МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ ІМЕНІ ІВАНА ПУЛЮЯ
UNIVERSITY OF THE NATIONAL EDUCATION COMMISSION, POLAND
TECHNICAL UNIVERSITY IN PRAGUE, CZECH REPUBLIC
OXFORD BROOKES UNIVERSITY, UNITED KINGDOM
ГО «АСОЦІАЦІЯ СПЕЦІАЛІСТІВ КІБЕРБЕЗПЕКИ»
ГО «АВТОМАТИЗАЦІЯ І КІБЕРБЕЗПЕКА»

ITSec-2026

Безпека інформаційних технологій

МАТЕРІАЛИ

XV Міжнародної науково-технічної конференції

27-29 травня 2026
м. Тернопіль (Україна)

УДК [003.26+004+519.816]:004.056:65(063)

ITSec: Безпека інформаційних технологій: матеріали XV Міжнар. наук.-техн. конф., м. Тернопіль, 27-29 трав. 2026 р. Тернопіль-Київ: ТНТУ-ДУІКТ, 2026. – 380 с.

Збірник містить тексти наукових матеріалів доповідей та тез учасників XV міжнародної науково-технічної конференції «ITSec: Безпека інформаційних технологій». Основною метою конференції є ознайомлення з сучасними досягненнями та висвітлення результатів наукових досліджень з усіх аспектів інформаційної та кібербезпеки та захисту інформації.

Призначено вченим, інженерам, докторантам наукових спеціальностей 05.13.21 – Системи захисту інформації, 21.05.01 – Інформаційна безпека держави, здобувачам вищої освіти спеціальності F5 – Кібербезпека та захист інформації (125 Кібербезпека та захист інформації), а також всім зацікавленим.

ОРГАНІЗАТОРИ КОНФЕРЕНЦІЇ

- Міністерство освіти і науки України
- Державний університет інформаційно-комунікаційних технологій
- Тернопільський національний технічний університет імені Івана Пулюя
- University of the National Education Commission, Poland
- Technical University in Prague, Czech Republic
- Oxford Brookes University, United Kingdom
- ГО «Асоціація спеціалістів кібербезпеки»
- ГО «Автоматизація і кібербезпека»

Співголови програмного комітету

Володимир ШУЛЬГА, доктор історичних наук, професор, ректор Державного університету інформаційно-комунікаційних технологій;

Микола МИТНИК, кандидат технічних наук, доцент, ректор Тернопільського національного технічного університету імені Івана Пулюя.

Члени програмного комітету

Олександр КОРЧЕНКО, член-кореспондент НАН України, доктор технічних наук, професор, перший проректор Державного університету інформаційно-комунікаційних технологій, голова ГО «Асоціація спеціалістів кібербезпеки»;

Павло МАРУЩАК, доктор технічних наук, професор, проректор з наукової роботи Тернопільського національного технічного університету імені Івана Пулюя;

Ірина УДОВИК, кандидат технічних наук, професор, декан факультету інформаційних технологій, Національного технічного університету «Дніпровська політехніка»;

Ігор БАРАН, кандидат технічних наук, доцент, декан факультету комп'ютерно-інформаційних систем і програмної інженерії Тернопільського національного технічного університету імені Івана Пулюя;

Євгенія ІВАНЧЕНКО, доктор технічних наук, професор, директор навчально-наукового інституту кібербезпеки та захисту інформації Державного університету інформаційно-комунікаційних технологій;

Libor DOSTALEK, Technical University in Prague, Czech Republic;

Mikolaj KARPINSKI, Professor, Doctor of Science, Head of Department of Software Engineering, University of the National Education Commission, Poland;

Inna SKARGA-BANDUROVA, Professor, Doctor of Science, Associate Professor of Artificial Intelligence School of Engineering, Computing and Mathematics, Oxford Brookes University, United Kingdom;

Юлія ХОХЛАЧОВА, кандидат технічних наук, професор, професор кафедри штучного інтелекту Державного університету інформаційно-комунікаційних технологій;

Юлія ТКАЧ, кандидат технічних наук, доктор педагогічних наук, професор, завідувач кафедрою кібербезпеки та математичного моделювання Національного університету «Чернігівська політехніка».

Голова організаційного комітету

Наталія ЗАГОРОДНА, кандидат технічних наук, доцент, завідувач кафедри кібербезпеки Тернопільського національного технічного університету імені Івана Пулюя.

Заступник голови організаційного комітету

Руслан КОЗАК, кандидат технічних наук, доцент, доцент кафедри кібербезпеки Тернопільського національного технічного університету імені Івана Пулюя.

Науковий секретар організаційного комітету

Марина ДЕРКАЧ, кандидат технічних наук, доцент, доцент кафедри кібербезпеки Тернопільського національного технічного університету імені Івана Пулюя.

Члени організаційного комітету

Ігор БОДНАРЧУК, кандидат технічних наук, доцент, завідувач кафедри комп'ютерних наук Тернопільського національного технічного університету імені Івана Пулюя;

Галина ОСУХІВСЬКА, кандидат технічних наук, доцент, завідувачка кафедри комп'ютерних систем та мереж Тернопільського національного технічного університету імені Івана Пулюя;

Михайло ПЕТРИК, доктор технічних наук, професор, завідувач кафедри програмної інженерії Тернопільського національного технічного університету імені Івана Пулюя;

Василь ЯЦИШИН, кандидат технічних наук, доцент, завідувач кафедри штучного інтелекту та аналізу даних Тернопільського національного технічного університету імені Івана Пулюя;

Михайло ПРИГАРА, кандидат технічних наук, доцент кафедри технології машинобудування Ужгородського національного університету;

Дмитро ТИМОЩУК, старший викладач кафедри кібербезпеки Тернопільського національного технічного університету імені Івана Пулюя;

Марія СТАДНИК, кандидат технічних наук, доцент, доцент кафедри кібербезпеки Тернопільського національного технічного університету імені Івана Пулюя;

Олег ЯРЕМА, асистент кафедри кібербезпеки Тернопільського національного технічного університету імені Івана Пулюя;

Тарас ЛЕЧАЧЕНКО, PhD, старший викладач кафедри кібербезпеки Тернопільського національного технічного університету імені Івана Пулюя;

Роман ЗОЛОТИЙ, кандидат технічних наук, доцент, доцент кафедри комп'ютерно-інтегрованих технологій Тернопільського національного технічного університету імені Івана Пулюя.

ЗМІСТ

Formalization of Cyber Resilience Assessment of Critical Infrastructure Facilities to Phishing Attacks Based on a Set-Theoretic Approach

UDC 004.056

Oleksandr Korchenko[1], Yuliia Khokhlachova[2], Serhiy Skvortsov[3]

State University of Telecommunications, Kyiv, Ukraine,
National University "Kyiv Aviation Institute",
[1]agkorchenko@gmail.com, [2]yuliiahohlachova@gmail.com, [3]ssamailer@gmail.com

In the context of rapid digitalization of critical infrastructure objects, their functioning increasingly depends on complex information and telecommunication systems, which significantly increases the level of cyber risks. One of the most widespread and at the same time effective threats remains phishing attacks, which are based on social engineering methods and aimed at obtaining confidential user information. For critical infrastructure objects, such attacks can lead to serious consequences, including disruption of technological processes, loss of access to critical resources, and cascading failures.

Phishing is a complex multi-level threat that combines technical mechanisms and behavioral aspects. The main target of such attacks is the user, which makes the human factor a key element in the cybersecurity system. Even with modern information security tools in place, human errors are most often the cause of successful attacks.

Traditional cybersecurity approaches based on prevention and detection of incidents prove to be insufficient in the modern threat environment. In this regard, the concept of cyber resilience becomes relevant, which implies the ability of a system not only to resist attacks but also to maintain critical functions during an incident, quickly recover after it, and adapt to new operating conditions.

The purpose of the study is to formalize the process of assessing the cyber resilience of critical infrastructure objects under phishing attacks based on a set-theoretic data model. The proposed approach allows the transition from qualitative analysis to quantitative evaluation of the effectiveness of cybersecurity measures.

In the proposed model, phishing is considered as an element of the set of cyber threats that affects authentication mechanisms and user behavior. A hierarchical structure is introduced, which includes sets of strategic goals, tasks, subtasks, and basic cyber resilience measures. Each element of this structure is assigned a unique identifier, which ensures a formalized mapping between threats and countermeasures.

Strategic goals of cyber resilience include anticipation, withstanding, recovery, and adaptation. To achieve these goals, a set of tasks and subtasks is defined, which are implemented through basic measures. Such decomposition allows a detailed analysis of the contribution of each element to the overall level of cyber resilience.

In the case of phishing attacks, key measures include the implementation of multi-factor authentication, systematic personnel training, the use of analytical monitoring and anomaly detection tools, restriction of user privileges, and segmentation of access to resources. Additionally, it is important to implement security policies that regulate user behavior.

The proposed model allows formalizing the contribution of each measure to cyber resilience through the use of functional dependencies. This makes it possible to determine optimal combinations of measures depending on the type of threat and system characteristics.

A key element of the model is the introduction of quantitative evaluation metrics, including detection time, response time, level of functionality preservation, and the degree of impact on critical processes. In the context of phishing attacks, it is important to determine the interval between credential compromise and its detection, as well as the effectiveness of incident localization measures.

The model also allows analyzing different response scenarios and evaluating their effectiveness. This creates a basis for making informed management decisions and optimizing resources.

The use of the set-theoretic approach ensures flexibility and scalability of the model. It can be adapted to various types of critical infrastructure objects and allows integrating new threats without changing the basic structure.

The practical significance lies in the possibility of using the model to improve cybersecurity effectiveness, optimize costs, and ensure stable system operation.

Thus, the proposed approach provides a systematic assessment of cyber resilience, enables the transition to quantitative analysis, and increases the effectiveness of countering phishing attacks.

Prospects for further research include expanding the model to other types of attacks, integration with decision support systems, and the development of software tools for automated cyber resilience assessment.

1. Bodeau D, Graubart R, McQuaid R, Woodill J. Cyber Resiliency Metrics, Measures of Effectiveness, and Scoring: Enabling Systems Engineers and Program Managers to Select the Most Useful Assessment Methods. (The MITRE Corporation, Bedford, MA), MITRE Technical Report 2018.
2. Bodeau, D., & Graubart, R. Cyber Resiliency Engineering Framework. MITRE Corporation Technical Report MTR110237 2011.
3. Bodeau, D., Brtis, J., Graubart, R., & Salwen, J. Cyber Resiliency Engineering Aid – The Updated Cyber Resiliency Engineering Framework and Guidance on Applying Cyber Resiliency Techniques. MITRE Technical Report MTR150264 2015.
4. Ross, R., Pillitteri, V., Graubart, R., Bodeau, D., & McQuaid, R. Developing Cyber-Resilient Systems: A Systems Security Engineering Approach. NIST Special Publication 800-160 2021, Vol. 2, Rev. 1.

Квантовий прямий безпечний зв'язок і квантове розділення секрету з використанням переплутаних квантових станів

УДК 003.26:621.39+530.145

Євген Василіу[1], Олександр Назаренко, Сергій Стайкуца

Державний університет інтелектуальних технологій і зв'язку,
[1]y.v_vasiliu@suitt.edu.ua

Квантові технології захисту інформації є одним із найбільш перспективних напрямів розвитку сучасної криптографії та систем захищеного зв'язку. Інтенсивний розвиток квантових обчислень створює потенційну загрозу для класичних криптографічних алгоритмів, що стимулює пошук нових підходів до забезпечення конфіденційності та цілісності інформації. У зв'язку з цим значну увагу привертають протоколи квантового прямого безпечного зв'язку (КПБЗ) та квантового розділення секрету (КРС), які використовують фундаментальні властивості квантової механіки для захищеного передавання даних [1-3].

Метою роботи є огляд та порівняльний аналіз протоколів квантового прямого безпечного зв'язку і квантового розділення секрету, що базуються на використанні переплутаних квантових станів кубітів та квантових систем більшої розмірності. Особливу увагу приділено пінг-понг протоколам зі станами Белла, багатокубітними ГХЦ-станами та кутритними переплутаними станами, а також аналізу їх інформаційної місткості та стійкості до некогерентних атак пасивного перехоплення.

На відміну від систем квантового розподілення ключів, у протоколах КПБЗ квантовий канал використовується для прямого передавання інформації, а не лише для формування спільного секретного ключа. Кодування класичної інформації виконується шляхом локальних унітарних операцій над частиною переплутаної квантової системи. Для протоколів із використанням станів Белла реалізується механізм квантового надщільного кодування, який дозволяє передавати два класичних біти за один цикл обміну квантовими частинками. Застосування багаточастинкових ГХЦ-станів та надщільного кодування забезпечує збільшення кількості переданих бітів та створює можливість реалізації багатокористувацьких схем захищеного квантового зв'язку.

Важливим елементом пінг-понг протоколів є наявність режиму контролю прослуховування каналу. У цьому режимі учасники виконують контрольні вимірювання квантових станів для виявлення спроб несанкціонованого втручання. Наявність підслуховування призводить до порушення квантових кореляцій між переплутаними частинками, що дозволяє виявити атаку з певною ймовірністю. Аналіз показує, що збільшення розмірності квантових систем та кількості частинок у багаточастинковому переплутаному стані підвищує рівень стійкості протоколів до атак перехоплення.

Окрему увагу приділено атакам пасивного перехоплення, які базуються на переплутуванні допоміжної квантової системи з передаваними кубітами та належать до класу некогерентних атак. Ефективність таких атак може оцінюватися за допомогою ентропії фон Неймана та ймовірності невиявлення перехоплення. Проведений аналіз демонструє, що протоколи з багато-

частинковими переплутаними станами характеризуються вищим рівнем стійкості порівняно з базовими схемами.

Для додаткового посилення стійкості може застосовуватися метод посилення таємності, який базується на використанні випадкових оборотних двійкових матриць та випадкової двійкової гами. Такий підхід забезпечує інформаційну незалежність переданих блоків від вихідного повідомлення навіть у разі часткового отримання інформації порушником.

Розглянуто також протоколи квантового розділення секрету, у яких відновлення конфіденційної інформації можливе лише за умови кооперації декількох учасників мережі. Використання переплутаних станів Белла та багатокубітних ГХЦ-станів дозволяє реалізувати схеми, у яких жоден із учасників окремо не може відновити секрет. Відновлення інформації здійснюється лише після обміну результатами вимірювань та інформацією про виконані кодувальні операції. Такий підхід забезпечує високий рівень захисту від внутрішніх і зовнішніх порушників та може бути використаний у розподілених інформаційних системах і критично важливих мережах.

Проведений огляд свідчить, що протоколи КПБЗ і КРС є перспективною основою для побудови майбутніх квантових мереж та систем захисту інформації. Серед основних переваг таких протоколів можна виділити можливість прямого передавання інформації, вбудоване виявлення підслуховування, підтримку багатокористувацьких сценаріїв та потенційно теоретико-інформаційний рівень захищеності. Водночас практична реалізація квантових систем зв'язку супроводжується значними технічними труднощами, пов'язаними з необхідністю використання ефективних джерел одиничних фотонів і переплутаних станів, квантової пам'яті, чутливих детекторів і захисту від впливу шумів та втрат у каналах передавання.

Подальший розвиток квантових технологій захисту інформації пов'язаний із вдосконаленням квантових каналів зв'язку, підвищенням завадостійкості протоколів, створенням масштабованих квантових мереж та інтеграцією квантових і постквантових криптографічних методів. Очікується, що розвиток супутникового квантового зв'язку та глобальних квантових мереж стане основою для формування нової глобальної безпечної інформаційної інфраструктури.

1. Vasiliu Y. Modern Quantum Technologies of Cryptographic Protection of Information. *Cybernetics and Systems Analysis*. – 2025. – Vol. 61, no. 4. – P. 671 – 684.
2. Pan D., Long G.L., Yin L., Sheng Y.B., Ruan D., Ng S.X., Lu J., Hanzo L. The Evolution of Quantum Secure Direct Communication: On the Road to the Qinternet. *IEEE Communications Surveys & Tutorials*. – 2024. – Vol. 26, no. 3. – P. 1898 – 1949.
3. Liu S., Lu Z., Wang P. et al. Experimental demonstration of multiparty quantum secret sharing and conference key agreement. *npj Quantum Information*. – 2023. – Vol. 9. – 92.

Інтелектуальна навчально-методична платформа систем безпеки у вигляді взаємодіючих агентів штучного інтелекту

УДК 004.89:004.056

Валерій Домарєв[1], Юрій Хлапонін[2]

[1]ПрАТ «Діпрозв'язок», domarev@ukr.net,
[2]Державний торговельно-економічний університет, y.khlaponin@knute.edu.ua

Сучасний світ перетворився на складну систему взаємопов'язаних процесів, де атака на один банк може вплинути на фінансову систему країни, а помилка в енергетичному управлінні — залишити без електроенергії цілі регіони. Традиційні системи безпеки, побудовані за принципом «латання дірок», більше не забезпечують цілісного захисту. Потрібна нова парадигма, яка поєднує системне мислення, логіко-лінгвістичний аналіз та штучний інтелект.

Мета роботи. Розробка та обґрунтування інтелектуальної навчально-методичної платформи AI MATRIX TEACHER, що реалізує логіко-лінгвістичну матричну модель систем безпеки у вигляді взаємодіючих агентів штучного інтелекту, та її застосування для підготовки фахівців у сфері кібербезпеки (рис. 1).

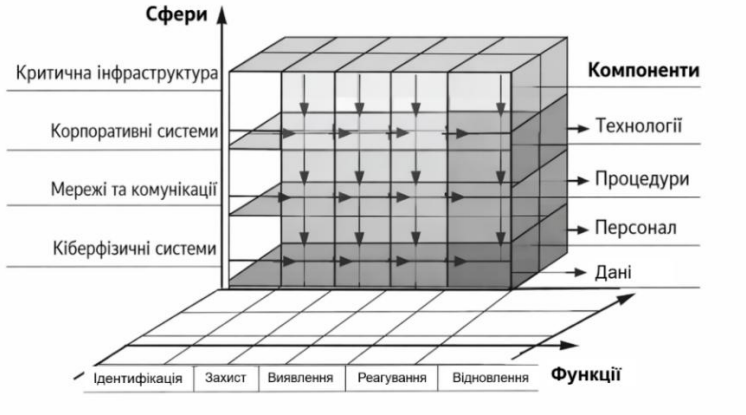


Рис. 1. Логіко-лінгвістична матрична модель систем безпеки

Наукова новизна. Запропонована платформа не є окремим програмним продуктом, а методологією мислення. Її ключова особливість — перехід від моделі накопичення знань до моделі розвитку інтелектуальних здібностей мислення. Емерджентні властивості взаємодії агентів ШІ створюють нові форми інтелектуального навчання, де штучний інтелект не замінює фахівця, а підсилює його творчі можливості.

Основні результати.

1. Розроблено логіко-лінгвістичну 3D-модель систем безпеки, що описує взаємодію ресурсів, загроз, заходів протидії, процесів управління, суб'єктів та інформаційних потоків (див. рис. 1)

2. Показано, що міжнародні стандарти (ISO 27001, NIST CSF, IEC 62443 тощо) можуть бути інтегровані у матрицю як рекомендації для конкретних перетинів «сфера — функція — компонент».

3. Реалізовано мультиагентну платформу, де ШІ-агенти автоматично виявляють прогалини у політиках та процедурах, формують звіти про невідповідності та пропонують адаптивні рішення.

4. Запропоновано методику використання платформи у навчальному процесі для формування системного мислення у студентів та підвищення кваліфікації фахівців.

Висновки. Інтелектуальна навчально-методична платформа AI MATRIX TEACHER створює нову парадигму підготовки кадрів у сфері інформаційної безпеки. Вона поєднує системний підхід, креативне мислення та можливості штучного інтелекту, забезпечуючи перехід від накопичення знань до розвитку інтелектуального творчого потенціалу. Це відкриває перспективи для формування компетенцій нового покоління фахівців, здатних діяти в умовах невизначеності та гібридних загроз.

1. Almuhammadi, S., & Alrehaili, A. (2023). Intelligent multi-agent system for cybersecurity education: A personalized learning approach.
2. Journal of Cybersecurity Education, Research and Practice, 2023(1), Article 4.
3. Chen, Y., Li, T., & Zhang, J. (2024). A logical-linguistic matrix model for security knowledge representation in intelligent tutoring systems. IEEE Transactions on Learning Technologies, 17(3), 512–526.
4. Nguyen, P. H., & Kolbe, N. (2022). Multi-agent reinforcement learning for cyber range training: A framework for adaptive security exercises. Computers & Security, 118, 102728.
5. Rodriguez, M., & Kotenko, I. (2025). AI-driven educational platform for security systems engineering: Integrating linguistic analysis and agent-based simulation. In Proceedings of the 25th IEEE International Conference on Advanced Learning Technologies (ICALT 2025), 187–191.
6. Венгер С.А., Марченко А.О. Інтелектуальна модель формування індивідуальної освітньої траєкторії здобувача на навчальній платформі. Сучасні інформаційні технології у сфері безпеки та оборони. – 2024. – № 1(49). – С. 160-170. DOI: 10.33099/2311-7249/2024-49-1-160-170.
7. Гречанінов В. Моделі та технології інтелектуального захисту інформаційних систем критичної інфраструктури для підвищення стійкості. Кібербезпека: освіта, наука, техніка. – 2025. – Т. 1, № 29. DOI: 10.28925/2663-4023.2025.29.948.

Когнітивний кіберконтроль та механізми уваги в системах кібермоніторингу: до проблеми прихованого впливу

УДК 004.056:004.8

Ткач Юлія[1], Шелест Михайло[2]

Національний університет «Чернігівська політехніка»,
[1]tkachym79@gmail.com, [2]mishel3141@gmail.com

Анотація. Сучасні системи кібермоніторингу об'єктів критичної інфраструктури функціонують в умовах стрімкого зростання обсягів даних, що надходять із різнорідних джерел [1]. У таких умовах ключовою проблемою стає не лише обробка інформації, а визначення її значущості для прийняття рішень.

Це зумовлює формування нового функціонального рівня, який можна охарактеризувати як когнітивний кіберконтроль — процес управління значущістю інформації та пріоритетами реагування.

У сучасних системах цей рівень реалізується через механізми attention, які визначають релевантність даних і тим самим впливають на поведінку системи [3]. Це ставить питання про їх стійкість до можливого прихованого впливу.

Вступ. Сучасні системи кібермоніторингу об'єктів критичної інфраструктури функціонують в умовах стрімкого зростання обсягів даних, що надходять із різнорідних джерел. У таких умовах ключовою проблемою стає не лише забезпечення збору та обробки інформації, а її відбір і інтерпретація — визначення того, які саме події є значущими для прийняття рішень.

Це призводить до формування нового функціонального рівня у системах кібербезпеки, який можна охарактеризувати як когнітивний кіберконтроль — процес управління значущістю інформації та пріоритетами реагування в умовах динамічного середовища.

У сучасних інформаційно-аналітичних системах, особливо тих, що використовують методи штучного інтелекту, зазначений рівень реалізується через механізми уваги (attention), які забезпечують відбір релевантних даних, формування контексту та визначення пріоритетів обробки інформації. Це означає, що вони виконують не лише допоміжну, а й керуючу функцію, впливаючи на поведінку системи в цілому.

За таких умов постає принципово важливе питання: наскільки стійкими є ці механізми до зовнішнього або прихованого впливу, і чи можуть вони виступати об'єктом цілеспрямованого керування.

Attention як функціональний рівень кібермоніторингу. У класичному підході кібермоніторинг розглядається як процес збору, обробки та аналізу даних. Однак у складних системах ключовим стає питання: *які саме дані впливають на рішення?* Це питання вирішується через механізми attention, які забезпечують фільтрацію інформаційних потоків, виділення релевантних сигналів, формування пріоритетів і визначення контексту прийняття рішень.

Фактично attention визначає "картину реальності", яку бачить система. Зміна цього рівня призводить до зміни поведінки всієї системи.

Attention як поверхня керування (control surface). У багаторівневій структурі систем кібермоніторингу attention може розглядатися як окремий

функціональний рівень, що забезпечує оцінювання значущості подій, впливає на процедури прийняття рішень і визначає подальші керуючі дії.

У такій інтерпретації attention можна розглядати як control surface — поверхню керування, через яку здійснюється вплив на поведінку системи. Це має принципове значення, оскільки:

- контроль над attention означає контроль над пріоритетами;
- контроль над пріоритетами означає контроль над рішеннями;
- контроль над рішеннями означає контроль над системою.

Потенційні вразливості та прихований вплив. Розгляд attention як окремого функціонального рівня дозволяє виявити новий клас потенційних вразливостей, пов'язаних із можливістю:

- зміни критеріїв відбору інформації;
- маніпуляції пріоритетами подій;
- прихованого зміщення фокусу системи;
- формування викривленої ситуаційної обізнаності.

Особливістю таких впливів є їх неявний характер: система може функціонувати коректно з технічної точки зору, але приймати субоптимальні або хибні рішення через зміну рівня attention.

Клептографічна інтерпретація. Зазначені механізми можуть бути розглянуті у ширшому контексті дослідження прихованих впливів у цифрових системах [4]. Можливість непомітного впливу на процес визначення значущості інформації дозволяє розглядати attention-рівень як потенційний об'єкт клептографічного впливу — управління через зміну пріоритетів обробки даних.

У цьому сенсі можна говорити про формування нового напрямку досліджень — клептографії рівня уваги (attention-level kleptography), яка розширює традиційні уявлення про приховані механізми контролю.

Висновки. У роботі запропоновано розглядати кібермоніторинг як систему формування значущості подій у динамічному середовищі. Показано, що механізми attention виступають ключовим рівнем керування, який визначає поведінку системи. Це відкриває можливість аналізу нових типів вразливостей, пов'язаних із прихованим впливом на процеси прийняття рішень.

1. Endsley M. R. Toward a Theory of Situation Awareness in Dynamic Systems // Human Factors. – 1995.
2. Kott A., Wang C., Erbacher R. Cyber Defense and Situational Awareness. – Springer, 2014.
3. Vaswani A. et al. Attention Is All You Need // NeurIPS. – 2017.
4. Шелест М. Є., Ткач Ю. М. Клептографія: від бекдору до політики довіри у цифрову епоху. – Львів: Новий світ-2000, 2025. – 303 с.

Оптимізація порогу прийняття рішення в системах IDS/IPS на основі моделей машинного навчання

УДК 004.8:004.056

Каріна Крушельницька[1], Дмитро Тимощук[2], Наталія Загородна[3]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]karina.kryshel@gmail.com, [2]dmytro.tymoshchuk@gmail.com,
[3]zagorodna_n@tntu.edu.ua

Цифровізація, розширення мережевої інфраструктури та зростання обсягів мережевого трафіку супроводжуються підвищенням кількості й складності кіберзагроз, що зумовлює активне використання методів машинного навчання в системах IDS/IPS. На відміну від сигнатурних підходів, які переважно ґрунтуються на відомих шаблонах атак і тому мають обмежену ефективність щодо нових загроз, алгоритми машинного навчання здатні аналізувати поведінкові ознаки трафіку та виявляти аномалії, характерні для нових або модифікованих атак, зокрема атак «нульового дня».

ML-класифікатори зазвичай формують не лише бінарну мітку класу, а й числову оцінку ймовірності або впевненості моделі, яку порівнюють із порогом класифікації τ. Без попереднього калібрування такі оцінки можуть бути зміщеними, особливо для дерев рішень, ансамблевих і бустингових моделей, що ускладнює прямий вибір порогу за шкалою ймовірностей. Хоча значення τ = 0,5 часто використовується як стандартне, у задачах IDS/IPS воно не завжди є оптимальним через незбалансованість класів і різну вартість помилок класифікації [1]. Для отримання більш достовірних імовірнісних оцінок перед вибором порогу доцільно застосовувати методи калібрування, а оптимальне значення τ визначати за метриками, релевантними до задачі виявлення вторгнень. Вибір порогу класифікації слід розглядати не лише як технічний параметр ML-моделі, а як важливе проєктне рішення, що визначає баланс між рівнем виявлення атак, кількістю хибних спрацювань і загальною ефективністю системи захисту.

Якість класифікації в задачах IDS/IPS доцільно оцінювати на основі матриці помилок, елементами якої є TP, TN, FP та FN. За умови, що позитивним класом вважається атака, TP відповідає правильно виявленій атаці, TN — правильно класифікованому нормальному трафіку, FP — хибній тривозі, коли легітимний трафік помилково визначено як атаку, а FN — пропущеній атаці. На основі цих величин обчислюють основні метрики оцінювання ефективності ML-класифікатора, наведені в таблиці 1.

Таблиця 1

Основні метрики оцінювання ML-класифікатора в задачах IDS/IPS

Метрика	*Формула*	*Що вимірює*	*Особливості застосування в IDS/IPS*
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Частку всіх правильних класифікацій	Може бути неінформативною при дисбалансі класів, оскільки високе значення може досягатися за рахунок домінування нормального трафіку

Precision	$\frac{TP}{TP+FP}$	Частку справжніх атак серед усіх зразків, класифікованих як атаки	Важлива для зменшення хибних тривог, особливо в IPS, де FP може спричинити блокування легітимного трафіку
Recall/TPR /Sensitivity	$\frac{TP}{TP+FN}$	Частку реально наявних атак, які були виявлені моделлю	Важлива для IDS/IPS, оскільки низьке значення Recall означає велику кількість пропущених атак
F1-score	$\frac{2 * Precision * Recall}{Precision + Recall}$	Гармонійне середнє між Precision і Recall	Доцільна як компромісна метрика, коли потрібно збалансувати хибні тривоги та пропущені атаки
ROC-AUC	Площа під ROC-кривою	Загальну здатність моделі розділяти класи при різних порогах	Може переоцінювати якість моделі при сильному дисбалансі класів
PR-AUC	Площа під Precision–Recall-кривою	Якість виявлення позитивного класу за різних порогів	Особливо корисна при дисбалансі класів, коли атаки становлять меншість
FPR	$\frac{FP}{FP+TN}$	Частку нормального трафіку, помилково класифікованого як атака	Критична для IPS, оскільки високий FPR може призводити до блокування легітимного трафіку та перевантаження системи реагування
TNR/Specificity	$\frac{TN}{TN+FP}$	Частку нормального трафіку, правильно класифікованого як нормальний	Важлива для оцінювання здатності системи не створювати хибних тривог; особливо актуальна для IPS, де FP може призводити до блокування легітимного трафіку
G-Mean	$\sqrt{TPR \cdot TNR}$	Збалансованість виявлення атак і правильного розпізнавання нормального трафіку	Корисна при дисбалансі класів, оскільки враховує якість класифікації як позитивного, так і негативного класу

Таким чином, у задачах IDS/IPS недостатньо оцінювати модель лише за показником Accuracy, оскільки ця метрика може бути неінформативною за умов дисбалансу класів. Більш обґрунтованим є комплексне використання Precision, TPR, TNR, F1-score, PR-AUC, FPR та G-Mean, що дає змогу оцінити баланс між здатністю системи виявляти атаки, рівнем хибних спрацювань і якістю розпізнавання нормального трафіку [2]. Оптимальне значення порогу класифікації доцільно визначати не довільно, а на основі цільового критерію, що відповідає функціональному призначенню IDS/IPS. Залежно від пріоритетів системи таким критерієм може бути максимізація F1-score, G-Mean або статистики Youden's J (J = TPR − FPR), досягнення заданого балансу між Precision і Recall, мінімізація очікуваної вартості помилок класифікації або максимізація Recall за умови допустимого рівня FPR.

Отже, універсального значення порогу τ не існує. Його вибір має залежати від операційної ролі системи, допустимого рівня ризику та вартості помилок класифікації. Для IDS пріоритетним може бути підвищення Recall з метою мінімізації пропущених атак, тоді як для IPS особливо важливо контролювати FPR, оскільки хибні спрацювання можуть призводити до блокування легітимного трафіку та порушення доступності сервісів. Тому поріг класифікації доцільно розглядати як керований проєктний параметр, що визначає практичну ефективність ML-моделі в конкретному середовищі розгортання.

1. Tymoshchuk, D., Zagorodna, N., Klots, Y., Yatskiv, V., Petliak, N. AutoML and explainable AI-based approach to enhance the efficiency and interpretability of IDS. CEUR Workshop Proceedings, 2025, 4163, pp. 231-246
2. Tymoshchuk, D., Sverstiuk, A., Klots, Y., Petliak, N., Titova, V. An explainable artificial intelligence approach for detecting network attacks. CEUR Workshop Proceedings, 2025, 4141, pp. 38-51

Застосування нечітких продукційних правил для контекстно-довірчого оцінювання кіберризиків у середовищі Інтернету речей

УДК 621.395.7 (043.2)

Підлісний Юрій[1], Шелест Михайло[2]

Національний університет «Чернігівська політехніка», [1]ypodlesny@ukr.net

Стрімкий розвиток Інтернету речей (IoT) призвів до масового впровадження інтелектуальних пристроїв у різних сферах, що супроводжується зростанням кіберзагроз через обмежені ресурси та спрощені механізми захисту [1].

Традиційні методики оцінювання ризиків орієнтовані на класичні інформаційні системи та недостатньо ефективні в IoT через динамічність середовища, гетерогенність пристроїв та неповноту даних [2]. У таких умовах доцільним є застосування методів нечіткої логіки, які дозволяють працювати з експертними оцінками та невизначеними параметрами [3].

У роботі запропоновано підхід до оцінювання кіберризиків у середовищі IoT на основі нечітких продукційних правил типу IF–THEN (табл.1). Наведена база правил є фрагментом знань, який може бути розширений або адаптований залежно від специфіки IoT-середовища та моделі загроз.

Таблиця 1

Фрагмент бази нечітких продукційних правил

№	Правило	Результат
1	IF V = High AND T = High AND D = Low THEN R = Critical	Критичний
2	IF V = Medium AND T = High THEN R = High	Високий
3	IF V = Low AND T = Medium AND D = High THEN R = Medium	Середній
4	IF S = High AND V = Low THEN R = Low	Низький
5	IF A = High AND T = Medium THEN R = High	Високий
6	IF C = High AND V = Medium THEN R = High	Високий
7	IF S = Low AND T = High THEN R = Critical	Критичний

Запропонована модель враховує технічні характеристики вузлів, стан середовища та достовірність даних, що забезпечує її застосування як у статичних, так і в динамічних IoT-системах.

Вхідними параметрами моделі є: рівень вразливості (V), інтенсивність загроз (T), рівень захищеності (S), критичність активу (C), мережева аномальність (A) та довіра до джерела даних (D). Вихідною змінною є інтегральний показник ризику R, що характеризує ступінь небезпеки для конкретного вузла, підсистеми або сегмента мережі.

$$R = \Psi(V, T, C, S, A, D) \quad (1)$$

де Ψ – оператор нечіткого логічного виведення.

Для отримання числового значення ризику використовується процедура дефазифікації методом центру ваги [3]:

$$R = \frac{\sum_{i=1}^{n} \mu_i x_i}{\sum_{i=1}^{n} \mu_i} \quad (2)$$

де μ_i – ступінь належності вихідного терма, x_i – відповідне значення шкали ризику.

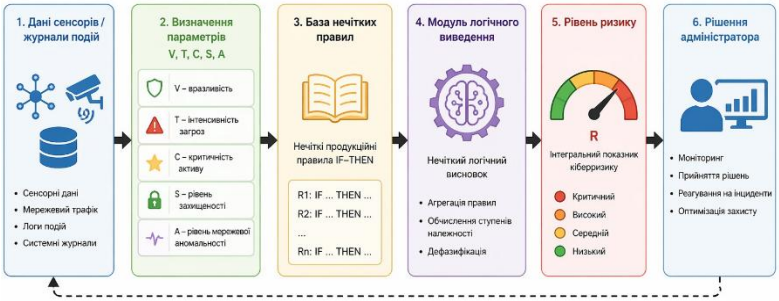


Рис.1 Структурна схема оцінювання кіберризику в IoT-середовищі

Наукова новизна. Полягає у переході від класичного оцінювання кіберризику на основі статичних параметрів до контекстно-залежної моделі, яка враховує не лише значення параметрів, але й ступінь довіри до джерел їх отримання. Запропонований підхід вводить мета-параметр довіри як фактор, що впливає на інтерпретацію вхідних даних, що дозволяє оцінювати ризик в умовах неповної, суперечливої або потенційно спотвореної інформації.

Контекст безпеки моделі. Особливу увагу слід приділити сценаріям навмисного спотворення вхідних параметрів, коли атакуючий впливає не безпосередньо на систему, а на дані, що використовуються для оцінювання ризику. У таких умовах параметр довіри до джерела інформації дозволяє враховувати можливість компрометації сенсорів, каналів передачі або систем моніторингу, що забезпечує більш стійке оцінювання ризику в умовах інформаційного впливу.

Практична цінність. Полягає у можливості використання запропонованої моделі в системах моніторингу безпеки та підтримки прийняття рішень у середовищах Інтернету речей. Модель дозволяє адаптивно оцінювати ризики з урахуванням динамічних змін середовища, неповноти даних та рівня довіри до джерел інформації, що особливо важливо для систем із розподіленою архітектурою та обмеженими ресурсами. Перспективним напрямом є дослідження атак, спрямованих на спотворення параметрів оцінювання ризику.

Висновки. Запропонований підхід дозволяє перейти від статичних моделей оцінювання вразливостей до адаптивного аналізу кіберризиків, який враховує як технічні характеристики системи, так і контекст її функціонування.

Введення параметра довіри до джерел даних розширює можливості моделі в умовах невизначеності та потенційного інформаційного впливу, що робить її придатною для застосування в сучасних кіберфізичних системах.

1. Roman R., Lopez J. Security in the Internet of Things: Current status and future challenges // Computer Networks. 2021.
2. Information security, cybersecurity and privacy protection — Guidance on managing information security risks : ISO/IEC 27005:2022. Geneva, 2022.
3. Zadeh L. Fuzzy sets // Information and Control. 1965. Vol. 8 / 3. P. 338–353.

Модифікація шифру Present

УДК:004.056.55

Володимир Лужецький[1], Тетяна Кирилащук[2]

Вінницький національний технічний університет,
[1]lva.kzi2002@gmail.com, [2]kgt0998@gmail.com

Алгоритм PRESENT є легковаговим блоковим шифром [1, 2], що базується на SP-мережі та складається з операцій підстановки та перестановки бітів. Операції підстановки реалізуються з використанням 16-ти однакових S-блоків, що складаються з логічних елементів, а перестановки бітів реалізуються фізичним розташуванням зв'язків. Крім того, для розгортання ключа використовується ще один S-блок. Оскільки S-блоки є необоротними, то для розшифрування використовуються інші S-блоки. Разом із перевагами алгоритму PRESENT існують і певні недоліки, пов'язані зі структурою S-блоків та кількістю раундів шифрування. У стандартній реалізації в кожному раунді використовуються однакові S-блоки, що значно спрощує апаратну реалізацію алгоритму, однак створює повторювану криптографічну структуру. Така регулярність може негативно впливати на стійкість алгоритму до окремих видів криптоаналізу, зокрема лінійного та диференціального. Ще одним недоліком є необхідність використання окремих інверсних S-блоків для процесу розшифрування. Для досягнення достатнього рівня нелінійності та дифузії даних використовується 31 раунд перетворень щоб компенсувати використання однакових S-блоків у всіх раундах.

З метою підвищення ефективності алгоритму шифрування у доповіді пропонується використовувати різні S-блоки, побудовані на основі латинських квадратів та перестановки змінних для S-блоків. Використання різних нелінійних перетворень у різних раундах дозволяє ускладнити криптоаналітичні залежності між ними та потенційно забезпечити необхідну криптографічну стійкість при меншій кількості раундів. Такий підхід може сприяти пришвидшенню процесу шифрування, зменшенню затримок обробки даних та підвищенню продуктивності алгоритму в ресурсно-обмежених системах, зокрема в IoT-пристроях та вбудованих інформаційних системах.

Пропонується для реалізації S-блоків використовувати латинські квадрати 4-го порядку. Оскільки, існує 576 таких латинських квадратів [3], то є можливість вибрати з них 16, які забезпечать побудову оборотних S-блоків. Як

наслідок, для розшифрування будуть використовуватись ті ж самі S-блоки, що і для зашифрування. Схему S-блоку на основі двох однакових латинських квадратів наведено на рис.1.

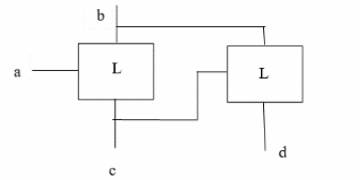


Рис.1 Схема S-блоку на основі латинських квадратів

Результат підстановки визначається як:

$$S'(x) = L(L(a,b),c),$$

де L - латинський квадрат, $a \parallel b$ - 4 вхідні біти, $c \parallel d$ – результат підстановки.

Проведені дослідження дозволили обрати 16 латинських квадратів 4x4, які є оборотними та забезпечують максимальну нелінійність S-блоків.

Додатково застосовується перестановка змінних:

$$\pi_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix};\ \pi_2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix};\ \pi_3 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix};\ \pi_4 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix};$$

$$\pi_5 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 1 & 2 \end{pmatrix};\ \pi_6 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix};\ \pi_7 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 2 & 1 \end{pmatrix};\ \pi_8 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}.$$

де π - перестановка індексів. Це означає, що перед застосуванням S-блоку змінюється порядок бітів, що дозволяє підвищити нелінійність та ускладнити алгебраїчну структуру. Ще одна модифікація шифру PRESENT полягає в тому, що накладання ключа після перестановки виконується не на основі операції XOR, а як операція, що описується латинським квадратом. Тобто, замість суматорів за модулем 2 використовуються логічні схеми, що реалізують латинські квадрати.

Таким чином, модифікація полягає не в зміні всієї структури PRESENT, а саме в заміні набору з 16 однакових S-блоків на набір з 16 різних S-блоків, сформованих на основі різних оборотних латинських квадратів. Такий підхід дозволяє зберегти легковагову структуру шифру, але водночас підвищити його криптографічну стійкість. Апаратна складність реалізації S-блоку на основі оборотних латинських квадратів складає 64 умовні одиниці, тоді як складність реалізації S-блоку шифру PRESENT дорівнює 95 умовних одиниць.

Отже, встановлено, що використання латинських квадратів у поєднані з перестановками змінних дозволяє досягти балансу між підвищенням криптостійкості та збереженням прийнятного рівня апаратної складності, що є важливим для ресурсно-обмежених середовищ, таких як IoT-пристрої та вбудовані системи.

1. Bogdanov A., Knudsen L., Leander G. та ін. PRESENT: An Ultra-Lightweight Block Cipher // Cryptographic Hardware and Embedded Systems – CHES 2007. Berlin : Springer, 2007. С. 450–466.
2. Wang Y., Ha Y. Compact FPGA implementation of PRESENT cipher with optimized S-box // IEEE Transactions on Very Large Scale Integration Systems. – 2011. – Vol. 19, №10. – P. 1864–1873.
3. A. Donald Keedwell and József Dénes. Latin Squares and their Applications. Elsevier, 2015. 439 с. URL: https://doi.org/10.1016/c2014-0-03412-0.

Розробка архітектури програмного застосунку для децентралізованої торгівлі електроенергією з використанням смарт-контрактів в блокчейн

УДК 004.4

Ганна Неласа[1], Вахтанг Чіхладзе[2],
Андрій Ублінських[3], Олег Неласий[4]

Інститут проблем моделювання в енергетиці ім. Г.Є. Пухова,
[1]annanelasa@gmail.com,
Національний технічний університет України «КПІ імені Ігоря Сікорського», [2]the.vaho1337@gmail.com,
Cytric, [3]andreo.ublin24@gmail.com,
Національний університет «Запорізька політехніка», [4]oleg.nelasy@gmail.com

В роботі представлено архітектуру програмного застосунку для децентралізованої торгівлі електроенергією з використанням смарт-контрактів в блокчейн [1]. На сьогодні смартконтракти набули широкого застосування та стали невід'ємною складовою блокчейн і фінансової індустрії.

Основна ідея полягає в токенізації 15-хвилинних порцій електроенергії, розділенні вартості електроенергії та прибутку на різні токени, та введенні фінансового посередника між виробниками та споживачами, який балансує момент розрахунків між ними, тобто сплачує виробнику за поставлений обсяг електроенергією відразу, а оплату від споживачів отримує пізніше при використанні відповідних порцій. Фінансовий посередник може отримувати дохід за свої послуги, оскільки протокол Pendle [2] довів створення ринку дохідності при розділенні базового активу та майбутнього доходу.

Відповідно в схемі використовуються токени:

- PT (PrincipalToken) - токен, який надає право користування електроенергією в заданий проміжок часу, «тіло» активу, тобто сума, що буде повернута після настання дати погашення. PT не приносить дохід, а лише гарантує повернення активу.
- YT (YieldToken) - токен, який містить у собі майбутній прибуток виробника-продавця.
- SY (Standardized Yield Token) виступає як «обгортка», що представляє повну вартість активу разом із майбутньою дохідністю.

Роботу виконано за держбюджетною темою «Розвиток розподіленої енергетики в умовах ринку електричної енергії України з використанням технологій та систем цифровізації. Розділ 1. Організаційні та математичні моделі взаємодії учасників децентралізованого ринку електроенергії» (ЦИФРОВІЗАЦІЯ), КПКВК 6541230.

1. Evdokimov, V.; Kudin, A.; Chikhladze, V.; Artemchuk, V. A Blockchain Architecture for Hourly Electricity Rights and Yield Derivatives. FinTech. – 2026, 5(1), №2. https://doi.org/10.3390/fintech5010002
2. Pendle Finance Documentation. Available URL: https://docs.pendle.finance (application date 09.05.2026).

Гібридний метод приховування водяних знаків на основі конформних відображень та сингулярного розкладу матриць

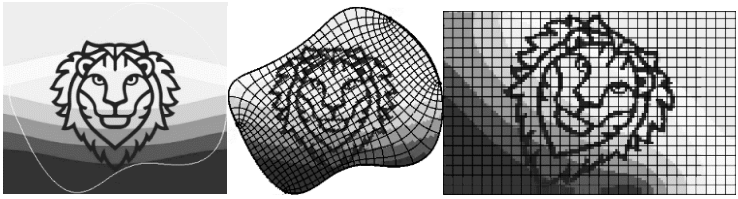
УДК 004.056.55 : 517.54

Андрій Бомба[1], Михайло Бойчура[2]

Національний університет водного господарства та природокористування, [2]m.v.boichura@nuwm.edu.ua

Традиційні методи вбудовування невидимих цифрових водяних знаків (ЦВЗ) в області є вразливими до геометричних атак, а використання частотних методів ускладнюється у випадку фігур складної геометрії. Метою роботи є підвищення захищеності зображень з криволінійною границею шляхом розроблення робастного методу вбудовування ЦВЗ, стійкого до топологічних деформацій.

Запропоновано підхід до накладання ЦВЗ на однозв'язну область (зображення) довільної форми. Він передбачає поєднання числових методів конформних відображень для перетворення досліджуваної області на прямокутник [1] та сингулярного розкладу матриць – для подальшого вбудовування прихованої інформації [2]. Окрім іншого, підхід передбачає побудову геометричного криптографічного ключа.



а) б) в)

Рис. 1. Початкове зображення (а), сітки фізичної області (б) та області комплексного потенціалу (в)

Алгоритм для вбудовування ЦВЗ складається з наступних кроків: визначаються межі криволінійної фігури, здійснюється відображення фізичної області на прямокутник, формується матриця яскравості пікселів, остання розбивається на блоки, до яких застосовуються сингулярні розклади,

модифікуються сингулярні числа блоків за допомогою бітової послідовності ЦВЗ, виконується зворотне відображення яскравості пікселів у фізичну область.

На рис.1, як приклад, зображено спеціальну схему: фігура – модельна область – область комплексного потенціалу.

Висновки. Запропонований підхід забезпечує високу робастність ЦВЗ до топологічних атак та стиснення JPEG. Використання конформної сітки як геометричного ключа забезпечує можливість надійно захищати приховану інформацію від вилучення.

1. Бомба А.Я., Бойчура М.В. Методи комплексного аналізу в задачах ідентифікації: монографія. Рівне: НУВГП, 2020. 188 с.
2. Liu R., Tan T. An SVD-based watermarking scheme for protecting rightful ownership. *IEEE Transactions on Multimedia*. – 2002. – V. 4, №1. – P. 121-128.

Проблематика узагальненої оцінки методів криптографічного захисту інформації

УДК 004.056.55

Віра Тітова[1], Володимир Анікін[2]

Хмельницький національний університет,
[1]titovav@khmnu.edu.ua, [2]anikin_volodymyr@khmnu.edu.ua

Існує значна розбіжність в підходах до узагальненої оцінки методів криптографічного захисту інформації (КЗІ). Аналіз досвіду проведення як вітчизняних, так і зарубіжних криптографічних конкурсів, в межах яких очевидною була необхідність строгого критеріального порівняння запропонованих методів-претендентів між собою, показує неузгодженість підходів до такої порівняльної оцінки. Часто таке порівняння методів криптографічного захисту фактично зводиться до порівняння засобів КЗІ, оскільки у якості критеріїв порівняння обирається, як приклад, швидкодія роботи, чи інша фізична величина, що емпірично вимірюється на прототипі засобу КЗІ. Подібний підхід простежувався, зокрема, у конкурсних процедурах, пов'язаних із вибором стандарту ДСТУ 7624:2014 «Калина», де поряд із криптографічними властивостями враховувалися також показники ефективності реалізацій.

З одного боку, такий підхід може бути виправданий з точки зору прикладної спрямованості порівняння, де той чи інший метод КЗІ не несе жодного практичного сенсу у суто теоретичному вимірі. Проте, з іншого боку, коли необхідність узагальненого порівняння виникає в ході певних науково-дослідних робіт, в межах яких, наприклад, є необхідність підібрати методи КЗІ як компонент деякої більш глобальної системи, а об'єктом оцінювання виступають не готові засоби, чи навіть їх прототипи, а виключно теоретичні абстракції, узагальнені характеристики яких необхідно порівняти, то згаданий емпірично-орієнтований підхід не може бути застосований.

Одним із аргументів, щодо принципової розбіжності у підходах порівняння методів та засобів КЗІ є те, що при дослідженні кількох різних реалізацій засобів

КЗІ, на базі одного і того ж самого методу, результати їх порівняння між собою можуть принципово відрізнятися. Це може бути зумовлено різними платформами, архітектурами, рівнями реалізацій, ступенями оптимізації та іншими чинниками, що є релевантними по відношенню до конкретного створеного засобу КЗІ.

Також в ході проєктування засобів КЗІ можуть бути створені додаткові вразливості, пов'язані із особливостями платформи, середовища чи будь-яких інших сторонніх каналів. Проте ці вразливості не будуть релевантні із теоретичними засадами функціонування методів КЗІ, що лежать в основі проблемних рішень.

Для вирішення зазначеної проблематики пропонується розробити критеріальну систему узагальненої оцінки методів КЗІ, яка б не опиралась в своїх розрахунках на ті чи інші емпіричні властивості прототипів, особливості реалізацій та інші практичні характеристики.

Принципово невірним було б рішення по створенню такої системи теоретичної оцінки виключно шляхом відкидання від інших загальноприйнятих систем оцінювання тих критеріїв, що базуються на емпіричних властивостях реалізацій засобів КЗІ, оскільки навіть суто теоретична узагальнена оцінка повинна зберігати широку різнобічність критеріїв, для того щоб результати такої оцінки були справді інформативними.

Для створення системи узагальненої оцінки методів КЗІ пропонується провести аналіз найбільш поширених критеріїв оцінки криптографічних примітивів, систем та протоколів, умовно розподілити їх на дві основних групи: ті, що спрямовані на теоретично-математичні властивості, та ті, що спрямовані на властивості конкретних реалізацій.

Перша група критеріїв може бути використана як базова складова запропонованого узагальненого методу оцінки.

Друга група потребує конвертації критеріїв у відповідники, що зберігають логічну сферу спрямування оцінки, проте опираються не на емпіричні властивості реалізації, а на деякі їх теоретично-математичні аналоги. Так, наприклад, критерій швидкодії роботи може бути приведено до критерію обчислювальної складності алгоритму тощо.

Для забезпечення гнучкості така критеріальна система може передбачати використання вагових коефіцієнтів. Вони можуть задаватися в усередненому варіанті або коригуватися відповідно до вимог конкретної задачі, якщо певний критерій має підвищену або знижену значущість.

Використання нормалізованих значень критеріїв і вагових коефіцієнтів дозволить отримати формалізований інтегральний показник, придатний для порівняння з іншими результатами, отриманими за тією самою процедурою.

Таким чином, проблематика узагальненої оцінки методів КЗІ є актуальною, в тому числі у вітчизняному науковому полі, оскільки станом на зараз є досить поширеною практика змішування оцінки методів та засобів КЗІ. Саме по собі таке змішування не є проблемою та може бути виправдане у вирішенні прикладних задач, в яких важливим є комплекс теоретичних та емпіричних властивостей об'єкту дослідження. Проте в інших задачах, наприклад у сфері теоретичного моделювання систем або протоколів КЗІ, де необхідна

узагальнена оцінка складових криптографічних елементів, без прив'язки до їх реалізацій, розділення підходів оцінки методів та засобів є необхідним. Наведені пропозиції щодо створення такої узагальненої системи оцінювання можуть лягти в основу окремого методу.

1. Prvulovic P, Radosavljevic N, Babic D, Drajic D. HERMEES: A Holistic Evaluation and Ranking Model for Energy-Efficient Systems Applied to Selecting Optimal Lightweight Cryptographic and Topology Construction Protocols in Wireless Sensor Networks. Sensors. 2025; 25(9):2732.
2. The concept of nonlinear cryptographic primitives, their steganographic applications, and related areas of use / Volodymyr Anikin, Serhii Lienkov, Ihor Muliar, Volodymyr Dzhulii, Oleksandr Sieliukov, Yaroslav Melnyk. EUREKA: Physics and Engineering. 2025. № 5. P. 190-204.

Сучасні підходи до безперервної автентифікації користувачів на основі динаміки рухів комп'ютерної миші

УДК 004.056.5:57.087.1

Олександр Корченко[1], Антон Герасименко[2], Імад Ірейфідж[3]

Державний університет інформаційно-комунікаційних технологій,
[1]agkorchenko@gmail.com, [2]anton.hrsmnk@gmail.com,
[3]dr.imad.education@gmail.com

Традиційні методи автентифікації (паролі, токени) стають вразливими до методів соціальної інженерії та витоку даних. Поведінкова біометрія пропонує концепцію безперервної автентифікації (Continuous Authentication), яка дозволяє верифікувати особу не лише в момент входу в систему, а протягом усієї сесії роботи. Рух миші є унікальним для кожної людини через індивідуальні особливості нейромоторної координації, що робить його перспективним об'єктом дослідження.

Методологія вилучення ознак (Feature Extraction) при автентифікації користувачів за рухом миші. Для ідентифікації користувача координати курсора x та y перетворюються на набір статистичних та кінематичних ознак. Основні параметри включають: швидкісні показники (миттєва швидкість v та прискорення a), траєкторні ознаки (кривизна руху, кутова швидкість та "джиттер", тобто мікротремтіння), часові ознаки (час реакції між появою стимулу та початком руху, тривалість пауз - dwell time), операційні ознаки (частота кліків, швидкість подвійного натискання, манера прокручування коліщатка).

Основні алгоритми машинного навчання та класифікації при автентифікації користувачів за рухом миші. Сучасні дослідження фокусуються на двох підходах до навчання моделей:

Однокласова класифікація (One-Class Classification): Модель навчається лише на даних легітимного користувача (наприклад, алгоритм Isolation Forest або One-Class SVM). Система шукає аномалії, які свідчать про те, що за комп'ютером перебуває інша особа.

Глибоке навчання (Deep Learning): Використання рекурентних нейронних мереж (LSTM) та згорткових мереж (CNN). Ці архітектури здатні вловлювати складні часові залежності в послідовності рухів, які неможливо описати простими статистичними формулами.

За даними актуальних досліджень, показник помилкового відхилення (FRR) та помилкового допуску (FAR) у таких системах варіюється в межах 2–7%, що є достатнім для використання у ролі другого фактору захисту.

Проаналізуємо недоліки традиційної автентифікації та поведінкової біометрії.

Таблиця 1

Недоліки традиційної автентифікації та поведінкової біометрії

Аспект	Традиційний пароль	Поведінкова біометрія
Стійкість до фішингу	Низька (можна виманити)	Абсолютна (неможливо передати)
Час дії	Тільки при вході	Безперервно (кожні 10-30 секунд)
Зручність (UX)	Вимагає зусиль від користувача	Повністю прозоро (Zero Friction)
Спроба злому	Брутфорс або підбір	Вимагає робота-маніпулятора, що імітує людину

Але поведінкова біометрія має і недоліки. Попри високу ефективність, існують фактори, що знижують точність розпізнавання:

- Hardware-залежність: Зміна роздільної здатності екрана або перехід з миші на тачпад радикально змінює поведінковий профіль.
- Психофізіологічний стан: Стрес, втома або хвороба користувача впливають на мікромоторику.
- Контекст діяльності: Рухи миші під час гри суттєво відрізняються від рухів під час роботи в офісних програмах.

Висновки: Динаміка рухів миші є надійним та маловитратним методом біометричної ідентифікації. Найбільш перспективним напрямом розвитку є створення гібридних моделей, що поєднують аналіз рухів миші з динамікою натискання клавіш (Keystroke Dynamics), що дозволяє досягти майже нульового показника помилок у корпоративних мережах.

1. Antal, M.; Egyed-Zsigmond, E. Intrusion detection using mouse dynamics. IET Biom. 2019, 8, 285–294.
2. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. The performance of LSTM and BiLSTM in forecasting time series. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 3285–3292.
3. Ahmed, A.A.E.; Traore, I. A new biometric technology based on mouse dynamics. IEEE Trans. Dependable Secur. Comput. 2007, 4, 165–179.

Підвищення обчислювальної ефективності криптосистеми Рабіна у кільці гауссових цілих чисел

УДК 519.7

Андрій Алілуйко[1]

Західноукраїнський національний університет, [1]aliluyko82@gmail.com

Криптосистема Рабіна є високоефективною завдяки швидкості шифрування та стійкості, що базується на складності факторизації та пошуку квадратного кореня за модулем. Проте її реалізація як у кільці цілих, так і гауссових чисел ($A = a + bi,\ a,b \in Z$) має недоліки: високу часову складність дешифрування через використання КТЗ (зокрема, пошук оберненого елемента) та обмеження у виборі параметрів (зокрема, вибір цілих чисел Блюма). Тому актуальним завданням є розробка нових підходів до реалізації системи в кільці гауссових чисел для оптимізації обчислювальних витрат.

Аналогічно до цілочисельної асиметричної криптографії, можна розглядати криптосистему Рабіна, коли повідомлення M та два прості числа p та q є цілими комплексними числами. Число $N = pq$ виступає відкритим ключем, а p та q – закритим. Шифрування відкритого повідомлення M відбувається за формулою $C = M^2 \bmod N$. При дешифруванні шифр тексту C вводяться додаткові допоміжні величини μ та ν: $\mu = C \bmod p,\ \nu = C \bmod q$. Для знаходження M необхідно знайти квадратні корені x та y за модулями p та q: $x^2 \bmod p = \mu,\ y^2 \bmod q = \nu.$

У результаті утворюються чотири системи рівнянь ($i = \overline{1,4}$):

$$\begin{cases} M_i \bmod p = \pm x, \\ M_i \bmod q = \pm y. \end{cases} \quad (1)$$

Один з розв'язків (1), пошук якого здійснюється на основі КТЗ, і буде шуканим відкритим повідомленням M.

Для уникнення ресурсномістких операцій пошуку оберненого елемента за комплексним модулем при застосуванні КТЗ запропоновано вибирати ключі p та q, які утворюють досконалу або модифіковану досконалу форму [1].

Для спрощення знаходження комплексних коренів за комплексним модулем при дешифруванні розроблено алгоритм розв'язування конгруенції $x^2 \equiv c \bmod \pi$, в якій комплексний модуль має норму виду $N(\pi) = 8k + 5, k \in Z$.

1. Aliluiko A., Kasianchuk M., Dziubanovska N., Netrebiak M. Construction of Perfect Form of Residue Number System for Gaussian Integers to Asymmetric Cryptosystems, *15th International Conference on Advanced Computer Information Technologies (ACIT)*, Sibenik, Croatia, 2025, pp. 471-475.

Заснований на DHT ефективний метод стеганоперетворення

УДК 004.056.5

Ірина Борисенко[1], Ігор Якименко[2]

Національний університет «Одеська політехніка», [1]borisenko.i.i@op.edu.ua,
Західноукраїнський національний університет, [2]iyakymenko@ukr.net

Значна кількість сучасних стегометодів використовує дискретне перетворення Адамара (DHT) зображення-контейнера для вбудовування повідомлення [1,2]. Існує багато схем приховування повідомлення в коефіцієнтах DHT, в основу яких покладено розбивку контейнера на блоки розміром *n* x *n*, а стегоалгоритми використовують різні стратегії приховування елементів повідомлення в кожному блоці [3]. Проте вбудовування відбувається переважно послідовно, а рішення приймається локально. У результаті алгоритм не завжди знаходить найкращий глобальний варіант розподілу блоків повідомлення по блоках контейнера. Саме ця проблема визначає втрату потенційної ефективності алгоритму. Якщо для кожного фрагмента повідомлення існує кілька можливих блоків контейнера, в яких він може бути розміщений із невеликою кількістю корекцій, то природно постає задача вибору не просто будь-якого допустимого варіанта, а найкращого з погляду всієї сукупності можливих вбудовувань.

Метою роботи є оптимізація розподілення блоків повідомлення по блокам контейнера, в сенсі їх подібності по заданому критерію, та дослідження стійкості одержаного стегоповідомлення до статистичних атак стегоаналізу.

Можна виділити основні критерії подібності: косинусна подібність (Cosine similarity), яка показує наскільки напрямок вектора повідомлення збігається з напрямком вектора контейнера; кореляція (Pearson correlation), яка враховує лінійну залежність між коефіцієнтами блоку контейнера і повідомлення; енергетичний критерій (Energy matching).

Для стеганографії найчастіше використовують кореляцію або косинусну подібність, оскільки вони показують, наскільки добре повідомлення відповідає структурі контейнера.

Алгоритм оптимізації (укрупнена схема).

1.Обчислити спектр повідомлення DHT для кожного блоку M_i.

2. Обчислити спектр контейнера для кожного блоку C_j.

3. Для кожного M_i знайти блок C_j, який має найбільшу: кореляцію або косинусну подібність.

4. Вбудувати M_i саме у цей найбільш подібний C_j.

Зауважимо, що для M_i може знайтись не один блок контейнера, що відповідатиме заданому критерію подібності, тому на третьому кроці алгоритму пропонується для кожного M_i знайти усі блоки C_j, які задовільнять заданому критерію з одночасною побудовою таблиці, рядки і стовпці якої відповідатимуть номерам блоків повідомлення та контейнера, а елементами таблиці будуть значення заданого критерія подібності. Проблема зводиться до вирішення задачі оптимального призначення, для розв'язку якої існує достатньо ефективних алгоритмів. Коли призначення визначено, тобто сформувались пари

$M_i \rightarrow C_j$, запускається процес вбудовування з одночасним формуванням ключа *K*, який є парами (i,j), де i,j – номери блоків.

Експериментальні дослідження показали: базове вбудовування (послідовне блок-за-блоком) дає збурення контейнера Δ RMS (Root Mean Square) =4,2, PSNR (дБ)=38, BER (%)=12; оптимізоване вбудовування Δ RMS=2,7 PSNR (дБ)=42, BER (%)=6.

Стійкість до деяких статистичних атак стеганоаналізу представлена графіками на рис.1.

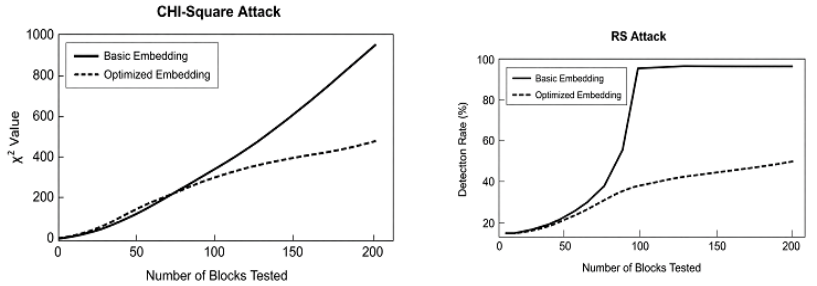


Рис.1. Графіки стійкості стего до атаки Хі-квадрат (зліва) та RS-атаки (справа)

Аналіз побудованих графіків показує, що оптимізоване вбудовування блоків повідомлення знижує статистичні відхилення, тому атака CHI-square менш ефективна. Це підтверджує, що адаптивний вибір блоків за критеріями подібності дає реальний виграш у стійкості. Оптимізоване вбудовування також знижує статистичну різницю між групами пікселів (Regular/Singular), тому RS-тест виявляє менше аномалій. Це підтверджує, що адаптивне вбудовування блоків за критеріями подібності підвищує стійкість не лише до CHI-square, а й до RS-атаки.

Алгоритм відновлення вбудованого повідомлення не вимагає вихідного зображення завдяки ключу *K* («сліпий»). Виконується зворотній процес: перетворення DHT обраного за ключем блока контейнера, застосування оберненого алгоритму вбудовування, а потім DHT до одержаного спектра.

Таким чином, оптимізація за критерієм подібності менше спотворює контейнер, а отже підвищує стійкість до статистичних атак стегоаналізу, дає вищу точність відновлення.

1. Zhang Y. Q., Zhong K., Wang X. Y. High-Capacity Image Steganography based on Discrete Hadamard Transform. IEEE Access. 2022. Vol. 10. P. 65141-65155. doi: 10.1109/ACCESS.2022.3181179
2. Helal S., Salem N. A Hybrid Watermarking Scheme Using Walsh Hadamard Transform and SVD. Procedia Computer Science. 2021. Vol. 194. P. 246-254. https://doi.org/10.1016/j.procs.2021.10.080
3. Prabha K., Sam I. S. A novel blind color image watermarking based on Walsh Hadamard Transform. Multimedia Tools and Applications. 2020. Vol. 79, No. 9. P. 6845-6869. doi: 10/1007/s11042-019-08212-w

Багатокритеріальне оцінювання ризиків інфраструктури навчальних кіберполігонів методом аналізу ієрархій

УДК 004.056

Андрій Сидор[1], Михайло Бойчура[2], Володимир Герус[3]

Національний університет водного господарства та природокористування,
[1]a.i.sydor@nuwm.edu.ua, [2]m.v.boichura@nuwm.edu.ua, [3]v.a.gerus@nuwm.edu.ua

Інтеграція практично-орієнтованих підходів до підготовки фахівців з кібербезпеки вимагає розгортання спеціалізованих тренувальних середовищ – кіберполігонів та платформ для проведення змагань у форматі Capture The Flag (CTF) [3]. Специфіка такої інфраструктури полягає в архітектурі "vulnerable-by-design" (вразливий за задумом). Наявність навмисно залишених прогалин у безпеці віртуальних машин, необхідність надання студентам широких прав доступу для тестування на проникнення, а також використання процедурно згенерованих середовищ створюють унікальний і вкрай динамічний ландшафт загроз.

Адміністрування таких навчальних лабораторій супроводжується постійним балансуванням між забезпеченням доступності сервісів та ізоляцією шкідливої активності учасників. Оскільки технічні та людські ресурси, виділені на підтримку академічних кіберполігонів [2], зазвичай жорстко обмежені, виникає гостра потреба у математично обґрунтованій пріоритезації загроз. Класичні методи оцінки ризиків [5] часто виявляються недостатньо гнучкими для врахування одночасно технічних, фінансових та репутаційних чинників академічного середовища, що зумовлює доцільність застосування багатокритеріальних експертних систем [6].

Метою даного дослідження є розробка комплексної моделі пріоритезації ризиків інфраструктури навчальних кіберполігонів на основі методу аналізу ієрархій (MAI) Томаса Сааті [1]. Враховуючи високу математичну трудомісткість та ризик людської суб'єктивності під час заповнення масивів матриць попарних порівнянь, у роботі запропоновано інноваційний підхід – використання ансамблю великих мовних моделей (LLM) у ролі незалежної експертної групи [4]. Синтез результатів оцінювання п'ятьма архітектурно різними моделями (Gemini, Qwen, Grok, ChatGPT та Deepseek) дозволяє нівелювати статистичні відхилення окремих нейромереж та отримати максимально об'єктивний вектор пріоритетів.

Сформована ієрархічна модель оцінювання складається з трьох рівнів [7]. Для всебічного аналізу інцидентів було виділено 8 критеріїв, що охоплюють технічні, ресурсні та репутаційні аспекти функціонування полігону: вплив на доступність, порушення цілісності, витік конфіденційних даних, масштаб ураження, складність виявлення інциденту, ймовірність його успішної експлуатації, ресурсоємність відновлення інфраструктури та загальні академічні ризики. У якості альтернатив досліджено 9 критичних сценаріїв: втеча з ізольованого середовища, компрометація скорингової платформи, внутрішня відмова в обслуговуванні ресурсоємними скриптами учасників, порушення мережевої ізоляції між командами, витік тренувальних завдань до

початку змагань, викрадення адміністративних облікових записів, експлуатація вразливостей керуючого ПЗ (гіпервізорів), зовнішня DDoS-атака та зараження мережі деструктивним програмним забезпеченням.

Для перевірки логічної транзитивності згенерованих нейромережами експертних суджень, на кожному етапі попарних порівнянь здійснювався розрахунок відношення узгодженості (CR). Згідно з методологією MAI, розраховані матриці вважалися математично коректними та допускалися до подальшого усереднення лише за умови виконання нерівності:

$$CR = CIRI \leq 0.1 \quad (1)$$

де RI – табличний індекс випадкової узгодженості для відповідної розмірності матриці, а CI – індекс узгодженості, який визначається за формулою:

$$CI = max\text{-}nn\text{-}1 \quad (2)$$

де max – найбільше власне значення матриці попарних порівнянь, n – розмірність матриці (кількість порівнюваних елементів). Завдяки строгому контролю параметра CR було забезпечено високу валідність кінцевих нормалізованих оцінок.

На етапі моделювання було сформовано уніфіковані запити (промпти) для кожної з п'яти залучених великих мовних моделей (Gemini, Qwen, Grok, ChatGPT, Deepseek). Моделі діяли як незалежні експерти, виконуючи попарне порівняння 8 критеріїв та 9 альтернатив за 9-бальною шкалою Сааті. Для уникнення математичних аномалій результати кожної ітерації автоматично перевірялися на узгодженість CR0.1. На основі отриманих локальних векторів пріоритетів було розраховано глобальні пріоритети для кожної загрози у розрізі кожної моделі. З метою мінімізації впливу специфічних галюцинацій або "зсувів" окремих нейромереж, фінальний вектор пріоритетів було обчислено як середнє арифметичне значення глобальних пріоритетів усього ансамблю моделей.

Після виконання всіх етапів ієрархічного синтезу та обчислення середнього арифметичного значень, отриманих від п’яти незалежних LLM-експертів, було сформовано підсумковий вектор глобальних пріоритетів. Результати ранжування досліджуваних загроз інфраструктури кіберполігону за рівнем їхньої критичності наведено у табл. 1.

Таблиця 1

Глобальні пріоритети загроз інфраструктури кіберполігону

Ранг	Назва загрози (Альтернатива)	Критерії впливу (>0,1)	Глобальний пріоритет
1	Зараження деструктивним ПЗ (Wiper / Ransomware Infection)	K1, K2, K5, K6, K8	0,2022840
2	Внутрішня відмова в обслуговуванні (Internal DoS / Resource Exhaustion)	K7, K8	0,1295191

3	Компрометація адміністративних доступів (Admin Credential Compromise)	К1, К3, К4, К5, К6, К7	0,1255742
4	Зовнішня DDoS-атака (External DDoS Attack)	К8	0,1237373
5	Вразливості керуючої інфраструктури (Management Infrastructure Vulnerabilities)	К2, К4, К5, К6	0,1031579
6	Компрометація скорингової системи (Scoring Platform Compromise)	К1, К3, К4, К5, К6	0,0951738
7	Витік завдань або рішень (Task / Flag Leakage)	К1, К3, К4, К7	0,0866986
8	Втеча з ізольованого середовища (Sandbox / VM Breakout)	К2, К3, К5	0,0730205
9	Порушення мережевої ізоляції (Network Isolation Failure)	К7	0,0608347

де: К1: Академічні та репутаційні ризики; К2: Ресурсоємність відновлення; К3: Складність виявлення; К4: Витік конфіденційних даних; К5: Масштаб ураження; К6: Порушення цілісності; К7: Ймовірність успішної експлуатації; К8: Вплив на доступність.

Дані таблиці вказують на те, що загроза зараження деструктивним ПЗ (0,2023) майже вдвічі випереджає наступні за рангом ризики. Це пояснюється тим, що при оцінюванні за критеріями "Масштаб ураження" та "Ресурсоємність відновлення" ансамбль моделей одностайно надав цьому сценарію найвищі бали. Водночас, порівняно низький пріоритет втечі з ізольованого середовища (0,0730) свідчить про те, що експерти вважають таку атаку складною в реалізації для пересічного учасника навчального процесу.

Для наочного представлення отриманої ієрархічної структури та виявлення розривів між найбільш критичними та другорядними ризиками було побудовано гістограму розподілу ваг (рис. 1). Така візуалізація дозволяє чітко розмежувати загрози за зонами їхнього впливу на безпеку полігону.

Графічне представлення результатів (рис. 1) підкреслює наявність так званої «червоної зони» – групи з п'яти загроз, чиї пріоритети перевищують значення 0,1. Саме ці напрямки потребують впровадження автоматизованих систем моніторингу та посиленого контролю прав доступу. Візуальний розрив між п'ятою та шостою позиціями вказує на логічну межу, після якої ризики переходять у категорію менш пріоритетних для першочергового фінансування систем захисту.

Проведене багатокритеріальне оцінювання дозволило виявити найбільш критичні вектори загроз для навчальних платформ. Згідно з усередненими даними (Таблиця 1), найвищий глобальний пріоритет має загроза зараження інфраструктури деструктивним ПЗ (0,2023), що зумовлено катастрофічним масштабом ураження та високою ресурсоємністю відновлення. Другу та третю позиції посідають внутрішня відмова в обслуговуванні (0,1285) та компрометація адміністративних доступів (0,1256), що пояснюється високою

ймовірністю випадкових помилок студентів під час сканування мережі та критичним впливом на цілісність навчального процесу.

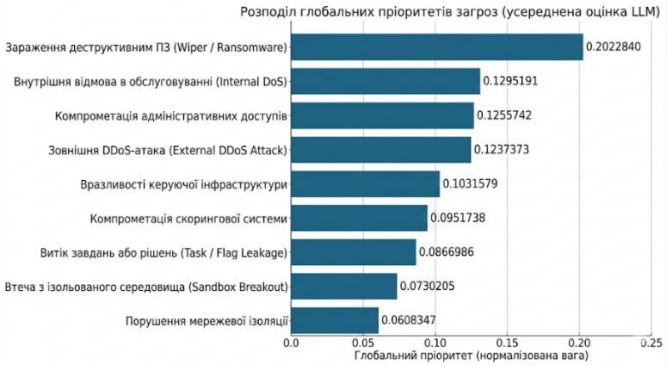


Рис.1. Узагальнена класифікація методів перехоплення інформації у СКК

Отримані результати формують науково обґрунтовану доказову базу для ефективного розподілу технічних та фінансових ресурсів при розгортанні спеціалізованих аудиторій та лабораторій кіберзахисту. Наявність такої верифікованої матриці ризиків є критично важливою при стандартизації безпекових процедур та узгодженні архітектури тренувальних комплексів у рамках співпраці з профільними державними структурами, зокрема Держспецзв'язком, що дозволяє вивести підготовку фахівців на якісно новий рівень захищеності.

1. Saaty T. L. Decision making with the analytic hierarchy process. International journal of services sciences. 2008. Vol. 1, No. 1. P. 83–98.
2. Yamin M. M., Katt B., Gkioulos V. Cyber ranges and security testbeds: Scenarios, functions, tools and architecture. Computers & Security. 2020. Vol. 88. P. 101636
3. Lyu Y., Dotson L., Draves N., Zhang A. CTF for education. arXiv preprint. 2026. arXiv:2601.17543. URL: https://arxiv.org/abs/2601.17543.
4. Zheng L., Chiang W. L., Ying Y. et al. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. Advances in Neural Information Processing Systems. 2023. Vol. 36. P. 46595–46623.
5. Гаврик С., Шишацький А., Сова О. та ін. Методика оцінювання кібербезпеки в інформаційно-телекомунікаційній системі спеціального призначення. Системи управління, навігації та зв'язку. 2020. Вип. 4 (62). С. 109–114.
6. Грищук Р. В., Даник Ю. Г. Основи кібербезпеки: навчальний посібник. Житомир: ЖВІ НАУ, 2016. 636 с.
7. Хохлачова Ю. Є., Венгерський П. С. Кількісна оцінка ризиків інформаційної безпеки на основі багатокритеріальних методів прийняття рішень. Сучасний захист інформації. 2020. № 3. С. 34–41.

Метричний аналіз та обчислювальна стійкість цільових словників паролів

УДК 004.056.52 (519.1)

Сергій Бабич[1], Андрій Сидор[2], Петро Голуб[3]

Національний університет водного господарства та природокористування, [1]s.v.babych@nuwm.edu.ua, [2]a.i.sydor@nuwm.edu.ua, [3]p.p.holub@nuwm.edu.ua

Дослідження спрямовані на перетворення хаотичної генерацію варіантів у керований і обчислювально ефективний процес [1] синтезу ефективних цільових словників. Аби дослідження тримати в межах актуальності, необхідно оцінювати ефективність перетворення даних з «відомостей про сутність» (вхідні дані у форматі OSINT – дослідження) у подальші токени (стадія 1), паролі (стадія 2) та їх об'єднання – безпосередньо словник (стадія 3).

Центральним об'єктом запропонованої моделі є парольний токен — мінімальна семантично-персональна одиниця, вилучена з OSINT-профілю [2]. Наукова новизна підходу полягає у відмові від розуміння токена як лінгвістичної морфеми (наприклад, у алгоритмах BPE чи SentencePiece) на користь суб'єктивно значущої сутності (ім'я, дата, місце). Для структурування OSINT-даних запропоновано семикласову таксономію (T1–T7) [3], що включає ідентифікатори, часові, реляційні, просторові, тематичні, числові та контекстуальні дані. Описавши об'єкт оцінки ефективності, варто перейти до компоненти дослідження, що стосується вже саме оцінки ефективності.

Аналітичний огляд досліджень у сфері оцінки ефективності словників паролів демонструє поступовий перехід від класичних методів до складних моделей, що враховують як структурні характеристики паролів, так і контекстні дані користувачів. Розвиток цієї тематики відбувся у працях, де були запропоновані моделі TarGuess, які інтегрують відкриті дані (OSINT), включно з персональними атрибутами та інформацією із соціальних мереж, для побудови цільових словників. Ефективність таких словників значно зростає завдяки використанню реальних цифрових слідів користувачів, що підтверджує важливість ймовірнісного та статистичного аналізу у вимірюванні їх стійкості. Водночас ці роботи піднімають питання етики та приватності, адже використання персональних даних у словниках може створювати додаткові ризики [4]. Трендом останніх років стало застосування нейромережевих моделей: запропоновано PassGAN, який використовує генеративні змагальні мережі для створення словників. Цей підхід дозволяє моделювати розподіл паролів без явних правил, що робить словники більш адаптивними та здатними відображати реальні патерни користувачів. Метричний аналіз у цьому випадку включає оцінку ентропії та варіативності згенерованих словників, що дозволяє порівнювати їх ефективність із класичними методами [5].

Підсумувавши дослідження інших авторів щодо оцінки ефективності словників паролів загалом та виокремивши необхідне для нашого дослідження, що стосується цільових словників, варто зауважити, що останнє десятиліття ознаменувалося переходом від простої частотної статистики до складних моделей представлення знань, де базові показники: CR (Coverage Rate) та DSR

(Dictionary Size Ratio) залишаються основними для оцінки компактності та влучності. Але варто згадати і використання математичної близькості за метрики Дамерау-Левенштейна [6], котра стала стандартом для моделювання транспозицій (перестановок сусідніх знаків), що відображає механічні помилки користувачів. Щодо порівняння моделей, звернемо увагу, що класичні ймовірнісні граматики (PCFG) розбивають пароль на сегменти [L][D][S], але ігнорують семантичний зв'язок. Нейромережеві підходи, такі як PassGAN та PASSLLM, демонструють здатність вивчати логіку витоків, проте в цільових атаках вони у 10 000 разів повільніші за комбінаторні методи та часто страждають від перенавчання (overfitting) [7]. Ідеї Representational Learning (PLR), в своєю чергою, використовують латентний простір паролів для генерації навколо опорних точок (pivots), проте вимагають значних обчислювальних ресурсів.

Як висновок, варто звернути увагу, що в межах даного представлення – окреслено бачення команди щодо оцінки ефективності сформованих словників, при формуванні яких буде використано запропоновану методику СРП з роботи [1]. Але, вже можна виокремити те, що перехід до інтегрованої моделі оцінки, що поєднує OSINT-профілювання, семантичну токенізацію та метод перманентної декомпозиції, дозволяє перетворити аудит парольної політики з хаотичного підбору на науково обґрунтовану процедуру. Запропонований підхід не лише підвищує точність генерації при цільовому використанні, а й забезпечує обчислювальну стійкість систем захисту в умовах зростання цифрового сліду користувачів.

1. С. В. Бабич, «Інформаційна технологія складання розкладу занять згідно перманентної декомпозиції», дисертація кандидата технічних наук, Хмельницький національний університет, Хмельницький, Україна, 2023.
2. M. Bazzell, Open Source Intelligence Techniques: Resources for Searching and Analyzing Online Information, 9th ed. IntelTechniques, 2022.
3. B. Rader, “Targeted password cracking with OSINT data,” Purdue School of Engineering & Technology, IUPUI, 2023.
4. M. Weir, S. Aggarwal, B. de Medeiros, and B. Glodek, “Password cracking using probabilistic context-free grammars,” in Proc. IEEE Symp. Security and Privacy (SP), Oakland, CA, USA, 2009, pp. 391–405.
5. Y. Xie, J. Wang, and J. Yan, “TarGuess II: A statistical framework for targeted password guessing,” in Proc. ACM Conf. Computer and Communications Security (CCS), London, UK, 2020, pp. 1–14.
6. F. J. Damerau, “A technique for computer detection and correction of spelling errors,” Communications of the ACM, vol. 7, no. 3, pp. 171–176, 1964.
7. B. Hitaj, P. Gasti, G. Ateniese, and F. Perez-Cruz, “PassGAN: A deep learning approach for password guessing,” in Proc. Int. Conf. Applied Cryptography and Network Security (ACNS), Bogota, Colombia, 2019, pp. 217–237.

Аналіз витоків паролів на наявність патернів та можливості їх субслівної токенізації

УДК 004.056

Сергій Бабич[1], Петро Голуб[2], Богдан Слив'як[3]

Національний водного господарства та природокристування,
[1]s.v.babych@nuwm.edu.ua, [2]p.p.holub@nuwm.edu.ua, [3]slyviak_ak23@nuwm.edu.ua

Паролі використовуються повсюдно як засіб автентифікації користувачів, відповідно складність паролів є одним із основних факторів захищеності інформаційних систем. Кореляції ж у рамках однієї із останніх збірок витоків паролів, а саме RockYou2024 демонструють стабільну наявність популярних патернів, що стосуються персональної інформації чи цифрового сліду людини [1]. Урядові та відомчі мережі, включно з індустріальними системами управління (SCADA), також залишаються вразливими до використання передбачуваних паролів, що становить загрозу критичні інфраструктурі [2, 3].

Метою роботи є дослідження тенденцій та кореляцій у масштабних витоках паролів, а також використання отриманих результатів для подальшого дослідження граматичних особливостей парольних фраз та їх субслівної токенізації, задля верифікації складності паролів базуючись на цифровому сліді людини. Відповідні результати можуть бути надзвичайно важливими для аудиту парольних політик всередині організації або створення релевантних цільових словників паролів на основі розвідки з відкритих джерел (OSINT) [4].

Новизна даного дослідження полягає у розширенні класичних підходів до побудови словників паролів та інтеграції українського соціокультурного і лінгвістичного контексту. Сучасні дослідження підтверджують, що традиційні моделі генерації паролів ігнорують вплив зовнішніх семантичних факторів (соціальних трендів, популярної лексики), що суттєво знижує їхню адаптивність [5]. Запропонований підхід субслівної токенізації враховуватиме унікальні регіональні аспекти, такі як транслітерацію кириличних слів, крос-розкладне введення та використання специфічних абревіатур.

Очікується, що це дозволить підвищити точність та швидкість підбору паролів. Даний фактор є важливим, оскільки грунтуючись на аналізі витоків паролів [2], варто відзначити, що за останні 15 років середня довжина паролів зросла з 8 до 10 символів, а їхня середня інформаційна ентропія — із 40 до 50 біт. Оцінка ентропії H здійснюється за класичною формулою Шеннона [3]:

$$E = -\sum_{i=1}^{n} P_i log_2 P_i \quad (1)$$

де P_i — імовірність появи i-го символу чи токена з доступного простору.

Загалом існує сукупність факторів, що дозволяє прослідкувати ускладнення паролів, проте залишається конструктивна слабкість через наявність передбачуваної інформації.

Для розв'язання поставленої задачі пропонується використовувати алгоритмічну модель генерації та скорингу токенів. Вона опрацьовує неструктурований масив вхідних OSINT-даних (імена, дати, локації, специфічний сленг) та виконує їх багаторівневу обробку. Процес включає вилучення базових сутностей, їх розбиття на субслівні одиниці (склади,

біграми), застосування алгоритмів транслітерації та додаткових мутацій. З метою формування оптимального словника, кожен токен отримує скорингову оцінку від 0 до 100 балів (табл. 1).

Таблиця 1

Критерії скорингової оцінки субслівної токенізації

Критерій	Опис параметрів токена	Макс. бал
Пріоритет джерела	Прямі збіги (імена, дати), ініціали	40
Локалізація	Транслітерація, крос-розкладне введення (kbswap), N-грами	30
Частотність	Мультипликатор ваги за частоту появи токена у вхідних OSINT-даних	20
Довжина	Оптимальний розмір токена для формування паролю (найвища вага 3-6 символів)	10

Отже, ускладнення паролів не усуває людського фактору, користувачі продовжують спиратися на власний цифровий слід та локальний контекст. Виокремлення граматичних особливостей формування парольних фраз і поєднання їх із субслівною токенізацією та OSINT-розвідкою дозволить створювати високорелевантні цільові словники паролів. Запропонована скорингова модель генерації токенів може бути ефективним інструментом для тестування на проникнення та комплексного аудиту безпеки інформаційних систем.

1. Слатвінська В., Бевза В., Вплив збою CrowdStrike на мега-витік паролів: чи є зв’язок? Ч. 1. Вісник Хмельницького національного університету. Технічні науки. – 2024. – № 4 (339). – С. 332-338. DOI: 10.31891/2307-5732-2024-339-4-52.
2. Rodrigues G.A.P. et al., From RockYou to RockYou2024: Analyzing Password Patterns Across Generations, Their Use in Industrial Systems and Vulnerability to Password Guessing Attacks. Journal of Internet Services and Applications. – 2025. – Vol. 16, No. 1. DOI: 10.5753/jisa.2025.5041.
3. Собина В.О., Тарадуда Д.В., Демент М.О., Захист інформації відомчої інформаційно-телекомунікаційної мережі за допомогою парольної системи. Проблеми надзвичайних ситуацій. – Харків: НУЦЗ України, 2021. – № 2 (34).
4. Різак М., Котик О., Використання OSINT для захисту персональних даних. The 14th International Scientific Conference «ITSec». – Тернопіль, 2025. – С. 165-167.
5. Yang X. et al., KAPG: Adaptive Password Guessing via Knowledge-Augmented Generation. – 2025.

Штучний інтелект та кібербезпека

УДК 004.08

Владислав Орбан[1]

Західноукраїнський національний університет, [1]v.orban@st.wunu.edu.ua

Сьогодні ми є свідками того, як технології формують не лише бізнес-середовище, а й архітектуру глобальної безпеки. І немає більш визначальної сили в сучасному цифровому світі, ніж штучний інтелект. Ще зовсім недавно системи кіберзахисту базувалися на сигнатурних методах, жорстких правилах брандмауерів та ручному аналізі логів. Зараз же ми спостерігаємо перехід до аналізу великих даних за допомогою різних моделей. ШІ став одночасно і помічником і загрозою для захисту інформаційних систем, і наша ключова мета як фахівців — зрозуміти, як ефективно володіти цим інструментом, щоб захистити нашу інфраструктуру від загроз нового покоління.

Перевагами ШІ є: традиційні системи виявлення вторгнень можуть генерувати лавини хибних спрацьовувань, серед яких буває важко виявити справжній інцидент. Саме тут на допомогу приходять алгоритми машинного навчання, де замість постійного оновлення сигнатур вірусів, модель тренується на масивах даних нормальної роботи мережі, і може виявляти навіть найменші відхилення, такі як: нестандартне використання протоколів, або ж спроби тунелювання трафіку, це дозволяє фіксувати Zero-Day атаки, тобто ті, що не відомі, наприклад, антивірусу. Штучний інтелект здатний не лише виявляти загрозу, а й автономно реагувати на неї. У разі фіксації інциденту система може миттєво ізолювати скомпрометований вузол, динамічно переналаштувати правила VLAN для блокування сегмента мережі або ж зупинити процеси, що виглядають підозріло, до того як спеціаліст з кібербезпеки встигне відкрити сповіщення про небезпеку.

Сучасні рішення використовують глибинне навчання для аналізу поведінки програм у оперативній пам'яті комп'ютера, ефективно блокуючи програми-вимагачі (Ransomware) на етапі спроби масового шифрування файлів. Але не варто забувати, що ШІ технології можуть допомогти хакерам зі зламами систем [2]. Через невпинний розвиток технологій, ми стикаємося з масштабуванням та ускладненням кібератак [3]. Великі мовні моделі дозволяють автоматизувати написання фішингових повідомлень. Зловмиснику не потрібно вивчати жертву власноруч, адже ШІ агрегує дані з відкритих джерел і генерує листи, які відрізнити від легітимних комунікацій майже неможливо. Ну і не варто забувати про технологію Deepfake, яку вже зараз використовують для омани корпоративних співробітників з метою несанкціонованого переказу коштів. Згідно даних представлених на світовому економічному форумі [1] загроза для ШІ становить 87%.

На фоні зростання масового використання технологій штучного інтелекту як і в роботі бізнесу, так і в звичайному житті, з'явилися віруси, які можуть адаптуватися до будь-якого середовища. За допомогою методів машинного навчання, такий код розуміє, коли він знаходиться в ізольованому середовищі для аналізу і «засинає», активуючи свій деструктивний функціонал лише після потрапляння на реальну робочу станцію.

Рис.1. Дослідження про загрози для ШІ від World Economic Forum

За допомогою ШІ можна за лічені хвилини проаналізувати величезні масиви коду або конфігурацій, виявляючи всі приховані вразливості, такі як SQL-ін'єкції або ж помилки в налаштуваннях доступу, що прискорює етап розвідки в десятки разів.

У зв'язку з тим, що все більше систем покладаються на системи безпеки на базі ШІ, зловмисник може атакувати саме його. Найнебезпечнішим методом є «отруєння даних». Таким чином, якщо зловмисник отримує доступ до датасетів, на яких тренується модель, він може непомітно внести зміни так, щоб навчити модель сприймати шкідливий трафік за безпечний. Це означає, що відтепер інженерам з кібербезпеки необхідно захищати не лише сервери, бази даних та мережеве обладнання. Ми маємо захищати саму математику, забезпечуючи цілісність наборів даних та стійкість ML-моделей до маніпуляцій.

Підсумовуючи, хочу зазначити, що штучний інтелект навряд зможе повністю замінити експертів з кібербезпеки, однак ті спеціалісти, які володіють знаннями та інструментарієм Data Science, зможуть з легкістю замінити тих, хто відмовляється адаптовуватися до нових реалій. У недалекому майбутньому кіберпростір належатиме тим, хто зможе найшвидше аналізувати дані та адаптуватися до нестандартних викликів нашого часу.

1. World Economic Forum. (2025, January 14). Global cybersecurity outlook 2026: Infographics. https://www.weforum.org/publications/global-cybersecurity-outlook-2026/infographics-global-cybersecurity-outlook-2026/
2. Deepstrike. (2025, January 1). Cybercrime statistics 2025: Trends, impact, and insights. https://deepstrike.io/blog/cybercrime-statistics-2025
3. Programs.com. (2024, December 31). AI cyberattack statistics: The role of artificial intelligence in modern cyber threats. https://programs.com/resources/ai-cyberattack-stats/

Методи шифрування на основі зміни модулів системи залишкових класів

УДК 004.056.55
Соломія Марчук[1], Mikolaj Karpinski[2], Михайло Касянчук[3]

[1,3]Західноукраїнський національний університет,
[2]University of the National Education Commission, Poland;
[1]sol.marchuk@gmail.com, [2]mikolaj.karpinski@uken.krakow.pl,
[3]kasyanchuk@ukr.net

Ключовими аспектами кібербезбеки є забезпечення конфіденційності, цілісності та доступності інформації [1]. Для шифрування перспективним напрямом є використання непозиційних систем числення. У системі залишкових класів (СЗК) цілі числа представляються як набір залишків від ділення цього числа на попарно взаємнопрості модулі [2]. Метою нашої роботи є розробка методів шифрування на основі зміни модулів СЗК.

