох щоб мінімізувати ризик соціального інжинірингу або дати можливість зловмиснику на повторний однотипний маневр.

Інформування клієнта (клієнтського додатку) про завершення і результати верифікації є нетривіальною задачею, адже користувач не повинен нічого додатково робити в кінці процесу верифікації і при цьому і сервер верифікації і клієнт мають отримати відповідні дані саме стосовно цього користувача. Це досягається за допомогою загальнодоступного веб-сервісу і методу GET протоколу HTTP, що на практиці означає звичайне переведення користувача на іншу сторінку і унікальними параметрами, що передаються в адресному рядку. Така сторінка містить інформацію про завершення верифікації і інформує користувача, що той може закрити вкладку браузера. Так, загальнодоступна реалізація API результатів верифікації може мати наступну послідовність:

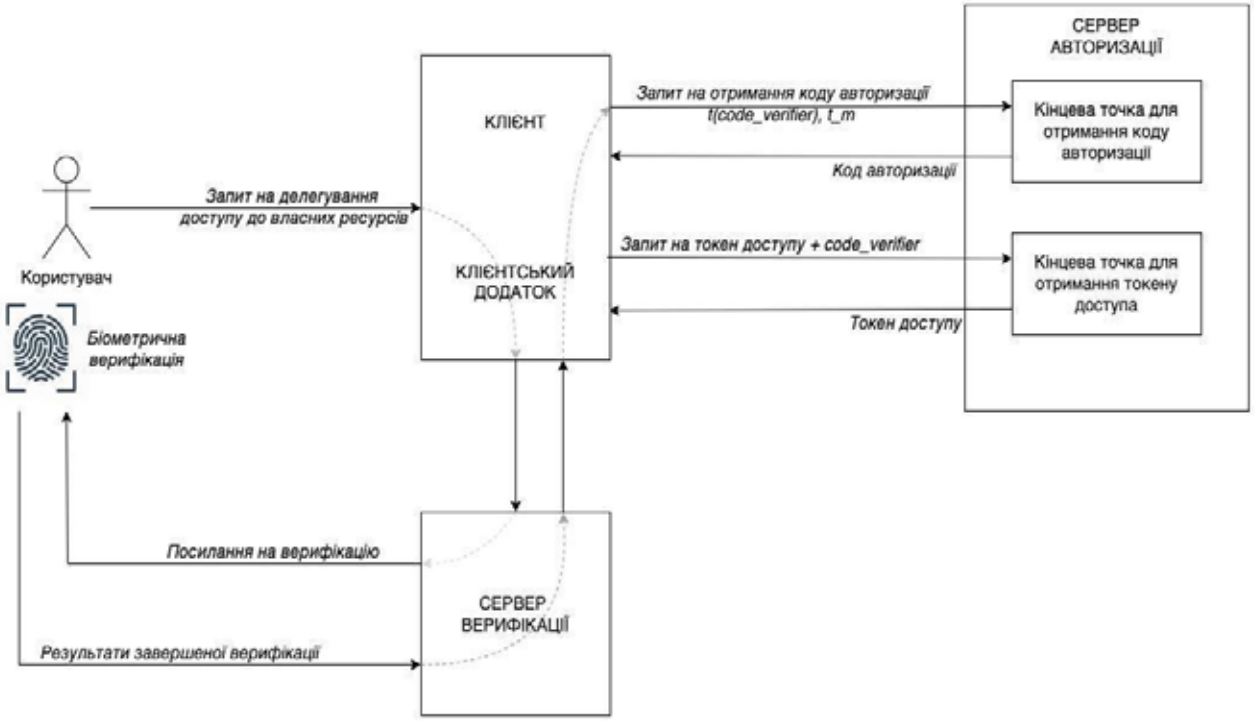


Рис. 7. Архутектура підсилення етапу делегації послідовності OAuth 2.0 PKCE

- Отримання інформації про користувача з URL
- Надсилання запиту до зовнішнього сервера верифікації
- Оновлення статусу перевірки на основі відповіді від сервера
- Парсинг результату з відповіді

Таким чином, запропонована архітектура підсилення PKCE біометричною верифікацію створює новий, більший високий, рівень кіберстійкості веб-додатків чи платформ де він буде застосований.

Висновки. OAuth 2.0 з PKCE є потужним інструментом для створення безпечних інтеграцій між сучасними бізнес веб-платформами, такими як ServiceNow та Salesforce, а також у веб-додатках, де неможливо гарантувати захист клієнтського секрету на пристрої користувача. Ця конфігурація протоколу дозволяє ефективно захищати дані при їх обміні між системами, знижуючи ризик несанкціонованого доступу. Впровадження PKCE у таких інтеграціях підвищує рівень безпеки, гарантуючи цілісність та конфіденційність інформації. Однак PKCE має потенційну вразливість, зокрема в процесі делегації прав доступу, де існує ймовірність соціального інжинірингу. Тому в середовищах, де вимагається високий рівень кіберстійкості, рекомендовано доповнювати цей процес верифікацією користувача з використанням біометричних технологій.

ЛІТЕРАТУРА:

1. The OAuth 2.0 Authorization Framework. Microsoft Internet Engineering Task Force (IETF). URL: https://datatracker.ietf.org/doc/html/rfc6749 (дата звернення 11.10.2024).

2. Proof Key for Code Exchange by OAuth Public Clients. URL: https://datatracker.ietf.org/doc/html/rfc7636 (дата звернення 11.10.2024)

3. Бодак В. В., Дорошенко А. Ю., Захист відкритих клієнтів за допомогою одного алгоритму авторизації. *Проблеми програмування*. 2022. № 3-4. С. 409–416.

4. Радівілова Т., Кіріченко Л., Пантєлєєв В., Мазепа А., Білодід В. «Аналіз методів автентифікації для вебзастосунків та реалізація вебзастосунку з інтегрованою системою автентифікації», СУЧАСНИЙ СТАН НАУКОВИХ ДОСЛІДЖЕНЬ ТА ТЕХНОЛОГІЙ В ПРОМИСЛОВОСТІ, 2024. (3(29), с. 76–90. doi: 10.30837/2522-9818.2024.3.076.

5. Authgear Team. (2024). PKCE in OAuth 2.0: How to Protect Your API from Authorization Code Grant Attacks. *Authgear Blog*. https://doi.org/10.48550/arXiv.2412.07012

6. Authgear Team. (2024). PKCE in OAuth 2.0: How to Protect Your API from Attacks. *Authgear Blog*. https://doi.org/10.48550/arXiv.2412.07012

7. Passport.js Team. (2020). PKCE Support for OAuth 2.0. *Medium*. https://medium.com/passportjs/pkce-support-for-oauth-2-0-e3a77013b278 (дата звернення 14.10.2024)

8. Identity Beyond Borders. (2020). What the heck is PKCE? *Medium*. https://medium.com/identity-beyond-borders/what-the-heck-is-pkce-40662e801a76

REFERENCES:

1. The OAuth 2.0 Authorization Framework. Microsoft Internet Engineering Task Force (IETF). Retrieved from: https://datatracker.ietf.org/doc/html/rfc6749

2. *PKCE Proof Key for Code Exchange by OAuth Public Clients*. Retrieved from: https://datatracker.ietf.org/doc/html/rfc7636

3. Bodak, V. V., Doroshenko, A. Yu. (2022). Zakhyst vidkrytykh kliientiv za dopomohoyu odnoho alhorytmu avtoryzatsii [Protection of open clients using a single authorization algorithm]. *Problemy prohramuvannya*. 3-4, 409–416. [in Ukrainian]

4. Radivilova, T., Kirichenko, L., Pantelieiev, V., Mazepa, A., Bilodid, V. (2024). «Analiz metodiv avtentifikatsii dlia vebzastosunkiv ta realizatsiia vebzastosunku z intehrovanoiu systemoiu avtentifikatsii» [Analysis of Authentication Methods for Web Applications and Implementation of a Web Application with an Integrated Authentication System] SUCHASNYI STAN NAUKOVYKH DOSLIDZHEN TA TEKHNOLOHII V PROMYSLOVOSTI, (3(29), s. 76–90. doi: 10.30837/2522-9818.2024.3.076 [in Ukrainian]

5. Authgear Team. (2024). *PKCE in OAuth 2.0: How to Protect Your API from Authorization Code Grant Attacks*. Authgear Blog. https://doi.org/10.48550/arXiv.2412.07012

6. Authgear Team. (2024). *PKCE in OAuth 2.0: How to Protect Your API from Attacks*. Authgear Blog. https://doi.org/10.48550/arXiv.2412.07012

7. Passport.js Team. (2020). *PKCE Support for OAuth 2.0*. Medium. Retrieved from: https://medium.com/passportjs/pkce-support-for-oauth-2-0-e3a77013b278

8. Identity Beyond Borders. (2020). *What the heck is PKCE?* Medium. Retrieved from: https://medium.com/identity-beyond-borders/what-the-heck-is-pkce-40662e801a76