У СЗК відкритий текст у вигляді цілого числа N представлений за допомогою невід'ємних залишків b_i від ділення цього числа на модулі p_i:

$$b_i = N \bmod p_i. \tag{1}$$

Попередньо вибрані параметри p_i повинні бути додатними і попарно взаємнопростими. Обов'язковою умовою є те, що їх добуток $P = \prod_{i=1}^{s} p_i$, де s - кількість модулів, повинен перевищувати значення числа N.

Зворотне перетворення в десяткову систему числення здійснюється за допомогою китайської теореми про залишки (КТЗ):

$$N = \left(\sum_{i=1}^{s} m_i M_i b_i \right) \bmod P, \tag{2}$$

де $M_i = \frac{P}{p_i} = p_1 p_2 \cdots p_{i-1} p_{i+1} \cdots p_s$ - добуток усіх модулів окрім і-го, коефіцієнти m_i є оберненими елементами за відповідними модулями і шукаються з конгруенцій:

$$m_i = M_i^{-1} \bmod p_i = \left(M_i \bmod p_i\right) - 1 \bmod p_i. \tag{3}$$

Із формули обчислення M_i випливає, що

$$(M_i m_i b_i) \bmod p_j = \{0, if\ i \neq j\ b_i, if\ i = j \tag{4}$$

Це означає, що кожен доданок в КТЗ (2) за модулем p_i рівний 0, окрім *i*-го, тому що M_i не містить в собі множника p_i.

В методі шифрування на основі зміни параметра M_{1i} обирається два набори модулів p_{1i} та p_{2i} таким чином, щоб кожен модуль з другого набору був більшим, ніж відповідний модуль з першого. Із цього випливає, що нові модулі

обов'язково більші, ніж залишки, отримані за модулями першого набору. При цьому набір p_{2i} повинен відповідати тим самим умовам, що і p_{1i}.

Для шифрування за допомогою КТЗ використовуються нові значення модулів p_i та параметрів M_i:

$$N' = \left(\sum_{i=1}^{s} m_{1i} M_{2i} b_{1i} \right) \bmod P_2 \quad . \tag{5}$$

При розшифруванні після ділення шифртексту N' на модулі p_{2i} отримуються змінені залишки b'_i. Істинні залишки обчислюються таким чином:

$$b_{1i} = (b'_i (m_{1i} M_{2i})^{-1}) \; mod \; p_{2i}. \tag{6}$$

Щоб отримати відкритий текст, потрібно знову використати КТЗ (2).

Для методу шифрування на основі зміни параметрів M_{1i} та m_{1i} аналогічно до попереднього обирається два набори модулів. Шифртекст отримується за допомогою КТЗ (2) при використанні нових значень p_i, M_i та m_i і набуває наступного вигляду:

$$N' = \left(\sum_{i=1}^{s} m_{2i} M_{2i} b_{1i} \right) \bmod P_2 \quad . \tag{7}$$

Оскільки m_{2i} та M_{2i} отримані з одного набору модулів, то при дешифруванні одразу отримуються значення початкових залишків b_{1i}. Відновлення відкритого тексту за допомогою КТЗ відбувається по модулях p_{1i}.

В методі шифрування на основі зміни одного модуля необхідно також обрати два набори, в яких відрізняється лише один модуль. Якщо перший набір складається із значень $p_{11}, p_{12}, p_{13}, \dots, p_{1s}$, а другий - з $p_{21}, p_{22}, p_{23}, \dots, p_{2s}$, то необхідно, щоб виконувались такі умови: $p_{21} > p_{11}, p_{12} = p_{22}, p_{13} = p_{23}, \dots, p_{1s} = p_{2s}$. Для шифрування рівняння КТЗ (2) набуває наступного вигляду:

$$N' = \left(\sum_{i=1}^{s} m_{2i} M_{1i} b_{1i} \right) \bmod P_2 \quad . \tag{8}$$

З рівності модулів випливає, що співвідношення (4) виконується для всіх елементів, крім першого рядка. Відновлення $b_{12}, b_{13}, \dots, b_{1s}$ відбувається за формулою (2). Значення b_{11} отримується за наступною формулою:

$$b_{11} = \left(\left(\sum_{i=1}^{s} b'_{1i} - \sum_{i=2}^{s} M_{1i} m_{2i} b_{1i} \right) \cdot (M_{11} m_{21})^{-1} \right) \bmod p_{21} \quad . \tag{9}$$

Обчисливши значення початкових залишків, відкритий текст можна відновити за допомогою КТЗ по модулях p_{1i}. Якщо змінити більше одного модуля, то при розшифруванні потрібно буде розв'язати систему діофантових рівнянь, в якій залишки b_i визначаються неоднозначно.

Дослідження підтвердило, що запропоновані методи забезпечують можливість ефективного шифрування та коректного відновлення відкритого тексту за рахунок властивостей КТЗ.

1. Chai, K. Y., & Zolkipli, M. F. (2021). Review on confidentiality, integrity and availability in information security. Journal of Information Security and Applications, 58, 102729.
2. Mohan Ananda P. Residue Number Systems: Theory and Applications. Birkhäuser. 2016. 351 p.

Вплив квантових обчислень на сучасну криптографію: загрози, виклики та напрямки розвитку постквантових алгоритмів

УДК 004.056.55

Михайло Касянчук[1], Юрій-Богдан Петренчук[2]

Західноукраїнський національний університет,
[1]kasyanchuk@ukr.net, [2]petrenchuk.yuriy@gmail.com

Сучасний розвиток квантових обчислень суттєво змінює модель загроз для криптографії. Алгоритми, які вважалися практично незламними для класичних комп'ютерів, можуть стати вразливими. Зокрема, алгоритм Шора забезпечує поліноміальний розв'язок задач факторизації великих чисел і обчислення дискретного логарифма, що створює загрозу для криптосистем RSA, DSA та ECC. Алгоритм Ґровера прискорює пошук у неструктурованих просторах, зменшуючи ефективну стійкість симетричних ключів завдяки квадратичному прискоренню. У результаті традиційні криптографічні підходи можуть втратити необхідний рівень безпеки у квантову епоху.

Вплив квантових обчислень на криптографію є одним із ключових викликів інформаційної безпеки XXI століття. Більшість сучасних криптографічних протоколів, що використовуються в Інтернеті, електронному урядуванні, банківських системах і блокчейн-інфраструктурі, базуються на математичних задачах, складність яких визначається обмеженнями класичних обчислень. Поява масштабованих квантових комп'ютерів змінює цю парадигму криптостійкості [1].

Алгоритм Шора, запропонований у 1994 році, дозволяє ефективно виконувати факторизацію великих чисел і обчислення дискретного логарифма на квантовому комп'ютері. Це означає, що криптосистеми на основі RSA та еліптичних кривих (ECDSA, ECDH) можуть стати вразливими після появи достатньо потужних квантових пристроїв. Дослідження показують, що для зламу RSA-2048 знадобляться тисячі логічних кубітів із корекцією помилок, хоча точні оцінки залежать від розвитку квантового обладнання.

Квантові обчислення впливають і на симетричну криптографію. Алгоритм Ґровера прискорює перебір ключів, зменшуючи складність атаки приблизно з

$O(2^n)$ до $O(2^{n/2})$. Таким чином, AES-128 у квантовій моделі забезпечує близько 64 біт ефективної безпеки, що робить AES-256 більш придатним для довготривалого використання. Водночас практична реалізація алгоритму Гровера потребує значних квантових ресурсів, тому збільшення довжини ключів залишається ефективним способом підвищення стійкості.

Особливу небезпеку становить стратегія “store now, decrypt later”, коли зашифровані дані можуть зберігатися для подальшого дешифрування після появи квантових комп’ютерів. Це критично для інформації з тривалим терміном зберігання, зокрема державних архівів, фінансових документів, медичних записів і військових комунікацій. Тому перехід до квантово-стійких алгоритмів розглядається як стратегічна необхідність уже сьогодні.

У відповідь на ці виклики Національний інститут стандартів і технологій США (NIST) у 2016 році розпочав процес стандартизації постквантової криптографії. У 2022–2024 роках було обрано перші алгоритми нового покоління, серед яких CRYSTALS-Kyber для обміну ключами та CRYSTALS-Dilithium для цифрового підпису. Вони базуються на складності задач Module-LWE (MLWE) та Module-SIS (MSIS), для яких наразі не відомо ефективних квантових алгоритмів розв’язання [2]. У процесі стандартизації також проводиться порівняння алгоритмів PQC за рівнями безпеки, розмірами ключів і продуктивністю, оскільки різні схеми можуть мати переваги в окремих сценаріях використання — від високопродуктивних мережевих протоколів до ресурсно-обмежених вбудованих систем.

Алгоритми на основі решіток (CRYSTALS-Kyber, Dilithium) демонструють кращий компроміс між продуктивністю й розмірами даних порівняно з деякими іншими підходами, тому вони й були серед лідерів впровадження. Проте оптимізація під апаратні архітектури (ARM, x86) та захист від побічних каналів залишається критичною. Рекомендована на практиці стратегія — використання гібридних схем, які поєднують класичні та PQC-компоненти в одному протоколі. Це забезпечує сумісність і поступовий перехід, зменшує ризик невдалого вибору нового алгоритму та гарантує захист у разі появи робочого квантового атакуючого апарату [3][4].

Під час оцінювання постквантових криптосистем необхідно враховувати математичну стійкість до класичних і квантових атак, результати криптоаналізу, вибір параметрів, продуктивність алгоритмів і розміри ключів, а також ризики реалізації, зокрема атаки через побічні канали.

Отже, квантові обчислення змінюють підходи до криптографічної безпеки: асиметричні алгоритми на основі факторизації та дискретного логарифму стають вразливими через алгоритм Шора, а симетричні схеми потребують збільшення параметрів через алгоритм Ґровера. У відповідь формується новий напрям — постквантова криптографія, що базується на математичних задачах, стійких до квантових атак, і поступово інтегрується в сучасну цифрову інфраструктуру.

1. Post-Quantum Cryptography PQC. [Електронний ресурс]. – Режим доступу: https://csrc.nist.gov/projects/post-quantum-cryptography/post-quantum-cryptography-standardization.

2. On the practical cost of Grover for AES key recovery. [Електронний ресурс]. – Режим доступу: https://csrc.nist.gov/csrc/media /Events/2024/fifth-pqc-standardization-conference/documents/papers/on-practical-cost-of-grover.pdf.
3. NIST Releases First 3 Finalized Post-Quantum Encryption Standards. [Електронний ресурс]. – Режим доступу: https://www.nist.gov/news-events/news/2024/08/nist-releases-first-3-finalized-post-quantum-encryption-standards.
4. Post-Quantum Cryptography: Anticipating Threats and Preparing the Future. [Електронний ресурс]. – Режим доступу:https://www.enisa.europa.eu/news/enisa-news/post-quantum-cryptography-anticipating-threats-and-preparing-the-future.

Аналіз енергетичної доцільності впровадження малопотужних апаратних вузлів у віртуалізовані навчальні середовища з кібербезпеки

УДК 004.056:004.738

Онищук Гліб[1] , Сергій Бабич[2]

Національний університет водного господарства та природокористування, [1]onyshchuk_ak23@nuwm.edu.ua, [2]s.v.babych@nuwm.edu.ua

Навчальні процеси з кібербезпеки вимагають розгортання складних ізольованих середовищ для тестування вразливостей, моделювання кібератак та налаштування засобів захисту. Традиційна модель забезпечення лабораторій передбачає використання повноцінних потужних робочих станцій («товстих» клієнтів), обчислювальні ресурси яких більшу частину часу використовуються неефективно, а їх обслуговування є складним та вкрай енергозатратним процесом [1].

Дане дослідження присвячене аналізу енергетичної та економічної доцільності заміни традиційних персональних комп'ютерів на малопотужні апаратні вузли («тонкі» або «нульові» клієнти) на базі інфраструктури віртуальних робочих столів (VDI). Актуальність переходу зумовлена глобальною необхідністю реалізації парадигми «зелених IT», що спрямована на зменшення енергоспоживання, а також критичною потребою в оптимізації витрат на супровід IT-інфраструктури в академічному та бізнес-середовищі [2, 3]. Принциповою особливістю цього підходу є дослідження показників малопотужних вузлів саме в умовах високодинамічних ресурсомістких навантажень, де архітектура VDI здатна одночасно знизити енерговитрати та безпрецедентно підвищити безпеку навчального середовища [1, 4].

Традиційна робоча станція разом з монітором споживає в середньому 150-200 Вт електроенергії [1]. Впровадження архітектури VDI передбачає перенесення всіх обчислювальних процесів на централізований сервер, тоді як на робочому місці студента залишається клієнтський пристрій, рівень енергоспоживання якого складає лише від 8 до 50 Вт [1, 2]. Навіть з урахуванням енерговитрат на живлення серверів та масивів зберігання даних,

сумарне зниження енергоспоживання для типової навчальної лабораторії на 30 місць становить близько 50% [1]. Крім того, апаратна простота «тонких» клієнтів (відсутність вентиляторів та жорстких дисків) збільшує середній час напрацювання на відмову (MTBF) до 70 000 годин у порівнянні з 30 000 годин для класичних комп'ютерів [1].

У контексті навчання з кібербезпеки архітектура малопотужних клієнтів також вирішує ключову проблему надійної ізоляції та швидкого відновлення стендів. Оскільки «тонкі» клієнти обмежують взаємодію із зовнішніми носіями, несанкціоноване копіювання даних або зараження локальної машини шкідливим кодом стає технічно неможливим [1, 4]. Сама ж інфраструктура VDI дозволяє ефективно використовувати технології «золотого образу» (golden image) та зв'язаних клонів (linked clones) [1]. Завдяки цьому після завершення лабораторної роботи користувацька сесія миттєво відключається, а змінена чи навіть скомпрометована віртуальна машина за кілька секунд повністю автоматично відновлюється до еталонного «чистого» стану [1, 5].

Отже, впровадження малопотужних апаратних вузлів у віртуалізовані навчальні середовища є обґрунтованим кроком, що забезпечує суттєве зниження сукупної вартості володіння (TCO) інфраструктурою [1]. Перехід від «товстих» клієнтів до спеціалізованих «тонких» пристроїв гарантує радикальне зменшення енергоспоживання на кінцевих робочих місцях, забезпечуючи при цьому централізоване керування та максимальний рівень ізоляції операційного середовища, який є критично необхідним для безпечного дослідження сучасних кіберзагроз [3, 4].

1. Rot A., Chrobak P. Benefits, Limitations and Costs of IT Infrastructure Virtualization in the Academic Environment. Case Study using VDI Technology. In Proceedings of the 13th International Conference on Software Technologies (ICSOFT 2018), 2018. P. 704-711.
2. Pattinson C., Cross R., Kor A. L. Thin-client and Energy Efficiency. JISC Green IT Technical Report, 2011. 23 p.
3. Akintunde R. A. The Benefits of Thin Clients to Business Organizations in Developing Countries “A Case Study of Nigeria”. Master’s Thesis, University of Oulu, 2017. 83 p.
4. Смирнов О.А. Сучасні підходи до побудови захищених віртуалізованих IT-інфраструктур у закладах вищої освіти. Сучасний захист інформації. – 2024. – Т. 15, №1. – С. 45-52.
5. Mell P., Grance T. The NIST Definition of Cloud Computing. NIST Special Publication 800-145, 2011. 7 p.

Гібридна модель захисту вебзастосунків на основі OWASP Top 10 і штучного інтелекту

УДК 004.056:004.8:004.77

[1]Савчук Костянтин, [2]Нємкова Олена

Національний університет Львівська Політехніка,
[1]kostiantyn.v.savchuk @lpnu.ua, [2]olena.a.niemkova@lpnu.ua

Вебзастосунки сьогодні є частиною майже кожного цифрового сервісу: інтернет-магазинів, банків, навчальних систем і державних порталів. Через це вони часто стають ціллю атак. У дослідженні розглянуто SQL-ін'єкції, XSS, CSRF та DDoS. Традиційні засоби захисту, зокрема WAF, IDS, IPS, статичні правила і сигнатури, що добре працюють проти відомих атак. Але їм важко помічати нові або змінені атаки, особливо у хмарних системах, мікросервісах та API.

Метою дослідження є побудова гібридної моделі веббезпеки, яка поєднує чіткі правила OWASP Top 10:2025 і методи штучного інтелекту [1]. Модель має два шари. Перший шар, Foundation, відповідає за базовий захист за правилами OWASP. Другий шар, Amplifier, додає три методи ШІ: HTTP2vec, графові нейронні мережі та Kitsune. Для порівняння різних методів виявлення атак було проаналізовано наукові праці та галузеві звіти. Перший шар закриває базові ризики вебзастосунків. Параметризовані запити та підготовлені оператори відокремлюють дані користувача від SQL-коду і допомагають проти ін'єкцій. Політика Content Security Policy обмежує виконання небезпечних скриптів і зменшує шкоду від XSS. SameSite cookies і короткоживучі токени захищають сесії. Багатофакторна автентифікація зменшує користь викраденого пароля. Захищене налаштування прибирає типові помилки, наприклад стандартні паролі, зайві сервіси або відкриті сховища. Для ризиків ланцюга постачання пропонуються сканування залежностей, Software Bill of Materials та перевірка процесу збірки.

Другий шар шукає те, що правила можуть пропустити. HTTP2vec працює з HTTP-запитами як з текстом. Модель навчається на нормальному трафіку, перетворює запити у числові вектори і знаходить запити, які сильно відрізняються від звичайних. Це корисно для виявлення SQLi, XSS та command injection, але така модель має великий розмір і потребує значних обчислювальних ресурсів [2]. Графові нейронні мережі показують зв'язки між користувачами, пристроями, сесіями та IP-адресами. Так легше помітити захоплення акаунтів або скоординовані атаки, хоча графові дані не завжди просто підготувати [3]. Kitsune працює на мережевому шлюзі або edge-пристрої, використовує невеликі автоенкодери і швидко бачить DDoS, сканування та MitM-атаки [4].

У моделі обидва шари з'єднані через SIEM і SOAR. Сигнали від HTTP2vec, графових нейронних мереж, Kitsune, WAF та IDS надходять до SIEM. SIEM збирає ці сигнали, порівнює їх і визначає, які сповіщення важливіші. Якщо підозрілий запит одночасно бачить модель HTTP2vec і WAF, довіра до такого сповіщення стає вищою. SOAR може автоматично заблокувати IP-адресу, відкликати токен або ізолювати систему. Якщо впевненість середня, сповіщення отримує аналітик. Його рішення потім можна використати для поліпшення моделей.

Порівняння можливостей двох підходів показало, що гібридний підхід поєднує їх сильні сторони. Правила добре зупиняють відомі атаки, а ШІ допомагає знайти нові або скоординовані загрози. Така модель важча для обходу, бо зловмиснику потрібно обійти і правила, і ШІ. У дослідженні також наведено галузеві дані: після додавання ШІ до SIEM і SOAR кількість

помилкових сповіщень може зменшитися приблизно на 60%, а розслідування інцидентів може стати приблизно на 40% швидшим [5]. Але варто підкреслити, що ці числа треба сприймати обережно, бо вони взяті з вторинних джерел, а не з контрольованого експерименту.

Для оцінювання таких систем доцільно не обмежуватися ROC-AUC. У безпекових даних шкідливий трафік часто становить менше ніж 0,1% від усього трафіку, тому ROC-AUC може приховати велику кількість зайвих сповіщень. Криві Precision-Recall та показник Average Precision краще показують, скільки сповіщень справді є корисними [6]. Для систем реального часу пропонується Numenta Anomaly Benchmark, бо він враховує, наскільки рано система помітила аномалію [7].

Висновки. У дослідженні запропоновано гібридну модель захисту вебзастосунків з двома шарами. Foundation дає базовий і зрозумілий захист за OWASP Top 10:2025, а Amplifier підсилює його трьома методами ШІ: HTTP2vec, графовими нейронними мережами та Kitsune. Головна ідея моделі проста: ШІ не замінює традиційний захист, а робить його сильнішим. Подальша робота має включати перевірку моделі в контрольованому середовищі, створення кращих відкритих наборів даних, навчання моделей без передавання чутливого трафіку та тестування стійкості ШІ до навмисного обходу.

1. OWASP Foundation. OWASP Top 10:2025. 2025. URL: https://owasp.org/Top10/2025/
2. Gniewkowski M. et al. HTTP2vec: Embedding of HTTP requests for detection of anomalous traffic. arXiv, 2021.
3. Bilot T., Legay A., Rønne P. B. Graph neural networks for intrusion detection: A survey // IEEE Access. 2023. Vol. 11. P. 45589-45612.
4. Mirsky Y. et al. Kitsune: An ensemble of autoencoders for online network intrusion detection // Proc. NDSS 2018.
5. Dilmegani C. Top 13 AI cybersecurity use cases with real examples in 2025. AIMultiple Research. 2025.
6. Saito T., Rehmsmeier M. The Precision-Recall plot is more informative than the ROC plot // PLOS ONE. 2015. Vol. 10(3). e0118432.
7. Lavin A., Ahmad S. Evaluating real-time anomaly detection algorithms - the Numenta Anomaly Benchmark // IEEE ICMLA. 2015. P. 38-44.

Fast squaring of multiword numbers using Mersenne modules

UDK 519.67

Andrii Tereshchenko[1], Valeriy Zadiraka[2]

V.M. Glushkov Institute of Cybernetics, [1]teramidi@ukr.net, [2]zvk140@ukr.net

In the era of post-quantum computing, it is necessary to increase the length of keys to increase cryptographic strength, which affects the performance of hardware-software cryptographic complexes. The performance of such complexes depends to a greater extent on the performance of the square operation [1] by modulus.

The multiword operation of squaring by Mersenne modulus is considered, which is one of the steps of the algorithm for implementing the multiword operation of

squaring based on the Mersenne transform. The operation of squaring by Mersenne modulus is performed element by element for each value of the result of the calculation of the direct Mersenne transform similarly to the Schönhage-Strassen algorithm [2].

A new algorithm for multiplication modulo a Mersenne prime is considered based on two half-length multiplications, instead of three multiplications, as in the Karatsuba method [3].

Lemma 1. The operation of squaring $\left\langle A^2 \right\rangle_{M_p}$ a number $A = A_1 \cdot 2^{p_0} + A_0$ of length p bits modulo a Mersenne prime $M_p = 2^p - 1$ can be implemented based on two operations of multiplication of numbers of length $p_0 = (p+1)/2$ and $p_1 = p_0 - 1$ bits.

Proof. The square of a number modulo $\left\langle A^2 \right\rangle_{M_p}$ can be represented in the form $\left\langle (A_1 \cdot 2^{p_0} + A_0)^2 \right\rangle_{M_p} = \left\langle A_1^2 \cdot 2^{2p_0} + 2 \cdot A_1 \cdot A_0 \cdot 2^{p_0} + A_0^2 \right\rangle_{M_p}$. Given that $\left\langle 2^{2p_0} \right\rangle_{M_p} = \left\langle 2^{p+1} \right\rangle_{M_p} = 2,$ the previous expression can be written in the form $\left\langle A^2 \right\rangle_{M_p} = \left\langle A_1^2 \cdot 2^{2p_0} + 2 \cdot A_1 \cdot A_0 \cdot 2^{p_0} + A_0^2 \right\rangle_{M_p}$. Let's add $3 \cdot A_1 \cdot A_0 - 3 \cdot A_1 \cdot A_0$ to the expression and get:

$$\left\langle A^2 \right\rangle_{M_p} = \left\langle A_1^2 \cdot 2^{2p_0} + 2 \cdot A_1 \cdot A_0 \cdot 2^{p_0} + A_0^2 + 3 \cdot A_1 \cdot A_0 - 3 \cdot A_1 \cdot A_0 \right\rangle_{M_p}.$$

We will use $M_0 = (2 \cdot A_1 + A_0) \cdot (A_1 + A_0),\ M_1 = A_1 \cdot A_0$ for replacement and get finally $\left\langle A^2 \right\rangle_{M_p} = \left\langle M_0 + 2 \cdot M_1 \cdot 2^{p_0} - 3 \cdot M_1 \right\rangle_{M_p}$.

Multiplications by 2 and 3 are not considered, as they can be replaced by addition operations. Multiplication by 2^{p_0} can be replaced by a bitwise cyclic shift to the left (towards the most significant bits).

The lemma is proven.

Similarly, it can be shown that in the case of dividing a multiword number into three sections, squaring modulo a Mersenne number can be performed based on three squaring operations and two multiplication operations of numbers of lengths that are a third of the original length of the number.

Lemma 2. The operation of squaring $\left\langle A^2 \right\rangle_{M_p}$ a number $A = A_2 \cdot 2^{p_1 + p_0} + A_1 \cdot 2^{p_0} + A_0$ of length p bits modulo a Mersenne number $M_p = 2^p - 1$ can be implemented based on three operations of squaring and two

operations of multiplying numbers of length $p_0 = p_1 = (p+1)/3$ and $p_2 = p_0 - 1$ bits.

$$\left\langle A^2 \right\rangle_{M_p} = \left\langle R_2 \cdot 2^{p_0+p_1} + R_1 \cdot 2^{p_0} + R_0 \right\rangle_{M_p}, \text{ where}$$

$$M_0 = (A_0 + (A_1 + A_2))^2, \quad R_0 = \frac{M_0 + M_1}{2} - M_2,$$
$$M_1 = (A_0 - (A_1 + A_2))^2,$$
$$M_2 = (A_1 - A_2)^2, \quad R_1 = \frac{M_0 - M_1}{2} - M_3,$$
$$M_3 = 2 \cdot (A_0 - A_2) \cdot A_2, \quad R_2 = \frac{M_0 - M_1}{2} - M_4.$$
$$M_3 = (2 \cdot A_0 - A_1) \cdot A_1.$$

For verification:

$$R_0 = (A_0)^2 + 4 \cdot A_1 \cdot A_2,$$
$$R_1 = 2 \cdot (A_2)^2 + 2 \cdot A_0 \cdot A_1,$$
$$R_2 = (A_1)^2 + 2 \cdot A_0 \cdot A_2.$$

Similarly to Lemma 1, multiplication and division by 2, multiplication by 2^{p_0} and $2^{p_0+p_1}$ can be disregarded in the total number of multiplication operations.

The algorithm for squaring a Mersenne number modulo by dividing the number into two sections (Lemma 1) is 33% more efficient than the Karatsuba method. The algorithm for squaring a Mersenne number modulo by dividing the number into three sections (Lemma 2) allows us to calculate the squaring operation by using three smaller-length squaring operations out of the five required multiplication operations. Squaring smaller-length numbers based on the fast Fourier transform is one of the reserves for optimizing the implementation of the multiword multiplication operation.

1. Zadiraka, V.K., Tereshchenko, A.M. An Efficient Algorithm for Squaring Multi–Word Numbers. *Cybern Syst Anal*, 61, 521-526 (2025).
2. Schönhage A., Strassen V. Schnelle Multiplikation großer Zahlen. *Computing*. – 1971. – № 7. – P. 281–292.
3. Karatsuba, A. A.; Ofman, Y. P. (1962). Multiplication of Many-Digital Numbers by Automatic Computers. Proceedings of the USSR Academy of Sciences (in Russian). 145: 293–294.

Система виявлення інформаційно-фінансових атак на криптовалютних ринках із застосуванням показника поглинання імпульсу

УДК 004.056.5:004.42:336.74

Ігор Цапро[1], Оксана Золотухіна[2]

Державний університет інформаційно-комунікаційних технологій,
[1]tsapro.ihor.work@gmail.com, [2]o.zolotukhina@duikt.edu.ua

Постановка проблеми. Стрімкий розвиток цифрових фінансових платформ, криптовалютних бірж та децентралізованих фінансових сервісів супроводжується збільшенням кількості кіберфінансових загроз, серед яких особливу небезпеку становлять координовані інформаційно-фінансові атаки: алгоритмічні маніпуляції ліквідністю, spoofing, wash trading та приховане накопичення позицій великими учасниками ринку [1]. Особливість таких атак

полягає у тому, що їх початкові ознаки часто не супроводжуються очевидною зміною ринкової ціни, а проявляються через аномальні торгові обсяги, зміну структури ліквідності та нетипову поведінку учасників ринку.

Метою дослідження є розробка програмної системи для виявлення потенційних інформаційно-фінансових атак на криптовалютних ринках шляхом статистичного аналізу часових рядів та використання показника поглинання імпульсу ринкових обсягів (Momentum Absorption Score, MAS) [2].

Актуальність дослідження обумовлена потребою створення програмних засобів раннього виявлення прихованих ринкових аномалій, які можуть передувати маніпулятивним або координованим фінансовим атакам.

Наукова новизна роботи полягає у поєднанні статистичного методу виявлення поглинання ринкових обсягів із принципами інженерії програмного забезпечення для побудови програмної системи моніторингу інформаційно-фінансових атак. На відміну від традиційних індикаторів технічного аналізу [3], показник MAS дозволяє виявляти ринкові стани, за яких значні обсяги торгів не супроводжуються суттєвою зміною ціни, що може сигналізувати про приховану ринкову боротьбу або підготовку до подальшого інформаційно-фінансового впливу.

Результати дослідження. В основу запропонованого рішення покладено аналіз часових рядів даних формату OHLCV, які включають ціну відкриття, максимальну та мінімальну ціну, ціну закриття та торговий обсяг. Формула показника сумарних ринкових обсягів має наступний вигляд (1):

$$MAS_{V,t} = (Z_{HL_t} < P_{25}(HL_t)) \wedge (Z_{V_t} > P_{75}(V_t)), \quad (1)$$

де Z_{HL_t} – це стандартна оцінка (z-score) різниці між максимальною та мінімальною цінами за час *t*, Z_{V_t} – це стандартна оцінка сумарних ринкових обсягів за час *t*, P_{25} – процентиль 25, P_{75} – процентиль 75.

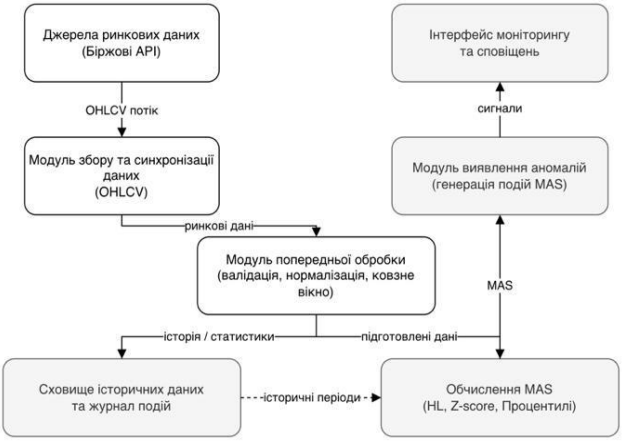


Рис.1. Схема програмної реалізації системи

Програмна система реалізована (Рис. 1) за модульною архітектурою та складається з компонентів збору ринкових даних, потокової обробки часових рядів, обчислення статистичних характеристик, генерації MAS-подій та журналювання результатів. Отримання даних здійснюється через біржові API у режимі реального часу. Для забезпечення обчислювальної ефективності застосовано механізми кешування проміжних розрахунків та планування виконання обчислювальних модулів. Архітектура системи забезпечує масштабованість, відмовостійкість та можливість інтеграції у середовища оперативного моніторингу фінансових ризиків.

Висновки. Розроблено програмну систему виявлення потенційних інформаційно-фінансових атак на криптовалютних ринках на основі показника поглинання імпульсу. Події MAS регулярно виникають у фазах низької волатильності перед значними імпульсними рухами ринку. Запропонований підхід дозволяє виявляти приховані ринкові аномалії, які не фіксуються класичними індикаторами технічного аналізу, що підвищує ефективність моніторингу потенційно маніпулятивної активності. Запропоноване рішення поєднує статистичний аналіз часових рядів та принципи інженерії програмного забезпечення, забезпечуючи можливість раннього виявлення прихованих інформаційно-фінансових загроз у режимі реального часу. Перспективами подальших досліджень є інтеграція алгоритмів машинного навчання для автоматичного відсіювання хибнопозитивних сигналів та підвищення точності детекції аномальної активності.

1. Cong, Lin William and Li, Xi and Tang, Ke and Yang, Yang, Crypto Wash Trading. *SSRN*. – 2023. DOI: http://dx.doi.org/10.2139/ssrn.3530220.
2. Цапро І. В. Застосування машинного навчання в задачі відсіювання неефективних торгових сигналів згенерованих показниками механістичного підходу. *Зв'язок*. – 2025. № 5 (177). – С. 79-86. DOI: 10.31673/2412-9070.2025.051067.
3. Han, Yufeng and Liu, Yang and Zhou, Guofu and Zhu, Yingzi, Technical Analysis in the Stock Market: A Review. *SSRN*. – 2021. DOI: http://dx.doi.org/10.2139/ssrn.3850494

Інтегрований контур захищеності вебсистем із криптографічною фіксацією та графово-нейромережевим оцінюванням подій

УДК 004.4:004.415.2 (043.2)
Ірина Замрій[1], Іван Шахматов[2], Діана Шахматова

Державний університет інформаційно-комунікаційних технологій, Київ, Україна, [1]i.zamrii@duikt.edu.ua, [2]i.shahmatov@duikt.edu.ua

Забезпечення захищеності сучасних вебсистем потребує формування єдиного контуру [1] обробки критичних подій. Метою дослідження є формалізація такого підходу для подій вебформ, транзакційних операцій та інших функціональних компонентів вебсистеми, щоб результат роботи механізмів захисту був не лише обчисленим, а й відтворюваним та доказово підтвердженим [3].

Формальну основу контуру захищеності вебсистем подамо у вигляді

$$K = (S, O, L, F, D, P, J), \tag{1}$$

де S - критичні події, O - пов'язані об'єкти й суб'єкти, L - зв'язки між ними, F - формування ознак, D - прийняття рішення, P - політики реагування, J - незмінний журнал.

Інтегрований механізм обробки подій можна подати як відображення:

$$M : S \to C \times A \times Q \times V, \tag{2}$$

де $C = \{NORM, ALERT, CRIT\}$ - множина класів стану події, $A = \{PASS, CHECK, ISOLATE, BLOCK\}$ - множина можливих дій реагування, Q - множина доказових записів, $V = \{0,1\}$ - множина результатів верифікації.

Для окремої критичної події s_i результат роботи механізму має вигляд $M(s_i) = (c_i, a_i, q_i, v_i)$, де c_i - визначений клас події, a_i - дія реагування, q_i - доказовий запис, $v_i \in \{0,1\}$ - результат перевірки коректності запису та його включення до журналу.

Послідовність роботи механізму подається як композицію функціональних перетворень:

$$M = J_q \circ Q_f \circ A_p \circ C_r \circ R_s \circ K_h, \tag{3}$$

де K_h виконує криптографічну фіксацію події, R_s формує оцінку ризику, C_r визначає клас події, A_p вибирає дію реагування відповідно до політики, Q_f формує доказовий запис, J_q додає цей запис до незмінного журналу.

Криптографічна фіксація події задається як $K_h(s_i) = (n_i, h_i, g_i)$, де n_i - нормалізоване подання події s_i, $h_i = H(n_i)$ - хеш нормалізованого подання, $g_i = Sign_{sk}(h_i)$ - цифровий підпис хешу, сформований із використанням закритого ключа sk. Оцінка ризику події визначається як $R_s(s_i) = r_i, r_i \in [0,1]$, де r_i - нормалізований рівень ризику події. Його значення залежить від типу події:

$$r_i = \begin{cases} b_i, & t_i = \text{FORM}, \\ z_i, & t_i = \text{PAY}, \\ r_i^0 \text{ для інших типів подій}, \end{cases} \tag{4}$$

де t_i - тип події, b_i - ризик подання вебформи за результатом графово-нейромережевого аналізу, z_i - ризик транзакції, r_i^0 - базова оцінка ризику [2].

Після прийняття рішення формується доказовий запис:

$$q_i = (h_i, g_i, m_i, w_i, l_i, r_i, c_i, a_i, d_i), \tag{5}$$

де h_i - хеш події, g_i - цифровий підпис, m_i - ідентифікатор версії моделі, w_i - контрольна сума параметрів моделі, l_i - контрольна сума політики реагування, r_i - оцінка ризику, c_i - клас події, a_i - дія реагування, d_i - час формування рішення. Такий запис фіксує не лише факт події, а й підстави прийнятого рішення. Завершальним етапом є включення доказового запису до незмінного журналу:

$$J_q(q_i) = (B_j, v_i), \tag{6}$$

де B_j - блок незмінного журналу, до якого включено запис q_i, $v_i \in \{0,1\}$ - результат перевірки коректності цього включення. Якщо змінюється зміст запису або порушується зв'язок між блоками журналу, результат верифікації набуває значення 0, що свідчить про порушення цілісності.

Запропонована формалізація описує єдиний контур забезпечення захищеності вебсистем, у якому критична подія проходить криптографічну фіксацію, ризикове оцінювання, класифікацію, вибір дії реагування та доказове журналювання. Використання композиції функціональних відображень дає змогу поєднати механізми верифікованого журналювання, контролю доступу та графово-нейромережевого аналізу подій, забезпечуючи простежуваність рішень, аудитну перевірність і контроль цілісності історії подій у вебсистемі.

1. Балацька В. Блокчейн-орієнтований підхід до забезпечення простежуваності та перевірюваності виконання політик КСЗІ. Кібербезпека: освіта, наука, техніка. 2026. Т. 4, № 32. С. 674–685. DOI: 10.28925/2663-4023.2026.32.1136.
2. Zamrii I., Shakhmatov I. Multi-View Graph Model with Representation Alignment and Adaptive Fusion for Better Spam Detection. Proceedings of the Workshop on Cryptology and Data Security (WCDS 2025), co-located with SMICS 2025, Lviv, Ukraine, October 16-18, 2025, CEUR Workshop Proceedings, 2026, Vol. 4191. P. 99-106.
3. Sha J., Wu J., Wang M., Pu Y., Lu S., Bilal M. QoE-aware edge server placement in mobile edge computing using an enhanced genetic algorithm. International Journal of Intelligent Networks. 2025. Vol. 6. P. 65–78.

Нормативні та методологічні засади впровадження ризико-орієнтованого підходу до кіберзахисту

УДК 004.056.55

Володимир Кононович[1], Дмитро Пастухов[2]

Державний університет інтелектуальних технологій і зв'язку,
[1]vl_kononovich@ukr.net, [2]edu.dvpastukhov@gmail.com

Розвиток інформаційно-комунікаційних систем та тотальна цифровізація як державних, так і приватних послуг кардинально змінили ландшафт кіберзагроз. Україна пройшла велику концептуальну трансформацію – відбувається перехід від застарілих комплексних систем захисту інформації (КСЗІ) до сучасних, гнучких та адаптивних ризико-орієнтованих підходів. З огляду на те, що базові системи КСЗІ мали застосовуватись до вересня 2025 року, наразі об'єкти, що підлягають захисту, повинні впровадити нові методології управління безпекою, які відповідають умовам інтенсивної кібернетичної та гібридної війни.

Фундаментальна проблема попередньої парадигми полягала в принципі загрозо-центричного підходу. Класичні системи будували як «рубіжну оборону» – захисту периметру від максимально можливої кількості гіпотетичних загроз, часто без глибокого урахування реальної ймовірності їх виникнення. Такий підхід вимагає значних, часто невиправданих капіталовкладень, що ставить під сумнів економічну доцільність реалізації КСЗІ спираючись на цей принцип. Застарілі методи захисту не здатні протистояти новітнім векторам атак, зокрема тим, що використовують інструменти штучного інтелекту для оптимізації соціальної інженерії та фішингу, де не діють жодні технічні методи захисту.

Новітній, актуальний підхід до безпеки інформації, систем та приватності (рис. 1) закріплені Законом України «Про внесення змін до деяких законів України щодо захисту інформації та кіберзахисту державних інформаційних ресурсів, об'єктів критичної інформаційної інфраструктури» (документ 4336-IX від 27 березня 2025) [1]. Новий підхід описаний у нормативному документі ДСТУ ISO 27005:2023 – побудова КСЗІ на об'єкті повинна спиратись на керуванні ризиками інформаційної безпеки [2].

Математична модель ризико-орієнтованого підходу визначається за наступною формулою оцінки ризику:

$$R = p \times h \tag{1}$$

де R – ризик, p – ймовірність загрози, h – розмір очікуваних збитків.

Розмір збитків, в сучасних умовах гібридних загроз, є комплексною величиною, яка виходить за межі вартості комп'ютерних систем та даних, що зберігаються в них. Він включає у себе також прямі фінансові витрати, репутаційні збитки, когнітивні втрати, штрафні санкції від регуляторів.

Ключовим елементом сучасного підходу є усвідомлення того факту, що досягнення ідеального рівня безпеки є утопією та економічним нонсенсом. Будь-який захід із кіберзахисту – чи то закупівля апаратних брандмауерів наступного покоління, впровадження систем EDR/XDR, інтеграція платформ розвідки загроз, чи то проведення регулярних тренінгів персоналу, має свою

визначену вартість. Завжди слід пам'ятати, що безпека не обмежується суто технічним рівнем захисту інформації, а є тільки її частиною. Не менш важливу роль відіграє когнітивний захист – повинні бути регулярні тренінги, де персонал повинен почати усвідомлювати наявність загроз з боку використання психологічних методик впливу на них для обходу наявних технічних систем захисту інформації [3].

Головним принципом прийняття рішень щодо управління ризиками стає суворе економічне правило: вартість впровадження та підтримки контрзаходів повинна бути меншою ніж можливі фінансові, операційні та репутаційні витрати від реалізації ризику. Якщо вартість захисту перевищує потенційні збитки, єдино правильним та раціональним управлінським рішенням є прийняття цього ризику керівництвом.

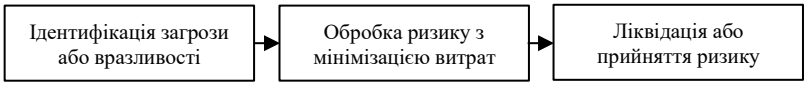


Рис.1. Процес обробки ризику з використанням ризико-орієнтованого підходу

Мінімізація витрат, а під час обробки ризику – пошук необхідних методик захисту, повинен відбуватись циклічно, допоки вартість контрзаходів не буде дешевшою, ніж потенційні витрати. У випадку коли вартість контрзаходу є більшою та його неможливо зменшити – ризик повинен бути прийнятим керівництвом.

На етапі перевірки всієї архітектури на об'єкті, якість побудованих систем захисту інформації можна оцінити завдяки рівню захищеності найслабшої ланки систем безпеки – вся система безпеки залежить від неї.

У якості примітки звернемо увагу на невірний переклад українською назви ДСТУ [2]: «Information security, cybersecurity and privacy». Термін «privacy» перекладено як «конфіденційність», а треба перекладати у даному контексті як «приватність», щоб не вводити читача в оману.

1. Закон України «Про внесення змін до деяких законів України щодо захисту інформації та кіберзахисту державних інформаційних ресурсів, об'єктів критичної інформаційної інфраструктури». URL: https://zakon.rada.gov.ua/laws/show/4336-20#Text (дата звернення: 13.05.2026).
2. ДСТУ ISO 27005:2023. Інформаційна безпека, кібербезпека та захист приватності. Настанова керування ризиками інформаційної безпеки (ISO/IEC 27005:2022, IDT).
3. Соціальна інженерія та кіберпсихологія: монографія / В. Г. Кононович, С. В. Стайкуца, М. М. Тодорова та ін.; за ред. В. Г. Кононовича, С. В. Стайкуци. Одеса: Астропринт, 2025. 388 с.

Архітектура самосуверенних цифрових двійників для приватного управління даними IoT-пристроїв

УДК 04.056:004.738.5

[1]Овсянко Дмитро, [2]Нємкова Олена

Національний університет Львівська Політехніка,
[1]dmytro.o.ovsianko@lpnu.ua, [2]olena.a.niemkova@lpnu.ua

Стрімке поширення IoT-систем супроводжується зростанням обсягів даних, кількості підключених пристроїв та вимог до безпеки їхньої взаємодії. Традиційні цифрові двійники зазвичай функціонують у межах централізованих хмарних платформ, де контроль над ідентичністю пристрою, політиками доступу та зберіганням даних фактично належить провайдеру системи. Такий підхід створює ризики єдиної точки відмови, ускладнює масштабування та обмежує автономію власника пристрою щодо управління власними даними [1]. Актуальність дослідження зумовлена необхідністю забезпечення приватності, цілісності та незалежного контролю над даними IoT-пристроїв в умовах активного впровадження концепцій Industry 4.0 та розумних середовищ, де користувач має зберігати суверенітет над власними цифровими активами.

Метою роботи є розроблення архітектури самосуверенного цифрового двійника (Self-Sovereign Digital Twin, SSDT) для приватного управління даними IoT-пристроїв на основі принципів децентралізованої ідентичності, верифікованих облікових даних та незмінного аудиту. Самосуверенний цифровий двійник розглядається як цифрове представлення фізичного об'єкта, яке має власну децентралізовану ідентичність, криптографічні ключі, набір атрибутів, поточний стан, історію змін та політики доступу [2]. Наукова новизна роботи полягає у тому, що, на відміну від відомих підходів, запропонована архітектура поєднує концепцію цифрового двійника з моделлю самосуверенної ідентичності, переносить функції управління ідентифікаторами та політиками доступу на периферійний рівень, а блокчейн-реєстр використовує виключно для криптографічних зобов'язань і службових записів, без зберігання первинних IoT-даних, що дозволяє забезпечити баланс між приватністю, прозорістю аудиту та продуктивністю системи.

Для досягнення поставленої мети запропоновано трирівневу архітектуру SSDT. Перший рівень (фізичний) представлений IoT-пристроями, які здійснюють збір телеметрії, формують повідомлення з часовими мітками та номерами послідовності, а також підписують дані за допомогою криптографічних ключів. Це дає змогу підтвердити автентичність джерела даних і виявляти спроби повторного використання повідомлень.

Другий рівень (рівень цифрового двійника) розміщується на периферійному обчислювальному шлюзі. Він відповідає за управління децентралізованими ідентифікаторами, зберігання верифікованих облікових даних, агрегацію телеметрії, підтримку поточного стану цифрового двійника та виконання політик доступу. Розміщення цього функціоналу на периферії дозволяє зменшити залежність від хмарної інфраструктури, скоротити затримки та забезпечити локальну обробку чутливих даних.

Третій рівень (блокчейн-рівень) виконує функцію розподіленого реєстру довіри. У ньому фіксуються операції реєстрації та оновлення децентралізованих ідентифікаторів, статуси облікових даних, факти відкликання повноважень і критичні події доступу. Використання приватного блокчейну дозволяє забезпечити незмінність журналу аудиту без необхідності зберігати повні набори IoT-даних у реєстрі. Замість цього до блокчейну можуть записуватися криптографічні зобов'язання, хеші або службові записи, необхідні для перевірки цілісності [3].

Окрему роль у запропонованій архітектурі відіграють механізми приватності. Для зменшення обсягу розкритих даних може застосовуватися селективне розкриття атрибутів та докази з нульовим розголошенням. Це дозволяє підтверджувати певні твердження про стан пристрою без передавання повного набору первинних даних. Наприклад, цифровий двійник може довести, що параметр перебуває в допустимому діапазоні, не розкриваючи його точного значення.

Для оцінювання безпеки системи доцільно використовувати модель загроз STRIDE. Основними загрозами для SSDT-системи є атаки типу Man-in-the-Middle, replay-атаки, підміна IoT-пристрою або цифрового двійника, підробка облікових даних, обхід політик доступу, масовані спроби несанкціонованого доступу та компрометація приватних ключів. Як контрзаходи можуть використовуватися взаємна автентифікація на рівні mTLS, підписання повідомлень на рівні застосунку, перевірка часових міток і номерів послідовності, реєстрація DID у блокчейн-реєстрі, перевірка статусу відкликання облікових даних, обмеження частоти запитів та зберігання ключів в апаратних модулях безпеки.

Висновки. У результаті дослідження запропоновано трирівневу архітектуру SSDT, яка поєднує функціональність цифрових двійників із принципами самосуверенної ідентичності, приватності за проєктуванням та незмінного аудиту. Отримані результати показують, що запропонований підхід забезпечує контроль власника пристрою над ідентичністю та даними, надає зовнішнім сервісам лише той обсяг інформації, який необхідний для конкретної взаємодії, та дозволяє верифікувати цілісність взаємодій без розкриття змісту IoT-даних. Перспективними напрямами подальших досліджень є оптимізація генерації доказів з нульовим розголошенням для ресурсно обмежених пристроїв, розроблення прототипу SSDT та експериментальна оцінка продуктивності запропонованої архітектури.

1. Tao F., Zhang M., Nee A. Y. C. Digital Twin Driven Smart Manufacturing. Academic Press, 2019. 340 p.
2. Mühle A., Grüner A., Gayvoronskaya T., Meinel C. A survey on essential components of a self-sovereign identity // Computer Science Review. 2018. Vol. 30. P. 80–86.
3. Androulaki E., Barger A., Bortnikov V. et al. Hyperledger Fabric: A distributed operating system for permissioned blockchains // Proceedings of the Thirteenth EuroSys Conference. 2018. P. 1–15.

Проблема узгодження експертних і алгоритмічних оцінок якості мастерингу аудіо

УДК 004.04

Євген Лабун[1], Богдан Худік[2]

Державний Університет Інформаційно Комунікаційних Технологій,
[1]e.labun@duikt.edu.ua, [2]b.khudik@duikt.edu.ua

Автоматизоване оцінювання якості мастерингу аудіо потребує узгодження технічних параметрів сигналу з тим, як результат сприймається експертами та слухачами. Мастеринг є завершальним етапом підготовки композиції до публікації, де приймаються рішення щодо гучності, динаміки, спектрального балансу, просторовості та цілісності звучання [6, 7]. Водночас відповідність окремим нормам не гарантує високої перцептивної якості: трек може мати коректні LUFS і true peak, але сприйматися як мутний, різкий, надмірно стиснений або неприродний [1, 4, 7, 8].

Актуальність роботи зумовлена тим, що в задачах машинного навчання експертні судження потрібно перетворювати на формалізовані цільові оцінки, а алгоритмічні метрики — на інформативні ознаки. ITU-R BS.1770-5 формалізує вимірювання гучності та true peak [1], ITU-R BS.1116-3 і BS.1534-3 задають основу контрольованих слухових тестів [2, 3], ITU-R BS.1387-2 описує об'єктивне оцінювання сприйманої якості аудіо [4], а ITU-T P.1401 визначає статистичні підходи до перевірки objective quality prediction models відносно суб'єктивних результатів [5].

Метою роботи є обґрунтування проблеми узгодження експертних і алгоритмічних оцінок якості мастерингу аудіо та визначення підходу, який може бути використаний на підготовчому етапі створення датасету для моделей машинного навчання. На відміну від простого переліку критеріїв, акцент зроблено на зв'язку між об'єктивними дескрипторами сигналу та суб'єктивними оцінками якості звучання.

Проблема узгодження проявляється у трьох аспектах. По-перше, технічні метрики мають різну перцептивну значущість залежно від жанру, матеріалу та контексту відтворення [6, 7]. По-друге, якість є багатовимірним сприйняттям: чистота, прозорість, тембральна природність, просторовість, відсутність артефактів і загальне враження не зводяться до одного числового показника [2–4, 8]. По-третє, об'єктивні моделі залежать від домену застосування, тому метрики, валідовані для кодування або інших задач, не можна автоматично переносити на оцінювання мастерингу без додаткової перевірки [4, 9].

Запропонований підхід передбачає парне представлення кожного аудіофрагмента: з одного боку обчислюються алгоритмічні ознаки (LUFS, true peak, RMS, crest factor, динамічний діапазон, спектральні характеристики, міжканальна кореляція, ознаки кліпінгу або надмірної компресії), з іншого — збираються експертні оцінки за перцептивними шкалами [1–3, 6, 8]. Такі оцінки доцільно зберігати не лише як інтегральний бал, а як набір пов'язаних міток: загальна якість, прозорість, тембральний баланс, просторовість, виразність динаміки та наявність артефактів.

Для організації слухового оцінювання доцільно використовувати логіку ITU-R BS.1116-3 у випадку малих відмінностей між варіантами мастерингу та ITU-R BS.1534-3 / MUSHRA-подібні процедури для проміжних рівнів якості [2, 3]. Для перевірки узгодженості прогнозів моделі з експертними оцінками варто враховувати розбіжності між оцінками слухачів, можливі нетипові судження та невизначеність суб'єктивних результатів [5].

У задачах машинного навчання об'єктивні метрики можуть виступати вхідними ознаками, експертні оцінки — цільовими значеннями, а випадки розбіжності між ними — матеріалом для аналізу обмежень моделі. Робота Shtern, Casas і Tzerpos демонструє релевантність ML саме для оцінювання якості музичного мастерингу відносно професійного еталона [6]. Дослідження якості популярної музики показує, що суб'єктивні оцінки якості пов'язані з об'єктивними ознаками сигналу, зокрема гучністю та динамічним стисненням, тоді як оцінка «подобається» має іншу природу [8]. Очікуване значення підходу полягає в тому, що він створює основу для датасету, у якому якість мастерингу описується не лише технічними параметрами, а й експертною перцептивною розміткою. Це дає змогу досліджувати, які об'єктивні ознаки корелюють із людським сприйняттям, у яких випадках алгоритмічна оцінка суперечить експертній і які фактори — жанр, стиль, тип обробки або особливості слухача — впливають на результат [5, 8, 9].

Отже, проблема узгодження експертних і алгоритмічних оцінок є ключовою для розробки методів автоматизованого оцінювання якості мастерингу аудіо. Запропонований підхід не передбачає готової навченої моделі, а визначає методологічну основу для подальшого формування датасету, проведення слухових тестів і побудови моделей, здатних пояснювати розбіжності між технічними метриками та людським сприйняттям.

1. ITU-R BS.1770-5. Algorithms to measure audio programme loudness and true-peak audio level. Geneva: ITU, 2023.
2. ITU-R BS.1116-3. Methods for the subjective assessment of small impairments in audio systems. Geneva: ITU, 2015.
3. ITU-R BS.1534-3. Method for the subjective assessment of intermediate quality level of audio systems. Geneva: ITU, 2015.
4. ITU-R BS.1387-2. Method for objective measurements of perceived audio quality. Geneva: ITU, 2023.
5. ITU-T P.1401. Methods, metrics and procedures for statistical evaluation of objective quality prediction models. Geneva: ITU, 2020.
6. Shtern M., Casas P., Tzerpos V. Evaluating Music Mastering Quality Using Machine Learning. CASCON, 2018. P. 126–135.
7. Katz B. Mastering Audio: The Art and the Science. 3rd ed. New York: Routledge / Focal Press, 2015.
8. Wilson A., Fazenda B. M. Perception of Audio Quality in Productions of Popular Music. J. Audio Eng. Soc. 2016. Vol. 64, No. 1/2. P. 23–34. DOI:10.17743/jaes.2015.0090.
9. Torcoli M., Kastner T., Herre J. Objective Measures of Perceptual Audio Quality Reviewed. arXiv:2110.11438, 2021.

Удосконалення автоматизованої генерації ознак мовними моделями для виявлення шахрайства у веб-застосунках

УДК 004.056:004.85

Вадим Яковець[1]

Ужгородський національний університет, [1]vadym.yakovets@uzhnu.edu.ua

Постановка проблеми. Виявлення шахрайства у веб-застосунках електронної комерції та банківських платіжних системах є практичною задачею прикладної кібербезпеки. Класифікатори цієї задачі працюють із сильно дисбалансованими класами, оскільки шахрайські транзакції зазвичай складають менше 0,1 % обсягу. Інженерія ознак потребує знань предметної галузі й переважно виконується вручну. Мовні моделі (CAAFE [1], LLM-FE [2]) дозволяють її автоматизувати, але у своєму дослідженні Küken J., Purucker L. та Hutter F. [3] показали важливу ваду: моделі майже завжди пропонують прості арифметичні оператори (додавання, множення) і дуже рідко пропонують операції групування з агрегацією. У задачах виявлення шахрайства саме операції групування дають доступ до агрегованих метрик користувача (частоти, обсягів, часових патернів), на яких будуються правила виявлення рідкісних подій. AML-датасети використано як наближений стенд, вони містять часові транзакції, контрагентів, суми та рідкісну позитивну мітку, але не покривають усіх веб-поведінкових ознак.

Метою роботи є удосконалення методології автоматизованої генерації ознак мовними моделями для класифікаторів виявлення шахрайства у веб-застосунках через зміну розподілу пропонованих операторів і запровадження вартісно-чутливого протоколу оцінювання.

Об'єктом дослідження є процеси автоматизованої генерації ознак для класифікаторів виявлення шахрайства; *предметом* є критерії відбору ознак з врахуванням вартості помилок та протокол часової вкладеної валідації для незалежної оцінки приросту якості.

Актуальність. Збитки від платіжного шахрайства в електронній комерції продовжують зростати, а ручна інженерія ознак для відповідних класифікаторів є затратною. Без виправлення зсуву моделей до простих операторів автоматична інженерія ознак мовними моделями не дає стабільного приросту якості в задачі виявлення платіжного шахрайства. У доступних українських публікаціях аналогічну задачу прицільно не розглядали.

Наукова новизна. Запропоновано протокол автоматизованої генерації ознак мовними моделями для виявлення шахрайства, що розвиває підхід CAAFE [1] та LLM-FE [2] трьома елементами: 1) шаблон системного запиту доповнено апріорними знаннями про оператори, що емпірично зсуває розподіл пропонованих моделлю операторів від простих арифметичних до групувань з агрегацією; 2) критерій прийняття кандидатної ознаки замінено: замість ROC-AUC використано очікувану вартість помилок із матрицею вартості, яка враховує робочий поріг; 3) фінальне оцінювання виконано на незалежному часовому тестовому розбитті з порогом, налаштованим лише на окремій калібрувальній вибірці.

Виклад розв'язку. Методологія включає класифікатор LightGBM на 27 базових ознаках, мовну модель як генератор кандидатних ознак із запитом, орієнтованим на предметну галузь і цикл прийому ознак з контролем на окремій калібрувальній вибірці, що схвалює нову ознаку за мінімумом очікуваної вартості. Часовий поділ 70/15/15. Оператори ознак навчаються лише на тренувальній вибірці, калібрувальна використовується для пошуку робочого порогу, а мітки тестової вибірки задіюються лише для фінального підрахунку метрик.

Результати пілотного експерименту. Публічні датасети IBM AML LI-Small [4] (6,92 млн транзакцій, частка шахрайства 0,0515 %) і LI-Medium (31,25 млн, 0,0513 %), стратифікована підвибірка 200 тис., по 20 ознак-кандидатів на модель. Протестовано сім моделей. Одну виключили за результатами перевірки запам'ятовування. До базової панелі увійшли шість моделей: Llama-3.1-8B, Mistral-v0.3, GPT-4o-mini, Claude-Haiku-4.5, Claude-Sonnet-4.6 та GPT-5.4. Усереднена частка простих операторів серед моделей, валідних в обох протоколах, наведена у табл. 1.

Таблиця 1.

Усереднена частка простих операторів за моделями, валідними в обох протоколах

Датасет	Моделей	CAAFE % простих	Новий % простих
LI-Small	4	82,2	1,3
LI-Medium	5	69,9	1,1

На LI-Small Mistral-7B-v0.3 і GPT-4o-mini виключено через невалідний формат відповіді, на LI-Medium виключено лише Mistral. У перевірці запам'ятовування за Kuken §3.4 (n = 100) усі шість моделей базової панелі залишаються нижче 50 %-го порогу. Відповідь моделі очікується у форматі JSON з полями {оператор, колонки, обгрунтування}, валідність визначається успішним парсингом і побудовою ознаки на тренувальній вибірці.

Обмеження: вартісний критерій прийому не залишив жодної ознаки (у калібрувальній вибірці було 17 шахрайських транзакцій на LI-Small і 12 на LI-Medium), тому ефект дав передусім шаблон запиту. На одному часовому відкладеному наборі зменшення вартості помилок не зафіксовано.

Висновки. У цьому пілоті підтверджено лише перший етап: запит справді змінює типи згенерованих ознак (табл. 1). Зменшення вартості помилок поки не показано, вартісний критерій прийому не залишив жодної кандидатної ознаки. Кількісне оцінювання приросту якості потребує повторних часових розбиттів.

1. Hollmann N., Müller S., Hutter F. LLMs for Automated Data Science: Introducing CAAFE for Context-Aware Automated Feature Engineering. *NeurIPS*. 2023. arXiv:2305.03403.
2. Abhyankar N., Shojaee P., Reddy C.K. LLM-FE: Automated Feature Engineering for Tabular Data with LLMs as Evolutionary Optimizers. arXiv preprint. 2025. arXiv:2503.14434.
3. Küken J., Purucker L., Hutter F. Large Language Models Engineer Too Many Simple Features for Tabular Data. 3rd Workshop on Table

Representation Learning at NeurIPS. 2024. arXiv:2410.17787v2.
4. Altman E. et al. Realistic Synthetic Financial Transactions for Anti-Money Laundering Models. *NeurIPS*. 2023. arXiv:2306.16424.

Sustainable information technology for auditable financial anomaly prediction aligned with the EU AI Act, ESG and CSRD standards

UDK 004.8:502.131.1]:657.6:[341.171:061.1EU]

Mykola Zlobin[1]

[1]Chernihiv Polytechnic National University, [1]mykolay.zlobin@gmail.com

The digital transformation of EU financial institutions requires a shift from experimental AI models to industrial, auditable, and sustainable AI systems. This transition is central to the goals of the AIFEU project, which focuses on artificial intelligence in EU financial institutions. In banking, AI is increasingly used for credit scoring, fraud detection, anomaly monitoring, and risk assessment. Creditworthiness and credit-scoring AI systems are classified as high-risk under Annex III of the EU AI Act, while fraud-detection systems, although treated differently in the legal classification, still require strong governance because they affect operational risk, customer protection, and institutional accountability. This creates a direct need for financial AI systems that are transparent, stable, and resource-aware. The scientific contradiction is clear: models optimized solely for predictive accuracy may become fragile in real-world conditions, suffer from backtest overfitting, and incur unnecessary computational and environmental costs.

This paper presents Sentinel as a sustainable information technology for predicting financial anomalies. Sentinel is not defined as a single predictive model. It is a modular information technology designed to support the full analytical cycle. It connects raw data processing, model stability diagnostics, sustainability evaluation, resource-aware training, and automated reporting. The architecture follows a regulation-first logic. This means auditability, traceability, stability, and sustainability are not added after model training. They are embedded in the technical architecture from the beginning. Sentinel consists of 4 functional modules: Adaptive data engine, Diagnostic algorithmic core, Green AI guard, and reporting layer.

The first module is the Adaptive Data Engine. It implements the Data processing method for financial datasets with extreme class imbalance, noise, and leakage risk. In the experiment, the credit-card fraud dataset contained only 0.18% fraudulent cases. This creates a serious risk because a model can appear accurate while failing to detect rare anomalies. Sentinel addresses this through leakage-safe preprocessing. The method includes robust scaling, stratified sampling, under-sampling, outlier removal, and out-of-fold target encoding for categorical variables. These operations protect the integrity of the input data and support the logic of Article 10 of the EU AI Act, which emphasizes data quality, data governance, data preparation, and bias mitigation for high-risk AI systems, also supporting ESG and CSRD compliance.

The second module is the Diagnostic algorithmic core. It implements a stability diagnostic model based on dispersion indicators: variance, interquartile range, and 95% confidence interval. Unlike standard validation relying only on average accuracy or error, this module evaluates model stability across folds and complexity levels. It

identifies the Knee point of complexity, where further model growth no longer improves generalization and increases instability risk. In the stability experiment, the Knee point was identified at 400 leaf nodes, with MAE = 242,906.

The third module is the Green AI guard. It implements the Sustainability scoring model, known as the P-score. P-score is not only a predictive metric. It is a multi-objective Model that combines predictive quality, training time, energy consumption, and CO_2 footprint. Unlike standard approaches that select the model with the lowest error, P-score penalizes resource-intensive configurations. In the Bitcoin LSTM experiment, the Balanced model achieved a P-score of 0.86, while the resource-intensive configuration achieved a P-score of 0.63.

The fourth module is the reporting layer. It transforms technical outputs into audit-ready documents. The first artifact is the Model utility bill. It summarizes data integrity, predictive performance, statistical stability, energy use, CO_2 footprint, and resource efficiency. The second artifact is the ESG Audit Trail. It records preprocessing versions, validation evidence, P-score results, training budget, final model decisions, and regulatory-relevant information.

The simulations verified the practical value of Sentinel. In the fraud detection experiment, the Adaptive data engine neutralized the initial 0.18% imbalance and achieved an ROC-AUC of 0.9787 with a Support Vector classifier. In the stability experiment, the Diagnostic core identified the Knee point at 400 leaf nodes and produced dispersion-based evidence through Variance, IQR, and 95% Confidence interval. In the Bitcoin experiment, P-score selected the more sustainable configuration and avoided a resource-intensive model that emitted 8.4 times more CO_2. In the LightGBM experiment, budget-aware training reduced the training cycle by up to 52% while improving predictive performance.

Sentinel enables EU financial institutions to meet legal and organizational expectations while improving operational performance. It supports EU AI Act-oriented documentation through data governance, stability evidence, robustness assessment, and audit-ready reporting.

Acknowledgments: This research is carried out within the framework of the ERASMUS+ Jean Monnet project «Artificial Intelligence in the EU Financial Institutions» (Project number 101127170 — AIFEU — ERASMUS-JMO-2023-HEI-TCH-RSCH) and TURBO project (Project number 101129315-TURBO-Erasmus-EDU-2023-CBHE). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or European Education and Culture Executive Agency (EACEA). Neither the EU nor the granting authority can be held responsible for them.

1. European Parliament and Council. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence. Official Journal of the European Union. – 2024.
2. Bailey D.H., Borwein J.M., López de Prado M., Zhu Q.J. The probability of backtest overfitting. Journal of Computational Finance. – 2016. – Vol. 20, №4. – P. 39–69.
3. Dal Pozzolo A., Caelen O., Johnson R.A., Bontempi G. Calibrating probability with undersampling for unbalanced classification. 2015 IEEE

Symposium Series on Computational Intelligence. – 2015. – P. 159–166.
4. Strubell E., Ganesh A., McCallum A. Energy and policy considerations for deep learning in NLP. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. – 2019. – P. 3645–3650.

Дослідження специфікації ZigBee та стандарту IEEE 802.15.4

УДК 004.7

Максим Марченко[1], Євгенія Іванченко[2],
Ігор Іванченко[3], Анна Васьковська[4]

Державний університет інформаційно-комунікаційних технологій,
[1]mv.marchenko@duikt.edu.ua, [2]e.ivanchenko@duikt.edu.ua,
[3]i.ivanchenko@duikt.edu.ua, [4]a.vaskovska@duikt.edu.ua

Розроблення протокольної структури передавання повідомлень у сенсорних мережах здійснювалося з 2000 року науковими колективами двох організацій:

- цільовою групою TG 15.4 комітету IEEE 802 зі стандартизації LAN/MAN, яка займалася створенням стандарту, що визначає протоколи рівнів PHY і MAC для LR-PAN. Перша редакція стандарту (IEEE 802.15.4-2003) була затверджена у жовтні 2003 р., а друга версія (IEEE 802.15.4-2006) - у вересні 2006 р.;
- групою розробки специфікації ZigBee Alliance, діяльність якої була спрямована на стандартизацію протоколів вищих рівнів LR-PAN. Перша редакція стандарту (ZigBee Specification 1.0) була затверджена у грудні 2004 р., а дві наступні версії (ZigBee Specification 2006 та ZigBee Specification 2007) - у грудні 2006 р. та листопаді 2007 р. відповідно.

Специфікації ZigBee передбачають використання двох верхніх рівнів стеку протоколів вузлів низькошвидкісних мереж: мережевого та прикладного, внаслідок чого загальна протокольна система взаємодії вузлів має чотирирівневу структуру (рис. 1). Рівні взаємодії, визначені специфікаціями ZigBee, функціонують як надбудова над рівнями, регламентованими стандартами IEEE 802.15.4.

Необхідність створення персональних мереж із невисокою швидкістю передачі даних (Low Rate PAN - LR PAN), на рівні кількох десятків кбіт/с, зумовлена глобальною тенденцією до автоматизації практично всіх сфер діяльності людини. Розвиток низькошвидкісних PAN відбувався паралельно з автоматизацією побутових пристроїв і впровадженням систем розподіленого контролю та управління (Distributed Control System - DCS), починаючи з 80-х років XX століття. Метою дослідження є аналіз архітектури, протокольної структури та особливостей функціонування низькошвидкісних персональних мереж LR PAN, побудованих на основі стандарту IEEE 802.15.4 та специфікацій ZigBee. Їхня науково-технічна основа формувалася з двох ключових складових:

- з одного боку, створення широкого спектра сенсорів (sensor) - елементів, призначених для вимірювання фізичних величин різної природи та формування електричних сигналів, які відображають значення цих величин;

- з іншого боку, розвиток малогабаритних приймально-передавальних та інтелектуальних пристроїв (мікроконтролерів), здатних здійснювати обробку й бездротову передачу електричних сигналів, отриманих від сенсорів.

Науково-технічний прогрес останньої чверті XX століття створив передумови для широкого впровадження DCS у різноманітних сферах гуманітарної та виробничої діяльності. Одним із підтверджень таких досягнень стала реалізація на межі століть проєкту «Інтелектуальний пил» (Smart Dust), метою якого була розробка сенсорних вузлів (Sensor Node - Mote) розміром приблизно 1 мм³.

Мережевий рівень забезпечує передавання повідомлень між віддаленими вузлами мережі, тобто виконує функції маршрутизації, тоді як прикладний рівень визначає ієрархічну та сервісну роль вузлів, а також профілі виконуваних функцій. Стандарти IEEE 802.15.4 регламентують взаємодію трансіверів, тоді як специфікації ZigBee визначають принципи взаємодії мікроконтролерів [1].

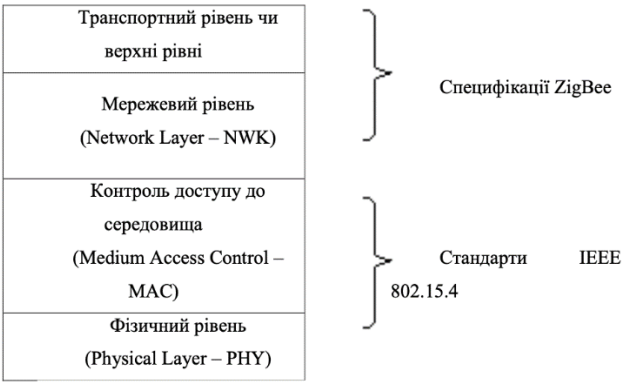


Рис. 1. Стек протоколів стандарту IEEE 802.15.4 і специфікації ZigBee

1. Охоронна сигналізація [Електронний ресурс]. – Режим доступу до ресурсу: http://www.florian-lviv.com/okhoronna-syhnalizatsiia.

Модель оцінювання кіберзахисту персональних даних у системах реєстрації заходів

УДК 004.056

Анна Васьковська[1], Євгенія Іванченко[2], Ігор Іванченко[3], Максим Марченко[4]

Державний університет інформаційно-комунікаційних технологій,
[1]a.vaskovska@duikt.edu.ua, [2]e.ivanchenko@duikt.edu.ua,
[3]i.ivanchenko@duikt.edu.ua, [4]mv.marchenko@duikt.edu.ua

У сучасному цифровому середовищі персональні дані є важливим інформаційним активом, а їх захист — одним із ключових завдань кібербезпеки. Бази реєстрації учасників спортивних подій обробляють значні обсяги конфіденційної інформації, зокрема ідентифікаційні дані, медичні довідки та платіжні реквізити, що робить їх привабливою ціллю для кіберзлочинців. Зі зростанням популярності онлайн-реєстраційних сервісів збільшується кількість кіберзагроз, серед яких фішингові атаки, компроментація облікових записів, ін'єкційні атаки, DDoS-атаки та експлуатація вразливостей веб-додатків. Успішна реалізація таких атак може призвести до витоку персональних даних, фінансових втрат і суттєвих репутаційних ризиків для організаторів спортивних подій.

Ефективний захист баз реєстрації потребує комплексного підходу, що включає використання багаторівневих механізмів безпеки відповідно до сучасних стандартів і рекомендацій, зокрема OWASP Top 10, NIST SP 800-53 та GDPR. Особливої актуальності це питання набуває в умовах підвищених кіберзагроз, характерних для України, де масові спортивні заходи проводяться в умовах постійного інформаційного протистояння. Враховуючи недостатній рівень впровадження сучасних механізмів захисту у більшості реєстраційних систем, актуальним завданням є розроблення моделі оцінювання рівня захищеності таких інформаційних ресурсів.

Метою цієї роботи є розроблення моделі оцінювання рівня захищеності персональних даних у базах реєстрації спортивних подій, що дозволяє виявляти вразливості та формувати рекомендації щодо підвищення кіберзахисту відповідно до сучасних стандартів інформаційної безпеки.

Сучасні реєстраційні системи для спортивних подій повинні відповідати високим вимогам інформаційної безпеки, оскільки обробляють значні обсяги конфіденційних даних. Одним із базових механізмів захисту є використання TLS (Transport Layer Security), який забезпечує шифрування каналу передачі даних між користувачем і сервером, а також гарантує цілісність інформації. Застосування HTTPS на основі актуальних версій TLS є стандартною вимогою для систем, що працюють із персональними та платіжними даними, оскільки дозволяють захистити їх від перехоплення та модифікації під час передачі.

Додатковими критичними елементами безпеки є багатофакторна автентифікація (MFA) та рольова модель доступу RBAC (Role-Based Access Control). MFA знижує ризик компрометації облікових записів шляхом використання кількох факторів автентифікації, тоді як RBAC обмежує доступ користувачів відповідно до їх ролей, мінімізуючи потенційні наслідки інцидентів безпеки. Сукупне застосування цих підходів забезпечує багаторівневий захист реєстраційних систем та підвищує стійкість до сучасних кіберзагроз.

Ще один критично важливий напрям забезпечення безпеки — моніторинг загроз і виявлення атак у режимі реального часу. Для цього застосовуються SIEM (Security Information and Event Management), IDS/IPS (Intrusion Detection/Prevention Systems) та журнали аудиту, які забезпечують фіксацію та аналіз усіх подій у системі.

Рис. 1. Алгоритм захисту персональних даних користувача в БР спортивних подіях

Попри наявність сучасних технологічних засобів захисту, організації, що проводять спортивні заходи, часто стикаються з обмеженістю ресурсів, відсутністю уніфікованих політик безпеки та впливом людського фактора. Це зумовлює необхідність впровадження моделей оцінювання рівня захищеності, які дозволяють об’єктивно визначати слабкі місця системи та формувати обґрунтовані заходи її вдосконалення.

1. Guide to Protecting Personally Identifiable Information (PII) // NIST. - URL:https://csrc.nist.gov/pubs/itlb/2010/04/guide-to-protecting-personally-identifiable-inform/final
2. Security and Privacy Controls for Information Systems and Organizations // NIST SP 800-53 Revision 5 Update 1. - URL: https://csrc.nist.gov/pubs/sp/800/53/r5/upd1/final

Parameter Selection for Friendly Fraud Detection in eCommerce

UDK 004.85:336.717

Dmytro Masiuk

Taras Shevchenko National University of Kyiv, masiukdmitry96@gmail.com

Friendly fraud has become one of the most significant challenges in modern eCommerce systems. Unlike traditional payment fraud, also called “true fraud”, friendly fraud occurs when legitimate customers dispute valid transactions through chargebacks, often claiming unauthorized usage. It has become increasingly simpler for fraudseter to commit fraud, as they can do so in a matter of clicks. Alongside this growth, merchants face increasing financial losses related to chargebacks and payment disputes. According to research in fraud analytics and financial machine learning, modern detection systems increasingly rely on intelligent parameter selection and behavioral analysis techniques to distinguish suspicious behavior from legitimate customer activity [1][3]. This report analyzes the importance of parameter selection in friendly fraud detection for increasing success metrics.

Friendly fraud is particularly difficult to detect because transactions are usually performed using valid customer credentials and legitimate payment instruments. Unlike stolen card fraud, friendly fraud often resembles normal customer behavior during the authorization stage. Fraud detection systems must therefore rely on subtle behavioral and transactional anomalies rather than obvious indicators of compromise [4].

The quality of such systems heavily depends on how good is selection of informative parameters capable of distinguishing fraudulent disputes from legitimate customer actions. Poor parameter selection may result in high false-positive rates, low precision-recall and thus high chargeback rates which lead to financial loss. Parameter selection represents one of the most important stages in the construction of intelligent fraud detection systems.

Parameters, also called features, describe measurable transaction characteristics that can be used for classification purposes. Transaction-related parameters describe financial and temporal characteristics of customer activity, such as purchase frequency or transaction amount. Transaction aggregation techniques can significantly improve fraud detection performance by identifying unusual transactional behavior over time [5].

For example, repeated purchases followed by refund requests within short intervals may indicate potentially abusive behavior. Behavioral analytics focuses on identifying deviations from normal user activity patterns. Common behavioral parameters include device and fingerprint consistency, IP address changes and login behaviour.

According to López de Prado, behavioral signals are especially valuable in modern financial machine learning systems because fraudulent activity often cannot be detected through transactional analysis alone [3]. Behavioral inconsistencies may therefore provide stronger predictive indicators than payment data itself, in some cases.

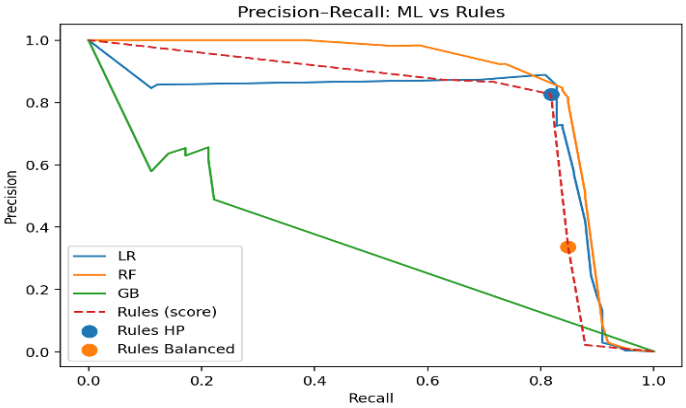


Fig.1. Rule based system with selected features versus untuned Machine Learning models

Historical behavioral information becomes particularly useful when working with highly imbalanced fraud datasets, where fraudulent transactions represent only a very small percentage of total activity [2]. In my previous paper [6] I briefly touched feature selection during an analysis of another topic. In order to determine strongest predictive features, Information Value(IV) was used.

$$IV = \sum_{i=1}^{n} (p_i - q_i) * \frac{p_i + \varepsilon}{q_i + \varepsilon}) \# (1)$$

It can be seen on the graph, that after the feature selection, even though quite simplistic, the rule-based system shows results on par with untuned ML models, while being much faster to run and more explainable. In my future papers I plan to analyze which features are the most important for different fraud types.

1. Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. Decision Support Systems, 50(3), 602–613.
2. Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2018). Credit card fraud detection: A realistic modeling and a novel learning strategy. IEEE Transactions on Neural Networks and Learning Systems, 29(8), 3784–3797.
3. López de Prado, M. (2018). Advances in financial machine learning. Wiley.
4. 4. Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. Artificial Intelligence Review, 34(1), 1–14.
5. Whitrow, C., Hand, D. J., Juszczak, P., Weston, D., & Adams, N. M. (2009). Transaction aggregation as a strategy for credit card fraud detection. Data Mining and Knowledge Discovery, 18(1), 30–55.
6. Gaina, G., & Masiuk, D. Evaluating rule-based vs. machine learning approaches for fraudulent transaction detection

Протокол розподіленого зберігання медичних даних

УДК 004.056

Микита Ціхоцький

Вінницький національний технічний університет, nik.tsikhotskiy15@gmail.com

Медичні інформаційні системи працюють з даними, для яких критичними є конфіденційність, цілісність, контрольований доступ і можливість відновлення. У розгалужених інфраструктурах зберігання повної копії медичного файлу на одному вузлі створює окрему точку компрометації, оскільки несанкціонований доступ до такого вузла може призвести до отримання всього об'єкта зберігання [1].

Метою роботи є розробка протоколу захищеного розподіленого зберігання медичних даних, який поєднує попереднє зашифрування файлу, пороговий розподіл зашифрованого вектору на частки та перевірку цілісності перед відновленням. Наукова новизна полягає у використанні порогової (*k*,*n*)-схеми з перекриттям блоків для контрольованого зберігання зашифрованого медичного файлу, за якого службові дані та контрольна частка залишаються у суб'єкта, що має право на відновлення.

У межах протоколу вхідний медичний файл F подається як послідовність байтів. У протоколі беруть участь уповноважений суб'єкт *M* та множина учасників зберігання $\mathbf{P} = \{P_1, P_2, \ldots, P_n\}$. Суб'єкт *M* готує файл до розподіленого зберігання, зберігає службові дані та має право на відновлення. Учасники P_i виконують роль вузлів, у яких розміщуються окремі частки розподіленого набору.

На першому етапі суб'єкт *M* виконує зашифрування файлу з використанням секретного ключа *K*: $\mathbf{C} = Enc_K(\mathbf{F})$, де C - зашифрований вектор медичного файлу [2]. Зашифрування виконується до розподілу, оскільки учасники зберігання не повинні працювати з відкритим медичним вмістом. У подальшій процедурі між вузлами розміщується не початковий файл, а частки його зашифрованого вектору.

Після зашифрування до вектору C застосовується порогова (*k*,*n*)-схема розподілу з перекриттям блоків [3]: $Share_{k,n}(\mathbf{C}) = \{\mathbf{S}, \mathbf{Q}, S_D, \mathbf{MB}\}$, де $\mathbf{S} = \{S_1, S_2, \ldots, S_n\}$ - множина часток, $\mathbf{Q} = \{Q_1, Q_2, \ldots, Q_n\}$ - множина порядкових номерів часток, S_D - частка, що зберігається у суб'єкта *M* та використовується під час відновлення, MB - службові дані розподіленого набору. Параметр *n* визначає загальну кількість сформованих часток, а параметр *k* - мінімальну кількість коректних часток, необхідну для відновлення зашифрованого вектору.

Для кожного учасника формується запис зберігання $\mathbf{R}_i = \{ID, q_i, \mathbf{S}_i\}$, де *ID* визначає належність частки до конкретного розподіленого медичного файлу, q_i задає її порядковий номер, а $\mathbf{S}_i$ є самою часткою. Розміщення часток у вузлах зберігання подається як $P_i \leftarrow \mathbf{R}_i, \quad i = 1, 2, \ldots, n$. Учасникам передаються лише

дані, необхідні для зберігання відповідної частки; ключ розшифрування, частка S_D і службові дані залишаються у суб'єкта *M*.

Службові дані MB формуються на етапі розподілу та містять ідентифікатор розподіленого набору *ID*, параметри *k* і *n*, таблицю відповідності між номерами часток і вузлами зберігання I_{SQ}, а також геш-значення часток і частки S_D. Таблиця I_{SQ} потрібна для перевірки відповідності отриманої частки очікуваному номеру та вузлу зберігання. Геш-значення фіксують стан часток у момент формування набору, тому під час відновлення дозволяють виявити пошкодження або підміну даних до їх об'єднання: $h_i = Hash(S_i)$, $h_D = Hash(S_D)$. Для відновлення суб'єкт *M* за ідентифікатором *ID* визначає потрібний розподілений набір і звертається до учасників зберігання. Отримані записи перевіряються за службовими даними: спочатку встановлюється відповідність *ID*, q_i та I_{SQ}, після чого перевіряється збіг геш-значень. До відновлення допускаються лише ті частки, для яких підтверджено належність до відповідного набору та цілісність. Після перевірки формується множина коректних часток $\mathbf{S}_{кор}$, і відновлення виконується лише за умови $\left|\mathbf{S}_{кор}\right| \geq k$.

Якщо порогова умова виконується, суб'єкт M відновлює зашифрований вектор з використанням множини коректних часток, частки S_D та службових даних [3]: $\mathbf{C} = \mathrm{Re}c_{k,n}(\mathbf{S}_{кор}, S_D, \mathbf{MB})$. Після цього виконується розшифрування зашифрованого вектору з використанням секретного ключа *K*: $F = Dec_K(\mathbf{C})$. У результаті відновлюється початковий медичний файл.

Запропонований протокол забезпечує зберігання медичного файлу без розміщення його повної копії на одному вузлі. Попереднє зашифрування обмежує доступ учасників зберігання до відкритого вмісту, порогова (*k*,*n*)-схема дозволяє відновити файл за наявності не менше ніж k коректних часток, а службові дані та частка S_D забезпечують контроль процедури відновлення. Використання порядкових номерів, таблиці відповідності та геш-значень дозволяє перевіряти належність і цілісність часток до запуску відновлення, що зменшує ризик використання пошкоджених або підмінених даних.

1. Лужецький В. Підходи до вирішення інструментальних завдань телемедицини. Матеріали III Всеукр. Медико-технічна наук.-практ. конф., м. Вінниця, 5–6 квіт. 2024 р. Вінниця: Едельвейс, 2024. С. 11–14.
2. Luzhetskyi V., Tsikhotskyi M. Image encryption and distribution method based on LFSR and counters. Information Technologies and Computer Engineering. 2025. Vol. 22, No. 3. P. 77-88.URL: https://doi.org/10.31649/vitce/3.2025.77 (дата звернення: 06.05.2026).
3. Лужецький В.А., Ціхоцький М.С. Розподіл секретного вмісту даних за (k,n)-схемою з використанням зашифрованих блоків. Вісник Вінницького політехнічного інституту. 2025. № 5. С. 113-120. URL: https://doi.org/10.31649/1997-9266-2025-182-5-113-120 (дата звернення: 08.05.2026).

Автоматизоване реагування на інциденти безпеки з використанням Suricata та SIEM Wazuh

УДК 004.056

Базилевський Д. В.[1], Цаволик Т. Г.[2]

Західноукраїнський національний університет, [1]bazilevskijdavid6@gmail.com

З огляду на стрімке зростання кількості кіберзагроз, спрямованих на критичну інфраструктуру та корпоративні мережі, автоматизація процесів реагування на інциденти стає важливим складником сучасної кібербезпеки [3]. Згідно зі звітами провідних організацій у сфері кібербезпеки, автоматизація атак зловмисниками вимагає від захисників впровадження систем, здатних реагувати на інциденти у режимі реального часу. Традиційний ручний моніторинг журналів подій уже не є ефективним через високу інтенсивність трафіку та складність векторів атак. Тому розробка та впровадження систем автоматизованого реагування на базі рішень з відкритим кодом є актуальним завданням для сучасних фахівців з інформаційної безпеки.

Автоматизоване реагування на інциденти безпеки — це процес виявлення, аналізу та нейтралізації загроз без безпосередньої участі людини на початкових етапах. Основна мета полягає у мінімізації часу перебування зловмисника в мережі (Dwell Time) та зменшенні навантаження на аналітиків центру моніторингу безпеки (SOC). Ключовими компонентами такої системи є: 1) джерела даних — мережеві сенсори, журнали ОС, антивірусні системи; 2) аналітичне ядро — системи класу SIEM (Security Information and Event Management), що корелюють події; 3) механізми дії — системи IDS/IPS (Intrusion Detection/Prevention System), що здатні блокувати трафік у режимі реального часу [5].

Використання Suricata IDS/IPS як активного засобу захисту. Suricata — це високопродуктивний механізм моніторингу мережевої безпеки, який підтримує виявлення вторгнень (IDS), запобігання вторгненням (IPS) та моніторинг мережевої безпеки (NSM) [1]. У режимі IPS (Inline mode) система забезпечує не лише фіксацію підозрілої активності, а й автоматичне блокування IP-адрес або розірвання з'єднань у реальному часі. Приклад правила для блокування SQL Injection:

```
drop http any any -> 10.10.20.2 80 \
    (msg:"LAB SQLi auth bypass in payroll_app.php"; \
    flow:established,to_server; \
    http.uri; content:"/payroll_app.php"; nocase; \
    http.request_body; pcre:"/(\%27|')(\s|\+)*(or|OR)(\s|\+)*(1=1|'1'='1)/i"; \
    classtype:web-application-attack; priority:1; \
    sid:1001030; rev:1;)
```

Рис. 1. Правило Suricata для виявлення SQL Injection

Команда drop ініціює негайне блокування трафіку, що є базовим етапом автоматизованого реагування [1].

Інтеграція з SIEM Wazuh для візуалізації та управління. Wazuh є платформою для моніторингу безпеки, яка інтегрує функції збору журналів, аналізу файлової цілісності та виявлення вразливостей. У даній роботі Wazuh

виступає як центральна консоль SOC-аналітика [2]. Suricata аналізує трафік та формує події у форматі eve.json, після чого Wazuh Agent передає їх на Wazuh Manager для обробки декодерами та правилами кореляції. Отримана інформація відображається на Wazuh Dashboard із зазначенням IP-адреси джерела атаки, типу загрози, часу та статусу реагування. Для ефективного реагування у Wazuh налаштовані спеціалізовані панелі моніторингу (Dashboards), які дозволяють швидко ідентифікувати найбільш атаковані сервіси, географічне розташування джерел атак (GeoIP) та динаміку спрацювань правил Suricata за рівнем критичності. Це забезпечує повну видимість мережевих інцидентів та дозволяє проводити цифрову криміналістику (Digital Forensics) після автоматичного блокування загроз [4].

Поєднання Suricata IPS та SIEM Wazuh створює багаторівневу систему захисту мережі. Suricata забезпечує швидку автоматичну реакцію на рівні мережевого трафіку, а Wazuh надає аналітичний інструментарій для моніторингу та розслідування інцидентів.

1. OISF. Suricata User Guide – Intrusion Detection and Prevention. [Електронний ресурс]. – Режим доступу: https://docs.suricata.io/en/latest/ips/index.html.
2. Wazuh Inc. Integration of Suricata with Wazuh for Network Threat Detection. [Електронний ресурс]. – Режим доступу: https://documentation.wazuh.com/current/proof-of-concept-guide/integrate-network-ids-suricata.html.
3. ENISA Threat Landscape Report 2025. [Електронний ресурс]. – Режим доступу: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2025.
4. MITRE Corporation. ATT&CK Framework – Command and Control, Initial Access and Web Attacks Techniques. [Електронний ресурс]. – Режим доступу: https://attack.mitre.org/
5. Scarfone K., Mell P. Guide to Intrusion Detection and Prevention Systems (IDPS). National Institute of Standards and Technology (NIST SP 800-94). – Gaithersburg: NIST, 2007. – 127 p. [Електронний ресурс]. – Режим доступу: https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-94.pdf.

Нормативно-правове регулювання кіберзахисту систем штучного інтелекту в Україні

УДК 004.8:342 (043.2)

Артем Жилін[1], Олександра Ценцеря[2]

Київський політехнічний інститут імені Ігоря Сікорського,
[1]zhylinartem@gmail.com, [2]o.tsentseria.s01@kpi.ua

Метою роботи є аналіз нормативно-правової бази кіберзахисту систем штучного інтелекту (далі – ШІ) в Україні та визначення напрямів її удосконалення.

Станом на 2026 рік нормативно-правове регулювання кіберзахисту систем ШІ в Україні перебуває на етапі формування та характеризується фрагментарністю, переважанням документів публічної політики над обов'язковими нормативно-правовими актами.

Таблиця 1

Нормативно-правове регулювання штучного інтелекту в Україні

1. Нормативно-правові акти (НПА)	
Суб'єкт правотворчості	Повна назва
КМУ	Постанова Кабінету Міністрів України (далі – КМУ) від 29.10.2024 № 1238 «Про реалізацію експериментального проекту щодо організації проведення досліджень високотехнологічних засобів методом "Sandbox"»
ДССЗЗІ	Наказ Адміністрації Державної служби спеціального зв'язку та захисту інформації України (далі – ДССЗЗІ) від 23.02.2026 № 154 «Про затвердження Рекомендацій з кіберзахисту інформаційно-комунікаційних систем, які використовують технології штучного інтелекту»
2. Документи публічної політики (ДПП)	
КМУ	Розпорядження КМУ від 02.12.2020 № 1556-р «Про схвалення Концепції розвитку штучного інтелекту в Україні»
КМУ	Розпорядження КМУ від 12.05.2021 № 438-р «Про затвердження плану заходів з реалізації Концепції розвитку штучного інтелекту в Україні на 2021–2024 роки»
КМУ	Розпорядження КМУ від 09.05.2025 № 457-р «Про затвердження плану заходів з реалізації Концепції розвитку штучного інтелекту в Україні на 2025–2026 роки»
Мінцифри	Дорожня карта з регулювання штучного інтелекту в Україні: Bottom-Up Підхід
Мінцифри	Біла книга з регулювання ШІ в Україні: бачення Мінцифри

Чинна нормативна база охоплює або кіберзахист інформаційно-комунікаційних систем загалом, без урахування специфіки ШІ-компонентів, або розвиток і впровадження ШІ без акцентованих вимог до кіберзахисту. Перетин цих двох сфер у єдиній обов'язковій нормі фактично відсутній, що й визначає мету досліджень.

На рівні нормативно-правових актів єдиним документом, що прямо адресує кіберзахист ШІ-систем, є Наказ ДССЗЗІ від 23.02.2026 № 154 «Про затвердження Рекомендацій з кіберзахисту інформаційно-комунікаційних систем, які використовують технології штучного інтелекту» [1]. Він затверджує рекомендації, а не обов'язкові вимоги, що унеможливлює його застосування як підстави для юридичної відповідальності. Постанова КМУ від 29.10.2024 № 1238 запроваджує механізм регуляторної пісочниці («Sandbox») для тестування високотехнологічних засобів, до яких можуть належати системи ШІ, однак без спеціальних вимог безпеки до таких систем це процедурний механізм,

а не регуляторний стандарт кіберзахисту. Документи публічної політики визначають безпеку та довіру до ШІ як пріоритети державної політики та фіксують напрями, а не норми [2-6]. При цьому орієнтація на поетапну імплементацію EU AI Act [7], зокрема в частині вимог до безпеки ризикових систем, простежується лише у [2] та [6].