ЗМІСТ

Валентина АБЛАМСЬКА, Наталія ДЯЧЕНКО, Валентин ГАЛУНЬКО
МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ВРАЗЛИВОСТІ ІНФОРМАЦІЙНОЇ СИСТЕМИ
ТА ПРОГНОЗУВАННЯ МОЖЛИВОСТЕЙ ЇЇ ЗАХИСТУ 3

Катерина АНТІПОВА, Ігор КАНДИБА, Світлана БОРОВЛЬОВА, Віктор РАЛЕНКО
ВИЯВЛЕННЯ ПЛАГІАТУ В ТЕКСТІ,
ЗГЕНЕРОВАНОГО ВЕЛИКИМИ МОВНИМИ МОДЕЛЯМИ 9

Михайло БЕРДНИК, Ігор СТАРОДУБСЬКИЙ
ВИКОРИСТАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ АДАПТАЦІЇ ПРОГРАМНОГО
ЗАБЕЗПЕЧЕННЯ ДО РІЗНИХ ОБЧИСЛЮВАЛЬНИХ ПЛАТФОРМ 16

Володимир БОГОМ'Я, Любов ЧЕРЕМІСІНА, Андрій ЯРМОЛАТІЙ
БЕЗПЕЧНА ОБРОБКА ДАНИХ В БІЗНЕС-ІНТЕРНЕТ ПЛАТФОРМАХ
ТА СИСТЕМАХ ЕЛЕКТРОННОЇ КОМЕРЦІЇ 29

Володимир БОГОМ'Я, Валентин ГАЛУНЬКО
ПРАВОВЕ РЕГУЛЮВАННЯ КІБЕРБЕЗПЕКИ В КОНТЕКСТІ ЗАХИСТУ
КРИТИЧНОЇ ІНФРАСТРУКТУРИ 35

Андрій БОРИСЕНКО
БЕЗПЕЧНЕ ПРОГРАМУВАННЯ: АВТОМАТИЗОВАНІ ІНСТРУМЕНТИ ДЛЯ ВИЯВЛЕННЯ
ВРАЗЛИВОСТЕЙ В КОДІ 43

Станислав ВЕДМЕДЄВ
РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ЗБИРАННЯ ФЕНОТИПІЧНИХ ДАНИХ
НАСІННЯ СОНЯШНИКУ 53

Vyacheslav GOREV, Yana SHEDLOVSKA, Ivan LAKTIONOV, Grygorii DIACHENKO
LMS PREDICTION AND KALMAN FILTERING FOR THE PROBABILITY
OF THE MAIZE DISEASE OCCURENCE 61

Serhii DOROZHYNSKYI, Vitaliy TUPKALO, Yuriy SHCHERBYNA
IMPORTANT ISSUES OF DATA PROCESSING AND PROTECTION
IN DISTRIBUTED SYSTEMS 68

Валентин ДЯЧЕНКО, Наталія ДЯЧЕНКО
ВРАХУВАННЯ РИЗИКІВ ПРИ ВИКОРИСТАННІ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ 75

Тімур ЖЕЛДАК, Олександр ВЛАДИКО
ВИКОРИСТАННЯ МАШИННОГО НАВЧАННЯ ДЛЯ КЛАСТЕРИЗАЦІЇ З ВИКОРИСТАННЯМ
ІНДЕКСУ НЕБЕЗПЕКИ ХВОСТОСХОВИЩ ТА ЇХ ВПЛИВУ
НА НАВКОЛИШНЄ СЕРЕДОВИЩЕ 81

Уляна ЖОРНОКУЙ, Наталія КОВАЛЕНКО
АНГЛОМОВНІ ЗАПОЗИЧЕННЯ У СФЕРІ КІБЕРБЕЗПЕКИ НА СУЧАСНОМУ ЕТАПІ
РОЗВИТКУ УКРАЇНСЬКОЇ МОВИ 92

Денис ІВАНОВ, Віта КАШТАН
ГЕОІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ВИЯВЛЕННЯ ТА МОДЕЛЮВАННЯ ЗАТОПЛЕНИХ
ЗОН МІСТА ДУБАЙ НА ОСНОВІ ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ 99

Антоніна КАШТАЛЬЯН
ХАРАКТЕРНІ ВЛАСТИВОСТІ ЦЕНТРАЛІЗАЦІЇ В АРХІТЕКТУРІ МУЛЬТИКОМП'ЮТЕРНИХ СИСТЕМ АНТИВІРУСНИХ КОМБІНОВАНИХ ПРИМАНОК І ПАСТОК.....114

Ольга КИВЛЮК
ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ЗАСАДИ ІНФОРМАЦІЙНОЇ ТА КІБЕРНЕТИЧНОЇ БЕЗПЕКИ В УМОВАХ ЦИФРОВІЗАЦІЇ СУСПІЛЬНИХ ПРОЦЕСІВ.....125

Павло КУНДЕНКО, Олександр СОКОЛ, Наталія КАЛІНІНА, Вікторія ГНАТУШЕНКО
НЕЙРОМЕРЕЖЕВИЙ ПІДХІД ВИЯВЛЕННЯ ВОДНИХ ОБ'ЄКТІВ НА МУЛЬТИСПЕКТРАЛЬНИХ ЗОБРАЖЕННЯХ.....132

Марина МОЛЧАНОВА
НЕЙРОМЕРЕЖЕВЕ ВИЗНАЧЕННЯ ЦІЛЕЙ ПРОПАГАНДИ У ТЕКСТОВОМУ КОНТЕНТІ З ВІЗУАЛЬНОЮ АНАЛІТИКОЮ РЕЗУЛЬТАТІВ.....144

Igor NOVYTSKYI, Yuliia SHEVCHENKO, Maksym TRYPUTEN
ADAPTIVE CONTROL SYSTEM OF THE SELF-GRINDING PROCESS OF ORES IN AUTOGENOUS MILLS.....151

Віктор ОЛЕВСЬКИЙ, Ірина УДОВИК, Дмитро ГРИЩАК, Юлія ОЛЕВСЬКА, Юлія ЖУРАВЛЬОВА
МЕТОДОЛОГІЧНІ ОСНОВИ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ АНАЛІЗУ БАГАТОКАНАЛЬНИХ АЕРОКОСМІЧНИХ ЗОБРАЖЕНЬ.....159

Сергій ПРИХОДЧЕНКО, Лариса КОРЯШКІНА, Ольга ШЕВЦОВА, Євгеній ЗАГИНАЙЛО
КОНЦЕПТУАЛЬНА МОДЕЛЬ АРХІТЕКТУРИ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ОБРОБКИ БАГАТОДЖЕРЕЛЬНИХ АКУСТИЧНИХ СИГНАЛІВ.....169

Павло РЕГІДА, Олег САВЕНКО
МЕТОД ВИЯВЛЕННЯ ЗЛОВМИСНОЇ АКТИВНОСТІ В ІНФІКОВАНИХ ПРОГРАМАХ.....178

Oleksii SERHIEIEV
APPLYING THE SYSTEMS APPROACH TO THE ANALYSIS OF REGIONAL-LEVEL MEDICAL LOGISTICS.....187

Олена СОБКО
НЕЙРОМЕРЕЖЕВИЙ ПОШУК І КЛАСИФІКАЦІЯ КІБЕРЗАЛЯКУВАНЬ У ТЕКСТОВИХ ПОВІДОМЛЕННЯХ.....197

Віталій ТУПКАЛО
БЛОКОВО-КОМПОЗИЦІЙНИЙ МЕТОД ФОРМУВАННЯ КОНТРОЛЬНИХ СИГНАТУР ФУНКЦІОНУВАННЯ ЦИФРОВИХ СИСТЕМ НА ОСНОВІ ВЛАСТИВОСТЕЙ СИГНАТУРНОГО ПОЛЯ $TSF[2^N,P^M(X)]$.....206

Любов ЧЕРЕМІСІНА, Наталія МАКСИМОВИЧ
УПРАВЛІННЯ СИСТЕМОЮ ОХОРОНИ ПРАЦІ ТА БЕЗПЕКОЮ ЖИТТЄДІЯЛЬНОСТІ ПЕРСОНАЛУ В УМОВАХ ВИРОБНИЧИХ КОМПАНІЙ, ЩО ПОВ'ЯЗАНІ З ЗАХИСТОМ ДАНИХ.....216