З огляду на виявлені прогалини, подолання структурної недостатності регулювання кіберзахисту систем ШІ потребує узгодженої дії на кількох рівнях нормативної ієрархії. На законодавчому рівні – прийняття Верховною Радою України спеціального закону про ШІ із розділом, що встановлює вимоги кіберзахисту ШІ-систем залежно від рівня ризику, до якого вони належать, відповідно до ризик-орієнтованої класифікації EU AI Act [7]. На підзаконному рівні – прийняття окремої постанови КМУ із мінімальними обов'язковими вимогами кіберзахисту для систем ШІ, що спиралися б на NIST AI RMF 1.0 в частині управління ризиками, або на ISO/IEC 42001:2023.

1. «Про затвердження Рекомендацій з кіберзахисту інформаційно-комунікаційних систем, які використовують технології штучного інтелекту». Наказ ДССЗЗІ від 23.02.2026 № 154. URL: https://cip.gov.ua/ua/docs/nakaz-administraciyi-derzhspeczv-yazku-vid-23-02-2026-154-pro-zatverdzhennya-rekomendacii-z-kiberzakhistu-informaciino-komunikaciinikh-sistem-yaki-vikoristovuyut-tekhnologiyi-shtuchnogo-intelektu (дата звернення: 14.05.2026).
2. Про схвалення Концепції розвитку штучного інтелекту в Україні: Розпорядження КМУ від 02.12.2020 № 1556-р. URL: https://zakon.rada.gov.ua/laws/show/1556-2020-%D1%80#Text (дата звернення: 15.05.2026).
3. Про затвердження плану заходів з реалізації Концепції розвитку штучного інтелекту в Україні на 2021–2024 роки: Розпорядження КМУ від 12.05.2021 № 438-р. URL: https://zakon.rada.gov.ua/laws/show/438-2021-%D1%80#Text (дата звернення: 15.05.2026).
4. Про затвердження плану заходів з реалізації Концепції розвитку штучного інтелекту в Україні на 2025–2026 роки: Розпорядження КМУ від 09.05.2025 № 457-р. URL: https://zakon.rada.gov.ua/laws/show/457-2025-%D1%80#Text (дата звернення: 15.05.2026).
5. Дорожня карта з регулювання штучного інтелекту в Україні: Bottom-Up Підхід. Мінцифри, 2023. URL: https://storage.thedigital.gov.ua/files/2/22/363bbcaec30bf9d4e598375feca c3227.pdf (дата звернення: 15.05.2026).
6. Біла книга з регулювання ШІ в Україні: бачення Мінцифри. Версія для консультацій. Мінцифри, 2024. URL: https://storage.thedigital.gov.ua/files/a/ba/d5da75c2613e331bb89258f950adcbae.pdf (дата звернення: 15.05.2026).
7. EU Artificial Intelligence Act: up-to-date developments and analyses. URL: https://artificialintelligenceact.eu/ (дата звернення: 16.05.2026).

Ризики компрометації API-ключів у сервісах Google Cloud із використанням генеративного ШІ та методи їх мінімізації

УДК 004.056.5:004.8

Михайло Рокош[1]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]mykhailo.rokosh@tntu.edu.ua

Активне впровадження генеративного штучного інтелекту у бізнес-процеси, розроблення прототипів і внутрішню автоматизацію створює не лише нові можливості, а й нові ризики для інформаційної та фінансової безпеки. У зоні найбільшого ризику перебувають малі команди, стартапи та користувачі без глибоких технічних знань, які користуються хмарними сервісами на кшталт Google AI Studio, Firebase або Google Cloud Console. У таких випадках API-ключ часто сприймається як допоміжний технічний параметр, хоча насправді він може створювати значні фінансові наслідки у разі компрометації.

Один із найпоширеніших сценаріїв пов'язаний із компрометацією API-ключа Google Maps, вбудованого у клієнтську частину вебсайту чи мобільного застосунку без належних обмежень. Якщо згодом у тому самому проєкті активується Gemini API, цей ключ може бути використаний для платних запитів до сервісів генеративного ШІ. У такому разі зловмисник після отримання доступу до ключа може виконувати запити за рахунок власника проєкту, що часто призводить до значних фінансових втрат, які можуть вимірюватися сотнями або навіть тисячами доларів за годину.

Дослідження Truffle Security у 2026 році показало, що старі Google API-ключі, які раніше сприймалися як відносно безпечні ідентифікатори для клієнтських сервісів, можуть створювати значно більші ризики після підключення Gemini. Дослідники повідомили про 2863 публічні Google API-ключі з доступом до Gemini [1]. Подібну проблему описала компанія CloudSEK: у 22 Android-застосунках було виявлено 32 жорстко вбудовані Google API-ключі, які потенційно відкривали доступ до сервісів Gemini. Сукупна аудиторія цих застосунків перевищувала 500 млн користувачів [2].

Окремою проблемою є неправильне розуміння бюджетних сповіщень. У Google Cloud бюджет за замовчуванням слугує механізмом сповіщення, а не автоматичним припиненням витрат. Документація Google Cloud прямо вказує, що встановлення бюджету не обмежує автоматично використання або витрати [3]. Для користувача без досвіду роботи з хмарною інфраструктурою слово "бюджет" може створити враження жорсткого фінансового ліміту, хоча насправді воно часто означає лише повідомлення про досягнення певної межі.

Для зменшення ризиків слід дотримуватись принципу "один ключ – одне призначення". Ключ, створений для Google Maps або Places API, не повинен використовуватися для інших сервісів. Для кожного середовища, сервісу та сценарію доцільно створювати окремий ключ із чітким найменуванням, що відображає його призначення. Також важливим є встановлення двох типів обмежень: обмеження за API та обмеження за джерелом запиту. Перше визначає, до яких саме API має доступ ключ, наприклад тільки до Places API або Maps JavaScript API. Друге означає, звідки дозволено використовувати ключ: з

конкретного домену вебсайту, IP-адреси сервера або мобільного застосунку. Google Maps Platform також рекомендує обмежувати API-ключі за застосунком і за доступними API [4].

Для запитів до Gemini API небажано використовувати пряме звернення з клієнтської частини вебзастосунку. Безпечнішою є архітектура, у якій ключ зберігається на сервері, у безсерверній функції або в захищеному сховищі секретів. Клієнтський застосунок має звертатися до власної серверної частини, а вона вже робить запит до Gemini API після перевірки користувача, обмеження частоти запитів та інших правил доступу. Якщо ключ потрапляє у зібраний JavaScript-код або мобільний застосунок без належних обмежень, його можуть знайти автоматизовані сканери.

Важливим доповненням до бюджетних сповіщень є квоти. Якщо API підтримує обмеження за кількістю запитів, їх потрібно встановлювати відповідно до реального сценарію використання, особливо для прототипів і тестових проєктів. Документація Google Cloud зазначає, що для окремих API можна явно обмежувати кількість запитів, щоб контролювати оплачуване використання [5]. Крім того, доцільно регулярно перевіряти список активованих API, історію використання, налаштування сповіщень, ролі доступу та застарілі тестові проєкти. Ключі, які більше не використовуються або могли бути відкриті у публічному коді, потрібно видаляти або замінювати.

Отже, розвиток генеративного ШІ суттєво прискорює створення цифрових продуктів, але також підвищує ризики, пов’язані з API-ключами. Основними заходами захисту є обмеження ключів за API і джерелом запиту, розділення проєктів за призначенням, серверне зберігання ключів для ШІ-сервісів, встановлення квот, регулярний аудит і видалення застарілих ключів. У контексті масового використання генеративного ШІ безпека API-ключів має розглядатися не як другорядне технічне налаштування, а як основа інформаційної та фінансової безпеки.

1. Truffle Security. Google API Keys Weren’t Secrets. But then Gemini Changed the Rules. URL: https://trufflesecurity.com/blog/google-api-keys-werent-secrets-but-then-gemini-changed-the-rules (дата звернення: 06.05.2026).
2. CloudSEK. Hardcoded Google API Keys in Top Android Apps Now Expose Gemini AI. URL: https://www.cloudsek.com/blog/hardcoded-google-api-keys-in-top-android-apps-now-expose-gemini-ai (дата звернення: 06.05.2026).
3. Google Cloud. Create, edit, or delete budgets and budget alerts. URL: https://docs.cloud.google.com/billing/docs/how-to/budgets (дата звернення: 07.05.2026).
4. Google Maps Platform. API security best practices. URL: https://developers.google.com/maps/api-security-best-practices (дата звернення: 07.05.2026).
5. Google Cloud. Capping API usage. URL: https://docs.cloud.google.com/apis/docs/capping-api-usage (дата звернення: 07.05.2026).

Система автоматизованої протидії інформаційним впливам на основі штучного інтелекту

УДК 004.8:004.056

Євгеній Волкотруб[1], Леонід Куперштейн[2]

Вінницький національний технічний університет,
[1]zhenvolkotrub@gmail.com, [2]kupershtein@vntu.edu.ua

Гібридна війна, яку росія веде проти України, включає не лише бойові дії на фронті, але й масштабні інформаційно-психологічні операції. Їхньою метою є формування вигідного для агресора інформаційного потоку, деморалізація суспільства та підрив довіри до державних інститутів. Соціальні мережі стали ключовим майданчиком для поширення дезінформації та маніпулятивного контенту. За даними EUvsDisinfo, лише протягом березня–листопада 2023 року дослідники зафіксували 596 рекламних повідомлень із дезінформацією та пропагандою, спрямованих на українську аудиторію [1]. Організація системної протидії таким впливам передбачає координацію дій та автоматизацію рутинних операцій [2].

Особливої уваги заслуговує платформа ok.ru (Одноклассники) — підсанкційна російська соціальна мережа, офіційно заблокована в Україні з 2017 року. Незважаючи на блокування, частина аудиторії продовжує відвідувати платформу через VPN-сервіси та проксі-сервери, а сама мережа залишається активним каналом поширення кремлівських наративів. Так, лише у 2023–2024 роках дослідники зафіксували десятки пропагандистських повідомлень антимобілізаційного характеру, розміщених саме в групах Одноклассників та VKontakte [1]. Уряд України та міжнародні організації фіксують, що попри формальне блокування, доступ до цих ресурсів зберігається для значної частини користувачів [3].

Традиційні методи ручного моніторингу та контрнаративу є ресурсомісткими і не дозволяють охопити весь масив шкідливого контенту в реальному часі. Великі мовні моделі (LLM) пропонують якісно новий підхід. Завдяки навчанню на великих корпусах текстів вони здатні розуміти контекст публікацій і генерувати природнзхомовні відповіді, наближені до людського стилю спілкування [4]. Інтеграція LLM у системи з веб-автоматизацією відкриває можливість масштабувати протидію на тисячі публікацій одночасно, що суттєво знижує операційне навантаження на фахівців з інформаційної безпеки.

Метою роботи є вдосконалення процесів протидії інформаційним впливам за рахунок розробки та впровадження автоматизованої інтелектуальної системи, яка у реальному часі аналізує контент соціальних мереж та генерує контекстно релевантні спростування або альтернативні наративи на основі генеративного штучного інтелекту.

Технологія протидії інформаційним впливам побудована за конвеєрним принципом і охоплює п'ять послідовних етапів, що утворюють замкнений цикл протидії (рис. 1).

Архітектура системи, що реалізує запропоновану технологію, включає такі складові: модуль веб-взаємодії, модуль генерації відповідей на основі LLM та

модуль управління акаунтами. Загальна архітектура передбачає конвеєрне опрацювання.

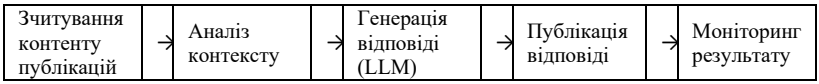


Рис. 1. Технологія застосування системи протидії інформаційним впливам

Модуль веб-взаємодії реалізовано на базі Selenium WebDriver із ChromeDriver. Для автентифікації в соціальній мережі застосовується механізм сесійних cookie, що дозволяє одночасно задіяти велику кількість акаунтів без повторного введення облікових даних. Для вилучення тексту публікацій застосовується багаторівневий парсинг за CSS-селекторами: основний текст публікації, підписи до медіа, тексти репостів та атрибути зображень. Це суттєво підвищує повноту аналізу, оскільки значна частина пропагандистського контенту поширюється у форматі зображень із текстовими підписами.

Для генерації відповідей використовується велика мовна модель Gemini від Google [5]. Gemini є мультимодальною моделлю, що підтримує розуміння тексту, зображень та структурованих даних. Доступ до моделі здійснюється через Gemini API, що забезпечує гнучку інтеграцію з автоматизованими системами без потреби у локальному розгортанні.

Ключовим елементом є спеціалізований промпт, що реалізує риторичну стратегію «Так, але…». Тобто модель спочатку демонструє поверхневу згоду з темою публікації (погода, кулінарія, свята), а потім переводить увагу читача на факти, які спростовують ворожий наратив. Така стратегія підвищує сприйнятливість аудиторії до контрнаративу, оскільки не сигналізує відразу про полемічний намір. Природність тексту додатково забезпечується повільним посимвольним введенням і випадковими затримками 3–6 с між публікаціями.

Модуль управління акаунтами забезпечує ротацію профілів із бібліотеки cookie-сесій. У разі недоступності LLM-сервісу система автоматично перемикається на резервну базу заздалегідь підготовлених відповідей, що гарантує безперервність процесу. Після кожної успішної публікації система зберігає знімок екрана для подальшого аудиту та верифікації.

Проведені випробування на платформі ok.ru підтвердили працездатність системи. Успішно опрацьовано понад 85% тестових публікацій із генерацією контекстно коректних відповідей. Середній час реакції на одну публікацію склав 8–12 секунд. Модуль моніторингу фіксує реакцію аудиторії на розміщені коментарі, що дозволяє оцінювати ефективність різних риторичних стратегій та коригувати промпт у реальному часі.

Розроблена система автоматизованої протидії інформаційним впливам поєднує можливості мовної моделі Gemini з технологіями веб-автоматизації та забезпечує протидію пропагандистських наративів у реальному часі. Конвеєрна п'ятиетапна технологія — від зчитування контенту до моніторингу результату — повністю автоматизує цикл протидії та знижує операційне навантаження на фахівців. Застосування риторичної стратегії підвищує природність та переконливість згенерованих відповідей. Масштабованість за рахунок багатоакаунтного підходу суттєво збільшує охоплення протидії. Перспективами

подальших досліджень є інтеграція мультимодальних LLM для аналізу зображень і відео, а також розробка модуля класифікації токсичного контенту.

1. EUvsDisinfo. How Russian Special Information Operations Try to Undermine Mobilisation in Ukraine. — 2024. URL: https://euvsdisinfo.eu/how-russian-special-information-operations-try-to-undermine-mobilisation-in-ukraine (дата звернення: 05.05.2026).
2. Kuperstein L. M., Lukichov V. V., Radetska A. O., Dudatyev A. V. System for Organizing Cyber Operations in the Context of Military Aggression // Science and Innovation. — 2025. — Vol. 21, № 3. — P. 86–98. https://doi.org/10.15407/scine21.03.086 (дата звернення: 05.05.2026).
3. Freedom House. Ukraine: Freedom on the Net 2023. — 2023. URL: https://freedomhouse.org/country/ukraine/freedom-net/2023 (дата звернення: 05.05.2026).
4. Baryshev Y., Kupershtein L., Maidanovych V., Voitovych O., Prokopenko S. Information System for the Fact-checker Support // CEUR Workshop Proceedings. — 2023. — Vol. 3646. — P. 127–138. URL: https://ceur-ws.org/Vol-3646/Paper_13.pdf (дата звернення: 05.05.2026).
5. Google. Gemini API Reference. — 2024. URL: https://ai.google.dev/gemini-api/docs (дата звернення: 05.05.2026).

Аналіз безпеки serverless-архітектур на основі моделювання подій

УДК 004.056:004.738.5 (043.2)

Петро Венгерський[1], Святослав Златоус[2]

Львівський національний університет імені Івана Франка,
[1]petro.venherskyy@lnu.edu.ua, [2]sviatoslav.zlatous@lnu.edu.ua

Сучасний розвиток хмарних обчислень зумовив широке впровадження serverless-архітектур, що базуються на концепції Function-as-a-Service (FaaS). Такий підхід дозволяє створювати масштабовані програмні системи без необхідності управління серверною інфраструктурою, що значно спрощує процес розробки та експлуатації додатків [1]. Водночас використання serverless-архітектур супроводжується появою нових викликів у сфері кібербезпеки, які пов'язані з динамічністю виконання функцій, складною структурою взаємодій між компонентами системи та обмеженим контролем користувачів над середовищем виконання [3]

У сучасних дослідженнях безпеки serverless-систем основна увага приділяється окремим аспектам їх функціонування, зокрема ізоляції функцій, управлінню доступом та аналізу продуктивності [3]. Водночас питання комплексного аналізу поведінки системи, що враховує взаємодії між функціями, подіями та сервісами хмарної інфраструктури, залишається недостатньо дослідженим [2]. Це ускладнює виявлення потенційних загроз безпеці, які можуть виникати у результаті нетипових сценаріїв функціонування системи.

У роботі запропоновано підхід до аналізу безпеки serverless-архітектур, що базується на дослідженні подій та взаємодій між компонентами системи. Основна ідея підходу полягає у використанні централізованого моніторингу

подій для формування моделі поведінки системи та виявлення відхилень від нормального режиму її функціонування.

Для перевірки ефективності запропонованого підходу було реалізовано експериментальне serverless-середовище у хмарній платформі AWS. Моніторинг функціонування системи здійснювався за допомогою сервісу AWS CloudWatch, який забезпечує збір та аналіз метрик і журналів подій.

Ingress Function Metrics (метрики інгрес лямбда-функції)

	Min	Max	Sum	Average (середнє)
Invocations (виклики)	23	23	23	23
Duration (тривалість)	821мс	821мс	821мс	821мс
Errors (помилки)	0	0	0	0
Throttles (тротлінг)	0	0	0	0

Processing Function Metrics (метрики процесінг лямбда-функції)

	Min	Max	Sum	Average (середнє)
Invocations (виклики)	15	127	339	56.5
Duration (тривалість)	4.81с	41.3с	116с	19.3с
Errors (помилки)	0	0	0	0

Consumer Function Metrics (метрики консюмер лямбда-функції)

	Min	Max	Sum	Average (середнє)
Invocations (виклики)	10	53	364	30.3
Duration (тривалість)	633мс	4.28с	28.2с	2.35с
Errors (помилки)	0	0	0	0

SQS Queue Metrics (метрики SQS черги)

	Min	Max	Sum	Average (середнє)
Messages sent (відправленні повідомлення)	0	127	340	0.39
Messages received (отримані повідомлення)	0	142	874	1.01
Messages deleted (видалені повідомлення)	0	0	0	0

DynamoDB Metrics (метрики DynamoDB бази данних)

	Min	Max	Sum	Average (середнє)
Consumed Read Capacity Units (Використані одиниці пропускної здатності читання)	0	0	0	0
Consumed Write Capacity Units (Використані одиниці пропускної здатності запису)	0	0.29	0.29	0
Put Item Successful Request Latency (Затримка успішного виконання операції запису)	32.2мс	32.2мс	32.2мс	32.2мс

Рис.1. Метрики функціонування serverless-системи у середовищі AWS CloudWatch

На рис. 1 наведено приклад панелі моніторингу AWS CloudWatch, що використовується для аналізу поведінки serverless-системи.

У процесі експерименту було досліджено поведінку системи у нормальному та аномальному режимах функціонування. У нормальному режимі середній час виконання функцій становив 120–250 мс, а частота викликів мала стабільний характер із незначними коливаннями. Структура взаємодій між компонентами залишалася сталою та відповідала визначеному сценарію обробки запитів.

Для моделювання аномалій було виконано штучне збільшення частоти викликів функцій у 2–3 рази, а також змінено послідовність взаємодій між компонентами системи. У результаті зафіксовано збільшення середнього часу виконання до 300–450 мс та появу нетипових послідовностей викликів функцій.

Аналіз отриманих результатів показав, що запропонований підхід дозволяє ефективно виявляти відхилення у функціонуванні serverless-систем. Зокрема, зміни у частоті викликів функцій, часі їх виконання та структурі взаємодій можуть бути використані як індикатори потенційних загроз безпеці.

1. Baldini I., Castro P., Chang K., et al. *Serverless computing: Current trends and open problems* // Research Advances in Cloud Computing. – 2017. – С. 1–20.
2. Jonas E., Schleier-Smith J., Sreekanti V., et al. *Cloud programming simplified: A Berkeley view on serverless computing* // arXiv:1902.03383. – 2019. – С. 1–28.
3. Zhang Y., Chen X., Li J. *Security and privacy in serverless computing: A systematic literature review* // ACM Computing Surveys. – 2023. – Vol. 55, No. 12. – С. 1–36.

Методи та алгоритми детектування кіберзагроз і реагування на інциденти інформаційної безпеки у мультихмарних середовищах

УДК 004.056.57

Венгерський Петро [1], Радченко Максим [2]

Львівський національний університет ім. Івана Франка,
[1]petro.venherskyy@lnu.edu.ua, [2]maksym.radchenko@lnu.edu.ua

Сучасні підприємства масово переходять до мультихмарних середовищ: понад 75 % організацій одночасно використовують ресурси Amazon Web Services, Microsoft Azure та Google Cloud Platform [1]. Різнорідність подій безпеки з хмарних джерел породжує критичну проблему alert fatigue - перевантаження аналітиків центрів оперативної безпеки хибними спрацюваннями. За даними результатами систематичного огляду, аналітики витрачають понад половину робочого часу на обробку хибних алертів, що безпосередньо збільшує середній час виявлення та усунення реальних інцидентів [2].

Метою роботи є розробка та програмна реалізація методів і алгоритмів детектування кіберзагроз і реагування на інциденти інформаційної безпеки у мультихмарних середовищах, що забезпечують зниження рівня хибних спрацювань та скорочення часу реагування.

Існуючі комерційні рішення класу CDR (Prisma Cloud, Microsoft Defender for Cloud, Google Chronicle) та відкриті системи (Wazuh, Falco) або не забезпечують формалізованої пре-фільтрації подій до їх запису в сховище, або реалізують post-ingestion фільтрацію без врахування семантичної схожості алертів від різних джерел. Це призводить до надлишкового накопичення дублюючих записів і знижує якість кореляції подій [3].

Наукова новизна роботи полягає у формалізації п'ятистадійного конвеєра пре-фільтрації подій безпеки як функціональної композиції із принципом short-circuit evaluation; розробці алгоритму оцінки семантичної схожості алертів на основі зваженої метрики Жаккара; 3) побудові FSM-моделі життєвого циклу інциденту з вбудованими SLA-метриками MTTA/MTTR.

Конвеєр пре-фільтрації формалізовано, як

$$F = F_5 \circ F_4 \circ F_3 \circ F_2 \circ F_1$$

де кожна стадія $F_i : \mathscr{E} \to \{0,1\}$ реалізує послідовно: порогову фільтрацію за рівнем критичності, фільтрацію за списком блокування `rule_ID`, зіставлення з шаблонами хибних спрацювань, пригнічення шуму методом ковзного часового вікна та оцінку схожості алертів. П'ята стадія є ключовим внеском роботи: для двох алертів *a* та *b* визначено зважену функцію схожості:

$$\mathrm{sim}(a, b) = 0{,}40 \cdot \mathbb{1}[\tau_a = \tau_b] + 0{,}30 \cdot \mathbb{1}[\theta_a = \theta_b] + 0{,}20 \cdot \mathbb{1}[\alpha_a = \alpha_b] + 0{,}10 \cdot J(T_a, T_b)$$

де τ - MITRE-тактика, θ - MITRE-техніка, α - ідентифікатор агента, $J(T_a, T_b)$ - індекс Жаккара токенів заголовку. При перевищенні порогу $\delta = 0{,}7$ виконується злиття семантично близьких алертів замість створення дублюючих записів.

Розроблено алгоритм об'єднання подій із гетерогенних джерел у єдиний нормалізований потік та механізм дедуплікації хмарних знахідок. Модель рушія детектування формалізує п'ять типів правил: порогове, паттерн, відсутність, кореляція, послідовність. Модель інциденту реалізовано як детермінований скінченний автомат $M = (Q, \Sigma, \delta, q_0, F)$ із шістьма станами та вбудованим контролем SLA.

Програмну реалізацію виконано у складі хмарно-нативної платформи uSafe. Основний модуль аналізу реалізовано у модулях alert_filters, alert_ingest та rule_engine. Також, модуль сканування хмарної безпеки, наприклад - AWS, охоплює 239 перевірок безпеки у 12 категоріях. Верифікацію алгоритмів проведено на синтетичних тестових наборах, що підтвердили коректність усіх стадій конвеєра.

Аналіз статистики FilterStats підтверджує ефективність розробленого конвеєра: pass rate складає ~14 %, тобто ~86 % подій відсіюється до їх запису в базу даних як події безпеки. Цей результат порівнянний із показниками ML-підходів (Carbon Filter: 82-84 % [3]) за суттєво вищої детермінованості та пояснюваності рішень. Розроблені методи впроваджено в продуктивному середовищі платформи uSafe та можуть бути застосовані у комерційних продуктах класу MSSP.

1. Illumio. The 2025 Global Cloud Detection and Response Report. URL: https://www.illumio.com/resource-center/2025-global-cloud-detection-and-response-report (дата звернення: 08.04.2026).
2. Tariq S., Chhetri M. B., Nepal S., Paris C. Alert Fatigue in Security Operations Centres: Research Challenges and Opportunities. ACM Computing Surveys. – 2025. – Vol. 57, No. 9. – DOI: 10.1145/3723158.
3. Yang L., Li Z., Chen H. Carbon Filter: Reducing False Positive Alerts in Security Operations Centers. IEEE Transactions on Information Forensics and Security. – 2022. – Vol. 17. – P. 2341–2354.

Виявлення несанкціонованого доступу та компрометації облікових записів завдяки SIEM системі

УДК 004.056

Степовенко Юрій[1], Ілля Фалендиш[2], Євгеній Юр'єв[3]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]urastepovenko1@gmail.com, [2]illia.falendysh02@gmail.com, [3]yurev05@gmail.com

У сучасному світі важко переоцінити роль кібербезпеки. Враховуючи кількість існуючих вірусів, атак та загроз комп'ютерні системи вимагають особливої уваги до своєї активності. Саме тому сучасні компанії часто вводять в експлуатацію SIEM системи, які завдяки централізованому збору логів та евристичному аналізу здатні виявляти аномалії в активності системи та повідомляти про це IT команду. Головною проблемою подібних систем є велика кількість хибних спрацювань - поведінка як інтернет трафіку, так і користувачів рідко буває стабільною та передбачуваною, проте автоматизований алгоритм чи штучний інтелект не можуть відрізнити істинні спрацювання від хибних.

Метою цієї роботи було розглянути критичну ланку подібних аномалій, а саме проаналізувати, яким чином відрізнити несанкціонований вхід до системи від легітимного завдяки SIEM системі, а також проаналізувати сучасні техніки та атаки, завдяки яким зловмисники проникають у приватні сервіси.

Найпростіший та найефективніший метод, щоб заволодіти обліковими даними користувача, фішинг. Сучасні фішингові методи важко охарактеризувати певними спільними рисами, адже в залежності від підготовки зловмисника, посилання у фішинговому листі буває важко відрізнити навіть підготованому користувачу. Деякі системи здатні помічати перехід за подібними посиланнями, і коли аналітик проаналізує його в ізольованому середовищі, та співставить приблизний час натискання на посилання з підозрілим входом, стане зрозуміло, що акаунт скомпрометовано. Проте бувають випадки, коли переходи за подібними посиланнями не фіксуються системою. Це відбувається тому, що не всі системи клієнта можуть мати встановлені агенти, які відслідковують підозрілу активність, або ж взагалі користувач перейде за посиланням з власного, а не корпоративного пристрою. В таких випадках, вхід на перший погляд не відрізнятиметься від звичайного. По-перше, необхідно дослідити, з якого пристрою виконувався вхід. По-друге, необхідно порівняти ID сесії при звичайних входах та під час підозрілого. Зазвичай одна сесія видається користувачеві на доволі довгий період, близько кількох тижнів. Агент, тобто інформація про користувача при вході, його браузер, операційна система та рендер, доволі сталі для конкретного акаунту. Також IP адреса може дати цінну інформацію про користувача, насамперед тому, що завдяки вебсайту https://ip2proxy.com/ можна дослідити чи використовує користувач VPN, а завдяки https://www.abuseipdb.com/ – отримати інформацію про зловмисну діяльність тих чи інших IP-адрес.

Що стосується методів протидії фішинговим атакам, то поряд з правилами цифрової гігієни, регулярної зміни пароля та аналізу листів на фішинг, безвідмовним способом захистити себе від несанкціонованого входу є впровадження двофакторної автентифікації. Проте все ще залишиться ризик перехоплення сесії та атак типу MITM, які не авторизуються до системи, а використовують вже існуючі сесії для обходу автентифікації. Якщо ж акаунт користувача вже скомпрометовано, необхідно розірвати з'єднання зі зловмисником. У деяких випадках компрометації, до акаунту входить лише один зловмисник з чистої IP адреси, що б замаскувати свій вхід. Інколи на акаунт "навалюються" десятки зловмисників, які виконують сканування, копіюють дані, встановлюють шкідливе ПЗ, і працюють на швидкість.

Проаналізуємо реальний інцидент безпеки, а саме несанкціонований вхід до SIEM системи. Імена скомпрометованих користувачів буде приховано, заради їх конфіденційності. У одному кейсі, було повідомлено про підозрілий вхід для певного користувача. Аналізуючи його активність, було помічено одну невдалу спробу входу через регіональну фільтрацію, з Британської IP адреси, проте сам користувач зазвичай має активність у Італії. Аналізуючи інші логіни з Італії, увагу привернув один неприродній вхід з нового девайсу. IP адреса у розглянутій моніторинговій системі позначалась як Італійська (рис. 1).

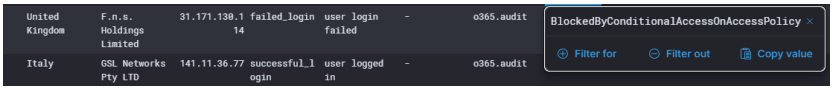


Рис. 1. Фіксація невдалого входу користувача

Проаналізувавши активність за допомогою IP2Proxy та AbuseIPDB було помічено, що насправді IP адреса є VPN, яка насправді походить з Німеччини, та має сумнівну репутацію (рис. 2).

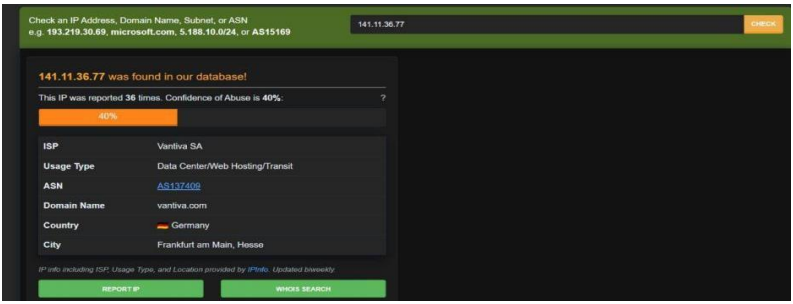


Рис. 2. Виявлена підозріла IP адреса

В результаті подальших перевірок було визначено, що користувач в результаті переходу за фішинговим посиланням ввів свої дані. Після виконання розриву сесій, зміни пароля та подальших перевірок атаку вдалось припинити.

Підбиваючи підсумки, навіть для такого простого процесу як авторизація можна використати десятки різних технік та атак, які дозволять обійти навіть найбільш досконалі системи захисту. Проте, якщо вчасно отримати повідомлення про підозрілі дії, можна виконати розрив сесії до виконання шкідливих дій. Тому актуальність використання моніторигових систем для аналізу стану безпеки зростатиме і надалі.

Розгортання системи моніторингу безпеки VPN-з'єднання на базі WireGuard

УДК 004.056

Каріна Крушельницька[1], Марина Деркач[2], Віталій Тимощук[3]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]karina.kryshel@gmail.com, [2]m_derkach@tntu.edu.ua, [3]Tymoshchuk@tntu.edu.ua

У сучасних мережевих інфраструктурах VPN-з'єднання забезпечують створення захищених каналів передачі даних через Інтернет. У свою чергу, системи моніторингу дозволяють контролювати віддалений доступ до ресурсів, поєднуючи у єдину безпечну мережу територіально віддалені офіси, а також відстежувати підозрілу активність та своєчасно виявляти загрози.

Для реалізації такого підходу при розгортанні системи моніторингу безпеки використано систему виявлення вторгнень на основі мережі Suricata, яка демонструє широкі аналітичні можливості для моніторингу мережевого

середовища, зокрема здатна перевіряти не лише структуровані дані пакетів, а й додаткові атрибути трафіку, такі як сертифікати TLS, HTTP-запити, DNS-транзакції, що дозволяє виявляти складні атаки на різних рівнях мережевої взаємодії [1]. Інструмент функціонує на прикладному рівні моделі OSI, забезпечуючи глибоку видимість і аналіз кількох пакетів одночасно, незважаючи на роботу на програмному рівні, Suricata зберігає повний доступ до інформації заголовків пакетів, що дозволяє детально аналізувати протоколи транспортного, мережевого та навіть прикладного рівнів, включно з можливістю оцінки шифрованих даних. Для VPN-з'єднання обрано протокол WireGuard, що характеризується гнучкістю та високою продуктивністю, забезпечуючи як конфіденційність, так і цілісність даних [2]. На відміну від інших протоколів WireGuard заплутує метадані пакетів, включаючи довжину передачі та IP-адреси відправників і одержувачів, тому ключі для кожного пакету узгоджуються в приватному порядку без участі третіх сторін, що робить його набагато швидше, а також є більш безпечним, оскільки немає потенційних витоків при обміні ключами з центральним сервером.

У розгорнутій системі реалізовано механізм регулярного моніторингу параметрів з'єднання, що включає IP- та MAC-адреси, ендпоінти та часові характеристики рукостискання. Аналіз даних дозволяє виявляти аномалії, зокрема несанкціоновані MAC-адреси, їх зміну або появу нових пристроїв. Також система виконує перевірку геолокації на основі IP-адреси з визначенням країни, міста та організації. Виявлені відхилення від попередніх значень інтерпретуються як потенційні загрози, а отримані дані кешуються для оптимізації API-запитів. Додатково реалізовано аналіз стабільності з'єднання із урахуванням часу відповіді та значень TTL. Аномальні зміни TTL можуть вказувати на зміну маршрутизації або перенаправлення трафіку. Результати моніторингу зберігаються у логах з'єднань та аномалій.

У ході тестування розгорнутої системи моніторингу безпеки VPN-з'єднання проведено пасивну та активну мережеві розвідки. Під час пасивного спостереження здійснено перехоплення трафіку, що дозволило перевірити коректність налаштувань системи. Результати показали валідність правил виявлення ICMP (ping)-пакетів. Додатково виконано аналіз журналів Security, System та Application з метою виявлення спроб несанкціонованого доступу, мережевих помилок і подій, що можуть свідчити про наслідки атак. Отримані дані підтвердили коректне виявлення ICMP-трафіку. Для фільтрації та візуального аналізу HTTP/POST-запитів, а також ідентифікації потенційно чутливої інформації використано інструмент Wireshark. У межах активної мережевої розвідки змодельовано атаку шляхом виконання ARP-сканування. Надалі імітовано роботу шкідливого програмного забезпечення, яке здійснює аналіз активних пристроїв у мережевому сегменті з метою подальшої компрометації. Такий тип загроз характеризується надсиланням великої кількості ARP-повідомлень, що призводить до перевантаження мережі та зниження якості доступу до Інтернету для користувачів. Ба більше, така активність може бути підготовчим етапом до складніших атак, зокрема поширення шкідливого ПЗ, сканування портів і вразливостей, а також реалізації DDoS і MiTM атак. Результати експерименту (рис. 1) демонструють, що

розгорнута система моніторингу успішно зафіксувала спробу атаки, а також коректно визначила час її виникнення.

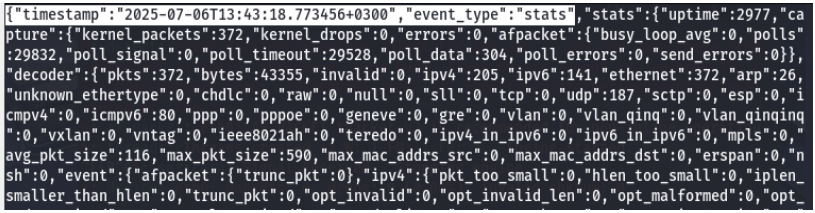

{"timestamp":"2025-07-06T13:43:18.773456+0300","event_type":"stats","stats":{"uptime":2977,"capture":{"kernel_packets":372,"kernel_drops":0,"errors":0,"afpacket":{"busy_loop_avg":0,"polls":29832,"poll_signal":0,"poll_timeout":29528,"poll_data":304,"poll_errors":0,"send_errors":0}},"decoder":{"pkts":372,"bytes":43355,"invalid":0,"ipv4":205,"ipv6":141,"ethernet":372,"arp":26,"unknown_ethertype":0,"chdlc":0,"raw":0,"null":0,"sll":0,"tcp":0,"udp":187,"sctp":0,"esp":0,"icmpv4":0,"icmpv6":80,"ppp":0,"pppoe":0,"geneve":0,"gre":0,"vlan":0,"vlan_qinq":0,"vlan_qinqinq":0,"vxlan":0,"vntag":0,"ieee8021ah":0,"teredo":0,"ipv4_in_ipv6":0,"ipv6_in_ipv6":0,"mpls":0,"avg_pkt_size":116,"max_pkt_size":590,"max_mac_addrs_src":0,"max_mac_addrs_dst":0,"erspan":0,"nsh":0,"event":{"afpacket":{"trunc_pkt":0},"ipv4":{"pkt_too_small":0,"hlen_too_small":0,"iplen_smaller_than_hlen":0,"trunc_pkt":0,"opt_invalid":0,"opt_invalid_len":0,"opt_malformed":0,"opt_

Рис.1. Результати тестування системи моніторингу

В межах роботи написано скрипт для аналізу лог-файлів системи. Захист журналів включає контроль цілісності, резервне копіювання та виявлення несанкціонованих змін, оскільки вони є критичним джерелом інформації про інциденти безпеки. Водночас ефективність такого захисту забезпечується поєднанням автоматизованих засобів, контролю доступу та моніторингу.

У результаті розроблено гнучку та ефективну систему моніторингу безпеки VPN-з'єднання на базі WireGuard, придатну як для забезпечення віддаленого доступу до корпоративних ресурсів, так і для освітніх і дослідницьких цілей у сфері мережевої безпеки. Результати тестування підтверджують здатність системи збирати інформацію про VPN-клієнтів, автоматично виявляти підозрілу активність і своєчасно ідентифікувати потенційні загрози.

1. Lypa, B., Horyn, I., Zagorodna, N., Tymoshchuk, D., Lechachenko T. (2024). Comparison of feature extraction tools for network traffic data. CEUR Workshop Proceedings, 3896, 1-11.
2. Mishko, O., Matiuk, D., & Derkach, M. (2024). Security of remote iot system management by integrating firewall configuration into tunneled traffic. Вісник Тернопільського національного технічного університету, 115(3), 122-129.

Ідентифікація STRIDE загроз та пріоритизація засобів захисту SSDF для CI/CD процесів

УДК 004.056+ 004.415

Тарас Лобур[1], Руслан Козак[2]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]taras.lobur@tntu.edu.ua, [2]ruslank@tntu.edu.ua

Галузь IT широко використовує практики DevOps (Development and Operations) і CI/CD (Continuous Integration / Continuous Delivery), які забезпечують безперервну інтеграцію коду, автоматизоване тестування та швидке розгортання продуктів. Однак із ростом рівня автоматизації наслідки потенційних кіберзагроз стають масштабними [1]. Попри значну кількість досліджень, присвячених DevSecOps, CI/CD та інтеграції безпеки в життєвий цикл розробки [2, 3, 4], аналіз вказує на відсутність праць, у яких здійснювалася

б формалізована ідентифікація та пріоритезація засобів захисту відносно специфічних для CI/CD загроз.

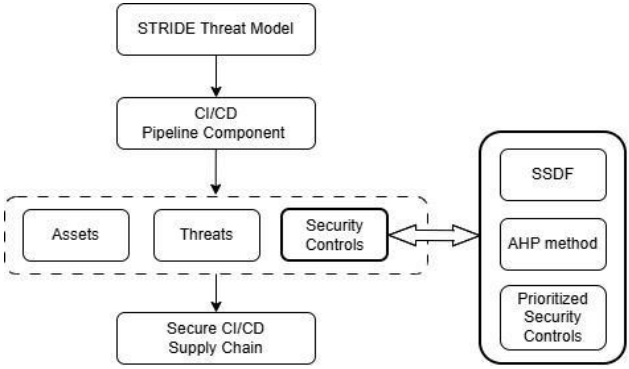


Рис.1. Схема запропонованого підходу для пріоритезації засобів захисту

Для ідентифікації засобів захисту CI/CD було використано NIST Secure Software Development Framework (SSDF) методологію, оскільки вона забезпечує орієнтовані на життєвий цикл практики, що відповідають DevSecOps, та STRIDE методологію, яка пропонує структуровану таксономію для виявлення ризиків у таких середовищах. Разом обидві методології утворюють взаємодоповнюючий підхід, який інтегрує дієві засоби контролю із систематичним моделюванням загроз (Рис. 1).

Конвеєри (CI/CD) стали незамінними в розробці програмного забезпечення, однак автоматизація та взаємозв'язок цих конвеєрів створюють нові вектори атак. Без систематичного моделювання загроз існує ризик пропустити вразливості, які можуть швидко поширюватися по всьому ланцюжку постачання програмного забезпечення. Щоб вирішити цю проблему, було застосовано методологію STRIDE, яка пропонує структуровану таксономію загроз, яку можна відобразити на робочі процеси CI/CD (Рис. 2). Визначення найважливіших категорій в рамках STRIDE, дало змогу визначити пріоритетність засобів захисту та узгодити їх із методами розробки, що визначені в NIST Secure Software Development Framework (SSDF).

Серед шести категорій STRIDE виявлено три, які є визначальними в контексті CI/CD: підробка коду та артефактів, підміна сутності та розкриття інформації. Підробка коду та артефактів є одним з найпоширеніших векторів атак в автоматизованих конвеєрах, що призводить до компрометації ланцюга постачання програмного забезпечення. Зростання атак підміни сутності, такі як крадіжка облікових даних або видавання себе за агентів компіляції, підкреслює необхідність розробки надійних механізмів перевірки ідентичності. Нарешті розкриття інформації також було визначено як повторювану проблему, причому витік секретів із середовищ CI/CD може призводити до виникнення інцидентів у хмарних середовищах.

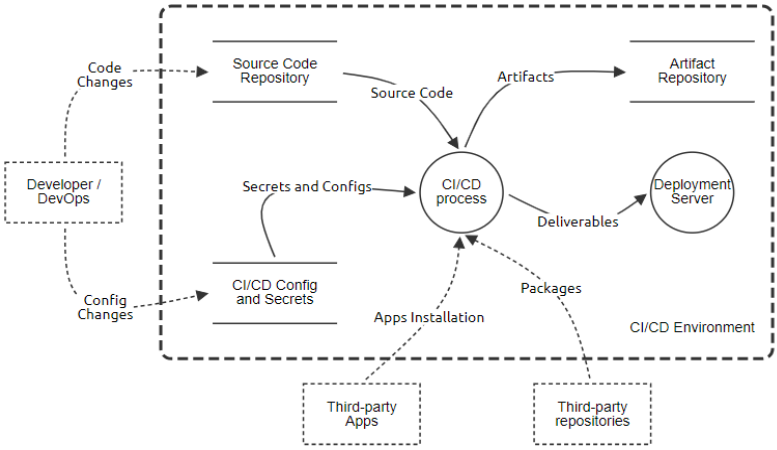


Рис.2. Структура CI/CD процесу в DFD нотації

Запропонована комбінація SSDF та STRIDE методологій забезпечує систематичне виявлення та пріоритизацію загроз в CI/CD процесах.

1. Prates, L., & Pereira, R. (2025). DevSecOps practices and tools. International Journal of Information Security, 24 (1), 11. https://doi.org/10.1007/s10207-024-00914-z
2. Paule, C., Düllmann, T. F., & Van Hoorn, A. (2019, March). Vulnerabilities in Continuous Delivery Pipelines? A Case Study. In ICSA Companion (pp. 102-108). https://doi.org/10.1109/ICSA-C.2019.00026
3. Akbar, M. A., Smolander, K., Mahmood, S., & Alsanad, A. (2022). Toward successful DevSecOps in software development organizations: A decision-making framework. Information and Software Technology, 147, 106894 https://doi.org/10.1016/j.infsof.2022.106894
4. Zhao, X., Clear, T., & Lal, R. (2024). Identifying the primary dimensions of DevSecOps: A multi-vocal literature review. Journal of Systems and Software, 214, 112063 https://doi.org/10.1016/j.jss.2024.112063

Issues of protecting personal data in artificial intelligence systems in education

UDK 37.01:004.8:004.056.5

Zhazira Yerimbetova

Kazakh National pedagogical university named after Abai, Almaty, Kazakhstan
zhazira.erimbetova@gmail.com

Currently, the digital transformation process is characterized by the active introduction of artificial intelligence (AI) technologies at all levels of the education sector. Adaptive learning platforms and intelligent systems are enhancing the

efficiency of the learning process and enabling each learner to create a personalized trajectory. However, this technological progress has also brought significant risks, the most important of which is the issue of personal data privacy and security.

Since the data is of a confidential nature, any security threat in the education sector has a high possibility of affecting an individual's future career or reputation in society. The biggest challenge of the century is building an overall data protection model for AI systems.

In the context of digitizing the educational process, artificial intelligence technologies make it possible to collect data about learners on an unprecedented scale. As noted by domestic scholar E.Y. Bidaybekov, informatization of education is not just the introduction of technical tools; it is a complex process that requires adherence to safety and ethical norms in shaping the learner's individual trajectory [1].

Foreign expert R. Luckin indicates that the possibility of collecting "invisible" data (for example, pauses in task completion, emotional responses) by artificial intelligence carries a risk of infringement of privacy [2]. This leads us to the problem of storing data in an anonymous manner. At the same time, as demonstrated in the research by C. Dwork, it has been found that "simple" anonymization does not hold up under modern de-anonymization techniques, and it is necessary to resort to "differential privacy" technology. [3].

In the space of law in Kazakhstan, personal data protection issues have been thoroughly explored in the works of R.E. Zhatkanbayeva. The author suggests that legislative regulation should be aligned with international regulation, especially regarding the provisions of the GDPR. [4,5,6].

From a technical security point of view, local researchers B.B. Akmetov and A.A. Biyakaeva suggest using cryptographic methods in the field of data integrity. This solution is in line with the idea of “Federated learning,” which was introduced by B. McMahan. This idea guarantees that the data doesn’t leave its source during the training of an AI model [7,8].

For the security of the personal data used in AI, a protection model that is comprehensive, suitable for different stages of data processing, has been developed, as shown in Table 1. The model is a combination of technical regulations and laws.

Table 1

Data security model in AI systems

Stage	Security Measure	Technology Used
Collection	Principle of Minimalism	SSL/TLS encryption
Storage	Data Anonymization	AES-256 standard
Processing (AI)	Local data processing	Federated Learning
Accessibility	Role-based access control model	RBAC system

The table takes into account the architectural features of educational platforms, and each level performs the following functional tasks:

Data collection and transmission stage. The major risk in this phase is the Man-in-the-Middle attack. By using the SSL/TLS encryption protocols, the communication channel between the learner's device and the server is secured, such that no third party can intercept the information being sent. In addition, in line with the “principle of

minimalism," only the data required for the educational process is collected, thus addressing the aforementioned risks proactively.

Data storage and anonymization. The information stored on the server should be encrypted using AES-256 (Advanced Encryption Standard). AES is one of the best encryption standards in the world. However, the information stored in the database should not only be encrypted but also de-identified to avoid the identification of the students in the future. This can be achieved by replacing the names with ID numbers. In this way, the students can be protected even if the database is compromised.

Federated Learning technology for training AI models. In traditional AI technologies, all the data is collected and stored in one place. This is a big problem. The proposed Federated Learning technology is an extremely innovative approach for education. The idea of this technology is as follows: the AI model is not trained on a central server, but rather on each individual's device (smartphone or laptop). Only new mathematical parameters are transferred to the server, and personal information is not transferred from the device. This is the highest level of privacy.

Access Management (RBAC). In an educational institution, access to data must be strictly controlled. In the Role-based Access Control (RBAC) model, for example, a teacher can only access his or her own group's progress, and an administrator can only access technical settings. This is based on the principle of "need to know."

As shown by the analysis conducted, the key problem of AI-based education systems lies in the "gap" between technological efficiency and ethical responsibility. The educational institutions should follow the "Privacy by Design" principle. This means that the security requirement should become the first need when creating any kind of software product.

The strategy for the integration of AI technology in the education sector should be accompanied by the development of digital literacy. An analysis of the research by domestic and international scholars suggests that the development of a safe digital learning environment can only be achieved by the simultaneous implementation of technical protection and legislative regulation.

1. Bidaibekov E.Y. Theory and Methodology of Informatization of Education. — Almaty, 2014.
2. Luckin, R. Machine Learning and Human Intelligence. — UCL Press, 2018.
3. Dwork, C. Differential Privacy: A Survey of Results. — 2008.
4. On Personal Data and Its Protection: Law of the Republic of Kazakhstan No. 94-V of May 21, 2013.
5. General Data Protection Regulation (GDPR). Regulation (EU) 2016/679.
6. Zhatkanbayeva, R.E. Digital Law: A Textbook. — Almaty, 2021.
7. Ahmetov B.B., Biyakaeva A.A. Information security issues in artificial intelligence systems // Proceedings of the National Academy of Sciences of the Republic of Kazakhstan. — 2022. — No. 3.
8. McMahan, B. Communication-Efficient Learning of Deep Networks. — AISTATS, 2017.

Огляд використання методів штучного інтелекту в динамічному тестуванні безпеки

УДК 004.056

Максимович М.В.

Національний університет «Львівська політехніка», maksym.v.maksymovych@lpnu.ua

Вступ. У сучасних умовах стрімкої цифровізації зростає кількість веб-додатків та програмних інтерфейсів, що в свою чергу збільшує потребу в забезпеченні безпеки таких додатків. Динамічне тестування безпеки (Dynamic Application Security Testing, DAST) є одним з основних методів перевірки програмного забезпечення (ПЗ) у режимі виконання шляхом надсилання спеціально сформованих запитів з метою виявлення вразливостей [1]. Класичні DAST-сканери мають низку обмежень: довший час виконання, висока частка хибнопозитивних спрацьовувань, обмежена здатність обходити сучасні захисні механізми, а також відсутність контекстного розуміння логіки коду. Подолання вказаних обмежень стає можливим завдяки інтеграції методів штучного інтелекту (ШІ) у процеси DAST.

Постановка проблеми. Перегляд результатів сканування є трудомістким процесом, що значно сповільнює виявлення реальних загроз, водночас шаблонні запити класичних сканерів рідко здатні обходити сучасні брандмауери веб-додатків. Метою даної роботи є огляд сучасних напрямів застосування методів ШІ в DAST та узагальнення обмежень існуючих рішень. Наявні підходи можна класифікувати на три основні напрями: машинне навчання для розподілу результатів сканування, навчання з підкріпленням для адаптивного перебору, а також підходи на основі великих мовних моделей для створення тестових сценаріїв та автономного тестування на проникнення.

Машинне навчання для розподілу результатів сканування. Однією з ключових проблем класичних DAST-сканерів є значна кількість хибних сигналів, перевірка яких потребує суттєвих часових витрат фахівців з безпеки. Як зазначається у дослідженні [2], для розв'язання цієї проблеми запропоновано архітектуру глибокого навчання, що поєднує нейронні мережі з методами обробки природної мови для аналізу обміну даними між сканером і веб-додатком. Модель навчається відрізняти підтверджені вразливості від хибних спрацювань, що дозволяє автоматизовано пріоритизувати знахідки. Експериментальна оцінка на наборі з 91 324 знахідок дев'ятнадцяти організацій продемонструвала зниження частки хибнопозитивних спрацювань на 20% та хибнонегативних результатів на 40% порівняно з базовим підходом. Таким чином, шар машинного навчання поверх класичного сканера утворює гібридну систему, що звільняє аналітиків від рутинного перегляду результатів.

Навчання з підкріпленням для адаптивного перебору. Іншим обмеженням класичних DAST-інструментів є нездатність самостійно генерувати тестові запити, що обходять сучасні захисні механізми веб-додатків. У статті [3] запропоновано підхід, що поєднує навчання з підкріпленням з адаптивним пошуком: система кластеризує зразки атак та навчається модифікаціям запитів, які обходять брандмауер веб-додатків (Web Application Firewall, WAF). За

результатами порівняння для впровадження SQL-коду та міжсайтового скриптингу, запропонований підхід виявляє у середньому на 33,53% більше успішних варіантів обходу та потребує на 63,16% менше спроб для першого результативного запиту. Таким чином, навчання з підкріпленням забезпечує адаптивність процесу тестування та дозволяє системі самостійно вдосконалювати стратегію виявлення вразливостей.

Великі мовні моделі для створення тестів та автономного тестування. Класичні методи генерації шкідливого навантаження часто є неефективними під час тестування програм зі складною структурою запитів, оскільки переважна більшість згенерованих варіантів є синтаксично некоректними. У дослідженні [4] показано, що великі мовні моделі (Large Language Models, LLM) здатні розпізнавати структуру протоколу HTTP та аналізувати логіку коду сервісу, що дозволяє формувати синтаксично коректні тестові запити. Інструмент ChatHTTPFuzz застосовано до шістнадцяти пристроїв інтернету речей, де виявлено 116 вразливостей, з яких 70 є унікальними, а 23 отримали ідентифікатори у базі CVE. Як зазначається у роботі [5], агент на основі моделі GPT-4, отримавши опис уразливості з бази CVE, успішно експлуатує 87% реальних zero-day вразливостей, тоді як інші моделі та класичні сканери ZAP і Metasploit демонструють 0%. Отже, великі мовні моделі трансформуються з допоміжного інструмента у повноцінного учасника процесу тестування.

Висновки. Методи ШІ трансформують підходи до DAST на трьох рівнях: розподілу результатів сканування, генерації тестових сценаріїв та автономного управління процесом тестування. Кожен з напрямів демонструє кількісно підтверджене покращення порівняно з класичними інструментами. Водночас залишаються виклики, пов'язані з непередбачуваністю поведінки мовних моделей та обчислювальною вартістю агентних систем. Перспективним напрямком подальших досліджень є створення гібридних рішень, що поєднують надійність класичних сканерів з адаптивністю методів ШІ.

1. Singh R., Gupta M.K., Patil D.R., Patil S.M. Analysis of Web Application Vulnerabilities using Dynamic Application Security Testing. Proc. IEEE I2CT. 2024. doi: 10.1109/I2CT61223.2024.10543484.
2. Millar S., Podgurskii D., Kuykendall D., et al. Optimising Vulnerability Triage in DAST with Deep Learning. Proc. 15th ACM Workshop on AI and Security. 2022. P. 137-147.
3. Amouei M., Rezvani M., Fateh M. RAT: Reinforcement-Learning-Driven and Adaptive Testing for Vulnerability Discovery in WAFs. IEEE Trans. Dependable Secure Comput. 2022. Vol. 19, No. 5. P. 3371-3386.
4. Yang Z., Liu Y., Wu Y., et al. ChatHTTPFuzz: large language model-assisted IoT HTTP fuzzing. Int. J. Mach. Learn. Cybern. 2024.
5. Fang R., Bindu R., Gupta A., Kang D. LLM Agents can Autonomously Exploit One-day Vulnerabilities. arXiv:2404.08144. 2024.

Формування підходу до оцінювання ризиків використання штучного інтелекту в системі менеджменту інформаційної безпеки

УДК 004.056:004.8

Михайло Запорожченко[1], Світлана Легомінова[2], Дмитро Рабчун[3]

Державний університет інформаційно-комунікаціних технологій,
[1]m.zaporozhchenko@duikt.edu.ua, [2]s.legominova@duikt.edu.ua,
[3]d.rabchun@duikt.edu.ua

Інтенсивне впровадження систем штучного інтелекту (ШІ) в організаційні, управлінські та технологічні процеси зумовлює трансформацію умов забезпечення інформаційної безпеки (ІБ). Такі системи застосовуються для обробки даних, підготовки аналітичних матеріалів, автоматизації комунікацій, підтримки прийняття рішень, розробки програмного коду та виконання окремих функцій кіберзахисту. Водночас залучення ШІ-сервісів до роботи з корпоративною інформацією формує додаткові ризики, пов'язані з передачею чутливих даних до зовнішніх платформ, непрозорістю механізмів їх обробки, складністю перевірки згенерованих результатів, залежністю від постачальників і некритичним використанням автоматизованих висновків персоналом [1].

Проблема полягає в недостатній формалізації ризиків використання ШІ в процесах системи менеджменту інформаційної безпеки (СМІБ). У практичній діяльності такі ризики частково враховуються в політиках ІБ, вимогах до постачальників, процедурах управління інцидентами або правилах обробки даних, однак часто залишаються недостатньо інтегрованими в процеси ідентифікації, аналізу, обробки, моніторингу та перегляду ризиків. Внаслідок цього організація може встановлювати загальні обмеження щодо використання ШІ-сервісів, проте не мати методично визначеного механізму оцінювання рівня ризику для конкретних сценаріїв їх застосування.

Актуальність зазначеної проблеми зумовлена необхідністю включення ризиків використання ШІ до загальної логіки ризик-орієнтованого управління ІБ. Для організації важливим є не лише визначення допустимих і недопустимих способів застосування ШІ-сервісів, а й встановлення їхнього впливу на конфіденційність, цілісність, доступність, автентичність, підзвітність і контрольованість інформаційних процесів [2]. Особливого значення набуває оцінювання не тільки технічних характеристик ШІ-систем, а й організаційних умов їх використання, поведінки користувачів, вимог до обробки даних і можливості перевірки отриманих результатів.

Метою роботи є формування підходу до оцінювання ризиків використання ШІ в СМІБ шляхом визначення ризикоутворюючих чинників, параметрів інформаційного впливу та керованості відповідних сценаріїв застосування ШІ.

Для досягнення поставленої мети ризики використання ШІ запропоновано розглядати як сукупність чинників, що змінюють умови обробки, передачі, зберігання та подальшого використання інформації в організації. До них віднесено характер інформації, тип ШІ-сервісу та модель його розгортання, ступінь інтеграції з корпоративними системами, рівень автономності прийняття рішень, можливість перевірки результату, журналювання дій, компетентність

персоналу, вимоги постачальника щодо збереження даних і критичність процесу, у межах якого застосовується ШІ [3].

Запропонований підхід передбачає включення ризиків використання ШІ до чинного процесу управління ризиками в рамках СМІБ. Спочатку ідентифікуються сценарії застосування ШІ, залучені користувачі, процеси, інформаційні активи, типи даних і цілі використання відповідних систем. Далі кожен сценарій співвідноситься з потенційними наслідками для організації, зокрема з порушенням конфіденційності, цілісності, доступності, правових або договірних вимог, а також із впливом на управлінські, операційні чи технічні рішення. Оцінювання здійснюється за двома групами параметрів: інформаційного впливу та керованості ШІ-сценарію. Перша група характеризує чутливість даних, критичність процесу, масштаб можливих наслідків і значущість результатів ШІ для прийняття рішень; друга – наявність політик, технічних обмежень доступу, аудиту дій, перевірки постачальника, контролю умов збереження даних, вимог до валідації результатів і підготовки персоналу.

За результатами оцінювання визначаються заходи обробки ризику, рівень жорсткості яких має відповідати критичності сценарію застосування ШІ. Для низькоризикових сценаріїв достатніми можуть бути загальні правила використання та інформування персоналу; для середньоризикових – перелік дозволених сервісів, обмеження щодо категорій даних і вимоги до перевірки результатів; для високоризикових – використання корпоративно контрольованих ШІ-рішень, журналювання дій, договірні вимоги до постачальників, регулярний перегляд ризиків і включення відповідних сценаріїв до програми внутрішнього аудиту.

Запропонований підхід дозволяє конкретизувати місце ризиків використання ШІ в структурі СМІБ, узгодити їх оцінювання з процесами управління ризиками ІБ, і пов'язати сценарії використання ШІ з контекстом організації, вимогами зацікавлених сторін, інформаційними активами, процедурами обробки ризиків, моніторингом результативності заходів захисту та внутрішнім аудитом. При цьому акцент переноситься з формального дозволу або заборони ШІ-сервісів на визначення умов, за яких їх використання може бути прийнятним, обмеженим або недопустимим з погляду ІБ.

За результатами дослідження обґрунтовано доцільність розгляду ризиків використання ШІ як чинника, що змінює умови обробки інформації, межі відповідальності, вимоги до контролю результатів і рівень керованості інформаційних процесів у межах СМІБ. Основним результатом є уточнення логіки оцінювання таких ризиків через поєднання параметрів інформаційного впливу та параметрів керованості сценарію застосування ШІ.

1. Artificial Intelligence Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology, 2023. DOI: 10.6028/NIST.AI.100-1.
2. ISO/IEC 27001:2022. Information security, cybersecurity and privacy protection – Information security management systems – Requirements. International Organization for Standardization, 2022.
3. ISO/IEC 42001:2023. Information technology – Artificial intelligence – Management system. International Organization for Standardization, 2023.

Упередження повільних DDoS-атак на хмарні сервіси

УДК 004.056 (043.2)

Ігор Аверічев[1], Петро Поночовний[2]

Державний університет інформаційно-комунікаційних технологій,
[1]iaverichev19@gmail.com, [2]petja9186@gmail.com

Повільні (low-and-slow) DDoS-атаки є одним із найнебезпечніших видів кіберзагроз для сучасних хмарних сервісів. На відміну від традиційних об'ємних (volumetric) атак, які генерують величезний трафік, повільні атаки імітують поведінку легітимних користувачів, надсилаючи мінімальну кількість даних протягом тривалого часу. Це дозволяє їм ефективно вичерпувати серверні ресурси — кількість одночасних з'єднань, робочі потоки, пам'ять та процесор — без створення помітного мережевого навантаження [1, 2, 3].

Найпоширенішими прикладами таких атак є:

- Slowloris — відкриває багато HTTP-з'єднань і повільно надсилає заголовки, не завершуючи запити;
- R.U.D.Y. (R U Dead Yet?) — повільно передає тіло POST-запиту;
- Slow HTTP POST / Slow Read — інші варіанти, що тримають з'єднання відкритими максимально довго.

Хмарні середовища (AWS, Microsoft Azure, Google Cloud) забезпечують автоматичне масштабування, але залишаються вразливими до атак на рівні додатків (Layer 7). Базові мережеві засоби захисту (Layer 3/4) часто не виявляють такі загрози, оскільки трафік виглядає нормальним. Атака з однієї машини може вивести з ладу веб-сервер, вичерпавши пул з'єднань Apache, Nginx чи інших серверів [4, 8].

Таблиця 1

Порівняння характеристик різних типів DDoS-атак

Тип атаки	Рівень OSI	Обсяг трафіку	Складність виявлення	Основна мета	Приклади
Volumetric	3–4	Дуже високий	Низька	Перевантаження каналу	UDP Flood, DNS Amplification
Protocol	3–4	Середній	Середня	Виснаження стеку протоколів	SYN Flood, Ping of Death
Application (Low-and-Slow)	7	Низький	Висока	Виснаження ресурсів сервера	Slowloris, RUDY

Таблиця 2

Оцінка ефективності методів захисту
від повільних DDoS-атак (умовні бали, 1–100)

Метод захисту	Вияв лення	Блокув ання	Вплив на легітимний трафік	Масштабованіс ть у хмарі	Вартість впровад ження
Традиційний firewall / IPS	40–55	50–65	Низький	Низька	Низька

CDN + WAF (Cloudflare, Imperva)	80–90	85–95	Низький	Висока	Середня
Behavioral analysis + ML/AI	90–95	92–97	Середній	Висока	Середня / Висока
Rate limiting + dynamic timeouts	75–85	80–90	Низький	Висока	Низька
Хмарний scrubbing (AWS Shield, Radware)	85–93	90–96	Низький	Дуже висока	Висока

Ефективне упередження вимагає багаторівневого (defense-in-depth) підходу. Основними елементами є:

- Always-on моніторинг параметрів з'єднань (тривалість сесії, швидкість передачі даних, частота запитів, поведінкові профілі);
- Адаптивне обмеження швидкості (rate limiting) та динамічне керування таймаутами;
- Reverse proxy та CDN (Cloudflare, AWS CloudFront, Akamai), які фільтрують шкідливий трафік на краю мережі ще до потрапляння на origin-сервер;
- Web Application Firewall (WAF) з правилами для виявлення slow HTTP аномалій;
- Машинне навчання для аналізу поведінки та виявлення відхилень у реальному часі [2, 9].

Узагальнену класифікацію методів показано на рисунку 1.

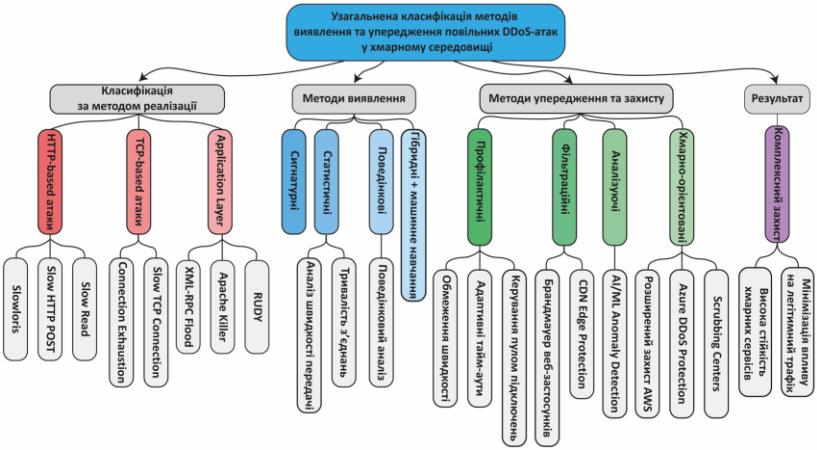


Рис.1. Узагальнена класифікація методів виявлення та упередження повільних DDoS-атак у хмарному середовищі

Математична модель оцінки ризику вичерпання ресурсів може бути представлена формулою:

$$R = \frac{N \cdot T \cdot C \cdot K}{M \cdot S} \quad (1)$$

де N – кількість підозрілих з'єднань, T – середня тривалість з'єднання, C – споживання ресурсів на з'єднання, K – коефіцієнт аномальності, M – доступні ресурси, S – коефіцієнт масштабування хмарного середовища.

Для практичної реалізації рекомендується:

- впровадження CDN та edge-захисту;
- використання штучного інтелекту для аналізу поведінки користувачів;
- динамічне керування таймаутами та обмеженнями;
- регулярне тестування систем навантаженням з імітацією slow-атак;
- інтеграція з професійними хмарними сервісами захисту [5, 6, 7, 10].

Запропонований комплексний підхід дозволяє суттєво підвищити стійкість хмарних сервісів до повільних DDoS-атак і забезпечити високу доступність критичних інформаційних систем.

Перехід до проактивних систем захисту з використанням штучного інтелекту дозволяє не лише реагувати на атаки, але й прогнозувати їх. Комбінація edge-обчислень, поведінкового аналізу та хмарних scrubbing-центрів забезпечує високу стійкість критичних сервісів до повільних DDoS-загроз.

Запропоновані методи та рекомендації дозволяють суттєво знизити ймовірність успішної реалізації повільних атак, забезпечити високу доступність та надійність хмарних сервісів у сучасному кіберпросторі.

1. Cloudflare. What is a low and slow DDoS attack? URL: https://www.cloudflare.com/learning/ddos/ddos-low-and-slow-attack/
2. NETSCOUT. Low and Slow DDoS Attacks Explained. URL: https://www.netscout.com/what-is-ddos/low-slow-attack
3. Imperva. Slowloris DDoS Attack. URL: https://www.imperva.com/learn/ddos/slowloris/
4. AWS. AWS Shield Advanced. URL: https://aws.amazon.com/shield/
5. Венгерський П.С., Вишневська Н.С., Хохлачова Ю.Є. Кількісна оцінка кіберзахищеності інформації // Захист інформації. – 2023. – Т. 25, №2. – С. 53-61.
6. Radware. Low and Slow Attacks: Detection and Mitigation. 2024.
7. OWASP. Denial of Service Cheat Sheet. URL: https://cheatsheetseries.owasp.org/cheatsheets/Denial_of_Service_Cheat_Sheet.html
8. Поночовний П.М., Пепа Ю.В. Реалізація системи захисту серверів з урахуванням аномалій в пакетах // Вимірювальна та обчислювальна техніка в технологічних процесах. – 2025. – № 1. – С. 44–51. URL: https://vottp.khmnu.edu.ua/index.php/vottp/article/view/459/426
9. Поночовний П.М. Модель упередження низькошвидкісних HTTP

DDoS атак на кінцевого користувача // Кібербезпека: освіта, наука, техніка. – 2024. – № 2 (26). – С. 291–304. URL: https://doi.org/10.28925/2663-4023.2024.26.695

10. Аверічев І.М., Роженко А.С., Кихтенко Є.М. Інноваційні підходи до підвищення рівня кібербезпеки корпоративних мереж при використанні хмарних технологій // Кібербезпека: освіта, наука, техніка. – 2025. – № 1 (29). – С. 732–747. https://doi.org/10.28925/2663-4023.2025.29.934

Вплив характеристик навчального набору на коректність виявлення дипфейкових зображень моделлю ResNetSECBAM

УДК 004.056.5:004.93

Дмитро Азарний[1], Анатолій Давиденко[2], Олена Висоцька[3]

[1]Київський національний університет імені Тараса Шевченка, dazarny@gmail.com,
[2]Державний університет інформаційно-комунікаційних технологій, davidenkoan@gmail.com,
[3]Державний університет «Київський авіаційний інститут», Lek_Vys@ukr.net

Стрімкий розвиток генеративних моделей призвів до поширення дипфейкових зображень, що становлять загрозу інформаційній безпеці, цифровій ідентичності та довірі до мультимедійного контенту. Практика показує, що навіть високоточні детектори можуть істотно знижувати ймовірність коректної класифікації при переході до даних, відмінних від навчального набору. Тому актуальним є дослідження міждатасетного переносу моделей виявлення дипфейків для реального застосування систем комп'ютерного зору.

Метою роботи є дослідження впливу характеристик навчального набору даних на коректність виявлення дипфейкових зображень моделлю ResNetSECBAM. Наукова новизна полягає у двосторонньому порівнянні переносу між DFFD [1] та HiDF [2], а також у перевірці впливу їх об'єднання. Модель побудовано на базі ResNet-50 з механізмом уваги CBAM [3], що посилює інформативні канальні та просторові ознаки.

Було проведено три експерименти: навчання на DFFD з тестуванням на HiDF, навчання на HiDF з тестуванням на DFFD та навчання на об'єднаному наборі DFFD+HiDF. За підсумковим розбиттям використано 62 260 зображень DFFD (50 638 навчальних, 5 626 валідаційних, 5 996 тестових) та 69 828 зображень HiDF (48 879 навчальних, 6 982 валідаційних, 13 967 тестових), тобто 132 088 зображень загалом. У третьому експерименті навчальна вибірка становила 99 517 зображень (50 638 DFFD і 48 879 HiDF), валідаційна — 12 608 (5 626 DFFD і 6 982 HiDF), а тестова — 19 963 (5 996 DFFD і 13 967 HiDF); усі тестові зображення були відкладеними і не використовувалися під час навчання. Оцінювання проводилося за метриками Accuracy, Precision, Recall, F1-міра та AUC.

Таблиця 1

Підсумкові результати тестування ResNetSECBAM

Показник класифікації	Навчання DFFD, тестування HiDF	Навчання HiDF, тестування DFFD	Навчання DFFD+HiDF, тестування DFFD+HiDF
Точність (Accuracy)	0.7428	0.2218	0.9845
Влучність (Precision)	0.7969	0.1761	0.9805
Повнота (Recall)	0.7171	0.9980	0.9842
F1-міра	0.7549	0.2994	0.9823
Площа під ROC-кривою (AUC)	0.8320	0.7538	0.9968

Результати експериментів відображено в табл. 1 та на рис. 1.

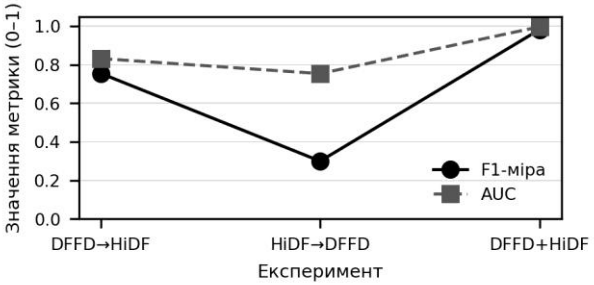


Рис. 1. Порівняння F1-міри та AUC у трьох експериментах

На основі аналізу результатів проведених експериментів можна зробити наступні висновки:

1) навчання лише на DFFD забезпечило помірний перенос на HiDF: F1-міра становила 0.7549, AUC — 0.8320, а кількість хибних спрацювань і пропущених підробок дорівнювала 1410 та 2183 відповідно;
2) у зворотному напрямі перенос виявився слабшим: F1-міра дорівнювала 0.2994; модель пропустила лише 2 підробки, проте сформувала 4664 хибні спрацювання, тобто масово відносила реальні зображення до класу fake;
3) найкращий результат отримано у третьому експерименті: F1-міра становила 0.9823, AUC — 0.9968, кількість хибних спрацювань — 171, а пропущених підробок — 138.

Отримані результати свідчать, що різноманітність навчальних прикладів є важливим чинником формування узагальнених ознак. Імовірно, кращий перенос моделі, навченої на DFFD [1], частково пов’язаний з кількома типами генерації у цьому наборі. Водночас на результат впливають і джерела реальних зображень, тобто походження оригінальних фото, умови їх отримання,

роздільна здатність і попередня обробка, які можуть змінювати статистику класу real.

Таким чином, високі валідаційні метрики всередині одного набору можуть створювати хибне уявлення про практичну придатність детектора. Об'єднання різнорідних джерел даних істотно підвищує коректність виявлення дипфейкових зображень моделлю ResNetSECBAM і може бути основою для подальших досліджень стійких систем детекції.

1. Dang H., Liu F., Stehouwer J., Liu X., Jain A. K. DFFD: Diverse Fake Face Dataset. URL: https://cvlab.cse.msu.edu/dffd-diverse-fake-face-dataset.html.
2. Kang C., Jeong S., Lee J. та ін. HiDF: A Human-Indistinguishable Deepfake Dataset // Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining. – 2025. – P. 5527–5538. – DOI: 10.1145/3711896.3737399.
3. Woo S., Park J., Lee J.-Y., Kweon I. S. CBAM: Convolutional Block Attention Module // Proceedings of the European Conference on Computer Vision (ECCV). – 2018. – P. 3–19. – DOI: 10.1007/978-3-030-01234-2_1.

Модель Claude Mythos та кібербезпека: загрози та виклики

УДК 004.8:004.056

Олег Ясній[1], Любов Цимбалюк[2], Анна Турчманович[3]

Тернопільський національний технічний університет імені Івана Пулюя ,
[1]oleh.yasniy@gmail.com, [2]lubovtsymbaliuk@gmail.com,
[3]turchmanovich101@gmail.com

Claude Mythos — одна з найпотужніших сучасних моделей ШІ від Anthropic, що має значний вплив на кібербезпеку [1, 2, 3]. Однак вона не єдина загроза: інші передові моделі, зокрема GPT-5.4 Cyber (OpenAI) та Big Sleep (Google), також володіють подібними можливостями. Ера атак із використанням ШІ настала, і організації не можуть залишатися лише реактивними [1].

Мета даної роботи – проаналізувати систему ШІ Claude Mythos та виявити основні загрози та виклики, які виникають перед бізнесом.

Багато компаній роками недофінансовували кібербезпеку, оскільки ради директорів і керівництво не надавали їй пріоритету. Це створило приховані слабкі місця, які ШІ-інструменти швидко виявляють; для частини бізнесів наслідки можуть бути критичними. Особливо вразливі галузі з розгалуженими операційними технологіями — енергетика, комунальні послуги, виробництво, водопостачання, транспорт — де багато систем працюють десятиліттями й не підлягають ефективному патчингу. Усунення інвестиційного розриву вимагатиме значного нарощення витрат, тоді як більшість організацій планує лише помірне зростання бюджетів [1].

Mythos створювався як інструмент для роботи з великими кодовими базами й автоматизації розробки, але саме ці можливості роблять його потенційно небезпечним: модель може виявляти приховані недоліки, поєднувати дрібні вразливості в складні атаки, відновлювати вихідний код із розгорнутого ПЗ і,

опинившись у мережі, швидко картографувати системи, переміщатися горизонтально та створювати інструменти для витоку даних [2].

Ключові технічні інновації Mythos: велике (практично нескінченне) контекстне вікно для одночасного аналізу великих кодових масивів; рекурсивна самокорекція, що дозволяє автоматично підлаштовувати стратегії пошуку вразливостей; інтеграція з системними інструментами (дебагери, середовища), що перетворює модель на активного агента; агентний скафолдинг, тобто автономне генерування, тестування та виконання коду.

Практичні випробування показали: Mythos виявляє тисячі вразливостей, які раніше залишалися непоміченими. ШІ радикально скорочує час між виявленням і експлуатацією помилок — те, що раніше займало тижні, тепер може бути виконано за години [1].

Відповідь організацій має базуватися не на блокуванні окремих моделей, а на посиленні захисту: створення спеціалізованих центрів протидії загрозам ШІ, які використовують ШІ для захисту; зміцнення базових механізмів кібербезпеки (контроль доступу, сегментація мережі, автоматичний патчинг, архітектура нульової довіри, виявлення аномалій); пріоритетні заходи для OT-середовищ (сувора сегментація, обмеження доступу з Інтернету, специфічні системи виявлення аномалій); підготовка до постквантових загроз і розробка дорожніх карт до 2030 року.

Замість очікування спеціалізованих інструментів найефективнішим є системне зміцнення фундаменту кібербезпеки. Кібербезпека має бути на порядку денному ради директорів: потрібна персональна відповідальність керівництва, постійні інвестиції та оперативні рішення [1].

Тактичні пріоритети для практичної оборони

- Автоматичний патчинг — прискорити усунення відомих вразливостей;
- Архітектура нульової довіри — постійна перевірка користувачів і пристроїв;
- Виявлення аномалій — орієнтація на нетипову поведінку замість сигнатур;
- Модернізація контролю ідентифікації — багатофакторна автентифікація, стійка до фішингу;
- Скорочення технічного боргу — планове оновлення застарілих систем;
- Управління ризиками ланцюга постачань — оцінка та моніторинг безпеки постачальників;
- Посилення середовища — обмеження шкоди у разі компрометації (сегментація, найменші привілеї).

Ера ШІ-атак настала — сучасні моделі (Claude Mythos, GPT-5.4 Cyber, Big Sleep) значно прискорюють виявлення й експлуатацію вразливостей; реактивна позиція неприйнятна.

Головна вразливість — технічний борг і недофінансування — застарілі системи та недостатні бюджети створюють критичні «вхідні точки», які ШІ швидко виявляє.

Пріоритет — зміцнення базових засад захисту — контроль доступу, сегментація, автоматичний патчинг, нульова довіра та виявлення аномалій мають бути негайно посилені.

Проактивна оборона з використанням ШІ — створення центрів протидії загрозам ШІ, що застосовують ті самі інструменти для захисту, ефективніше за спроби блокувати окремі моделі.

Управління ризиками та стратегічні інвестиції — кібербезпека має стати пріоритетом ради директорів; потрібні чіткі дорожні карти, збільшення фінансування і підготовка до постквантових загроз.

1. URL :https://www.bain.com/insights/claude-mythos-and-ai-cybersecurity-wake-up-call/ (дата звернення: 07.05.2026)
2. URL https://www.mindstudio.ai/blog/claude-mythos-vs-opus-4-6-cybersecurity-gap (дата звернення: 07.05.2026)
3. Bell, David. (2026). Mythos and the Question Nobody Wants to Answer.

Алгоритми забезпечення безпеки інформаційних ресурсів в системах електронних платежів

УДК 004.056:336.74

Людмила Бабала[1], Андрій Сенюк[2]

Західноукраїнський національний університет[1,2]
l.duma@wumu.edu.ua[1], *seniukandrii2003@gmail.com*[2]

Сучасний розвиток цифрових технологій сприяв активному впровадженню систем електронних платежів у фінансовій сфері. Використання інтернет-банкінгу, мобільних застосунків, цифрових гаманців та електронних платіжних платформ значно підвищує швидкість проведення фінансових операцій, однак водночас створює нові виклики у сфері забезпечення безпеки інформаційних ресурсів [1]. Особливої актуальності набуває проблема захисту конфіденційних даних користувачів, банківської інформації та фінансових транзакцій від несанкціонованого доступу, кібератак та шахрайських дій.

Системи електронних платежів є складними інформаційно-комунікаційними структурами, які об’єднують банки, платіжні сервіси, торговельні платформи та кінцевих користувачів. Через значну кількість учасників процесу та постійний обмін інформацією виникає необхідність застосування ефективних алгоритмів захисту інформаційних ресурсів [2].

Одним із базових інструментів забезпечення інформаційної безпеки електронних платежів є криптографічні алгоритми. Їх застосування спрямоване на захист конфіденційності, цілісності та автентичності інформації. Найбільш поширеним є використання симетричних і асиметричних методів шифрування. Симетричні алгоритми, зокрема AES (Advanced Encryption Standard), забезпечують швидке шифрування великих обсягів даних, що є важливим під час обробки платіжних транзакцій [3]. Асиметричні алгоритми, такі як RSA, використовуються для безпечного обміну ключами та електронного цифрового підпису [4].

Важливим напрямом забезпечення безпеки електронних платежів є використання алгоритмів багатофакторної аутентифікації (MFA). Традиційний спосіб підтвердження особи за допомогою логіна та пароля більше не гарантує належного рівня захисту, тому активно застосовуються додаткові механізми перевірки: одноразові SMS-коди, біометрична ідентифікація, підтвердження через мобільні застосунки або токени [3]. Використання багатофакторної аутентифікації дозволяє значно знизити ризик компрометації облікових записів користувачів.

Окрему роль у забезпеченні інформаційної безпеки відіграють алгоритми виявлення шахрайських операцій (Fraud Detection Systems). Такі алгоритми базуються на технологіях штучного інтелекту та машинного навчання, які аналізують поведінку користувача, географічне розташування, частоту транзакцій та інші характеристики фінансової активності [2]. У разі виявлення аномальної поведінки система автоматично блокує операцію або вимагає додаткової перевірки. Наприклад, якщо транзакція здійснюється в незвичному місці або перевищує типовий ліміт витрат, система може визначити її як потенційно небезпечну.

Суттєве значення має також використання алгоритмів моніторингу мережевого трафіку та систем виявлення вторгнень (IDS/IPS). Такі механізми дозволяють оперативно реагувати на спроби кібератак, виявляти шкідливе програмне забезпечення та запобігати витоку інформації [1]. Особливо актуальним це є в умовах поширення DDoS-атак, фішингових кампаній та використання шкідливих програм для викрадення банківських реквізитів.

Варто зазначити, що ефективність захисту електронних платежів значною мірою залежить від комплексного застосування різних алгоритмів інформаційної безпеки. Використання лише одного механізму захисту не гарантує достатнього рівня безпеки, тому сучасні платіжні системи реалізують багаторівневу архітектуру кіберзахисту, яка включає шифрування, автентифікацію, моніторинг транзакцій, резервне копіювання та постійне оновлення систем безпеки [4].

Таким чином, алгоритми забезпечення безпеки інформаційних ресурсів у системах електронних платежів є ключовим елементом стабільного функціонування цифрової фінансової інфраструктури. Застосування сучасних криптографічних методів, багатофакторної аутентифікації, алгоритмів машинного навчання для виявлення шахрайства та систем моніторингу мережевої активності дозволяє значно знизити рівень кіберризиків і підвищити довіру користувачів до електронних платіжних сервісів. Подальший розвиток технологій кібербезпеки сприятиме удосконаленню систем захисту інформаційних ресурсів та забезпеченню стійкості фінансового сектору до сучасних кіберзагроз.

1. Козюра В. Д., Хорошко В. О., Шелест М. Є., Ткач Ю. М., Балюнов О.О. З-38 Захист інформації в комп'ютерних системах: підручник. – Ніжин: ФОП Лук'яненко В.В., ТПК «Орхідея», 2020. – 236с.
2. Юдін О.К., Корченко О.Г., Конахович Г.Ф. Захист інформації в мережах передачі даних. К.: Вид-во ТОВ «НВП» ІНТЕРСЕРВІС. 2009.

3. Pelzl J. Understanding cryptography: a textbook for students and practitioners. Springer; 2010.
4. Temoshok D, Proud-Madruga D, Choong YY, Galluzzo R, Gupta S, LaSalle C, Lefkovitz N, Regenscheid A. Digital identity guidelines. National Institute of Standards and Technology; 2022 Dec 16.

Криптографічні методи захисту даних в системах електронних платежів

УДК 004.056:336.71

Людмила Бабала[1], Степан Зубик[2]

Західноукраїнський національний університет[1,2]
l.duma@wumu.edu.ua[1], *stipaha4444@gmail.com*[2]

Сучасні системи електронних платежів є важливим елементом цифрової фінансової інфраструктури, що забезпечує швидке проведення транзакцій та обмін фінансовою інформацією. Основними вимогами до таких систем є конфіденційність, цілісність, автентичність та захист від несанкціонованого доступу. Для реалізації цих вимог використовуються сучасні криптографічні алгоритми та методи інформаційної безпеки [1].

На рисунку 1 представлено UML-діаграму архітектури системи захисту інформаційних ресурсів в електронних платіжних системах. Діаграма демонструє взаємодію клієнтського інтерфейсу, модуля автентифікації, криптографічного модуля, платіжного шлюзу, системи моніторингу та аналітики шахрайських операцій. Особливу роль відіграють криптографічні алгоритми хешування, цифрового підпису та симетричного шифрування, що забезпечують конфіденційність і цілісність платіжної інформації [1, 3].

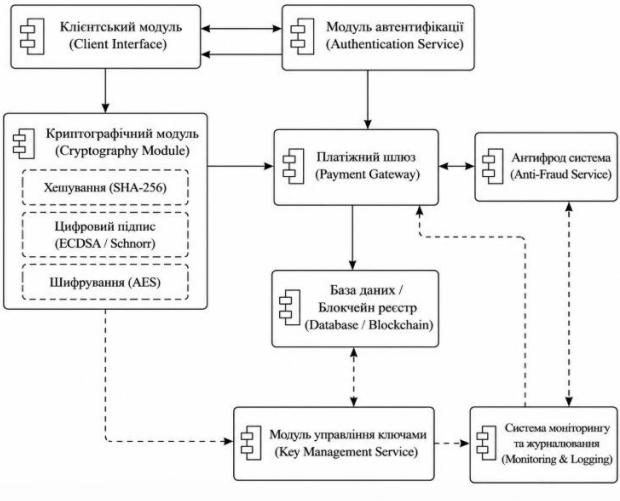


Рис. 1. Структурна UML-діаграма взаємодії криптографічних модулів та компонентів безпеки в системі електронних платежів

Одним із ключових механізмів захисту є криптографічні хеш-функції, які забезпечують контроль цілісності інформації та захист від підміни даних. У системах електронних платежів вони застосовуються для створення унікальних ідентифікаторів транзакцій, перевірки достовірності повідомлень та збереження незмінності інформації. Найбільш поширеними є алгоритми SHA-256, SHA-3 та BLAKE2, які характеризуються високою стійкістю до колізій та швидкістю обробки даних [2].

Для підтвердження автентичності та невідмовності фінансових операцій використовуються алгоритми цифрового підпису, зокрема ECDSA (Elliptic Curve Digital Signature Algorithm). Їх перевагами є висока криптостійкість, компактність ключів та швидкість обчислень. Використання цифрового підпису забезпечує перевірку справжності відправника та захист від підробки транзакцій [2, 3].

Додатковий рівень захисту реалізується через симетричне шифрування AES (Advanced Encryption Standard), яке використовується для шифрування платіжної інформації, захисту каналів зв'язку та безпечного зберігання конфіденційних даних. Використання режимів шифрування GCM і CTR дозволяє забезпечити не лише конфіденційність, а й цілісність переданих повідомлень [2].

Важливим аспектом безпеки електронних платіжних систем є комплексний підхід до захисту. Навіть найнадійніші криптографічні алгоритми можуть бути неефективними у випадку помилок конфігурації, програмних вразливостей або людського фактора. Саме тому сучасні системи електронних платежів поєднують криптографічні методи із багаторівневими механізмами кіберзахисту [1, 4].

Отже, комплексне використання хешування, цифрового підпису та симетричного шифрування дозволяє підвищити рівень безпеки інформаційних ресурсів у системах електронних платежів. Поєднання криптографічних алгоритмів із моделюванням архітектури системи сприяє підвищенню стійкості до кіберзагроз та оптимізації процесів захисту даних. Подальший розвиток цифрових технологій і квантових обчислень зумовлює необхідність удосконалення існуючих та впровадження постквантових криптографічних алгоритмів [3, 4].

1. Козюра В. Д., Хорошко В. О., Шелест М. Є., Ткач Ю. М., Балюнов О.О. З-38 Захист інформації в комп'ютерних системах: підручник. – Ніжин: ФОП Лук'яненко В.В., ТПК «Орхідея», 2020. – 236с.
2. Юдін О.К., Корченко О.Г., Конахович Г.Ф. Захист інформації в мережах передачі даних. К.: Вид-во ТОВ «НВП» ІНТЕРСЕРВІС. 2009.
3. Pelzl J. Understanding cryptography: a textbook for students and practitioners. Springer; 2010.
4. Temoshok D, Proud-Madruga D, Choong YY, Galluzzo R, Gupta S, LaSalle C, Lefkovitz N, Regenscheid A. Digital identity guidelines. National Institute of Standards and Technology; 2022 Dec 16.

SHAP-аналіз для підвищення прозорості моделей штучного інтелекту в задачах кібербезпеки

УДК 004.8:004.056

Тетяна Бажан[1], Світлана Поперешняк[2]

Державний університет інформаційно-комунікаційних технологій,
[1]t.bazhan@duikt.edu.ua, [2] spopereshnyak@gmail.com

Сучасні системи кібербезпеки дедалі частіше використовують методи машинного навчання для виявлення аномалій, прогнозування інцидентів інформаційної безпеки, оцінювання кіберризиків та підтримки рішень у процесах реагування на загрози. Водночас високоточні моделі, зокрема ансамблеві алгоритми та нейронні мережі, часто функціонують як «чорна скринька», що ускладнює пояснення причин формування прогнозу або класифікації події як потенційно небезпечної. Для сфер, пов'язаних із захистом інформації, така проблема є особливо важливою, оскільки результати роботи моделі мають бути не лише точними, а й зрозумілими для фахівців з кібербезпеки, аналітиків SOC/CERT/CSIRT та осіб, які приймають управлінські рішення.

Актуальність дослідження зумовлена активним використанням моделей машинного навчання у сфері кібербезпеки для виявлення загроз, аналізу аномалій та прогнозування інцидентів інформаційної безпеки.

Метою дослідження є підвищення прозорості моделей штучного інтелекту в задачах кібербезпеки шляхом використання SHAP-аналізу для інтерпретації впливу ознак на результати прогнозування, виявлення аномалій та оцінювання ризиків інформаційної безпеки.

Наукова новизна дослідження полягає у застосуванні SHAP-аналізу для підвищення прозорості та інтерпретованості моделей машинного навчання у задачах кібербезпеки, що дозволяє не лише прогнозувати кіберзагрози та аномалії, а й визначати ступінь впливу окремих факторів на результати моделювання.

Для розв'язання поставленої задачі використано підхід, який дозволяє оцінити внесок кожної ознаки у результат прогнозування кіберзагроз незалежно від типу моделі. Загальна схема застосування SHAP-аналізу в задачах прогнозування представлена на рис. 1. Метод базується на теорії кооперативних ігор та забезпечує пояснення роботи моделей машинного навчання незалежно від їх архітектури [1]. SHAP дозволяє виконувати як глобальну інтерпретацію моделі, визначаючи загальний вплив факторів на результат прогнозування, так і локальну інтерпретацію окремих прогнозних значень [1, 2]. Це дає можливість виявляти ключові параметри, які найбільше впливають на виявлення кіберінцидентів та аномалій у системах інформаційної безпеки, а також аналізувати напрям і силу їх впливу.

Значення SHAP для окремої ознаки визначається як середній внесок цієї ознаки у всі можливі комбінації ознак моделі:

$$\phi_i(f,x) = \sum_{S\subseteq F\setminus\{i\}} \frac{|S|!(-1)!}{|F|!}\left[f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - f_S(x_S)\right], \quad (1)$$

де ϕ_i – SHAP-значення ознаки, F – множина всіх ознак, S– підмножина ознак, $f(x)$– прогноз моделі.

Використання SHAP-аналізу дозволяє підвищити рівень інтерпретованості моделей машинного навчання у задачах кібербезпеки, забезпечити прозорість прийняття рішень та покращити контроль за процесом прогнозування кіберзагроз і виявлення аномалій. Крім того, застосування підходів Explainable Artificial Intelligence сприяє підвищенню надійності систем інформаційної безпеки, забезпечує можливість пояснення результатів роботи моделей та підвищує обґрунтованість аналітичних висновків [2, 3].