Олеся ЧЕРНЯКОВА
ПСИХОТЕХНОЛОГІЇ ВИРІШЕННЯ КОНФЛІКТІВ РОЗРОБНИКІВ СИСТЕМ ЗАХИСТУ ІНФОРМАЦІЇ В УМОВАХ КІБЕРЗАГРОЗ.....222

Максим ШАБАН
АПРОБАЦІЯ АПАРАТУ РЕГУЛЯРНИХ ВИРАЗІВ ПРИ АНАЛІЗІ ФУНКЦІОНАЛЬНОГО ПРОФІЛЮ ЗАХИСТУ.....227

Максим ШАБАН, Наталія ДЯЧЕНКО
ОРГАНІЗАЦІЯ ПРАВИЛ ПЕРЕВІРКИ ФУНКЦІОНАЛЬНОГО ПРОФІЛЮ ЗАХИСТУ 230

Геннадій ШВАЧИЧ, Павло ІЩУК
ІДЕНТИФІКАЦІЯ ДИНАМІКИ ДОВКІЛЛЯ: ВИКЛИКИ ТА ПЕРСПЕКТИВИ
РОЗПОДІЛЕНОГО МОДЕЛЮВАННЯ.....235

Андрій ЯРМОЛАТІЙ, Любов ЧЕРЕМІСІНА, Валентин ГАЛУНЬКО
ПОБУДОВА ВЕБ-ДОДАТКІВ ІЗ СУЧАСНИМИ СИСТЕМАМИ АУТЕНТИФІКАЦІЇ
НА БАЗІ OAUTH 2.0 PKCE ТА БІОМЕТРИКИ ДЛЯ ПІДВИЩЕННЯ
КІБЕРСТІЙКОСТІ ЦИХ ДОДАТКІВ.....242

CONTENTS

Valentyna ABLAMSKA, Natalia DIACHENKO, Valentyn HALUNKO
MATHEMATICAL MODELING OF THE VULNERABILITY OF THE INFORMATION SYSTEM AND FORECASTING OF ITS PROTECTION POSSIBILITIES.....3

Kateryna ANTIPOVA, Ihor KANDYBA, Svitlana BOROVLOVA, Viktor RALENKO
PLAGIARISM DETECTION IN TEXTS GENERATED BY LARGE LANGUAGE MODELS.....9

Mykhailo BERDNYK, Igor STARODUBSKYI
THE USE OF MACHINE LEARNING METHODS FOR ADAPTING SOFTWARE TO DIFFERENT COMPUTING PLATFORMS.....16

Volodymyr BOHOMYA, Liubov CHEREMISINA, Andrii YARMOLATII
SECURE DATA PROCESSING IN BUSINESS-INTERNET PLATFORMS AND E-COMMERCE SYSTEMS.....29

Volodymyr BOHOMIA, Valentyn HALUNKO
LEGAL REGULATION OF CYBERSECURITY IN THE CONTEXT OF CRITICAL INFRASTRUCTURE PROTECTION.....35

Andrii BORYSENKO
SECURE PROGRAMMING: AUTOMATED TOOLS FOR DETECTING VULNERABILITIES IN CODE.....43

Stanyslav VEDMEDEV
DEVELOPMENT OF SOFTWARE FOR COLLECTION OF PHENOTYPIC DATA OF SUNFLOWER SEEDS.....53

Vyacheslav GOREV, Yana SHEDLOVSKA, Ivan LAKTIONOV, Grygorii DIACHENKO
LMS PREDICTION AND KALMAN FILTERING FOR THE PROBABILITY OF THE MAIZE DISEASE OCCURENCE.....61

Serhii DOROZHYNSKYI, Vitaliy TUPKALO, Yuriy SHCHERBYNA
IMPORTANT ISSUES OF DATA PROCESSING AND PROTECTION IN DISTRIBUTED SYSTEMS.....68

Valentyn DIACHENKO, Natalia DIACHENKO
CONSIDERING RISKS WHEN USING INFORMATION TECHNOLOGY.....75

Timur ZHELDAK, Alexander VLADYKO
USING MACHINE LEARNING FOR CLUSTERINGUSING THE HAZARD INDEX OF TAILINGSAND THEIR ENVIRONMENTAL IMPACT.....81

Uliana ZHORNOKUI, Nataliia KOVALENKO
ENGLISH BORROWINGS IN THE CYBERSECURITY SECTOR ON THE CURRENT STAGE OF THE UKRAINIAN LANGUAGE DEVELOPMENT.....92

Denys IVANOV, Vita KASHTAN
GEOINFORMATION TECHNOLOGY FOR DETECTING AND MODELING FLOODED AREAS OF DUBAI BASED ON CONVOLUTIONAL NEURAL NETWORK.....99

Antonina KASHTALIAN
CHARACTERISTIC PROPERTIES OF CENTRALIZATION IN THE ARCHITECTURE OF MULTI -COMPUTER SYSTEMS OF ANTIVIRAL COMBINED BAITS AND TRAPS.....114

Olga KYVLIUK
THEORETICAL AND METHODOLOGICAL FOUNDATIONS OF INFORMATION AND CYBER SECURITY IN THE CONTEXT OF DIGITALISATION OF SOCIAL PROCESSES.....125

Pavlo KUNDENKO, Oleksander SOKOL, Nataliia KALININA, Viktoriia HNATUSHENKO
NEURAL NETWORK APPROACH TO DETECTING WATER OBJECTS IN MULTISPECTRAL IMAGES.....132

Maryna MOLCHANOVA
NEURAL NETWORK IDENTIFICATION OF PROPAGANDA GOALS IN TEXT CONTENT WITH VISUAL ANALYTICS OF RESULTS.....144

Igor NOVYTSKYI, Yuliia SHEVCHENKO, Maksym TRYPUTEN
ADAPTIVE CONTROL SYSTEM OF THE SELF-GRINDING PROCESS OF ORES IN AUTOGENOUS MILLS.....151

Viktor OLEVSKYI, Iryna UDOVYK, Dmytro HRYSHCHAK, Yuliia OLEVSKA, Yuliia ZHURAVLOVA
METHODOLOGICAL FOUNDATIONS OF INFORMATION TECHNOLOGIES FOR THE ANALYSIS OF MULTI-CHANNEL AEROSPACE IMAGES.....159

Serhii PRYKHODCHENKO, Larysa KORIASHKINA, Olha SHEVTSOVA, Yevhenii ZAHYNAILO
CONCEPTUAL MODEL OF THE SOFTWARE ARCHITECTURE FOR PROCESSING MULTI-SOURCE ACOUSTIC SIGNALS.....169

Pavlo REHIDA, Oleg SAVENKO
METHOD FOR DETECTING MALICIOUS ACTIVITY IN INFECTED PROGRAMS.....178

Oleksii SERHIEIEV
APPLYING THE SYSTEMS APPROACH TO THE ANALYSIS OF REGIONAL-LEVEL MEDICAL LOGISTICS.....187

Olena SOBKO
NEURAL NETWORK SEARCH AND CLASSIFICATION OF CYBERBULLYING IN TEXT MESSAGES.....197

Vitaliy TUPKALO
BLOCK-COMPOSITION METHOD OF FORMING CONTROL SIGNATURES OF DIGITAL SYSTEMS OPERATION BASED ON THE PROPERTIES OF THE SIGNATURE FIELD TSF[2^N,P^M(X)].....206

Liubov CHEREMISINA, Natalia MAKSYMOVYCH
MANAGEMENT OF THE OCCUPATIONAL HEALTH AND SAFETY SYSTEM AND PERSONNEL LIFE SAFETY IN PRODUCTION COMPANIES RELATED TO DATA PROTECTION.....216

Olesia CHERNIAKOVA
PSYCHOTECHNOLOGIES FOR CONFLICT RESOLUTION OF DEVELOPERS OF INFORMATION PROTECTION SYSTEMS IN THE CONDITIONS OF CYBER THREATS.....222

Maksym SHABAN
TESTING OF REGULAR EXPRESSION APPARATUS IN ANALYSIS OF FUNCTIONAL PROFILE OF PROTECTION.....227

Maksym SHABAN, Natalia DIACHENKO
ORGANIZATION OF RULES FOR VERIFICATION OF THE FUNCTIONAL PROFILE OF PROTECTION.....230

Hennadii SHVACHYCH, Pavlo ISHCHUK
IDENTIFICATION OF ENVIRONMENTAL DYNAMICS: CHALLENGES AND PROSPECTS OF DISTRIBUTED MODELING.....235

Andrii YARMOLATII, Liubov CHEREMISINA, Valentyn HALUNKO
BUILDING WEB APPLICATIONS WITH MODERN AUTHENTICATION SYSTEMS BASED ON OAUTH 2.0 PKCE AND BIOMETRICS TO ENHANCE THE CYBER RESILIENCE OF THESE APPLICATIONS.....242

INFORMATION TECHNOLOGY: COMPUTER SCIENCE, SOFTWARE ENGINEERING AND CYBER SECURITY

Випуск 4

Коректура • Ірина Миколаївна Чудеснова

Комп'ютерна верстка • Андрій Олександрович Філатов

Підписано до друку: 30.12.2024. Формат 60x84/8. Гарнітура Arial.
Папір офсет. Цифровий друк. Ум. друк. арк. 29,99. Замов. № 0125/109. Наклад 300 прим.

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