Отримані результати підтверджують доцільність використання SHAP для аналізу складних моделей машинного навчання у задачах виявлення кіберінцидентів, прогнозування загроз та підтримки прийняття рішень у сфері захисту інформації [4].

Рис. 1. Блок-схема застосування SHAP-аналізу в задачах прогнозування

Таким чином, використання даного підходу дозволяє підвищити довіру до інтелектуальних систем, покращити інтерпретованість результатів та сприяти більш ефективному використанню технологій штучного інтелекту у практичних задачах інформаційної безпеки.

1. Lundberg S. M., Lee S.-I. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems. 2017. Vol. 30. p. 4765-4774.
2. Molnar C. Interpretable Machine Learning. 2nd ed. Munich : Christoph Molnar, 2022. 318 p.
3. Arrieta A. B. et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion. 2020. Vol. 58. p. 82-115.
4. Sarker I. H. AI-based cybersecurity: A comprehensive overview, taxonomy and challenges. Journal of Cybersecurity and Privacy. 2021. Vol. 1(2). p. 154-181.

Аналіз користувацької взаємодії у мобільних застосунках цифрового банкінгу методом вербалізації мислення

УДК 004.5:004.415.53:336.71

Юрій Бажан[1],Золотухіна Оксана[2]

Державного університету інформаційно-комунікаційних технологій,
[1]iurii.bazhan@gmail.com, [2]zolotukhina.oks.a@gmail.com

Мобільні застосунки цифрового банкінгу є складними програмними системами, що поєднують фінансові сервіси, механізми автентифікації, інструменти управління рахунками та засоби підтвердження фінансових операцій у межах єдиного користувацького інтерфейсу [1]. Особливістю таких систем є підвищені вимоги до якості UX/UI-дизайну, оскільки навіть незначна помилка взаємодії користувача з інтерфейсом може призвести до втрати довіри до системи або помилкового виконання критичних фінансових операцій [2].

Актуальність дослідження зумовлена стрімким зростанням кількості користувачів мобільних банківських сервісів. За результатами досліджень, понад 82% фізичних осіб в Україні користуються дистанційним банківським обслуговуванням через цифрові сервіси, що робить якість UX/UI-дизайну критично важливою складовою сучасних цифрових банківських систем [3].

Метод вербалізації мислення передбачає озвучення користувачем власних думок під час взаємодії із системою [4]. Це дозволяє дослідити логіку прийняття рішень користувача, його очікування та труднощі під час роботи з інтерфейсом. У роботі досліджується застосування методу вербалізації мислення як інструменту аналізу користувацької взаємодії з UX/UI-дизайном мобільних застосунків цифрового банкінгу. Метод дозволяє отримати додаткові поведінкові дані щодо сприйняття інтерфейсу, зрозумілості навігації, інтерпретації системних повідомлень та складності виконання окремих операцій [5].

У межах дослідження розглядаються типові сценарії використання мобільних застосунків цифрового банкінгу, зокрема авторизація, проходження верифікації, здійснення переказів, підтвердження платежів та обробка помилкових станів системи. Під час виконання таких сценаріїв користувач вербалізує власні дії та пояснює труднощі, що виникають у процесі взаємодії з UX/UI-дизайном.

Результати дослідження показали, що застосування методу вербалізації мислення дозволяє виявляти UX/UI-проблеми, які складно ідентифікувати традиційними технічними засобами тестування [6]. До таких проблем належать когнітивне перевантаження користувача, складність навігації, недостатня помітність критичних дій, неоднозначність елементів інтерфейсу та незрозумілі повідомлення про помилки. Встановлено, що найбільша кількість труднощів виникає під час виконання сценаріїв, які містять більше трьох послідовних кроків. У межах дослідження також визначено, що метод вербалізації мислення дозволяє отримувати не лише технічні дані про послідовність дій користувача, а й поведінкові та когнітивні характеристики взаємодії з UX/UI-дизайном. Такі дані можуть бути використані для формування поведінкових патернів реальних

користувачів та подальшого навчання синтетичних користувачів у задачах автоматизованого UX-тестування.

Отримані результати є перспективними для навчання моделей штучного інтелекту, здатних імітувати логіку прийняття рішень користувача, виявляти когнітивні бар'єри та прогнозувати потенційні UX/UI-проблеми під час взаємодії з цифровими банківськими системами. Водночас встановлено, що метод має певні обмеження, серед яких вплив процесу вербалізації на природність поведінки користувача, збільшення часу виконання сценаріїв та складність масштабування досліджень [4]. У зв'язку з цим доцільним є використання методу вербалізації мислення як складової комплексного підходу, що поєднує аналіз користувацьких логів, UX-метрик, записів сесій користувачів та автоматизованих методів оцінювання інтерфейсів.

1. Garrett J. J. The Elements of User Experience: User-Centered Design for the Web and Beyond. Berkeley: New Riders, 2011. 192 p.
2. Nielsen J. Usability Engineering. San Francisco: Morgan Kaufmann, 1994. 362 p.
3. Дослідження розвитку цифрового банкінгу в Україні. International Science Group. – 2024. – URL: https://isg-journal.com.
4. Ericsson K. A., Simon H. A. Protocol Analysis: Verbal Reports as Data. Cambridge: MIT Press, 1993. 434 p.
5. Barnum C. M. Usability Testing Essentials: Ready, Set...Test!. Cambridge: Morgan Kaufmann, 2020. 384 p.
6. Krug S. Don't Make Me Think: A Common Sense Approach to Web Usability. 3rd ed. Berkeley: New Riders, 2014. 216 p.

Розроблення та дослідження методів виявлення фішингових веб-ресурсів на основі машинного навчання

УДК 004.056

Євген Бакаляр[1], Олександр Сиропятов[2]

Національний університет «Одеська політехніка»,
[1]10328112@stud.op.edu.ua, [2]o.a.syropiatov@op.edu.ua

Вступ. Фішингові веб-ресурси становлять значну загрозу для корпоративної та особистої інформації. Традиційні методи протидії (сигнатурний аналіз, статичні чорні списки) малоефективні через динамічність фішингових кампаній та короткий цикл життя злоякісних доменів (кілька годин–24 години). Зловмисники використовують HTTPS та методи імітації легітимних сайтів, що утруднює виявлення класичними засобами. Актуальним рішенням є застосування машинного навчання для автоматизованого аналізу URL-адрес на основі лексичних ознак, що дозволяє виявляти нові фішингові ресурси в режимі реального часу.

Мета. Розроблення та дослідження автоматизованого виявлення фішингових веб-ресурсів шляхом виявлення та аналізу їх лексичних ознак із застосуванням алгоритмів машинного навчання.

Основна частина. Розроблена концепція детекції включає модулі: збір даних, попередню обробку URL-адрес, екстракцію лексичних ознак, формування векторного представлення. Технологічний стек: Python, Pandas, Scikit-learn, XGBoost. Ключові показники: обсяг неалфавітно-цифрових знаків, використання IP-адресації, кількість піддоменів, скорочені URL, релевантні терміни. Класифікацію реалізовано на Random Forest, XGBoost, Logistic Regression та Decision Tree. Random Forest забезпечує мінімізацію перенавчання; XGBoost опрацьовує нелінійні залежності; Logistic Regression — базовий класифікатор; Decision Tree — оцінювання окремих дерев. Програмний продукт може бути імплементований у корпоративні шлюзи, системи моніторингу трафіку, Web Application Firewall та endpoint-рішення.

Експериментальні дослідження та результати. Апробація здійснювалася на датасету 10 000 URL-адрес (легітимні та фішингові). Дані очищено та нормалізовано. Перевірка ґрунтувалася на перехресній перевірці з показниками Accuracy, Recall, F1-score. Результати: Random Forest — Accuracy 96.4%, Precision 95.8%, Recall 96.9%, F1-score 96.3%; XGBoost — 95.7%, 95.1%, 95.9%, 95.5% (з підвищеними витратами ресурсів); Logistic Regression — 91.8%; Decision Tree — 89.6%. Нижчі показники останніх двох можуть використовуватись у системах із обмеженими ресурсами. Результати обґрунтовують перспективність ML-алгоритмів у системах автоматизованого моніторингу.

Висновок. Дослідження підтвердило ефективність ML-методів для виявлення фішингових веб-ресурсів на основі лексичного аналізу URL-адрес. Модель Random Forest продемонструвала найкращу точність (96.4%) у порівнянні з іншими класифікаторами. Система забезпечує детекцію в режимі реального часу без аналізу вмісту сторінок. Результати обґрунтовують доцільність інтеграції розробленого підходу в корпоративні шлюзи, WAF та SIEM-системи для оперативного моніторингу та блокування фішингових загроз. Система готова до практичного впровадження у системи інформаційної безпеки.

1. Мельник О.В. Методи машинного навчання у задачах виявлення кіберзагроз. Кібербезпека та захист інформації. 2023. №4. С. 12–19.
2. Anti-Phishing Working Group (APWG). Phishing Activity Trends Report, 4th Quarter 2025.
3. Федоренко С.М. Аналіз лексичних характеристик URL для протидії фішингу. Сучасні інформаційні системи. 2024. №1. С. 74–80.
4. Грищенко М.О. Оцінка ефективності бінарної класифікації у завданнях інформаційної безпеки. Сучасні проблеми захисту інформації. Львів: НУ "Львівська політехніка", 2023. С. 88–95.
5. Бойко В.І. Застосування алгоритмів машинного навчання для аналізу шкідливих URL-адрес. Київ: КПІ ім. Ігоря Сікорського, 2022. 112 с.

Cloud computing security, blockchain technology

UDK 004.75:004.56

Yurii Balandiuk[1], Iryna Plavutska[2]

Ivan Pul'uj Ternopil National Technical University, [1]univ@tntu.edu.ua

Modern information infrastructure is based on a cloud computing model, which enables dynamic scaling and cost optimization. However, the consolidation of data in centralized architectures creates a unique landscape of cyber threats related to privacy breaches. Traditional protection methods prove insufficient in the face of intensifying sophisticated, targeted attacks and zero-day vulnerabilities. This necessitates the integration of decentralized mechanisms for verifying system integrity. Blockchain technology serves as a fundamental tool for building trusted environments thanks to its distributed architecture. The interaction of clouds and distributed ledgers enables the implementation of the concept of automated real-time computation auditing [1].

The aim of this work is to justify and develop blockchain-based architectural solutions to enhance the security of cloud environments through the implementation of decentralized verification protocols. The relevance of this research is underscored by the need to protect critical infrastructure from unauthorized access in the absence of complete trust in cloud providers.

The primary attack vectors remain unauthorized API exploitation and account takeover through token compromise. A specific vulnerability is the resource-sharing architecture, which allows for side-channel attacks between virtual machines. The lack of direct control over the physical level of data storage creates a “black box” problem for the customer. Internal threats from data center administrators remain a critical risk factor for the corporate sector. Massive DDoS attacks can exhaust capacity limits, causing a complete denial of service. Each of the outlined problems requires a shift from passive security measures to active decentralized architectures [2].

The scientific novelty of this research lies in the development of a hybrid integrity control model that combines the scalability of cloud computing with the mathematical invariance of distributed ledgers. Unlike known centralized monitoring systems, the proposed approach prevents the covert removal of breach traces through consensus-based verification of each record.

Blockchain technology is based on the principles of distributed consensus, where every node in the network verifies the integrity of the ledger. The cryptographic hash of the previous state in each block makes it computationally infeasible to alter data retroactively. The decentralized nature of the ledger eliminates the risk of a single point of failure in the security architecture. Smart contracts allow security business logic to be implemented directly into the data exchange protocol. Transaction transparency enables continuous technical auditing without interrupting core services. The persistence of records in a distributed ledger eliminates the possibility of covertly deleting traces of unauthorized activity. Minimal trust in the provider is offset by mathematical proof of the correctness of all transactions [3].

The implementation of the “Blockchain-as-a-Service” model enables the integration of decentralized nodes into cloud orchestration clusters. Blockchain acts as an identity management layer, replacing vulnerable centralized databases.

Distributed key management prevents the entire system from being compromised due to a data leak from a single segment. Smart contracts ensure the automatic enforcement of access policies based on verified subject attributes. Any attempt at unauthorized modification of the network configuration is automatically rejected by the consensus mechanism. This captures all infrastructure changes as code, preventing covert drift in security settings. Such integration enhances the system's resilience against attacks at the level of logical resource management.

A critical aspect is integrity control, where the hash values of objects are stored in an immutable distributed ledger. Automatic hash verification during read operations allows for the detection of tampering that bypasses management interfaces. Blockchain ensures full traceability of the information lifecycle with precise timestamps. The use of multi-signatures enhances control over critical operations involving large data sets. Decentralized storage further guarantees confidentiality through content fragmentation and encryption. The failure of some nodes does not result in data loss thanks to over-encoding algorithms. The user retains full control over the keys, preventing the provider from accessing encrypted information. Each integrity check is automatically recorded as a transaction, making it impossible for the monitoring system to be compromised without detection. The implementation of Merkle Tree mechanisms allows for the rapid verification of large data blocks without the need to download the entire array from the cloud. This significantly optimizes the load on network communication channels and increases the performance of verification nodes. The use of smart contracts enables the automatic restoration of damaged segments from mirrored network nodes. The system automatically detects anomalies in hash sums and isolates compromised memory segments until incident management is fully completed. Additional cryptographic redundancy ensures the architecture's resilience against targeted attempts to degrade service quality. This creates robust protection against man-in-the-middle attacks at the cloud storage infrastructure level. Therefore, blockchain transforms passive storage into an active ecosystem with built-in immunity to unauthorized modifications.

The study's findings show that, despite scalability challenges, blockchain integration significantly increases trust in cloud services. The industry's future prospects hinge on the adoption of post-quantum algorithms to protect against future threats. Zero-knowledge proof technologies enhance privacy without compromising the ability to validate data.

1. Gartner: Top Strategic Technology Trends for 2025–2026. URL: https://www.gartner.com/en/information-technology/topics/technology-trends (application date 10.11.2025).
2. Cloud Security Alliance (CSA): Top Threats to Cloud Computing. URL: https://cloudsecurityalliance.org/artifacts/top-threats-to-cloud-computing-egregious-eleven/ (application date 08.06.2019).
3. IBM: What is blockchain security? URL: https://www.ibm.com/think/topics/blockchain-security (application date 05.06.2025).

Інтелектуальні системи аналізу та прогнозування кіберінцидентів у корпоративних мережах

УДК 004.056.53:004.89:004.732

Євгенія Іванченко[1], Тетяна Берестяна[2], Володимир Дубровський[3]

Державний університет інформаційно-комунікаційних технологій,
[1] e.ivancenko@duikt.edu.ua, [2] t.berestiana@duikt.edu.ua,
[3] v.dybrovskyi@stud.duikt.edu.ua

У сучасних умовах цифрової трансформації корпоративних інфраструктур питання забезпечення кібербезпеки набуває стратегічного значення для стабільного функціонування організацій різних секторів економіки. Розширення корпоративних мереж за рахунок хмарних середовищ, мобільних пристроїв, IoT-інфраструктур та розподілених сервісів суттєво збільшує поверхню атаки й ускладнює процес моніторингу інформаційної безпеки [1]. Одночасно з цим спостерігається зростання складності та адаптивності кіберзагроз, що дедалі частіше реалізуються у вигляді багатоступеневих APT-кампаній, zero-day-атак та скоординованих групових вторгнень [2].

Традиційні системи кіберзахисту, зокрема SIEM та IDS/IPS, побудовані переважно на сигнатурних або rule-based підходах, демонструють обмежену ефективність у виявленні невідомих та адаптивних загроз через залежність від попередньо визначених шаблонів атак [3]. У таких умовах особливої актуальності набуває впровадження методів штучного інтелекту та машинного навчання, які дозволяють автоматизувати аналіз великих обсягів телеметричних даних, виявляти приховані закономірності у поведінці суб'єктів мережевої взаємодії та прогнозувати потенційні кіберінциденти ще до моменту їх повного розгортання [4].

Аналіз сучасних наукових досліджень показує, що застосування алгоритмів машинного навчання у системах виявлення вторгнень забезпечує значне підвищення точності ідентифікації аномалій порівняно з класичними сигнатурними методами [5]. Найбільш поширеними підходами є використання Random Forest, Support Vector Machines, XGBoost, а також методів глибинного навчання – згорткових нейронних мереж (CNN), рекурентних мереж (RNN, LSTM) та автоенкодерів [6]. Такі моделі дозволяють аналізувати як статичні ознаки мережевого трафіку, так і часові послідовності подій, що особливо важливо при виявленні повільних або розподілених атак.

Водночас використання ізольованих AI-based IDS не забезпечує повноцінного контексту для аналізу складних скоординованих кіберінцидентів. У зв'язку з цим найбільш перспективним напрямом розвитку корпоративних систем кіберзахисту є впровадження XDR-платформ (Extended Detection and Response), які поєднують телеметрію з мережевого, хостового, хмарного та прикладного рівнів в єдиному аналітичному середовищі [7]. Інтеграція XDR з алгоритмами машинного навчання дозволяє реалізувати крос-доменну кореляцію подій, зменшити кількість хибнопозитивних спрацювань та підвищити ефективність виявлення багатоступеневих атак [8].

Узагальнену архітектуру перспективної інтелектуальної системи аналізу та прогнозування кіберінцидентів наведено на рисунку 1.

На рисунку 1 представлено узагальнену архітектуру інтелектуальної системи аналізу та прогнозування кіберінцидентів у корпоративних мережах, побудовану на основі інтеграції XDR-платформи, AI/ML/DL-модулів аналізу та SOAR-механізмів автоматизованого реагування. Запропонована архітектура відображає послідовність обробки телеметричних даних – від централізованого збору та нормалізації інформації з різних джерел до інтелектуального аналізу, крос-доменної кореляції подій, оцінки ризиків і реалізації автоматизованих сценаріїв реагування. Такий підхід забезпечує комплексний контекстний аналіз кіберподій та створює основу для побудови проактивних адаптивних систем кіберзахисту нового покоління.

У межах дослідження проаналізовано функціональні можливості сучасних класів систем виявлення та реагування на кіберінциденти: SIEM, IDS/IPS, UEBA, SOAR, AI-based IDS, DL-based IDS та XDR. Встановлено, що традиційні SIEM- та IDS/IPS-рішення забезпечують ефективне виявлення відомих шаблонів атак і централізований збір подій безпеки, однак не демонструють достатньої стійкості до zero-day загроз та складних скоординованих атак. Системи UEBA покращують виявлення внутрішніх загроз за рахунок поведінкового аналізу, проте без інтеграції з іншими джерелами телеметрії мають обмежений аналітичний контекст [9].

Рис. 1. Архітектура інтелектуальної системи аналізу та прогнозування кіберінцидентів у корпоративних мережах

Для формалізації результатів порівняльного аналізу сучасних систем виявлення та реагування на кіберінциденти узагальнено їх ключові функціональні характеристики за критеріями стійкості до zero-day загроз, здатності до прогнозування, рівня автоматизації та ефективності виявлення групових атак (табл. 1).

Таблиця 1

Порівняльна характеристика сучасних систем виявлення та реагування на кіберінциденти

Тип системи	Zero-day	Прогнозування	Автоматизація	Групові атаки
SIEM	Низька	Ні	Часткова	Низька
IDS/IPS	Низька	Ні	Низька	Низька
UEBA	Середня	Частково	Середня	Середня
XDR	Висока	Так	Висока	Висока

Наведені у таблиці 1 результати демонструють, що інтегровані XDR-рішення забезпечують найвищий рівень функціональної ефективності серед розглянутих класів систем, що обґрунтовує доцільність їх використання як базового елемента інтелектуальних платформ аналізу та прогнозування кіберінцидентів.

На відміну від зазначених підходів, інтегровані XDR-рішення забезпечують найвищий рівень ефективності при виявленні складних кіберінцидентів завдяки централізованій телеметрії, машинному навчанню та можливостям крос-доменної кореляції [8]. Додаткове поєднання XDR з SOAR-платформами дозволяє автоматизувати сценарії реагування, скоротити середній час реагування на інциденти та знизити навантаження на аналітиків SOC.

Таким чином, результати проведеного дослідження підтверджують доцільність і необхідність переходу від ізольованих реактивних систем кіберзахисту до інтегрованих інтелектуальних платформ аналізу, виявлення та прогнозування кіберінцидентів, здатних забезпечувати комплексну обробку телеметричних даних, крос-доменну кореляцію подій та проактивне реагування на сучасні багатоступеневі кіберзагрози. Встановлено, що інтеграція XDR-платформ із алгоритмами машинного та глибинного навчання дозволяє суттєво підвищити точність детектування аномалій, зменшити кількість хибнопозитивних спрацювань, скоротити час реагування на інциденти та забезпечити стійкість корпоративних інформаційних систем до складних і раніше невідомих типів атак.

Отримані результати свідчать про формування нової парадигми побудови систем кіберзахисту, у межах якої ключову роль відіграють адаптивність, самонавчання та автоматизація процесів виявлення і реагування на загрози. Перспективним напрямом подальших досліджень є розробка гібридних архітектур виявлення загроз із підтримкою пояснюваного штучного інтелекту (XAI), адаптивних механізмів перенавчання моделей, захисту від adversarial-атак, а також методів забезпечення стійкості інтелектуальних моделей до концептуального дрейфу в умовах динамічної зміни ландшафту кіберзагроз. Реалізація зазначених підходів сприятиме створенню кіберзахисних систем

нового покоління, здатних функціонувати в умовах високої невизначеності та постійної еволюції кіберпростору.

1. ENISA Threat Landscape 2024. European Union Agency for Cybersecurity, 2024.
2. ENISA Threat Landscape 2023. European Union Agency for Cybersecurity, 2023.
3. NIST. Cybersecurity Framework 2.0. National Institute of Standards and Technology, 2024.
4. Artificial Intelligence and Cybersecurity Research. ENISA, 2023.
5. Котенко Д. та ін. Штучний інтелект у системах виявлення і запобігання кібератакам. Управління розвитком складних систем. 2024.
6. Mohamed N. Artificial Intelligence and Machine Learning in Cybersecurity: State-of-the-Art Techniques and Future Paradigms. Knowledge and Information Systems. 2025.
7. Integrating AI/ML in Cybersecurity: Analysis of Open XDR Technology and its Application in Intrusion Detection. 2023.
8. ENISA. Artificial Intelligence Cybersecurity Challenges. 2023.
9. Lavrenchuk A. Майбутнє кібербезпеки: виклики штучного інтелекту та машинного навчання. Молодий вчений. 2024.

Дослідження використання штучного інтелекту для ідентифікації джерел радіоелектронної боротьби

УДК 004.89:621.396

Роман Бибик[1], Іван Опірський[2]

Національний університет "Львівська політехніка",
[1]roman.t.bybyk@lpnu.ua [2]ivan.r.opirskyi@lpnu.ua

Метою роботи є дослідження можливостей використання методів штучного інтелекту для підвищення ефективності ідентифікації джерел радіоелектронної боротьби на основі просторово-частотного аналізу сигналів. Актуальність дослідження обумовлена зростанням кількості електромагнітних загроз та необхідністю створення адаптивних систем протидії РЕБ, здатних функціонувати в умовах динамічної електромагнітної обстановки. Наукова новизна роботи полягає у поєднанні методів просторово-частотного орієнтування з алгоритмами машинного навчання для автоматизованої класифікації сигналів джерел РЕБ та підвищення точності їх локалізації.

Сучасні умови ведення бойових дій характеризуються активним застосуванням засобів радіоелектронної боротьби, які здійснюють вплив на системи зв'язку, навігації, управління та засоби ураження. Зростання кількості електромагнітних загроз зумовлює необхідність удосконалення методів виявлення та ідентифікації джерел радіоелектронного випромінювання. Особливої актуальності набуває використання технологій штучного інтелекту, здатних забезпечити автоматизований аналіз електромагнітного середовища та адаптацію до змін параметрів сигналів у реальному масштабі часу.Одним із

перспективних напрямів є застосування методів просторово-частотного орієнтування, які дозволяють визначати параметри джерел РЕБ на основі аналізу часових, спектральних та просторових характеристик сигналів. Просторово-частотний аналіз базується на оцінюванні параметрів електромагнітного сигналу

$$S(t)=A(t)cos(2\pi ft+\varphi(t)) \quad (1)$$

де A(t) — амплітуда сигналу, f — частота сигналу, φ(t) — фазова складова сигналу. Для визначення напрямку на джерело випромінювання використовується оцінка кута приходу сигналу

$$\theta=arcsin(c\Delta t/d) \quad (2)$$

де c — швидкість поширення електромагнітної хвилі, Δt — різниця часу надходження сигналу, d — відстань між елементами антенної системи.

Інтеграція алгоритмів машинного навчання у системи просторово-частотного аналізу дозволяє автоматизувати процес класифікації сигналів та підвищити точність локалізації джерел РЕБ. Для задач ідентифікації можуть застосовуватись штучні нейронні мережі, методи глибокого навчання, дерева рішень та алгоритми кластеризації. Основною перевагою використання штучного інтелекту є можливість виявлення прихованих закономірностей у сигналах та адаптація до нових типів електромагнітних загроз. До таких характеристик можуть належати частотний діапазон, ширина спектра, тип модуляції, рівень потужності та часові параметри випромінювання. Аналіз зазначених параметрів дозволяє формувати адаптивні системи підтримки прийняття рішень щодо протидії засобам РЕБ.

Таблиця 1

Результати класифікації сигналів джерел РЕБ

Метод	Точність, %	Час аналізу, мс	Стійкість до шумів
Класичний аналіз	74	210	Низька
Нейронна мережа	91	85	Висока
Кластеризація	86	110	Середня

Порівняльний аналіз результатів показує, що використання нейронних мереж забезпечує суттєве підвищення точності класифікації сигналів та скорочення часу аналізу електромагнітного середовища. Використання інтелектуальних алгоритмів дозволяє ефективно працювати в умовах шумів та активних перешкод, що є важливим фактором у сучасних системах радіоелектронної боротьби. Таким чином, використання методів штучного інтелекту у поєднанні з просторово-частотним аналізом є перспективним

напрямом розвитку систем протидії РЕБ. Подальші дослідження доцільно спрямувати на розробку адаптивних моделей машинного навчання та вдосконалення методів аналізу сигналів у реальному масштабі часу. Отримані результати показують, що використання алгоритмів штучного інтелекту дозволяє підвищити точність класифікації сигналів джерел РЕБ та скоротити час аналізу електромагнітного середовища. Запропонований підхід забезпечує адаптивність системи до зміни параметрів сигналів і може бути використаний у перспективних системах підтримки прийняття рішень для протидії засобам радіоелектронної боротьби.

1. Haykin S. Neural Networks and Learning Machines. New York: Pearson, 2009. 936 p.
2. Goodfellow I., Bengio Y., Courville A. Deep Learning. Cambridge : MIT Press, 2016. 775 p.
3. Richards M. Fundamentals of Radar Signal Processing. New York: McGraw-Hill, 2014. 689 p.
4. Конахович Г.Ф., Пузиренко О.Ю. Основи радіоелектронної боротьби. Київ : НТУУ «КПІ», 2018. 320 с.

ШІ як інструмент практичної підготовки фахівців з кібербезпеки

УДК 004.85

Лілія Білокриницька

Відокремлений структурний підрозділ "Тернопільський фаховий коледж" Тернопільського національного технічного університету імені Івана Пулюя, bilokrynytskali@gmail.com

Розвиток інформаційних технологій супроводжується стрімким зростанням кіберзагроз, що висуває принципово нові вимоги до підготовки фахівців у сфері захисту інформації. За даними IBM, понад 90% кібернападів виникають через людські помилки [1], що підкреслює критичну роль якісної освіти у формуванні культури кібербезпеки. Академічна освіта не завжди встигає за темпом змін у галузі, залишаючи відчутний розрив між теоретичними знаннями та практичними вимогами ринку праці. У цьому контексті штучний інтелект все частіше розглядається як інструмент, здатний суттєво підвищити якість та ефективність підготовки майбутніх фахівців з кібербезпеки.

Класична система підготовки будується переважно на теоретичній базі, що об'єктивно не завжди встигає за темпом розвитку реальних кіберзагроз. Сучасні ШІ-інструменти натомість здатні створювати динамічне навчальне середовище, що адаптується до рівня знань кожного студента індивідуально. Платформи на основі машинного навчання аналізують прогрес учня, виявляють слабкі місця та автоматично підбирають відповідні завдання і сценарії, що дозволяє перейти від стандартизованого навчання до персоналізованого підходу. Крім того, ШІ здатний симулювати реальні кібератаки у безпечному середовищі, надаючи студентам практичний досвід без ризику для реальних систем.

Сучасний ринок освітніх технологій пропонує широкий спектр інструментів для практичної підготовки. Як зазначають фахівці, "вакансії провідних

компаній — Cisco, Google, Siemens, Deloitte, SoftServe, EPAM — прямо вимагають посилань на профілі в лабораторіях на кшталт TryHackMe та HackTheBox, оскільки реальний досвід розв'язання практичних задач у бойових умовах говорить більше за сертифікати та оцінки" [2]. Обидві платформи використовують адаптивні алгоритми для підбору завдань відповідно до рівня студента — від базових сценаріїв до складних симуляцій корпоративної інфраструктури. Актуальність цього підходу підтверджує і статистика: минулого року понад 67% фішингових атак покладалися на штучний інтелект [3], що вимагає від фахівців глибокого розуміння принципів роботи ШІ-інструментів як у захисті, так і в нападі.

Водночас надмірна залежність від автоматизованих інструментів несе в собі серйозні педагогічні ризики. Українські дослідники зазначають, що "відсутність глибоких знань і критичного мислення призводить до того, що студенти легко обходять інтелектуальні виклики, отримуючи сертифікати без реальних компетенцій" [4]. У кібербезпеці це є особливо небезпечним — фахівець без справжнього розуміння матеріалу стає слабкою ланкою у системі захисту організації. "Штучний інтелект автоматизує дедалі більше когнітивних завдань, але людські навички — критичне мислення, емпатія, етичне судження та творчість — залишаються поза його досяжністю" [5] — а саме ці якості є вирішальними під час реагування на нестандартні інциденти.

Штучний інтелект кардинально змінює підходи до підготовки фахівців з кібербезпеки — від пасивного засвоєння теорії до активної взаємодії з реальними сценаріями загроз. За прогнозами, ринок ШІ у кібербезпеці досягне 46,3 мільярда доларів до 2027 року [6] — що свідчить про масштаб трансформації галузі та зростаючу потребу у фахівцях, здатних ефективно працювати з цими технологіями. Оптимальна модель підготовки поєднує можливості штучного інтелекту з традиційними педагогічними підходами — саме такий синтез здатний сформувати фахівця, готового до реальних викликів цифрового світу.

1. synchron.ua — Кібератаки на бізнес України 2025. https://synchron.ua/cyberattacks-on-ukrainian-business-2025-uk/
2. oksim.ua — ТОП-10 платформ для практики з кібербезпеки.
3. https://www.oksim.ua/top-10-platform-dlya-praktiki-z-kiberbezpeki/3.synchron.ua — Кібератаки на бізнес України 2025. https://synchron.ua/cyberattacks-on-ukrainian-business-2025-uk/
4. kreschatic.kiev.ua — Чому студенти вразливі перед штучним інтелектом, 2025. https://kreschatic.kiev.ua/tehno/2025/05/15/chomu-studenty-vrazlyvi-pered-shtuchnym-intelektom-i-yak-universytety-mozhut-cze-zminyty.html
5. focus.ua — Епоха штучного інтелекту: яка освіта і які професії вціліють. https://focus.ua/uk/ukraine/751608-epoha-shtuchnogo-intelektu-yaka-osvita-i-yaki-profesiji-vciliyut
6. bdo.ua — Роль штучного інтелекту в кібербезпеці. https://www.bdo.ua/en-gb/insights-1/information-materials/2024/the-role-of-ai-in-cybersecurity-anticipating-and-preventing-attacks

Багаторівневий захист мобільного застосунку

УДК 004.056.5:004.9

Любомир Боценюк

Ужгородський національний університет, liubomyr.botseniuk@uzhnu.edu.ua

Сьогодні смартфон фактично став персональним сховищем конфіденційної інформації: від фото й листування до банківських застосунків та приватних фінансових нотаток. Саме тому захист даних на телефоні є не менш важливим, ніж захист у мережі. Реальні ризики часто виникають у повсякденних ситуаціях: підглядання через плече, випадковий скріншот, запис екрана, залишений застосунок у перемикачі останніх програм або короткочасний доступ до пристрою без власника. Навіть якщо застосунок працює офлайн і не передає дані на сервер, загроза витоку не зникає — вона просто зміщується на локальне зберігання та поведінку інтерфейсу.

У межах проєкту створено мобільний офлайновий фінансовий застосунок для обліку та керування коштами. Він дозволяє створювати кілька гаманців, задавати назву, валюту та суму, додавати курс обміну й одразу бачити перерахунок у гривні. Застосунок підраховує загальний баланс по всіх гаманцях, підтримує редагування й видалення записів, зміну порядку гаманців, а також має режим приватності: суми можна швидко замаскувати або контрольовано відображати — окремо по кожному гаманцю чи одразу для всіх.

Захист реалізований за принципом багаторівневості (defense-in-depth), коли безпека не зводиться до одного механізму, а розподіляється на кілька незалежних шарів.

Перший рівень — автентифікація доступу: вхід через PIN-код і, за потреби, біометрію (Face ID/Touch ID) з коректною обробкою сценаріїв, коли користувач просто скасовує біометричну перевірку.

Другий рівень — захист від підбору коду: ліміт невдалих спроб і прогресивне блокування (1 секунда → 2 → 5 → 10 → 30…), що робить атаки типу brute-force практично неефективними, при цьому користувач бачить прозорий таймер блокування.

Третій рівень — контроль сесії: автоматичне блокування при неактивності та під час повернення застосунку з фону, щоб фінансова інформація не залишилася відкритою без нагляду.

Четвертий рівень — захист від витоку через інтерфейс: заборона скріншотів і запису екрана, а також захист мініатюр у перемикачі застосунків там, де це підтримується платформою.

Окремо важливий базовий принцип — безпечне зберігання: чутливі дані та ключі доступу зберігаються в захищеному сховищі пристрою в зашифрованому форматі, щоб навіть у разі доступу до файлів вони не читалися у відкритому вигляді.

Додатковим напрямом у межах проєкту став модуль Face ID, який розширює класичну ідею «пустити / не пустити». У більшості застосунків біометрія — це бінарна перевірка: користувач пройшов автентифікацію і отримав доступ. У запропонованому рішенні підхід інший: Face ID виступає інструментом ідентифікації, тобто дозволяє не просто підтвердити факт входу, а визначити,

хто саме зайшов у систему. Це дає можливість прив'язувати сесію до конкретної особи, а також застосовувати розмежування доступу залежно від ролі. Концепція працює так: під час реєстрації для користувача створюється локальний профіль з ідентифікатором та роллю, а під час входу здійснюється біометричне розпізнавання, яке визначає належність до одного із зареєстрованих профілів. Після успішної ідентифікації система може надавати різні рівні доступу до функцій і даних. Таким чином, біометрія перетворюється на основу керування правами, а не лише на зручний «замок» для входу.

Логічним продовженням розвитку застосунку є поглиблення цього механізму: розширення системи ролей і політик доступу, додавання сценаріїв «підвищеного підтвердження» для критичних дій (повторна ідентифікація), а також гнучке налаштування правил приватності під конкретного користувача. Перспективними є й режими спільного використання одного пристрою кількома людьми, ведення безпечного локального журналу подій (без фінансових деталей, але з фіксацією фактів входу/виходу) та посилення захисту локального зберігання і резервних копій, щоб навіть у разі компрометації середовища дані залишалися недоступними без ключів і підтвердження особи.

У підсумку багаторівневий підхід дозволяє зробити офлайновий фінансовий застосунок не лише зручним, а й стійким до типових реальних загроз. Він поєднує механізми операційної системи, продуману логіку доступу та приватність інтерфейсу користувача (UI), формуючи просте правило: навіть якщо один бар'єр не спрацює, інші шари все одно зменшать ризик витоку та несанкціонованого доступу.

Безпека транспортної інфраструктури України в умовах повномасштабної війни як пріоритетне завдання Державної спеціальної служби транспорту

УДК 656.078:351.862 (477)

Володимир Будз[1], Сергій Костиря[2], Станіслав Шумлянський[3]

Науково-дослідний центр Державної спеціальної служби транспорту,
[1]budzwolodymyr@gmail.com, [2]kostyrya81@gmail.com,
[3]s.shumlianskyi@dsst.gov.ua

Критична інфраструктура (КІ) забезпечує повноцінне функціонування суспільства. До неї належать енергетичні мережі, транспортні вузли, зв'язок, системи водопостачання, охорона здоров'я, банківська система, оборонна промисловість, а також інформаційні технології, в умовах війни вона стає головною мішенню для ворога [1, с. 75]. Руйнування транспортної інфраструктури разом з іншими діями ворога створили безпрецедентні труднощі для національної економіки [2, с. 167], а безпека транспортних маршрутів є одним із ключових викликів для логістичної діяльності в умовах війни [2, с. 168]. Зокрема, за даними СБУ, на кінець січня 2026 року було нейтралізовано понад 14 тисяч масштабних кібератак на Україну за час повномасштабного вторгнення [3]. В умовах повномасштабної російсько-української війни захист КІ України набуває стратегічного значення, оскільки

захищена КІ зможе забезпечити стабільне функціонування української держави та не допустити зростання соціальної напруги в умовах повномасштабної війни.

Одним із найбільш важливих елементів КІ України є транспортна інфраструктура (ТІ). У цьому ракурсі *мета дослідження* – виявити заходи підвищення безпеки ТІ України, яка перебуває у сфері безпекової відповідальності Державної спеціальної служби транспорту (далі – Держспецтрансслужби), яка виконує функцію підтримки безпеки ТІ через її охорону, оборону, технічне прикриття, відновлення пошкоджених об'єктів та будівництво нових, а також проводить розмінування вибухонебезпечних предметів на об'єктах ТІ. Безперебійне функціонування ТІ України створює підґрунтя для ефективної логістики у сфері оборони та економіки, зокрема через військові та гуманітарні вантажні перевезення, забезпечує соціальну стабільність та соціальну захищеність українського населення.

ТІ України, яка перебуває у сфері безпекової відповідальності Держспецтрансслужби охоплює: автомобільні та залізничні мости, мережу шляхів (дороги, автомагістралі, залізниці, річкові, морські, повітряні), транспортні вузли (порти, вокзали, аеропорти), тунелі, а також розмаїті інженерні споруди, які забезпечують функціонування ТІ. В умовах повномасштабної російсько-української війни зазначені об'єкти ТІ стають одними із пріоритетних цілей для нанесення ракетних ударів, диверсійних дій, кібератак та інформаційно-психологічного впливу через явну чи фейкову загрозу їх замінування та можливих терористичних актів. У цьому ракурсі ключовим викликом щодо безпечного функціонування ТІ України є її стійкість до комбінованих загроз, які можуть включати фізичне руйнування/знищення об'єктів ТІ через ракетні удари та диверсійні дії разом із кібератаками на цифрові системи управління транспортними потоками та пасажирськими перевезеннями.

Для створення безпечних умов використання ТІ Держспецтрансслужба на міжвідомчому рівні забезпечує резервування транспортних маршрутів, приймає участь у створенні моделей альтернативних логістичних шляхів, комплектує мобільні інженерні підрозділи, які здатні швидкими темпами відновлювати пошкоджені об'єкти ТІ, розробляє та впроваджує сучасні інженерні технології швидкого відновлення об'єктів ТІ, у тому числі – через наукові розробки.

Сучасна ТІ України функціонує на основі цифрових технологій, автоматизованих систем управління транспортним рухом, які є вразливими для кібератак ворога. Особливо небезпечними є кібератаки на автоматизовані системи керування залізничним рухом, інформаційні мережі логістичних центрів, цифрові канали координації військових і гуманітарних перевезень. Кібератаки на ТІ у першу чергу спрямовані на її дестабілізацію, порушення логістики, блокування управлінських процесів та створення умов для транспортної кризи. На цій основі Держспецтрансслужбі слід зосередити увагу також на кібербезпеці ТІ, оскільки автоматизовані системи управління рухом транспорту та цифрові диспетчерські мережі можуть бути об'єктами кібератак.

Безпека ТІ повинна включати постійний моніторинг кіберзагроз, резервне копіювання даних, підготовку спеціалістів з кібербезпеки, створення багаторівневої системи кіберзахисту через сегментацію цифрових мереж,

створення рівнів доступу до інформаційних ресурсів, багатофакторну автентифікацію персоналу, ізоляцію критично важливих систем управління від відкритих мереж передачі даних, дублювання серверної інфраструктури, використання захищених мобільних центрів управління та створення альтернативних каналів передачі інформації, підвищення рівня знань із кібербезпеки персоналу ТІ, використання систем штучного інтелекту для аналізу ризиків щодо можливих кіберзагроз.

Концепцію безпеки ТІ України слід будувати на основі системного підходу як багаторівневу модель організаційних, міжвідомчих, технічних, наукових, інформаційних та кібербезпекових заходів, які створюють надійні умови використання ТІ України.

1. Ковальов К. Є. Сучасні виклики та загрози для критичної інфраструктури України під час воєнного стану. *Приватне та публічне право*. – 2025. – № 1. – С.75-80. DOI: https://doi.org/10.32782/2663-5666.2025.1.12
2. Коваль К. П., Тєлєгін О. О. Логістика в умовах війни: ключові виклики та шляхи протидії загрозам. *Проблеми і перспективи економіки та управління*. – 2025. – № 1 (41). – С. 167-178. DOI: https://doi.org/10.25140/2411-5215-2025-1(41)-167-178
3. СБУ нейтралізувала понад 14 тисяч масштабних кібератак на Україну за час повномасштабного вторгнення. URL: https://ssu.gov.ua/novyny/sbu-neitralizuvala-ponad-14-tysiach-masshtabnykh-kiberatak-na-ukrainu-za-chas-povnomasshtabnoho-vtorhnennia (дата звернення: 01.05.2026).

Graph-Based Model for Risk Assessment of Access to Corporate Databases in Network Infrastructure

UDK 004.056.5:004.7

Oleksandr Budzynskyi[1], Yurii Shchavinsky[2]

State University of Information and Communication,
[1]oleksandr.email@gmail.com, [2]y.shchavinskyi@duikt.edu.ua

Modern corporate databases are characterized by a large number of interconnected nodes, services and data exchange channels. Under such conditions, ensuring cybersecurity becomes one of the key tasks. Traditional security mechanisms are focused mainly on the analysis of individual events or network traffic, and cannot fully assess the risks of multi-stage attacks implemented by sequentially compromising network nodes. In this regard, the use of graph models becomes promising, since they allow formalizing the structure of the network infrastructure and assessing risks at the level of possible access paths to the database.

The relevance of the topic of using graph models to improve the effectiveness of corporate database protection is substantiated in many studies. In work [1], the authors showed the need to use structural graph attack models to detect complex lateral movement scenarios. In paper [2], a model for assessing the cyber-physical security

of industrial systems based on “AND/OR” attack graphs is proposed, which allows identifying critical infrastructure components and minimal ways to compromise the system. In study [3] a combination of Bayesian attack graphs with process mining is proposed for dynamic risk assessment in real time.

Unlike the above approaches, in the proposed work, the graph model is combined with an AI-based node anomaly assessment based on Isolation Forest, LSTM, and Autoencoder, as well as an exponential risk model that takes into account the cumulative impact of node criticality along the attack path to the corporate database. This allows for adaptive risk assessment in conditions of variable network activity and dynamic network segmentation. The purpose of the study is to develop a graph-based model for assessing access risks to corporate databases in network infrastructure by considering attack propagation paths, node criticality, and AI-driven anomaly detection mechanisms.

The corporate network is proposed to be represented as a directed graph $G=(V,E)$, where V is the set of network nodes (workstations, servers, routers, databases), and E is the set of edges defining possible transitions between nodes.

This model makes it possible to describe potential attack propagation routes and formalize the relationships between infrastructure components. To evaluate the state of nodes, a combined AI-based approach using Isolation Forest, LSTM, and Autoencoder models is applied. The anomaly level of a node is determined by the following expression:

$$q_v(t) = \sigma(\alpha s_{IF}(v,t) + \beta s_{LSTM}(v,t) + \gamma s_{AE}(v,t)), \quad (1)$$

where s_{IF}, s_{LSTM}, and s_{AE} are the outputs of machine learning models, $\sigma(\cdot)$ is the sigmoid normalization function, and $q_v(t) \in [0,1]$ characterizes the suspiciousness level of a node. α+β+γ=1 - model impact factors

The probability of spreading an attack between adjacent nodes u-v of the network is defined as:

$$p_{u,v}(t) = \sqrt{q_u(t) \cdot q_v(t)} \quad (2)$$

Accordingly, the probability of attack traversal along a path $\pi$is calculated as the product of transition probabilities:

$$P(\pi,t) = \prod_{(u,v)\in\pi} p_{uv}(t) \quad (3)$$

Unlike conventional approaches, the proposed model makes it possible to assess not individual incidents but the risk of reaching a critical resource through a sequence of interconnected network nodes. For integrated risk assessment, an exponential risk model is proposed:

$$R(\pi,t) = P(\pi,t) \cdot I(DB) \cdot \exp(\lambda \sum_{v\in\pi} I(v)) \quad (4)$$

where $I(v)$ denotes the criticality of a node, and λ is the sensitivity coefficient of the model, $I(DB)$ – database criticality

With an exponential function, even a small increase in the number of critical nodes can significantly increase the overall risk level. Model analysis shows that the most

dangerous routes are critical infrastructure nodes. Reducing the number of alternative database access paths significantly reduces the overall risk of system compromise. The results obtained confirm the effectiveness of the graph approach for assessing the risks of accessing corporate databases. The proposed model provides a comprehensive analysis of the network infrastructure, takes into account the structural features of the attack propagation and can be applied in SOC and SIEM systems for automatic management of cybersecurity policies. Prospects for further research include integrating the model into real corporate networks, using industrial datasets, and optimizing the parameters of the artificial intelligence model to increase the accuracy of risk assessment.

1. Stergiopoulos G.. Gritzalis D.. Limnaios E.,. Cyber-Attacks on the Oil & Gas Sector: A Survey on Incident Assessment and Attack Patterns. *IEEE Access* – 2020. p. 1-37. URL: http://doi.org/10.1109/ACCESS.2020.3007960
2. Barrère M., Hankin C., Nicolaou N., Eliades D., Parisini T. Measuring cyber-physical security in industrial control systems via minimum-effort attack strategies. *Journal of Information Security and Applications* - 2020. V. 52. - p. 1-5. URL: http://doi.org/52. 17. 10.1016/j.jisa.2020.102471
3. Vitale F.,, Guarino S., Perone S., Rak M., Mazzocca N. Dynamic Risk Assessment by Bayesian Attack Graphs and Process Mining. *Accepted to the 2026 IEEE International Conference on Cyber Security and Resilience.* 20 Apr 2026. p. 1-6. URL: https://doi.org/10.48550/arXiv.2604.18080

Архітектурний підхід Policy-as-Code для захисту LLM-інференс пайплайнів від атак prompt injection

УДК 004.056.5, 004.822

О.П. Вахула[1],

[1]Національний університет «Львівська політехніка», Львів;
oleksandr.p.vakhula@lpnu.ua

Інтеграція великих мовних моделей (LLM) у корпоративні системи обробки інформації відкриває нові вектори атак, що не охоплюються класичним апаратом статичного аналізу. Атаки типу *prompt injection* дозволяють зловмисникам маніпулювати поведінкою моделі через структурований вхід природною мовою, обходячи системні інструкції або витягуючи конфіденційні дані. Регуляторні вимоги - зокрема EU AI Act (статті 12, 13, 15) та директива NIS2 - висувають додаткову вимогу: кожне автоматизоване рішення у сфері безпеки має бути *аудитопридатним та простежуваним*, що принципово несумісне з «чорноскриньковими» ML-класифікаторами загального призначення.[1,2]

Існуючі підходи до фільтрації промптів поділяються на два полюси: прості ключово-словникові фільтри з детермінованою, але негнучкою логікою і ML-класифікатори, що забезпечують семантичне розуміння ціною непрозорості та ресурсомісткості. Жоден із підходів не задовольняє одночасно критеріям точності, швидкодії та аудитопридатності. Розрив між цими полюсами визначив

мету дослідження: розробити *детерміноване, версіоноване та аудитопридатне* рішення для виявлення prompt injection у LLM-інференс пайплайнах, придатне для регульованих середовищ.[3]

Запропонований підхід ґрунтується на специфікації загроз у вигляді правил *Policy-as-Code* мовою Rego для Open Policy Agent (OPA). Ключовою архітектурною ідеєю є *інвертований розподіл відповідальності*: OPA виступає єдиним авторитетним пунктом прийняття рішень (Policy Decision Point), тоді як ML-компонент - за наявності, відіграє лише допоміжну роль постачальника числових ознак, але не приймає фінального рішення. Таке розмежування гарантує: кожен заблокований запит трасується до конкретного правила і категорії загрози, а бібліотека правил є версіонованою, тестованою і розгортається в наявний CI/CD-пайплайн без навчання окремої моделі. Архітектура охоплює: таксономію п'яти категорій атак на основі OWASP LLM Top 10, модуль інспекції промптів між клієнтом і моделлю, та бібліотеку Rego-правил з покриттям 33 юніт-тестами (всі PASS).

Оцінювання проведено на датасеті з 305 зразків (155 атак, 150 легітимних запитів) з порівнянням трьох методів. Результати наведено у таблиці.

Таблиця 1

Порівняльні результати методів виявлення prompt injection (n = 305)

Метод	P	R	F1	Затримка, мс
OPA/Rego (запропонований)	0.88	0.85	0.84	1.19
Ключовий фільтр	0.86	0.80	0.79	0.01
ML-класифікатор (toxic-comment)	0.25	0.50	0.34	21.74

Посекційний аналіз виявив диференціацію ефективності OPA залежно від категорії загрози: Recall = 1.00 для *data exfiltration*, 0.93 для *goal hijacking*, 0.89 для *role switch*, 0.75 для *context manipulation* і 0.10 для *harmful content*. Низьке покриття останньої категорії пояснюється використанням у датасеті промптів з корпусу AdvBench із натуралістичним формулюванням, що виходить за межі keyword-орієнтованих шаблонів, цю обмеженість визнано як відому слабкість підходу, а не замовчано. Критично низька Precision ML-класифікатора (0.25) спричинена доменним зміщенням: модель, навчена на токсичних коментарях у соціальних мережах, класифікує переважну більшість легітимних технічних запитів як шкідливі.[4]

Порівняно з обома базовими лініями підхід OPA/Rego забезпечує якісно вищий баланс: перевищує ключовий фільтр за F1 (+6.3%) та Recall при збереженні прийнятної Precision; демонструє на порядок нижчу затримку порівняно з ML (1.19 мс проти 21.74 мс); виключає хибнопозитивні спрацювання на структурованих категоріях атак, що є критичним для виробничих середовищ. Практична цінність підходу полягає у його придатності для on-premise розгортань у регульованих галузях без хмарних залежностей та без додаткового навчання моделей. Подальші дослідження спрямовані на

гібридну ML+OPA архітектуру, в якій ML надає ймовірнісні ознаки, а OPA залишається єдиним авторитетним джерелом рішень.[5]

1. OWASP Foundation. OWASP Top 10 for Large Language Model Applications, v2.0. 2025. URL: https://owasp.org/www-project-top-10-for-large-language-model-applications/
2. Regulation (EU) 2024/1689 of the European Parliament and of the Council (EU AI Act). Official Journal of the European Union, 2024.
3. Vakhula O., Opirskyy I. AI Development Security as Code (AISaC): A Policy-Based Approach for Securing AI Engineering Pipelines. CEUR Workshop Proceedings, Vol. 4024, 2025, pp. 170–185.
4. Open Policy Agent. Policy Language (Rego). URL: https://openpolicyagent.org/docs/policy-language (дата звернення: 25.04.2026).
5. Fernández Saura P. et al. On Automating Security Policies with Contemporary LLMs. arXiv:2506.04838, 2025.

Розробка архітектури захищеного менеджера облікових даних з підвищеною стійкістю до GPU-атак

УДК 004.056

Михайло Вдовін[1], Олена Головачова[2]

Національний університет «Одеська політехніка», 9650041@stud.op.edu.ua[1], holovachova@op.edu.ua[2]

Актуальність теми роботи. Зростання кількості вебсервісів, онлайн-платформ та інших цифрових ресурсів призводить до необхідності створення та використання великої кількості облікових даних. У результаті користувачі часто застосовують слабкі або однакові паролі для різних сервісів, що значно підвищує ризик несанкціонованого доступу до їх даних. Менеджери облікових даних є надійним рішенням для централізованого, безпечного зберігання великої кількості облікових даних користувача.

Метою роботи є розробка архітектури захищеного локального менеджера облікових даних, який забезпечує надійне зберігання шляхом використання криптографічних алгоритмів AES-GCM та Argon2.

Однією з найнебезпечніших загроз є офлайн-перебір пароля після отримання файлу з даними або резервної копії. У такому сценарії зловмисник може використовувати потужні графічні процесори (GPU) або спеціалізованих схем (ASIC) для перебору великої кількості варіантів.

Для протидії цьому типу атак в архітектурі програми доцільно застосовувати функції виведення ключів, стійкі до апаратного прискорення. Найбільш розповсюдженим рішенням є Argon2 [1], який спеціально розроблено з урахуванням загроз, пов'язаних з GPU та ASIC.

Менеджер облікових даних будується за такою логікою: 1) Користувач створює головний пароль; 2) Пароль не зберігається напряму, а обробляється алгоритмом Argon2 (створюється хеш для перевірки введеного пароля при вході

та створюється ключ для шифрування даних); 3) Усі дані користувача шифруються та зберігаються в зашифрованому вигляді. Майстер-пароль користувача проходить через Argon2 → отримується ключ шифрування → цим ключем через AES-GCM шифрується база паролів.

На відміну від класичних швидких хешів, Argon2 навмисно сповільнює обробку кожного пароля. Для користувача це означає незначне збільшення часу входу, а для зловмисника – різке зростання очікування при кожній спробі підбору. Саме така асиметрія і є основою криптографічного захисту.

GPU-атаки базуються на здатності графічних процесорів виконувати велику кількість однотипних операцій паралельно. Саме тому зловмисник може перевіряти мільйони варіантів за короткий час, якщо алгоритм хешування є недостатньо складним.

Архітектура менеджера облікових даних розділена на чотири окремі модулі: 1) модуль автентифікації – при першому вході відповідає за створення майстер-пароля, перевірку його надійності та створює файл автентифікації, в якому зберігаються параметри KDF алгоритму і хеш пароля. В подальшому виконує функції перевірки введеного пароля та його перетворення на криптографічний ключ; 2) модуль збереження даних – забезпечує зчитування введених користувачем даних, їх шифрування та збереження у файл; 3) модуль завантаження даних – відповідає за зчитування та розшифрування даних зі сховища; 4) модуль генерації паролів – створює паролі, відповідно до заданих користувачем параметрів. Архітектура наведена на рис.1.

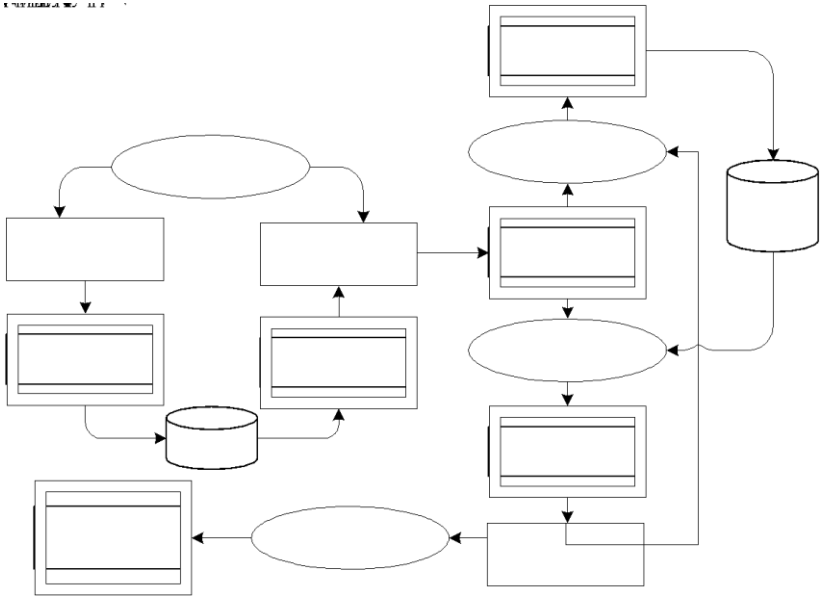


Рис. 1. Архітектура менеджера облікових даних

Argon2 знижує ефективність таких атак завдяки трьом властивостям:1) Argon2 є алгоритмом [2], який вимагає значного обсягу оперативної пам'яті для обчислення. GPU мають високу обчислювальну потужність, але пам'ять для кожного потоку обмежена. Якщо алгоритм потребує багато пам'яті на одну спробу, то кількість одночасних перевірок різко зменшується; 2) Хоча GPU добре підходять для задач із незалежними обчисленнями, Argon2 формує залежності між блоками пам'яті так, що атака стає менш ефективною; 3) Argon2 дозволяє регулювати обсяг пам'яті кількість проходів, ступінь паралелізму. Це дозволяє підібрати параметри, при яких вхід для користувача залишиться зручним, але GPU-атака стане не вигідною.

Запропонована архітектура поєднує зручність для користувача та високий рівень криптографічного захисту.

1. Argon2: вебсайт – URL: https://www.argon2.com (дата звернення: 11.04.2026).
2. Biryukov A., Dinu D., Khovratovich D., Josefsson S. RFC 9106 – Argon2 memory-hard function for password hashing and proof-of-work applications. – 2021.

Machine learning methods for automated assessment in distance learning

UDK 004.89:37.018.43

Kostiantyn Radchenko[1], Wei Shenlai[2]

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", [1]radchenko.kostiantyn@lll.kpi.ua, [2]radche001@gmail.com

The rapid transition to distance education has intensified the need for automated assessment technologies capable of supporting instructors in evaluating large volumes of student work [1]. In many learning management systems (LMS), assessment remains predominantly manual, requiring significant time and leading to potential subjectivity in scoring. Machine learning (ML) methods provide new opportunities for automating evaluation processes and generating individualized feedback for learners [2]. Unlike rule-based approaches, ML models can recognize linguistic patterns, semantic relationships, and contextual nuances in student responses. This study examines the application of machine learning methods for automated scoring and real-time feedback generation in LMS environments, with particular attention to adaptability, interpretability, and efficiency.

One of the most challenging aspects of distance learning is the evaluation of open-ended student responses [3]. Manual assessment demands substantial instructional resources and may result in inconsistent scoring across different evaluators. Existing automated systems typically focus on multiple-choice or short-answer items that can be assessed using pattern-matching techniques. However, such approaches cannot adequately capture semantic depth, logical structure, or argumentative quality in extended written responses. The central objective of this work is the development of

an intelligent grading mechanism capable of evaluating free-text answers while preserving interpretability and transparency in the scoring process.

Previous research on automated assessment methods can generally be divided into two categories: rule-based systems, which rely on predefined keyword lists and scoring templates; machine learning systems, which learn to evaluate responses based on features extracted from human-scored data.

Rule-based approaches are straightforward to implement but lack adaptability and scalability across diverse linguistic styles [2]. Machine learning approaches, particularly those combining TF-IDF representations with semantic embeddings, achieve greater scoring accuracy but may introduce challenges related to model interpretability. Additionally, existing work rarely addresses mechanisms for generating real-time formative feedback for students. Therefore, there is a need for hybrid models that integrate the robustness of linguistic rules with the flexibility of machine learning to support both automated scoring and meaningful feedback delivery in LMS environments.

The proposed method includes three core components: feature extraction, model training, and feedback generation.

First, student responses undergo preprocessing through tokenization, lemmatization, and stop-word removal, after which each text is represented using a hybrid vector that combines TF-IDF weights and Word2Vec semantic embeddings.

Second, a regression-based prediction model (e.g., Random Forest or Linear Regression) is trained on human-scored samples to learn a mapping between textual features and score values.

Finally, after the model predicts a score, the system analyzes low-weight or missing semantic features to automatically generate feedback in the form of corrective suggestions or conceptual hints. This mechanism provides personalized feedback while ensuring interpretability, as instructors may review feature contributions and the rationale underlying the score.

The proposed model was experimentally evaluated on a dataset of manually graded student responses obtained from an online learning course. Performance was compared with a baseline TF-IDF + Cosine similarity scoring method. Results indicate that the proposed approach achieved a higher correlation with human-assigned scores and demonstrated stable performance across multiple evaluation runs. Additionally, the system was able to process responses in real time and produce feedback summaries for learners, which reduced instructor workload and enhanced transparency in assessment practices.

In conclusion, this study presents a machine learning-based approach for automated scoring and feedback generation in distance learning environments. The proposed model integrates TF-IDF features, Word2Vec embeddings, and regression-based prediction to achieve both accurate and interpretable evaluation results. Implemented as a microservice, the system exhibits high scalability and usability within modern learning platforms. Future research will focus on expanding the dataset, incorporating transformer-based language models, and improving the personalization of generated feedback through more advanced linguistic analysis techniques.

1. Nestulya S., Shara S. Distance learning as a relevant educational technology in higher education institutions, *Scientific Bulletin of Mukachevo State University. Series: Pedagogy and Psychology*, vol. 9, no. 1, pp. 39–46, 2023. doi: 10.52534/msu-pp1.2023.39.
2. Vajjala S., Meurers D. Readability Assessment for Text Simplification: From Analyzing Documents to Identifying Sentential Simplifications, *International Journal of Applied Linguistics*, vol. 165, pp. 159–189, 2014. doi: 10.1075/itl.165.2.04vaj.
3. Zupanc K., Bosnić Z. Automated essay evaluation with semantic analysis, *Knowledge-Based Systems*, vol. 120, pp. 118–132, 2017. doi: 10.1016/j.knosys.2017.01.006.

Алгоритмічні ІПСО: генеративні моделі, мікротаргетинг і захист критичних аудиторій

УДК 004.8

Верголяс О.О.

старший лейтенант, кандидат юридичних наук, старший викладач,
Харківський національний університет повітряних сил ім. І. Кожедуба,
arnijazov@gmail.com, ORCID: https://orcid.org/0000-0002-9780-1298

Розглянуто алгоритмічні ІПСО, що використовують генеративні моделі, синтетичні медіа, bot-swarming і мікротаргетинг. Визначено типові сценарії таких атак, індикатори їх виявлення та базову модель реагування. Обґрунтовано, що протидія ШІ-підсиленим ІПСО потребує не окремих спростувань, а узгодженого циклу моніторингу, правової оцінки, стратегічних комунікацій і міжвідомчого обміну індикаторами [1-5].

Генеративний ШІ змінив не лише інструменти ІПСО, а й темп їх виконання. Якщо раніше інформаційна операція потребувала автора, редактора, монтажера і мережі поширення, то нині значну частину циклу можна автоматизувати: мовна модель готує варіанти тексту, генератор зображень створює візуальний супровід, синтез голосу дає псевдоавтентичне аудіо, а бот- або рекламна інфраструктура перевіряє, який меседж краще діє на конкретну групу. Для України це не абстрактна технологічна тема: російська агресія поєднує кібероперації, дезінформацію, психологічний тиск на родини військовослужбовців і спроби розколоти довіру між військом, владою та суспільством.

Алгоритмічні ІПСО доцільно розглядати як зміну архітектури впливу. Її основними компонентами є synthetic identity, deepfake-контент, LLM-мікротексти та bot-swarming. Синтетичні акаунти імітують локальну присутність і створюють враження органічної дискусії. Deepfake може бути не лише повністю синтетичним відео, а й змішаною формою: справжній фрагмент виступу, змінена аудіодоріжка, обрізаний контекст або поширення через канал із готовою аудиторією [3; 5]. LLM-мікротексти діють тонше: короткі повідомлення для військових, родин військовослужбовців, медиків, енергетиків, волонтерів чи місцевих громад маскуються під коментар, «інсайд» або побутову пораду. Крос-платформні хвилі переносять наратив між Telegram,

короткими відео, коментарями під новинами та псевдомедійними обговореннями.

Виявлення таких операцій не можна будувати на одному «детекторі ШІ». Надійнішою є сукупність сигналів: стилометрія тексту, синхронність публікацій, графові ознаки неорганічного поширення, мультимодальні артефакти і темп зміни наративу. Якщо після спростування повідомлення не зникає, а швидко змінює формулювання, канал або аудиторію, ймовірно, діє адаптивний контур впливу. Такі індикатори мають накопичуватися у hash-банках медіа та бібліотеках ознак, сумісних із практиками NIST і CISA [1-3].

Протидія починається не з публічного спростування, а з процесу. Якщо установа не знає, хто фіксує інцидент, хто проводить первинну перевірку, хто готує повідомлення і хто має право говорити публічно, вона програє перші години. Технічний блок має охоплювати перевірку походження медіа, provenance/watermark-маркування офіційного контенту, пасивне детектування дипфейків і крос-платформну кореляцію. Організаційний блок має спиратися на SOC-StratCom-OSINT-ланцюжок: аналітики фіксують сигнал, OSINT перевіряє маршрут поширення, StratCom готує реакцію, юридичний блок оцінює ризики обмеження контенту або звернення до платформи [2-4].

Для України раціональною є тришарова модель реагування. Тактичний рівень (0-2 години) забезпечує моніторинг, первинну верифікацію і фіксацію цифрових артефактів. Оперативний рівень (2-24 години) відповідає за кореляцію індикаторів, запуск контрнаративів, юридичну оцінку і координацію із ЗСУ, СБУ, НКЦК та іншими органами. Стратегічний рівень охоплює міжвідомчий обмін індикаторами, зв'язок із платформами, міжнародними партнерами та аналіз тенденцій [2; 4]. Для військових, родин військових, медиків, енергетиків, освітян і місцевих громад потрібні різні канали довіри; тому trusted voices має бути мережею локально авторитетних комунікаторів, а pre-bunking - способом пояснити типові маніпуляції до того, як конкретний фейк набере масштаб.

Практичні кейси підтверджують цю логіку. Deepfake-відео із закликом до «складання зброї» у березні 2022 року було технічно слабким, але одночасне поширення через сайт, телевізійний рядок, Telegram і соціальні мережі створило коротке вікно паніки. DFRLab також описує фальшиві відео із Валерієм Залужним і відеодзвінок із підробленим образом Петра Порошенка для учасників Інтернаціонального легіону; ці випадки важливі переходом від масової дезінформації до вузького таргетування [5].

Отже, ШІ в ІПСО є новою якістю загрози, а не просто новим інструментом для старої пропаганди. Ефективна модель протидії має поєднувати NIST-сумісні технічні практики, FIMI-орієнтовану аналітику, правову пропорційність і локалізовану комунікацію з критичними аудиторіями. Її результатом має бути не ідеальна цензура, а швидке відокремлення небезпечної операції від інформаційного шуму, збереження доказів і точне спростування [1-5].

1. National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). 2023. URL: https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf (дата звернення:

10.11.2025).
2. National Institute of Standards and Technology. Generative AI Profile for the NIST AI RMF (NIST AI 600-1). 2024. URL: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf (дата звернення: 10.11.2025).
3. NSA; CISA; FBI. Contextualizing Deepfake Threats to Organizations. 2023. URL: https://media.defense.gov/2023/Sep/12/2003298925/-1/-1/0/CSI-deepfakethreats.pdf (дата звернення: 10.11.2025).
4. European External Action Service. 2nd Report on Foreign Information Manipulation and Interference (FIMI). 2024. URL: https://www.eeas.europa.eu/sites/default/files/documents/2024/EEAS-2nd-Report%20on%20FIMI%20Threats-January-2024_0.pdf (дата звернення: 10.11.2025).
5. Osadchuk R. AI tools usage for disinformation in the war in Ukraine. Digital Forensic Research Lab. 09.07.2024. URL: https://dfrlab.org/2024/07/09/ai-tools-usage-for-disinformation-in-the-war-in-ukraine/ (дата звернення: 10.11.2025).

Захист CI/CD-конвеєрів автоматизованого оновлення Docker-контейнерів на основі криптографічного підписування образів

УДК 004.056:004.75

Віталій Тимощук[1], Дмитро Тимощук[2]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]tymoshchuk@tntu.edu.ua, [2]dmytro.tymoshchuk@gmail.com

Розробка програмного забезпечення дедалі частіше базується на використанні контейнеризації та автоматизованих CI/CD-конвеєрів. Docker-контейнери дають змогу швидко збирати, доставляти й оновлювати застосунки, однак водночас створюють нові ризики для безпеки. Особливо критичною є проблема довіри до контейнерного образу, який автоматично завантажується з реєстру та запускається у продуктивному середовищі. У типовій схемі CI/CD новий образ збирається після зміни коду, публікується в container registry, після чого сервер отримує його за тегом, наприклад latest. Такий підхід є зручним, але не гарантує, що завантажений образ справді відповідає легітимній версії застосунку.

Основна загроза полягає в тому, що контейнерний образ може бути підмінений у реєстрі, у процесі доставки або через помилкову конфігурацію CI/CD-конвеєра. У такому випадку продуктивне середовище може запустити шкідливий або застарілий образ як легітимне оновлення. Це безпосередньо пов’язано з атаками на ланцюг постачання програмного забезпечення, кількість яких суттєво зростає внаслідок активного використання автоматизованих засобів збірки, сторонніх залежностей і container registry [1]. Традиційні засоби мережевого захисту не завжди здатні виявити такі атаки, оскільки скомпрометований образ може виглядати як звичайне оновлення.

Метою роботи є підвищення безпеки автоматизованого оновлення Docker-контейнерів шляхом використання криптографічного підписування

контейнерних образів, перевірки їх автентичності перед розгортанням і прив'язки образів до незмінного SHA-256 дайджесту. Запропонований підхід орієнтований на інтеграцію в наявний CI/CD-процес без суттєвої зміни його логіки.

У межах запропонованої моделі контейнерний образ після збірки не лише публікується в реєстрі, а й підписується за допомогою інструмента Cosign. Для цього використовується keyless-підхід Sigstore, що дає змогу виконувати підписування без зберігання довготривалих приватних ключів. У середовищі GitHub Actions автентифікація може здійснюватися через OIDC, після чого сервіс Fulcio видає короткоживучий сертифікат для підписування конкретного артефакту [2]. Це знижує ризики компрометації ключів і спрощує впровадження механізмів підпису в CI/CD.

Важливою складовою підходу є використання адресації образів за криптографічним дайджестом, а не лише за тегом. Теги на кшталт latest є змінними й можуть у різні моменти часу вказувати на різні образи. Натомість SHA-256 дайджест однозначно ідентифікує вміст контейнерного образу. Якщо образ змінено хоча б мінімально, його дайджест також зміниться. Тому використання формату repository@sha256:<digest> забезпечує отримання сервером саме того артефакту, який відповідає зафіксованому дайджесту, а в поєднанні з перевіркою цифрового підпису дає змогу підтвердити, що цей артефакт був зібраний і підписаний у CI/CD-конвеєрі [3].

Запропонований процес складається з двох основних етапів. На першому етапі CI/CD-конвеєр виконує збірку Docker-образу, публікує його в container registry, отримує SHA-256 дайджест і підписує образ за допомогою Cosign. До підпису можуть додаватися анотації, які фіксують контекст збірки: ідентифікатор workflow, номер запуску, commit hash і час створення образу. Це підвищує прозорість і дає змогу встановити, з якої версії вихідного коду та в межах якого CI/CD-процесу був сформований артефакт.

На другому етапі, перед розгортанням, сервер виконує обов'язкову верифікацію підпису. Перевіряється не лише факт наявності підпису, а й відповідність підписувача очікуваному репозиторію та workflow. Якщо образ не має валідного підпису або був підписаний неавторизованим джерелом, процес розгортання зупиняється. Таким чином, навіть у разі компрометації контейнерного реєстру зловмисник не зможе непомітно впровадити власний образ, оскільки він не пройде криптографічну перевірку.

Окрему увагу приділено rollback-атакам. Така атака полягає у спробі повернути систему до старішої, але раніше підписаної версії образу, яка може містити вже відомі вразливості. Сам по собі цифровий підпис підтверджує автентичність артефакту, але не підтверджує його актуальність. Тому в запропонованій моделі передбачено політику контролю версій і часу підписання. Відкат дозволяється лише до попередньо відомих і перевірених стабільних версій, а не до довільного старого образу. Такий підхід дає змогу зберегти можливість легітимного rollback у разі помилки релізу та зменшує ризик використання rollback як каналу атаки.

Для перевірки ефективності підходу було змодельовано кілька типових сценаріїв: підміну образу в container registry, зміну тегу latest, спробу запуску

непідписаного образу, несанкціоноване додавання стороннього образу до довіреного реєстру та спробу зловмисного rollback. У базовому підході, де використовується лише тег без перевірки підпису, підміна образу може призвести до запуску шкідливого контейнера. У захищеній моделі розгортання прив'язується до конкретного SHA-256 дайджесту та перед запуском обов'язково виконується перевірка підпису. У результаті підроблені, непідписані або підписані неавторизованим суб'єктом образи блокуються ще до запуску.

Практична перевага запропонованого рішення полягає в тому, що воно не потребує радикальної перебудови CI/CD-процесу. Підписування й перевірка додаються як окремі автоматизовані кроки. При цьому накладні витрати на виконання криптографічних операцій є незначними порівняно із загальним часом збірки та розгортання контейнерного застосунку. Отже, модель може бути використана в DevSecOps-практиках як додатковий рівень захисту програмних артефактів [4].

Водночас криптографічне підписування не усуває всі можливі ризики. Якщо шкідливий код потрапив до образу ще до моменту підписання, наприклад через компрометацію CI-середовища, залежностей або runner-а, такий образ може бути формально валідно підписаний. Тому запропонований підхід доцільно поєднувати з іншими засобами безпеки: скануванням образів на вразливості, мінімізацією прав доступу в CI/CD, перевіркою сторонніх залежностей, code review та політиками контролю запуску контейнерів у середовищі виконання [5].

Отже, використання Cosign/Sigstore, адресації за SHA-256 digest і обов'язкової верифікації підпису перед розгортанням суттєво підвищує рівень довіри до автоматизованих оновлень Docker-контейнерів. Запропонований підхід забезпечує перевірку автентичності та цілісності контейнерних артефактів, а також створює основу для контрольованого керування їх версіями за рахунок використання незмінних digest і політик дозволених релізів. Це зменшує ризики атак на ланцюг постачання програмного забезпечення та може бути інтегроване в сучасні CI/CD-конвеєри без суттєвої зміни їх логіки.

1. Sonatype. *State of the Software Supply Chain Report: 10 Year Look*. URL: https://www.sonatype.com/state-of-the-software-supply-chain/2024/10-year-look.
2. Newman Z., Meyers J. S., Torres-Arias S. *Sigstore: Software Signing for Everybody*. CCS '22: ACM SIGSAC Conference on Computer and Communications Security, 2022. DOI: 10.1145/3548606.3560596.
3. RFC 6234. *US Secure Hash Algorithms: SHA and SHA-based HMAC and HKDF*. URL: https://datatracker.ietf.org/doc/html/rfc6234.
4. GitHub. *Sigstore Cosign: Code Signing and Transparency for Containers and Binaries*. URL: https://github.com/sigstore/cosign.
5. Koishybayev I., Nahapetyan A., Zachariah R., Muralee S., Reaves B., Kapravelos A., Machiry A. *Characterizing the Security of GitHub CI Workflows*. 31st USENIX Security Symposium, 2022. P. 2747–2763.

Проблеми та обмеження виявлення аномалій у кіберфізичних системах критичної інфраструктури

УДК 004.056:004.8

Ігор Воробець

Тернопільський національний технічний університет імені Івана Пулюя, ihor_vorobets0809@tntu.edu.ua

Питання безпеки критичної інфраструктури набуває особливої актуальності в умовах сучасних воєнних і кіберзагроз, пов'язаних з російсько-українською війною. Об'єкти енергетики, промисловості, транспорту та зв'язку функціонують на основі кіберфізичних систем (КФС), які забезпечують інтеграцію цифрових алгоритмів керування з фізичними процесами та стають цілями кіберфізичних атак, що можуть призводити до порушення їх функціонування та виникнення аварійних ситуацій.

До причин їх вразливості можна віднести складну ієрархічну структуру, а також використання відкритих мережевих протоколів. Атаки на КФС можуть бути спрямовані не лише на порушення конфіденційності інформації, а й на вплив на фізичний стан об'єктів, що призводить до катастрофічних наслідків - від зупинки виробничих процесів до масштабних техногенних аварій [1]. У зв'язку з цим постає необхідність систем моніторингу для раннього виявлення інцидентів безпеки у таких системах.

Для вирішення цієї проблеми часто застосовуються методи машинного навчання (machine learning, ML), зокрема алгоритми виявлення аномалій, що дозволяють ідентифікувати відхилення від нормального функціонування системи [2], зокрема аномальні зміни технологічних параметрів, несанкціоновану зміну режимів роботи обладнання, підміну даних сенсорів або нетипову мережеву активність.

Завдяки автоматизованому аналізу великих обсягів даних методи ML дозволяють виявляти складні закономірності та відхилення у функціонуванні КФС. Однак їх ефективність обмежується низкою чинників, особливо пов'язаних із підготовкою даних для навчання моделей.

Метою роботи є аналіз проблеми дефіциту даних при навчанні ML-моделей у задачах виявлення аномалій у КФС і систематизація підходів до її вирішення в контексті забезпечення безпеки об'єктів критичної інфраструктури.

Специфіка функціонування таких систем полягає в тому, що реальні аномальні події й інциденти безпеки трапляються рідко та мають різні прояви. Це унеможливлює формування збалансованих наборів навчальних даних, що, у свою чергу, спричиняє зниження здатності моделей до узагальнення і ризик некоректної роботи системи виявлення аномалій.

У доповіді проведено аналіз підходів до вирішення проблеми дефіциту аномальних даних при навчанні ML-моделей для виявлення аномалій. Розглянуті підходи доцільно розділити на дві групи.

Перша група складається з підходів, які мінімізують потребу в аномальних даних для навчання моделей. До них належать методи однокласової класифікації (one-class classification, OCC), відповідно до яких ML-модель навчається на нормальних даних та визначає аномалії як відхилення від

визначеної «норми». Використання методів OCC не вимагає наявності аномальних даних при навчанні моделі та загалом є ефективним для задач із суттєвим дисбалансом даних. Водночас моделі на основі OCC часто показують високу чутливість до змін у даних та підвищену кількість помилкових спрацювань [3]. Також можна розглянути методи навчання з малою кількістю прикладів (few-shot learning, FSL), що дозволяють навчати моделі в умовах обмеженої кількості аномальних даних, здійснюючи формування узагальнених представлень даних й оцінюючи подібність між ними. Основною перевагою методів FSL є можливість виявляти аномалії за обмеженої кількості прикладів аномальних даних. Проте їх ефективність значною мірою залежить від формування представлень і вибору архітектури моделі [4].

Підходи другої групи полягають у штучному розширенні навчальної вибірки шляхом генерування синтетичних аномальних даних. Це можна здійснити за допомогою генеративних моделей, які дозволяють моделювати сигнали у КФС і таким чином створювати нові приклади аномалій. Генеративні моделі дозволяють збільшити кількість і різноманітність аномальних прикладів, однак вони можуть характеризуватися нестабільністю навчання, а синтетичні дані не завжди повною мірою відтворюють статистичні характеристики реальних даних [5].

Таким чином, у доповіді проаналізовано проблему дефіциту навчальних даних як одного з ключових обмежень застосування методів виявлення аномалій для попередження кіберфізичних атак на КФС критичної інфраструктури. Розглянуто підходи до вирішення цієї проблеми, які поділено на дві групи: на основі зменшення залежності від аномальних даних, зокрема методи OCC та FSL, і підходи на основі штучного розширення навчальної вибірки з використанням генеративних моделей.

1. Cyber-physical systems security–a survey / A. Humayed et al. IEEE Internet of Things Journal. 2017. Vol. 4, no. 6. P. 1802–1831. URL: https://doi.org/10.1109/jiot.2017.2703172 (дата звернення: 09.05.2026).
2. Deep learning for anomaly detection: a review / G. Pang et al. ACM Computing Surveys. 2021. Vol. 54, no. 2. P. 1–38. URL: https://doi.org/10.1145/3439950 (дата звернення: 09.05.2026).
3. Seliya N., Abdollah Zadeh A., Khoshgoftaar T. M. A literature review on one-class classification and its potential applications in big data. Journal of Big Data. 2021. Vol. 8, no. 1. URL: https://doi.org/10.1186/s40537-021-00514-x (дата звернення: 09.05.2026).
4. Ntalampiras S., Potamitis I. Few-shot learning for modeling cyber physical systems in non-stationary environments. Neural Computing and Applications. 2022. URL: https://doi.org/10.1007/s00521-022-07903-0 (дата звернення: 09.05.2026).
5. Generative adversarial networks for synthetic data generation: a systematic review of techniques, applications, and evaluation methods / R. Thinakaran et al. International Journal of Innovative Research and Scientific Studies. 2025. Vol. 8, no. 5. P. 286–293. URL: https://doi.org/10.53894/ijirss.v8i5.8655 (дата звернення: 09.05.2026).

Готовність IT-інфраструктури до епохи квантових обчислень

УДК 621.395.7 (043.2)

Павло Воробець[1]

Національний університет "Львівська політехніка", [1]pavlo.a.vorobets@lpnu.ua

Стрімкий розвиток квантових обчислень формує новий клас загроз для сучасної IT-інфраструктури. Ключовою проблемою є потенційна здатність квантових алгоритмів, зокрема алгоритму Шора, ефективно розв'язувати задачі факторизації великих чисел та обчислення дискретного логарифму, що лежать в основі більшості сучасних криптографічних систем. У результаті під загрозою опиняються базові механізми захисту інформації, які використовуються в IT-інфраструктурі, включаючи протоколи TLS, VPN-з'єднання, інфраструктуру відкритих ключів (PKI), а також системи управління ідентифікацією та доступом. Компрометація цих механізмів може призвести до порушення конфіденційності, цілісності та автентичності даних.

Додаткову небезпеку становить сценарій "harvest now, decrypt later", за якого зашифровані дані можуть накопичуватися зловмисниками з метою їх подальшого розшифрування після досягнення достатнього рівня розвитку квантових технологій. Це особливо критично для даних із тривалим життєвим циклом, таких як фінансова, медична або державна інформація [1]. Незважаючи на усвідомлення потенційних ризиків, рівень готовності сучасної IT-інфраструктури до квантових загроз залишається низьким. Більшість організацій продовжує використовувати криптографічні алгоритми, стійкість яких базується на обчислювальній складності задач, вразливих до квантових алгоритмів.

Однією з ключових проблем є відсутність повної інвентаризації криптографічних залежностей в інфраструктурі. У багатьох випадках організації не мають чіткого розуміння, де саме і яким чином використовуються криптографічні механізми — у протоколах передачі даних, внутрішніх сервісах, API, системах зберігання або сторонніх рішеннях. Додатковим ускладненням є висока ступінь інтегрованості криптографії в IT-системи. Криптографічні механізми часто жорстко вбудовані в архітектуру застосунків, мережевих сервісів і платформ, що суттєво ускладнює їх заміну або оновлення без впливу на працездатність систем. Також значним обмеженням є відсутність практичного досвіду впровадження постквантових алгоритмів у продуктивних середовищах. Хоча відповідні стандарти вже розробляються, їх інтеграція в існуючі IT-ландшафти потребує часу, ресурсів і зміни підходів до проектування безпеки. У сукупності ці фактори свідчать про те, що більшість сучасних IT-інфраструктур не готова до швидкої адаптації в умовах появи квантових загроз, що підвищує ризики для довгострокового захисту даних [2].

Квантові обчислення по-різному впливають на криптографічні алгоритми залежно від їх типу. Найбільш критичний вплив спостерігається для асиметричних алгоритмів, таких як RSA, ECC та Diffie–Hellman. Їх безпека базується на складності задач факторизації великих чисел та обчислення дискретного логарифму, які можуть бути ефективно розв'язані за допомогою

квантових алгоритмів. Це означає, що у разі появи масштабованих квантових комп'ютерів такі алгоритми можуть бути повністю скомпрометовані.

Натомість симетричні алгоритми шифрування, зокрема AES, є менш вразливими до квантових атак. Використання алгоритму Гровера дозволяє лише квадратично прискорити перебір ключів, що фактично зменшує ефективну довжину ключа вдвічі. У практичному вимірі це означає необхідність переходу на довші ключі (наприклад, використання AES-256 замість AES-128) для збереження належного рівня безпеки [3]. Подібний підхід застосовується і до криптографічних хеш-функцій. Квантові алгоритми знижують їх стійкість до пошуку колізій, однак не призводять до повної компрометації. Відповідно, підвищення рівня безпеки досягається шляхом використання хеш-функцій із більшою довжиною вихідного значення.

З огляду на зазначені ризики, ключовим завданням є підготовка IT-інфраструктури до поступового переходу на квантово-стійкі механізми захисту. Одним із базових принципів такої підготовки є впровадження підходу crypto-agility — здатності систем гнучко змінювати криптографічні алгоритми без суттєвих змін архітектури. Першочерговим кроком є інвентаризація криптографічних залежностей, включаючи TLS, VPN, PKI, API та системи управління секретами. Це дозволяє визначити критичні компоненти інфраструктури та пріоритети їх модернізації. Важливим аспектом є централізоване управління ключами та сертифікатами, що спрощує їх ротацію та подальший перехід на нові алгоритми. Додатково доцільним є впровадження гнучких конфігурацій безпеки, зокрема підтримки змінюваних криптографічних наборів і гібридних рішень.

У висновку слід зазначити, що підготовка до квантової епохи не є одноразовим заходом, а тривалим процесом трансформації IT-інфраструктури. Впровадження принципів гнучкості, централізації управління криптографією та поступової міграції дозволить знизити ризики та забезпечити стійкість систем у майбутньому.

1. Kumar M. Post-quantum cryptography Algorithm's standardization and performance analysis. Array. – 2022. – Vol. 15. – Article 100242.
2. Воробець П. А., Горпенюк А. Я., Опірський І. Р. Перехід до постквантових криптографічних систем: виклики, стандартизація та перспективи // Безпека інформації. – 2024. – Т. 30, № 2. – С. 306–315
3. Alagic G., et al. Status Report on the Fourth Round of the NIST Post-Quantum Cryptography Standardization Process. NIST IR 8545. – 2025. – 54 p.

З історії становлення національної системи захисту інформації. 1992–1999 рр.

УДК 94:35.083.8(045)

Валерій Ворожко

СБ України, wp06vv@gmail.com

Першим законодавчим актом, що стверджував інформаційний суверенітет України, став Закон України «Про інформацію», прийнятий ВР України 2 жовтня 1992 р. Цей Закон визначив режим доступу до інформації, поділивши її на відкриту інформацію та інформацію з обмеженим доступом, закріпив за державою право й обов'язок здійснювати контроль за режимом доступу до інформації. Указом Президента України від 1 грудня 1992 р. № 593/92 була створена Державна служба України з питань технічного захисту інформації (далі – ДС ТЗІ), на яку покладалися функції щодо реалізації державної політики, організаційного, нормативного, інженерно-технічного забезпечення технічного захисту інформації. На той час уся діяльність у цій сфері здійснювалася на підставі нормативних документів, затверджених Держтехкомісією СРСР/РФ. ДС ТЗІ підписала декілька угод про співробітництво з Держтехкомісією РФ та перебувала у фарватері політики РФ у сфері технічного захисту інформації.

Одним із питань на підготовчому етапі формування власної системи охорони державної таємниці (далі – СОДТ) було визначення державного органу як спеціально уповноваженого органу державної влади з головним завданням реалізації державної політики у цій сфері діяльності. У той період існувала думка, що формування СОДТ, аналогічної радянському режиму, могло б створити передумови для зловживань щодо застосування таємної інформації, порушень прав і свобод людини. Демократизація правовідносин у сфері, пов'язаній з державною таємницею, повинна була передбачати розширення прав і, водночас, підвищення відповідальності керівників усіх рівнів за режим секретності та персоніфікацію питань щодо віднесення відомостей до державної таємниці та їхнього засекречування. Тому було прийнято рішення про створення окремого спеціального уповноваженого органу державної влади з питань охорони державної таємниці відповідно до досвіду США та в цілому євроатлантичної СОДТ. Ця ідея тоді викликала спротив у проросійської частини української верхівки, яка з часом значно посилилася й після 1999 р. українську СОДТ значною мірою повернули до радянсько-російської моделі.

Визначений ВР України політичний курс щодо формування СОДТ був реалізований відповідними рішеннями уряду держави. Постановою КМУ від 4 травня 1993 р. № 327 було створено Державний комітет України з питань державних секретів, який функціонально був побудований схожим до Управління з нагляду за інформаційною безпекою (Information Security Oversight Office, далі – ISOO), у складі NARA (Національне агентство з питань документації і архівів) США. ISOO відповідає за впровадження та реалізацію президентських виконавчих наказів у сфері таємниць; розробляє стандарти для засекречування і розсекречування документів та типові інструкції з охорони таємних відомостей, займається підвищенням кваліфікації спеціалістів РСО; проводить перевірки в установах на їхню відповідність політиці уряду у галузі інформаційної безпеки; реагує на скарги та пропозиції установ і громадян щодо засекречування та розсекречування інформації; вносить пропозиції Президентові щодо змін у політиці інформаційної безпеки; збирає відповідні звіти з установ, узагальнює їх і готує щорічний звіт з інформаційної безпеки Президенту США тощо. ВР України вже на етапі прийняття законів, які

регламентували діяльність СБ України, залишила за цим державним органом лише функції спеціальної компетенції як правоохоронного органу.

У жовтні 1997 р. постановою КМУ № 1126 затверджена «Концепція технічного захисту інформації в Україні», яка визначала основи державної політики у сфері захисту інформації інженерно-технічними заходами і засобами. Реалізація Концепції забезпечувала єдність принципів формування і проведення державної політики в усіх сферах життєдіяльності особи, суспільства та держави. Зазначено, що необхідність розвитку ТЗІ зумовлена зростанням загроз для інформації, спричинених лібералізацією суспільних та міжнародних відносин, кризовим станом економіки, застосуванням технічних засобів оброблення інформації та засобів зв'язку іноземного виробництва, поширенням засобів несанкціонованого доступу до інформації та впливу на неї. На виконання положень «Концепції технічного захисту інформації в Україні», 16 лютого 1998 р. постановою № 180 Кабінетом Міністрів України затверджене «Положення про забезпечення режиму секретності під час обробки інформації, що становить державну таємницю, в автоматизованих системах». Це «Положення» визначило основи організації, завдання, права, обов'язки суб'єктів правовідносин та порядок забезпечення режиму секретності під час обробки секретної інформації в автоматизованих системах.

Забезпечення безпеки мікроконтролерних систем у робототехнічних комплексах

УДК 004.056

Гануля Роман , Козбур Ігор

Тернопільський національний технічний університет імені Івана Пулюя, Hanularoma@gmail.com Kozbur.igor@gmail.com

Актуальність проблеми. Мікроконтролерні системи є базовим рівнем керування у сучасних робототехнічних комплексах, включаючи промислові роботи, автономні системи та IoT-пристрої. Зростання кількості підключених пристроїв, а також інтеграція робототехніки у промислові та побутові сфери призводять до значного збільшення кіберзагроз, що можуть суттєво порушити роботу систем, викликати втрату даних або фізичну шкоду. Найпоширенішими проблемами є несанкціонований доступ до апаратного забезпечення, підміна прошивки, перехоплення та модифікація даних у комунікаційних каналах, а також атаки побічними каналами (side-channel attacks), що використовують інформацію про споживання енергії чи електромагнітне випромінювання. За даними NIST [1], більшість сучасних IoT-пристроїв мають критичні вразливості на рівні прошивки та автентифікації, що робить їх потенційною мішенню для атак на робототехнічні комплекси. З огляду на це, забезпечення безпеки мікроконтролерних систем є надзвичайно актуальним завданням, яке поєднує в собі апаратні, програмні та комунікаційні аспекти захисту.

Мета роботи. Метою цього дослідження є аналіз основних загроз безпеці мікроконтролерних систем у робототехнічних комплексах та визначення ефективних методів їх захисту. Крім того, робота спрямована на розробку та дослідження комплексного підходу, що поєднує апаратні та програмні засоби

безпеки з урахуванням специфіки робототехнічних систем, де критичною є робота у реальному часі та обмежені ресурси.

Основні загрози. До ключових загроз мікроконтролерних систем належать:

- фізичний доступ до мікроконтролера – зчитування пам'яті, спроби модифікації внутрішніх даних або заміна компонентів;
- атаки на прошивку (firmware tampering) – несанкціоноване змінення коду, що може призвести до функціональних збоїв або створення несанкціонованих каналів доступу;
- перехоплення даних у комунікаційних інтерфейсах (UART, SPI, I2C) – атаки, що дозволяють отримати або змінити інформацію, що передається між мікроконтролером та периферійними пристроями;
- атаки побічними каналами – використання енергоспоживання, електромагнітного випромінювання або акустичних сигналів для отримання конфіденційної інформації;
- недостатній криптографічний захист та контроль доступу – більшість вбудованих систем не забезпечують надійного шифрування або автентифікації, що збільшує ризик компрометації даних [2].

Методи забезпечення безпеки. Ефективний захист мікроконтролерних систем передбачає використання комплексного підходу, що включає:

- Апаратні механізми захисту – використання TrustZone або Secure Enclave, захист пам'яті (*Read-Out Protection*), контролери доступу до периферії;
- Криптографічний захист – шифрування даних за алгоритмами AES та RSA, цифровий підпис прошивки для підтвердження цілісності коду [3,4];
- Безпечне оновлення прошивки (Secure Boot) – перевірка автентичності та цілісності коду перед запуском, що запобігає виконанню зміненого або шкідливого програмного забезпечення [5];
- Захист комунікацій – використання протоколів TLS/DTLS для забезпечення безпечного обміну даними, автентифікація пристроїв у мережі, контроль доступу до критичних команд управління [6];
- Моніторинг та аудит безпеки – відстеження спроб несанкціонованого доступу, ведення журналів подій, оцінка ризиків на основі логів роботи системи.

Наукова новизна. У роботі запропоновано комплексний підхід до захисту мікроконтролерних систем у робототехнічних комплексах, що поєднує апаратні та програмні засоби безпеки з урахуванням обмежених ресурсів, необхідності роботи у реальному часі та специфічних вимог промислових та автономних роботів. Підкреслено важливість інтеграції різних рівнів захисту: фізичного, програмного та мережевого, а також адаптивного реагування на потенційні загрози.

Висновки. Мікроконтролерні системи у робототехнічних комплексах є критично важливими та водночас вразливими до широкого спектра кіберзагроз, включаючи фізичний доступ, модифікацію прошивки, перехоплення даних та атаки побічними каналами. Більшість сучасних вбудованих систем не забезпечують достатнього рівня криптографічного захисту та контролю доступу. Ефективний захист досягається шляхом комплексного підходу: апаратні механізми (TrustZone, Secure Enclave, Read-Out Protection), програмні засоби (шифрування даних, цифровий підпис прошивки, Secure Boot) та захищені комунікаційні протоколи (TLS/DTLS) з автентифікацією пристроїв. Впровадження цих методів підвищує надійність і стійкість систем до кібератак, а подальші дослідження повинні зосереджуватися на оптимізації захисту для ресурсно-обмежених мікроконтролерів, безпеку робототехнічних комплексів протягом усього життєвого циклу системи.

1. NISTIR 8259. Foundational Cybersecurity Activities for IoT Device Manufacturers. https://nvlpubs.nist.gov/nistpubs/ir/2018/NIST.IR.8259.pdf
2. OWASP IoT Top 10. https://owasp.org/www-project-internet-of-things/
3. ARM TrustZone Documentation. https://developer.arm.com/documentation/ 100690/latest
4. FIPS 197: Advanced Encryption Standard (AES). https://csrc.nist.gov/publications/detail/fips/197/final
5. Secure Boot. Trusted Computing Group. https://trustedcomputinggroup.org/ resource/secure-boot/
6. RFC 6347: Datagram Transport Layer Security (DTLS). https://datatracker.ietf.org/doc/html/rfc6347

Аналіз методів тестування генераторів псевдовипадкових чисел відповідно до стандартів NIST та ISO

УДК 004.056

Олег Гарасимчук

Національний університет "Львівська політехніка", oleh.i.harasymchuk@lpnu.ua

Генератори псевдовипадкових чисел (ГПВЧ) є невід'ємним елементом сучасних інформаційних систем. Їхнє застосування охоплює значно ширший спектр завдань, ніж традиційно прийнято вважати: від застосувань для моделювання різноманітних процесів до використання в задачах кібербезпеки [1]. Якість вихідних послідовностей ГПВЧ безпосередньо визначає достовірність результатів усіх перерахованих процесів. Хоча в криптографії до ГПВЧ пред'являються найвищі вимоги (необоротність, стійкість до відновлення стану, перевірка за NIST SP 800-22), аналогічна якість випадковості є критичною і для не-криптографічних завдань, де погані ГПВЧ призводять до нерепродукованих результатів фаззингу чи помилкових висновків ML-моделей.

Попри широке практичне застосування, методологія оцінювання якості ГПВЧ у не-криптографічних контекстах залишається недостатньо

систематизованою. Більшість існуючих публікацій або орієнтовані виключно на криптографічне застосування генераторів, або обмежуються поверхневим описом окремих тестових пакетів. Разом із тим дослідники фіксують, що жодний ізольований метод тестування не здатний повністю верифікувати якість ГПВЧ: кожен з існуючих підходів має власні обмеження щодо виявлення прихованих статистичних закономірностей.

Основу нормативної бази тестування ГПВЧ формує декілька ключових міжнародних документів, які визначають методологічні вимоги та критерії оцінювання якості генераторів (таблиця 1).

Таблиця 1

Порівняльна характеристика основних стандартів тестування ГПВЧ

Стандарт	Область застосування	Ключові методи / вимоги
NIST SP 800-22 Rev.1a	Статистичне тестування послідовностей	15 статистичних тестів: frequency, runs, DFT, approximate entropy, cumulative sums та ін.
NIST SP 800-90B	Оцінювання джерел ентропії	Вимірювання мін-ентропії; IID- та non-IID-треки; валідація джерел випадковості
ISO/IEC 18031:2011	Вимоги до механізмів генерації	Загальна архітектура RBG; вимоги до сідування *(seeding)* та ресідування генераторів
ISO/IEC 20543:2019	Тестування в рамках ISO/IEC 15408 та 19790	Методи аналізу для сертифікації RBG у складі криптографічних модулів

Порівняльний аналіз методів тестування виявляє низку системних обмежень. По-перше, NIST SP 800-22 [2] розроблявся переважно для верифікації криптографічних генераторів, тому його критерії якості не повністю відповідають вимогам фаззингу чи тестування ML-систем. По-друге, тривалість виконання стандартизованих тестів (зокрема, процедур оцінювання ентропії за NIST SP 800-90B [3]) становить більше години навіть для одного джерела, що унеможливлює застосування в умовах ресурсно-обмежених IoT-пристроїв. Фахівці з безпеки вже вказують на доцільність оновлення керівних документів NIST для врахування вимог безпеки систем машинного навчання.

На основі аналізу можна виділити такі практичні рекомендації для кібербезпеки.

- Для QA-тестування і фаззингу достатньо базового пакета NIST SP 800-22 разом із TestU01 (мінімум SmallCrush). Для довгоперіодних генераторів варто додати спектральний аналіз та перевірку автокореляції.
- Для IDS/IPS та систем виявлення аномалій обов'язковою є оцінка мін-ентропії згідно з NIST SP 800-90B, оскільки саме вона визначає реальну непередбачуваність джерела.
- У тестуванні ML-систем слід доповнювати класичні статистичні методи ML-аналізом прихованих патернів, бо частина атак на ГПВЧ не виявляється стандартними тестами.

- Для IoT і вбудованих рішень доцільно використовувати полегшені набори тестів або спеціалізовані фреймворки з підтримкою паралельного виконання, щоб врахувати обмежені ресурси.
- Загалом найкращий результат дає комбінований підхід, який поєднує статистичне та ентропійне тестування, оскільки вони виявляють різні типи слабких місць генераторів.

Отже, генератори псевдовипадкових чисел є критичними не лише для криптографії, але й для багатьох інших задач кібербезпеки. Їх якість визначає надійність результатів у тестуванні, пентесті, виявленні аномалій і ML-системах. Чинні стандарти (NIST та ISO) задають базові підходи до оцінювання, але повну картину забезпечує лише поєднання статистичних, ентропійних і ML-методів. Подальші дослідження мають бути спрямовані на створення адаптивних тестів для ресурсно-обмежених систем, уточнення критеріїв якості ГПВЧ для конкретних застосувань кібербезпеки та інтеграцію методів машинного навчання у стандартизовані процедури оцінювання.

1. Марія Хомік, Олег Гарасимчук. Застосування генераторів псевдовипадкових чисел та послідовностей в кібербезпеці, методи їх побудови та оцінки якості // Захист інформації. – 2023. – Т. 25, № 3. – С. 147–159. DOI: https://doi.org/10.18372/2410-7840.25.17940.
2. NIST SP 800-22 Rev. 1a: A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications / Rukhin A., Soto J., Nechvatal J., et al. – Gaithersburg, MD: National Institute of Standards and Technology, 2010. – 127 p. – DOI: 10.6028/NIST.SP.800-22r1a.
3. NIST SP 800-90B: Recommendation for the Entropy Sources Used for Random Bit Generation / Bassham L. III, et al. – Gaithersburg, MD: National Institute of Standards and Technology, 2018. – 88 p. – DOI: 10.6028/NIST.SP.800-90B.

Розробка застосунку для забезпечення конфіденційності користувачів шляхом анонімізації метаданих у мультимедійних файлах

УДК 004.056.5

А.А. Гринько[1], Г.В. Шаповалов[2]

Національний університет «Одеська політехніка»
10252755@stud.op.edu.ua, shapovalov@op.edu.ua

Глобальна цифровізація та масове розповсюдження мультимедійного контенту актуалізували проблему захисту персональної інформації в інформаційному просторі. Кожен медіафайл супроводжується масивом метаданих стандартів EXIF, XMP та IPTC, які часто містять критично чутливі відомості: GPS-координати, серійні номери обладнання та часові маркери. Такі дані стають підґрунтям для деанонімізації особи та підготовки атак із використанням соціальної інженерії. Більшість загальнодоступних застосунків

проводять лише поверхневу обробку, ігноруючи глибоко вкладені теги, що створює ілюзію безпеки.

Метою роботи є розробка та програмна реалізація застосунку для ОС Windows, що забезпечує анонімізацію мультимедійних файлів шляхом модифікації заголовків без деградації якості контенту. Для досягнення мети проаналізовано специфікації медіаконтейнерів, змодельовано алгоритми деструкції метаданих EXIF, XMP та ID3, а також спроєктовано архітектуру системи з модулем логування на базі SQLite.

Архітектура застосунку базується на модульному патерні MVC, де рівень представлення на PyQt6 ізольований від логіки бінарного парсингу. Центральний контролер координує передачу дескрипторів файлів до сервісних модулів, які здійснюють пряму модифікацію заголовків медіаконтейнерів без перекодування. На низькому рівні взаємодія з файловою системою NTFS оптимізована шляхом буферизованого читання та використання атомарних операцій заміни файлів.

У результаті виконання роботи розроблено програмне забезпечення, яке дозволяє ефективно видаляти ідентифікаційні маркери без втрати якості медіафайлів. Використання асинхронної архітектури на базі Python дозволило досягти високої швидкодії при пакетній обробці даних. Реалізований підхід забезпечує надійний захист приватності в умовах масового розповсюдження контенту в мережі.

1. Information technology — Digital compression and coding of continuous-tone still images — Part 1: Requirements and guidelines : ISO/IEC 10918-1:1994 : standard. Revised 2021. URL: https://www.iso.org/standard/18902.html.
2. ExifTool by Phil Harvey : official website. URL: https://exiftool.org/.
3. Про захист персональних даних : Закон України від 01.06.2010 № 2297-VI. URL: https://zakon.rada.gov.ua/laws/show/2297-17.
4. Metadata Working Group. Guidelines For Handling Image Metadata : version 2.0. URL: https://s3.amazonaws.com/software.tagthatphoto.com/docs/mwg_guidance.pdf.

Комбінований метод захисту авторського права в зображеннях

УДК 004.056.5

Ірина Борисенко[1], Артем Грушевський[2]

Національний університет «Одеська політехніка»,
[1]borisenko.i.i@op.edu.ua, [2]10328113@stud.op.edu.ua

Традиційні методи захисту, такі як текстові метадані або видимі логотипи, виявилися малоефективними, оскільки вони легко видаляються або спотворюються зловмисниками. Це зумовлює гостру необхідність у розробці та впровадженні методів прихованого маркування на основі цифрових водяних знаків (ЦВЗ), які інтегруються безпосередньо в структуру графічного файлу на математичному рівні та демонструють високу стійкість до деструктивних впливів обробки сигналів.

Метою роботи є розробка комбінованого стеганографічного методу для захисту авторського права на цифрові зображення з застосуванням ЦВЗ, стійкого до стиснення.

В основі розробленого методу лежить комбінований підхід DWT-SVD як найбільш збалансований інструмент, що забезпечує одночасну візуальну невидимість (прозорість) вбудованого водяного знаку та його робастність (стійкість) до атак стиснення. Алгоритм вбудовування передбачає перехід зображення-контейнера в колірний простір YCbCr, де модифікації піддається лише канал яскравості (Y). До цього каналу застосовується дискретне вейвлет-перетворення на базі фільтра Хаара, що дозволяє виділити низькочастотну область апроксимації (LL), яка містить основну енергетичну складову зображення. Паралельно з цим, авторський водяний знак (логотип) піддається криптографічному скремблюванню за допомогою хаотичного перетворення Арнольда. Це перетворює впізнаваний підпис на візуальний шум, який неможливо відновити без знання секретного ключа. Математичне ядро системи реалізує модифікацію сингулярних чисел матриці LL-піддіапазону контейнера шляхом додавання до них сингулярних чисел зашифрованого знаку з певним коефіцієнтом інтенсивності α. Для реалізації «напівсліпої» схеми вилучення програма автоматично генерує та зберігає файл матричних ключів формату .npz. Це дозволяє проводити верифікацію авторства без наявності оригінального зображення, використовуючи лише захищене фото та збережені ортогональні матриці.

Експериментальні дані розробленого методу підтвердили його високу ефективність. PSNR = 41,8 дБ, що свідчить про ідеальну візуальну стійкість вбудованого ЦВЗ. Тестування на робастність продемонструвало, що водяний знак зберігає свою структуру та залишається впізнаваним після агресивного стиснення JPEG з фактором якості до 20%.

1. Мокін В. Б., Капшук О. В. Методи та засоби стеганографічного захисту інформації : монографія. Вінниця : ВНТУ, 2021. 180 с. URL: https://ir.lib.vntu.edu.ua/

Розробка безпечної системи таємного голосування

УДК 004.056.5

Володимир Гудиш [1], Валерій Трушевський [2]

Львівський національний університет імені Івана Франка, [1]volodymyr.hudysh@lnu.edu.ua, [2]valeriy.trushevsky@lnu.edu.ua

Голосування є ключовим механізмом народного волевиявлення в демократичному суспільстві. Проте сучасні виклики, зокрема пандемія COVID-19, повномасштабне збройне вторгнення Росії в Україну та вимушена міграція мільйонів громадян як всередині країни, так і за кордон, унеможливили або суттєво ускладнили участь у традиційному виборчому процесі. Ці обставини актуалізували запит на системи дистанційного електронного голосування як засіб забезпечення конституційного права на волевиявлення за будь-яких умов.

Більшість наявних рішень змушені балансувати між конкуруючими вимогами. Традиційні паперові системи не забезпечують математично доведеної верифікованості результатів, а процес підрахунку голосів залишається непрозорим для незалежних спостерігачів. Сучасна криптографія виборів демонструє фундаментальну напругу між верифікованістю результатів та захистом від примусу [6]: системи, що ставлять абсолютний пріоритет на верифікованості, як правило, надають виборцю докази, якими може скористатися зловмисник або роботодавець. Водночас повне усунення будь-якої верифікованості перетворює криптографію на "чорну скриньку", якій виборець змушений сліпо довіряти. Тому завданням є не вибір одного з полюсів, а свідоме позиціювання на цьому спектрі: досягнення достатнього рівня верифікованості при одночасному збереженні анонімності та недопущенні можливості довести конкретний вибір виборця стороннім особам.

З метою вирішення зазначених проблем розроблено безпечну систему таємного голосування, яка забезпечує анонімність виборця, верифікованість результатів та захист від широкого спектру криптографічних і мережевих атак без необхідності довіряти будь-якому окремому компоненту системи. Запропоновано та реалізовано систему з елементами підходу Zero Trust, що поєднує: гомоморфне шифрування ElGamal на еліптичній кривій SECP256R1 [1] для агрегації голосів без розшифрування; три незалежні рівні доказів з нульовим розголошенням (ZKP): диз'юнктивний OR-доказ коректності голосу, Sum-доказ одиничного вектора бюлетеня [9] та доказ коректності дешифрування Чаума–Педерсена [2]; сліпі підписи за RFC 9474 (RSABSSA-SHA384-PSS-Randomized) [3] для унеможливлення зв'язування виборця з бюлетенем; порогове дешифрування за схемою Шаміра (2-з-3) [4] із захистом ключових часток у HashiCorp Vault з шифруванням Fernet [7]; а також окремий клієнтський застосунок із вбудованим Tor для анонімної маршрутизації трафіку.

Система складається з трьох незалежних серверів та клієнтського застосунку. *Main Server* (Python/Flask) відповідає за автентифікацію виборців і ведення публічного Bulletin Board. *Validator Server* (TypeScript/Node.js, бібліотека @cloudflare/blindrsa-ts) верифікує право голосу та видає сліпий підпис, не маючи змоги асоціювати його з конкретним виборцем. *Election Agency* (Python/Flask) верифікує ZKP-докази і сліпий підпис та публікує агреговані результати. Клієнтський застосунок (Python/customtkinter) надсилає бюлетень безпосередньо до Election Agency через Tor (.onion-адреса) [8], повністю виключаючи Main Server з ланцюжка передачі голосу.

Бюлетень формується як unit vector — масив шифротекстів ElGamal, де кожна позиція містить зашифроване значення 0 або 1. Sum-доказ гарантує, що сума елементів вектора дорівнює рівно 1 [9]. Tracking Code, що є хешем SHA-384 від JSON-об'єкта бюлетеня, обчислюється на стороні клієнта і публікується на Bulletin Board: виборець у будь-який момент може переконатися у врахуванні свого голосу (Individual Verifiability). На етапі підрахунку Election Agency виконує гомоморфне додавання шифротекстів окремо для кожного кандидата, збирає частки приватного ключа з двох серверів і розшифровує агреговані суми. Для кожного кандидата публікується Decryption Proof — доказ рівності

дискретних логарифмів (Чаум–Педерсен) [2], що підтверджує коректне використання приватного ключа. Будь-який аудитор може самостійно відтворити агрегацію з Bulletin Board і верифікувати кожен доказ без знання приватного ключа (Universal Verifiability).

Усі сервери функціонують за протоколом HTTPS. Реєстрація виборців захищена підтвердженням електронної пошти та CAPTCHA. Передбачено rate limiting для захисту від DoS-атак. Схема 2-з-3 за Шаміром із зберіганням часток у HashiCorp Vault [7] унеможливлює компрометацію ключа при доступі до одного сервера.

Розроблена система реалізує ключові складові End-to-End Verifiability — Individual Verifiability та Universal Verifiability: виборець підтверджує врахування голосу через Tracking Code і Bulletin Board; математика гарантує чесність розшифрування через Chaum–Pedersen Decryption Proof [2, 5]. Застосування мережі Tor та сліпих підписів RFC 9474 [3] забезпечує анонімність виборця навіть за умови компрометації окремих вузлів інфраструктури.

1. National Institute of Standards and Technology. Digital Signature Standard (DSS). FIPS PUB 186-5. NIST, 2023. 36 p.
2. Chaum D., Pedersen T. Wallet databases with observers. CRYPTO 1992. Lecture Notes in Computer Science, vol. 740. – Springer, 1993. – P. 89–105.
3. Denis F. et al. RFC 9474: RSA Blind Signatures. – IETF, 2023. URL: https://www.rfc-editor.org/rfc/rfc9474 (дата звернення: 01.05.2026).
4. Shamir A. How to share a secret. Communications of the ACM. – 1979. – Vol. 22, №11. – P. 612–613.
5. Cortier V., Galindo D., Glondu S., Izabachène M. Election verifiability for Helios under weaker trust assumptions. ESORICS 2014. Lecture Notes in Computer Science, vol. 8713. – Springer, 2014. – P. 347–364.
6. Jafar U., Ab Aziz M.J., Shukur Z. Blockchain for Electronic Voting System — Review and Open Research Challenges. Sensors. – 2021. – Vol. 21, №17. – P. 5874.
7. Python Cryptography Library. Fernet (symmetric encryption). — URL: https://cryptography.io/en/stable/fernet/
8. The Tor Project. About Tor. — URL: https://support.torproject.org/about-tor/
9. Cramer R., Gennaro R., Schoenmakers B. A secure and optimally efficient multi-authority election scheme. Eur. Trans. Telecommun. – 1997. – Vol. 8, №5. – P. 481–490.

Модель гібридної системи виявлення вторгнень на основі криптографічних перетворень та методів штучного інтелекту

УДК 621.395.7 (043.2)

Аліна Давлетова[1]

Західноукраїнський національний університет, [1]a7davletova@gmail.com

Забезпечення безпеки у сучасних розподілених інформаційних системах потребує комплексного підходу, що виходить за межі стандартного шифрування. Традиційні засоби захисту часто не здатні ідентифікувати загрози, які не порушують цілісність даних, але створюють часові аномалії або статистичні відхилення в каналі зв'язку. Актуальною є розробка систем виявлення вторгнень (IDS), які поєднують математичний апарат криптографії з інтелектуальним аналізом ознак сеансу.

Метою роботи є розробка та дослідження моделі гібридної IDS, яка поєднує методи асиметричного шифрування, завадостійкого кодування та ансамблевого машинного навчання для комплексної оцінки безпеки сеансів зв'язку в умовах динамічної зміни параметрів передачі даних..

На відміну від відомих підходів, де криптографічні механізми та інтелектуальна детекція використовуються переважно окремо або комбінуються в межах однорідних моделей, наприклад, rule-based та machine learning-based IDS [1–3], запропонована система дозволяє враховувати як структурні ознаки порушення даних, так і часові аномалії. Такий підхід узгоджується з сучасними тенденціями розвитку гібридних IDS [4, 5], проте розширює їх функціональні можливості за рахунок інтеграції криптографічного контролю передачі даних у процес формування вектору ознак, що дозволяє одночасно враховувати повторне відтворення повідомлень та їх підміну. В основі розробки лежить багаторівнева модель (рисунок 1):

Рис.1. Узагальнена схема гібридної системи виявлення вторгнень

1) Криптографічний рівень - забезпечення конфіденційності даних шляхом застосування алгоритму асиметричного шифрування RSA, що дозволяє локалізувати потенційні втручання та аналізувати успішність дешифрування для кожного блоку.
2) Рівень контролю цілісності - реалізація завадостійкого кодування на основі коду Хеммінга над небінарними полями дозволяє не лише відновлювати поодинокі спотворення, а й формувати метрики, що слугують індикаторами активного втручання.
3) Інтелектуальний рівень – включає модуль машинного навчання (ML) на базі алгоритму Random Forest, що аналізує вектор ознак (Features), включаючи часові затримки (Input Delay) та метрики успішності декодування, а також модуль аналізу історії (AI), що здійснює ретроспективну оцінку ризику на основі бази даних попередніх сесій.
4) Рівень прийняття рішень – етап, на якому за допомогою гібридного алгоритму, що комбінує «жорсткі» правила криптографічного ядра (crypto), прогнози моделі машинного навчання Random Forest та результати аналізу історії, формується фінальне рішення.

Для оцінки ефективності запропонованого рішення проведено серії з 3000 експериментів у режимі змішаних сценаріїв, де випадковим чином генерувалися різні типи впливів, зокрема помилки передачі, часові затримки та комбіновані атаки.

Такий підхід дозволив оцінити здатність IDS працювати в умовах невизначеності та часткового перекриття ознак різних типів загроз.

Таблиця 1

Результати експериментального дослідження гібридної IDS

Компонент системи	Базова точність детекції, %	Оптимізована точність детекції, %	Усереднена F-міра	Повнота детекції атак
Crypto	78,73	81,23	0,583	0,84
AI	48,06	56,41	0,508	0,61
ML	78,86	99,47	0,992	1,00
Фінальне рішення	72,86	92,29	0,874	1,00

Результати експериментів підтвердили, що криптографічне ядро системи стабільно виявляє прямі порушення цілісності даних, тоді як накопичення статистики аномалій дозволило адаптувати систему до специфічних завад у каналі зв'язку.

Обраний вектор ознак продемонстрував високу роздільну здатність для класифікації станів системи, а гібридна інтеграція компонентів забезпечила підвищення точності фінального рішення на 19,43% при повному виключенні пропуску вторгнень (Recall = 1,00).

1. Viswanathan C., Kirthika G. Hybrid Machine Learning-Based Intrusion Detection System for Cybersecurity in Autonomous Vehicles. 2025. 1-6. https://doi.org/10.1109/ICCDS64403.2025.11209720..
2. Joshi V.R., Assa-Agyei K., Al-Hadhrami T., Qasem, S.N. Hybrid AI Intrusion Detection: Balancing Accuracy and Efficiency. Sensors. 2025. 25(24), 7564. https://doi.org/10.3390/s25247564
3. Mamatha P., Balaji S., Anuraghav S.S. Development of Hybrid Intrusion Detection System Leveraging Ensemble Stacked Feature Selectors and Learning Classifiers to Mitigate the DoS Attacks. Int J Comput Intell Syst 18, 20 (2025). https://doi.org/10.1007/s44196-025-00750-6
4. Bharti J., Singh S. A Machine Learning Based Hybrid Encryption System to Prevent Cloud Data Breach. Journal of Information Systems Engineering and Management. 2025. 10. https://doi.org/708-717. 10.52783/jisem.v10i50s.10350.
5. Ahmad Z., Khan A.S., Wai Shiang C., Abdullah J., Ahmad F. Network intrusion detection system: A systematic study of machine learning and deep learning approaches. Trans Emerg Telecommun Technol. 2021. 32(1):e4150. https://doi.org/10.1002/ett.4150

Оцінювання ризику атак соціальної інженерії в банківських установах

УДК 004.056.5 (336.71)

Дар'я Семидетнова[1], Ірина Вінковська[2], Дар'я Курінська[3]

Національний університет «Одеська політехніка», [1]9560456@stud.op.edu.ua, [2]vinkovska.i.s@op.edu.ua, [3]10252746@stud.op.edu.ua

В умовах сучасної цифровізації фінансових послуг соціальна інженерія залишається однією з найнебезпечніших загроз, оскільки спрямована не лише на технічні вразливості, а й на експлуатацію психологічних особливостей людини. Соціальна інженерія, що базується на маніпулюванні довірою, страхом або цікавістю співробітників, дозволяє зловмисникам отримувати доступ до конфіденційних даних, систем дистанційного банківського обслуговування та внутрішніх мереж. Враховуючи умови стресу та багатозадачності, в яких працюють працівники банківських установ, а також значні потенційні збитки від успішних атак, питання оцінювання рівня вразливості персоналу до методів соціальної інженерії набуває особливої актуальності [1].

Метою роботи є дослідження та оцінювання рівня ризику атак соціальної інженерії в банківських установах на основі аналізу поведінкових факторів і результатів сценарного опитування персоналу.

У ході дослідження встановлено, що найбільш ефективним підходом до оцінювання ризику є використання сценарних ситуацій, які моделюють типові випадки взаємодії співробітників із потенційними загрозами. Такий підхід дозволяє оцінити реакцію працівників на фішингові повідомлення, телефонні дзвінки, підроблені службові запити та інші методи психологічного впливу. При формуванні сценаріїв особлива увага приділялась доступності формулювань, відсутності складної технічної термінології та мінімізації психологічного тиску на респондентів.

Проведення сценарного опитування дало змогу отримати більш об'єктивні результати щодо поведінки працівників у потенційно небезпечних ситуаціях. Аналіз відповідей показав, що найбільшу складність для респондентів становлять ситуації, пов'язані з терміновими службовими запитами, повідомленнями від нібито керівництва та необхідністю швидкого прийняття рішень без додаткової перевірки інформації.

Для кількісного визначення рівня ризику було використано адитивну модель оцінювання на основі системи зважених балів. Кожна відповідь класифікується залежно від рівня потенційної небезпеки: 0 балів – безпечна поведінка; 1 бал – незначна необачність; 3 бали – критична вразливість. Загальний показник ризику розраховується за формулою:

$$R = \sum_{i=1}^{n} W_i$$

де R – загальний показник рівня ризику, n – кількість питань, W_i – вага обраної відповіді.

Отримані результати дозволили виокремити три основні рівні ризику. Низький рівень (0-5 балів) свідчить про достатню обізнаність працівника у сфері кібербезпеки та здатність розпізнавати потенційні загрози. Середній рівень (6-15 балів) вказує на наявність окремих прогалин у знаннях і певну схильність до психологічного впливу. Високий рівень ризику (16-45 балів) характеризує критичну вразливість співробітника, що може створювати загрозу для інформаційної безпеки банківської установи.

Результати аналізу підтверджують, що людський фактор залишається одним із найуразливіших елементів системи інформаційної безпеки. Навіть за умови використання сучасних технічних засобів захисту недостатній рівень підготовки персоналу може призвести до витоку конфіденційної інформації або несанкціонованого доступу до внутрішніх ресурсів банку. Саме тому регулярне оцінювання рівня готовності працівників до протидії атакам соціальної інженерії є важливою складовою забезпечення кібербезпеки [2].

Крім того, використання сценарного підходу дозволяє не лише визначити загальний рівень ризику, а й виявити найбільш проблемні аспекти поведінки персоналу. Це створює можливість для вдосконалення програм навчання та підвищення рівня обізнаності працівників щодо правил кібергігієни й безпечної роботи з інформацією. Проведене дослідження демонструє доцільність застосування методів оцінювання поведінкових ризиків як одного з напрямів зміцнення системи інформаційної безпеки банківських установ.

Отримані результати можуть бути використані як інформаційна основа для проведення внутрішнього аудиту безпеки, оцінювання ефективності навчальних заходів і формування рекомендацій щодо мінімізації ризиків соціальної інженерії.

Запропонований підхід дозволяє перевести поняття «людський фактор» у вимірювані показники, що підвищує ефективність управління інформаційною безпекою в банківській сфері.

Таким чином, оцінювання ризиків соціальної інженерії є важливим елементом сучасної системи захисту банківських установ. Комплексний аналіз поведінкових факторів працівників сприяє своєчасному виявленню потенційних загроз та підвищенню загального рівня стійкості організації до кіберінцидентів. Подальший розвиток методів оцінювання ризику дозволить удосконалити механізми підготовки персоналу та забезпечити більш ефективний захист інформаційних ресурсів у фінансовій сфері.

1. Бурячок В.Л. Толубко В.Б., Хорошко В.О., Толюпа С.В. Інформаційна та кібербезпека: соціотехнічний аспект. Львів: «Магнолія 2006». – 2018 – 320 с.
2. Ванацький Д. І. Соціальна інженерія. Вісник Київського інституту бізнесу та технологій, 2018. Вип. № 2 (36). – С. 26.
3. Сілін Є.С., Кадубовський О.А. Основи кібербезпеки: навчальний посібник. Дніпро, 2023. – 200 с.

Поетапне впровадження SOC 2 Type 2 для зберігання великих даних на підприємстві

УДК 004.056

Олег Дейнека[1], Олег Гарасимчук[2]

Національний Університет "Львівська Політехніка",
[1]oleh.r.deineka@lpnu.ua , [2] oleh.i.harasymchuk@lpnu.ua

Стрімке зростання обсягів корпоративних даних у поєднанні з підвищенням вимог до їх захисту та підзвітності ставить перед підприємствами принципово нові завдання. Дані перетворились на ключовий стратегічний ресурс, без якого неможливе ефективне функціонування державних і приватних організацій, фінансових установ, охорони здоров'я, промислових та наукових структур [1]. Разом з тим ринок комерційних рішень для управління даними пропонує переважно закриті, вартісні продукти без відкритої документації алгоритмів і процедур, що унеможливлює їх адаптацію під специфіку конкретного підприємства. Це зумовлює потребу у формалізованій, відкритій і відтворюваній методиці, яка б забезпечувала відповідність стандарту SOC 2 Type 2 і могла бути впроваджена незалежно від обраної технологічної платформи.

Стандарт SOC 2 Type 2 (Service Organization Control 2) є одним з найавторитетніших міжнародних стандартів у сфері контролю безпеки сервісних організацій. Він базується на п'яти критеріях довіри (Trust Services Criteria): безпека, доступність, цілісність обробки, конфіденційність та приватність. На відміну від Type 1, що фіксує стан засобів контролю на певний момент часу, Type 2 підтверджує їх операційну ефективність протягом аудиторського періоду – зазвичай від 6 до 12 місяців. Саме тому впровадження SOC 2 Type 2 вимагає не разового технічного налаштування, а системного, безперервного управління процесами зберігання, класифікації та доступу до даних [2,4].

Метою роботи є розроблення формалізованої методики безпечного зберігання та обробки великих обсягів даних, що забезпечує відповідність вимогам SOC 2 Type 2 та може бути застосована незалежно від технологічної платформи.

У рамках наукового дослідження розроблено методику безпечного зберігання великих обсягів даних, що відповідає вимогам SOC 2 Type 2. Методика поєднує організаційні, технічні та процедурні аспекти управління даними і ґрунтується на поетапній архітектурі обробки інформації: від збору та первинного збереження сирих даних – до їх очищення, семантичного збагачення і формування готових аналітичних структур для потреб звітності й аудиту. На кожному рівні реалізуються заходи контролю якості, класифікації та захисту інформації, а результати класифікації безпосередньо визначають застосування політик шифрування і розмежування доступу [3].

Автоматизована ідентифікація чутливих даних – персональної інформації, фінансових відомостей, клієнтських записів – здійснюється із залученням інтелектуальних інструментів аналізу тексту, що суттєво знижує вплив людського фактору на безпеку. Оркестрація інтеграційних потоків, моніторинг

виконання та реагування на порушення забезпечуються засобами автоматизації й сповіщень, тоді як управління секретами та ідентифікація користувачів реалізовані відповідно до вимог SOC 2 щодо контролю доступу та захисту облікових даних [2, 3]

Для успішного впровадження методики на підприємстві запропоновано поетапний системний підхід, що враховує специфіку бізнесу, організаційну структуру та технічні вимоги. Запропонований підхід формалізовано у вигляді послідовного алгоритму впровадження методики, що складається з наступних етапів:

1. Ідентифікація замовника та визначення стратегічних цілей. Налагодження комунікації з технічними і нетехнічними спеціалістами підприємства, збір вимог, узгодження бачення та фіксація критеріїв ефективності (включають показники продуктивності, рівня доступності, точності класифікації даних та відповідності вимогам безпеки) й очікуваних результатів.

2. Формування проєктної команди. Визначення ключових ролей – бізнес-аналітика, керівника проєкту та архітектора рішень – відповідальних за координацію, технічний дизайн і управління проєктом.

3. Планування та декомпозиція завдань. Розподіл проєкту на послідовні цілі та проведення повного аудиту даних підприємства: виявлення локальних і хмарних сховищ, класифікація типів, обсягів і форматів даних, оцінка їх критичності відповідно до вимог SOC 2 Type 2.

4. Розроблення архітектурної візії рішення. Проєктування технічної архітектури з урахуванням бізнес-процесів та наявної IT-інфраструктури: вибір технологічних платформ, принципи організації шарів зберігання і трансформації даних, інтеграційні та захисні механізми, підтримка гібридних сценаріїв і масштабованості.

5. Формалізація проєктного плану та затвердження. Підготовка детального плану з розподілом ролей, заходами контролю якості та безпеки, оцінкою ресурсів і строками; отримання офіційного затвердження від замовника.

6. Розроблення пілотного рішення. Запуск прототипу на обмеженому наборі даних для оцінки працездатності архітектури, виявлення потенційних проблем і збір емпіричних даних щодо функціонування системи.

7. Комплексне тестування. Тестування інтеграційних процесів, трансформації та обробки даних; верифікація коректності метаданих, налаштувань доступу й класифікації чутливих даних; оцінювання ефективності інструментів автоматизованого виявлення конфіденційної інформації.

8. Аналіз результатів та внесення коректив. Оцінювання продуктивності, масштабованості та відповідності системи вимогам безпеки за критеріями часу відповіді, рівня доступності (uptime/SLA), повноти обробки даних, дотримання нормативних вимог щодо класифікації, зберігання і видалення даних, а також експлуатаційної придатності системи.

9. Навчання персоналу та внутрішня документація. Підготовка IT-персоналу і кінцевих користувачів, розроблення документації щодо правил роботи з системою, визначення ролей і стандартів обробки даних.

10. Масштабування та постійна підтримка. Розширення системи на інші підрозділи або додаткові джерела даних; безперервний моніторинг, регулярний

аудит та оновлення відповідно до змін у бізнес-процесах і нормативних вимогах безпеки.

Варто зазначити, шо процес має ітераційний характер і передбачає повернення до попередніх етапів у разі виявлення невідповідностей або змін у вимогах

Галузева специфіка підприємства суттєво визначає пріоритизацію критеріїв Trust Services Criteria. Для фінансового сектору на перший план виходять конфіденційність і цілісність транзакційних журналів; для закладів охорони здоров'я – суворі вимоги до приватності пацієнтських даних; для телекомунікаційних компаній – безперервна доступність сервісів та стійкість до відмов; для виробничих підприємств з Industrial IoT – цілісність потокових даних промислових сенсорів.

Запропонована методика передбачає адаптивний вибір пріоритетних засобів контролю на основі профілю ризиків конкретної галузі, що відрізняє її від універсальних «коробкових» рішень [1, 3].

Впровадження запропонованої методики дозволяє підприємствам різних галузей структуровано підготуватись до сертифікаційного аудиту за стандартом SOC 2 Type 2, сформувати прозору та відтворювану систему управління великими обсягами корпоративних даних, мінімізувати ризики несанкціонованого доступу та витоку інформації.

Кожен етап методики корелює з відповідними критеріями Trust Services Criteria та забезпечує їх поступову імплементацію на організаційному і технічному рівнях.

На відміну від закритих комерційних продуктів, методика є відкритою, адаптованою до різних технологічних середовищ і розрахованою на практичне впровадження силами внутрішньої команди підприємства без залежності від конкретного постачальника хмарних сервісів.

Адаптація реалізується шляхом вибору релевантного набору контролів на основі оцінки ризиків (risk-based approach).

1. Дейнека О. Р., Гарасимчук О. І. Виклики та стратегії зберігання великих обсягів даних у сучасному світі. Захист інформації. 2024. Т. 25, № 4. С. 197–207.
2. Дейнека О. Р., Бортнік Л. Л. Методологія збору, обробки, зберігання та класифікації даних відповідно до вимог SOC 2 Type 2. Комп'ютерні системи та мережі. 2024. Т. 6, № 1. С. 36–43. https://doi.org/10.23939/csn2024.01.036
3. Дейнека О. Р., Гарасимчук О. І. Дослідження проблем класифікації та безпечного зберігання даних. Безпека інформації. 2023. Т. 29, № 2. С. 147–153.
4. AICPA. Trust Services Criteria for Security, Availability, Processing Integrity, Confidentiality, and Privacy (SOC 2). American Institute of CPAs, 2022. 112 p

Модель інтегрального оцінювання рівня кіберзахищеності корпоративних мереж

УДК 004.056.5:004.7

Денис Трухан[1]

Державний університет інформаційно-комунікаційних технологій,
[1]d.truhan@stud.duikt.edu.ua

Зростання кількості кіберінцидентів у корпоративних інформаційних системах актуалізує потребу у кількісних методах оцінювання кіберзахищеності. На практиці адміністратор безпеки часто має справу не з окремою вразливістю або окремою загрозою, а з множиною взаємопов'язаних факторів: одна вразливість може використовуватися кількома типами атак, а один захисний механізм може одночасно знижувати ризик для групи вразливостей. Тому ізольоване оцінювання загроз, вразливостей і захисних заходів призводить до завищення або заниження реального рівня ризику [2].

Метою роботи є стислий виклад моделі інтегрального оцінювання рівня кіберзахищеності корпоративних мереж, яка враховує взаємне перекриття загроз, вразливостей та захисних механізмів. На відміну від підходів, що ґрунтуються лише на усередненні CVSS-оцінок або якісних матрицях ризику, запропонований підхід розглядає елементи кіберзахисту як єдину тріаду «загрози – вразливості – захист».

Корпоративну мережу доцільно подати у вигляді кортежу $N = < A, T, V, P, C, M, W >$, де A – множина активів; T – множина кіберзагроз; V – множина вразливостей; P – множина захисних механізмів; C – матриця покриття, що відображає можливість реалізації загрози через конкретну вразливість; M – матриця нейтралізації, що відображає вплив захисних механізмів на вразливості; W – система вагових коефіцієнтів. Таке подання дозволяє перейти від описового аналізу до формалізованого обчислення інтегрального показника.

Для визначення взаємного перекриття між елементами тріади використовується ідея подібності множин. Зокрема, перекриття двох загроз можна оцінити через частку спільних вразливостей, через які вони реалізуються. Якщо дві загрози використовують однакові або близькі набори вразливостей, їхній внесок у сукупний ризик не повинен дублюватися повністю. Аналогічно оцінюється перекриття вразливостей та захисних механізмів.

$$\Omega_T(t_i, t_j) = \frac{|V(t_i) \cap V(t_j)|}{|V(t_i) \cup V(t_j)|}, \qquad (1)$$

де $\Omega_T(t_i, t_j)$ – коефіцієнт перекриття двох загроз; $V(t_i)$ та $V(t_j)$ – множини вразливостей, через які можуть бути реалізовані відповідні загрози. Значення коефіцієнта належить інтервалу $[0; 1]$: 0 означає відсутність спільних вразливостей, а 1 – повний збіг множин.

На основі матриць C і M та коефіцієнтів перекриття формується інтегральний показник кіберзахищеності $I(N)$. Він поєднує критичність активів, залишкову уразливість, ефективність захисних механізмів і ступінь дублювання ризику. У практичній інтерпретації $I(N)$ набуває значень від 0 до

1, де значення, близьке до 1, відповідає високому рівню захищеності, а значення, близьке до 0, – критичному стану системи.

Процедура обчислення показника складається з п'яти етапів: інвентаризації активів і побудови матриць взаємозв'язків; обчислення коефіцієнтів перекриття; призначення вагових коефіцієнтів; визначення часткових метрик для активів; розрахунку інтегрального показника та його інтерпретації за шкалою рівнів захищеності. Перевагою такої процедури є можливість повторного перерахунку після появи нових вразливостей або впровадження додаткових засобів захисту.

Таблиця 1

Порівняння підходів до оцінювання кіберзахищеності

Метод	**Оцінка** / $I(N)$	**Достовірність, %**	**Облік перекриття**
Пентест (еталон)	0,71	100	-
Запропонована модель	0,68	95,8	повний
CVSS v3.1	0,54	73,9	відсутній
NIST SP 800-30	0,58	81,7	частковий
ISO/IEC 27005	якісна оцінка	64,8	відсутній

Як видно з табл. 1, у тестовому сценарії запропонована модель наблизилася до еталонної оцінки, отриманої за результатами пенетраційної перевірки, і продемонструвала вищу достовірність порівняно з підходами, які не враховують структурні взаємозв'язки між загрозами, вразливостями та засобами захисту. Основна причина покращення полягає в усуненні подвійного зарахування взаємопов'язаних ризиків.

Практичне значення моделі полягає у можливості використовувати її як інструмент підтримки прийняття рішень під час аудиту корпоративної мережі. За допомогою інтегрального показника можна не лише визначати поточний рівень кіберзахищеності, а й порівнювати різні варіанти впровадження захисних механізмів, ранжувати їх за очікуваним впливом і прогнозувати залишковий ризик.

Отже, запропонований підхід дозволяє розглядати кіберзахищеність корпоративної мережі як системну характеристику, що залежить від взаємодії загроз, вразливостей і захисних механізмів. Подальші дослідження доцільно спрямувати на автоматизацію заповнення матриць C і M на основі даних SIEM, CVE/NVD та результатів сканування інфраструктури, а також на адаптацію моделі для хмарних і гібридних середовищ.

1. NIST Special Publication 800-30 Revision 1. Guide for Conducting Risk Assessments. Gaithersburg: National Institute of Standards and Technology, 2012. 95 p.
2. CVSS v3.1 Specification Document. Forum of Incident Response and Security Teams (FIRST). 2019. URL: https://www.first.org/cvss/v3.1/specification-document (дата звернення: 10.01.2024).

3. ISO/IEC 27005:2018. Information technology - Security techniques - Information security risk management. Geneva: International Organization for Standardization, 2018. 87 p.
4. MITRE ATT&CK: Design and Philosophy. MITRE Corporation Technical Report. Bedford, MA, 2020. 57 p.

Система нечіткого логічного виводу вразливостей та загроз інформаційної безпеки

УДК 004.056

Володимир Джулій[1], Денис Вишневський[2]

Хмельницький національний університет,
[1]dzhuliivm@khmnu.edu.ua, [2]vushnya5495@gmail.com

Аналіз проведеного дослідження поточного стану в області інформаційної безпеки показує, що темпи розвитку інформаційних та комп'ютерних технологій значно випереджають процес створення програмно-апаратного забезпечення в області інформаційної безпеки. Пріоритетними, в даній ситуації, є задача аналізу, класифікації, виявлення діючих механізмів та засобів проведення атак і загроз інформаційній безпеці системи, які можуть призвести до отримання несанкціонованого доступу до конфіденційних даних, порушення функціонування інформаційної системи, визначення заходів протидії атакам та загрозам, оцінка заданої шкоди, розробка нормативно-правової бази, механізмів захисту та критеріїв інформаційної безпеки системи протидії.

Актуальною залишається задача проектування та розробки методу, системи прогнозування, виявлення вразливостей, загроз безпеки інформації. Як інструмент для досягнення поставленої задачі пропонується використовувати нечітку інформаційно-аналітичну систему прогнозування вразливостей та загроз інформаційної безпеки, що надає функціональні можливості, які представленні на рис. 1.

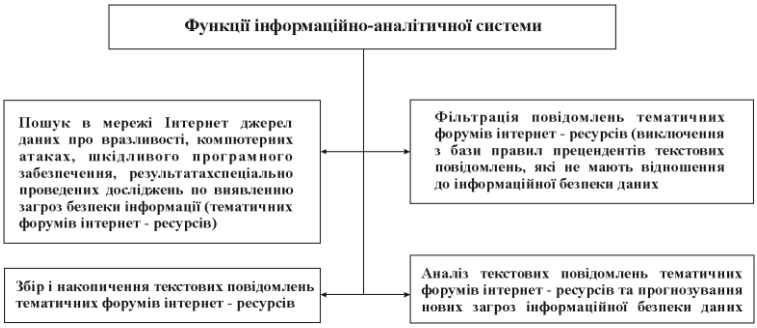


Рис.1. Функції інформаційно-аналітичної системи аналізу потоку повідомлень тематичних інтернет-форумів

Список тематичних джерел форумів містить адреси інтернет- ресурсів, на яких розміщуються текстові публікації про шкідливе програмне забезпечення, вразливості та комп'ютерні атаки [1].

На початковому етапі роботи інформаційно-аналітичної системи список формується експертним шляхом, із загальної кількості форумів тематичних інтернет – ресурсів виділяються ті, тематика яких дозволяє інтернет - інформацію віднести до хакерських (інформація містить результати спеціалізованих досліджень з виявлення вразливостей та загроз інформаційної безпеки конфіденційних даних, повідомлення про комп'ютерні атаки, вразливості, шкідливе програмне забезпечення). Автоматизоване виявлення нових форумів тематичних інтернет-ресурсів, в даній ситуації, можливо, шляхом проведення аналізу різноманітних форумів інтернет-ресурсів, з використанням запропонованих критеріїв відбору текстових повідомлень, що належать до заданої предметної області, для якої проводиться аналіз з використання онтології [2]. Для реалізації фізичної системи, потрібно реалізувати в матеріальні сутності всі елементи логічного представлення.

Для фізичного представлення моделі використовується діаграма UML розгортання, відображається загальна топологія та конфігурація інформаційно-обчислювальної системи, а також розподіл за окремими вузлами компонентів. Вузлами діаграми інформаційно-обчислювальної системи є персональний комп'ютер користувача, сервер адміністратора.

Покращення якості прогнозування виникнення вразливостей та загроз безпеки інформації з використанням систем логічного нечіткого виводу може сприяти збільшення кількості вхідних змінних, використанню більш точних нечітких правил продукцій, також велике значення має визначення функцій приналежності вихідних та вхідних параметрів системи логічного нечіткого виводу, необхідно враховувати статистичні показники потоку текстових повідомлень форумів тематичних інтернет-ресурсів.

1. Ленков С.В**.** Метод прогнозування вразливостей інформаційної безпеки на основі аналізу даних тематичних інтернет-ресурсів / С.В. Лєнков, В.М. Джулій, А.М. Берназ, І.В. Муляр, І.В. Пампуха // Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка. – К.: ВІКНУ, 2023. – Вип. №78. – С. 123-134.
2. Ленков С.В. Метод протидії поширенню та виявлення шкідливої інформації в соціальних мережах/ С.В. Ленков, В.М. Джулій, Л.В. Солодєєва // Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка. – К.: ВІКНУ, 2022. – Вип. №77. – С. 103-117.

Структура мережі розподілу квантово-захищених ключів у мережах магістральної топології

УДК 004.056

Володимир Джулій[1], Максим Вовкович[2]

Хмельницький національний університет, [1]dzhuliivm@khmnu.edu.ua, [2]maksym.vovkovyhc@gmail.com

На основі запропонованих вимог до структури мережі та методу розподілу квантово-захищених ключів у мережах магістральної топології сформована архітектура квантової мережі квантового розподілення ключів змішаної топології.

Квантова мережа квантового розподілу ключів будується на основі довірених вузлів мережі, до кожного з вузлів можуть підключатися системи захисту інформації-користувачі, зовнішні пристрої. Кожен вузол квантової мережі щонайменше містить один напівкомплект квантових пристроїв і з'єднаний щонайменше квантовим каналом зв'язку не менш ніж із одним сусіднім вузлом квантової мережі [1].

Кожен вузол квантової мережі квантового розподілу ключів містить один і більше модулів генерації квантових ключів мережі, реалізованих квантовими пристроями. На даний момент через відсутність стандартизованого протоколу квантового розподілу ключів між двома кінцях квантового каналу мережі повинні розташовуватися пристрої, реалізовані одним виробником і виконувати один протокол квантового розподілу ключів. На кожному сегменті квантової мережі квантового розподілу ключів можливе застосування різних квантових пристроїв.

Максимальна довжина квантового каналу квантової мережі кожного сегмента визначається граничними значеннями втрат на обраному сегменті протоколу квантового розподілу ключів квантового каналу.

Для початку роботи квантової мережі розподілу ключів, початку генерації квантових ключів для розподілу квантово-захищених ключів необхідно розподілити попередньо відповідні набори ключів: на пари сусідніх вузлів квантової мережі; для аутентифікації квантових пристроїв класичного каналу.

Розподіл квантово-захищених ключів між парами мережі вузлів квантової мережі реалізується відповідно до запропонованого методу. Обчислення магістральної підмережі квантової мережі проводиться вузлом, на який надійшов запит квантово-захищеного ключа з урахуванням топології мережі квантового розподілу ключів [2].

Для визначення пар мережі цільових вузлів квантової мережі, на які розподіляються квантово-захищені ключі відповідно до запиту системою захисту інформації - користувача, необхідно підтримувати базу відповідності підключених систем захисту інформації-користувачів і вузлів квантової мережі в актуальному стані для всієї квантової мережі. Ідентифікація мережі вузлів квантової мережі повинна бути унікальною для визначення, однозначно, цільових вузлів квантової мережі. Мережа протоколу квантового розподілу ключів може будуватися підключенням нових вузлів квантової мережі ітераційним шляхом до існуючої квантової мережі протоколу квантового розподілу ключів. Для підключення до мережі нового вузла квантової мережі необхідно:

1. Під’єднати вузол квантової мережі до квантового каналу мережі. Квантовий канал під’єднаний може бути до до вузла квантової мережі, до комутатора, до модуля генерації квантових ключів мережі.

2. Завантажити в новий вузол квантової мережі, попередньо згенеровані розподілені ключі для побудови в мережі автентифікованого каналу квантових

пристроїв у складі пари мережі вузлів квантової мережі. Попередньо розподілені ключі, створюються з використанням датчика випадкових чисел, з вузлів квантової мережі, з доставкою довіреним кур'єром до нового вузла квантової мережі. Після проведення успішного сеансу квантового розподілення ключів з новим вузлом квантової мережі, даний вузол вважається підключеним до квантової мережі протоколу квантового розподілення ключів.

3. Класичні квантові ключі захисту квантово-захищених ключів для проведення розподілу квантово-захищених ключів для систем захисту інформації-користувачів можуть бути використані квантово-захищені ключі, розподілені із захистом на квантових ключах по квантовій мережі протоколу квантового розподілення ключів між новим вузлом квантової мережі та необхідними цільовими вузлами квантової мережі.

Запропоновано архітектуру квантової мережі квантового розподілення ключів змішаної топології, та наведені вимоги до функцій квантової мережі, отримані на основі запропонованих підходів розподілу квантово-захищених ключів та виявлених, при цьому, особливостей квантових пристроїв. Така квантова мережа надає можливість розподіляти квантово-захищені ключі, передавати згенеровані ключі у пари систем захисту інформації-користувачів, на основі запитаного системами захисту інформації-користувачами забезпечення ключами, вирішуючи задачу зміни та доставки ключів шифрування.

1. Квантова криптографія. Пояснення [Електронний ресурс] // Quantum Xchange. – 2019. – Режим доступу: https://quantumxc.com/blog/quantum-cryptography-explained/ (дата звернення 02.03.2024).
2. Satish Kumar. Quantum Cryptography [Електронний ресурс] // Tutorialspoint. – 2023. – Режим доступу: http://surl.li/fjebs (дата звернення 05.03.2024).

A study of methods for detecting hidden threats in multimedia objects on web resources

UDK 004.056.5:004.932

Dmytro Denysiuk[1], Bohdan Savenko[2]

Khmelnytskyi National University,
[1]denysiuk@khmnu.edu.ua, [2]savenko_bohdan@ukr.net

Multimedia objects on web resources are not only interface elements or user-generated content, but also potential carriers of hidden cyber threats. Mechanisms for loading images, video files, avatars, banners, and graphic containers create a separate attack surface that often falls outside the scope of traditional web request scanning. Hidden data can be embedded in the pixel area, frequency coefficients, EXIF/XMP metadata, service segments, overlay areas, or polyglot files [1, 2].

The problem is that such modifications do not always alter the appearance of the multimedia object and may not violate the format’s basic characteristics. Therefore, checking only the file extension, MIME type, signature, or antivirus database does not

provide sufficient reliability in detection. For web resources, this creates a risk of bypassing download filters, covertly transmitting service markers, storing script fragments, or preparing multi-stage attacks [1, 3].

The main groups of features characterizing hidden cyber threats in multimedia objects of web resources have been identified, and the feasibility of their comprehensive use within the framework of multimodal analysis has been determined. Formally, a multimedia object is presented as a multi-level information container:

$$M = \{P, F, S, Meta, B\}, \quad (1)$$

where M is a multimedia object of a web resource; P is a pixel matrix or a sequence of frames; F is a set of frequency features; S are structural elements of the file container; $Meta$ is EXIF/XMP metadata; B are behavioral features recorded during the opening, decoding, previewing, or transformation of the object.

Model (1) provides a formalized description of a multimedia object as a container comprising heterogeneous groups of features. Component P reflects changes in the pixel domain, particularly those characteristic of LSB steganography; F describes anomalies in the frequency domain, specifically in DCT or DWT coefficients; S characterizes the integrity of the file structure, the consistency of signatures, MIME type, and actual content; Meta covers hidden or atypical service fields; B reflects behavioral manifestations that arise during the processing of the object in a controlled environment.

To detect hidden threats, it is advisable to use statistical, structural, frequency-based, neural network, and behavioral methods. Statistical analysis is used to study entropy, histograms, correlations between pixels, and noise residuals. It is effective for detecting some steganographic changes, but depends on the compression method, image quality, and type of hiding [3].

Structural analysis makes it possible to verify whether a container conforms to the declared format, as well as to detect redundant data, duplicate signatures, anomalous service blocks, suspicious metadata, and polyglot structures [2]. Frequency analysis is used to detect changes in DCT and DWT representations, which is relevant for JPEG images and other lossy compression formats. Neural network methods enable the automatic extraction of complex latent features that are difficult to formalize manually.

Behavioral analysis complements static analysis by allowing us to assess how the software environment reacts to a suspicious object. In a sandbox, honeypot, or decoy environment, it is possible to capture events that do not occur during a standard file analysis but arise when the file is opened, decoded, previewed, or transformed.

For practical application in web security, it is advisable to develop a comprehensive risk assessment:

$$R(M) = w_1 A_{stat} + w_2 A_{struct} + w_3 A_{freq} + w_4 A_{nn} + w_5 A_{beh} \quad (2)$$

where $R(M)$ - the overall risk of a multimedia object; A_{stat} - statistical anomalies; A_{struct}- structural abnormalities; A_{freq}- frequency deviations; A_{nn}- evaluation of a neural network detector; A_{beh}- behavioral abnormalities; w_1, w_2, w_3, w_4, w_5 - weighting factors for the respective groups of characteristics.

A multimodal model makes it possible to assess a multimedia object as a potential carrier of a cyber threat by taking into account a set of heterogeneous features, including pixel-based, frequency-based, structural, metadata-related, and behavioral characteristics. Such representation allows the object to be analyzed not only as an image, video, or graphic file, but also as a complex information container in which hidden modifications may appear at different levels.

To protect web resources, it is advisable to combine statistical, structural, frequency-based, neural network, and behavioral methods within an integrated risk assessment framework. This combination increases the reliability of detecting hidden modifications in multimedia content, since each group of methods covers a specific class of indicators and compensates for the limitations of the others.

This approach reduces the detection system's reliance on a single type of indicator and enables it to respond more effectively to various methods of concealing threats. As a result, the assessment of multimedia content becomes more comprehensive and better adapted to the detection of hidden cyber threats in web environments.

1. Almehmadi L., Basuhail A., Alghazzawi D., Rabie O. Framework for Malware Triggering Using Steganography. Applied Sciences. – 2022. – Vol. 12, No. 16. – Article 8176. DOI: https://doi.org/10.3390/app12168176
2. Koch L., Oesch S., Chaulagain A., Adkisson M., Erwin S., Weber B. Toward the Detection of Polyglot Files. Proceedings of CSET 2022 – 15th Workshop on Cyber Security Experimentation and Test. – 2022. – P. 120–128. DOI: https://doi.org/10.1145/3546096.3546106
3. Denysiuk D., Savenko O., Lysenko S., Savenko B., Kashtalian A. Method for Detecting Steganographic Changes in Images Using Machine Learning. Proceedings of the 2023 IEEE 13th International Conference on Dependable Systems, Services and Technologies (DESSERT). – Athens, Greece, 2023. – P. 1–6. DOI: https://doi.org/10.1109/DESSERT61349.2023.10416453

Optimizing uav routes under conditions of restricted access to confidential objects

UDC 623.746.4-519:004

Юлія Ткач[1], Ігор Дюба[2]

Національний університет "Чернігівська політехніка",
[1]tkachym79@gmail.com, [2] idyuba@gmail.com

Introduction. Modern scenarios for the use of Unmanned Aerial Vehicles (UAVs) require solving complex routing problems in dynamic environments. This issue becomes particularly acute when planning flights over territories containing restricted access objects or critical infrastructure. Traditional approaches based on the complete bypass of "No-Fly Zones" lead to significant time and energy expenditures.

Problem Statement. This study considers the applied task of constructing a UAV route from point A to point C through a transit area B. Area B contains objects whose coordinates are confidential. It is necessary to minimize the travel time while ensuring

the non-disclosure of information regarding these objects by maintaining a safe distance D. The primary contradiction lies in the need for route optimization without providing the planning algorithm with the precise coordinates of the critical objects.

Research Methodology. A dual-level data access model was used to solve this problem. Unlike the classic model, which addresses security issues by bypassing the entire perimeter of area B, the proposed approach allows movement directly through area B using adaptive trajectory analysis.

The non-disclosure criterion is defined as preventing the UAV from approaching critical points closer than distance D. Furthermore, the pathfinding algorithm operates under conditions of informational uncertainty regarding the exact location of the objects, which satisfies confidentiality requirements.

To formalize the problem of safe UAV passage through area B, we introduce the following notations:

$P(t) = \big(x(t), y(t)\big)$ state vector (coordinates) of the UAV at a given moment in time t;

$Oi = (x_i, y_i)$ coordinates of the i - th critical object in area B, where $i\epsilon\{x_i, y_i\}$;

D - minimally permissible distance (radius of the non-disclosure zone).

The criterion of non-disclosure of information about objects within the framework of the two-level access model is defined as a system of restrictions on the trajectory of movement:

$$\forall t \in \left[T_{start}, T_{end}\right], \quad \forall i: \sqrt{(x(t) - x_i)^2 + (y(t) - y_i)^2} \geq D$$

At the same time, according to the confidentiality terms, the exact values of O_i are not transmitted to the global planning system. Instead, a 'contact function' is used, which is activated only when approaching the boundary D.

The route construction algorithm is based on iterative approximation to the target point C with dynamic adjustment of the movement vector when entering a restricted access zone.

Stages of algorithm implementation:

1. Global planning: Construction of a straight trajectory A → C. Determination of entry and exit points for region B.

2. Local monitoring: When located in region B, the system continuously analyzes the gradient of the "disclosure threat field."

3. Two-level verification:

- Level 1 (Zone access): Permission to stay in region B to shorten the path.
- Level 2 (Object access): Prohibition on entering radius D.

4. Adaptive correction: If the current trajectory leads to the violation of the condition |P(t) - Oi| < D, the algorithm calculates the tangent to the circle of radius D and changes course H with minimal deviation from the target.

Experimental Part. A series of 10,000 numerical experiments were conducted. Modeling conditions included:

Four objects were generated in area B following a uniform distribution.

Two types of routes were constructed: a guaranteed bypass (maximum) and an optimized route through area B with accuracy H.

Forbidden and restricted access zones were defined for the objects in area B.

Analysis of Results. Statistical data processing demonstrated that as the number of iterations increases, the distribution of results converges to a normal distribution.

The findings indicate that utilizing the dual-level model allows for an average flight time gain of 19% compared to the maximum bypass trajectory. Such an indicator is critically important for the operational planning of UAV missions, particularly in electronic warfare, reconnaissance, and monitoring tasks under time-sensitive conditions.

Conclusions.

1. The application of an adaptive trajectory model instead of rigid "No-Fly Zones" significantly increases the efficiency of UAV utilization.

2. The dual-level access model ensures reliable protection of confidential information regarding critical objects even during the physical transit of the vehicle through their location zone.

3. The results confirm the algorithm's stability under the random distribution of objects, allowing for its recommended integration into the intelligent control systems of autonomous unmanned complexes.

1. Zhou, Y., Ma, L., & Wen, M. (2015). Task-Constrained RBAC Model and Its Privilege Redundancy Analysis. 2nd ISCE, pp. 489–492.
2. Ren, M., et al. (2024). Conception of Foreign Heterogeneous EW UAV Cross Domain Cooperative Operations. ICAUS 2023, vol 1171.
3. Zhang, Y., et al. (2023). Modeling and simulation of UAVs swarm electromagnetic operation. Systems Engineering and Electronics, Vol. 45(7).

Кібербезпека систем розпізнавання мовлення в реальному часі

УДК 004.056

Олег Єгоров[1], Тарас Кравченко[2]

Український державний університет науки і технологій,
[1]egoroffoleg@ukr.net, [2]t.o.kravchenko@ust.edu.ua

Розвиток технологій синтезу мовлення та deepfake-голосу створює нові кіберзагрози для систем розпізнавання мовлення. Синтетичний голос може використовуватися для підміни особи, обходу голосової автентифікації і несанкціонованого доступу до інформаційних ресурсів. За таких умов традиційні підходи до розпізнавання мовлення, орієнтовані переважно на точність і швидкодію, не забезпечують достатнього рівня захисту даних.

Актуальність дослідження зумовлена поширенням голосових сервісів і одночасним зростанням доступності засобів генерації синтетичного мовлення. Для систем реального часу особливо важливим є своєчасне виявлення deepfake-голосу без істотного збільшення затримки обробки, оскільки від цього залежить безпека доступу до даних і надійність роботи мовних сервісів.

Метою роботи є підвищення рівня захисту даних у системах розпізнавання мовлення реального часу шляхом виявлення deepfake-голосу та синтетичного мовлення на основі інтелектуального аналізу аудіопотоку.

Ефективне виявлення deepfake-голосу та синтетичного мовлення доцільно розглядати як окремий функціональний модуль, інтегрований у загальну систему розпізнавання мовлення. Такий модуль має працювати паралельно з основними етапами обробки аудіосигналу та виконувати попередню оцінку його автентичності ще до завершення процедури розпізнавання. Це дає змогу не лише підвищити надійність інтерпретації мовних команд, а й своєчасно запобігти обробці потенційно небезпечного вхідного сигналу.

Для виявлення синтетичного мовлення можуть використовуватися ознаки різної природи: спектральні характеристики сигналу, особливості часової структури, неприродна стабільність тембру, атипові переходи між фонемами, а також відхилення в ритміко-інтонаційній організації мовлення. У поєднанні з інтелектуальними методами аналізу такі параметри дозволяють формувати оцінку ймовірності штучного походження голосу. У межах захищеної архітектури системи розпізнавання мовлення результати такого аналізу можуть використовуватися для адаптивного керування подальшими діями системи. При підвищеному рівні ризику доцільно ініціювати додаткову перевірку користувача, блокувати доступ до окремих функцій або переводити обробку в режим підвищеного контролю. Цей підхід дозволяє поєднати вимоги до швидкодії з вимогами до безпеки без погіршення користувацького досвіду.

Отже, виявлення deepfake-голосу та синтетичного мовлення є важливою складовою підвищення кіберзахищеності систем розпізнавання мовлення в реальному часі. Інтеграція інтелектуальних механізмів аналізу аудіосигналу в архітектуру таких систем дозволяє своєчасно виявляти ознаки підробленого голосу, зменшувати ризик несанкціонованого доступу та підвищувати рівень захисту даних. Практичне значення цього підходу полягає у можливості створення адаптивних мовних сервісів, у яких забезпечуються не лише точність і швидкодія, а й стійкість до актуальних кіберзагроз.

Аналіз та виявлення аномалій у мережевому трафіку з використанням SLIPS

УДК 004.056.53 (004.75)

Анатолій Жуков[1], Сергій Чернишук[2]

Житомирський військовий інститут імені С.П. Корольова,
[1]anzhukov@ukr.net, [2]sergiy.chernyshuk@ukr.net

В умовах еволюції цілеспрямованих атак (APT) та поширення складного шкідливого програмного забезпечення, традиційні методи захисту, що базуються виключно на сигнатурах, стають недостатніми. За даними звіту IBM Cost of a Data Breach 2025, середня вартість витоку даних у світі сягнула рекордних показників, що підкреслює необхідність впровадження про активних засобів захисту [1]. Одним із найбільш перспективних інструментів для вирішення цієї проблеми є поведінковий аналіз трафіку, реалізований у відкритих проектах, таких як Stratosphere Linux IPS (SLIPS) [4].

Stratosphere Linux IPS - це модульна система виявлення вторгнень IDS/IPS, що базується на аналізі поведінки мережевих потоків. На відміну від класичних IDS, SLIPS орієнтована на виявлення патернів шкідливої активності, таких як

робота ботнетів та командно-контрольних каналів [2]. Система використовує наступні парадигми аналізу:

1. Профілювання тривалих з'єднань та виявлення періодичних сигналів дозволяє системі Slips ідентифікувати приховані канали керування C&C, оскільки вона фокусується не на аналізі окремих пакетів, а на тривалих поведінкових патернах мережевих пристроїв. Використовуючи потужності фреймворку Zeek, Slips відстежує часові інтервали активності та накопичує докази шкідливої поведінки, що дає змогу виявляти шкідливе ПЗ, яке підтримує зв'язок із сервером зловмисника через характерні низькочастотні «маяки».

2. Аналіз за допомогою нейронних мереж та машинного навчання у реальному часі є фундаментом Slips як першої вільної системи IDS/IPS, що базується на методах машинного навчання для детекції шкідливих дій [3]. Система застосовує адаптивні моделі для класифікації трафіку, зокрема аномалій у HTTPS-з'єднаннях, із підтримкою механізмів обробки дрейфу даних, а через субмодуль реалізує концепцію федеративного навчання для підвищення точності прогнозів без передачі конфіденційних даних

3. Інтеграція з Threat Intelligence та автоматизована перевірка загроз забезпечується шляхом постійного оновлення даних із понад 40 спеціалізованих джерел розвідки та можливості перевірки IP-адрес на зовнішніх платформах, таких як VirusTotal або RiskIQ. Особливої ефективності цьому підходу додає наявність P2P-модуля, який дозволяє вузлам Slips автоматично та безпечно обмінюватися індикаторами компрометації з іншими довіреними вузлами в мережі, формуючи спільну систему протидії кіберзагрозам [1].

Система поєднує декілька підходів, що дозволяє досягти високої точності при мінімізації помилкових спрацювань. Гібридні моделі, що поєднують глибоке навчання DL та класичне машинне навчання ML, демонструють найкращі результати в сучасних мережевих середовищах [4].

Порівняльну характеристику модулів SLIPS представлено у таблиці 1.

Таблиця 1

Характеристика модулів аналізу Stratosphere Linux IPS

Модуль аналізу	Тип виявлення	Перевага	Складність
Behavioral Module	Поведінковий	Виявляє невідомі C&C канали	Середня
Machine Learning	Алгоритмічний	Адаптація до нових типів трафіку	Висока
TI Module	Репутаційний	Миттєве виявлення відомих загроз	Низька
Ensemble Logic	Гібридний	Найнижчий рівень false positives	Висока

Ключовою особливістю SLIPS є використання алгоритмів машинного навчання для аналізу мережевих подій. Система перетворює мережеві потоки у текстове представлення, що дозволяє застосовувати методи обробки природної мови для пошуку аномалій. Для IoT-мереж, інтегрованих у Linux-інфраструктуру, особливо ефективним є поєднання методів Deep Learning, що забезпечує точну класифікацію підозрілих патернів. Це дозволяє виявляти активність шкідливого ПЗ навіть у зашифрованому трафіку без його дешифрації.

Сучасний стек технологій для аналізу трафіку за допомогою SLIPS включає інструменти захоплення Zeek і Suricata та обробки даних у реальному часі. Використання алгоритмів типу Isolation Forest підтвердило свою ефективність для швидкого виявлення аномалій у вебтрафіку. Платформа SLIPS використовує Redis для швидкого обміну даними між модулями та пропонує інструменти візуалізації інцидентів, такі як Kalipso.

Особливе значення має відкритість датасетів, на яких тренується SLIPS, наприклад, Stratosphere IoT Dataset, що дозволяє дослідникам адаптувати систему під специфічні потреби Linux-серверів або IoT-інфраструктур.

Stratosphere Linux IPS представляє собою сучасний приклад інтелектуальної системи захисту, яка успішно поєднує статистичний аналіз з методами глибокого навчання. Застосування SLIPS дозволяє значно підвищити рівень виявлення складних кіберзагроз за рахунок аналізу довгострокової поведінки вузлів мережі. Перспективами подальшого розвитку є вдосконалення модулів аналізу зашифрованого трафіку та автоматизація реагування на інциденти IPS mode в хмарних середовищах.

1. IBM Security. Cost of a Data Breach Report 2025. IBM Corporation, 2025.
2. Alsoufi M. A., et al. Anomaly-based intrusion detection model using deep learning for IoT networks. Computer Modeling in Engineering & Sciences. – 2024. – Vol. 141, №1. – P. 823–845.
3. Chua W., et al. Web traffic anomaly detection using Isolation Forest. Informatics. – 2024. – Vol. 11, №4. – P. 83.
4. Garcia S. Stratosphere Linux IPS: Behavioral-based Intrusion Detection System. Stratosphere Laboratory, 2025. URL: https://github.com/stratosphereips/StratosphereLinuxIPS

Cybersecurity for small and medium-sized businesses: a practical framework for organizations with limited resources

UDK 004.056.53

Iurii Zhurov

SMB Cybersecurity Advisors LLC, iurii.zhurov@smbsecurityadvisors.com

Over the past decade, the number of cyber incidents targeting small and medium-sized businesses (SMBs) has grown at an accelerating rate. According to the Verizon Data Breach Investigations Report (2024), 46% of cyberattacks in the United States are directed at SMBs, while 60% of businesses that suffer a significant cyber incident cease operations within six months [1]. At the same time, the SMB segment — over 33 million enterprises accounting for 44% of U.S. GDP [2] — remains critically underprotected. The root cause lies not in the absence of standards, but in their inapplicability under conditions of limited resources.

Existing information security frameworks — NIST Cybersecurity Framework 2.0, ISO/IEC 27001:2022, and CIS Controls v8 — were designed primarily for large organizations with dedicated security departments, certified personnel, and substantial implementation budgets. For SMBs that typically lack any dedicated cybersecurity specialist, full implementation of these standards is practically unfeasible.

The objective of this paper is to develop and validate a practice-oriented framework for risk assessment and information security policy implementation, adapted to the specific organizational and resource constraints of SMBs, enabling achievement of a baseline cybersecurity posture without dedicated security personnel or significant financial investment.

Analysis of existing approaches revealed a systemic gap between academic standards and the real capabilities of SMBs. Specifically, ISO/IEC 27001 implementation requires 6 to 18 months and external consultants; NIST CSF, despite its declared scalability, provides no concrete tools for organizations without an IT department; CIS Controls v8 encompasses 18 control groups, of which only a subset is adapted to the SMB level [3-5].

Based on this analysis, the SMB Cybersecurity Framework: Assessment, Policy & Implementation was developed. The framework implements a four-stage methodology: (1) asset inventory and threat identification accounting for the enterprise's industry-specific context; (2) risk prioritization using a likelihood × impact matrix; (3) implementation of a minimum viable set of security policies — password policy, access control, data protection, and incident response; (4) a phased implementation roadmap with 30/90/180-day horizons and an employee security awareness program.

A distinguishing feature of the proposed approach is its prioritization of organizational over purely technical security measures. Practical consulting experience demonstrates that the vast majority of successful cyber incidents against SMBs are enabled by organizational vulnerabilities — absence of policies, inadequate access control, low staff awareness — rather than technical infrastructure deficiencies. Addressing organizational vulnerabilities is a necessary precondition for the effectiveness of any technical security controls.

A comparative analysis of the proposed framework against existing approaches is presented in Table 1.

Table 1

Comparative analysis of information security frameworks

Criterion	*NIST CSF 2.0*	*ISO/IEC 27001*	*CIS Controls v8*	*SMB Framework (proposed)*
SMB-oriented	No	No	Partially	209
No dedicated security staff required	No	No	Partially	220
Minimal budget	No	No	Partially	Yes
Phased implementation	Yes	No	Yes	Yes
Staff awareness training	Partially	Partially	Yes	Yes
Open access	Yes	No	Yes	Yes
Implementation complexity (1-10, lower is better)	7	9	6	2

The framework is distributed in open access and designed for independent implementation by SMB enterprises without external specialist involvement. Application of the approach across enterprises in multiple sectors confirmed its

practical feasibility and effectiveness in achieving a baseline cybersecurity posture under constrained resources.

The following conclusions can be drawn from this research: 1) existing information security frameworks exhibit a systemic gap with the actual needs of SMBs, making their full implementation impractical under resource constraints; 2) the proposed framework provides a practically achievable path to baseline cybersecurity, bridging the gap between academic standards and the real capabilities of small businesses; 3) prioritizing organizational security measures over technical ones is the key principle of effective cybersecurity for SMBs; 4) a phased approach with a 30/90/180-day horizon enables enterprises to systematically build their security posture in accordance with available resources.

1. Verizon Data Breach Investigations Report 2024. Basking Ridge: Verizon, 2024. 100 p.
2. U.S. Small Business Administration. Small Business Facts. Washington: SBA, 2024.
3. NIST Cybersecurity Framework 2.0. National Institute of Standards and Technology. Gaithersburg: NIST, 2024. 32 p.
4. ISO/IEC 27001:2022. Information security management systems — Requirements. Geneva: ISO, 2022. 19 p.
5. CIS Controls Version 8. Center for Internet Security. East Greenbush: CIS, 2021. 114 p.

Аналіз векторів загроз корпоративній безпеці засобами автоматизованого інструмента OSINT

УДК 004.056

Владислав Загороднюк[1], Артем Соколов[2]

Національний університет «Одеська політехніка»,
[1]10252811@stud.op.edu.ua, [2]sokolov.a.v@op.edu.ua

Експоненційне зростання обсягів даних з відкритих джерел вимагає перегляду підходів до забезпечення корпоративної кібербезпеки [1]. Традиційні методи моніторингу демонструють низьку ефективність, оскільки фізично не здатні впоратися з потоками інформації та генерують багато хибних тривог. Автоматизація підготовки кібератак зловмисниками зумовлює потребу в симетричних рішеннях, що робить превентивну стратегію безпеки критично важливим фундаментом для бізнесу [2].

Метою роботи є розробка автоматизованого інструмента OSINT, який використовує передові архітектури штучного інтелекту, зокрема багатоагентні моделі та Retrieval-Augmented Generation (RAG), для виявлення та комплексного аналізу векторів загроз.

До ключових векторів загроз належать: 1) експлуатація зовнішньої поверхні атаки (тіньове IT); 2) глибоке ШІ-профілювання для атак соціальної інженерії; 3) масові витоки конфіденційних даних у публічних репозиторіях.

Наукова новизна дослідження полягає в інтеграції багатоагентної архітектури LangGraph із великою мовною моделлю Gemini 3.1, яка виступає як

центральне когнітивне ядро для оркестрації розвідувального циклу [3]. Запропоновано гібридну математичну модель кількісної оцінки кіберризику (CRQ), що поєднує детерміновані метрики з методами неконтрольованого та ансамблевого машинного навчання [4].

Застосунок реалізовано мовою Python (FastAPI, Celery, Redis). Конвеєр обробки даних включає три автономні ШІ-агенти: *розвідник* (збір даних про субдомени), *агент збагачення* (мапування вразливостей через API Shodan і Censys) та *агент оцінки* (підсумковий аналіз за парадигмою ReAct). Модуль RAG здійснює семантичний пошук у векторній базі FAISS, що містить звіти Threat Intelligence, унеможливлюючи алгоритмічні галюцинації моделі.

Математична модель обчислює загальний показник кіберризику R_i на основі комбінації задокументованих вразливостей, частоти спроб атак (TEF), базової оцінки втрат від компрометації активу, динамічного коефіцієнта критичності, а також коригувального фактора впливу алгоритмів машинного навчання F_{ML}. Для розрахунку фактора F_{ML} інтегровано алгоритм Gradient Boosted Decision Trees (GBDT) для класифікації активів та просторову кластеризацію DBSCAN для детектування аномалій тіньового ІТ.

На основі контрольної вибірки з понад 70 000 записів визначено вагові коефіцієнти моделі. Для забезпечення прозорості рішень (Explainable AI) побудовано графік важливості ознак алгоритму GBDT (рис. 1). Він дозволяє кількісно оцінити вплив конкретних технічних артефактів (вразливостей, відкритих портів) на формування коригувального фактора F_{ML} у загальній моделі оцінки кіберризику.

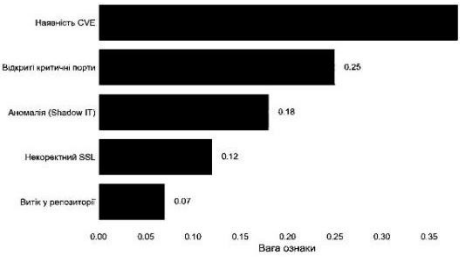


Рис.1. Графік важливості ознак

Аналіз графіка показує, що найбільш вагомим фактором ризику є наявність зафіксованих CVE (вага 0,38), тоді як відкриті критичні порти (0,25) та виявлені аномалії тіньового ІТ (0,18) посідають наступні за значущістю позиції. Для порівняння, лінійні моделі, такі як логістична регресія, продемонстрували значно гірші показники класифікації (F1-score на рівні 0,74) через нездатність ефективно фіксувати складні приховані зв'язки між технічними атрибутами мережевих активів [4]. Водночас алгоритм GBDT забезпечив мінімізацію хибних тривог завдяки вбудованим механізмам роботи з фрагментарними даними та інформаційним шумом.

Наскрізне тестування розробленого прототипу на прикладі реальної інфраструктури (домен op.edu.ua) підтвердило високу стабільність та

обчислювальну потужність архітектури, яка здатна безперебійно обробляти до 10 000 атомарних OSINT-запитів на годину. Завдяки синергії математичної моделі та інтелектуальних агентів, експериментальна точність ідентифікації загроз досягла 93,3%, а інтегральний показник F1-score склав 0,92. Крім того, автоматизація повного життєвого циклу розвідки дозволила скоротити час обробки 1 ГБ неструктурованих даних з 2-4 годин до 5-10 хвилин (приріст швидкодії у 12-48 разів).

Отже, такий підхід суттєво зменшує навантаження на аналітиків Security Operations Center (SOC) під час ручного сортування подій (на 58%) та допомагає автоматично формувати конкретні рекомендації щодо захисту периметра, забезпечуючи ефективний перехід компаній до проактивного управління корпоративними кіберризиками.

1. Ланде Д. В. OSINT у кібербезпеці : навч. посіб. Київ: ТОВ «Інжиніринг», 2024. 522 с.
2. Рибка Д. OSINT у забезпеченні національної безпеки. Науковий вісник Ужгородського національного університету. Серія: Право. 2023. Вип. 78. С. 344-348.
3. Фролов Д. І., Дягілєва М. С. Застосування технологій машинного навчання в кібербезпеці. Радіоелектроніка та молодь у XXI столітті. Харків : ХНУРЕ, 2025. №4. С. 97–99.
4. Чевардін В. Є., Юрченко О. В., Залужний О. В. Аналіз конкурентних атак на моделі машинного навчання систем кіберзахисту. Системи і технології зв'язку, інформатизації та кібербезпеки. 2023. №4. С. 9-15.

Метод шифрування на основі збільшення кількості модулів у системі залишкових класів

УДК 004.056.55

Віктор Залізняк[1], Павло Басістий[2], Михайло Касянчук[3]

[1, 3]Західноукраїнський національний університет,
[2]Тернопільський національний педагогічний університет
імені Володимира Гнатюка,
[1]viktor.zalizniak@gmail.com, [2]basi@ukr.net, [3]kasyanchuk@ukr.net

У сучасних криптографічних системах важливим завданням є підвищення стійкості алгоритмів шифрування без істотного зниження їхньої продуктивності [1]. Одним із перспективних напрямів розв’язання цієї задачі є використання системи залишкових класів (СЗК), яка дає змогу подавати числові дані у вигляді набору залишків за системою попарно взаємно простих модулів [2]. Такий підхід забезпечує незалежність обчислень за кожним модулем і створює передумови для побудови швидкодіючих криптографічних перетворень.

Класичне шифрування в СЗК ґрунтується на представленні відкритого тексту як набору залишків та подальшому відновленні відповідного десяткового числа за допомогою китайської теореми про залишки (КТЗ). У такій схемі модулі є ключами, а криптостійкість істотно залежить від складності їх

визначення з боку криптоаналітика. Водночас базова схема може бути посилена шляхом зміни параметрів КТЗ, зокрема через розширення системи модулів.

Суть запропонованого методу полягає у введенні додаткового модуля системи. Для нього формується окремий залишок. Він може бути заданий довільно або виконувати контрольну функцію. Шифрування виконується за розширеною формулою КТЗ. При цьому виникає питання вибору добутку модулів. Якщо використовувати добуток лише початкової системи модулів, додатковий залишок не повністю впливає на результат і не завжди може виконувати контрольну функцію. Натомість використання добутку всіх модулів, включаючи додатковий, розширює діапазон можливих значень шифротексту та збільшує простір ключів.

Практичне значення такого підходу полягає в тому, що збільшення кількості модулів у СЗК ускладнює структуру криптографічного перетворення. Криптоаналітик має враховувати більшу кількість можливих комбінацій модулів, залишків та проміжних параметрів КТЗ. Крім того, додатковий модуль може використовуватися як механізм контролю цілісності повідомлення після розшифрування. Це особливо важливо для систем, у яких потрібно поєднати захист інформації з виявленням помилок передавання або обробки даних.

Отже, метод шифрування на основі збільшення кількості модулів у СЗК є ефективним способом підвищення криптографічної стійкості та розширення простору ключів. Подальші дослідження доцільно спрямувати на оптимальний вибір додаткових модулів, оцінювання стійкості до криптоаналітичних атак та реалізацію методу в апаратних і гібридних криптографічних системах.

1. 1. Nieles M., Dempsey K., Pillitteri V. *An Introduction to Information Security*. Gaithersburg : NIST, 2017. 101 p..
2. 2. P.V. Ananda Mohan. *Residue number systems: Theory and applications*. Birkhäuser, Basel. 2016. 351 p.

Штучний інтелект як інструмент підтримки прийняття рішень у системах кібербезпеки

УДК 004.056:004.8

Роман Золотий[1], Ігор Чихіра[2], Віктор Устенко[3]

Тернопільський національний технічний університет імені Івана Пулюя, [1]zolotyy@gmail.com, [2]ig.vi.chi@gmail.com, Івано-Франківська філія Університету "Україна", [3]ustenkotoda@gmail.com

Зростання кількості кіберінцидентів і обсягів даних безпеки актуалізує використання штучного інтелекту як інструменту підтримки прийняття рішень у кібербезпеці. На відміну від суто сигнатурних засобів та засобів на основі правил, AI може швидко аналізувати журнали, мережевий трафік, події кінцевих пристроїв, хмарних сервісів та IoT-компонентів. Це важливо з огляду на поширення атак на доступність, програм-вимагачів, загроз даним, соціальної інженерії та експлуатації вразливостей [1].

Метою роботи є обґрунтування моделі використання штучного інтелекту для аналізу подій безпеки, оцінювання ризиків, пріоритезації інцидентів і формування рекомендацій для фахівця. Відповідно до NIST Cybersecurity Framework 2.0 управління кібербезпекою охоплює ідентифікацію активів, захист, виявлення, реагування, відновлення та управління ризиками [2]. AI доцільно розглядати як аналітичний шар цих процесів, а не як заміну експерта.

Наукова новизна підходу полягає в трактуванні AI як допоміжного інтелектуального контуру прийняття рішень, що поєднує технічні події з контекстом критичності активів, історією інцидентів, рівнем ризику та політиками реагування. Відповідно до NIST AI Risk Management Framework, довіра до AI-систем має ґрунтуватися на керованості, надійності, безпечності, прозорості, пояснюваності та підзвітності [3].

Запропонована модель передбачає чотири етапи: збирання й нормалізацію даних із SIEM, IDS/IPS, EDR, мережевих сенсорів і хмарних журналів; AI-аналіз подій із виявленням аномалій та класифікацією сценаріїв атак; оцінювання ризику з урахуванням критичності активу, потенційного впливу й достовірності спрацювання; формування рекомендацій щодо перевірки облікового запису, ізоляції вузла, зміни правил доступу або посилення моніторингу.

Таблиця 1

Функції штучного інтелекту в системі підтримки кібербезпеки

Функція AI	Призначення	Результат для фахівця
Виявлення аномалій	пошук нетипової активності в трафіку та журналах	раннє попередження про можливий інцидент
Класифікація подій	віднесення подій до типових сценаріїв атак	зменшення кількості ручного аналізу
Оцінювання ризику	урахування критичності активу та впливу інциденту	визначення пріоритету реагування
Формування рекомендацій	пропозиція дій відповідно до політик безпеки	підтримка прийняття рішення експертом

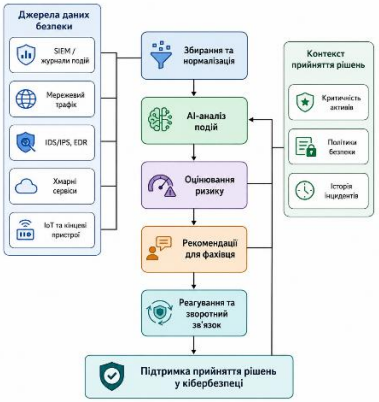


Рис.1. Штучний інтелект у системі підтримки прийняття рішень з кібербезпеки

Цінність підходу полягає у скороченні часу між появою ознак атаки та рішенням щодо реагування. AI-модуль може об'єднати невдалу автентифікацію, нетиповий трафік, звернення до критичного сервера та зміну поведінки користувача в один інцидент із підвищеним пріоритетом. Водночас остаточне рішення має ухвалювати фахівець з урахуванням організаційного контексту. Основними ризиками впровадження AI є залежність від якості навчальних даних, хибнопозитивні або хибнонегативні спрацювання, складність пояснення рішень і вразливість моделей до маніпуляцій. Оскільки ISO/IEC 27005 розглядає управління ризиками як безперервний процес [4], AI-рішення мають супроводжуватися аудитом даних, перевіркою моделей, контролем доступу та журналюванням рекомендацій.

Отже, штучний інтелект є перспективним інструментом підтримки прийняття рішень у кібербезпеці, що підвищує швидкість аналізу, покращує пріоритезацію інцидентів і сприяє обґрунтованому реагуванню. Подальші дослідження доцільно спрямувати на пояснювані AI-моделі для SOC-середовищ, оцінювання довіри до автоматизованих рекомендацій і захист AI-модулів від цілеспрямованих атак.

1. Kaur R., Gabrijelcic D., Klobucar T. Artificial intelligence for cybersecurity: Literature review and future research directions. Information Fusion. 2023. Vol. 97. Article 101804.
2. Rjoub G. et al. A survey on explainable artificial intelligence for cybersecurity. IEEE Transactions on Network and Service Management. 2023. Vol. 20(4). P. 5115-5140.
3. Capuano N., Fenza G., Loia V., Stanzione C. Explainable artificial intelligence in cybersecurity: A survey. IEEE Access. 2022. Vol. 10. P. 93575-93600.
4. Sarker I.H. et al. Cybersecurity data science: an overview from machine learning perspective. Journal of Big Data. 2020. Vol. 7. Article 41.

Analysis of modern methods for detecting phishing domains and links

UDK 004.056.5:004.738.5

Ivan Azarov[1], Anna Korchenko[2],
Illia Azarov[3], Kyrylo Davydenko[4]

State University of Communication and Informational Technologies,
[1]azarovphone@gmail.com,
NTU Dnipro Polytechnic, [2]annakor@ukr.net, [4]kirilldavy@gmail.com,
Kyiv Aviation Institute, [3]azarovforce@gmail.com

Phishing remains one of the most common cyberattack vectors, utilizing social engineering and technical masking techniques to steal confidential data.

With the emergence of automated tools, attackers have gained the ability to mass-generate deceptive URLs and clone legitimate web resources, which reduces the effectiveness of traditional security measures.

The use of comprehensive approaches to analyzing domains and links is critical for countering phishing attacks, particularly zero-day attacks [1].

The main objective of this study is to analyze the methods and criteria for detecting phishing domains and URLs, as well as to evaluate their effectiveness, speed, and practical feasibility in modern cybersecurity systems.

The process of identifying phishing resources involves checking various levels: from analyzing the URL structure to examining the webpage content. For a comprehensive review, 10 fundamental detection methods were identified.

1. Use of blacklists. The method is based on comparing the requested URL with global databases of known malicious links (e.g., Google Safe Browsing, PhishTank). If the domain is found in the database, access to it is automatically blocked. This is the oldest and most basic approach to filtering internet traffic.

2. Use of trusted site lists. The method relies on utilizing global rankings of the world's most popular and trusted domains (e.g., the top 1 million domain lists by Tranco or Cisco Umbrella). Resources included in this list are considered safe by default and are bypassed by the system without deep inspection. This approach works as a fast pre-filter to separate the bulk of legitimate traffic [2].

3. Lexical analysis of URLs. The method examines the structure of the link itself: URL length, presence of special characters, use of an IP address instead of a domain, and typosquatting. It operates at the text level, detecting statistical anomalies that are atypical for legitimate sites.

4. Analysis of DNS and WHOIS records. The method evaluates a domain's reputation based on data regarding its registration, age, time-to-live (TTL), and history of IP address changes. Phishing domains typically have an age of a few days and are registered in cheap or free zones.

5. Hosting and network reputation analysis. Verifies whether the domain's IP address belongs to a specific Autonomous System (ASN) and evaluates the hosting provider's reputation. Attackers often choose specific, abuse-tolerant data centers or networks compromised by botnets.

6. Analysis of SSL/TLS certificates. Checks the presence, type, and issuer of the cryptographic certificate. Today, domain name mismatches within the certificate or the mass use of free solutions (Let's Encrypt) are analyzed, since the presence of basic HTTPS is no longer a guarantee of security.

7. Analysis of HTML content and page structure. The method involves downloading the page code to search for hidden iframes, suspicious login collection forms, and external link mismatches (when a form sends data to a third-party domain). The method records actual malicious activity on the site.

8. Visual analysis of the page. Takes a screenshot of the web page and compares its graphic elements (logos, structure, color scheme) with reference samples of well-known brands. It uses neural networks to evaluate the percentage of visual similarity [3].

9. Analysis of redirection chains. Tracks all HTTP redirects from the initial link to the final page. Phishers often use multiple redirects through shortening services to hide the final malicious domain from security filters [4].

10. Search indexing analysis. Checks whether the domain is present in the organic search results of search engines (Google, Bing) and what its ranking is. Most phishing sites do not have time to get indexed or are quickly blocked by search algorithms [5].

Based on the comprehensive analysis conducted, the following conclusions can be drawn: there is no single universal method for detecting phishing. The simplest and fastest methods (blacklists, trusted lists of the top million domains) are excellent primary filters due to their ease of implementation, but their standalone detection effectiveness is insufficient against modern zero-day threats. Methods of visual analysis and HTML content inspection demonstrate the highest accuracy, but they are too slow for stream processing. Therefore, an effective modern phishing detection system must be combined and include the following parameters: fast lists and lexical analysis filter out mass traffic, after which resource-intensive content or visual analysis using artificial intelligence is applied to suspicious links.

1. Li, W., Manickam, S., Chong, Y. W., Leng, W., & Nanda, P. (2024). A state-of-the-art review on phishing website detection techniques. IEEE Access, 12, 187976-188012.
2. Azeez, N. A., Misra, S., Margaret, I. A., Fernandez-Sanz, L., & Abdulhamid, S. I. M. (2021). Adopting automated whitelist approach for detecting phishing attacks. Computers & Security, 108, 102328.
3. Kustiawan, Y. A., & Ghauth, K. I. (2025). Feature Engineering for Phishing Website Detection Using Machine Learning: A Systematic Review. *IEEE Access*, *13*, 192080-192104.
4. Lamina, O. A., Ayuba, W. A., Adebiyi, O. E., Michael, G. E., Samuel, O. O. D., & Samuel, K. O. (2024). AI-powered phishing detection and prevention. *Path of Science*, *10*(12), 4001-4010.
5. Alazaidah, R., BaniSalman, M., Alqawasmi, K. E., Abu Zaid, A., Hazaimeh, Y., Alshraiedeh, F. S., & Qumsiyeh, E. (2026). Identifying key features for phishing website detection through feature selection techniques. Frontiers in Computer Science, 7, 1687867.

Симетричний ієрархічний криптоалгоритм на основі Китайської теореми про залишки

УДК 004.056.55

Ігор Якименко[1], Степан Івасьєв[2]

Західноукраїнський національний університет,
[1] jiz@wunu.edu.ua, [2]isv@wunu.edu.ua

Сучасний розвиток інформаційно-комунікаційних технологій супроводжується постійним зростанням обсягів передавання даних та підвищенням вимог до їх криптографічного захисту. Особливо актуальними є задачі забезпечення конфіденційності інформації у багаторівневих інформаційних системах, де необхідно поєднувати високу криптостійкість із швидкодією процесів шифрування та дешифрування. У зв'язку з цим значний інтерес становлять криптографічні методи, побудовані на основі систем залишкових класів (СЗК) [1, 2] та Китайської теореми про залишки (КТЗ) [3].

У роботі [4] досліджено теоретичні основи симетричних криптоалгоритмів у СЗК та показано, що криптостійкість таких методів визначається кількістю модулів і їх розрядністю. Подальший розвиток даного напряму пов'язаний із

використанням ієрархічних структур СЗК. У праці [5] запропоновано симетричний криптоалгоритм на основі ієрархічної СЗК, який забезпечує поетапне зменшення розрядності модулів та операндів на кожному рівні шифрування і характеризується комбінаторною криптостійкістю до криптоаналітичних атак.

Незважаючи на існуючі дослідження, питання побудови ефективних симетричних ієрархічних криптосистем на основі Китайської теореми про залишки залишаються актуальними, зокрема в частині оптимізації швидкодії розшифрування та забезпечення багаторівневого захисту інформації.

Метою роботи є розроблення симетричного ієрархічного криптоалгоритму на основі Китайської теореми про залишки, який забезпечує багаторівневе шифрування інформації та підвищує ефективність процесу розшифрування за рахунок використання модульного представлення даних.

В симетричній ієрархічній криптосистемі на основі КТЗ обом абонентам повинні бути відомі модулі p_{ji} – ключі, де j – ієрархічний рівень, i – кількість модулів. Відкрите повідомлення S у цифровій формі на першому ієрархічному рівні розбивається на блоки S_{1i}, для кожного з яких виконується умова $S_{1i} < p_{1i}$, де p_{1i} – модулі (ключі) першого рівня. Шифрування відбувається згідно КТЗ:

$$S_1 = \left(\sum_{i=1}^{k} \quad S_{1i} M_{1i} m_{1i}\right) P_1 \,, \tag{1}$$

де $P_1 = \prod_{i=1}^{k} \quad p_{1i}$, $M_{1i} = \frac{P_1}{p_{1i}}$, m_{1i} шукається з виразу $m_{1i} = M_{1i}^{-1} p_{1i} \; = 1$, k – кількість модулів, S_1 – шифротекст першого рівня.

Після чого S_1 розбивається на блоки S_{2i} , для яких виконується умова $S_{2i} < p_{2i}$ і передається на другий ієрархічний рівень і ширується згідно співвідношення:

$$S_2 = \left(\sum_{i=1}^{k} \quad S_{2i} M_{2i} m_{2i}\right) P_2 \,, \tag{2}$$

де $P_2 = \prod_{i=1}^{k} \quad p_{2i}$, $M_{2i} = \frac{P_2}{p_{2i}}$, m_{2i} шукається з виразу $m_{2i} = M_{2i}^{-1} p_{2i} \; = 1$, k – кількість модулів, S_2 – шифротекст другого рівня.

Відповідно на j – тому ієрархічномі рівні отримується шифротекст S_j згідно формули:

$$S_j = \left(\sum_{i=1}^{k} \quad S_{ji} M_{ji} m_{ji}\right) P_j \,, \tag{3}$$

де $P_j = \prod_{i=1}^{k} \quad p_{ji}$, $M_{ji} = \frac{P_j}{p_{ji}}$, m_{ji} шукається з виразу $m_{ji} = M_{ji}^{-1} p_{ji} \; = 1$, k – кількість модулів, S_j – шифротекст j – того ієрархічного рівня передається по відкритому каналі зв'язку до отримувача.

Дешифрування відбувається у зворотньому порядку з j – того ієрархічного рівня до першого згідно формули:

$$S_{ji} = S_j mod p_{ji} \tag{4}$$

На першому рівні отримуємо блоки S_{1i}, які в результаті конкатенації відновлюють вхідне повідомлення S. Запропонований симетричний метод шифрування доцільно використовувати, коли потрібне швидке дешифрування, оскільки на цьому етапі необхідно виконання тільки операції пошуку залишків.

У роботі запропоновано симетричний ієрархічний криптоалгоритм на основі КТЗ, який реалізує багаторівневе шифрування інформації шляхом

послідовного перетворення даних у системі взаємно простих модулів. Особливістю методу є використання ієрархічної структури ключів, що дозволяє організувати багаторівневий захист інформації та підвищити гнучкість криптографічної системи.

Використання модульного представлення даних сприяє зменшенню обчислювальної складності окремих операцій та створює передумови для паралельної обробки інформації.

Дослідження показали, що запропонований метод доцільно використовувати в інформаційних системах, де важливими є швидке дешифрування даних, багаторівневий доступ та підвищені вимоги до захисту інформації.

1. Omondi A., Premkumar B. Residue number systems: theory and implementation. London: Imperial College Press, 2007. 296 p.
2. Singh N. An overview of Residue Number System. Devices, Circuits & Communication: Proceedings of the National Seminar. 2008, pp. 132-135.
3. Verma S., Garg D. Improvement in rebalanced CRT RSA. *International Arab Journal of Information Technology*. Vol.12, No. 6, 2015, pp.524-532.
4. M. M. Kasianchuk, I. Z. Yakymenko, Ya. M. Nykolaychuk Symmetric Cryptoalgorithms in the Residue Number System. Cybernetics and Systems Analysis, 2021, Vol 57, Issue 2, p.184-189. https://doi.org/10.1007/s10559-021-00358-6
5. Yakymenko I., Kasianchuk M., Martyniuk O., Martyniuk S. A Symmetric Cryptoalgorithm Based on a Hierarchical Residue Number System. International Journal of Computing, 2025. 24(1), pp. 92-101. https://doi.org/10.47839/ijc.24.1.3880

Оцінка ефективності Counter-OSINT стратегій за допомогою теорії інформації

УДК 004.056.5:519.72

Валерія Івкова[1], Іван Опірський[2]

Національний університет «Львівська політехніка» ,
[1]valeriia.s.ivkova@lpnu.ua, [2]ivan.r.opirskyi@lpnu.ua

Стрімка діджиталізація та гібридні загрози дозволяють ворожим аналітикам реконструювати детальні цифрові профілі через OSINT-технології з мінімальними витратами ресурсів, що зумовлює перехід від пасивної цифрової гігієни до проактивних Counter-OSINT стратегій.

З позиції теорії інформації, ефективність таких стратегій доцільно вимірювати обсягом когнітивних та часових ресурсів, необхідних атакуючій стороні для відновлення цілісного "корисного сигналу" цифрової ідентичності з наявних даних. У цьому контексті виділяють два концептуальні підходи: обфускацію та компартменталізацію.

Метод обфускації фокусується на зниженні співвідношення сигнал/шум шляхом генерації надлишкової хибної інформації. Проте сучасні автоматизовані OSINT-засоби та алгоритми машинного навчання здатні

ефективно фільтрувати цей "шум", виокремлюючи справжні «точки дотику», що дозволяє порівняно легко відновити граф ідентичності [1].

Метод компартменталізації - заснований на військовій моделі «need-to-know», передбачає логічну та технічну ізоляцію цифрових ідентичностей на рівнях операційної системи, браузера та мережевого трафіку. Цей підхід фізично розриває зв'язки між фрагментами даних, повністю фрагментуючи граф ідентичності та позбавляючи аналітика "точок дотику" [2 с.73].

З позиції теорії інформації, цифрову ідентичність об’єкта можна представити як систему станів X. Разом із тим ефективність протидії OSINT оцінюється через збільшення невизначеності (ентропії) для аналітика.

Ефективність обфускації базується на штучному підвищенні ентропії системи шляхом внесення надлишкового шуму N. Проте зловмисник може використовувати алгоритми фільтрації для мінімізації відстані Кульбака-Лейблера [3] (D_{KL}), яка вимірює розбіжність між розподілом ймовірностей зашумлених даних $P(x)$ та реальним профілем $Q(x)$:

$$D_{KL}(P\|Q) = \sum_{x\epsilon X} P(x)\ log\ log\ \frac{P(x)}{Q(x)} \tag{1}$$

При використанні автоматизованих OSINT-засобів значення D_{KL} стрімко зменшується, що вказує на низьку стійкість обфускації до машинного аналізу.

На відміну від обфускації, компартменталізація спрямована на мінімізацію взаємної інформації $I(X;Y)$ між різними сегментами ідентичності X та Y:

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \tag{2}$$

де $H(X,Y)$ — спільна ентропія сегментів. Стратегія вважається успішною, якщо $I(X;Y) \to 0$, що означає повну відсутність кореляції між фрагментами даних. Це фізично унеможливлює автоматизоване поєднання профілю через відсутність точок дотику.

Ефективність стратегії E визначається як функція від складності відновлення сигналу C:

Для обфускації:

$$cobf \approx log\left(\frac{S}{N}\right) \tag{3}$$

де S/N — співвідношення сигнал/шум. Витрати атакуючої сторони зростають лише лінійно.

Для компартменталізації:

$$C_{comp} = \sum_{i=1}^{n} Ti \tag{4}$$

де Ti — час на окреме ручне дослідження кожного ізольованого сегмента n.

Математичне моделювання доводить, що компартменталізація забезпечує експоненціальне зростання витрат зловмисника, оскільки розриває логічні

зв'язки в графі ідентичності, тоді як обфускація лише тимчасово ускладнює обробку даних.

Таким чином, головна перевага компартменталізації над обфускацією полягає в тому, що вона ускладнює можливість автоматизованої кореляції даних, що змушує атакуючу сторону переходити від швидкого збору інформації до складного, ресурсомісткого та часто безрезультатного ручного інтелектуального розслідування.

Проте, поряд із високою ефективністю, метод компартменталізації накладає значні обмеження на користувача та потребує підвищених апаратних ресурсів для підтримки декількох ізольованих середовищ. Це створює поріг входження для пересічного користувача, проте є виправданим компромісом у контексті експоненціального зростання витрат атакуючої сторони

1. Cheng, H., Qiang, C., Cong, L., Xiao, J., Liu, S., Zhou, X., Wang, H., Ruan, M., & Lv, C. (2025). A Novel Data Obfuscation Framework Integrating Probability Density and Information Entropy for Privacy Preservation. *Applied Sciences*, *15*(3), 1261. https://doi.org/10.3390/app15031261
2. Івкова В.С., & Опірський І.Р., (2025). Дослідження можливості інтеграції методу компартменталізації у захист інформації у відкритих джерелах. *Computer systems and network*, *7*(2), 71–83. https://doi.org/10.23939/csn2025.02.071
3. Lopes, A. O., & Mengue, J. K. (2022). On information gain, Kullback-Leibler divergence, entropy production and the involution kernel. *Discrete & Continuous Dynamical Systems*, 0. https://doi.org/10.3934/dcds.2022026

Сучасний стан кібербезпеки в Україні

УДК 004.056.53

Кардашук Володимир[1]

Східноукраїнський національний університет, [1]kardashuk1@snu.edu.ua

Стрімкий розвиток інформаційних технологій суттєво випереджає створення надійних умов захисту цифрового кіберпростору. Для сучасного економічного розвитку критичним є забезпечення особистої безпеки, бізнесу та держави. Актуальним завданням сьогодення є створення комплексу заходів спрямованих на захист інформаційних ресурсів від кібератак, небажаного доступу, інших дій на модифікацію або знищення важливих даних. Сучасне інформаційне оточення швидко змінюється, що вимагає динамічно реагувати на загрози, адаптуватись до викликів та кібернетичних атак. Все більша кількість загроз, методів та стратегій використовує вразливості кіберзахисту.

Аналіз сучасного стану кібербезпеки в Україні показав недостатню координацію між суб'єктами кібербезпеки, що ускладнює координацію на кіберзагрози [1].

В останні роки підмічена тенденція еволюціонування загроз з незмінними мотивами фінансової вимоги [2].

Сучасні дослідження в сфері кібербезпеки спрямовані в таких новітніх сферах як штучний інтелект, блокчейн, квантові обчислення на системи кіберзахисту. Відмічено, що штучний інтелект є водночас найбільшою загрозою і найпотужнішим інструментом для захисту інформаційних систем.

Прогнозується, що до 2031 року буде зберігатись тенденція перевищення збитків (12 трлн. дол.) від кіберзлочинності над витратами (1 трлн. дол.) на захист, і це не тимчасовий дисбаланс, а фундаментальна проблема [3].

Актуальним постає питання подальшого розвитку Національної стратегії кібербезпеки в Україні як стратегічного пріоритету. На системи кіберзахисту свій вплив відіграють і зростаючий ринок хмарних сервісів таких як Google Cloud Platform, Microsoft, AWS, Azure. Хмари у 2026 році відіграють роль цифрового фундаменту інформаційного розвитку випереджаючи можливість кіберзахисту. В цій боротьбі поки перевага не на стороні кіберзахисту.

Що стосується застосування квантових обчислень для криптографічного перетворення, то дослідження в цій галузі ще тільки на початковому етапі розвитку не тільки в Україні, а і у світі. Формується новітня галузь постквантової криптографії на базі штучного інтелекту, як відповідь кіберзагрозам. Лідерство в цьому напрямку зберігають такі фірми, як Google (Quantum AI), Microsoft (Azure Quantum), Amazon (Braket, чіп Ocelot), Thale. Національний інститут стандартів і технологій США (NIST) планує у 2027 році закінчити розробку стандартів постквантової криптографії.

В поточній ситуації в Україна актуальним стає питання також об'єднанню урядів, корпорацій, інвесторів, компаній, незалежних експертів, стартапів, аналітичних центрів і наукових кіл для спільної протидії сучасним кіберзагрозам та обміну передовими рішеннями, практиками та тенденціями. Цим важливим питання був присвячений Київський Міжнародний Форум з Кіберстійкості 2026. Така всебічна робота спрямована на збільшення міцності в кібердомені для України та країн Європейського союзу через кіберстійкість на основі досвіду та знань, отриманих під час кібервійни, шляхом зміцнення міжнародного співробітництва та синергії з глобальними гравцями в державному та приватному секторах.

Крім того, загрози направлені проти кричної інфраструктури, вони також додатково впливають на процеси ухвалення рішень та суспільне сприйняття. Це спонукає дослідження таких кібератак та їх прогнозування для зміцнення кіберстійкості. Серед кібернетичних атак гостро стоїть питання протидії FIMI (Foreign Information Manipulation and Interference) нового покоління.

У секторах енергетики та телекомунікацій основна увага зосереджена на сучасних технологіях швидкого відновлення в умовах кібератак. В цьому питанні важливу ключову роль у зміцненні цифрової стійкості в Україні відіграють програми та масштабовані рішення для захисту критичної інфраструктури.

Стратегічно важливим кроком для України стало приєднання у жовтні 2025 року до Угоди про визнання загальних критеріїв (Common Criteria Recognition Arrangement – CCRA), яка встановлює єдині підходи до оцінювання засобів захисту інформації та підтвердження рівня довіри до такої оцінки, а також методології її проведення [4].

Стосовно кібергігієни рекомендується створювати надійні паролі, бути обережним з підозрілими повідомленнями, використовувати перевірені ресурси для фінансових операцій та відповідально ставитися до поширення особистої інформації в мережі.

Зростає фінансова підтримка партнерів на оновлення IT-інфраструктури, захист цифрових продуктів та навчання фахівців, що сприяє посиленню кібербезпеки та розвитку цифрової інфраструктури України.

1. Живило Є.О. Пріоритетні напрями національної стратегії кібербезпеки в контексті інтеграції до трирівневої моделі кібероборони. Державне будівництво. – № 1 (37). – 2025. – С. 229-251.
2. Олег Полігенько. Що відбувається на ринку кібербезпеки — колонка. URL: https://ain.ua/2026/04/15/shho-vidbuvajetsia-na-rinku-kiberbezpeki-kolonka/ (дата звернення: 04.05.2025).
3. Глобальна індустрія кібербезпеки станом на 2026 рік. Основні показники, тенденції та прогнози. URL: https://niss.gov.ua/doslidzhennya/natsionalna-bezpeka/hlobalna-industriya-kiberbezpeky-stanom-na-2026-rik-osnovni (дата звернення: 04.05.2025).
4. Приєднання до Угоди про визнання Common Criteria значно розширило інструментарій захисту інформації – перелік засобів та продуктів. URL: https://cip.gov.ua/ua/news/priyednannya-do-ugodi-pro-viznannya-common-criteria-znachno-rozshirilo-instrumentarii-zakhistu-informaciyi-perelik-zasobiv-ta-produktiv (дата звернення: 04.05.2025).

Метод шифрування растрових зображень засобами асиметричних криптосистем

УДК 004.056.5

Євгеній Кацубо

Національний університет «Одеська політехніка», 10252737@stud.op.edu.ua

В епоху глобальної цифровізації та безперервного обміну даними класичні асиметричні алгоритми попіксельного шифрування не здатні приховати просторову надлишковість та автокореляцію елементів, що призводить до візуалізації контурів об'єктів у шифротексті. Це зумовлює необхідність розробки комплексних гібридних моделей, які поєднують криптографічну стійкість факторизації великих чисел із методами нелінійного просторового маскування та внутрішньоблокового алгебраїчного перемішування.

Метою роботи є розробка та дослідження розширеної моделі модифікованого асиметричного алгоритму, що інтегрує двовимірну квадратичну координатну маску та механізми алгебраїчного зчеплення для досягнення максимальної ентропії шифротексту.

Запропонований у роботі метод розглядає зображення як матрицю, що поділяється на незалежні блоки по чотири суміжні пікселі. Диференційована обробка полягає в тому, що крайні пікселі піддаються експоненціюванню у

першому кільці лишків, а внутрішні — у другому, що повністю унеможливлює частотний криптоаналіз. Для руйнування візуальних патернів застосовується позиційно залежна квадратична маска, яка додає унікальне псевдовипадкове зміщення на основі глобальних координат блоку [1].

Для внутрішніх елементів реалізовано механізм алгебраїчного зчеплення через знаходження їхньої суми та різниці в модульному просторі, що створює потужний лавинний ефект. Використання цих операцій формує нерозривну залежність між сусідніми пікселями, тому будь-яка спроба локальної модифікації руйнує весь зашифрований блок, надійно захищаючи цілісність документа. Повна математична оборотність системи при використанні мультиплікативного оберненого елемента забезпечує бездоганне побітове відновлення оригінального зображення без жодних структурних втрат чи артефактів.

Проведені дослідження підтверджують, що поєднання векторного розбиття, двокільцевої обробки та позиційно-залежної маски забезпечує абсолютно рівномірний розподіл інтенсивностей на гістограмі шифротексту. Це повністю перетворює початкові візуальні дані на гетерогенний псевдовипадковий шум та експоненціально ускладнює проведення диференціальних і статистичних атак, роблячи запропонований підхід ефективним методом для конфіденційного електронного документообігу.

1. Zhang B., Liu L. Chaos-Based Image Encryption: Review, Application, and Challenges. *Mathematics*. 2023. Vol. 11, no. 11. P. 2585. DOI: https://doi.org/10.3390/math11112585

Порівняльний аналіз SDN-систем за критеріями кібербезпеки

УДК 004.056:004.7

Юрій Кльоц[1], Олексій Федоров[2]

Хмельницький національний університет, [1]klots@khmnu.edu.ua, [2] fedorovoo@khmnu.edu.ua

Програмно-конфігурована мережа є архітектурою, у якій функції керування трафіком логічно відокремлюються від функцій його передавання. На відміну від традиційної мережевої інфраструктури, де рішення щодо маршрутизації або комутації приймаються розподілено на кожному пристрої, у SDN значна частина логіки керування концентрується в контролері. Це забезпечує централізоване застосування політик, гнучке програмне керування трафіком, швидке впровадження змін конфігурації та можливість побудови систем автоматизованого виявлення аномалій [1].

З погляду кібербезпеки SDN-архітектура має подвійний характер. З одного боку, централізоване керування спрощує реалізацію політик сегментації, ізоляції вузлів, динамічного блокування потоків, перенаправлення трафіку на засоби аналізу та збирання телеметрії з мережевих пристроїв. З іншого боку, контролер, southbound-канали та програмно-керовані комутатори стають критичними об’єктами захисту. Компрометація контролера або порушення цілісності правил потоків може призвести до порушення роботи всієї мережевої

інфраструктури. Тому під час порівняння SDN-систем доцільно оцінювати не лише їхню продуктивність, масштабованість і підтримку протоколів, а й здатність забезпечувати виявлення шкідливих дій на рівні кінцевих пристроїв, контролера та програмно-керованого обладнання [2].

Для порівняльного аналізу SDN-систем доцільно використовувати такі критерії: архітектурна роль системи, підтримувані протоколи керування, масштабованість, наявність відкритих API, можливість інтеграції з системами моніторингу й аналізу трафіку, підтримка журналювання подій, можливість отримання статистики потоків, підтримка механізмів контролю доступу, захищеність керуючих інтерфейсів, придатність до виявлення атак на кінцеві пристрої, а також придатність до виявлення атак на сам SDN-контролер і мережеве обладнання.

Порівняння SDN-систем доцільно виконувати не за кількістю підтримуваних протоколів, а за придатністю до виявлення шкідливих дій у трьох зонах: на кінцевих пристроях, на рівні контролера та на рівні програмно-керованого обладнання. Перша зона охоплює сканування, spoofing, DDoS-активність і порушення політик доступу. Друга зона стосується захисту API, застосунків контролера, автентифікації та перевантаження Packet-In подіями. Третя зона пов’язана з контролем flow-таблиць, southbound-з’єднань, стану портів і коректності топологічної інформації.

Для порівняння SDN-систем використаємо такі критерії: OF — OpenFlow / відкритість, API — відкритість і програмованість API, T — можливість телеметрії та отримання статистики (3 – так, 2 – частково, 1 – ні), E — виявлення дій кінцевих вузлів, C — захист і контроль контролера, D — контроль програмно-керованого обладнання, SI — інтеграція із зовнішнім аналізом, Σ — узагальнена придатність. Результати порівняння представлено в таблиці 1.

Таблиця 1

Порівняння SDN-систем за безпековими критеріями

Система	Клас	OF	API	T	E	C	D	SI	Σ
OpenDaylight	SDN-контролер	+	3	3	3	3	3	3	3
ONOS	SDN-контролер	+	3	3	3	3	3	3	3
Ryu	фреймворк	+	3	2	3	2	2	2	2
Floodlight	SDN-контролер Ctrl	+	2	2	2	1	2	2	2
Open vSwitch	програмний комутатор vSw	+	2	3	2	1	2	3	2
VMware NSX	SDN/NFV-платформа	-	2	3	3	3	3	3	3
Omada	NMS	-	1	2	2	2	2	1	2

Результати порівняння свідчать, що системи програмно-конфігурованих мереж, незалежно від призначення, мають спільну особливість: централізація керування підвищує гнучкість адміністрування, але водночас формує нові критичні точки вразливості. SDN-контролер, northbound- і southbound-

інтерфейси, програмно-керовані комутатори та кінцеві пристрої утворюють багаторівневу поверхню атаки, тому безпека SDN-інфраструктури не може обмежуватися лише фільтрацією трафіку або налаштуванням політик доступу.

Наявність механізмів централізованого керування, телеметрії, журналювання, аналізу потоків і програмного застосування політик створює передумови для побудови ефективних засобів захисту. Водночас ці можливості не усувають вразливостей автоматично, тому розглянуті системи потребують додаткових механізмів виявлення атак, кореляції подій, контролю цілісності правил, аналізу поведінки кінцевих пристроїв і захисту каналів взаємодії між компонентами SDN-архітектури.

Отже, результати порівняння підтверджують актуальність розроблення методів захисту та виявлення атак у програмно-конфігурованих мережах. Особливо важливими є методи, що враховують стан контролера, зміни в таблицях потоків, поведінку програмно-керованого обладнання, активність кінцевих вузлів і параметри службової взаємодії між компонентами SDN.

1. Odarchenko, R.; Iavich, M.; Iashvili, G.; Fedushko, S.; Syerov, Y. Assessment of Security KPIs for 5G Network Slices for Special Groups of Subscribers. Big Data Cogn. Comput. 2023, 7, 169. https://doi.org/10.3390/bdcc7040169
2. Cherednichenko O., Sharonova N., Pliekhova G., Babkova N. Intelligent Methods of Secure Routing in Software-Defined Networks. CEUR Workshop Proceedings. 2024. Vol. 3664. P. 342–351. https://doi.org/10.31110/COLINS/2024-1/024

Аналіз методологічного забезпечення оцінювання кіберстійкості інформаційних ресурсів

УДК 004.056:006.01

Олександра Ковальчук[1], Євгенія Іванченко[2]

Державний університет інформаційно-комунікаційних технологій,
[1]oa.kovalchuk@duikt.edu.ua, [2]e.ivanchenko@duikt.edu.ua

У сучасних умовах зростання складності кібератак традиційні підходи до безпеки стають недостатніми. Актуальним є перехід від кібербезпеки до кіберстійкості (Cyber Resilience) – здатності інформаційних ресурсів (ІР) не лише протистояти загрозам, але й адаптуватися до них та швидко відновлюватися. Проте на практиці виникає проблема об'єктивної оцінки реального рівня такої стійкості.

Метою є аналіз методологій управління кіберстійкістю ІР для виявлення їхніх переваг та недоліків. Для досягнення поставленої мети необхідно здійснити порівняльний аналіз п’яти провідних фреймворків за критеріями фокусу оцінки, динаміки тестування та джерел даних.

Фундаментом для побудови систем кіберстійкості виступають міжнародні стандарти, зокрема ISO/IEC 27001 закладає основу через впровадження системи управління інформаційною безпекою (СУІБ) та забезпечує функцію

протистояння (Withstand) через систематичне управління ризиками [1]. ISO/IEC 27031 визначає готовність ІКТ-інфраструктури до забезпечення безперервності бізнесу, допомагає визначити рівень готовності для досягнення цільових метрик безперервності [2]. NIST CSF описує життєвий цикл стійкості через п'ять основних функцій: ідентифікація, захист, виявлення, реагування та відновлення [3]. NIST SP 800-160: пропонує інженерний підхід, розглядаючи безпеку як невід'ємну властивість архітектури (Security by Design) та включає принципи надмірності та «м'якої деградації» [4].

Для переведення підходів, описаних у стандартах, у практичну площину застосовуються спеціалізовані методики оцінювання. Для детального аналізу було обрано п'ять методологій, що різняться за фокусом та інструментарієм - CRR [6], CAF [8], C-RA[9], CRI [7] та IT Governance [10] (табл. 1).

Таблиця 1

Порівняльний аналіз методик оцінки кіберстійкості інформаційних ресурсів

Методологія	*Фокус оцінки*	*Динаміка*	*Джерело даних*
CRR (CISA)	Зрілість процесів (якісна)	Статична	Опитування, документація
CAF (UK NCSC)	Досягнення цілей (якісна)	Статична	Аудит доказів
C-RAF (НКМА)	Гібридний (зрілість + тех.)	Динамічна	Тестування (iCAST), опитування
CRI (Академічні)	Кількісні метрики (MTTR)	Статична	Технічні показники
IT Governance	Управлінська зрілість	Статична	Консалтинговий аудит

Проведений аналіз провідних фреймворків показав, що якісні моделі (CRR, IT Governance) зосереджені на оцінці зрілості процесів, але базуються на суб'єктивних опитуваннях, що дає уявлення про «паперову стійкість». CRA базується на якості і достатності зібраних доказів для проведення оцінки. Кількісні моделі (CRI) закладають до об'єктивність оцінки через метрики часу (MTTR, RTO), проте часто не враховують організаційний контекст та бізнес-пріоритети. Динамічне тестування застосовується тільки у вузькоспеціалізованому фреймворку C-RAF, де використовується моделювання атак на основі Threat Intelligence (iCAST).

Ключовим недоліком фреймворків є відсутність механізмів перевірки адекватності та достовірності вихідних даних. Якісні моделі залежать від компетентності відповідей при самооцінці, кількісні моделі - від актуальності метрик, що використовуються для оцінки, що може призвести до формування хибного уявлення про захищеність ІР.

Отже, аналіз показав, що жодна з розглянутих методологій не вирішує проблему адекватності вхідних даних. Перспективою досліджень є розробка механізму, що поєднав би кількісні архітектурні метрики з оцінкою зрілості процесів та обов'язковим коефіцієнтом валідації вхідних даних.

1. ISO/IEC 27001:2022. Information security, cybersecurity and privacy

protection — Information security management systems — Requirements. URL: https://www.iso.org/standard/27001
2. ISO/IEC 27031:2025. Information technology — Security techniques — Guidelines for information and communication technology readiness for business continuity. URL: https://www.iso.org/standard/27031
3. The NIST Cybersecurity Framework (CSF) 2.0. URL: https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.29.pdf
4. NIST Special Publication 800-160, Volume 2. URL: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-160v2r1.pdf
5. MITRE ATT&CK Framework. URL: https://attack.mitre.org/\
6. Cyber Resilience Review. URL: https://www.cisa.gov/resources-tools/services/cyber-resilience-review-crr
7. Cyber Resilience Index. URL: https://www3.weforum.org/docs/WEF_Cyber_Resilience_Index_2022.pdf
8. Cyber Assessment Framework. URL: https://www.ncsc.gov.uk/collection/cyber-assessment-framework
9. Cyber Resilience Assessment Framework. URL: https://uploads-ssl.webflow.com/59d28ad983887e000196f803/5fecc1fe13498132b4fa835b_HKMA CFI - Cyber Resilience Assessment Framework - Dec 2016.pdf
10. IT Governance Cyber Resilience Framework. URL: https://www.itgovernance.co.uk/cyber-resilience-framework

Мережева безпека IoT-пристроїв у кіберфізичних системах розумного міста

УДК 004.056:004.738.5:681.5

Андрій Микитишин[1], Сергій Козак[2]
Роман Ніколайчук[3]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]mikitishin@gmail.com, [2]serhii_kozak1710@tntu.edu.ua,
[3]romanikolaychuk2017@gmail.com

Концепція розумного міста передбачає використання кіберфізичних систем, у яких IoT-пристрої поєднують фізичні об'єкти інфраструктури з цифровими сервісами, сенсорними мережами, edge-вузлами, хмарними платформами та системами підтримки рішень. Вони забезпечують телеметрію, дистанційне керування, автоматичне реагування й інтеграцію з сервісами освітлення, транспорту, екологічного моніторингу, енергоменеджменту, водопостачання та безпеки громадських просторів. Ключовою проблемою є мережева безпека IoT-пристроїв, оскільки через мережеву взаємодію передаються дані між сенсорами, шлюзами, платформами оброблення й операторськими інтерфейсами. Розподіленість вузлів, різнорідні протоколи, обмежені ресурси, фізична доступність частини обладнання, бездротові канали й залежність від хмарних сервісів підвищують ризик перехоплення або підміни телеметрії, відмови сервісу, втрати керованості виконавчими механізмами та помилкових

управлінських рішень[1]. Для IoT-систем це означає потребу в моделі, що враховує маршрути передавання даних, сегментацію, правила доступу, журнали подій, захищене оновлення та взаємодію з хмарними платформами. Базові вимоги IoT-безпеки передбачають відмову від універсальних стандартних паролів, безпечне зберігання облікових даних, керування вразливостями, оновлення програмного забезпечення й мінімізацію поверхні атаки [2].

Основними загрозами є перехоплення трафіку, MITM-атаки, несанкціоноване підключення вузлів, сканування портів, слабкі паролі, підміна ідентифікаторів, атаки на MQTT, CoAP, HTTP, Modbus TCP та порушення доступності.

Таблиця 1

Основні рівні мережевого захисту IoT-пристроїв

Рівень захисту	Об'єкти захисту	Типові ризики	Захисні заходи
Рівень пристроїв	сенсори, контролери, виконавчі модулі	несанкціонований доступ, слабкі паролі, вразлива прошивка	унікальні облікові дані, безпечне оновлення, вимкнення зайвих сервісів
Мережевий рівень	канали зв'язку, шлюзи, маршрутизатори, протоколи	перехоплення трафіку, MITM-атаки, DDoS, підміна пакетів	шифрування, VPN, TLS, сегментація, фільтрація трафіку
Рівень платформ	edge-вузли, хмарні сервіси, API, панелі керування	компрометація доступу, витік даних, порушення доступності	контроль доступу, журналювання, резервування, SIEM-моніторинг

Практична реалізація має починатися з інвентаризації IoT-пристроїв і класифікації їх за критичністю. Для кожного вузла визначають функцію, мережеву адресу, протоколи обміну, рівень доступу, залежності та наслідки компрометації. Далі формують окремі сегменти для критичних пристроїв, тестових вузлів, адміністративного доступу, публічних сервісів і платформ оброблення даних, що зменшує ризик горизонтального поширення атаки.

Технічний захист має включати TLS для прикладних протоколів, VPN для віддаленого адміністрування, сертифікати пристроїв для взаємної автентифікації, контроль цілісності повідомлень і керування ключами. Для обмежених IoT-вузлів слід застосовувати легковагові, але криптографічно стійкі механізми, поєднані з контролем доступу, оновленнями та журналюванням.

Моніторинг мережевої активності є важливим, бо багато сенсорних і виконавчих вузлів мають передбачувані шаблони поведінки. Ознаками компрометації можуть бути різке збільшення з'єднань, звернення до невідомих адрес, зміна частоти передавання даних, нетипові команди або спроби автентифікації з невідомих джерел. З огляду на актуальні атаки на доступність, експлуатацію вразливостей і загрози даним моніторинг потрібно інтегрувати з реагуванням на інциденти [3].

Для підвищення стійкості слід застосовувати принцип мінімально необхідного доступу: пристрій взаємодіє лише з ресурсами, потрібними для його функцій, а адміністративні інтерфейси ізолюються від публічних мереж і захищаються багатофакторною автентифікацією. Європейський підхід до критичних суб'єктів акцентує безперервність функцій, урахування взаємозалежностей і здатність до відновлення після інцидентів [4].

Отже, мережева безпека IoT-пристроїв є базовою умовою надійності кіберфізичних систем розумного міста. Запропонований підхід охоплює автентифікацію, шифрування, сегментацію, моніторинг, журналювання та реагування. Подальші дослідження доцільно спрямувати на автоматизоване виявлення аномалій трафіку, адаптацію Zero Trust і моделі оцінювання ризиків для різних IoT-мереж.

1. Humayed A., Lin J., Li F., Luo B. Cyber-physical systems security - A survey. IEEE Internet of Things Journal. 2017. Vol. 4(6). P. 1802-1831.
2. Mosenia A., Jha N.K. A comprehensive study of security of Internet-of-Things. IEEE Transactions on Emerging Topics in Computing. 2017. Vol. 5(4). P. 586-602.
3. Khan M.A., Salah K. IoT security: Review, blockchain solutions, and open challenges. Future Generation Computer Systems. 2018. Vol. 82. P. 395-411.
4. Dritsas E., Trigka M. A survey on cybersecurity in IoT. Future Internet. 2025. Vol. 17(1). Article 30.

AI bots as a factor reducing the cyber resilience of virtual communities on social networking services

UDK 004.056:004.738.5(045)

Vadym Kolesnyk

Kharkiv National University of Radio Electronics, vadym.kolesnyk@nure.ua

The cyber resilience of virtual communities on social networking services is determined by their ability to maintain functionality, uphold trust among participants, ensure communicative continuity, adapt to disruptive influences, and recover from them. One factor reducing cyber resilience is AI bots, which can mimic human behavior, automatically generate messages, perpetuate specific narratives, provoke conflicts, and complicate the detection of coordinated inauthentic activity [1–2].

The urgency of the issue is heightened by the fact that AI is already considered one of the defining factors of the modern cyberthreat landscape. In particular, ENISA notes that AI is used to enhance social engineering, generate content, and increase the effectiveness of destructive campaigns [3]. In the case of virtual communities, the danger lies not only in the spread of misinformation but also in the gradual degradation of the socio-technical environment.

The goal is to identify sets of indicators that can be used to assess the impact of AI bots on the cyber resilience of virtual communities on social networking services.

Existing research has primarily focused on bot detection, disinformation analysis, automated moderation, network resilience, and the assessment of behavioral

anomalies [3]. Such approaches are important because they enable the identification of suspicious accounts, the classification of destructive content, and the detection of specific signs of coordinated influence. At the same time, their limitation lies in their fragmentary nature: they do not always reveal whether a community maintains its cyber resilience after exposure to AI bots.

It is advisable to view AI bots as a factor in the systemic destabilization of a virtual community, affecting the four interrelated dimensions of its cyber resilience (Fig. 1).

Fig. 1. Dimensions of a virtual community's cyber resilience

The communicative dimension encompasses an increase in the proportion of repetitive or formulaic posts, a decline in the depth of discussions, a rise in the number of contentious debates, the spread of toxic reactions, and a decrease in substantive reciprocity among participants. The behavioral dimension includes abnormal posting frequency, synchronized actions by groups of accounts, unnatural activity, sudden spikes in comments or reactions, and uniformity in communication patterns. Trust-related indicators are associated with a decline in real user participation, a suspicious increase in the number of new accounts, the fragmentation of discussions, and a weakening of social support within the community. Moderation-related indicators characterize an increase in the number of complaints, longer response times from moderators, the reappearance of destructive content, and an increase in the number of repeat violations.

Unlike approaches that assess only signs of the presence of bots or bot-like activity, the proposed approach focuses on evaluating their impact on the functioning of the community. This allows us to view AI bots not merely as a technical or informational risk, but as a cybersecurity factor that can alter the structure of interactions, participant behavior, communication channels, and the effectiveness of moderation mechanisms. In this context, structural stability is just one component of broader cyberresilience: a community may maintain formal activity but lose the trust of its members, engage in destructive communication, and lose its ability to recover.

A limitation of the proposed approach is the difficulty of reliably distinguishing between AI bots, ordinary automated accounts, and overly active real users. An

additional challenge is the uneven access to data across different platforms, particularly to moderation logs, interaction histories, and account metadata.

Further research should focus on building a model of a virtual community in which the proportion of AI bots, the intensity of their activity, the level of toxicity, and the speed of moderation vary. This will allow us to identify empirical threshold conditions under which AI bots do not yet disrupt community functioning, as well as critical states under which significant destructive effects, fragmentation of interactions, and a decline in cyberresilience occur.

1. Molodetska K. Information Influence on the Virtual Community: Implementation Features and Method of Detection in Social Internet Services / K. Molodetska, S. Veretiuk, I. Rahimova, S. Milevskyi, V. Khvostenko // 2023 7th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). Ankara, Turkiye, 2023. P. 1–6. DOI: 10.1109/ISMSIT58785.2023.10305001.
2. Doshi J., Novacic I., Fletcher C., Borges M., Zhong E., Marino M. C., Gan J., Mager S., Sprague D., Xia M. Sleeper Social Bots: a new generation of AI disinformation bots are already a political threat. arXiv. 2024. DOI: 10.48550/arXiv.2408.12603.
3. ENISA. ENISA Threat Landscape 2025. Luxembourg: Publications Office of the European Union, 2025.

Аналіз атак витоку системних інструкцій у великих мовних моделях

УДК 004.056.5

Віктор Кольченко[1]

Національний університет «Львівська політехніка»,
[1]viktor.v.kolchenko@lpnu.ua

Стрімкий розвиток великих мовних моделей (LLM) та їх інтеграція в інтелектуальних агентів створили нові вектори загроз, серед яких витік системних інструкцій (Prompt Leakage) стає однією з найбільш критичних проблем безпеки. Ця вразливість полягає в ненавмисному розкритті прихованих налаштувань, які визначають логіку поведінки, обмеження безпеки та операційні параметри моделі [1]. В сучасній екосистемі ШІ системні інструкції перетворюються на найцінніший актив інтелектуальної власності, що створює прямі стимули для їх викрадення.

Мета дослідження полягає у проведенні аналізу атак на витік системних інструкцій, дослідженні внутрішніх механізмів їх реалізації для забезпечення безпеки інтелектуальної власності в екосистемах ШІ-агентів.

Системні інструкції у LLM функціонують як прихований шар управління, що задає роль моделі, її тон та межі дозволеної взаємодії. У сучасних агентних системах вони еволюціонують від простих текстових промптів до модульних пакетів навичок, які поєднують робочі процеси, використання інструментів та специфічні доменні знання. Сутність атак типу Prompt Leakage базується на експлуатації фундаментальної здатності моделей до повторення контексту, що

є необхідним для виконання корисних завдань, таких як узагальнення тексту. Зловмисники намагаються обійти фільтри безпеки та змусити модель «забути» про заборону на розголошення інструкцій, використовуючи неоднозначність між даними користувача та керуючими командами. Особливо небезпечними є багатоходові діалоги, де використовується схильність моделі погоджуватись з користувачем та «flip-flop» ефект, що дозволяє підвищити успішність витоку з 17,7% до понад 86% [2].

Методи витоку включають різноманітні стратегії, починаючи з евристичних атак, таких як відновлення здатності моделі до повторення контексту через вгадування початкових токенів (наприклад, «You are ChatGPT») [3]. Більш складні агентні підходи використовують навчання з підкріпленням та кооперативні команди агентів для автоматизованого пошуку вразливостей цільової моделі [1]. Окрему категорію становлять атаки через сторонні канали, такі як PROMPTPEEK, що експлуатують механізм спільного використання KV-cache у багатокористувацьких середовищах для покрокового відновлення токенів чужих запитів [4]. Також поширюються методи «крадіжки навичок», де зловмисники використовують рольові сценарії (наприклад, роль адміністратора) та ін'єкцію «ланцюжка думок» для ексфільтрації пропрієтарних алгоритмів [5].

Методи захисту еволюціонують у напрямку багатошарової оборони. Програмні методи включають «захист сендвічем», дублювання інструкцій, XML-тегування та переписування запитів для видалення шкідливих компонентів [1]. Проте більш надійними є архітектурні рішення, зокрема SysVec (System Vectors), що пропонує кодувати системні промпти як внутрішні вектори активації, повністю видаляючи їх із текстового контексту, що робить їх недоступними для прямого копіювання [3]. Іншим інноваційним підходом є пробінг інтентів, який дозволяє виявити намір моделі здійснити витік через аналіз прихованих станів останнього токена вхідної послідовності ще до початку генерації відповіді з точністю понад 90% [1]. Для вихідної фільтрації застосовуються системи, які поєднують семантичний аналіз за допомогою потужних моделей-суддів із перевіркою посимвольного збігу [5].

Попри ці заходи, існують суттєві проблеми та обмеження методів захисту. Стохастична природа LLM та відсутність чіткої межі між інструкціями та даними роблять повне запобігання витокам наразі недосяжним. Більшість методів аналізу внутрішніх станів вимагають доступу до моделі за принципом «білої скриньки», що неможливо для розробників, які працюють через сторонні API [3]. Крім того, адаптивні атаки через переклад на інші мови або складне перефразування часто обходять фільтри схожості, а жорсткі обмеження можуть негативно впливати на корисність моделі та швидкість її відповіді [1].

Аналіз витоків системних інструкцій підкреслює необхідність впровадження ешелонованої оборони. Вона повинна поєднувати мінімізацію привілеїв агентів, архітектурне приховування інструкцій, безперервний моніторинг аномалій у внутрішніх станах та ретельну фільтрацію вихідних даних [3, 5]. Тільки комплексна комбінація цих підходів дозволить забезпечити стійкість комерційних ШІ-систем до сучасних автоматизованих атак.

1. Sternak T., Runje D., Granoša D., Wang C. Automating Prompt Leakage Attacks on Large Language Models Using Agentic Approach – 2025. – URL: https://arxiv.org/pdf/2502.12630
2. Agarwal D., Fabbri A., Risher B., Laban P., Joty S., Wu C.-S. Prompt Leakage Effect and Mitigation Strategies for Multi-turn LLM Applications // Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track. – Miami, Florida, USA, 2024. – C. 1255–1275. – URL: https://aclanthology.org/2024.emnlp-industry.94/
3. Cao B., Li C., Cao Y., Ge Y., Wang T., Chen J. You Can't Steal Nothing: Mitigating Prompt Leakages in LLMs via System Vectors // Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security (CCS '25). – Taipei, Taiwan, 2025. – URL: https://arxiv.org/abs/2509.21884
4. Wu G., Zhang Z., Zhang Y., Wang W., Niu J., Wu Y., Zhang Y. I Know What You Asked: Prompt Leakage via KV-Cache Sharing in Multi-Tenant LLM Serving // Network and Distributed System Security Symposium (NDSS) 2025. – San Diego, CA, USA, 2025. – URL: https://www.ndss-symposium.org/ndss-paper/i-know-what-you-asked-prompt-leakage-via-kv-cache-sharing-in-multi-tenant-llm-serving/
5. Wang Z., Zhang R., Liu Y., Liu C., Zhao Q., Li H., Xu G. Black-Box Skill Stealing Attack from Proprietary LLM Agents: An Empirical Study // arXiv. – 2026. – URL: https://arxiv.org/abs/2604.21829

Еволюція стратегії ЄС щодо протидії іноземному втручанню та маніпулюванню інформацією (FIMI)

УДК 327:004.056

Сергій Кондратюк

Державний університет інформаційно-комунікаційних технологій, s.kondratiuk@duikt.edu.ua

У період 2015-2026 рр. стратегія Європейського Союзу у сфері протидії іноземним маніпуляціям і втручанням (FIMI) пройшла глибоку еволюцію — від поодиноких реакцій на дезінформаційні кампанії до створення узгодженої, багаторівневої екосистеми проактивної протидії. У березні 2015 р. на саміті ЄС у Брюсселі було ухвалене рішення про необхідність протидії постійним дезінформаційним кампаніям Росії, що зумовило створення East StratCom Task Force у межах Європейської служби зовнішніх справ (EEAS).

До 2022 р. на рівні ЄС було напрацьовано базову систему захисту від FIMI. Вона включала Платформу EUvsDisinfo; План дій проти дезінформації 2018 р., який формалізував FIMI як гостру внутрішню загрозу демократії ЄС; та Кодекс практики щодо дезінформації. Важливим етапом стало створення у середині 2016 р. Центру аналізу гібридних загроз ЄС (EU Hybrid Fusion Cell) у складі EEAS для аналізу розвідувальної інформації та OSINT. У квітні 2017 р. в Гельсінкі засновано Європейський центр передового досвіду з протидії гібридним загрозам (Hybrid CoE) — міжнародний хаб для країн ЄС та НАТО.

Внаслідок повномасштабної російської агресії проти України у 2022 р. гібридні загрози стрімко загострилися. FIMI еволюціонували у складні багатовимірні операції. Ключовим чинником стало використання Кремлем концепції «асиметрії витрат» — організація надзвичайно дешевих інформаційних кампаній із залученням ботів, місцевих проксі та ШІ, для захисту від яких потрібні були колосальні ресурси з боку європейських урядів. Крім того, у 2025–2026 рр. посилилася синергія російських дестабілізаційних кампаній та китайських «тактик відхилення». Важливим тригером змін стали резолюції Європарламенту (зокрема Russiagate–2024), які викрили внутрішні вразливості, механізми «корумпування» європейських політичних еліт та нелегальне фінансування радикальних партій.

Після 2022 р. ЄС перейшов до комплексної, проактивної стратегії, в основі якої лежать такі механізми:

1.Координаційна рамка та оперативне реагування. Запроваджений у 2022 р. інструментарій EU Hybrid Toolbox діє як координаційна рамка для узгодження застосування різних механізмів, таких як Cyber Diplomacy Toolbox (кіберсанкції) та заходи протидії FIMI. У травні 2024 р. Рада ЄС затвердила Гібридні команди швидкого реагування (EU Hybrid Rapid Response Teams, HRRTs) для надання оперативної вузькопрофільної допомоги державам-членам, які вперше були розгорнуті у Молдові напередодні виборів 2025 р.

2.Двоєдина система внутрішньої безпеки. З 2025 р. захист від FIMI забезпечується комплексним підходом. Стратегія ProtectEU (квітень 2025) відповідає за «жорстку» безпеку: захист критичної інфраструктури, протидію саботажу та використанню організованої злочинності. Водночас Європейський щит демократії (European Democracy Shield, листопад 2025) забезпечує «м'яку» безпеку, створюючи Європейський центр демократичної стійкості, захищаючи інформаційний простір і незалежні медіа.

3.Доктрина стримування FIMI. Концептуальним зрушенням, закріпленим у 4-му звіті EEAS (березень 2026), стало впровадження Доктрини стримування (FIMI Deterrence Playbook). Це парадигмальний зсув від простого спростування фейків до демонтажу самої тіньової інфраструктури. ЄС спрямовує санкції та правоохоронні інструменти проти посередників, PR-агентств та розробників ботнетів, атакуючи ланцюги постачання, щоб зробити гібридні операції фінансово невигідними.

4.Аналітика на базі ШІ. Європейська служба зовнішніх справ активно використовує системи штучного інтелекту для моніторингу та ідентифікації патернів скоординованої неавтентичної поведінки в реальному часі. Однак зловмисники також масштабують генерацію фейків (у 2025 р. 27% проаналізованих інцидентів містили ШІ-контент), перетворюючи це на постійну технологічну гонку.

Висновки. Еволюція стратегії Європейського Союзу відображає перехід від фрагментарних реактивних ініціатив до комплексної моделі проактивного стримування. Сучасна архітектура поєднує юридичну відповідальність платформ, єдині координаційні протоколи та розбудову соціальної стійкості. Протидія маніпуляціям інформацією сьогодні стала фундаментальною

складовою захисту демократичної цілісності та безпеки всієї європейської спільноти.

1. European Council Conclusions on external relations (19 March 2015). https://www.consilium.europa.eu/en/press/press-releases/2015/03/19/conclusions-russia-ukraine-european-council-march-2015/
2. FIMI and disinformation as global threats. EUvsDisinfo, 30.01.2026. https://euvsdisinfo.eu/fimi-and-disinformation-as-global-threats/
3. 4th EEAS Report on Foreign Information Manipulation and Interference Threats : [Annual report] / European External Action Service (EEAS). Brussels, 2026. 12 March. URL: https://www.eeas.europa.eu/sites/default/files/2026/documents/EEAS%204th%20Threat%20Report_web%20version_1.pdf
4. European Commission. Communication from the Commission on ProtectEU: a European Internal Security Strategy. COM(2025) 148 final. April 2025.
5. European Commission. Joint Communication on the European Democracy Shield. November 2025.
6. Annual Report 2025 / The Soufan Center. New York, 2026. 24 p. URL: thesoufancenter.org (дата звернення: 10.05.2026).
7. EUvsDisinfo. (2020, April 22). “To Challenge Russia’s Ongoing Disinformation Campaigns”: The Story of EUvsDisinfo. https://euvsdisinfo.eu/to-challenge-russias-ongoing-disinformation-campaigns-the-story-of-euvsdisinfo/

Методологія збагачення подій SIEM результатами аналізу мережевого трафіку засобами машинного навчання

УДК 004.056.5

Юрій Коровайченко[1], Євгеній Педченко[2], Сергій Гахов[3]

Державний університет інформаційно-комунікаційних технологій,
[1]y.korovaichenko@duikt.edu.ua, [2]e.pedchenko@duikt.edu.ua,
[3]s.gakhov@duikt.edu.ua

Операційні центри безпеки в реальній корпоративній інфраструктурі стикаються з протиріччям: платформи моніторингу подій безпеки, такі як SIEM (IBM QRadar, Wazuh, Rapid7 InsightIDR, Palo Alto XSIAM) - консолідують сотні тисяч, або навіть мільйони подій на добу, але реальне покриття мережевого рівня при цьому залишається неповним. Більшість кореляційних правил спирається на стандартні джерела подій, а саме: журнали кінцевих точок та застосунків. Мережевий потік або передається як сирий NetFlow без подальшої обробки, або взагалі залишається поза полем кореляційного аналізу. На практиці це означає, що аналітик першої лінії вимушений працювати з

інтерфейсами мінімум двох рішень: SIEM та NDR і вручну зіставляти те, що система SIEM могла б зробити автоматично.

Метою даної роботи є обґрунтування методології збагачення подій структурованими метаданими машинного навчання в кореляційну логіку SIEM-систем для підвищення ефективності виявлення кіберзагроз в операційних центрах безпеки.

Наявний досвід інтеграції аналізу машинним навчанням мережевого трафіку з платформами моніторингу вже накопичений: відомі підходи на базі Zeek у поєднанні з машинним навчанням в Elastic Security, а також передача алертів від NDR-рішень до SIEM через Syslog або API. Проте в більшості реалізацій зовнішній компонент машинного навчання залишається ізольованим: до платформи надходить готовий алерт без структурованих метаданих класифікації: без оцінки впевненості, без інформації про ознаки, що обумовили рішення про подію. Наявні інтеграції, наприклад, у рішеннях Flowmon і Corelight з тими самими SIEM-платформами працюють за схожою логікою. Це суттєво обмежує можливості побудови складних кореляційних правил, які враховували б не лише факт виявленої аномалії, а й її характеристики.

На відміну від існуючих підходів, запропонована методологія будується на ідеї перенесення контексту машинного навчання всередину події SIEM. Джерелом даних слугують структуровані метадані мережевих сесій. зокрема JSON-логи Corelight на базі Zeek-аналізатора, або записи потоків Flowmon з поведінковими мітками. Міжмережеві екрани, наприклад, Palo Alto Networks NGFW з розширеним логуванням сесій, також можуть бути задіяні як додаткове джерело контексту. З цих даних формується ознаковий вектор: тривалість з'єднання, співвідношення вхідного/вихідного трафіку, кількість пакетів, розподіл їхніх розмірів, значення TTL, ентропія корисного навантаження. Перелік ознак уточнюється за результатами feature importance у процесі навчання.

Класифікацію в запропонованій методології виконує модель Random Forest. Цю модель обрано, насамперед, через стійкість до незбалансованих класів, що типово для трафіку в SOC, та можливість відстежити, які саме ознаки вплинули на конкретне рішення моделі. Швидкодія в умовах потокової обробки також вкладається в допустимі межі. Модель повертає мітку класу та оцінку впевненості, яка може бути безпосередньо використана як динамічний поріг у кореляційних правилах.

Вибір ознак ґрунтується на їхній здатності відображати поведінкові характеристики сесії незалежно від корисного навантаження: тривалість, байтове співвідношення та розподіл розмірів пакетів дозволяють виявляти аномалії на рівні патерну з'єднання, тоді як ентропія навантаження та значення TTL чутливі до тунелювання і спуфінгу відповідно. Остаточний склад вектора визначається за результатами аналізу feature importance на навчальній вибірці.

Етап збагачення події в SIEM реалізується по-різному залежно від платформи: для IBM QRadar - через Custom Properties або DSM-розширення, для Wazuh - через декодери з полями типу data.ml.*, для InsightIDR та XSIAM - через API-інтеграції та custom log parsers. Підсумок: кореляційні правила отримують поля ml_class, ml_score і ml_features_summary безпосередньо в тілі

події. Жодного перемикання між консолями, аналітик бачить висновок системи машинного навчання там, де і всі інші події, а система аналізує збагачені дані.

Попередні лабораторні тести на відкритому датасеті CIC-IDS2017 підтвердили принципову працездатність підходу та прийнятні метрики класифікації для більшості представлених класів атак. Повна валідація методології в умовах реального виробничого середовища запланована на наступному етапі дослідження. Відповідно, кількісні оцінки впливу на середній час виявлення наразі не наводяться.

В роботі представлено методологію збагачення подій структурованими метаданими машинного навчання в кореляційну логіку SIEM-систем, що базується на класифікації мережевого трафіку моделлю Random Forest з поверненням мітки класу та оцінки впевненості. Ознаковий вектор формується з мережевих метаданих сесій, а результати класифікації передаються в тіло події через поля ml_class, ml_score та ml_features_summary. Це усуває необхідність ручного зіставлення даних між консолями та створює передумови для скорочення часу виявлення загроз.

1. Sharafaldin I., Habibi Lashkari A., Ghorbani A.A. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP). 2018. P. 108–116. URL: https://www.scitepress.org/Papers/2018/66398/
2. Breiman L. Random Forests. Machine Learning. 2001. Vol. 45. P. 5–32. URL: https://link.springer.com/article/10.1023/A:1010933404324
3. Palo Alto Networks. XSIAM Platform Architecture. Technical Whitepaper, Palo Alto Networks. 2024. URL: https://www.paloaltonetworks.com/resources/techbriefs/cortex-xsiam

The Eastin-Knill Theorem: Fundamental Limitations of Quantum Fault Tolerance

UDK 621.395.7 (043.2)

Yevgen Kotukh[1]

Yevhenii Bereznyak Military Academy, [1]yevgenkotukh@gmail.com

The construction of a large-scale universal quantum computer remains constrained by the inherent fragility of quantum states, which are continuously subject to decoherence, dephasing and energy dissipation. Reliable quantum information processing therefore relies on quantum error correction (QEC) codes, in which a single logical qubit is encoded into an entangled state of many physical qubits. Within the framework of fault-tolerant quantum computing (FTQC), logical operations on encoded data must be implemented in a manner that prevents the uncontrolled propagation of physical errors. The principal mechanism that ensures this property is transversality: a transversal logical gate acts as a tensor product of local unitaries on the constituent subsystems, guaranteeing that a single physical fault propagates to at most one qubit per code block [1].

In 2009 B. Eastin and E. Knill established a no-go theorem stating that no nontrivial finite-dimensional quantum error correcting code can support a continuous group of transversal logical gates [1]. An immediate corollary is that no finite-dimensional QEC code admits a universal set of logical gates implemented exclusively by transversal operations. This result imposes a fundamental constraint on every known approach to scalable FTQC and motivates a broad research program aimed at characterizing, quantifying and circumventing this limitation [2,3]. A QEC code is defined as an isometric embedding of a logical Hilbert space into the Hilbert space of a composite physical system. A transversal logical operator is, by definition, a tensor product of local unitaries acting on individual subsystems or fixed-size groups of subsystems. The original argument of Eastin and Knill proceeds by considering the intersection of an arbitrary Lie group of logical unitaries with the group of locality-preserving operators on the code. This intersection is a closed subset of a Lie group and, by Cartan's closed-subgroup theorem, is itself a Lie subgroup [1].

The infinitesimal generators of any continuous transversal symmetry must therefore be expressible as sums of local Hermitian operators. The Knill-Laflamme conditions, however, require that every correctable local operator act on the code subspace as a scalar multiple of the identity, since otherwise local degrees of freedom would carry information about the logical state. Consequently, the Lie algebra of transversal logical operators reduces to global phases, the corresponding Lie group has dimension zero, and the set of transversal logical gates is a finite discrete subgroup of the unitary group. Universality, which requires a dense subset of the unitary group, is therefore unattainable by transversal means alone in any finite-dimensional code that corrects local errors [1,3]. The Eastin-Knill theorem admits a natural reformulation in the language of covariant quantum codes, i.e. codes for which a continuous symmetry transformation on the logical system is implemented by a symmetry transformation on the physical system.

Table 1

Comparison of approximate Eastin-Knill bounds for covariant QEC against local erasure noise

Bound / construction	Symmetry group	Scaling of infidelity ε	Reference
Original no-go (exact)	Continuous $\{U(1), SU(2), \ldots\}$	$\varepsilon = 0$ is forbidden	[1]
Representation-theoretic bound	Compact Lie group	$\varepsilon \geq \Omega(1/n)$	[3]
Metrological bound (QFI)	$U(1)$	$\varepsilon \geq (\Delta H_1)^2/(4F_Q)$	[4]
Thermodynamic code	$U(1)$	$\varepsilon \sim 1/n$ (erasure)	[4,5]
Single-shot min-entropy bound	Universal unitary	Necessary and sufficient	[6]

In this formulation the theorem states that exact covariant codes with respect to a continuous symmetry cannot correct local erasure exactly in finite dimensions [2,4]. In [4] recast covariant QEC as a quantum metrological protocol in which the estimation of an unknown rotation angle on the logical system is mapped to the estimation of the corresponding angle on the physical system, and derived analytic lower bounds on the worst-case entanglement infidelity ε in terms of the regularized symmetric-logarithmic-derivative quantum Fisher information of the noise channel [4]

$$\varepsilon \geq \ell_1\left(\frac{(\Delta H_L)^2}{4F_S^{reg}(N_S, H_S)}\right), \quad (1)$$

where $\ell_1(x) = \left(1+4x-\sqrt{(1+4x)}\right)/\left(2(1+4x)\right)$, $(\Delta H_L)^2$ is the variance of the logical Hamiltonian and $4F_S^{reg}$ is the regularized quantum Fisher information of the physical noise channel.

The bound holds whenever the Hamiltonian-in-Kraus-span condition is satisfied and is asymptotically saturated by a family of so-called thermodynamic codes for erasure noise [4,5]. An independent representation-theoretic approach in [3] established an approximate Eastin-Knill theorem of the form $\varepsilon \geq \Omega(1/n)$ for codes admitting a universal transversal action of a compact Lie group on n physical subsystems [3]. Since universality cannot be achieved by transversal gates alone in any single finite-dimensional code, fault-tolerant architectures combine transversal Clifford operations with additional resources that lift the gate set to universality. Three principal strategies are established in recent papers. First, magic-state distillation provides high-fidelity ancillary states that, together with transversal Clifford gates and gate teleportation, realize a non-Clifford gate such as the T gate; in conventional surface-code architectures this procedure has been estimated to account for a substantial fraction of the total qubit overhead [7,8]. Second, code switching transfers logical information between two codes whose transversal gate sets are mutually complementary, jointly spanning a universal set without resorting to distillation [9]. Third, concatenated and triorthogonal code constructions implement universal fault-tolerant gates by combining transversal operations at different concatenation levels with intermediate error correction [10]. Recent experimental work has demonstrated that these strategies are reaching the regime of practical fault tolerance. In 2025 high-fidelity logical magic states were prepared via code switching in the two-dimensional color code, with logical error rates comparable to or below the underlying two-qubit physical gate error rate, completing a universal fault-tolerant gate set together with previously demonstrated transversal Cliffords, state preparation and measurement [11].

Conclusions. The Eastin-Knill theorem constitutes one of the central structural results of fault-tolerant quantum computing. Its rigorous mathematical content that finite-dimensional codes correcting local errors admit only a discrete group of transversal logical unitaries has been substantially refined over the past five years through the development of approximate and covariant formulations, quantitative metrological bounds, and explicit code constructions that nearly saturate these bounds. The combined progress in theory and experimental realization of magic-state preparation and code switching indicates that the

practical implications of the theorem, while fundamental, do not preclude scalable universal fault-tolerant quantum computation.

1. 1. Eastin B., Knill E. Restrictions on transversal encoded quantum gate sets. Physical Review Letters, 2009, vol. 102, no. 11, art. 110502.
2. 2. Hayden P., Nezami S., Popescu S., Salton G. Error correction of quantum reference frame information. PRX Quantum, 2021, vol. 2, art. 010326.
3. 3. Faist P., Nezami S., Albert V. V., Salton G., Pastawski F., Hayden P., Preskill J. Continuous symmetries and approximate quantum error correction. Physical Review X, 2020, vol. 10, art. 041018.
4. 4. Zhou S., Liu Z.-W., Jiang L. New perspectives on covariant quantum error correction. Quantum, 2021, vol. 5, art. 521.
5. 5. Kubica A., Demkowicz-Dobrzański R. Using quantum metrological bounds in quantum error correction: a simple proof of the approximate Eastin-Knill theorem. Physical Review Letters, 2021, vol. 126, art. 150503.
6. 6. Alexander R. A new approximate Eastin-Knill theorem. npj Quantum Information, 2025, vol. 11, art. 156.
7. 7. Bravyi S., Kitaev A. Universal quantum computation with ideal Clifford gates and noisy ancillas. Physical Review A, 2005, vol. 71, art. 022316.
8. 8. Litinski D. Magic state distillation: not as costly as you think. Quantum, 2019, vol. 3, art. 205.
9. 9. Yoder T. J., Takagi R., Chuang I. L. Universal fault-tolerant gates on concatenated stabilizer codes. Physical Review X, 2016, vol. 6, art. 031039.
10. 10. Bravyi S., Haah J. Magic-state distillation with low overhead. Physical Review A, 2012, vol. 86, art. 052329.
11. 11. Quantinuum collaboration. Experimental demonstration of high-fidelity logical magic states from code switching. arXiv:2506.14169, 2025.

Пояснюване AI/ML-виявлення аномалій у мікросервісних та мультихмарних середовищах

УДК 004.056.5:004.75:004.8

Віталій Криворучко[1]

Державний університет інформаційно-комунікаційних технологій,
[1]hamandes@gmail.com

У мікросервісних та мультихмарних середовищах засоби моніторингу формують великі обсяги телеметрії: метрики ресурсів, мережеві події, журнали доступу, HTTP-коди помилок та дані про взаємодію сервісів. Традиційний пороговий контроль часто виявляє лише факт відхилення і не пояснює, чому стан є ризиковим. Для кібербезпеки це ускладнює розмежування інциденту, деградації сервісу та звичайного коливання навантаження [1].

Мультихмарне середовище розглядається як сукупність сервісів, розгорнутих у різних хмарних доменах, регіонах або провайдерах. Метою роботи є розроблення підходу до пояснюваного AI/ML-виявлення аномалій,

який поєднує аналіз телеметрії, графову модель взаємодії сервісів та механізм формування пояснення для аналітика безпеки.

Наукова новизна полягає у переході від ізольованої оцінки метрик до комбінованої оцінки інцидентного стану вузла за трьома компонентами: невідповідність телеметрії, зміна топологічного оточення та ймовірність ризикової події.

Схема не обмежується бінарним рішенням, а формує пояснення типу «вузол - ознака - зв'язок», що наближує модель до використання у SOC/DevSecOps-процесах [2, 3].

Стан вузла v у момент часу t описується вектором x(v,t)={r_cpu, r_mem, r_io, r_net, l_p95, e_5xx, d_auth}, де всі ознаки нормовано до шкали [0;1]. Інфраструктура подається графом G_t=(V_t,E_t,X_t), де V_t - множина сервісів, E_t - залежності між ними, X_t - матриця атрибутів вузлів. Реконструкційна модель формує оцінку x_hat(v,t), після чого локальна похибка визначається як

$$e_{rec}(v,t) = \frac{\| x(v,t) - \hat{x}(v,t) \|_2}{\sqrt{m}},$$

де m - кількість ознак. Щоб похибка була сумісною з іншими складовими ризику, вона нормується через 95-й перцентиль похибок нормального режиму:

$$E_{rec}(v,t) = min\left(1; \frac{e_{rec}(v,t)}{q_{95}}\right).$$

Топологічне відхилення визначається через відстань Жаккара між поточним оточенням вузла N_t(v) та еталонним оточенням N_ref(v):

$$D_{top}(v,t) = 1 - \frac{\left|N_t(v) \cap N_{ref}(v)\right|}{\left|N_t(v) \cup N_{ref}(v)\right|}.$$

Ймовірність інцидентного стану позначається як R_inc(v,t) і може бути отримана з класифікатора або каліброваної моделі ризику. Підсумкова оцінка ризиковості вузла задається формулою

$$S(v,t) = \lambda_1 E_{rec}(v,t) + \lambda_2 D_{top}(v,t) + \lambda_3 R_{inc}(v,t),\ \lambda_1 + \lambda_2 + \lambda_3 = 1.$$

де всі складові знаходяться у межах [0;1]. Вузол вважається ризиковим, якщо S(v,t) ≥ θ. Для оцінювання пояснюваності введено показник EC=N_full/N_detected, де N_full - кількість виявлених інцидентів, для яких сформовано повне пояснення «вузол - ознака - зв'язок», а N_detected - загальна кількість виявлених інцидентів. Це дозволяє оцінити не лише точність, а й практичну корисність моделі.

У контрольному сценарії підхід підвищив F1-score з 0,78 до 0,89, зменшив нормований рівень хибних спрацювань на 26% та скоротив час інтерпретації інциденту на 32%. Значення EC=0,86 означає, що для 86% ризикових станів сформовано повне пояснення.

Таблиця 1

Оцінювання результатів виявлення ризикових станів

Показник	Базовий моніторинг	Запропонований підхід
F1-score	0,78	0,89
False Positive Rate	1,00	0,74
Час інтерпретації інциденту	1,00	0,68
Explainability Coverage	-	0,86

Висновки. Запропоновано формалізовану схему пояснюваного AI/ML-виявлення аномалій у мікросервісних та мультихмарних середовищах. Нормовані компоненти E_rec, D_top та R_inc роблять оцінку S(v,t) інтерпретованою і придатною для порівняння вузлів. На практиці це зменшує хибні спрацювання, підвищує якість виявлення інцидентів і пояснює причини ризикового стану.

1. Chandola V., Banerjee A., Kumar V. Anomaly detection: A survey. ACM Computing Surveys. - 2009. - Vol. 41, №3. - P. 1-58.
2. Ahmed M., Mahmood A.N., Hu J. A survey of network anomaly detection techniques. Journal of Network and Computer Applications. - 2016. - Vol. 60. - P. 19-31.
3. Ying R., Bourgeois D., You J., Zitnik M., Leskovec J. GNNExplainer: Generating Explanations for Graph Neural Networks. Advances in Neural Information Processing Systems. - 2019. - P. 9240-9251.

Метрикове оцінювання зменшення технічного боргу JavaScript-коду як передумови підвищення безпеки програмних систем

УДК 004.41:004.8

Ірина Замрій [1], Олексій Кулаков[2]

Державний університет інформаційно-комунікаційних технологій,
[1]i.zamrii@duikt.edu.ua, [2]o.kulakov@stud.duikt.edu.ua

У сучасних JavaScript-проєктах технічний борг проявляється як зниження зв'язності модулів, зростання міжмодульних залежностей та ускладнення трасування змін. У контексті кібербезпеки такі властивості є непрямими чинниками ризику, оскільки ускладнюють аудит коду, статичний аналіз, локалізацію потенційно вразливих ділянок і безпечне внесення виправлень. Роботи з LLM-рефакторингу підкреслюють перспективність автоматизованих змін, але вказують на потребу їх метрикової та функціональної валідації [1–3].

Метою роботи є розроблення компактної процедури метрикового оцінювання зменшення технічного боргу JavaScript-коду після LLM-згенерованих змін для підвищення безпеки програмних систем. Наукова новизна полягає у використанні інтегрального показника, який поєднує зв'язність, залежність і трасувальну невпорядкованість, а рішення про прийняття зміни пов'язує з додатним приростом якості. Такий підхід

узгоджується з розвитком спеціалізованих моделей для коду [4]. Для кількісної перевірки прийнятності кожної LLM-зміни технічний борг подано як інтегральний індекс $S(k)$, що агрегує три нормовані характеристики структури коду:

$$S(k) = 0{,}4 \cdot (1 - Coh(k)) + 0{,}4 \cdot Dep(k) + 0{,}2 \cdot TrEnt(k), \quad (1)$$

де $S(k)$ − інтегральний індекс технічного боргу на k-тій ітерації, $Coh(k)$ − внутрішня зв'язність, $Dep(k)$ − нормована залежність між модулями, $TrEnt(k)$ − ентропія трасування вимог або тестів до коду. Менше значення $S(k)$ відповідає нижчому рівню боргу та кращій структурній готовності коду до аудиту, тестування і безпечного супроводу. Така інтерпретація узгоджується з практикою вимірювання підтримуваності [5].

Експериментальна процедура передбачала такі етапи: 1) обчислення початкових метрик; 2) формування промпту з описом структурних недоліків; 3) отримання LLM-рекомендацій; 4) внесення змін без зміни зовнішньої поведінки; 5) повторне обчислення $S(k)$. Для перевірки використано відкриті бібліотеки axios, commander.js і qs-main.

Таблиця 1

Зміна інтегрального індексу технічного боргу

Проєкт	$Coh_0 \rightarrow Coh_2$	$Dep_0 \rightarrow Dep_2$	$S_0 \rightarrow S_2$	Зменшення S
axios	0,252→0,254	0,101→0,098	0,348→0,346	0,5%
commander.js	0,082→0,097	0,385→0,353	0,521→0,502	3,6%
qs-main	0,104→0,120	0,125→0,077	0,408→0,383	6,2%

Отримані значення показують, що вплив LLM-згенерованих змін залежить від початкового стану системи. Для axios, який має відносно впорядковану структуру, зменшення S є мінімальним. Для commander.js та qs-main ефект виразніший: зменшується Dep, а Coh зростає. У безпековому контексті це означає послаблення непрямих передумов кіберризику, пов'язаних із надмірною зв'язаністю компонентів.

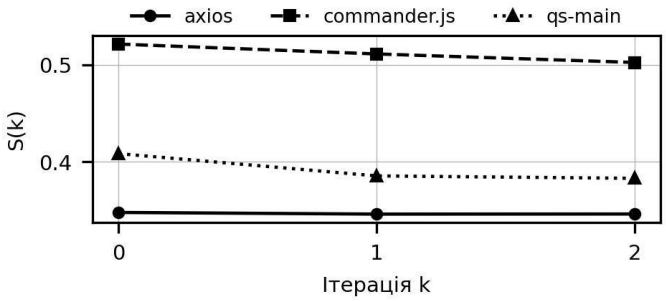


Рис. 1. Динаміка інтегрального індексу $S(k)$ після LLM-згенерованих змін

Практична значущість підходу полягає в тому, що LLM-зміни не приймаються автоматично, тобто кожна ітерація проходить кількісну перевірку,

а негативний або нульовий приріст $\Delta S(k) = S(k-1) - S(k)$ є підставою для відхилення зміни. Це зменшує ризик формального рефакторингу, коли код виглядає простішим локально, але створює нові залежності або ускладнює безпековий аудит.

Висновки. Запропонована процедура дозволяє формалізувати зменшення технічного боргу як вимірюваний процес, у якому LLM виконує роль генератора кандидатних змін, а метрики − роль критерію прийняття. Пілотні результати засвідчили монотонне або майже монотонне зменшення $S(k)$, причому найбільший ефект спостерігається у проєктах з вищою початковою залежністю модулів. Отже, підхід може розглядатися як допоміжний засіб підвищення кіберстійкості, оскільки спрощення структури коду полегшує аудит, тестування і контроль безпечних змін.

1. Martinez S., Xu L., Elnaggar M., Alomar E.A. Software Refactoring Research with Large Language Models: A Systematic Literature Review. Journal of Systems and Software. 2025. URL: https://www.sciencedirect.com/science/article/abs/pii/S0164121225004315.
2. Tornhill A., Borg M., Hagatulah N., Söderberg E. ACE: Automated Technical Debt Remediation with Validated Large Language Model Refactorings. FSE Companion '25: Proceedings of the 33rd ACM International Conference on the Foundations of Software Engineering. 2025. P. 1318–1324. DOI: 10.1145/3696630.3730565.
3. Cordeiro J., Noei S., Zou Y. LLM-Driven Code Refactoring: Opportunities and Limitations. 2025 IEEE/ACM Second IDE Workshop (IDE). 2025. P. 32–36. DOI: 10.1109/IDE66625.2025.00011.
4. Rozière B., et al. Code Llama: Open Foundation Models for Code. arXiv:2308.12950. 2023. URL: https://arxiv.org/abs/2308.12950.
5. SonarSource. Understanding measures and metrics. SonarQube Server Documentation. 2026. URL: https://docs.sonarsource.com/sonarqube-server/user-guide/code-metrics/metrics-definition.

Ризик-орієнтований підхід до захисту IoT-пристроїв у муніципальних системах

УДК 004.056:004.738.5:352.07

Олександр Голотенко[1], Сергій Кульчицький[2], Данило Стухляк[3]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]golotenko@gmail.com, [2]serhii_kulchytskyi0212@tntu.edu.ua,
[3]itaniumua@gmail.com

Активне впровадження IoT-пристроїв у муніципальних системах є важливим напрямом цифрової трансформації громад. Такі пристрої застосовуються для екологічного моніторингу, керування освітленням, обліку ресурсів, контролю енергоспоживання, транспортних потоків і роботи комунальних об'єктів. Водночас їхня значна кількість, територіальна

розподіленість, бездротові канали зв'язку, обмежені обчислювальні ресурси та фізична доступність створюють підвищені кіберризики. За цих умов однаковий рівень захисту для всіх компонентів є неефективним, оскільки не враховує різної критичності пристроїв і можливих наслідків інциденту.

Метою роботи є обґрунтування ризик-орієнтованого підходу до захисту IoT-пристроїв у муніципальних системах, який дає змогу визначати пріоритетність заходів кібербезпеки з урахуванням критичності пристрою, рівня експозиції, потенційного впливу компрометації та наявних механізмів захисту. ISO/IEC 27005 розглядає управління ризиками як процес ідентифікації, аналізу, оцінювання та оброблення ризиків [1].

NIST Cybersecurity Framework 2.0 доповнює цю логіку безперервним циклом ідентифікації активів, захисту, виявлення подій, реагування, відновлення та управління ризиками [2].

Наукова новизна підходу полягає в адаптації ризик-орієнтованого управління до муніципального IoT-середовища, де пріоритет захисту визначається не лише технічними параметрами пристрою, а й його роллю у функціонуванні конкретного міського сервісу.

Сенсор температури в приміщенні, лічильник енергоспоживання на комунальному об'єкті та контролер вуличного освітлення можуть належати до одного класу IoT, але мати різний рівень критичності й різні вимоги до захисту.

Таблиця 1

Критерії ризик-орієнтованого захисту IoT-пристроїв

Критерій	Зміст оцінювання	Приклад захисного рішення
Критичність пристрою	вплив на роботу муніципального сервісу	пріоритетне резервування та моніторинг
Рівень експозиції	доступність із мережі або фізичного середовища	сегментація, VPN, обмеження портів
Вразливість конфігурації	паролі, прошивка, відкриті сервіси	оновлення, унікальні облікові дані
Потенційний вплив	наслідки для даних, процесів і безперервності	журналювання, резервне відновлення
Наявний захист	поточні технічні й організаційні заходи	посилення контролю доступу

Запропонований підхід передбачає п'ять етапів: інвентаризацію IoT-пристроїв, класифікацію за критичністю, ідентифікацію загроз і вразливостей, оцінювання ризику та визначення пріоритетних захисних заходів.

Для кожного пристрою доцільно фіксувати призначення, місце встановлення, тип підключення, протоколи обміну, відповідальний підрозділ і пов'язаний муніципальний сервіс.

Рівень ризику може визначатися через критичність, імовірність реалізації загрози, експозицію та потенційний вплив інциденту за експертною шкалою від 1 до 5 балів.

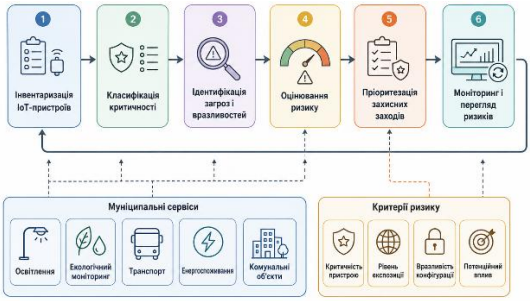


Рис.1. Ризик-орієнтований підхід до захисту IoT-пристроїв у муніципальних системах

З урахуванням сучасного ландшафту загроз особливої уваги потребують атаки на доступність, експлуатація відомих вразливостей, компрометація облікових даних і підміна даних. Можемо визначити загрози даним, атаки на доступність, соціальну інженерію та експлуатацію вразливостей серед провідних напрямів кіберзагроз Отже, ризик-орієнтований підхід дає змогу перейти від фрагментарного технічного захисту до системного управління безпекою муніципальних IoT-сервісів, раціонально розподіляти ресурси громади та підвищувати стійкість міської цифрової інфраструктури. Подальші дослідження доцільно спрямувати на кількісну модель оцінювання ризиків, автоматизацію інвентаризації активів та інтеграцію цього підходу з платформами моніторингу кібербезпеки [3,4].

1. Alavi A.H. et al. Internet of Things-enabled smart cities: State-of-the-art and future trends. Measurement. 2018. Vol. 129. P. 589-606.
2. Elmaghraby A.S., Losavio M.M. Cyber security challenges in smart cities: Safety, security and privacy. Journal of Advanced Research. 2014. Vol. 5(4). P. 491-497.
3. Tok Y.C., Chattopadhyay S. Identifying threats, cybercrime and digital forensic opportunities in smart city infrastructure via threat modeling. Forensic Science International: Digital Investigation. 2023. Vol. 45. Article 301540.
4. Kumar V. et al. Challenges in the design and implementation of IoT testbeds in smart-cities: A systematic review. arXiv:2302.11009. 2023.

Development of an artificial intelligence-driven managed detection and response framework for proactive enterprise cyber defense

UDK 004.056.53(043.2)

Kyrylo Kurchak

National University of Water and Environmental Engineering,
kurchak_ak21@nuwm.edu.ua

The continuous evolution of enterprise computing paradigms, characterized by the widespread adoption of hybrid cloud architectures, distributed microservices, and extensive remote-work infrastructures, has radically expanded the corporate attack

surface. Modern corporate networks generate an unprecedented volume of multi-layered telemetry streams originating from endpoints, cloud logging facilities, database management systems, and network perimeter devices [1, 5]. Security Operations Centers (SOCs) tasked with monitoring these vast data landscapes are increasingly overwhelmed by the sheer scale, velocity, and complexity of incoming alerts, a phenomenon that introduces critical operational risks [2]. Traditional security monitoring infrastructures rely heavily on static, rule-based signature detection mechanisms that evaluate events in isolation, failing to correlate disparate activities across different environment layers [3]. Consequently, human security analysts are subjected to pervasive alert fatigue, wherein thousands of low-fidelity warnings obscure the subtle, highly distributed indicators of compromise characteristic of advanced persistent threats (APTs) and sophisticated multi-stage cyber campaigns [3, 5].

This operational bottleneck is further exacerbated by the severe global shortage of specialized cybersecurity personnel capable of conducting deep forensic investigations under tight time constraints. When security analysts are forced to manually investigate, parse, and triage millions of continuous data events every day, the time elapsed between an initial perimeter breach and its subsequent identification — commonly referred to as threat dwell time — extends to unacceptable levels [1, 3]. Sophisticated adversarial groups exploit this visibility gap by deploying stealthy, living-off-the-land techniques, utilizing legitimate system administration tools to execute malicious actions that easily evade isolated endpoint detection rules. The fundamental challenge within modern enterprise cyber defense lies in the inability of conventional security frameworks to rapidly distinguish high-fidelity threat indicators from intense background operational noise, thereby preventing proactive containment and leaving corporate digital assets highly vulnerable to devastating operational disruptions, data exfiltration, and ransomware deployment.

The objective of the work is to design and evaluate an artificial intelligence-driven Managed Detection and Response (MDR) framework that automates high-volume telemetry correlation, accelerates threat containment, and optimizes analyst operational workflow. To achieve this objective, the system utilizes continuous behavioral modeling and cross-layer event correlation to minimize threat dwell time. The realization of this goal requires the systematic execution of several interconnected technical tasks, beginning with the establishment of a highly scalable, multi-layered telemetry ingestion architecture capable of normalizing heterogeneous log schemas into a unified data structure. Following this, automated behavioral analysis models must be developed to leverage machine learning algorithms for establishing dynamic operational baselines for all network entities and user accounts. Additionally, the framework incorporates an intelligent cross-environment orchestration layer designed to correlate separate localized anomalies into a single, comprehensive incident timeline. Finally, the operational performance of the proposed architecture must be validated through rigorous experimental testing against real-world attack vectors to measure improvements in critical security metrics, specifically focusing on the reduction of mean time to detect (MTTD) and mean time to respond (MTTR).

The relevance of this study is underscored by the rapidly escalating frequency, speed, and sophistication of automated cyber threats targeting critical corporate and

state infrastructures. In an era where adversarial actors routinely utilize automated exploit kits, artificial intelligence-driven scanning tools, and rapidly mutating malware variants, the traditional model of episodic or business-hours security monitoring is entirely obsolete [2]. Proactive, uninterrupted 24/7 defensive vigilance is a mandatory operational requirement to safeguard corporate data integrity, preserve financial stability, and maintain compliance with increasingly stringent global regulatory mandates [1, 2]. Organizations that fail to maintain continuous visibility over their digital assets face catastrophic financial liabilities, legal repercussions, and permanent reputational damage following a successful breach.

Managed Detection and Response services have emerged as a critical architectural solution for organizations seeking to achieve robust security coverage without incurring the prohibitive capital and operational expenses associated with building and maintaining a sophisticated, internal 24/7 SOC [3, 5]. However, as corporate networks scale exponentially with the integration of IoT devices and multi-cloud environments, human-centric MDR service designs encounter definitive physical limitations [2, 4]. The incorporation of advanced artificial intelligence models into the core of MDR platforms is highly relevant and urgent, as it provides the only viable computational mechanism to scale threat detection capabilities alongside expanding enterprise data volumes, ensuring that security analysts are empowered with pre-filtered, context-rich intelligence rather than being buried under unmanageable alert volumes [4].

The scientific novelty of this research consists in the development of a unified multi-environment telemetry correlation engine based on the CORTAI platform architecture [4]. Unlike conventional security frameworks that analyze endpoint events, cloud access logs, and network netflow data in isolated, parallel silos, the proposed model mathematically projects multi-layer telemetry into a single, cohesive investigation timeline. This approach advances the state of the art by replacing rigid, retrospective correlation rules with dynamic, context-aware risk scoring that continuously adapts to evolving adversarial techniques based entirely on live behavioral deviations.

Furthermore, the scientific innovation lies in the design of a temporal behavioral modeling approach within the CORTAI architecture that explicitly evaluates the structural and temporal links between independent, seemingly low-severity anomalies occurring across disparate enterprise sectors [4]. By mapping these distributed anomalies into a unified threat progression graph, the system is capable of detecting stealthy, slow-rate attack campaigns that purposefully operate below the detection thresholds of isolated security controls. This formulation successfully shifts the defensive paradigm from historical signature-matching to predictive intent modeling, enabling the early identification of multi-stage adversarial pathways before the final payload execution or data exfiltration phase can occur.

The proposed solution implements a comprehensive, multi-layered Managed Detection and Response architecture natively integrated with the CORTAI security intelligence platform [4]. Telemetry collection is executed via widespread deployment of lightweight endpoint detection and response (EDR) agents, cloud-native API logging facilities, and continuous network flow collectors, ensuring complete visibility across the hybrid corporate infrastructure [1, 2, 3]. These distributed data

sources feed raw log entries into localized ingestion nodes, which apply structural normalization and statistical parsing to transform heterogeneous data formats into a standardized, unified schema. As iThe core analytical engine of the CORTAI platform processes the normalized telemetry streams by applying unsupervised machine learning algorithms to establish dynamic, historical baseline behaviors for every user account, host device, and network interface within the organization [4].

Fig.1. Operational flow of the AI-augmented CORTAI platform from raw telemetry ingestion to predictive visibility

As illustrated in Fig. 1, the CORTAI engine performs automated event correlation by programmatically linking distributed anomalies across the network into pre-assembled investigation dossiers [4]. When a series of localized anomalies collectively exceeds the cumulative security threat score threshold, the platform automatically triggers localized containment playbooks designed to neutralize the threat instantly without requiring manual analyst intervention [5]. These playbooks include executing automated actions such as isolating compromised host devices from the network, revoking compromised user credentials, and modifying perimeter firewall rules [3, 5]. Concurrently, the system delivers a comprehensive, pre-filtered, and highly context-rich evidence dossier directly to senior security analysts for formal validation, preserving valuable human cognitive resources and maintaining an efficient human-in-the-loop oversight model for high-impact mitigation decisions [2].

The design and implementation of the artificial intelligence-driven Managed Detection and Response framework significantly enhances enterprise cybersecurity posture by systematically optimizing operational defensive metrics. By deploying the CORTAI platform as the central analytical engine, the framework achieves a drastic reduction in both the mean time to detect (MTTD) and the mean time to respond (MTTR) when faced with sophisticated, multi-vector adversarial campaigns [4]. The automation of low-level alert triage and context enrichment successfully mitigates the pervasive issue of analyst alert fatigue, allowing human security experts to allocate their specialized cognitive resources exclusively toward proactive threat hunting, deep forensic investigations, and strategic incident response planning [2, 4]. Experimental validation confirms that the proposed multi-layered correlation architecture effectively minimizes false-positive rates, maximizes threat detection coverage across highly complex hybrid infrastructures, and maintains continuous, robust 24/7 enterprise defense capabilities.

1. Hautala T. New Managed Detection and Response System to Lower the Likelihood of Malicious Activities in IP Network. Master's Thesis, Metropolia University of Applied Sciences, Helsinki, 2022. 46 p.
2. Rajgopal P.R., Karanam L. MDR Service Design: Building Profitable 24/7 Threat Coverage for SMBs. International Journal of Applied Mathematics. – 2025. – V. 38, № 2s.
3. Fortinet. What Is Managed Detection and Response (MDR)? URL: https://www.fortinet.com/uk/resources/cyberglossary/managed-detection-and-response (application date: 30.04.2026).
4. MCK. The Role of AI in Managed Detection: How Artificial Intelligence Strengthens MDR Security. URL: https://www.mck.com/blog/role-of-ai-in-managed-detection (application date: 30.04.2026).
5. Rapid7. What Is MDR? Managed Detection and Response Security. URL: https://www.rapid7.com/fundamentals/what-is-managed-detection-and-response-mdr/ (application date: 30.04.2026).

Модель оцінювання кіберризиків IoT-інфраструктури розумного міста

УДК 004.056:004.738.5:004.77

Віталій Левицький[1], Олег Тотоськo[2]
Олександр Добруцький[3]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]levytskyylv@gmail.com, [2]totosko@gmail.com, [3]oleksandrdobruckij@gmail.com

Цифровізація міської інфраструктури передбачає впровадження IoT-пристроїв, сенсорних мереж, edge-вузлів, хмарних платформ і диспетчерських систем для моніторингу довкілля, транспорту, енергоресурсів, відеоспостереження та аварійного реагування. Зростання кількості підключених компонентів формує кіберфізичне середовище, у якому порушення безпеки даних має інформаційні, організаційні й технологічні наслідки. Традиційні підходи до управління ризиками орієнтовані переважно на корпоративні мережі й серверну інфраструктуру [1], тоді як міські IoT-системи мають малопотужні пристрої, неоднорідні протоколи, бездротові канали та залежність від хмарних сервісів. За NIST CSF 2.0 управління кіберризиками має охоплювати виявлення активів, захист, моніторинг, реагування й відновлення [2]. Метою роботи є розроблення моделі оцінювання кіберризиків IoT-інфраструктури розумного міста для пріоритезації захисту компонентів з урахуванням критичності, експозиції, вразливостей і зрілості захисних заходів. Новизна полягає в поєднанні логіки «загроза — вразливість — наслідок» із параметрами IoT-середовища: відкритістю, залежностями, сегментацією та рівнем захисту. Модель передбачає інвентаризацію активів, профіль загроз, оцінювання вразливостей, визначення впливу на сервіси, розрахунок інтегрального ризику та пріоритезацію заходів. До активів належать сенсори, виконавчі пристрої, шлюзи, edge-вузли, мережеве обладнання, API, бази даних, хмарні сервіси та операторські панелі.

Інтегральний показник ризику і-го компонента подається як $R_i = P_i \times I_i \times E_i \times D_i \times (1 - C_i)$, де P_i — імовірність загрози, I_i — вплив, E_i — експозиція, D_i — залежність від інших сервісів, C_i — зрілість захисту. P_i, I_i, E_i та D_i оцінюються за шкалою 1–5, C_i — у межах 0–1.

Таблиця 1

Складові оцінювання кіберризику IoT-компонента

Показник	Зміст оцінювання	Шкала
P_i	імовірність реалізації загрози з урахуванням вразливості та доступності атаки	1–5
I_i	вплив на конфіденційність, цілісність, доступність і безперервність сервісу	1–5
E_i	рівень відкритості компонента до мережевої або фізичної взаємодії	1–5
D_i	критичність залежностей від інших пристроїв, платформ і сервісів	1–5
C_i	зрілість захисних заходів: автентифікація, оновлення, сегментація, журналювання	0–1

Реалізація можлива як матриця ризиків або модуль муніципальної системи моніторингу кібербезпеки. Реєстр активів фіксує тип пристрою, місце розгортання, мережевий сегмент, відповідальний підрозділ і критичність сервісу. Профіль загроз охоплює несанкціонований доступ, компрометацію облікових даних, атаки на канал зв'язку, підміну телеметрії, відмову в обслуговуванні й незакриті вразливості, а також базові вимоги IoT-безпеки [3].

Особливістю моделі є врахування каскадних інцидентів: компрометація шлюзу може спричинити втрату телеметрії, помилкові управлінські рішення або недоступність суміжних сервісів. Тому D_i відображає роль компонента в цифровій екосистемі. Це важливо з огляду на атаки на доступність, програми-вимагачі, загрози даним та експлуатацію слабких місць цифрових сервісів [4].

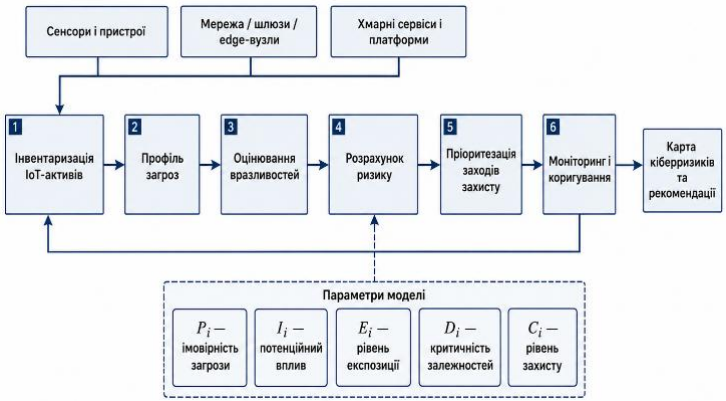


Рис.1. Модель оцінювання кіберризиків IoT-інфраструктури розумного міста

Результати можуть використовуватися для карти кіберризиків, сегментації мережі, посилення автентифікації, централізованого оновлення, резервування критичних сервісів і журналювання подій. Модель не замінює аудит, але є інструментом попереднього оцінювання та регулярного моніторингу. Подальші дослідження доцільно спрямувати на калібрування вагових коефіцієнтів, інтеграцію з SIEM/SOC-рішеннями та адаптацію до окремих міських сервісів..

1. Nurse J.R.C., Creese S., De Roure D. Security risk assessment in Internet of Things systems. IT Professional. 2017. Vol. 19(5). P. 20-26.
2. Kandasamy K. et al. IoT cyber risk: a holistic analysis of cyber risk assessment frameworks, risk vectors, and risk ranking process. EURASIP Journal on Information Security. 2020. Vol. 2020. Article 8.
3. Parsons E.K. et al. A survey on cyber risk management for the Internet of Things. Applied Sciences. 2023. Vol. 13(15). Article 9032.
4. Gharaibeh A. et al. Smart cities: A survey on data management, security, and enabling technologies. IEEE Communications Surveys & Tutorials. 2017. Vol. 19(4). P. 2456-2501.

Еволюція методів виявлення шкідливих URL: від евристик до трансформерних архітектур

УДК 519.6

Петро Венгерський[1], Володимир Лесик[2]

Львівський національний університет імені Івана Франка,
[1]petro.venherskyy@lnu.edu.ua, [2]volodymyr.lesyk@lnu.edu.ua

Стрімке зростання кількості вебресурсів та кіберзагроз зумовлює актуальність задачі виявлення шкідливих URL-адрес, що використовуються для фішингу, поширення шкідливого програмного забезпечення та координації бот-мереж. Традиційні підходи, засновані на чорних списках і сигнатурному аналізі, мають обмежену ефективність через нездатність виявляти нові або модифіковані атаки [1-3].

Метою роботи є систематизація еволюції методів виявлення шкідливих URL-адрес - від евристичних підходів до сучасних моделей машинного та глибокого навчання, а також визначення перспектив подальшого розвитку інтелектуальних систем захисту.

Еволюцію технологій детекції можна розділити на кілька послідовних етапів. Технологічний прогрес у цій галузі характеризується переходом від жорстких алгоритмів до гнучких нейромережевих архітектур [3] (рис. 1).

Початковий етап базувався на використанні ручних правил і статичних баз даних, що забезпечували швидку, але обмежену за узагальненням детекцію відомих аномалій [3].

Наступним кроком стало впровадження класичного машинного навчання (ML). Алгоритми логістичної регресії, SVM та ансамблеві методи (зокрема Random Forest) автоматизували класифікацію на основі видобутих ознак URL-адрес: лексичних, хостових, контентних та зовнішніх [1,3,4]. Лексичні ознаки дозволяють здійснювати швидку класифікацію без доступу до мережевих

ресурсів [4,5], тоді як хостові та контентні характеристики забезпечують глибокий аналіз веб інфраструктури [6]. Головною перевагою цих моделей стала висока інтерпретованість та обчислювальна ефективність.

Рис.1. Узагальнена схема еволюції технологій детекції URL

Зростання обсягів даних зумовило перехід до методів глибокого навчання (DL). Архітектури згорткових (CNN) та рекурентних (LSTM) нейронних мереж дозволяють автоматично витягувати складні нелінійні залежності з URL-рядків і контенту вебсторінок [7-9]. Це зменшує залежність від ручної інженерії ознак і підвищує здатність моделей до узагальнення.

Сучасний етап розвитку представлений трансформерними архітектурами (моделі типу BERT), які забезпечують аналіз контексту та семантичних залежностей у URL-адресах [8,10]. Такі підходи демонструють найвищі показники точності (до 99,97%) [3], однак їх практичне впровадження ускладнюється жорсткими вимогами до обчислювальних ресурсів.

Паралельно також розвиваються альтернативні підходи, зокрема асоціативна класифікація та методи видобування правил, що дозволяють формувати інтерпретовані залежності між ознаками [11, 12]. Їх інтеграція з моделями глибокого навчання відкриває можливості створення унікальних гібридних систем.

Основними викликами сучасних систем є залежність від якості даних, обчислювальна складність моделей та швидка еволюція кіберзагроз [3, 13]. У зв'язку з цим перспективним напрямом є створення адаптивних систем, здатних динамічно змінювати ваги ознак і поєднувати різні підходи до аналізу [6, 14].

Еволюція методів виявлення шкідливих URL-адрес демонструє поступовий перехід від простих евристик до складних інтелектуальних моделей. Найбільш перспективним напрямом є розробка адаптивних гібридних систем, що поєднують інтерпретованість класичних методів, здатність до узагальнення моделей глибокого навчання та контекстний аналіз трансформерів, з можливістю динамічного коригування ваг ознак у реальному часі.

1. Sahoo D., Liu C., Hoi S.C.H. Malicious URL Detection using Machine Learning: A Survey. arXiv preprint. – 2017. – arXiv:1701.07179. – doi:10.48550/arXiv.1701.07179.
2. Aljabri M., Altamimi H., Albelali S., Al-Harbi M., Alhuraib H., Alotaibi N., Alahmadi A., Alhaidari F., Mohammad R., Salah K. Detecting

Malicious URLs Using Machine Learning Techniques: Review and Research Directions. IEEE Access. – 2022. – doi:10.1109/ACCESS.2022.3222307.
3. Tian Y., Yu Y., Sun J., Wang Y. From Past to Present: A Survey of Malicious URL Detection Techniques, Datasets and Code Repositories. arXiv preprint. – 2025. – arXiv:2504.16449. – doi:10.48550/arXiv.2504.16449.
4. Dhotre A. Malicious URLs Detection using Lexical Features based on Machine Learning. IJSRD. – 2023. – Vol. 11, Issue 8.
5. Joshi A., Lloyd L., Westin P., Seethapathy S. Using Lexical Features for Malicious URL Detection – A Machine Learning Approach. arXiv preprint. – 2019. – arXiv:1910.06277. – doi:10.48550/arXiv.1910.06277.
6. Hamadouche S., Boudraa O., Gasmi M. Combining Lexical, Host, and Content-based Features for Phishing Websites Detection using Machine Learning Models. EAI Endorsed Transactions on Scalable Information Systems. – 2024. – Vol. 11, no. 6. – doi:10.4108/eetsis.4421.
7. Do N., Selamat A., Krejcar O., Herrera-Viedma E., Fujita H. Deep Learning for Phishing Detection: Taxonomy, Current Challenges and Future Directions. IEEE Access. – 2022. – Vol. 10. – P. 36543–36564. – doi:10.1109/ACCESS.2022.3151903.
8. Turk F., Kilicaslan M. Malicious URL Detection with Advanced Machine Learning and Optimization-Supported Deep Learning Models. Appl. Sci. – 2025. – Vol. 15, no. 18. – Art. 10090. – doi:10.3390/app151810090.
9. Kibriya H., Amin R., Alshamrani S.S. et al. Lightweight Malicious URL Detection using Deep Learning and Large Language Models. Sci. Rep. – 2025. – Vol. 15. – Art. 43044. – doi:10.1038/s41598-025-26653-2.
10. Elsadig M., Ibrahim A.O., Basheer S., Alohali M.A., Alshunaifi S., Alqahtani H., Alharbi N., Nagmeldin W. Intelligent Deep Machine Learning Cyber Phishing URL Detection Based on BERT Features Extraction. Electronics. – 2022. – Vol. 11, no. 22. – Art. 3647. – doi:10.3390/electronics11223647.
11. Kumi S., Lim C., Lee S.-G. Malicious URL Detection Based on Associative Classification. Entropy. – 2021. – Vol. 23, no. 2. – Art. 182. – doi:10.3390/e23020182.
12. Jeeva S.C., Rajsingh E.B. Intelligent Phishing URL Detection using Association Rule Mining. Hum. Cent. Comput. Inf. Sci. – 2016. – Vol. 6. – Art. 10. – doi:10.1186/s13673-016-0064-3.
13. Alsaedi M., Ghaleb F.A., Saeed F., Ahmad J., Alasli M. Cyber Threat Intelligence-Based Malicious URL Detection Model Using Ensemble Learning. Sensors. – 2022. – Vol. 22, no. 9. – Art. 3373. – doi:10.3390/s22093373.
14. Rafsanjani A.S., Kamaruddin N.B., Behjati M., Aslam S., Sarfaraz A., Amphawan A. Enhancing Malicious URL Detection: A Novel Framework Leveraging Priority Coefficient and Feature Evaluation. IEEE Access. – 2024. – Vol. 12. – P. 85001–85026. – doi:10.1109/ACCESS.2024.3412331.

Способи витоку персональних даних, методи обробки та захист від витоків

УДК 004.056

Ірина Волобуєва[1], Лідія Тимошенко[2]

Національний університет «Одеська політехніка»,
[1]10252809@stud.op.edu.ua, [2]l.m.timoshenko@op.edu.ua

Витік персональних даних є однією з актуальних проблем сучасної інформаційної безпеки. Він передбачає несанкціоноване поширення, копіювання або викрадення інформації, за допомогою якої можна ідентифікувати особу. До таких відомостей належать імена, адреси, електронні адреси, номери телефонів, паролі, банківські реквізити та інші дані. Особливо небезпечними є витоки облікових даних, оскільки паролі залишаються одним із найпоширеніших засобів автентифікації, а їх повторне використання в різних сервісах підвищує ризик несанкціонованого доступу [1].

Актуальність дослідження зумовлена тим, що після витоку персональні дані часто потрапляють до публічних або напівпублічних масивів, зокрема у форматі email:password. Такі набори можуть використовуватися для підбору паролів, компрометації облікових записів, шахрайства та подальших кібератак. Тому важливим є не лише захист персональних даних, а й аналіз уже наявних масивів витоку для виявлення типових слабких місць у поведінці користувачів і політиках безпеки.

Метою роботи є аналіз способів витоку персональних даних, наборів даних формату email:password, виявлення слабких і повторюваних паролів шляхом розроблення програмного застосунку, визначення доменної концентрації записів та оцінювання рівня ризику витоку для подальшого формування рекомендацій з покращення інформаційної безпеки.

До основних способів витоку персональних даних належать технічні, організаційні та пов’язані з діями користувачів. Технічні способи витоку можуть бути пов’язані з вразливостями серверів, шкідливим програмним забезпеченням, фішинговими атаками, SQL-ін’єкціями та іншими засобами отримання несанкціонованого доступу. Організаційні способи витоку виникають у разі недостатнього контролю доступу, відсутності належних політик безпеки, слабкого журналювання дій користувачів або неналежного зберігання інформації. Також поширеним способом витоку залишається людський фактор: використання слабких паролів, повторення однакових комбінацій у різних сервісах, довіра до фішингових повідомлень і нехтування базовими правилами інформаційної безпеки [2].

Окрему увагу слід приділити шаблонності поведінки користувачів. Люди часто створюють паролі за передбачуваними схемами: використовують дати народження, прості цифрові послідовності, імена, короткі слова або однакові комбінації для різних акаунтів. Така поведінка спрощує підбір паролів і підвищує ймовірність компрометації облікових записів після витоку даних [3].

Наукова новизна роботи полягає в поєднанні аналізу способів витоку персональних даних із практичним оцінюванням ризику на основі характеристик набору email:password. На відміну від робіт, у яких переважно

розглядаються окремі аспекти захисту даних або поведінки користувачів, у цій роботі пропонується підхід, що враховує слабкість паролів, повторне використання парольних комбінацій, доменну концентрацію записів та актуальність витоку.

Для розв'язання поставленої задачі запропоновано створення програмного застосунку для аналізу публічних наборів даних формату email:password. Застосунок доцільно реалізувати мовою Python із використанням бібліотек pandas, regex, matplotlib та бази даних SQLite. Основні етапи його роботи передбачають імпорт текстового файлу, розбір рядків на електронну адресу, домен і пароль, перевірку коректності формату, підрахунок унікальних паролів і доменів, визначення поширених слабких паролів, обчислення повторного використання паролів, групування записів за доменами та формування підсумкових звітів.

Для наочного подання логіки роботи застосунку використано блок-схему. Загальна структура роботи програмного застосунку наведена на рис. 1.

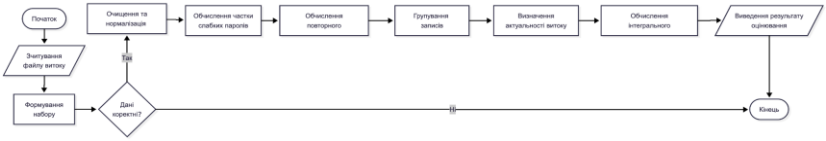


Рис.1. Узагальнена блок-схема застосунку

Вона відображає послідовність основних етапів обробки даних: зчитування файлу витоку, формування набору записів, перевірка коректності даних, очищення та нормалізація записів, обчислення показників ризику, групування за доменами та виведення результату оцінювання.

У результаті роботи обґрунтовано актуальність аналізу публічних наборів персональних даних формату email:password та запропоновано структуру програмного застосунку для їх опрацювання. Очікуваними результатами є визначення поширеності слабких паролів, рівня повторюваності парольних комбінацій, доменної концентрації записів та інтегрального ризику витоку. Отримані результати можуть бути використані для формування рекомендацій щодо посилення політик інформаційної безпеки, підвищення рівня захисту облікових даних і зменшення ризику несанкціонованого доступу.

1. Тимошенко Л., Вакуліна С., Волобуєва І. Локальний менеджер паролів із генеруванням для мобільних пристроїв, Кібербезпека в сучасному світі: актуальні виклики : матеріали VI міжнар. наук.-практ. конф., м. Одеса, 28 листопада 2025 р. Одеса, 2025. С. 40.
2. Корнєва В.Д. Способи захисту ІТ-індустрії від витоку інформації. Інформаційна безпека та комп'ютерні технології : матеріали VII міжнар. наук.-практ. конф., м. Київ: Державний торговельно-економічний університет, 2023 р. Київ, 2023. С. 31–33.
3. Заболоцький П.В. Шаблонність людського мислення та комп'ютерна безпека. Міжнародний науковий журнал «Грааль науки», №24, лютий 2023. С. 345–347.

Використання штучного інтелекту для автоматизації виявлення та пріоритезації інцидентів інформаційної безпеки

УДК 004.056:004.8

Ірина Лозова[1], Михайло Різак[2], Євгеній Педченко[3]

Державний університет інформаційно-комунікаційних технологій,
[1]illozovaya@gmail.com, [2]advokat.rizak@gmail.com, [3]ympedchenko@gmail.com

Сучасний розвиток інформаційних технологій супроводжується постійним зростанням кількості кіберінцидентів та ускладненням механізмів атак. Центри моніторингу безпеки (SOC) змушені обробляти значні обсяги подій безпеки, що призводить до перевантаження аналітиків та збільшення часу реагування на інциденти. Традиційні підходи, засновані на статичних правилах кореляції, не забезпечують достатньої швидкості та точності аналізу подій. У зв'язку з цим актуальним є використання технологій штучного інтелекту (ШІ) та платформ класу SOAR (Security Orchestration, Automation and Response), які дозволяють автоматизувати процеси виявлення, аналізу та пріоритезації інцидентів інформаційної безпеки [1].

Метою роботи є дослідження можливостей використання штучного інтелекту для автоматизації виявлення та пріоритезації інцидентів інформаційної безпеки, а також практична реалізація SOAR-сценарію у середовищі Make.com із використанням зовнішніх Threat Intelligence-сервісів та AI-модуля.

Питання автоматизації процесів реагування на інциденти активно досліджуються у сучасній науковій літературі. У роботі [4] розглядаються принципи побудови SOAR-рішень для автоматизації поведінкових honeypot-систем та механізмів реагування на загрози. В роботі [2] описано можливості сервісу VirusTotal щодо отримання репутаційних характеристик IP-адрес, доменів та файлів через API. Документація [3] містить опис можливостей великих мовних моделей Anthropic Claude для автоматизації аналізу кіберзагроз і формування рекомендацій щодо реагування. У роботі [5] запропоновано багаторівневу агентну AI-модель для автоматизації процесів SOC. Аналіз публікацій показує, що поєднання SOAR-платформ, сервісів Threat Intelligence та AI-модулів дозволяє значно підвищити ефективність роботи SOC та скоротити час первинного аналізу інцидентів.

У межах роботи було реалізовано автоматизований сценарій реагування на інциденти у середовищі Make.com. Сценарій включав декілька взаємопов'язаних етапів: приймання подій через Webhook, збагачення даних за допомогою сервісу VirusTotal, аналіз інциденту за допомогою AI-модуля Anthropic Claude, маршрутизацію результатів та автоматичне формування повідомлень (рис.1).

На першому етапі модуль Webhook приймав вхідний JSON-об'єкт, який містив IP-адресу, опис події та ідентифікатор системи. Далі за допомогою HTTP-запиту до API сервісу VirusTotal виконувалось отримання репутаційних характеристик IP-адреси, зокрема кількості malicious, suspicious, harmless та undetected-спрацювань [2].

Після цього дані передавались до AI-модуля Anthropic Claude Simple Text Prompt, який виконував аналіз події та формував структуровану JSON-відповідь із полями risk_level, summary та action. Модель здійснювала автоматичну оцінку рівня ризику інциденту та визначала рекомендовану дію: alert або log [3]. Для подальшої обробки результатів використовувався модуль Parse JSON, який перетворював відповідь моделі у структуровані поля.

Рис.1. Загальна схема сценарію в Make.com

На етапі маршрутизації застосовувався Router, який залежно від значення action виконував різні сценарії реагування. Якщо подія визнавалась критичною, система автоматично надсилала повідомлення до Telegram. Для менш критичних подій інформація записувалась до Google Sheets з метою ведення журналу аудиту.

У процесі тестування використовувалися як інциденти, пов'язані з потенційними brute-force атаками, так і приклади легітимної активності користувачів. Для тестового інциденту модель Claude сформувала значення risk_level = 62 та action = alert, що дозволило автоматично маршрутизувати подію до Telegram-каналу аналітика без додаткової ручної перевірки (рис. 2).

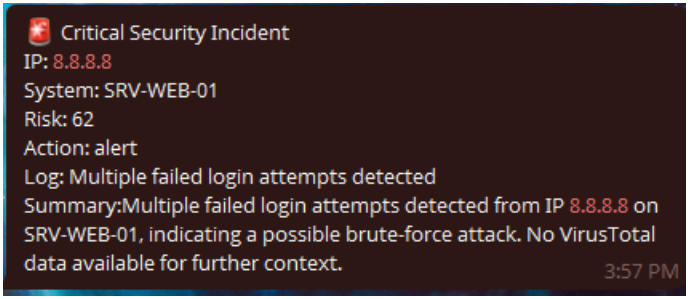


Рис. 2. Приклад фінального сповіщення в Telegram

Отримані результати показали, що запропонований підхід дозволяє автоматизувати значну частину первинного аналізу інцидентів інформаційної безпеки. Інтеграція Threat Intelligence та AI-моделі забезпечує швидке оцінювання ризику та скорочує час реагування на інциденти. Крім того, використання SOAR-підходу дозволяє знизити навантаження на SOC-аналітиків та підвищити ефективність процесів моніторингу безпеки.

Таким чином, результати дослідження підтверджують доцільність використання штучного інтелекту та SOAR-платформ для автоматизації виявлення та пріоритезації інцидентів інформаційної безпеки. Подальші дослідження можуть бути спрямовані на інтеграцію додаткових джерел Threat Intelligence, удосконалення моделей оцінювання ризику та розширення автоматизованих механізмів реагування.

1. SANS Institute. AI-Driven SecOps: Unifying Controls, Automating Response, and Advancing the Modern SOC Using Cortex XSIAM. 2025. URL: https://www.sans.org/white-papers/ai-driven-secops-unifying-controls-automating-response-advancing-modern-soc-using-cortex-xsiam.
2. VirusTotal. API Overview. URL: https://docs.virustotal.com/reference/overview.
3. Anthropic. Anthropic Claude Documentation. URL: https://platform.claude.com/docs/en/home.
4. Umesh Bartwal, Subhasis Mukhopadhyay, Rohit Negi, Shashank Shukla. Security Orchestration, Automation, and Response Engine for Deployment of Behavioural Honeypots. 2022. URL: https://doi.org/10.1109/DSC54232.2022.9888808.
5. Jay Roy, Sandeep Singh. AgentSOC: A Multi-Layer Agentic AI Framework for Security Operations Automation. 2026. URL: https://doi.org/10.1109/ICAIC67076.2026.11395783.

Методологія забезпечення мережевої ізоляції та динамічного масштабування ресурсів у середовищі змагального кіберполігону

УДК 004.056:378.147

Богдан Маліцький[1], Михайло Євдокімов[2], Данило Куташ[3], Василь Різак[4]

Ужгородський національний університет,
[1]bohdan.malitskyi@uzhnu.edu.ua, [2]mykhailo.yevdokimov@uzhnu.edu.ua,
[3]danylo.kutash@student.uzhnu.edu.ua, [4]vrizak@uzhnu.edu.ua

Проведення CTF-чемпіонату регіонального масштабу ставить низку технічних вимог до інфраструктури: ізоляція вразливих сервісів від основної мережі закладу освіти, контрольований віддалений доступ для учасників з різних міст, можливість швидко повертати середовище у вихідний стан між раундами та одночасна робота з десятками учасників. Виконати ці вимоги на базі діючого навчального кіберполігону можливо лише за умови комплексної адаптації його архітектури – на рівні платформи віртуалізації, мережі, доступу та змагальної системи. Метою роботи є розробка та впровадження методології,

що забезпечує мережеву ізоляцію та динамічне масштабування ресурсів кіберполігону, на базі кафедри твердотільної електроніки та інформаційної безпеки Ужгородського національного університету для проведення першого чемпіонату Закарпатської області з кібербезпеки (CTF-змагань) ТЕІВ-2026.

Архітектуру кіберполігону адаптовано за чотирма напрямками. На рівні віртуалізації виконано перехід на серверний гіпервізор Proxmox VE, який підтримує одночасну роботу з віртуальними машинами і контейнерами та керується з єдиного інтерфейсу; цей перехід продиктований необхідністю обслуговувати багатьох учасників одночасно та швидко повертати середовища у вихідний стан через знімки. Базова апаратна конфігурація серверного вузла включає 12 ядер CPU, 64 ГБ RAM та 1 ТБ сховища; під час змагань на ньому одночасно функціонували 14 віртуальних машин. На рівні платформи розгорнуто змагальну систему CTFd, як окрему віртуальну машину з декількома контейнеризованими сервісами, відокремленими від основного трафіку через reverse proxy. На рівні мережі до існуючих навчальних сегментів додано окремі домени для роботи учасників і для розміщення вразливих сервісів, ізольовані firewall-правилами. На рівні доступу впроваджено VPN-підключення з автентифікацією за сертифікатами (OpenVPN), що дозволяє відключати конкретного учасника без впливу на інших.

В межах адаптованої інфраструктури підготовлено 69 практичних завдань у семи категоріях; статичні артефакти (forensic-образи, мережевий трафік, криптографічні артефакти) розміщено централізовано на змагальній платформі, а інтерактивні завдання – у персональних ізольованих середовищах із поверненням до вихідного стану між раундами. Експлуатація під час змагання: 24 години безперервної роботи, до 58 одночасних VPN-підключень, 7 811 спроб розв'язання завдань – без зафіксованих збоїв. Спостереження за телеметрією під час пікової фази зафіксували завантаження процесора до 50 % та використання оперативної пам'яті близько 90 % (рис. 1). Передзмагальне навантажувальне тестування (Load, Stress та Endurance Testing загальною тривалістю понад 78 годин) із застосуванням Apache Bench, iperf3, stress-ng, hping3, siege та Nmap визначило верхню межу стабільної роботи стенду: 850 одночасних HTTP-з'єднань, до 1,8 Гбіт/с внутрішнього трафіку, 14 000 подій журналювання за хвилину при деградації продуктивності в межах 4–6 % після стрес-фази.

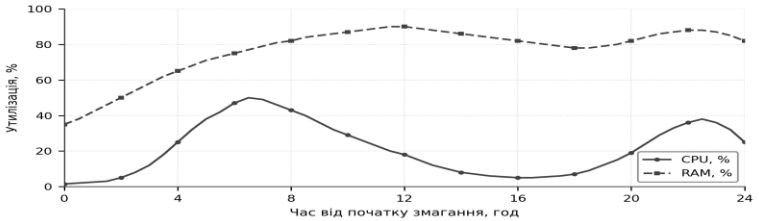


Рис.1. Динаміка утилізації CPU та RAM серверного вузла протягом 24 годин змагання

Запропонована методологія належить до класу локальних навчально-змагальних кіберполігонів і концептуально близька до інших академічних

рішень: автономної контейнерної платформи UNIWA на основі OpenStack та Docker [1] та платформи EMP для bootcamp-формату на 6–11 віртуальних машинах [2]. Відмінність полягає у комбінованому використанні віртуальних машин і контейнерів на базі гіпервізора рівня закладу освіти (Proxmox VE) замість хмарних платформ, що знижує бар'єри впровадження для регіональних університетів [3].

Результати апробації підтверджують працездатність запропонованої методології. Контейнеризація CTFd-сервісів і сегментація мережі firewall-правилами забезпечили повну ізоляцію вразливих сервісів від штатних навчальних середовищ протягом 24 годин роботи. Механізм знімків Proxmox VE дозволив швидко повертати середовища у вихідний стан між раундами без впливу на сусідні команди. Фактичне навантаження під час змагання (CPU до 50 %, RAM близько 90 %) залишилося нижчим за граничні показники стенду (850 одночасних HTTP-з'єднань, 1,8 Гбіт/с трафіку), що свідчить про наявний запас потужності. Розроблена методологія придатна для проведення CTF-змагань регіонального масштабу на базі кіберполігонів закладів освіти з open-source стеком віртуалізації.

1. Chouliaras N., Kantzavelou I., Maglaras L., Pantziou G., Ferrag M. A. A novel autonomous container-based platform for cybersecurity training and research. PeerJ Computer Science. 2023. Vol. 9. Article e1574. DOI: 10.7717/peerj-cs.1574.
2. Arnold D., Ford J., Saniie J. Architecture of an Efficient Environment Management Platform for Experiential Cybersecurity Education. Information. 2025. Vol. 16, No. 7. Article 604. DOI: 10.3390/info16070604.
3. Schafeitel-Tähtinen T., Lazarov W. Teaching and Learning Cybersecurity Using Capture the Flag: Effectiveness Comparison Between University Students in Finland and Czechia. Computer Applications in Engineering Education. 2025. Vol. 33, No. 5. Article e70082. DOI: 10.1002/cae.70082.

Privacy and information security in social media

UDK 004.056.5:004.738.5

Andrii Manko[1], Zhanna Babiak[2]

Ternopil Ivan Puluj National Technical University,
[1]andreymanko4@gmail.com, [2]b.janna73@gmail.com

Social media have become the primary channel for interpersonal and mass communication: according to DataReportal estimates, in 2025 they were actively used by more than 5.2 billion people — about 64 % of the world population. Modern platforms accumulate unprecedented volumes of personal data: contacts, geolocations, biometric templates (face and voice recognition), interaction history, behavioural patterns and psychometric profiles. The combination of these data with artificial intelligence tools, in particular AI-based open-data aggregation and generative models (deepfake), creates qualitatively new risks to the privacy and information security of users, organisations and the state as a whole [1, 2].

The objective of the work is to systematise current threats to privacy and information security in social media and to substantiate a complex of technical and organisational measures that mitigate these risks both at the level of the platform provider and at the level of the end user.

The relevance is driven by the rapid growth in the number and scale of incidents. According to ENISA Threat Landscape 2024, social media remain among the top three vectors for phishing and account takeover; large-scale leaks such as the 2021 Facebook incident (533 million records) and recurring compromises of the open APIs of X (Twitter), Telegram and LinkedIn demonstrate the systemic nature of the problem [3]. An additional factor in the Ukrainian context is the hybrid information operations under the conditions of the war: social media are actively used for coordinated disinformation, adversarial OSINT, profiling of military personnel from open posts and targeted phishing against the civilian population [2, 3].

The scientific novelty consists in systematising threats with regard to modern capabilities of AI-based open-data aggregation and generative synthetic content models, and in formulating a two-level «platform — user» protection model that combines technical means, regulatory requirements and behavioural practices of users [4].

As a result of the study, five main classes of threats have been identified. 1) Personal data leaks caused by attacks on platforms, open APIs and third-party applications (SQL injections, authentication vulnerabilities, cloud-storage misconfigurations). 2) Social engineering: phishing through direct messages and links, spoofing of trusted contacts, account takeover via password guessing and SIM-swap attacks. 3) Profiling and de-anonymisation through AI aggregation of open sources (OSINT) — combining data fragments from several platforms makes it possible to re-identify a person even in cases of formal anonymisation [2]. 4) Manipulative content: video and voice deepfakes, disinformation narratives, coordinated inauthentic behaviour (CIB) of botnets. 5) Disclosure of metadata and geolocation through EXIF data of photos, check-ins, messenger activity and background telemetry collection by client applications.

To mitigate the listed threats, a two-sided complex of measures is proposed. On the platform side: implementation of the «privacy by design» principle, minimisation of collected data, end-to-end message encryption (E2EE), restriction of OAuth token scopes, anomaly detection based on machine-learning models (graph neural networks, ensembles of outlier-detection algorithms), automated labelling of synthetic content with digital watermarks (in particular C2PA manifests), regular independent security audits and bug-bounty programmes. On the user side: multi-factor authentication with FIDO2 hardware keys, regular audit of third-party application permissions, restriction of public profile telemetry (private account, disabled geotags), use of different e-mail addresses for different platforms, and a critical attitude towards unverified content, especially audio and video recordings concerning sensitive topics.

At the organisational level, enterprises and public institutions should implement acceptable-use policies for social media, regular cyber-hygiene training, DLP solutions and formalised DPIA procedures for any integrations with external platforms and APIs, in line with GDPR and Law No. 2297-VI.

Of particular interest are emerging privacy-enhancing technologies (PETs) that can be deployed at the platform level: differential privacy for analytical queries, federated learning that enables training of ML models without centralising user data, homomorphic encryption for confidential computation, and zero-knowledge proofs for age or identity verification without disclosing the underlying attributes. Pilot deployments of these mechanisms by major operators (Apple, Google, Meta) demonstrate the feasibility of combining service functionality with strong privacy guarantees, although wider adoption still requires industry-level standardisation, economic incentives and clear regulatory expectations.

Conclusions. Security in social media is a complex socio-technical problem at the intersection of technical, legal and behavioural aspects, which cannot be solved by one party alone. The combination of platform-side protection mechanisms, requirements of GDPR and Law No. 2297-VI, organisational policies and conscious user behaviour substantially reduces the risks of leaks and manipulation. Further research should focus on ML models for detecting coordinated inauthentic behaviour and on adapting regulatory mechanisms to the challenges of generative AI.

1. Farooq A., Salminen J., Martin J. D., Aldous K., Jung S.-G., Jansen B. J. Exploring Social Media Privacy Concerns: A Comprehensive Survey Study Across 16 Middle Eastern and North African Countries. IEEE Access. 2024. Vol. 12. P. 147087–147105.
2. Hlavatska A., Anhelska O., Opirskyi I. Research of OSINT technology as a new threat of person deanonymisation in cyberspace. Cybersecurity: Education, Science, Technique. 2024. Vol. 1, Iss. 25. P. 19–50. (in Ukrainian)
3. ENISA Threat Landscape 2024. European Union Agency for Cybersecurity. URL: https://www.enisa.europa.eu/topics/cyber-threats/threats-and-trends (application date: 07.05.2026).
4. Law of Ukraine «On the Protection of Personal Data» of 01.06.2010 No. 2297-VI. URL: https://zakon.rada.gov.ua/laws/show/2297-17 (application date: 07.05.2026).

Застосування блокового шифру «Кипарис» для шифрування приватних даних у блокчейн-транзакціях

УДК 004.056.55

Марія Родінко

Харківський національний університет імені В. Н. Каразіна, mariia.rodinko@karazin.ua

Публічні блокчейни за своєю концепцією є прозорими – всі транзакції доступні для перегляду, що створює проблему конфіденційності для застосувань, що оперують чутливими даними. Виникає потреба в ефективному симетричному шифруванні, яке забезпечить приватність без втрати продуктивності мережі. Метою дослідження є розробка методу інтеграції перспективного постквантового блокового шифру «Кипарис» у блокчейн-

інфраструктуру для забезпечення конфіденційності даних у транзакціях із збереженням верифікованості та цілісності.

Блоковий шифр «Кипарис» [1] було розроблено з урахуванням вимог до постквантових примітивів. Перевагами алгоритму є висока швидкодія, стійкість до диференціального криптоаналізу, постквантова стійкість [2].

Запропонована схема криптографічного захисту базується на поєднанні симетричного шифрування, постквантових примітивів та децентралізованих технологій для забезпечення цілісності даних.

Нижче описані ключові етапи циклу гібридного шифрування.

1) *Генерація ключа*. Створення випадкового сесійного ключа (256/512 біт залежно від необхідного рівня стійкості) із використанням криптографічно стійкого генератора псевдовипадкових чисел.
2) *Шифрування даних алгоритмом «Кипарис»*. Використання ARX-архітектури алгоритму забезпечує високу швидкість обробки на пристроях з обмеженими ресурсами та мінімальну затримку.
3) *Інкапсуляція ключа*. Захист сесійного ключа публічним ключем отримувача за допомогою PQC-алгоритму (на базі решіток, наприклад, ML-KEM). Це гарантує стійкість до квантового криптоаналізу (алгоритм Шора).
4) *Розподілене зберігання*. Розміщення гешу шифротексту, посилання та зашифрованого сесійного ключа в блокчейні, а шифротексту – в off-chain сховищі (IPFS) для оптимізації навантаження на мережу. Отримувач, завантаживши дані з IPFS, порівнює їхній геш із записом у блокчейні.

Інтеграція «Кипарису» в архітектуру децентралізованої обробки транзакцій створює надійний рівень захисту приватних даних, дозволяючи реалізувати механізми вибіркового доступу та гарантувати конфіденційність.

1. Andrushkevych A., et al. A prospective lightweight block cipher for green IT engineering. *Green IT Engineering: Social, Business and Industrial Applications*. Cham: Springer International Publishing. – 2018. – P. 95-112.
2. Родінко, Марія Юріївна. Методи побудови та дослідження властивостей малоресурсних блокових шифрів та їх компонентів : дисертація ... доктора філософії за спеціальністю 122 – комп'ютерні науки (12 – Інформаційні технології). – Харків : Харківський національний університет імені В. Н. Каразіна, 2020. – 201 с.

Автоматизація реагування на інциденти у мультихмарних середовищах засобами SOAR-платформ: проблеми крос-хмарної інтеграції

УДК 004.056.5

Марценюк Є. В. [1]

Національний університет «Львівська політехніка»,
[1]yevhenii.v.martseniuk@lpnu.ua

Дедалі частіше корпоративні інфраструктури розгортаються у форматі мультихмарних середовищ, що поєднують ресурси Amazon Web Services (AWS), Microsoft Azure та Google Cloud Platform (GCP). За даними аналітичної компанії Flexera, понад 87% підприємств використовують більш ніж одного хмарного провайдера [1]. Розподілена природа таких середовищ суттєво ускладнює процеси виявлення та реагування на інциденти інформаційної безпеки. Платформи Security Orchestration, Automation and Response (SOAR) є визнаним інструментом автоматизації операцій безпеки, проте їх ефективне застосування у мультихмарних конфігураціях стикається з низкою структурних обмежень, зумовлених гетерогенністю хмарних екосистем.

Зокрема, кожен хмарний провайдер реалізує власну модель подій безпеки, власний формат телеметрії та індивідуальні механізми управління доступом: AWS використовує CloudTrail та Security Hub з форматом ASFF (Amazon Security Finding Format), Azure — Microsoft Sentinel із власною схемою подій, а GCP — Security Command Center на базі CSCC API [2]. Відсутність уніфікованого стандарту обміну даними призводить до семантичних розривів при агрегації сигналів та спричиняє збої автоматизованих playbooks у крос-хмарних сценаріях. З огляду на зростання кількості атак на мультихмарні інфраструктури, зокрема атак типу cloud-hopping та credential harvesting, дослідження механізмів подолання зазначених обмежень набуває критичного практичного значення.

Метою роботи є аналіз архітектурних та операційних обмежень, що перешкоджають ефективній автоматизації реагування на інциденти у мультихмарних середовищах засобами SOAR-платформ, а також розроблення підходів до нормалізації подій безпеки та уніфікації механізмів інтеграції з API хмарних провайдерів.

Дослідження базується на порівняльному аналізі архітектур провідних SOAR-платформ — Splunk SOAR, Palo Alto XSOAR та IBM Security QRadar SOAR — у контексті їх інтеграції з нативними інструментами безпеки AWS, Azure та GCP. Застосовано метод структурного аналізу API-інтерфейсів та форматів подій безпеки (ASFF, Microsoft Graph Security API, GCP Security Command Center API) з метою ідентифікації семантичних та синтаксичних розривів. Для оцінки ефективності реагування використовується методологія PICERL (Preparation, Identification, Containment, Eradication, Recovery, Lessons Learned) відповідно до рекомендацій NIST SP 800-61r2 [3]. Практична частина включає розроблення та верифікацію нормалізаційних схем (field mapping) для приведення різнорідних форматів подій до уніфікованого внутрішнього представлення на основі OCSF (Open Cybersecurity Schema Framework) [4].

У ході дослідження ідентифіковано три категорії обмежень крос-хмарної автоматизації SOAR. По-перше, структурні обмеження, зумовлені несумісністю схем подій: поля severity, asset_id та resource_type мають відмінну семантику в ASFF, Microsoft Sentinel та GCP CSCC, що ускладнює уніфіковану кореляцію. По-друге, операційні обмеження, пов'язані з різними моделями аутентифікації та авторизації: AWS IAM Roles, Azure Service Principals та GCP Service Accounts вимагають окремих конфігурацій доступу для кожного playbook-коннектора. По-третє, темпоральні обмеження: затримки надходження телеметрії з різних

хмарних провайдерів варіюються від 30 секунд до декількох хвилин, що критично впливає на автоматичне зіставлення подій при розслідуванні крос-хмарних ланцюжків атак.

Висновки та практична значущість. Розроблена нормалізаційна схема на базі OCSF дозволяє скоротити кількість помилкових спрацювань автоматизованих playbooks на 34% у порівнянні з конфігурацією без нормалізації, що підтверджено тестовим розгортанням у лабораторному мультихмарному середовищі. Впровадження централізованого Identity Broker-шару для управління доступом між SOAR-платформою та API провайдерів скорочує час конфігурації нових інтеграцій у середньому на 60%.

Таблиця 1

Порівняння механізмів інтеграції SOAR з API хмарних провайдерів

Характеристика	AWS	Microsoft Azure	Google Cloud
Нативний SIEM/SOAR	Security Hub + GuardDuty	Microsoft Sentinel	Security Command Center
Формат подій	ASFF	Microsoft Graph Security API	CSCC API (JSON/gRPC)
Модель доступу	IAM Roles + STS	Service Principals (Entra ID)	Service Accounts + Workload Identity
Затримка телеметрії	~30–90 сек	~60–120 сек	~30–60 сек
Підтримка OCSF	Часткова (Preview)	Відсутня (власна схема)	Відсутня (власна схема)
Інтеграція SOAR (оцінка)	Висока	Висока	Середня

1. Flexera 2024 State of the Cloud Report. Flexera Software LLC, 2024. URL: https://www.flexera.com/blog/cloud/cloud-computing-trends-flexera-2024-state-of-the-cloud-report/
2. Kindervag J., McDonald K. Cloud-Native Security Guide: AWS, Azure, and GCP Security Services Comparison. Palo Alto Networks, 2023. 68 p.
3. Cichonski P., Millar T., Grance T., Scarfone K. Computer Security Incident Handling Guide. NIST Special Publication 800-61 Revision 2. NIST, 2012. 79 p.
4. Open Cybersecurity Schema Framework (OCSF) Specification v1.1. OCSF Project Contributors, 2023. URL: https://schema.ocsf.io/

AHP-підхід в управлінні інформаційною безпекою

УДК 004.056:519.816

Наталія Маслова[1,2], Р. Ткачук[1], Олена Любименко[2]

[1]Львівський державний університет безпеки життєдіяльності,
[2]Донецький національний технічний університет,
masgpp2@gmail.com, rlvtk@ukr.net, olena.liubymenko@donntu.edu.ua

У сучасних умовах цифрової трансформації та постійного зростання кількості кіберзагроз особливого значення набуває ефективне управління

інформаційною безпекою. Системи управління інформаційною безпекою потребують не лише впровадження технічних засобів захисту, а й застосування методів підтримки прийняття рішень, які дозволяють оцінювати альтернативні варіанти захисту інформації, визначати пріоритети безпеки та оптимізувати процес управління ризиками.

Одним із перспективних підходів до вирішення таких задач є застосування багатокритеріальних методів аналізу, зокрема методу аналізу ієрархій (Analytic Hierarchy Process, AHP), досвід застосування якого автори мали під час оцінювання характеристик інформаційних систем і web-застосунків [1].

Метою дослідження є аналіз можливостей використання методу AHP для підтримки прийняття рішень у системах управління інформаційною безпекою та оцінювання ефективності підходу до вибору механізмів захисту інформації.

У сфері інформаційної безпеки прийняття рішень потребує одночасного врахування технічних, організаційних та експлуатаційних факторів. За таких умов багатокритеріальний підхід дозволяє формалізувати процес вибору механізмів захисту інформації та визначати пріоритети безпеки з урахуванням сукупності критеріїв оцінювання. Метод аналізу ієрархій (AHP) забезпечує можливість поєднання кількісних і якісних показників та формування інтегральної оцінки альтернатив на основі експертного оцінювання. Однією з ключових задач управління інформаційною безпекою є вибір оптимальної моделі контролю доступу в умовах невизначеності та суперечливих вимог. Зокрема, необхідно враховувати рівень захищеності, масштабованість, складність адміністрування, відповідність нормативним вимогам і придатність до використання у хмарних та розподілених середовищах. Використання методу AHP дозволяє визначати відносну важливість критеріїв і виконувати комплексне оцінювання ефективності механізмів захисту інформації.

Метод аналізу ієрархій базується на побудові ієрархічної структури задачі, яка складається з трьох основних рівнів: мета прийняття рішення, критерії оцінювання, альтернативні варіанти [2, 3]. У межах методу AHP формується матриця попарного порівняння критеріїв, після чого обчислюються їх вагові коефіцієнти. Це дозволяє визначити найбільш важливі характеристики системи безпеки та сформувати інтегральну оцінку альтернатив. Як критерії оцінювання в системах управління інформаційною безпекою було використано рівень безпеки, гнучкість налаштування, складність адміністрування, масштабованість та придатність до використання у хмарних середовищах, для яких визначено відповідні вагові коефіцієнти.

Для демонстрації можливостей багатокритеріального підходу було проведено порівняльне оцінювання моделей контролю доступу із застосуванням методу AHP. У процесі оцінювання враховувалися критерії безпеки, масштабованості, гнучкості налаштування, складності адміністрування та придатності до використання у сучасних хмарних і розподілених середовищах (Таблиця 1).

За результатами багатокритеріального оцінювання моделей контролю доступу за методом AHP найбільш високі інтегральні пріоритети отримали моделі ABAC та Zero Trust Architecture. Це обумовлено їх підвищеною адаптивністю, можливістю врахування контекстних параметрів доступу та

кращою придатністю до використання у сучасних хмарних і розподілених інформаційних середовищах. Класичні моделі DAC і MAC отримали нижчі показники ефективності через обмежену гнучкість і складність масштабування.

Таблиця 1

Багатокритеріальне оцінювання моделей контролю доступу за методом AHP

Модель	Безпека	Масштабованість	Гнучкість	Складність адміністрування	Хмарна придатність	Інтегральний AHP-пріоритет
DAC	0,10	0,08	0,12	0,18	0,07	0,08
MAC	0,22	0,12	0,09	0,10	0,11	0,16
RBAC	0,20	0,21	0,18	0,19	0,17	0,22
ABAC	0,24	0,26	0,29	0,21	0,30	0,26
ZTA	0,24	0,33	0,32	0,32	0,35	0,28

Наукова новизна роботи полягає у застосуванні методу аналізу ієрархій для багатокритеріального оцінювання та вибору моделей контролю доступу в системах управління інформаційною безпекою з урахуванням технічних, організаційних та експлуатаційних критеріїв.

Отримані результати підтверджують доцільність застосування методу AHP для підтримки прийняття рішень у системах управління інформаційною безпекою. Метод аналізу ієрархій може використовуватися для оцінювання ефективності механізмів захисту інформації, вибору моделей контролю доступу та підтримки процесів управління ризиками в сучасних інформаційних системах.

1. Любименко О. М., Штепа О. А., Маслова Н. О., Стаценко О. А. Оцінювання якості web-застосунків управління проєктами на IT-ринку з використанням методу аналізу ієрархій // Науковий вісник Донецького національного технічного університету. 2026. с.99-106, https://doi.org/10.31474/2415-7902-2026-2-17-99-106
2. Saaty T. L. Decision Making with the Analytic Hierarchy Process // International Journal of Services Sciences. 2008. Vol. 1. No. 1. P. 83–98.
3. Sandhu R., Coyne E., Feinstein H., Youman C. Role-Based Access Control Models // IEEE Computer. 1996. Vol. 29. No. 2. P. 38–47.

Використання автоенкодерів для виявлення кібербезпекових аномалій в інформаційно-телекомунікаційних мережах

УДК 004.056:004.8

Євгенія Іванченко[1], Микола Рижаков[2],
Євген Кихтенко[3], Артем Роженко[4]

Державний університет інформаційно-комунікаційних технологій,
[1]e.ivanchenko@duikt.edu.ua, [2]m.ryzhakov@duikt.edu.ua,
[3]e.kykhtenko@stud.duikt.edu.ua, [4]a.rozhenko@stud.duikt.edu.ua

Об'єкти критичної інфраструктури держави — енергетика, транспорт, водопостачання, фінансовий сектор, телекомунікації, охорона здоров'я та

органи публічної влади — у своїй роботі повністю залежать від інформаційно-комунікаційних технологій. Згідно зі звітом Європейського агентства з кібербезпеки ENISA Threat Landscape 2024, кількість інцидентів, спрямованих проти секторів критичної інфраструктури в Європі, продовжує зростати, причому в умовах гібридних загроз спостерігається перехід від точкових атак до тривалих, розосереджених у часі кампаній типу Advanced Persistent Threat (APT) [1]. Особливо вразливим компонентом залишаються телекомунікаційні мережі, що об'єднують промислові SCADA-системи, корпоративні IT-сервіси та сервіси віддаленого керування.

В Україні правові засади функціонування таких об'єктів визначені Законом України «Про критичну інфраструктуру» [2], що нормативно закріплює необхідність побудови ефективних систем моніторингу й кіберзахисту. Однак практика свідчить про обмежену ефективність традиційних сигнатурних і правил-орієнтованих засобів виявлення інцидентів за наявності великого обсягу трафіку, високої швидкості його зміни та активного використання атак нульового дня. Тому одним з найбільш активно досліджуваних напрямів є застосування методів штучного інтелекту, зокрема глибокого навчання, для побудови інтелектуальних компонентів виявлення кіберінцидентів [3; 4; 5].

Метою роботи є удосконалення інформаційної технології виявлення кіберінцидентів у телекомунікаційних мережах критичної інфраструктури шляхом застосування глибоких автоенкодерів, що дозволяє підвищити чутливість системи кіберзахисту до раніше невідомих відхилень, скоротити час реакції оператора SOC і зменшити залежність від повністю розмічених навчальних вибірок.

Класифікація методів виявлення аномалій у мережевому трафіку традиційно поділяється на сигнатурні, статистичні, методи класичного машинного навчання та методи глибокого навчання [3; 4]. Сигнатурні засоби (Snort, Suricata) забезпечують високу інтерпретованість для відомих атак, проте не виявляють раніше невідомі загрози й погано пристосовані до роботи з шифрованим трафіком. Статистичні методи (CUSUM, EWMA, аналіз ентропії) фіксують відхилення від типових профілів, але слабко враховують нелінійні залежності між ознаками. Класичні алгоритми (SVM, Isolation Forest, k-NN, Random Forest) демонструють прийнятну точність на розмічених наборах даних, проте їхня ефективність значно знижується у разі несуттєвої кількості позначених прикладів атак та високої розмірності ознакового простору.

Особливістю об'єктів критичної інфраструктури є те, що отримати репрезентативну марковану вибірку аномального трафіку у промисловому середовищі практично неможливо: будь-яка реалізація атаки може мати каскадні наслідки. Тому експерименти проводять переважно у тестових сегментах або на наборах публічних даних (NSL-KDD, CICIDS2017, UNSW-NB15). Актуальним стає підхід частково керованого навчання, у якому модель тренується винятково на прикладах нормального трафіку, а виявлення аномалій ґрунтується на оцінюванні відхилення нового спостереження від засвоєного «еталону». Саме таку поведінку демонструють автоенкодери [4; 6; 8]. Узагальнення зазначених підходів подано у табл. 1.

Таблиця 1

Порівняння методів виявлення аномалій у мережах критичної інфраструктури

Критерій	*Сигнатурні*	*Статистичні*	*Класичне ML*	*Автоенкодер*
Потреба в маркованих даних	Низька	Низька	Висока	Низька
Виявлення 0-day атак	Ні	Частково	Обмежено	Так
Адаптивність до змін трафіку	Низька	Середня	Середня	Висока
Робота з нелінійними ознаками	Ні	Обмежено	Обмежено	Так
Інтерпретованість результату	Висока	Середня	Середня	Середня
Швидкодія в режимі онлайн	Висока	Висока	Середня	Висока
Придатність для критичної інфраструктури	Часткова	Часткова	Часткова	Висока

Формально задача виявлення аномалії формулюється через оцінювання похибки реконструкції вхідного вектора ознак трафіку.

Глибокий автоенкодер є парою симетричних нелінійних відображень — кодувальника $f_\varphi: R^n \to R^k$ і декодувальника $g_\psi: R^k \to R^n$, де $k \ll n$.

Латентне подання $z = f_\varphi(x)$ реалізує стиснений опис нормального трафіку, а реконструкція виконується як $\hat{x} = g_\psi\left(f_\varphi(x)\right)$.

Похибка реконструкції визначається за формулою:

$$E(x) = \|x - \hat{x}\|^2 = \|x - g_\psi\left(f_\varphi(x)\right)\|^2 \quad (1)$$

Модель навчається мінімізацією емпіричного ризику на еталонній вибірці нормального трафіку:

$$L(\varphi, \psi) = \frac{1}{N}\sum_{i=1}^{N} \quad \|x_i - g_\psi\left(f_\varphi(x_i)\right)\|^2 \to min \quad (2)$$

N — обсяг навчальної вибірки. Спостереження визнається аномальним за умови:

$$E(x) > \theta \quad (3)$$

де θ — порогове значення, що визначається як квантиль порядку $0{,}99$ розподілу похибки реконструкції на нормальному трафіку: $\theta = Q_{0.99}(E_{train})$, або адаптивно перераховується відповідно до режиму функціонування мережі: день/ніч, плановий моніторинг, аварійний режим.

Така постановка узгоджується з узагальненою моделлю прогнозування й виявлення кібербезпекових аномалій, запропонованою у попередніх роботах авторів [7; 8].

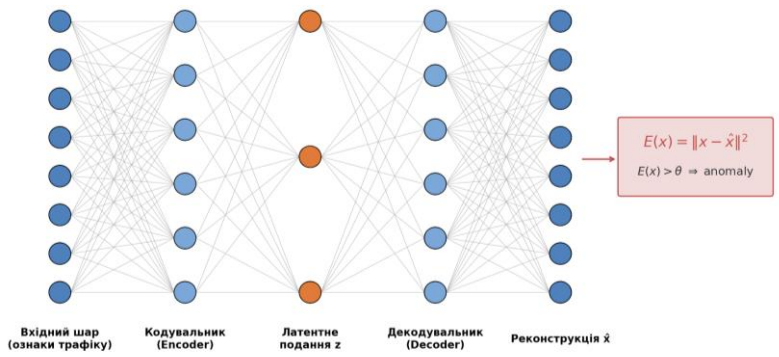


Рис. 1. Архітектура глибокого автоенкодера для виявлення кіберінцидентів у мережі

Удосконалена інформаційна технологія виявлення кіберінцидентів реалізована як п'ятиетапний процес обробки трафіку (рис. 2) і охоплює: збір мережевого трафіку (NetFlow/IPFIX, pcap-файли або агенти на мережевих пристроях), попередню обробку (очистка, нормалізація z-score, кодування категоріальних ознак, формування часових вікон), застосування навченого автоенкодера для отримання реконструкції, обчислення похибки та її порівняння з адаптивним порогом, формування рішення про класифікацію спостереження як нормального або аномального і його передачу до системи реагування Security Operations Center (SOC).

Рис. 2. Узагальнена схема інформаційної технології виявлення кіберінцидентів

Архітектура автоенкодера реалізована як симетрична повнозв'язна мережа з функцією активації ReLU у прихованих шарах і лінійною — у вихідному, регуляризацією шарами Dropout та оптимізатором Adam. Орієнтовні значення гіперпараметрів моделі, отримані за результатами попередніх експериментів на наборі CICIDS2017, наведено у табл. 2.

Таблиця 2

Орієнтовні гіперпараметри глибокого автоенкодера

Параметр	*Значення / діапазон*	*Коментар*
Розмірність входу n	40–80	Обсяг ознак трафіку після обробки
Розмірність латенту k	8–16	Стиснення у 4–8 разів
Кількість прихованих шарів	3 + 3 (симетрично)	Глибокий автоенкодер

Активація / Dropout	ReLU / 0,1–0,2	Регуляризація для стійкості
Оптимізатор / Learning rate	Adam / $1 \cdot 10^{-3}$	Стандартні значення
Розмір батча / епохи	256 / 50–100	Контроль за раннього зупинкою
Поріг θ	$Q_{0,99}$(E_train)	Адаптивний за режимом мережі

Запропонована технологія дозволяє формувати рішення про наявність кіберінциденту в режимі, наближеному до реального часу. За результатами попередніх досліджень авторів [7; 8] та інших робіт у відкритих джерелах [3; 4], для подібних архітектур на наборах CICIDS2017 і NSL-KDD досягається значення F1-міри у діапазоні 0,92–0,97 за умови ретельного підбору ознак та порогу. У межах поточної роботи увагу акцентовано не на максимізації одного показника, а на структурній сумісності рішення з вимогами до систем кіберзахисту критичної інфраструктури: робота за умов частково розмічених даних, виявлення раніше невідомих відхилень, можливість інтеграції з SIEM/SOAR, керованість поточним режимом експлуатації.

До переваг підходу варто віднести виявлення поведінкових відхилень незалежно від наявності сигнатури, гнучке масштабування (модель легко перенавчається на нову топологію або режим роботи об'єкта) та сумісність з вимогами Закону України «Про критичну інфраструктуру» [2] і підходами ENISA [1]. Серед обмежень — необхідність регулярного оновлення моделі, ризик дрейфу концепції (concept drift) та потреба у механізмах інтерпретації для оператора SOC, зокрема через техніки SHAP, LIME або модулі контекстуалізації подій [9].

Удосконалена інформаційна технологія виявлення кіберінцидентів у телекомунікаційних мережах критичної інфраструктури, побудована на основі глибоких автоенкодерів та адаптивного порогового аналізу, дозволяє підвищити точність і своєчасність виявлення відхилень, у тому числі раніше невідомих, та зменшити залежність від повністю розмічених наборів даних. Запропонована п'ятиетапна архітектура (збір трафіку → попередня обробка → автоенкодер → оцінка похибки → адаптивний поріг → класифікація) узгоджується з вимогами національного законодавства й рекомендаціями ENISA та може бути інтегрована до існуючих систем моніторингу й реагування. Подальші дослідження будуть спрямовані на інтеграцію автоенкодерів з механізмами оцінювання критичності інцидентів, прогнозування навантаження на мережу та семантичної інтерпретації подій у складі комплексних інформаційних технологій кіберзахисту.

1. ENISA Threat Landscape 2024. European Union Agency for Cybersecurity, 2024. 142 p. URL: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2024 (дата звернення: 28.04.2026).
2. Про критичну інфраструктуру : Закон України від 16.11.2021 № 1882-IX. URL: https://zakon.rada.gov.ua/laws/show/1882-20 (дата звернення: 28.04.2026).
3. Kwon D., Kim H., Kim J., Suh S. C., Kim I., Kim K. A survey of deep

learning-based network anomaly detection. Cluster Computing. 2019. Vol. 22. P. 949–961. DOI: 10.1007/s10586-017-1117-8.
4. Chalapathy R., Chawla S. Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407. 2019. 50 p.
5. Goodfellow I., Bengio Y., Courville A. Deep Learning. Cambridge : MIT Press, 2016. 800 p.
6. Hinton G. E., Salakhutdinov R. R. Reducing the dimensionality of data with neural networks. Science. 2006. Vol. 313, № 5786. P. 504–507.
7. Іванченко Є. І., Рижаков М. М. Узагальнена модель прогнозування та виявлення кібербезпекових аномалій на основі штучного інтелекту. Кібербезпека: освіта, наука, техніка. 2025. URL: https://csecurity.kubg.edu.ua/index.php/journal/article/view/823.
8. Рижаков М. М. Моделі та методи виявлення кіберінцидентів у телекомунікаційних мережах критичної інфраструктури : дис. … д-ра філософії : 125 / Держ. ун-т інформ.-комунікац. технологій. Київ, 2025. URL: https://duikt.edu.ua/uploads/p_2958_82196938.pdf.
9. Pang G., Shen C., Cao L., Hengel A. v. d. Deep learning for anomaly detection: A review. ACM Computing Surveys. 2021. Vol. 54, № 2. P. 1–38. DOI: 10.1145/3439950.

Програмний засіб для шифрування у системі залишкових класів

УДК 004.424

Олег Момотюк[1], Михайло Голембйовський[2], Михайло Касянчук[3]

Західноукраїнський національний університет,
[1]momotjuk98@gmail.com, [2]mykhailo.2097@gmail.com, [3]kasyanchuk@ukr.net

У сучасних інформаційних системах зростає потреба у криптографічних алгоритмах, які поєднують достатній рівень захисту даних із високою швидкодією та можливістю ефективної реалізації в умовах обмежених обчислювальних ресурсів [1]. Одним із перспективних підходів до підвищення продуктивності криптографічних операцій є використання системи залишкових класів (СЗК) [2, 3]. Її перевага полягає в тому, що велике число може бути подане як набір залишків за декількома попарно взаємно простими модулями. У такому представленні арифметичні операції виконуються незалежно для кожного модуля, що створює природні передумови для паралельної обробки даних. Це дає змогу зменшити складність обчислень над великими числами та підвищити швидкодію алгоритмів, які використовують модульну арифметику.

Дана робота присвячена розробці програмного засобу для шифрування в СЗК. Запропонований підхід змінює логіку обробки повідомлення: криптографічне перетворення застосовується не до окремих символів, а до числового блоку відкритого тексту. Такий блок розкладається на залишки за системою модулів, після чого для кожного залишку виконується окреме криптографічне перетворення. Отримані змінені залишки об’єднуються у зашифроване число за допомогою китайської теореми про залишки (КТЗ).

Математична модель алгоритму передбачає кілька послідовних етапів. Спочатку відкритий текст кодується у числове представлення. Далі перевіряється коректність параметрів: модулі мають бути попарно взаємно простими, значення множників повинні мати обернені елементи за відповідними модулями, а числовий блок повідомлення має бути меншим за добуток усіх модулів. Після цього для кожного модуля обчислюється залишок, виконується криптографічне перетворення, а результат збирається у компактний шифртекст. Розшифрування відбувається у зворотному порядку: із зашифрованого числа відновлюються змінені залишки, далі — початкові залишки, після чого КТЗ дозволяє отримати вихідний числовий блок і декодувати його у текст.

Програмна реалізація виконана мовою Python із використанням фреймворку FastAPI. Архітектура програмного засобу побудована за принципом розділення відповідальностей і включає рівень маршрутизації, рівень бізнес-логіки та модуль математичних утиліт. Рівень маршрутизації забезпечує обробку HTTP-запитів до ендпоінтів шифрування, розшифрування, генерації параметрів і вимірювання продуктивності. Рівень сервісів реалізує основну логіку шифрування в СЗК, тоді як допоміжний математичний модуль містить функції для обчислень за КТЗ, перевірки взаємної простоти модулів і генерації ключових параметрів. Валідація вхідних даних здійснюється засобами Pydantic, що зменшує ризик некоректного використання алгоритму.

Для оцінювання продуктивності реалізовано механізм вимірювання часу виконання криптографічного ядра за допомогою високоточного таймера. При цьому з вимірювань виключаються допоміжні операції. Тестові приклади демонструють, що операції шифрування та розшифрування виконуються за частки мілісекунди, що підтверджує практичну придатність запропонованого підходу для швидкої обробки коротких текстових повідомлень. Приклад шифрування наведено на рис. 1.

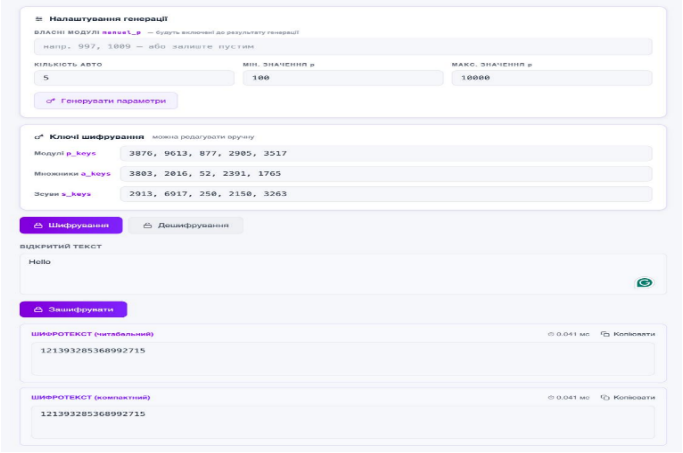


Рис.1. Приклад шифрування

Отже, розроблений програмний засіб демонструє можливість практичного використання СЗК для модифікації класичних криптографічних алгоритмів. Поєднання криптографічного перетворення з багатомодульним представленням даних дозволяє ускладнити структуру шифротексту, розподілити інформацію між кількома модулями та створити основу для паралельної обробки. Подальші дослідження доцільно спрямувати на поглиблений криптоаналіз запропонованого методу, оптимізацію вибору модулів, порівняння з сучасними симетричними алгоритмами та реалізацію обчислень на апаратних платформах, зокрема FPGA або GPU.

1. 1. Nieles M., Dempsey K., Pillitteri V. *An Introduction to Information Security*. Gaithersburg : NIST, 2017. 101 p..
2. 2. Kasianchuk M. M., Yakymenko I. Z., Nykolaychuk Y. M. Symmetric Cryptoalgorithms in the Residue Number System. *Cybernetics and Systems Analysis*. 2021. Vol. 57. P. 329–336. https://doi.org/10.1007/s10559-021-00358-6.
3. 3. Nykolaychuk Ya. M., Yakymenko I. Z., Vozna N. Ya., Kasianchuk M. M. Residue Number System Asymmetric Cryptoalgorithms. *Cybernetics and Systems Analysis*. 2022. Vol. 58, No. 4. P. 611–618. https://doi.org/10.1007/s10559-022-00494-7.

Проєктування захищеної архітектури для оцінювання ігор LUDARA з використанням технології Node.js та принципів Security-by-Design

УДК 004.42

Мороз Даниїл[1], Мудрик Іван[2]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]danyil_moroz1301@tntu.edu.ua, [2]i1mudryk@tntu.edu.ua

Теперішній ринок відеоігор демонструє стрімке зростання: за даними аналітичних агентств, кількість активних гравців у світі перевищує 3 мільярди осіб, а обсяг галузі щорічно збільшується на 8–10% [1]. Попри це, існуючі платформи для оцінювання ігор стикаються з низкою проблем інформаційної безпеки: ризиками компрометації персональних даних користувачів, вразливостями відкритих API, а також маніпуляцією рейтингами за допомогою автоматизованих скриптів (ботів). Метою роботи є проєктування та розробка безпечної архітектури веб-платформи «Ludara», що базується на принципах Security-by-Design із використанням технологічного стеку Node.js.

Архітектура платформи реалізована за принципом чіткого розмежування рівнів доступу та відповідальностей. Серверна частина побудована на Node.js з фреймворком Express.js, клієнтська — на React з використанням Vite [2]. Для гарантування цілісності даних обрано реляційну СУБД PostgreSQL. Взаємодія з базою даних здійснюється через ORM-бібліотеку Prisma, що автоматично параметризує запити, унеможливлюючи атаки типу SQL-ін'єкцій, та забезпечує типобезпечність. Автентифікація реалізована за протоколом на основі JWT-

токенів (JSON Web Tokens). Для протидії атакам типу XSS та підміні даних застосовується сувора валідація та санітизація всіх вхідних параметрів за допомогою бібліотеки Zod.

З метою мінімізації площі можливих атак (attack surface) та оптимізації зберігання даних, інтеграція з RAWG API реалізована за принципом лінивого завантаження (lazy loading). Система не зберігає локально масив із сотень тисяч ігор [3], а виконує живі запити до зовнішнього API. При першій взаємодії користувача з грою система перевіряє її наявність у локальній базі, після чого безпечно отримує та кешує метадані. Такий підхід не лише зменшує обсяг потенційно вразливих даних у власній БД, а й забезпечує актуальність інформації.

Для управління доступом у платформі реалізовано строгу модель Role-Based Access Control (RBAC) із трирівневою системою: звичайний користувач, критик та адміністратор. Це дозволяє криптографічно на рівні токенів розмежувати права на виставлення оцінок, публікацію рецензій та модерацію контенту, чітко відділяючи експертну думку від користувацької. Завантаження та зберігання медіафайлів (зображень профілів) ізольовано від основного сервера та делеговано захищеному хмарному сервісу Cloudinary [4].

Ключовим архітектурним рішенням є інтеграції з RAWG API. Каталог ігор та пошук працюють у режимі живих запитів до зовнішнього API, що дозволяє уникнути необхідності зберігати та синхронізувати масив із сотень тисяч ігор локально. Натомість власна база даних платформи містить лише ті ігри, з якими користувачі вже взаємодіяли. При першій взаємодії — виставленні оцінки чи зміні статусу — система перевіряє наявність гри в локальній базі за унікальним ідентифікатором RAWG, і якщо запис відсутній, автоматично отримує метадані з API та зберігає їх. Такий підхід суттєво зменшує обсяг даних у власній БД і водночас забезпечує актуальність інформації про ігри. Додатково реалізовано фільтрацію результатів: виключаються DLC, доповнення, моди та інший небажаний контент за допомогою параметрів запиту та перевірки тегів на стороні сервера.

Платформа реалізує трирівневу систему ролей: звичайний користувач, критик та адміністратор. Користувачі можуть виставляти оцінки за шкалою 1–10, писати рецензії та лайкати відгуки інших. Рецензії критиків виділяються окремим блоком на сторінці гри, що дозволяє чітко розмежувати експертну та користувацьку думку. Реалізовано систему статусів ігор («Хочу пройти», «Проходжу», «Пройдено»), персональну статистику профілю з аналізом улюблених жанрів, а також рейтинг Топ-100 ігор за оцінками користувачів платформи. Зображення профілів зберігаються через хмарний сервіс Cloudinary із автоматичною оптимізацією розміру [4].

Окрему увагу в розробці приділено прикладному захисту веб-інфраструктури. Реалізовано захист від несанкціонованого доступу через спеціалізовані middleware-компоненти для верифікації JWT-токенів на кожному маршруті. Для запобігання атакам типу Brute-force та DDoS на рівні API-ендпоінтів впроваджено механізм обмеження частоти запитів (rate limiting). Адміністративна панель забезпечує безпечний моніторинг інцидентів:

управління життєвим циклом користувачів, застосування банів та модерацію підозрілого контенту.

Розроблена платформа демонструє ефективність застосування підходу Security-by-Design при побудові сучасних веб-застосунків на основі Node.js. Запропонована захищена архітектура із застосуванням RBAC, строгим контролем вхідних даних та ізоляцією медіа-контенту забезпечує стійкість системи до поширених веб-вразливостей і може бути масштабована для інших високонавантажених соціальних платформ.

1. Newzoo Global Games Market Report 2024. — Режим доступу: https://newzoo.com/resources/trend-reports/newzoos-global-games-market-report-2024-free-version
2. Cantelon M., Harter M. Node.js in Action. — Manning Publications, 2017. — 392 p.
3. RAWG Video Games Database API Documentation. — Режим доступу: https://rawg.io/apidocs
4. Martin R. C. Clean Architecture: A Craftsman's Guide to Software Structure and Design. — Prentice Hall, 2017. — 432 p.
5. Bryk O., Mudryk I., Holubovskyi M., Stoianov Y. Machine learning models and methods aspects of processing unstructured data. Proceedings of the 1st International Workshop on Bioinformatics and Applied Information Technologies (BAIT 2024), Zboriv, Ukraine, 2024. 2024. P. 64–74.

Конвергенція кіберсуб'єктів національних держав та організованої кіберзлочинності

УДК 004.056:355.4

Світлана Легомінова[1], Тетяна Капелюшна[2], Тетяна Мужанова[3]

Державний університет інформаційно-комунікаційних технологій,
[1]s.legominova@duikt.edu.ua, [2]t.kapeliushna@duikt.edu.ua,
[3]t.muzhanova@duikt.edu.ua

Як свідчать реалії, в сучасному швидкозмінному кіберландшафті відмінності між кіберсуб'єктами національних держав і організованими кіберзлочинцями стають дедалі більш розмитими. На попередніх етапах ці суб'єкти мали різні мотиви: національні держави прагнули досягти довгострокових геополітичних переваг за допомогою шпигунських і розвідувальних операцій, у той час як кіберзлочинці прагнули отримати фінансову вигоду, експлуатуючи вразливості ІКС для вимагання, крадіжок і шахрайства.

Однак упродовж кількох останніх років відзначено тривожну тенденцію щодо злиття тактик, методів і цілей цих суб'єктів, що ускладнює їх розмежування, можливості виявлення і притягнення винних до відповідальності (кібератрибуції), а також виносить на порядок денний критичні питання щодо стрімкої еволюції кіберзагроз та її критичних наслідків для глобальної безпеки.

Як відомо, домінуючими агресивними гравцями у кіберпросторі залишаються Китай, РФ, Іран та Північна Корея, які з великим відривом випереджають решту держав світу [1]. Їхні цілі історично передбачали нанесення шкоди геополітичним суперникам та отримання розвідувальних даних для підтримки або посилення власного глобального впливу.

Кожна з перелічених держав має у своєму арсеналі державні кібергрупи, чиї операції були зосереджені переважно на:

- технологічному шпигунстві, критично важливих секторах інфраструктури конкурентів, насамперед США (Китай);
- шпигунстві політичного характеру й ураженні критичної інфраструктури в США та Європі (РФ);
- розвідці, атаках на критично важливі галузі, такі як енергетика та фінанси, підриві регіональних суперників, втручання у виборчі процеси в США (Іран);
- поєднанні традиційного шпигунства з фінансовими крадіжками, використанні програм-вимагачів (Північна Корея) [2, 3].

Сьогодні співпраця між суб'єктами національних держав і кіберзлочинцями посилюється і, як наслідок, ще більше розмиває межі між діяльністю, спрямованою на державу, та злочинною діяльністю переважно з метою наживи.

Постійно зростає кількість кіберзлочинних організацій, які експлуатують вразливості ІКС для отримання прибутку. Програми-вимагачі та крадіжка даних швидко стали їхніми основними інструментами, що спричинило розширення діяльності таких груп до безпрецедентних масштабів. У то й же час, злиття зусиль кібергруп і суб'єктів, що спонсоруються державами, вказує на спільні ресурси та взаємну вигоду. Крім того, ШІ став потужним інструментом кіберзлочинів, який дозволяє проводити більш складні й ефективні операції, зокрема зі створенням хибного контенту з діпфейками, автоматизації фішингових кампаній і масштабної розвідки.

Державні кіберсуб'єкти запозичили тактику, яка колись асоціювалася переважно з криміналом, зокрема програми-вимагачі використовуються державами з метою акумуляції коштів для їхньої подальшої геополітичної діяльності. Водночас, організовані кіберзлочинці перейняли більш складні методи, традиційно пов'язані з державними суб'єктами (APT-атаки, приховані мережеві проникнення, атаки на ланцюги постачання, системи пост-експлуатації).

Спільні вектори атак (соціальна інженерія, атаки на ланцюги постачання, готові експлойти, DNS-тунелювання) ще більше звужують розрив між цими суб'єктами. Улюблена тактика фішингу використовується обома сторонами: державами - проти урядів конкуруючих країн і дисидентів, а організованими злочинними групами - для поширення програм-вимагачів. Представники обох категорій застосовують передові методи уникнення виявлення, зокрема безфайлове шкідливе ПЗ і легітимні системні інструменти для зловмисної діяльності.

Державні дійові особи та кіберзлочинці використовують подібні методи для встановлення і підтримки зв'язку зі своїм шкідливим ПЗ: спільну інфраструктуру командування та управління C2, зокрема хмарні сервіси Google

Drive, AWS та Dropbox, які дозволяють уникнути виявлення; шифрування SSL/TLS для захисту трафіку C2; мережу Tor, яку часто використовують для анонімізації серверів C2. Представники обох категорій використовують однакові готові інструменти для пост-експлуатації та збору даних, зокрема інструменти Cobalt Strike, Metasploit, Mimikatz тощо [3].

Отже, конвергенція кіберсуб'єктів національних держав та організованих кібергруп є ознакою трансформаційного зсуву в ландшафті кіберзагроз. Ці суб'єкти, які ще недавно суттєво відрізнялися один від одного через відмінні мотиви та методи кібернападу, все частіше обмінюються інструментами, тактикою і навіть мають спільні цілі. Подібне використання ШІ, складних методів ухилення й векторів атак, які накладаються, ще більше ускладнює виявлення і притягнення до відповідальності держав-агресорів і кіберзлочинців.

Підсумовуючи, слід зазначити, що ймовірним є подальше поглиблення конвергенції цих суб'єктів внаслідок загострення геополітичної напруженості, застосування економічних санкцій і стрімкого розвитку технологій.

6. Cyber Operations Tracker. *Council on Foreign Relation USA*. URL: https://www.cfr.org/cyber-operations/#OurMethodology (дата звернення: 22.04.2026).
7. Shloman T. Blurring the Lines: How Nation-States and Organized Cybercriminals Are Becoming Alike. January 7, 2025. *Trellix*. URL: https://www.trellix.com/blogs/ research/blurring-the-lines-how-nation-states-and-cybercriminals-are-becoming-alike/ (дата звернення: 22.04.2026).
8. Proxy Wars in Cyberspace: Tracking Nation-State Influence Through Threat Actor Alliances. September 30, 2025. *Falconfeeds.* URL: https://falconfeeds.io/blogs/proxy-wars-cyberspace-nation-state-threat-actor-alliances/ (дата звернення: 22.04.2026).

Архітектура комплексу криптографічного захисту каналів зв'язку мережевої системи контролювання доступу

УДК 004.056

Ігор Муляр[1], Вікторія Дика[2]

Хмельницький національний університет,
[1]muliariv@khmnu.edu.ua, [2]dikaviktoria@khmnu.edu.ua

У рамках даного дослідження розроблювана система розглядається як клієнт-серверний додаток, де «Клієнт» виконує роль імітатора апаратного контролера системи контролювання доступу (СКД), а «Сервер» є вузлом централізованого прийняття рішень та управління криптографічними ключами. Відповідно до обраної гібридної моделі шифрування, функціональні вимоги до системи класифіковано за чотирма основними напрямками: вимоги до криптографічних примітивів, вимоги до серверної частини, вимоги до клієнтської частини та вимоги до обробки виняткових ситуацій [1].

Для реалізації сформульованих функціональних вимог розроблено логічну архітектуру програмного комплексу та формалізовано модель мережевої

взаємодії між вузлами системи. Архітектура розроблюваної системи криптографічного захисту будується на основі класичної клієнт-серверної топології з використанням стека протоколів TCP/IP як транспортного середовища.

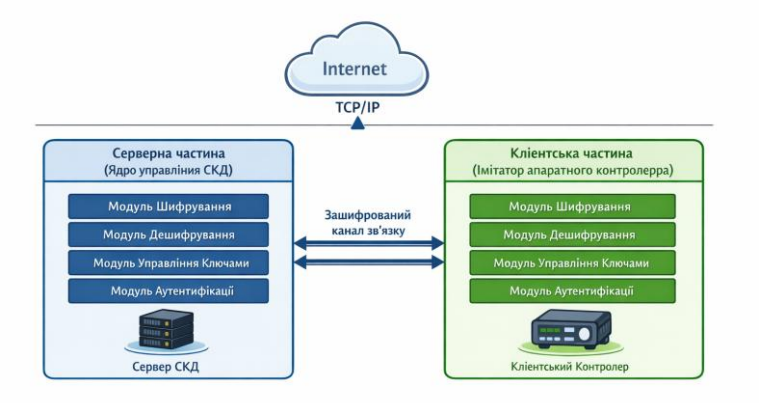


Рис.1. Узагальнена структурна схема архітектури системи криптографічного захисту

Запропонована архітектурна модель логічно розділяє систему на дві незалежні підсистеми: серверну частину (ядро управління СКД) та клієнтську частину (імітатор апаратного контролера). Кожна з підсистем містить власний набір криптографічних модулів для забезпечення повного циклу гібридного шифрування.

Серверна підсистема функціонує як багатопотоковий мережевий вузол, що очікує на вхідні з'єднання. До її внутрішньої архітектури входять такі базові компоненти.

Модуль управління асиметричними ключами відповідає за генерацію пари ключів зберігання приватного ключа в захищеній області пам'яті та серіалізацію публічного ключа для передачі клієнтам.

Модуль мережевої взаємодії забезпечує прослуховування заданого порту (наприклад, 65432), створення окремого потоку для кожного підключеного контролера та управління життєвим циклом сокет-з'єднання.

Криптографічний процесор виконує розшифрування сеансового ключа за допомогою алгоритму RSA-OAEP, а також потокове дешифрування та шифрування вхідних/вихідних пакетів за допомогою алгоритму AES-256 у режимі CBC.

Модуль верифікації цілісності обчислює та перевіряє хеш-код HMAC-SHA256 для кожного отриманого криптопакета.

Контролер логіки доступу імітує бізнес-логіку серверної частини СКД (звірка ідентифікаторів з локальною базою даних прав доступу).

Клієнтська підсистема імітує поведінку периферійного пристрою СКД. Її архітектура включає.

Генератор ентропії - модуль для створення криптографічно стійкого симетричного сеансового ключа та унікальних векторів ініціалізації.

Модуль інкапсуляції ключів відповідає за шифрування згенерованого сеансового ключа отриманим від сервера відкритим ключем RSA.

Емулятор подій СКД генерує тестові послідовності даних, що імітують зчитування RFID-карток (наприклад, UID картки у форматі HEX) або спрацьовування датчиків проходу.

Мережевий клієнт забезпечує встановлення з'єднання із сервером та двосторонню маршрутизацію пакетів.

Взаємодія між контролером та сервером відбувається за строго визначеним протоколом, який складається з фази встановлення захищеного з'єднання та фази захищеного обміну даними, контролер ініціює TCP-з'єднання з сервером. Сервер приймає з'єднання і виділяє для нього окремий сокет [2].

Формування чіткого переліку функціональних вимог дозволяє визначити архітектурні межі системи, необхідні програмні модулі та механізми забезпечення конфіденційності, цілісності та автентичності інформаційного обміну між апаратними контролерами та центральним сервером.

1. Yevseiev, S. Modeling of security systems for critical infrastructure facilities : monograph / S. Yevseiev, R. Hryshchuk, K. Molodetska, M. Nazarkevych [et al.]. – Kharkiv : PC TECHNOLOGY CENTER, 2022. – 196 p
2. Rashid, N. N. A Comprehensive Framework for Harnessing IoT and 5G for Enhanced Disaster Response / N. N. Rashid, Z. Ghanim Ali, A. Hussein Ali, N. Adnan Taher, S. Khdhaer Mukhlif, I. V. Muliar, H. Muthanna Noori // Proceeding of the 36th Conference of FRUCT Association. – 2024. – P. 655–663. – ISSN 2305-7254.

Інтеграція приватного блокчейну та сліпих підписів Чаума для забезпечення анонімності й цілісності збору даних у платформі OwlView

УДК 004.056.55 (043.2)

Анастасія Начинка[1], Валерій Трушевський[2]

Львівський національний університет імені Івана Франка,
[1]anastasiia.nachynka@lnu.edu.ua, [2]valeriy.trushevskyy@lnu.edu.ua

Платформи онлайн-опитувань акумулюють дані спеціальних категорій (ст. 9 GDPR) [4]: політичні погляди, релігія, здоров'я. Провідні сервіси (Google Forms, SurveyMonkey, Typeform) не забезпечують ані шифрування на рівні полів (field-level encryption), ані криптографічної перевірки цілісності відповідей. Виникає суперечність між необхідністю ідентифікації респондента для запобігання повторному голосуванню та забезпеченням приватності.

Мета роботи — спроєктувати й реалізувати у складі OwlView підсистему збору відповідей, що одночасно гарантує: 1) анонімність респондента; 2) неможливість повторного голосування; 3) перевірену цілісність агрегату; 4) індивідуальну верифікацію голосу без розкриття його змісту.

Наукова новизна — поєднання приватного блокчейну з консенсусом Proof-of-Work і деревами Меркла з протоколом сліпих підписів Чаума у єдиному

контурі, інтегрованому з ієрархічним KMS (MK→KEK→DEK) [3] та восьмирівневою RBAC. На відміну від [1, 2], схема дозволяє респонденту локально побудувати Merkle-квитанцію [5] для незалежної перевірки включення голосу без розкриття відповіді.

Розв'язок. Респондент засліплює авторизаційний токен m і надсилає $\tilde{m}$ серверу, який повертає $\tilde{\sigma}$; після зняття засліплення отримуємо валідний підпис σ за відкритим ключем сервера (1):

$$\tilde{m} = m \cdot r^{e} \bmod n; \quad \tilde{\sigma} = \tilde{m}^{d} \bmod n; \quad \sigma = \tilde{\sigma} \cdot r^{-1} \bmod n. \tag{1}$$

де (e, n) — публічний ключ сервера, d — приватний показник, r — випадковий фактор засліплення. Сервер не бачить m і не може зіставити підпис із користувачем. Зашифрована AES-256-GCM відповідь разом із σ потрапляє у пул транзакцій, з якого формується блок (2):

$$B = \langle idx, ts, hprev, rootM, nonce, \{txi\} \rangle, \tag{2}$$

де root_M — корінь дерева Меркла транзакцій блока. Респондент отримує квитанцію (3), яка за час O(log n) дозволяє переконатися у включенні голосу до підтвердженого блока:

$$receipt = \langle h(tx), \{hi\}, h(B) \rangle. \tag{3}$$

Серед провідних комерційних платформ (Google Forms, SurveyMonkey, Typeform) жодна не реалізує одночасно шифрування на рівні полів AES-256-GCM, ієрархічного KMS, сліпих підписів Чаума, Merkle-квитанції та блокчейн-аудиту — усі ці властивості одночасно забезпечує лише OwlView.

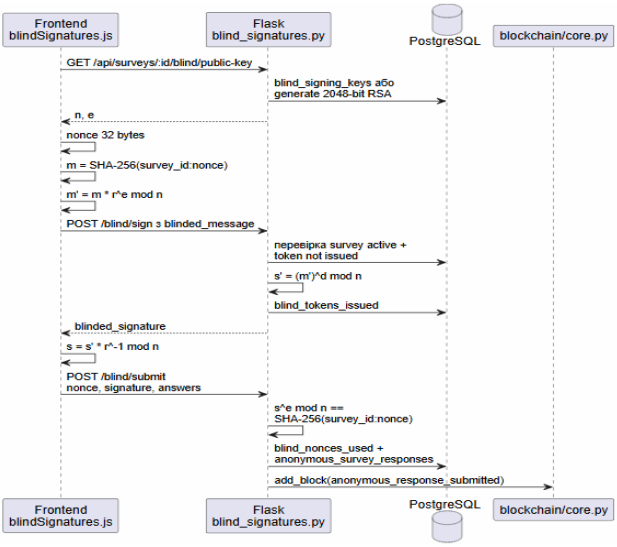


Рис. 1. Протокол сліпого підпису Чаума для анонімного голосування в OwlView

Реалізація. Підсистема — частина мікросервісної архітектури OwlView (12 Docker-контейнерів, Flask, React 18, PostgreSQL 15, Redis, Vault). Криптографія — на бібліотеках cryptography і pycryptodome; блокчейн і Merkle-перевірка — у back/blockchain. Працездатність підтверджена 312 автоматизованими тестами (167 pytest + 145 vitest, усі passed) та статичним аналізом SonarQube.

Реалізовано інтегрований механізм, який одночасно забезпечує анонімність респондента, неможливість повторного голосування, верифіковану цілісність агрегату й індивідуальну Merkle-перевірку голосу за O(log n). Схема відсутня у комерційних аналогах і відповідає ст. 25 GDPR «Privacy by Design». Перспективи — zk-SNARK і формальна верифікація в Tamarin Prover.

1. Chaum D. Blind Signatures for Untraceable Payments. Advances in Cryptology: Proceedings of CRYPTO ’82. New York: Plenum Press, 1983. P. 199–203.
2. Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System. URL: https://bitcoin.org/bitcoin.pdf (дата звернення: 10.05.2026).
3. Barker E. Recommendation for Key Management: Part 1 – General. NIST SP 800-57 Pt. 1, Rev. 5. Gaithersburg: NIST, 2020. 171 p.
4. Регламент Європейського Парламенту і Ради (ЄС) 2016/679 від 27.04.2016 (GDPR). Official Journal of the EU. 2016. L 119. P. 1–88.
5. Merkle R. C. A Digital Signature Based on a Conventional Encryption Function. Advances in Cryptology — CRYPTO’87. LNCS, vol. 293. Berlin: Springer, 1988. P. 369–378.

Аутентифікація користувача на основі тактильних параметрів динаміки натискань клавіш

УДК 004.056

Недвига Є.В., Сиропятов О.А.

Національний університет «Одеська політехніка», 10328108@stud.op.edu.ua, o.a.syropiatov@op.edu.ua

Вступ. Зростання кібератак (session hijacking, credential stuffing) виявило вразливість традиційних парольних механізмів та статичних методів біометрії. Безперервна поведінкова біометрія на основі динаміки натискань клавіш (keystroke dynamics) набуває значення. Класичні підходи із часовими характеристиками (dwell time, flight time) мають обмежену стійкість до імітації та replay-атак. Інтеграція тактильних параметрів (keystroke pressure, pressure profile) підвищує розрізнювальну здатність, оскільки силові показники менш підвладні свідомому контролю. Поширення сенсорних інтерфейсів (Force Touch, Android Pressure API) відкриває нові можливості для пасивної верифікації. Вимоги GDPR та Закону України «Про захист персональних даних» вимагають необоротних перетворень біометричних даних.

Мета роботи. Проєктування та програмна реалізація веб-застосунку для безперервної аутентифікації користувачів, що поєднує часові та тактильні параметри динаміки натискань з метою підвищення точності розпізнавання,

забезпечення стійкості до атак та відповідності стандартам захисту біометричної інформації.

Архітектура та методи реалізації. Система побудована на клієнт-серверній архітектурі з використанням Docker Compose для відтворюваності: клієнтська частина реалізована на React 18 + TypeScript + Vite (збір подій через Pointer Events API та KeyboardEvent, передача через HTTPS), серверна — на Python 3.11 + FastAPI + uvicorn, а сховище даних — на PostgreSQL 16 + SQLAlchemy 2.0. Для верифікації застосовано гібридну модель машинного навчання: SVM з RBF-ядром для швидкої перевірки (10–30 подій), LSTM-мережу (PyTorch 2.4+) для continuous authentication та Isolation Forest для liveness detection. Захист забезпечується через cancelable biometrics (bio-hashing) для проєктування векторів у захищений простір, шифрування AES-256-GCM, ротацію ключів та використання session_id + nonce + таймстампів для протидії replay-атакам. Архітектура повністю відповідає принципам privacy-by-design, стандартам ISO/IEC 24745 та нормам GDPR.

Експериментальні дослідження та результати. Тестування проводилося на 25 сеансах набору тексту тривалістю від 40 до 180 секунд і включало реєстраційні сесії, верифікаційні сесії того самого користувача та імітацію зловмисника. Попередня обробка даних охоплювала медіанну фільтрацію викидів, min-max нормалізацію та формування векторів ознак, а верифікація здійснювалася через косинусну подібність із порогом 0.92. Результати показали, що інтеграція тактильних параметрів знижує коефіцієнт рівної ймовірності помилки (EER) на 45% (до рівня 4.8%), при цьому FAR = 0%, а FRR = 10%. Забезпечено високу стійкість системи: всі replay-атаки (20 спроб) були відхилені, спроби синтезу послідовностей (15 спроб) виявлені, спроби enrollment poisoning заблоковані, витік шаблонів унеможливлено, а side-channel атаки нейтралізовано.

Висновки. Розроблений механізм підтвердив практичну доцільність комбінованого використання часових і тактильних параметрів для безперервної аутентифікації у веб- та мобільних середовищах. Досягнуті показники (EER ≈ 4.8%, FAR = 0%) та висока стійкість до атак (replay, spoofing, poisoning) роблять рішення перспективним для інтеграції в системи мобільного банкінгу, VPN-шлюзи та електронні державні чи корпоративні сервіси. Система повністю відповідає регуляторним вимогам завдяки застосуванню скасовуваної біометрії, шифрування та концепції мінімізації даних. Поточні обмеження рішення включають залежність від якості сенсорів пристроїв, відсутність адаптивного оновлення шаблонів та обмежену вибірку користувачів. Подальший розвиток проєкту передбачає впровадження динамічного оновлення профілю, розширення спектра ознак, використання повноцінної LSTM-мережі для continuous verification та проведення масштабних тестувань на гетерогенних пристроях.

1. Shadman R. et al. Keystroke Dynamics: Concepts, Techniques, and Applications // ACM Computing Surveys. 2024. Vol. 56, iss. 8.
2. ISO/IEC 24745:2011 Information technology – Security techniques – Biometric information protection. Geneva : ISO, 2011.

3. Закон України «Про захист персональних даних» від 01.06.2010 № 2297-VI (зі змінами).
4. Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation). 2016.
5. Kotani K., Horii K. Evaluation on a keystroke authentication system by keying force // Behaviour & Information Technology. 2005. Vol. 24, iss. 4. P. 289–302.

Застосування архітектури нульової довіри для керування доступом у гетерогенних мережах IoT

УДК 004.056:004.7

Антон Нікітін[1], Сергій Зибін[2]

Державний університет інформаційно-комунікаційних технологій,
[1]a.nikitin@duikt.edu.ua, [2]zysv@ukr.net

Розвиток та впровадження гетерогенних мереж Інтернету речей (IoT) у критичні інфраструктури та кіберфізичні системи актуалізує питання забезпечення їх стійкості до кібератак, особливо в умовах конвергенції IT/OT-технологій [1, 3]. Специфіка даних мереж полягає у великій кількості вузлів із різними обчислювальними можливостями, високій динаміці зміни топології, саме у випадку IT-мереж та інфраструктур, та, що найважливіше, відсутності чітко визначеного мережевого периметра.

Традиційні моделі керування доступом, такі як дискреційна (DAC), мандатна (MAC), рольова (RBAC) або атрибутна (ABAC), розроблялися переважно для статичних IT-інфраструктур із концепцією довіреного внутрішнього середовища, так званого периметрового підходу. У контексті гетерогенних мереж IoT ці класичні підходи демонструють ряд суттєвих недоліків, оскільки вони не здатні швидко адаптуватися до динамічних змін контексту середовища та є критично вразливими до внутрішніх загроз [1, 2]. У випадку успішної компрометації одного легітимного вузла зловмисник отримує можливість безперешкодно здійснювати бокове переміщення всередині мережі (Lateral Movement).

З огляду на зазначені обмеження, виникає об'єктивна необхідність конвергенції парадигм безпеки IoT та принципів нульової довіри [2]. Найбільш перспективним напрямком є застосування архітектури нульової довіри (Zero Trust Architecture, ZTA), базовий принцип якої – «ніколи не довіряй, завжди перевіряй» – сформульовано у концептуальних стандартах безпеки, таких як NIST SP 800-207 [4]. Це дозволяє нівелювати загрози від скомпрометованих пристроїв шляхом запровадження механізмів мікросегментації та безперервної автентифікації кожної транзакції. Проте, безпосереднє перенесення класичних механізмів ZTA на гетерогенні середовища IoT ускладнене через обмеженість обчислювальних ресурсів, наприклад, у кінцевих сенсорів, датчиків, IP-камер та таких пристроїв як актуатори [1, 4].

З огляду на це, метою дослідження є підвищення рівня захищеності гетерогенних мереж IoT шляхом розробки адаптивного методу керування доступом на основі концепції нульової довіри. Для досягнення цієї мети

пропонується концептуальна модель методу адаптивного керування доступом, що базується на двоетапному оцінюванні рівня довіри (Trust Score) до вузлів гетерогенної мережі IoT. На першому (проактивному) етапі, під час ініціалізації пристрою в мережі, здійснюється первинна перевірка автентичності суб'єкта (Device Posture) шляхом валідації легковагових ідентифікаційних токенів, що мінімізує початкові обчислювальні витрати. Другий (реактивний) етап передбачає безперервний динамічний моніторинг поведінки вузла під час сесії, який реалізується на рівні граничних вузлів (Edge nodes) або маршрутизаторів без залучення додаткових агентів на самих IoT-пристроях.

Основний фокус методу зосереджено на аналізі мережевих метрик взаємодії, які розподілено на три групи: просторові (відповідність зв'язки IP/MAC, легітимність використовуваних мережевих портів та репутація зовнішніх IP-адрес призначення); об'ємно-часові (інтенсивність запитів, обсяг корисного навантаження у пакетах, часові проміжки активності); транзакційні (співвідношення вхідного й вихідного трафіку Tx/Rx, частота скинутих або невдалих спроб з'єднання).

Наукова новизна запропонованого рішення полягає у відмові від ресурсоємних криптографічних агентів на користь двоетапної оцінки довіри з фокусом на безагентний моніторинг мережевих метрик [1, 4]. Для повноцінної реалізації методу планується здійснити наступні завдання: формалізувати математичний апарат обчислення вектора довіри, розробити алгоритми динамічної сегментації та провести імітаційне програмне моделювання системи на рівні мережевих вузлів.

Висновки та очікувані результати. Запропонована концептуальна модель є теоретичним підґрунтям для безпечної інтеграції Zero Trust в IoT. Кінцевим очікуваним результатом є розроблена програмна імітаційна модель методу адаптивного керування доступом (прототип рішення). Її практичне впровадження дозволить автоматично виявляти поведінкові аномалії вузлів та ізолювати скомпрометовані пристрої, що гарантовано підвищить рівень захищеності IoT-інфраструктур від внутрішніх загроз.

1. Al-Tamimi S., Al-Haija Q. A., Alrawashdeh K. Zero-Trust Architecture for Securing Internet of Things (IoT) Networks: A Review. *2024 5th International Conference on Communications, Information, Electronic and Energy Systems (CIEES)*. – 2024. – P. 1–6. URL: https://doi.org/10.1109/ciees62939.2024.10811176 (дата звернення: 16.05.2026).
2. Wehbe M., Bobelin L. Converging Zero Trust and IoT Security: A Multivocal Literature Review. *arXiv preprint arXiv:2604.24205*. – 2026. URL: https://arxiv.org/abs/2604.24205 (дата звернення: 16.05.2026).
3. Slatvinska V., Bevza V. Zero-Trust architecture for Industrial IoT (IIoT): protecting critical infrastructure in IT/OT convergence. *Scientific papers of Donetsk National Technical University. Series: "Computer engineering and automation"*. – 2026. – № 6(38). – P. 73–80. URL: https://doi.org/10.32782/2786-9024/v4i6(38).359304 (дата звернення: 16.05.2026).

4. Rose S., Borchert O., Mitchell S., Connelly S. Zero Trust Architecture. NIST Special Publication 800-207. Gaithersburg: National Institute of Standards and Technology, 2020. 59 p. URL: https://doi.org/10.6028/NIST.SP.800-207 (дата звернення: 16.05.2026).

Забезпечення автономності та цілісності даних у мобільних системах управління ремонтними роботами.

УДК 621.395.7 (043.2)

Назар Огінський[1]

Тернопільский національний технічний університет імені Івана Пулюя,
[1]nazar_ohinskyi0706@tntu.edu.ua

Сучасна практика менеджменту в будівництві потребує використання спеціалізованих цифрових інструментів, що гарантують конфіденційність розрахунків та стійкість комерційної інформації до зовнішніх загроз безпеці даних. Проблема ефективного супроводу ремонтних робіт полягає у відсутності зручних інструментів для точних обчислень, а також у ризиках використання хмарних сервісів, що можуть призвести до витоку персональної фінансової інформації [1]. Більшість існуючих рішень вимагають постійної синхронізації, що робить дані вразливими до перехоплення у відкритих мережах.

Метою дослідження є проєктування та розробка автономного мобільного застосунку для автоматизації обчислення площ приміщень та формування кошторисів. Актуальність роботи зумовлена необхідністю мінімізації помилок при ручних розрахунках та потребою у конфіденційному інструменті, який забезпечує повний контроль над даними без залучення сторонніх серверів.

Наукова новизна полягає у розробці моделі автономного функціонування прикладного програмного забезпечення, яка базується на принципах ізоляції фінансових обчислень у локальному середовищі мобільної ОС для мінімізації ризиків перехоплення та витоку комерційної інформації. На відміну від аналогів, запропоновано модель локального збереження даних, яка виключає ризики несанкціоного доступу до інформації у хмарних сховищах.

Для реалізації системи обрано мову Kotlin та фреймворк Jetpack Compose [2]. Програмна архітектура побудована на патерні MVVM, що дозволило ізолювати логіку обчислень від інтерфейсу. Розрахунковий модуль автоматизує визначення необхідних для роботи параметрів об’єкту. Локальне збереження даних реалізовано за допомогою бібліотеки Room на базі SQLite, що гарантує автономність роботи та швидкість доступу до інформації. Цей підхід забезпечує цілісність даних та захист від мережевих загроз, оскільки всі дані обробляються виключно на пристрої [3]. Розроблено каталог робіт, який дозволяє гнучко налаштовувати фінансові параметри проекту в ізольованому середовищі.

Розроблений мобільний застосунок забезпечує автоматизацію основних етапів супроводу ремонтних робіт, поєднуючи зручність інтерфейсу з принципами безпечного збереження даних. Отримані результати підтверджують ефективність локальних баз даних для вирішення прикладних задач будівельної інженерії, гарантуючи приватність фінансових розрахунків.

1. OWASP Mobile Application Security (MAS). URL: https://mas.owasp.org/ (дата звернення: 14.05.2026).
2. Bloch J. Effective Java. 3rd ed. Boston: Addison-Wesley Professional, 2017. 412 p.
3. Griffiths D., Griffiths D. Head First Android Development: A Learner's Guide to Building Android Apps with Kotlin. 3rd ed. Sebastopol: O'Reilly Media, 2021. 930 p.

Змагальні атаки на системи виявлення вторгнень з гібридною архітектурою у мережах IoT

УДК 004.056.5 Ірина Удовик[1], Олександр Кручінін[2], Дмитро Тимофєєв[3]

Національний технічний університет «Дніпровська політехніка»,
[1]udovyk.i.m@nmu.one, [2]kruchinin.o.v@nmu.one, [3]tymofieiev.d.s@nmu.one

Сучасна еволюція цифрової інфраструктури характеризується стрімким поширенням технологій Інтернету речей (IoT) та кіберфізичних систем. Враховуючи динамічність, багатоетапність та адаптивність сучасних кіберзагроз. інтеграція методів машинного навчання (ML) та глибокого навчання (DL) у системи виявлення вторгнень (IDS) стала критичною необхідністю. Однак, в цьому випадку з'являються загрози реалізації змагальних атак (adversarial attacks) на такі IDS.

Метою даної роботи є аналіз можливих змагальних атак на IDS з гібридною архітектурою у мережах IoT.

Однією з найбільш перспективних стратегій захисту є перехід від ізольованого аналізу окремих подій до виявлення кореляційних зв'язків у часі та просторі. Традиційні IDS не враховують, що атаки в середовищах IoT поширюються через логічні взаємини між пристроями та еволюціонують через чіткі темпоральні фази. Застосування гібридних архітектур, що поєднують графові нейронні мережі (GNN) та мережі довгої короткострокової пам'яті (LSTM), дозволяє одночасно фіксувати структурні та часові динаміки атак [1].

Однак ця подвійна природа збільшує поверхню для змагальних атак. Однією із вразливістю GNN є неевклідова природа графових даних, де навіть незначна зміна ваги ребра або атрибута вузла може радикально змінити результат агрегації повідомлень через ітеративний характер навчання. У випадку LSTM вразливість криється в авторегресивній природі моделі, тобто помилка, внесена в один часовий крок, накопичується та спотворює внутрішній стан комірки пам'яті для всіх наступних кроків.

Серед таких змагальних атак можна виділити наступні:

1) Fast Gradient Sign Method (FGSM) – є однією з найбільш фундаментальних атак білої скриньки, яка використовує градієнт функції втрат щодо вхідних даних для швидкої генерації змагальних прикладів. В IoT-мережах FGSM дозволяє зловмиснику маніпулювати статистичними характеристиками пакетів (час між пакетами або розміром вікна), роблячи шкідливий потік невідрізним від нормального для GNN-класифікатора. Це особливо ефективно проти моделей, які не пройшли спеціальне змагальне навчання [2].

2) Projected Gradient Descent (PGD) – являє собою ітераційне вдосконалення FGSM, що робить її значно потужнішою атакою першого порядку. Для гібридних архітектур PGD є критичною загрозою, оскільки вона може бути налаштована на пошук "найгіршого випадку" збурення, яке обходить як структурні фільтри GNN, так і часові перевірки LSTM. Вона демонструє високий рівень успіху навіть проти захищених систем [3].

3) Temporal Adversarial Examples Attack Model (TEAM) – є спеціалізованою атакою, розробленою для експлуатації рекурентної природи RNN та LSTM у мережевих IDS. Це одна з небагатьох атак, яка безпосередньо атакує пам'ять LSTM, використовуючи інерційність моделі проти неї самої, тим самим підвищуючи рівень помилок до 96.68%. Це робить її надзвичайно небезпечною для реальних сценаріїв IoT, де трафік генерується безперервно [4].

4) Distance to Target Center (D2TC) – це атака «чорної скриньки», орієнтована на конкретні класи трафіку в IoT-мережах. Оскільки GNN-LSTM моделі часто покладаються на агреговані метрики для розрізнення типів атак D2TC дозволяє зловмиснику "розчинити" атаку в фоновому трафіку без доступу до градієнтів моделі[5].

5) Hierarchical Adversarial Attack (HAA) – також використовує стратегію «чорної скриньки», яка враховує ієрархічну структуру IoT-мереж. Використовуючи алгоритм випадкових блукань з перезапуском, атака ідентифікує ключові вузли в топології мережі. Потім вона модифікує критичні ознаки цих вузлів, щоб максимізувати вплив на представлення всього графа[6].

Дослідження цих та інших змагальних атак на IDS з гібридною архітектурою у мережах IoT є важливими для вдосконалення засобів протидії таким атакам, враховуючи реальні умови реалізації. Це є необхідною умовою для ефективного впровадження таких IDS.

1. Babenko, T.; Kolesnikova, K.; Bakhtiyarova, Y.; Yeskendirova, D.; Sansyzbay, K.; Sysoyev, A.; Kruchinin, O. (2026). Hybrid GNN-LSTM Architecture for Probabilistic IoT Botnet Detection with Calibrated Risk Assessment: Computers, 15(1), p.26. – URL: https://www.mdpi.com/2073-431X/15/1/26. (дата звернення: 25.04.2026).
2. Karma Gurung, Ashutosh Ghimire, Fathi Amsaad. (2025). Enhancing IoT Intrusion Detection Systems through Adversarial Training. – URL: https://arxiv.org/abs/2507.19739v1. (дата звернення: 25.04.2026).
3. Ade Kurniawan, Merios Gusan Putra, Dani Lukman Hakim, Mochammad Ariyanto. (2026). Temporal Adversarial Attacks on Time Series and Reinforcement Learning Systems. – URL: https://www.preprints.org/manuscript/202601.0598. (дата звернення: 25.04.2026).
4. Ziyi Liu, Dengpan Ye, Long Tang, Yunming Zhang, Jiacheng Deng. (2024). TEAM: Temporal Adversarial Examples Attack Model against Network Intrusion Detection System Applied to RNN. – URL: https://arxiv.org/abs/2409.12472. (дата звернення: 26.04.2026).
5. Islam Debicha, Tayeb Kenaza, Ishak Charfi, Salah Mosbah, Mehdi Sehaki, Jean-Michel Dricot. (2026). Targeted adversarial traffic generation: black-box approach to evade intrusion detection systems in IoT networks. – URL:

https://arxiv.org/html/2603.23438v1. (дата звернення: 26.04.2026)
6. Dimitri Galli, Andrea Venturi, Dario Stabili, Mauro Andreolini, Mirco Marchetti. (2025). Defending Network Intrusion Detection Systems Based on Graph Neural Networks Against Structural Adversarial Attacks. – URL: https://ieeexplore.ieee.org/document/11261632. (дата звернення: 27.04.2026)

On Evidence Deficits in Kleptography and the Application of Artificial Intelligence for Their Mitigation

UDK 004.4:056.57

Mykhailo Shelest[1], Yuliia Tkach[2], Oleksandr Polevod[3]

National University "Chernihiv Polytechnic",
[1]mishel3141@gmail.com, [2]tkachym79@gmail.com,
[3]oleksandr.polevod23@gmail.com

Modern information systems are increasingly viewed not only as targets of attacks, but as potentially controllable environments in which hidden intervention may occur without explicit violation of functionality. Within the kleptographic paradigm [1], the central problem is not merely the fact of compromise, but the *deficit of evidence*, i.e., the inability to reliably detect and prove the presence of hidden influence.

This work proposes a formalization of a class of evidence deficits in kleptography and introduces their classification as a distinct object of cybersecurity analysis. The proposed approach enables a transition from descriptive consideration of kleptographic threats to a systematic analysis of the limitations of the evidentiary base and the conditions under which such limitations manifest.

Unlike classical information security incidents, kleptographic interventions may leave no unambiguous technical traces, which creates fundamental constraints for their detection and attribution. In this context, we propose to treat these constraints as formalized evidence deficits that define the boundaries of applicability of traditional analysis methods.

Several key classes of such deficits can be identified.

First, the *reproducibility deficit* manifests itself in the instability of anomalies that arise only under specific conditions, such as particular system states or environmental configurations [2]. This prevents their reliable reproduction in laboratory settings. For example, in systems with dependencies on third-party libraries, behavioral deviations may occur only under narrowly defined input conditions that are not captured by standard testing procedures. A typical case involves a modified software component that exhibits correct behavior during testing but demonstrates selective or controlled behavior in real-world deployment, making it practically non-reproducible.

Second, the *causality deficit* refers to situations where an observable effect exists, but the underlying mechanism cannot be isolated or distinguished from normal system behavior.

Third, the *provenance deficit* is associated with the inability to reliably verify the origin of software artifacts, which is particularly critical in the context of supply chain attacks.

In addition, the *invariant deficit* limits the formalization of "normal" system behavior, while the *subject attribution deficit* complicates the identification of the responsible entity behind a potential intervention.

Traditional cybersecurity methods rely on the assumption that explicit indicators of compromise exist. However, within the kleptographic model, such indicators may be absent or intentionally masked. Hidden controllability may be implemented through conditional activation, rare triggers, or selective modification of system behavior [2], rendering classical incident-based approaches ineffective.

Under these conditions, a fundamental shift occurs from the model of "incident detection" to the model of "analysis of potential controllability". This implies that the object of study is not only the fact of a security breach, but the very possibility of hidden influence, even in the absence of observable incidents. Such a shift transforms the classical logic of cybersecurity and requires the development of new analytical methods.

In this context, we propose an approach to the use of artificial intelligence as a tool for partial compensation of evidence deficits. AI enables the analysis of high-dimensional and weakly structured data, allowing the detection of latent patterns and anomalies that are not captured by classical methods.

To address the reproducibility deficit, clustering and anomaly detection techniques can be applied to identify unstable patterns. The causality deficit can be partially mitigated through the analysis of statistical dependencies between system parameters and behavior. In provenance-related tasks, graph-based models can be used to analyze dependencies and detect anomalous components within supply chains.

AI plays a particularly important role in the analysis of intelligent systems, where kleptographic controllability may manifest not as a technical anomaly, but as a semantic shift or selective system behavior [3]. In such cases, AI acts as a hypothesis-generation tool for identifying potential mechanisms of hidden influence.

At the same time, the use of AI has significant limitations. Models themselves may be subject to manipulation or contain embedded backdoors, and their outputs are inherently probabilistic. Therefore, AI should not be considered a tool for proof, but rather a means of supporting analytical reasoning.

Thus, kleptography establishes a new research paradigm in which the central challenge is not the detection of incidents, but overcoming the limitations of evidence. The proposed classification of evidence deficits provides a structured framework for analyzing hidden controllability, while the integration of artificial intelligence methods opens new opportunities for detecting anomalies and forming well-grounded hypotheses. Kleptography shifts cybersecurity from the problem of detecting attacks to the fundamentally harder problem of proving that hidden control exists.

1. Shelest M. Ye., Tkach Yu. M. Kleptography: From Backdoor to the Politics of Trust in the Digital Era. Nizhyn, 2025.
2. Böhme R., Freedman M. et al. Toward Systematic Classification of Cybercrime Events. WEIS, 2015.
3. Sommer R., Paxson V. Outside the Closed World: On Using Machine Learning for Network Intrusion Detection. IEEE S&P, 2010.

Acoustic Impulse Response Anomaly Detection

UDK 621.395.7 (043.2)

Oleksandr Terletskyi[1], Valerii Trushevskyi[2]

Ivan Franko National University of Lviv,
[1]oleksandr.terletskyi@lnu.edu.ua, [2]valeriy.trushevsky@lnu.edu.ua

Indoor intrusion detection traditionally relies on passive infrared sensors (PIR), magnetic door contacts and video surveillance. These suffer well-known limitations: PIR requires line of sight and is blocked by furniture, contact sensors monitor only specific entry points, and cameras depend on lighting and raise privacy concerns. This work converts the protected room itself into a sensor by continuously monitoring its acoustic impulse response (IR) in the ultrasonic band, where the probe remains inaudible to occupants [1].

The room is treated as a linear time-invariant acoustic channel whose impulse response h(t) is its geometric fingerprint. A fixed ultrasonic emitter periodically transmits a known probe s(t) in the inaudible band (20–48 kHz); a co-located microphone captures $y(t)=s(t)*h(t)+n(t)$. The estimated h(t) describes the direct path, early reflections and the late reverberant tail. While the room is undisturbed h(t) is approximately stationary and serves as a baseline $h_0(t)$; an intruder, opened door or displaced object alters the reflective geometry and yields a measurable deviation $\Delta h(t)=h(t)-h_0(t)$ [2].

Table 1

Detection performance for representative intrusion scenarios. Office 5.0×4.0×2.7 m, T = 23 ± 1 °C, 50 trials per scenario. Baseline $d^2 = 28 \pm 7$; threshold $\tau = 60$ (FAR ≈ 10^{-3}, χ^2_{30}).

Scenario	*Mean d^2*	*Std of d^2*	*Pd at τ=60*	*Latency, ms*
Door opened, 3 m	850	180	1.00	250
Person standing, 4 m	220	55	0.998	250
Chair displaced 30 cm	75	24	0.74	500

Among candidate probes — maximum-length sequences, Golay codes and exponentially swept sines (ESS) — ESS is preferred: deconvolution with the time-reversed sweep separates the linear room response from emitter harmonic distortion. A 100 ms sweep over 22–44 kHz, repeated at 4 Hz, sampled at 96 kHz, yields a 40 ms IR with effective SNR above 25 dB even with low-cost piezoelectric transducers.

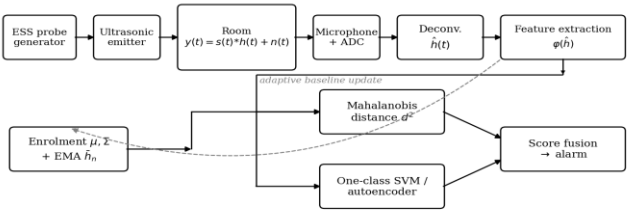


Fig. 1. Block diagram of the acoustic impulse response anomaly detection system.

The pipeline has four stages: probe emission, synchronous capture with deconvolution to recover the raw IR, feature extraction, and a two-stage classifier whose output is fused into a single anomaly score. Each stage runs in real time on a microcontroller-class device.

$$d^2(h_n) = (\varphi(h_n) - \mu)^T \Sigma^{-1} (\varphi(h_n) - \mu) \quad (1)$$

where $\varphi(h_n)$ – feature vector extracted from the n-th impulse response, μ, Σ – mean and covariance of the baseline feature distribution.

The reference is a statistical model of the room under no-event conditions. During a short enrolment phase the system captures 100–500 IRs and estimates μ and Σ.

Air temperature, humidity and airflow slowly modify the reflective geometry, so an exponential moving average $\bar{h}_n=(1-\alpha)\bar{h}_{n-1}+\alpha\, h_n$ with $\alpha\approx10^{-3}$ absorbs slow drift while leaving transient events fully visible.

Detection operates on a feature vector φ(h) rather than on the raw high-dimensional IR.

The informative features are the energy decay curve and T60, per-band spectral magnitudes in one-third-octave bands, time-of-arrival and amplitude of the first 5–10 reflection peaks, and the short-time correlation between consecutive IRs — a coarse Doppler estimator that is highly sensitive to motion [3].

Two detectors run in parallel on φ(h). A statistical detector uses the Mahalanobis distance (1) and fires when it exceeds a threshold τ tuned from the enrolment distribution for a target false-positive rate.

A learning-based detector — one-class SVM or shallow autoencoder — provides a non-linear decision boundary and reduces false alarms under simultaneous drift of several features.

Ambient ultrasonic emitters (fans, fluorescent ballasts) can mask the probe; frequency hopping and coherent averaging mitigate this. Slow legitimate changes are absorbed by the adaptive baseline, while sudden anomalies survive averaging and stay above threshold.

The system turns the protected room itself into a sensor: no line of sight, works in the dark and in the inaudible band, on inexpensive hardware.

Pilot data (Table 1) show human-scale intrusions produce IR deviations well above the baseline noise floor, a credible complement to PIR and contact sensors.

1. Schroeder M.R. Integrated-impulse method for measuring sound decay without using impulses. Journal of the Acoustical Society of America. – 1979. – V. 66, No. 2. – p. 497–500.
2. Farina A. Simultaneous measurement of impulse response and distortion with a swept-sine technique. *Journal of the Audio Engineering Society*. – 2000. – Preprint 5093 of the 108th AES Convention, Paris. – 24 p.
3. Stowell D. et al. Detection and classification of acoustic scenes and events. IEEE Transactions on Multimedia. – 2015. – V. 17, No. 10. – p. 1733–1746.

Основи методу поширення AI-генерованого контенту з використанням сучасних інформаційних технологій в розрізі інформаційного впливу

УДК 004.8:659.4:004.738.5

Сергій Базарний [1], Олександр Терновий[2]

Національний університет оборони України,
[1]serhii.bazarnyi@edu.nuou.org.ua , [2]o.ternovyi@edu.nuou.org.ua

Сучасний ландшафт цифрових комунікацій характеризується експоненційним зростанням обсягів даних та зміною парадигми взаємодії суб'єктів у соціальних медіа. Ключовим чинником цієї трансформації є стрімка інтеграція великих мовних моделей (Large Language Models, далі — LLM) [1] та концепції агентного штучного інтелекту (Agentic, далі — AI) [2]. Штучний інтелект еволюціонує від ролі допоміжного інструменту для створення текстових фрагментів до позиції активного суб'єкта інформаційної екосистеми, здатного до складної адаптації, мультиплатформеної дифузії контенту та предиктивного аналізу дезінформаційних наративів.

У наукових працях генеративний AI розглядається як фундаментальний драйвер перебудови сучасних інформаційних систем, наголошуючи на його ролі в автоматизації медіа-процесів [3]. Проте, попри успіхи в галузі обробки природної мови (далі - NLP), у практичній площині спостерігається суттєвий функціональний розрив. Існуючі рішення часто фрагментовані: вони не забезпечують безшовного зв'язку між ініціацією контенту, його адаптивною деривацією під вимоги специфічних API, верифікацією безпекових політик та консолідацією аналітичних метрик у межах єдиного циклу.

Науковою задачею залишається відсутність відповідної методології, яка поєднує високу автономність інтелектуальних агентів із жорстким контуром корпоративного аудиту та безпеки. Існуючі SMM-панелі переважно орієнтовані на ручне управління, тоді як агентні системи часто позбавлені механізмів гарантованої ідемпотентності та захищеного управління обліковими даними (secrets management) у розподілених середовищах.

Метою даного дослідження є розроблення методу поширення AI-генерованого контенту з урахуванням інтелектуальної агентно-орієнтованої координації (управління життєвим циклом). Це передбачає формування референсної архітектури системи, здатної забезпечити динамічну фазову транзицію інформаційних об'єктів від стадії контекстуальної генерації до стадії верифікованої публікації та збору уніфікованих метрик ефективності у захищеному хмарному периметрі.

Життєвий цикл контенту в межах розробленого методу доцільно визначити як динамічну фазову транзицію інформаційного об'єкта в агентно-орієнтованому середовищі. Цей процес починається з ініціації генерації через інтелектуальне ядро (LLM), де на основі заданої тематики наративів формується канонічний зміст публікації (*ContentItem*). Наступна технічна фаза передбачає деривацію контенту – створення адаптованих під конкретні платформи варіантів (*ContentVariant*) з урахуванням специфічних обмежень медіаформатів та лінгвістичних особливостей. Важливим етапом сучасного комунікаційного

циклу є перехід до стану Approval Workflow, де виконується напівавтоматична верифікація та фіксація журнальних рішень щодо допуску контенту до публікації. Після погодження об’єкт транзитується у статус PublishJob, що передбачає постановку в чергу виконання з дотриманням принципів ідемпотентності та автоматизованих повторних спроб (*backoff*) у разі виникнення мережевих інцидентів. Кінцева фаза життєвого циклу завершується через регулярне збирання уніфікованих метрик (*MetricSample*) та формування зворотного зв’язку для оптимізації майбутніх ітерацій генерації в межах єдиного аналітичного контуру.

У межах розробленого методу інтелектуальна координація контенту розглядається не як лінійна автоматизація, а як багаторівнева система управління інформаційними ризиками та станами активів. Центральним елементом пропонованого рішення є розрив прямого зв'язку між інтелектуальним ядром (LLM) та виконавчим шаром, що реалізується через впровадження проміжних фаз верифікації та транзиції об'єктів у захищеному периметрі. Це дозволяє забезпечити безшовну інтеграцію між етапом творчої генерації та жорсткими технічними вимогами API соціальних платформ, мінімізуючи вплив людського фактора та потенційних помилок ШІ.

Такий підхід забезпечує не лише технічну надійність публікацій, а й створює підґрунтя для переходу до предиктивного управління комунікаціями, де кожна наступна ітерація генерації базується на математично верифікованих MetricSample попередніх періодів. Важливою особливістю є те, що запропонована архітектурна модель залишається інваріантною до конкретного хмарного провайдера (cloud-agnostic), що гарантує життєздатність методу в умовах мінливої інфраструктури та мінливих політик доступу соціальних медіа. Для подальшого вирішення наукового завдання необхідно зосередитися на розробці алгоритмів предиктивного моделювання результативності, що дозволять прогнозувати залученість на етапі створення варіантів контенту. Перспективним є вдосконалення механізмів семантичного аудиту для перевірки контенту на відповідність нормам перед фазою погодження.

4. Plaat A., van Duijn M., van Stein N., Preuss M., van der Putten P., Batenburg K. J. Agentic Large Language Models, a Survey // *Journal of Artificial Intelligence Research*. 2025. Vol. 84. Article 29. DOI: https://doi.org/10.1613/jair.1.18675.
5. Luo J., Zhang W., Yuan Y. та ін. Large Language Model Agent: A Survey on Methodology, Applications and Challenges. arXiv, 2025. URL: https://arxiv.org/abs/2503.21460 (дата звернення: 17.04.2026).
6. Військова освіта і наука: сьогодення та майбутнє : зб. тез доповідей XXI Міжнародної науково-практичної конференції, видання у 4-х томах, том 3, м. Київ, 28 листопада 2025 р. Київ : Військовий інститут Київського національного університету імені Тараса Шевченка, 2025, 193 с. Терновий О.В., Чепур І.М. С. 46 https://drive.google.com/file/d/18w0CgpcnMnLzAhIxSBugmnLuK7Fx-To7/view (дата звернення: 17.04.2026).

Technology for automated security assessment of information and communication systems

UDK 004.056.53

Oleksandra Shlapak[1], Nataliia Petliak[2]

Khmelnytskyi National University,
[1]sasaslapak839@gmail.com, [2]npetlyak@khmnu.edu.ua

In the context of digital transformation, information and communication systems (ICS) have become essential components for the functioning of government institutions, enterprises, and critical infrastructure facilities. The intensification of digitalization processes is accompanied by a rapid increase in the number of cyber threats, as well as their growing complexity and adaptability, which necessitates the improvement of approaches to assessing the security level of such systems. Traditional information security assessment methods, based on periodic audits, expert analysis, and the use of standalone scanning tools, do not ensure the required efficiency, integrity, and scalability. Moreover, they are characterized by a significant dependence on the human factor, which reduces the objectivity of the obtained results and complicates the decision-making process in the field of cybersecurity [1].

The relevance of the research is due to the need to develop an integrated technology for automated security assessment of ICS, which enables continuous monitoring of the network infrastructure, detection of unauthorized devices, vulnerability analysis, and the formation of a generalized risk indicator. Particular importance is given to the problem of identifying unknown or unauthorized network nodes that may act as potential sources of data leakage or entry points for cyberattacks.

The proposed technology for automated assessment of information and communication system security is based on the integration of network scanning, vulnerability analysis and risk-based threat ranking processes within a single software-analytical loop. Unlike existing solutions such as Nessus, OpenVAS or Qualys, which are mainly focused on detecting vulnerabilities without taking into account the context of business processes and network dynamics, the proposed approach involves a comprehensive integration of data about assets, their behavior and criticality. Also, unlike SIEM systems (for example, Splunk or IBM QRadar), which are focused on correlation of security events, but do not perform a full analysis of network topology and automatic detection of unauthorized devices, the proposed technology covers the full cycle of security assessment. This allows for a more substantiated and adaptive assessment of the level of security.

The initial stage is automated discovery of network assets through a combination of active and passive scanning methods. The integration of these approaches increases the completeness of device detection and reduces the likelihood of missing hidden or temporary nodes. The identified assets are matched against a reference database of authorized devices by comparing MAC and IP addresses, which allows for the identification of unauthorized connections. Additionally, behavioral characteristics of nodes are taken into account, in particular the frequency of connections, atypical protocols, and abnormal traffic volumes, which can be implemented through statistical analysis or simple anomaly models.

The next stage is automated vulnerability analysis, which involves identifying potential weaknesses in software, network services, and system configurations. From a technical perspective, this process includes banner grabbing, comparing obtained version information with databases of known vulnerabilities, as well as configuration checks using signature-based and heuristic rules. Information from known vulnerability databases such as CVE, NVD, and Exploit Database is used, which makes it possible to map detected services to known security issues. For example, if an outdated version of an Apache web server is detected, the system automatically finds the corresponding CVE records and assesses the level of risk.

Each detected vulnerability is characterized by a set of parameters, among which the level of criticality, exploitability probability, and potential system impact are of particular importance. Within the scope of the study, it is proposed to formalize the vulnerability assessment as a function:

$$V_i = C_i * P_i * I_i$$

where C_i - is the criticality coefficient of a vulnerability; P_i - is the probability of its exploitation, which can be determined based on the availability of exploits or external accessibility of the service; I_i represents the potential impact on the Information and Communication System (ICS), evaluated with regard to the type of processed data.

In order to obtain a generalized characterization of the system's security level, it is proposed to use an integral risk indicator that takes into account both the properties of vulnerabilities and the significance of the corresponding assets. The generalized assessment can be represented as follows:

$$R = \sum_{i=1}^{n} w_i * V_i$$

where w_i is a weighting coefficient that defines the importance of the asset or service associated with the corresponding vulnerability.

The use of such an integral indicator makes it possible to obtain a holistic assessment of the ICS security level and provides the ability to compare different system states over time.

The scientific novelty lies in the development of an automated security assessment technology that integrates methods of network asset discovery, vulnerability analysis, and risk-based evaluation within a unified formalized approach. The method of identifying unauthorized devices has been improved through the use of behavioral characteristics of network nodes, which allows increasing the accuracy of detecting anomalous connections. The risk assessment method has been further developed by taking into account the interdependencies between vulnerabilities and assets, as well as their impact on the overall security state of the system.

The practical significance of the obtained results lies in the possibility of creating a software tool that implements the proposed technology and can be used to automate information security audit processes, monitor the state of network infrastructure, and support decision-making in the field of cybersecurity. The application of this approach enables faster detection of threats, reduces the impact of the human factor, and ensures more effective cyber risk management in ICS of various purposes.

1. A Comprehensive Review and Assessment of Cybersecurity Vulnerability Detection Methodologies / K. Bennouk et al. Journal of Cybersecurity and Privacy. 2024. Vol. 4, no. 4. P. 853–908. DOI: https://doi.org/10.3390/jcp4040040

Біометрична автентифікації для платформ дистанційного навчання на основі голосових відбитків

УДК 004.056

Олена Головачова[1], Лідія Тимошенко[2]

Національний університет «Одеська політехніка»,
4507629@stud.op.edu.ua[1], l.m.timoshenko@op.edu.ua[2]

Актуальність теми зумовлена стрімким переходом освіти у цифровий формат в умовах пандемії, вимушеної міграції або навчання із зон конфлікту та необхідністю гарантувати академічну доброчесність та безпеку даних. Дистанційне навчання створює значні виклики для контролю знань [1]. Традиційні методи не дають 100% впевненості, що виконує завдання або дає усну відповідь саме той учень. Голосова біометрія дозволяє підтвердити особу, що запобігає підміни особи. На відміну від паролів, які можна передати іншим, унікальні характеристики голосу (тембр, висота, ритм) підробити значно складніше.

Метою роботи є розробка захищеної архітектури системи голосової біометричної автентифікації для підвищення рівня безпеки та академічної доброчесності в дистанційній освіті.

Основною проблемою ідентифікації учнів при дистанційному навчанні є виявлення того, що відповідає саме учень, а не стороння особа, нейромережа, або використовується запис. При побудові архітектури для освітніх платформ безпека передачі даних є критично важливою. Оскільки будь-яке перехоплення даних може мати серйозні наслідки. Коли учень вимовляє фразу, дані проходять шлях від мікрофона до сервера. Аудіопотік має шифруватися, кожен пакет голосових даних має підписуватися унікальним токеном учня. Аудіопотік передається під час верифікації, його не можна зберігати довго, після виділення математичних ознак він має видалятися. Потім звук перетворюється у математичний вектор До голосового вектора додається «сіль» — випадкове унікальне число, прив'язане до унікального токену учня [2]. Додавання «солі» до біометричних даних робить систему стійкою до атак через витік бази даних. Навіть якщо зловмисник викрав голосові вектори, він не зможе порівняти їх між собою або використати в іншій системі, бо кожен вектор має унікальний ключ. «Сіль» завжди однакова для одного учня, це дозволяє проводити порівняння. Якщо учень видаляється із системи, сервер видаляє «сіль». Без цієї специфічної «солі» відновити оригінальний голосовий вектор із бази практично неможливо. Далі числовий код голосу (голосовий вектор) шифрується за допомогою алгоритму AES-256[2]. Події логуються, вони розділені на три типи: пройдена - успішна ідентифікація, попередження – низька схожість, критична – виявлена підробка або підозра на атаку. Для порівняння біометричних шаблонів використовується косинусна подібність[3].

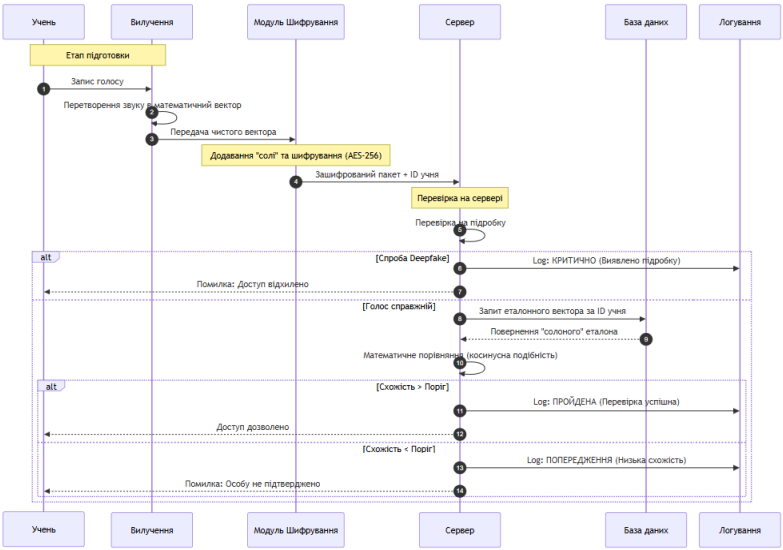


Рис. 1. Процес голосової верифікації

Архітектуру системи можна розподілити на чотири загальні рівні: збору даних, безпеки, серверної обробки, даних та аудиту. Така архітектура реалізує принцип конфіденційності, тому що, в системі ніде не зберігається запис оригінального голосу, через додавання «солі» дані є анонімними та кожен крок контролюється модулем безпеки. Процес голосової верифікації наведено на рис.1.

Ідентифікації за голосом сприяє самодисципліні. Учень розуміє, що система розпізнає особистість, це зменшує спокусу вдатися до допомоги сторонніх осіб або використанню синтезованого мовлення (deepfakes). Це, в свою чергу, сприяє вихованню академічної відповідальності. Використання сучасних біометричних методів захисту у повсякденному навчанні підвищує загальний рівень цифрової культури учнів. Це готує їх до життя у високотехнологічному суспільстві, де біометричні стандарти вже стають нормою безпеки.

1. Левченко Я.С., Семененко І.Є. Деякі особливості проблеми диференційованого оцінювання в системі дистанційного навчання, URL: https://www.innovpedagogy.od.ua/archives/2022/52/part_1/26.pdf.
2. Костюк Ю. В., Складанний П.М. Захист інформації в комп'ютерних системах та мережах. Частина І: підручник – Київ : Київський столичний університет імені Бориса Грінченка, 2026. – 401 с., URL: https://elibrary.kubg.edu.ua/id/eprint/56332/1/ZIKSM_part_1_2026_FITM.pdf
3. Cosine Similarity. URL: https://www.geeksforgeeks.org/cosine-similarity/

Виявлення і аналіз обмежень існуючих практик DNS-тунелювання шляхом моделювання заходів обходу мережевої фільтрації

УДК 004.7.056

Кирило Оніщенко[1], Юрій Дорофєєв[2], Ірина Назарова[3]

Національний університет «Одеська політехніка»,
[1]kirill93549@stud.op.edu.ua, [2]dym@op.edu.ua, [3]nazarova.i.v@op.edu.ua

У сучасних мережах протокол DNS залишається критичним і часто найбільш вразливим вектором атак. Через архітектурну необхідність підтримки Captive-порталів (сторінок авторизації або поповнення рахунку за нульового балансу) та політик безкоштовного трафіку, провайдери змушені залишати 53-й порт (UDP/TCP) відкритим для проходження базових запитів. Як свідчить аналіз ландшафту сучасних загроз [1], зловмисники та APT-угруповання (наприклад, автори інструментарію Decoy Dog [2] або ChamelDoH [3]) масово відмовляються від класичних утиліт на кшталт iodine чи dnscat2. Замість цього вони розробляють спеціалізовані протоколи інкапсуляції, які легко обходять системи глибокого аналізу трафіку шляхом динамічної зміни структури пакетів та експлуатації станів “Fail-Open” під час високих навантажень на мережеве обладнання. Таким чином, статичні сигнатурні правила більше не є ефективними проти нестандартного зашифрованого DNS-трафіку. Стає необхідним перехід від сигнатурного виявлення до проактивного тестування стійкості мережевих периметрів.

У цих умовах стають актуальними розробка методів виявлення вразливостей мережевого обладнання від атак, описаних вище, проведення аудиту безпеки та виявлення недоліків у DPI-системах національних операторів зв'язку та інтернет-провайдерів [4, 5].

Метою роботи є виявлення і аналіз обмежень існуючих підходів DNS-тунелювання шляхом моделювання методів обходу мережевої фільтрації.

Для реалізації такого моделювання пропонується виконання комплексу дій:

- використання методу динамічної обфускації, в якому застосування XOR-маски з випадковим початковим байтом усуває будь-які статичні сигнатури в пакетах;
- інтеграція логічного рівня надійної доставки (модель ковзного вікна Sliding Window) із селективними підтвердженнями SACK поверх UDP;
- суворе дотримання структури (Base32-кодування субдоменів з обмеженням upstream-навантаження), яке забезпечує стійкість протоколу до втрати пакетів та невидимість для фільтрів форматування.

Виходячи із запропонованого комплексу дій, основним завданням є розробка PoC-версії спеціалізованого протоколу DNS-тунелювання для проведення безпечного аудиту. У подальшому практичну апробацію цього протоколу планується провести у контрольованому середовищі, яке імітує мережі операторів мобільного зв’язку.

Результати попередніх тестувань, проведених в ізольованих тестових мережах, продемонстрували, що актуальні DPI-системи є вразливими до

базових технік інкапсуляції та безперешкодно пропускають трафік тестових DNS-тунелів. Цей факт підтверджує наявність вразливостей і обґрунтовує необхідність глибокого моделювання методів обходу фільтрації, для чого проєктований протокол виступатиме основним інструментом тестування.

Попередній аналіз підтверджує, що традиційні DPI-рішення провайдерів, орієнтовані на сигнатурний пошук, є критично вразливими до методів динамічної обфускації та нестандартного шифрування на рівні DNS. Для надійного захисту інфраструктури та закриття цих "сліпих зон" необхідно змістити фокус із прямої інспекції корисного навантаження на евристичні моделі. Ефективна протидія сучасним прихованим каналам зв'язку потребує впровадження систем поведінкового аналізу на базі машинного навчання [6], здатних своєчасно виявляти аномалії та нетипові патерни DNS-комунікації.

1. Duan R., Liu D. Understanding DNS Tunneling Traffic in the Wild. Unit 42, Palo Alto Networks. 2023. URL: https://unit42.paloaltonetworks.com/dns-tunneling-in-the-wild/
2. Decoy Dog Is No Ordinary Pupy – Infoblox Reveals Shift in Malware Tactics After Initial Discovery. Infoblox. 2023. URL: https://www.infoblox.com/news/news-events/press-releases/decoy-dog-is-no-ordinary-pupy-infoblox-reveals-shift-in-malware-tactics-after-initial-discovery/
3. Mayer D. ChamelGang and ChamelDoH: A DNS-over-HTTPS implant. Stairwell Threat Research. 2023. URL: https://blog.netmanageit.com/content/files/2023/06/Report-ChamelGang-and-ChamelDoH-A-DNS-over-HTTPS-implant.pdf
4. Amirov N., Isik B., Tuncer B. I., Bahtiyar S. DNS Tunneling: Threat Landscape and Improved Detection Solutions. arXiv preprint arXiv:2507.10267. 2025. URL: https://arxiv.org/abs/2507.10267
5. Salat L., Davis M., Khan N. DNS Tunnelling, Exfiltration and Detection over Cloud Environments. Sensors. 2023. Vol. 23, №5. P. 2760. DOI: 10.3390/s23052760
6. Ali F., Afaq M., Niazi M., Behzad M. From Graphs to Gates: DNS-HyXNet, A Lightweight and Deployable Sequential Model for Real-Time DNS Tunnel Detection. arXiv preprint arXiv:2512.09565. 2025. URL: https://arxiv.org/abs/2512.09565

GPU-Adapted Compact Hashing with Bitonic Sort for Neighborhood Search in SPH

UDC 004.94 (043.2)

Ostap Hrytsyshyn[1], Valeriy Trushevskyy[2]

Ivan Franko National University of Lviv, [1]ostap.hrytsyshyn@lnu.edu.ua, [2]valeriy.trushevsky@lnu.edu.ua

Smoothed Particle Hydrodynamics (SPH) is a Lagrangian particle-based method widely used for simulating fluid dynamics. Each particle interacts only with neighbors

located within the kernel support radius h. A naive search over all particle pairs requires $O(n^2)$ operations, which becomes prohibitive for large n, making efficient neighborhood search one of the most important performance bottlenecks of any practical SPH solver [1].

Uniform grids and spatial hashing reduce the cost to approximately O(n), but on GPU architectures additional considerations are required: memory access patterns must be coalesced, dynamic memory allocation should be avoided, and parallel sorting algorithms must be carefully chosen. This paper presents a GPU-adapted compact hashing scheme combined with Bitonic Sort, designed to maintain memory locality and predictable runtime for real-time SPH simulations [2, 3].

Compact Hashing. Following the approach of Ihmsen et al. [2], a uniform grid is constructed over the simulation domain with cell size equal to the kernel support radius h. Each particle is mapped to a grid cell based on its position. The 3D-to-1D hash function is given by:

$$c = (\lfloor x/d \rfloor p_1 \oplus \lfloor y/d \rfloor p_2 \oplus \lfloor z/d \rfloor p_3) \bmod m, \quad (1)$$

where d is the cell size, p_1, p_2, p_3 are large prime numbers, and m is the hash table size. To ensure predictable memory access and avoid runtime overhead from dynamic allocation, the hash table is pre-allocated with size equal to the number of particles. While this does not eliminate hash collisions, it minimizes allocation overhead and enables efficient parallel processing.

Bitonic Sort on GPU. After hash indices are computed, particles must be sorted by their grid index to ensure spatial locality during neighbor search. For GPU execution we adopt Bitonic Sort introduced by Batcher [4], which is highly structured and well-suited to parallel architectures. A sequence is called bitonic if it first increases and then decreases. The algorithm recursively builds bitonic subsequences and merges them into a sorted result using compare-and-swap operations executed in parallel by GPU threads.

The overall complexity is $O(n \log^2 n)$ but each level executes fully in parallel, which gives a wall-clock advantage when n is large. After sorting, particles in the same or adjacent grid cells are stored contiguously in memory, dramatically improving cache utilization during the neighbor query step.

Neighborhood Query. For each particle, only the 27 (or 9 in 2D) cells in the immediate vicinity of its grid cell are evaluated. The cost of the query scales linearly with the number of particles, i.e. O(n). The pseudocode of the integrated procedure is presented in Algorithm 1.

Algorithm 1. GPU-adapted neighborhood search

```
1: for all particles i in parallel do
    2:   compute hash c_i from particle position x_i via Eq. (1)
3: end for
4: BitonicSort(particles by c_i)  // parallel on GPU
5: for all particles i in parallel do
    6:   for all 27 cells in vicinity of c_i do
        7:     collect candidates with ||x_j − x_i|| < h
    8:   end for
9: end for
```

Results. The proposed scheme was implemented using compute shaders and tested on a benchmark with up to 75,000 fluid particles. The GPU-based implementation completes one full simulation step in approximately 10 ms. For comparison, a multi-threaded CPU implementation requires about 250 ms, and a single-threaded CPU implementation requires about 690 ms for an equivalent particle count. The reported speedup factor exceeds 25× compared to the multi-threaded CPU baseline and 65× compared to the single-threaded baseline, confirming that the dominant share of the gain comes from the parallel neighborhood search [5].

Pre-allocation of the hash table with size equal to the number of particles eliminates allocation overhead and stabilizes per-step timings. The locality enforced by sorting reduces irregular memory access in the subsequent density and pressure passes, which is particularly important for IISPH-based incompressible solvers where multiple Jacobi iterations are executed per step.

Conclusions. This paper presented a GPU-adapted neighborhood search procedure for SPH simulations based on compact hashing and Bitonic Sort. The combination of pre-allocated hash tables, structured parallel sorting and cell-based neighbor queries enables linear scaling with the number of particles and real-time performance for ensembles of tens of thousands of particles. Future work includes adaptive hash table sizing and extension to multi-GPU configurations.

1. Monaghan J. J. Smoothed particle hydrodynamics. Reports on Progress in Physics. – 2005. – Vol. 68. – P. 1703–1759.
2. Ihmsen M., Akinci N., Becker M., Teschner M. A parallel SPH implementation on multi-core CPUs. Computer Graphics Forum. – 2011. – Vol. 30, No. 1. – P. 99–112.
3. Ihmsen M., Cornelis J., Solenthaler B., Horvath C., Teschner M. Implicit incompressible SPH. IEEE Transactions on Visualization and Computer Graphics. – 2014. – Vol. 20, No. 3. – P. 426–435.
4. Batcher K. E. Sorting networks and their applications. Proceedings of the AFIPS Spring Joint Computer Conference. – 1968. – Vol. 32. – P. 307–314.
5. Hrytsyshyn O., Trushevskyy V. Application of IISPH for incompressible fluid dynamics simulation. Вісник Львівського університету. Серія прикладна математика та інформатика. – 2024. – Вип. 33. – С. 55–68.

Автоматизація процесів інтеграції та розгортання вебзастосунків

УДК 004.42

Петришин Ярослав[1], Мудрик Іван[2]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]aroslav_petryshyn1608@tntu.edu.ua, [2]i1mudryk@tntu.edu.ua

Сучасні вебзастосунки відрізняються складною багаторівневою архітектурою, яка включає серверну частину, клієнтський застосунок, бази даних та сервіси реального часу. За таких умов ручне тестування й розгортання програмного продукту стає неефективним і ризикованим. Методологія CI/CD (Continuous Integration / Continuous Deployment) забезпечує автоматизацію процесів перевірки, складання та доставки змін до виробничого середовища,

суттєво підвищуючи надійність і швидкість випуску нових версій [1]. Сучасні вебзастосунки еволюціонували до рівня розподілених cloud-native систем, що вимагає переходу від класичних моделей CI/CD до парадигми DevSecOps. У 2026 році автоматизація розгортання вже не обмежується лише доставкою коду, а включає обов'язкові етапи динамічного аналізу безпеки (DAST), сканування образів на вразливості та автоматизоване керування інфраструктурою. Це дозволяє нівелювати ризики, пов'язані з людським фактором, та забезпечити високу доступність сервісів у високонавантажених середовищах.

Типовий процес автоматизації складається з кількох послідовних етапів: статичний аналіз коду (лінтинг), автоматичне тестування (юніт-, інтеграційні та end-to-end тести), складання артефактів та їх доставка на цільовий сервер. Інструменти на кшталт GitHub Actions, GitLab CI або Jenkins дозволяють описати ці етапи декларативно у вигляді YAML-конфігурацій, що запускаються автоматично при кожному коміті або pull request. Це забезпечує швидкий зворотний зв'язок для розробника і унеможливлює потрапляння некоректного коду в основну гілку [1].

Контейнеризація засобами Docker є невід'ємною складовою сучасних процесів автоматизації розгортання, оскільки забезпечує відтворюваність середовища на всіх етапах — від локальної розробки до виробництва. Docker Compose дозволяє декларативно описати мультисервісну інфраструктуру, а збірка образів безпосередньо в процесі доставки усуває залежність від ручного налаштування сервера. У поєднанні з процес-менеджерами це забезпечує автоматичний перезапуск сервісів і мінімізує час простою [2].

Використання Docker та оркестрації залишається фундаментом відтворюваності. Проте акцент змістився на оптимізацію розміру образів та впровадження Infrastructure as Code (IaC) за допомогою Terraform або Pulumi. Це дозволяє декларативно описувати не лише сервіси (через Docker Compose), а й усю хмарну мережеву інфраструктуру, бази даних та системи кешування (наприклад, Redis), мінімізуючи розбіжності між середовищами розробки та продуктиву.

Дослідження ефективності автоматизованих процесів інтеграції та розгортання показують, що команди скорочують час від коміту до розгортання в середньому на 60–80% порівняно з ручними підходами. Крім того, частота виробничих інцидентів, пов'язаних із помилками розгортання, зменшується на 40–50% завдяки обов'язковому проходженню автоматичних перевірок на кожному етапі.

Аналіз впровадження сучасних автоматизованих платформ (базуючись на метриках DORA) показує, що перехід до повної автоматизації дозволяє досягти: збільшення частоти розгортання (Deployment Frequency) у 5–10 разів; скорочення середнього часу відновлення сервісу (MTTR) на 60%; зниження частки невдалих змін (Change Failure Rate) до рівня менше 5% завдяки жорстким автоматизованим фільтрам якості.

Особливої ваги набуває захист ланцюгів постачання програмного забезпечення (Software Supply Chain Security), оскільки за даними останніх досліджень 2025 року [2, 4], вразливості в інфраструктурі збірки є критичним фактором ризику для Cloud-Native систем.

Таким чином, впровадження автоматизованих процесів інтеграції та розгортання є критично важливою практикою для підтримки стабільності й масштабованості сучасних вебплатформ. Описані підходи є актуальними для широкого класу застосунків і можуть бути адаптовані відповідно до потреб конкретного проєкту [2]. Трансформація процесів інтеграції та розгортання у цілісну екосистему DevSecOps є критичною умовою для масштабованості вебплатформ. Актуальність дослідження полягає у переході від простого скриптування до створення інтелектуальних систем доставки, що здатні самостійно адаптуватися до навантажень та безпекових викликів.

1. Мачужак А. В. Дослідження методології DevOps для розробки та підтримки веб-застосунків : кваліфікаційна робота магістра. – Тернопіль : ТНТУ, 2023. – 114 с.
2. Спасітєлєва С. О. Безперервна інтеграція та безперервна доставка (CI/CD) як практика безпечної розробки ПЗ // Кібербезпека: освіта, наука, техніка. – 2023. – № 21. – С. 193–210.
3. Evolution of DevSecOps and Its Influence on Application Security: A Systematic Literature Review // MDPI: Applied Sciences. – 2025. – Vol. 13, No. 12. – P. 548–565.
4. Research Directions in Software Supply Chain Security // ACM Transactions on Software Engineering and Methodology. – 2025. – Vol. 34, No. 5. – P. 112–134.

Інструментальні засоби аналізу впливу характеристик комерційних SPAD-детекторів на стійкість протоколу BB84+decoy-state

УДК 004.056.55:535.14

Олексій Пирогов[1], Василь Різак[2]

Ужгородський національний університет,
[1]oleksii.pyrohov@uzhnu.edu.ua, [2]vrizak@uzhnu.edu.ua

QKD вийшов на промисловий рівень: китайська мережа CN-QCN (China Quantum Communication Network) охоплює понад 10 000 км волокна, 145 магістральних вузлів та 20 метромереж у 80 містах [1]. Транскордонний сегмент EuroQCI розгортається з 2026 р. З 20 квітня 2025 р. набрав чинності Закон № 4336-IX про кіберзахист державних інформаційних ресурсів, але не містить вимог до QKD-систем; галузевого стандарту на QKD в Україні немає. Вибір SPAD-детектора критично впливає на максимальну дальність каналу та параметри секретного ключа, але відкритий інструмент аналізу стійкості відсутній.

BB84 [2] із decoy-state розширенням [3] — найпоширеніший протокол комерційних QKD-систем на волокні. Утім, чи здатна така система генерувати секретний ключ і на якій відстані ще зберігається стійкий режим — визначає не протокол, а характеристики SPAD: η_d (квантова ефективність детектування), вакуумний yield Y_0 (зумовлений темновим рахунком), мертвий час, післяімпульсація, похибка оптичного вирівнювання e_{det} — саме вони задають

параметричні межі стійкості. Проте бракує відкритого інструменту для GLLP-розрахунку (Gottesman-Lo-Lütkenhaus-Preskill) [4] з MC-верифікацією та візуалізацією зони стійкості.

Мета роботи — розробити інструментальні засоби, що поєднують аналітичний розрахунок стійкості BB84+decoy-state за GLLP [4], Monte-Carlo верифікацію з реалістичним моделюванням SPAD та інтерактивну параметричну карту $L_{\_max}(\eta_d, Y_0)$ з пресетами комерційних SPAD-модулів 2004–2025.

Реалізація базується на Python (NumPy, SciPy) з GUI на pywebview та 6 пресетами комерційних SPAD. BB84+decoy-state із трьома інтенсивностями (μ, ν, вакуум); загальне пропускання каналу та приймача $\eta_{total} = 10^{(-\alpha L/10)} \cdot \eta_{Bob} \cdot \eta_d$. Q_x та QBER E_x обчислюються стандартними виразами для BB84+decoy-state [3]. Нижня межа однофотонного yield за двома інтенсивностями [3]:

$$Y_1^L = \frac{\mu}{\mu\nu - \nu^2}\left[Q_\nu e^\nu - Q_\mu e^\mu \left(\frac{\nu}{\mu}\right)^2 - \frac{\mu^2-\nu^2}{\mu^2} Y_0\right]. \quad (1)$$

Звідки $Q_1^L = \mu\, e^{(-\mu)}\, Y_1^L$; $e_1^U = (E_\nu Q_\nu e^\nu - e_0 Y_0)/(\nu\, Y_1^L)$, де $e_0 = 1/2$. Швидкість секретного ключа за GLLP:

$$R \geq q \cdot \{-Q_\mu f_{EC} h(E_\mu) + Q_1^L \cdot [1 - h(e_1^U)]\}, \quad (2)$$

$q = 1/2$, $f_{EC} = 1{,}16$. L_{max} (де $R = 0$) обчислюємо методом Brent на сітці $\eta d \in [0{,}02;\ 0{,}95]$, $Y0 \in [10^{-9};\ 10^{-3}]$ при $\alpha = 0{,}21$ дБ/км, $\eta Bob = 0{,}45$, $e_{det} = 0{,}033$, $\mu = 0{,}5$, $\nu = 0{,}1$. Формули (1)–(2) асимптотичні (finite-key межі — предмет наступних версій). Monte-Carlo (MC) узгоджується з аналітикою ($|\Delta| \leq 6$ % при $L \leq 100$ км, $5 \cdot 10^7$ імпульсів на точку); для $L > 100$ км точна MC-верифікація потребує $> 10^9$ імпульсів на точку через рідкісну decoy-статистику.

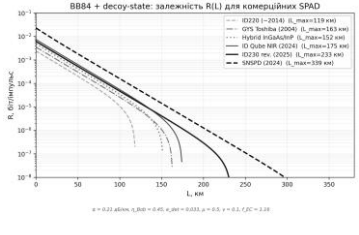


Рис. 1. Межі стійкого режиму QKD-каналу для 6 комерційних SPAD.

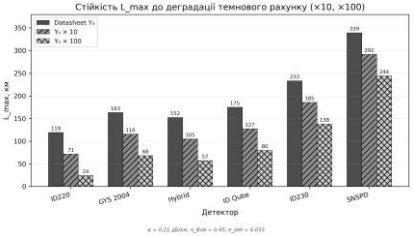


Рис. 2. Стійкість L_{max} до деградації темнового рахунку (×10, ×100).

Параметрична розгортка $L_{max}(\eta_d, Y_0)$ показує, що вакуумний yield обмежує стійкість сильніше, ніж квантова ефективність. Порівняння ID230 та ID Qube ілюструє цей ефект: η_d відрізняється в 1,2×, Y_0 — у 20×, що дає різницю L_{max} у 58 км. Hybrid, попри вищу $\eta d = 20$ %, поступається GYS 2004 (152 проти 163 км) через більший $Y_0 = 2 \cdot 10^{-6}$, що підтверджує визначальну роль темнового рахунку. Аналіз стійкості до деградації Y_0 (Рис. 2) показує: при зростанні

темнового рахунку в 100× ID230 зберігає 138 км (59 % від номіналу), тоді як ID220 — лише 24 км (20 %).

На відміну від NuQKD та OpenQKDSecurity, інструмент орієнтований на інженерну задачу: відкритий код, пресети комерційних SPAD 2004–2025 за публічними datasheet, GLLP+MC та інтерактивна карта $L_{max}(\eta_d, Y_0)$ як критерій вибору детектора.

Інструмент валідовано (MC-аналітика $|\Delta| \leq 6$ % при $L \leq 100$ км). Сучасні SPAD (ID Qube, ID230) забезпечують 175–233 км стійкого QKD-каналу проти 163 км зразка 2004. Результати можуть використовуватись для обґрунтованого вибору SPAD при проектуванні QKD-ланок. Перспектива — розширення інструмента на інші протоколи QKD та finite-key аналіз.

1. Chen H.Z., Li M.H., Wang Y.Z. et al. Implementation of carrier-grade quantum
2. communication networks over 10000 km. npj Quantum Information. 2025. Vol. 11. Art. 137. DOI: 10.1038/s41534-025-01089-8.
3. Bennett C.H., Brassard G. Quantum cryptography: Public key distribution and coin tossing. Proc. IEEE Int. Conf. Comput. Syst. Signal Process., Bangalore, 1984. P. 175–179.
4. Lo H.-K., Ma X., Chen K. Decoy state quantum key distribution. Phys. Rev. Lett. 2005. Vol. 94. P. 230504.
5. Gottesman D., Lo H.-K., Lütkenhaus N., Preskill J. Security of quantum key distribution with imperfect devices. Quant. Inf. Comput. 2004. Vol. 4. P. 325–360. arXiv:quant-ph/0212066.

Архітектура системи верифікації відкритих джерел за допомогою OSINT-технологій

УДК 004.056

Олена Пирч[1], Катерина Федоренко[2]

Хмельницький національний університет,
[1]pyrchov@khmnu.edu.ua, [2]katefedorenko8080@gmail.com

В українських реаліях дезінформація має виражений гібридний характер і поєднує психологічні, інформаційні та кіберкомпоненти. Російська збройна агресія проти України супроводжується масштабними інформаційно-психологічними операціями, спрямованими як на внутрішню, так і на зовнішню аудиторію. Основними цілями таких кампаній є підрив довіри до державних інститутів, посилення почуття нестабільності та беззахисності, деморалізація населення, дискредитація Збройних сил України та міжнародних партнерів. Для їх реалізації активно використовуються анонімні Telegram-канали, фейкові акаунти у соціальних мережах, медіаресурси та мережі ботів [1].

Архітектура системи верифікації відкритих джерел побудована як багаторівнева модульна система, у якій кожен рівень відповідає за окремий аспект перевірки допису у відкритому Telegram-каналі на наявність дезінформації. Така структурна організація системи забезпечує не лише логічну послідовність процесу, а й методологічну узгодженість, оскільки кожен модуль

опирається на чітко визначену теоретичну основу та виконує свою функцію в межах загального циклу аналітичної верифікації.

Концепція архітектури системи верифікації відкритих джерел відповідає адаптованій моделі розвідувального циклу, у межах якої виділено чотири ключові етапи: оцінка джерела, контекстна перевірка, семантичний аналіз та інтеграційна верифікація. Кожен з етапів виконує автономну аналітичну функцію, але водночас утворює логічну послідовність, де результати попереднього рівня слугують вхідними даними для наступного. Такий підхід забезпечує системність, масштабованість та можливість багаторазової повторної перевірки отриманих результатів [2].

Архітектура системи верифікації відкритих джерел на наявність дезінформації є комплексною моделлю аналітичної верифікації, у якій поєднано класичні теоретичні засади комунікаційної достовірності, контекстної кореляції та когнітивного аналізу. Кожен модуль виконує окрему функцію, але водночас є частиною єдиного аналітичного потоку, що перетворює сирі дані з відкритих джерел на доказову аналітичну інформацію.

Модуль первинної OSINT-ідентифікації Telegram-поста, у загальній архітектурі системи виконує функцію вхідної ланки для подальшого аналітичного опрацювання. На даному етапі «сирий» контент із Telegram перетворюється на структурований набір даних, придатний для подальшої контекстної, семантичної та інтегрованої перевірки. Модуль забезпечує послідовне отримання метаданих, завантаження медіафайлів, витягнення EXIF-інформації, виконання зворотного пошуку зображення та формування базової часової триангуляції появи інформації у різних джерелах. Завдяки цьому подальші модулі працюють не з неструктурованим повідомленням, а з узгодженим набором полів, які можна обробляти автоматизованими методами.

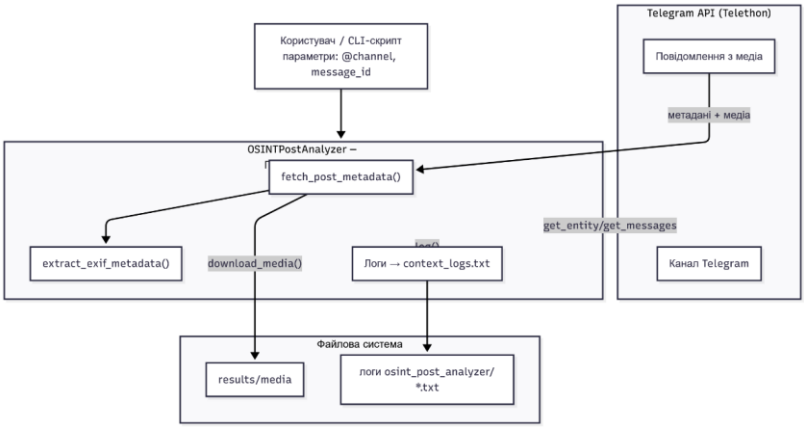


Рис. 1. Отримання Telegram-поста та первинна обробка

Модуль працює в асинхронному режимі з використанням async await. Це дає можливість паралельно звертатися до Telegram та зовнішніх сервісів і не

блокувати виконання програми. Узагальнену схему взаємодії класу з Telegram API, сервісами SerpAPI і imgbb та файловою системою подано на рисунку 1 та наведено основні потоки даних і місця збереження результатів.

На основі проведеного дослідження розроблено архітектуру системи верифікації відкритих джерел, яка охоплює декілька взаємопов'язаних модулів. Усі ці елементи об'єднано у єдину систему інтегрованої оцінки, що забезпечує можливість об'єктивного визначення ступеня достовірності досліджуваного матеріалу. Важливою складовою системи є застосування Source Credibility Matrix, яка дозволяє кількісно характеризувати надійність каналу на основі його історії, активності, стабільності публікацій та виявлених ознак маніпулятивності. Дані елементи створюють теоретично обґрунтований інструмент для оцінювання джерел інформації, що особливо важливо в умовах інформаційних загроз, спрямованих на українське суспільство.

1. What is OSINT: Open-source intelligence? European data. URL: https://data.europa.eu/en/publications/datastories/what-osint-open-source-intelligence.
2. Як ефективно використовувати OSINT-інструментарій? Команда «OsintFlow» ділиться досвідом // Державна прикордонна служба України. URL: https://dpsu.gov.ua/uk/news/43174-YAk-efektivno-vikoristovuvati-OSINT-instrumentariy-Komanda-OsintFlow-dilitsya-dosvidom.

Сучасні підходи до трансформації систем охорони праці на основі штучного інтелекту та предикативної аналітики

УДК 004.056.5:57.087.1
Михайло Пригара[1], В'ячеслав Шматуха[2], Володимир Щербина[3]

[1]Ужгородський національний університет, misha_prigara@ukr.net,
[2]Київська школа економіки, vvshmatukha@gmail.com,
[3]Державний університет «Київський авіаційний інститут», smya@kai.edu.ua

Традиційні системи управління охороною праці (СУОП) тривалий час базувалися на реактивному підході — аналізі інцидентів, що вже відбулися. Однак у 2025–2026 роках девіз Всесвітнього дня охорони праці «Революційні підходи до здоров'я і безпеки: роль ШІ та цифровізації» підкреслив глобальний перехід до проактивних стратегій. Використання ШІ дозволяє не лише фіксувати порушення, а й передбачати їх, створюючи динамічне безпечне середовище.

1. Ключові напрями використання ШІ в ОП

Сучасні підходи можна класифікувати за технологічними доменами:

- Комп'ютерний зір (Computer Vision) для моніторингу в реальному часі
- Це найбільш поширений напрям у 2026 році. Камери з ШІ інтегруються в існуючі системи відеоспостереження для:
- Автоматичного контролю використання засобів індивідуального

захисту (ЗІЗ): касок, жилетів, масок, рукавичок.

- Детекції небезпечних зон: сповіщення працівника та оператора при вході людини в зону роботи кранів або навантажувачів.
- Аналізу ергономіки: відстеження рухів працівника для запобігання скелетно-м'язовим розладам через неправильні пози чи перевантаження.

Б. Предиктивна аналітика та прогнозування ризиків

Алгоритми машинного навчання обробляють великі масиви даних (Big Data), включаючи звіти про дрібні інциденти (near-misses), погодні умови, стан обладнання та навіть психофізіологічні показники працівників. ШІ виявляє патерни, що передують аваріям, дозволяючи менеджменту втрутитися до моменту виникнення нещасного випадку.

В. Інтеграція з носимими пристроями (Wearables) та IoT

Розумні годинники та сенсори на одязі моніторять життєво важливі показники (пульс, температура тіла, рівень втоми). Це критично для робіт у замкнених просторах, при екстремальних температурах або у нічні зміни. При виявленні критичного рівня втоми система рекомендує зробити перерву.

2. Трансформація навчання та інструктажів

ШІ змінює підхід до підготовки персоналу:

- Адаптивне навчання: Генеративний ШІ створює персоналізовані курси на основі помилок, які працівник допускав раніше.
- VR/AR-симуляції: Тренування навичок безпечної роботи у віртуальному середовищі з ШІ-інструктором, який моделює критичні ситуації в режимі реального часу.

3. Переваги та економічна ефективність

За даними міжнародних звітів 2025-2026 рр., підприємства, що впровадили ШІ-платформи безпеки, демонструють:

- Зниження травматизму на 25–40%.
- Скорочення витрат на страхові виплати та компенсації.
- Прискорення аудитів: підготовка до перевірок з охорони праці стає швидшою на 40% завдяки автоматизованому збору даних.

4. Виклики та етичні аспекти

Незважаючи на технологічний прогрес, залишаються відкритими питання:

- Конфіденційність даних: необхідність балансу між моніторингом безпеки та приватністю працівника.
- Психологічний тиск: ризик виникнення стресу у персоналу через відчуття «постійного нагляду» алгоритмом.
- Технологічна залежність: ризик втрати навичок самостійної оцінки небезпеки людиною.

Висновок. Станом на 2026 рік штучний інтелект перестав бути футуристичним концептом і став фундаментальним інструментом у сфері охорони праці. Перехід від «контролю після факту» до «запобігання до події» є ключовим вектором розвитку. Майбутнє СУОП полягає у синергії людського досвіду та обчислювальної потужності ШІ, де технології виступають не як заміна інженеру з ОП, а як його високотехнологічний асистент.

1. Healthy Workplaces Summit 2025: discover key takeaways, photos and resources on safe digital work / EU-OSHA. Bilbao, 2025. URL: https://osha.europa.eu/en/highlights/healthy-workplaces-summit-2025-discover-key-takeaways-photos-and-resources-safe-digital-work (дата звернення: 06.05.2026).
2. Mishiba, Takenori. 2024. “Transforming Occupational Health and Safety Regulation: Strategic Pathways in the Era of Industry 4.0.” Journal of Occupational Health Law and Emerging Vision 3, no. 2: 151–169. https://doi. org/10.57523/jaohlev.pp.24-016 (дата звернення: 06.05.2026).
3. World Day for Safety and Health at Work 2025: Revolutionizing Health and Safety: The Role of AI and Digitalization at Work / International Labour Organization (ILO). 2025. URL: https://www.ilo.org/safeday (дата звернення: 06.05.2026).

Mitigating AI-driven security risks in educational software systems

UDK 621.395.7 (043.2)

Stepan Prokipchyn[1]

State University of Information and Communication Technologies,
[1]s.prokipchyn@stud.duikt.edu.ua

The growing use of autonomous AI agents in educational software introduces new cybersecurity challenges due to their ability to interact with external systems and act on behalf of users. The objective of this work is to analyze access control-related risks in such systems and propose practical mitigation strategies. The relevance of the study is driven by the increasing integration of AI into critical educational processes. The scientific novelty is the structured analysis of these risks across different layers of AI usage within educational systems.

The vulnerability surface of AI systems is often defined as prompt injection, data poisoning and hallucination [1, 2]. However, any LLM is prone to these kinds of risks. What makes agentic systems especially vulnerable to these attacks is the main strength of the ReAct pattern – the ability to interact with external systems (load data, perform actions).

In the context of educational systems, the work covers three classes of AI agent use. Internal automation agents – purpose-built agents that automates operational scenarios, running scripted flows (e.g., automated syllabus review per instructional design, suggesting additional materials for students based on their results, AI-assisted learning, etc.). AI-assisted coding and engineering – in 2026 this has become an industry standard, with a significant portion of code generated by AI agents. AI-generated code in system-critical domains like security can lead to vulnerabilities. AI features in the product – production-side features powered by AI agents that students and educators interact with directly. Exposing AI to actual users without proper guardrails not only allows the system to be tricked or abused by dishonest individuals but can also lead to unexpected costs and overall system instability. All three layers are prone to access control-related security risks which are not new but are rather elevated by AI [3].

Credential and token leakage. AI agents can read files, project configuration, MCP definitions, and environment variables. Tokens are especially dangerous: they can be used from anywhere, are often broader-scoped than intended, and persist long after they're useful. Before AI, tokens could be leaked due to security misconfiguration or application bugs. Nowadays, an agent running within the security perimeter can read the token and publish it due to hallucination or prompt injection [2]. To mitigate this risk, long-lived access tokens must be avoided in favor of temporary credentials, SSH keys and other authentication mechanisms limited in lifetime and scope of use.

Over-privileged access. Even short-lived, machine-scoped credentials have some access assigned. Often "minimal" permissions can be broader than intended. Hallucinations can cause the agent to deviate from the intended task to unexpected destructive or corruptive actions (e.g., removing a student user instead of sending an assignment) [2]. LLMs are stochastic in principle. Usually this is a strength, but in some rare cases this can lead to unwanted behavior. Mitigation always requires following the principle of the least privilege. Mitigation includes stricter permission scoping for agent service accounts, the use of fine-grained tokens, and tracking of AI identities. A comprehensive audit layer adds observability to the security plane.

Uncontrolled AI actions. Agents may execute unintended actions, especially when used in auto-approve mode. Bugs in agent tools or in their underlying MCPs can lead to unauthorized data changes. For agents to be useful, they must be provided a certain level of write access. We can forbid destructive actions via fine-grained permission configuration, but most tasks will still require data modification, email sending, API requests, etc. Even within a hardened security perimeter, a hallucinating agent can do harm. There is no general solution to this risk. However, secure-by-design implementation of tools and MCPs can significantly reduce it. In addition to that, a secondary observer AI agent can be running in the background and validating the primary agent intent. Such observers are less prone to hallucination because their context is often limited to an action that the primary agent is trying to execute and a brief explanation of intent. If the observer is unsure or the potential risk of the action is higher than a certain threshold – a human-in-the-loop pattern can be triggered, requiring explicit operator approval.

The work shows that key security risks: credential leakage, over-privileged access, and uncontrolled actions are amplified by agent autonomy. Mitigation requires least-privilege access, short-lived credentials, audit mechanisms, and human-in-the-loop controls. These results provide practical guidance for securing AI-enabled educational systems and form a basis for further research on controllable agentic architectures.

1. OWASP Gen AI Security Project. OWASP Top 10 for LLM Applications 2025. URL: https://genai.owasp.org/resource/owasp-top-10-for-llm-applications-2025 (date of access: 08.05.2026)
2. Tang Y., et al. Security of LLM-based agents regarding attacks and defenses. Journal of Network and Computer Applications, 2025. DOI: https://doi.org/10.1016/j.inffus.2025.103941.
3. Gao Y., Wu S. A Four-Layer Security Governance Framework for LLM-Based AI Agents. Journal of Artificial Intelligence Practice, 2025. DOI: http://dx.doi.org/10.23977/jaip.2025.080406.

Управління інформаційною безпекою в умовах впровадження великих мовних моделей у CRM-системи

УДК 004.056

Ігор Ралік[1]

Тернопільський національний технічний університет, [1]tntuihor@gmail.com

Впровадження великих мовних моделей у CRM-системи створює нові виклики для управління інформаційною безпекою на корпоративному рівні [1]. Перехід від статичних комунікаційних шаблонів до динамічної генерації тексту вимагає розробки рішень для контролю за поведінкою алгоритмів штучного інтелекту.

Генерація персоналізованих комунікаційних сценаріїв у реальному часі несе ризики порушення корпоративних політик, розголошення чутливої інформації або надання клієнтам юридично некоректних даних. Відповідно, ефективне управління інформаційною безпекою вимагає механізмів, які б регулювали генеративний процес, забезпечуючи його контрольованість та запобігаючи появі небажаних відповідей.

У запропонованій архітектурі CRM-системи управління безпекою реалізується превентивно через механізм динамічного структурованого промпту, за інформаційну безпеку відповідають два блоки, а саме:

- Блок системних обмежень. Виконує функцію системного метарегулятора, який визначає універсальні рамки безпеки. Управління ризиками тут відбувається через суворі інструкції, такі як заборону на використання чутливої інформації, уникнення припущень щодо особистих даних клієнта та обмеження на інтерпретацію юридичної інформації. Цей рівень управління спрямований на мінімізацію ризиків, безпосередньо пов'язаних з автоматизованою генерацією тексту.
- Блок бізнес-правил. Забезпечує управління дотриманням вимог на рівні конкретної організації. Цей компонент містить політики підтримки, внутрішні процедури, регламенти та юридичні обмеження компанії. Завантажуючись одноразово як системний промпт, він гарантує повну відповідність згенерованих сценаріїв бізнес-політиці, забезпечуючи функціональну та операційну надійність взаємодії.

Управління інформаційною безпекою у CRM-системах має базуватися на проактивному вбудовуванні політик безпеки безпосередньо в логіку генерації відповідей. Поєднання блоку універсальних обмежень та суворого дотримання доменних бізнес-правил дозволяє зберігати контроль над системою, відкриваючи перспективи для подальшої розробки механізмів контролю генеративних сценаріїв.

1. HOCHI C. The Impact of Large Language Model Integration on Customer Relationship Management in Small and Medium-Sized Enterprises: An Empirical Study of a Medium-Sized Printing Company. Journal of International Social Science. 2025. p. 177.

Оцінювання допустимості альтернатив реагування на кіберінциденти в органах військового управління

УДК 004.056.5

Геннадій Рибачок[1]

[1]*Національний університет оборони України, ryba4okgen@gmail.com*

Кіберінциденти у секторі безпеки і оборони дедалі частіше мають не лише технічний, а й управлінський зміст [1]. Для органів військового управління Збройних Сил України важливо встановити не тільки факт порушення функціонування інформаційної системи, а й допустимі варіанти реагування з урахуванням часу, ресурсів, повноважень, режимно-безпекових обмежень і залишкового ризику. За таких умов технічно можлива дія не завжди є управлінсько прийнятною, тому в системі підтримки прийняття рішень потрібна окрема процедура оцінювання допустимості альтернатив реагування. Сучасні стандарти й рекомендації з управління кіберінцидентами визначають загальну логіку підготовки, виявлення, аналізу, реагування, відновлення та післяінцидентного удосконалення [2, 3, 4].

Водночас вони не розв'язують повною мірою завдання вибору конкретної альтернативи дій у військово-управлінському контексті. Після оцінювання інциденту суб'єкт управління має отримати не загальну вказівку на потребу реагування, а множину дій, з якої вилучено варіанти, неприйнятні за наявних умов. Метою тез є обґрунтування методичного підходу до оцінювання допустимості альтернатив реагування на кіберінциденти в системі підтримки прийняття рішень органів військового управління Збройних Сил України. Наукова новизна підходу полягає в розмежуванні технічно можливих і управлінсько допустимих альтернатив реагування шляхом попередньої перевірки за повноважними, часовими, ресурсними, режимно-безпековими, ризиковими та управлінськими обмеженнями.

Альтернатива реагування розглядається як можливий варіант дій щодо нейтралізації кіберінциденту, сформований з урахуванням результату оцінювання інциденту, початкового управлінського стану, вимог до ефективності реагування, допустимого залишкового ризику та профілю умов реагування. Базова множина альтернатив не повинна одразу ототожнюватися з рекомендованим рішенням. Вона є початковим простором можливих дій, який потребує перевірки на допустимість і здійсненність. У загальному вигляді перевірку доцільно подати як формування множини допустимих альтернатив:

$$A_i feas = \{a_{ij} \in A_i^0 \mid Feas_i(a_{ij}, \Pi_i, E_i req, Risk_i adm) = 1\},$$

де a_{ij} — j-та альтернатива реагування для і-го кіберінциденту; A_i^0 — базова множина альтернатив; Π_i — профіль умов реагування; вимоги до ефективності реагування визначають очікуваний результат нейтралізації; допустимий залишковий ризик задає межу прийнятності наслідків; предикат Feas визначає допустимість і здійсненність альтернативи. Якщо значення предиката дорівнює 1, альтернатива може бути передана на критеріальне оцінювання.

До основних груп умов допустимості належать повноважна й процедурна допустимість, часова здійсненність, ресурсна забезпеченість, режимно-безпекова сумісність, ризикова прийнятність та управлінська придатність.

Повноважна умова не допускає варіантів, що виходять за межі компетенції суб'єкта рішення або порушують порядок ескалації. Часова умова відсікає дії, які не забезпечують досягнення потрібного рівня нейтралізації в наявних часових межах. Ресурсна умова перевіряє наявність сил, засобів, фахівців і доступу до необхідних даних. Режимно-безпекова умова враховує обмеження щодо інформації, режимів роботи систем і недопущення вторинних ризиків. Ризикова умова пов'язує альтернативу з допустимим залишковим ризиком.

Управлінська придатність показує, чи не погіршує дія керованість, інформаційний обмін або виконання завдань більше, ніж сам інцидент. Для об'єктів критичної інфраструктури та систем, що забезпечують військове управління, такий підхід дає можливість відокремити технічно можливі дії від управлінсько прийнятних [1, 4]. Наприклад, локалізація інциденту може бути технічно швидкою, але неприйнятною, якщо вона зупиняє критичний процес, потребує неузгодженого втручання або створює надмірний залишковий ризик. Саме тому оцінювання допустимості має передувати критеріальному вибору рекомендованої альтернативи. Якщо після перевірки множина допустимих альтернатив є порожньою, система підтримки прийняття рішень не повинна імітувати наявність готового рішення. У такому випадку результатом має бути висновок про відсутність допустимих альтернатив у наявному профілі умов реагування з пропозицією уточнення даних, перегляду ресурсних меж, формування комбінованих варіантів або ескалації рішення. Отже, оцінювання допустимості альтернатив є необхідною проміжною ланкою між формуванням базового простору дій і вибором рекомендованого рішення щодо нейтралізації кіберінциденту.

Його використання підвищує відтворюваність підготовки управлінського рішення, забезпечує пояснюваність відбору альтернатив і зберігає принципову межу: інформаційна технологія формує обґрунтовану рекомендацію, а остаточне рішення залишається за уповноваженим суб'єктом військового управління.

1. Про основні засади забезпечення кібербезпеки України : Закон України від 05.10.2017 № 2163-VIII. URL: https://zakon.rada.gov.ua/go/2163-19 (дата звернення: 15.05.2026).
2. ISO/IEC 27035-1:2023. Information technology — Information security incident management — Part 1: Principles and process. URL: https://www.iso.org/standard/78973.html (дата звернення: 15.05.2026).
3. Nelson A., Rekhi S., Souppaya M., Scarfone K. Incident Response Recommendations and Considerations for Cybersecurity Risk Management. NIST SP 800-61 Rev. 3. 2025. URL: https://doi.org/10.6028/NIST.SP.800-61r3 (дата звернення: 15.05.2026).
4. Pascoe C., Quinn S., Scarfone K. The NIST Cybersecurity Framework (CSF) 2.0. NIST, 2024. URL: https://doi.org/10.6028/NIST.CSWP.29 (дата звернення: 15.05.2026).

Оцінювання методів захисту агентних систем на основі великих мовних моделей

УДК 004.8:004.056.5

Роман Шклярський[1], Даниїл Журавчак[2]

Національний університет "Львівська політехніка",
[1]roman.shkliarskyi.asp.2025@lpnu.ua, [2]danyil.y.zhuravchak@lpnu.ua

Процес розвитку великих мовних моделей пройшов три етапи: від простих інтерфейсів завершення тексту через базові виклики API до автономних агентів з постійною пам'яттю, доступом до файлової системи та можливістю перегляду вебсторінок [1]. OWASP та NIST визначають prompt injection як основну загрозу для таких систем [2].

Архітектурна причина вразливості полягає в тому, що трансформер не розрізняє керуючі інструкції та дані користувача. Коли LLM інтегрується у виробничий конвеєр, стохастична модель поєднується з детермінованим середовищем виконання з підвищеними привілеями без механізмів перевірки походження команд [1]. Аудити 2025 року виявили вразливості в комерційних агентах для написання коду, де непрямий prompt injection призвів до виконання зловмисного коду та витоку облікових даних [3].

Атаки типу prompt injection поділяються на прямі та непрямі. Прямі передбачають формування запиту для обходу налаштувань безпеки моделі. Непрямі діють через вміст, який агент обробляє в ході роботи: вебсторінки, електронні листи або записи баз даних [2]. Успішність ручних непрямих атак становить 60–80%, атак на основі градієнтної оптимізації досягає 90–95% [4].

Системи отримання даних (RAG) вразливі до окремого класу атак. Дослідження PoisonedRAG показало, що введення п'яти шкідливих документів у базу з мільйонів записів дозволяє маніпулювати відповідями моделі з успішністю до 97% на наборі Natural Questions [5]. Показники успішності атаки за наборами даних наведено на рис. 1.

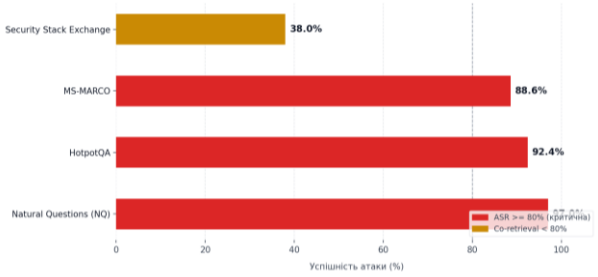


Рис.1. Успішність атаки отруєння бази знань RAG залежно від набору даних (PoisonedRAG, n = 5 документів)

Атака вимагає, щоб шкідливий текст одночасно потрапляв до топ-k результатів пошуку і містив інструкції для формування потрібної відповіді [5]. Гібридне отримання даних, що поєднує BM25 і векторний пошук, знизило

успішність атак на наборі Security Stack Exchange з 38% до 0% [6]. Архітектура RAGShield застосовує криптографічну перевірку документів перед індексуванням відповідно до NIST SP 800-53 [7].

Витік системних промптів є окремим вектором атак у багатоорендних середовищах. PROMPTPEEK показав, що спільне кешування ключ-значення дозволяє відновити вміст системного промпту іншого орендаря через аналіз часових характеристик запитів [8]. Фреймворк PLeak демонструє, що часткова реконструкція конфіденційних інструкцій можлива і без прямого доступу до інфраструктури [9]. Система StruQ розділяє вхідні дані на привілейований канал інструкцій і ненадійний канал даних, знижуючи успішність атак на 90–95% [4]. Підхід XOA забороняє моделі безпосередньо спостерігати ненадійні дані: LLM генерує сценарій, який виконується в ізольованому середовищі, а модель отримує лише кінцевий результат [10].

Фільтрація за ключовими словами та вирівнювання через RLHF не вирішують проблему на архітектурному рівні. Розмежування каналів інструкцій і даних, верифікація походження документів у RAG-конвеєрах та ізоляція середовища виконання є підходами з підтвердженою ефективністю в умовах контрольованого тестування. Стандартизація методів оцінювання захисту залишається відкритою проблемою, оскільки наявні дослідження використовують несумісні набори даних і метрики.

1. Sehwag V. et al. Clawed and Dangerous: Can We Trust Open Agentic Systems? arXiv. 2025.
2. Prompt Injection Attacks on Agentic Coding Assistants: A Systematic Analysis of Vulnerabilities in Skills, Tools, and Protocol Ecosystems. arXiv. 2025.
3. Prompt Injection Attacks in Large Language Models and AI Agent Systems. MDPI Information. 2025. Vol. 17, No. 1.
4. Chen S. et al. StruQ: Defending Against Prompt Injection with Structured Queries. USENIX Security Symposium. 2025.
5. Zou W. et al. PoisonedRAG: Knowledge Corruption Attacks to Retrieval-Augmented Generation of Large Language Models. USENIX Security Symposium. 2025.
6. Semantic Chameleon: Corpus-Dependent Poisoning Attacks and Defenses in RAG Systems. arXiv. 2025.
7. RAGShield: Provenance-Verified Defense-in-Depth Against Knowledge Base Poisoning in Government RAG Systems. arXiv. 2025.
8. PROMPTPEEK: Prompt Leakage via KV-Cache Sharing in Multi-Tenant LLM Serving. NDSS Symposium. 2025.
9. Cao Y. PLeak: Prompt Leaking Attacks against Large Language Model Applications. ACM CCS. 2024.
10. Williams D. Execute-Only Agents: Architectural Defense Against Prompt Injection for AI Agents. URL: https://people.cs.vt.edu/djwillia/papers/agenticos26-xoa.pdf (дата звернення: 05.05.2026).

Формалізація атак підміни інструкцій у великих мовних моделях та методи їх виявлення

УДК 004.8:004.056.5

Роман Шклярський[1], Даниїл Журавчак[2]

Національний університет "Львівська політехніка",
[1]roman.shkliarskyi.asp.2025@lpnu.ua, [2]danyil.y.zhuravchak@lpnu.ua

Атаки типу prompt injection потрапили до переліку OWASP Top 10 для LLM-застосунків у 2025 році як основна загроза [1]. Причина полягає в архітектурній особливості трансформера: модель не розрізняє керуючі інструкції оператора та дані користувача, що дозволяє зловмиснику підмінити поведінку системи через текстовий вхід. Традиційні метрики оцінювання вразливостей, зокрема CVSS, не враховують стохастичну природу LLM, тому виникає потреба у формалізованих підходах до класифікації та вимірювання таких атак [2].

Перший систематичний підхід до формалізації запропонували Liu et al. у дослідженні, представленому на USENIX Security 2024 [3]. Атака описується як задача оптимізації: зловмисник прагне максимізувати ймовірність цільової відповіді R при заданому запиті Q та впровадженій інструкції. Для вимірювання ефективності атак введено метрику Attack Success Variation (ASV), яка дозволяє кількісно порівнювати п'ять типів атак із десяти різних механізмів захисту на десяти провідних моделях.

Паралельно з формалізацією дослідники розробляли таксономії для класифікації атак. Робота SoK, опублікована у 2026 році, вводить тривимірну таксономію за векторами доставки (прямі та непрямі), модальностями атаки (текстові та обфусковані) і поведінкою поширення (persistent та transient) [4]. Аналіз понад 30 CVE у комерційних агентах, зокрема Claude Code та GitHub Copilot, показав, що непрямі атаки через протокол MCP призводять до виконання довільного коду та витоку облікових даних. Окремо виділено клас атак "Confused Deputy", де агент використовує власний авторизований доступ до інструментів для виконання дій на користь зловмисника [4].

Дослідження OpenClaw розширює таксономію до трьох рівнів: ланцюжок постачання, активація та виконання [5]. Серед семи категорій загроз найкритичнішою є вихід із ізольованого середовища, для якого базовий рівень захисту становить лише 17% [5]. Атаки через кодування (Base64, шістнадцяткове представлення) обходять статичні фільтри і класифіковані як середня загроза за MITRE ATLAS AML.TA0008 [5].

Серед архітектурних методів захисту виділяються два підходи з підтвердженою ефективністю. Система StruQ розділяє вхідні дані на привілейований канал інструкцій і ненадійний канал даних за допомогою зарезервованих роздільників та спеціалізованого налаштування моделі [6]. Підхід SecAlign формулює захист як задачу оптимізації переваг через Direct Preference Optimization: модель навчається на трійках {вхід, бажана відповідь, небажана відповідь}, що дозволяє розрізняти оригінальні інструкції від зовнішніх даних [6].

Таблиця 1

Порівняння методів захисту від атак підміни інструкцій

Метод захисту	Технічний підхід	Середній рівень захисту (ASV)	Основне обмеження
StruQ	Розмежування каналів за роздільниками	95.0% -- 98.0%	Потребує переналаштування моделі
SecAlign	Оптимізація переваг (DPO)	~100.0%	Вразливий до адаптивних алгоритмів
InstruCoT	Міркування за ланцюжком думок (CoT)	92.5%	Збільшена затримка виведення
HITL	Перехоплення людиною в контурі	19% → 92.0%	Вузьке місце для автономних задач

Для відомих векторів атак ASR наближається до 0%, хоча метод залишається вразливим до адаптивних алгоритмів. Порівняння методів захисту наведено у табл. 1. Метрика ASV, запропонована Liu et al., дозволяє стандартизувати порівняння захисних рішень, проте не враховує ефект підсилення ризику через автономність агента [3]. Наведені дослідження показують, що формалізація атак типу prompt injection перейшла від окремих описів до кількісних фреймворків з уніфікованими метриками. Таксономії SoK та OpenClaw охоплюють як текстові, так і агентні вектори атак, а AIVSS забезпечує інструмент для оцінювання ризику в контексті конкретного розгортання. Архітектурні підходи StruQ і SecAlign демонструють вищу ефективність, проте обидва вимагають переналаштування моделі, що обмежує їх застосування з комерційними закритими системами. Стандартизація бенчмарків оцінювання залишається відкритою проблемою.

1. LLM01:2025 Prompt Injection. OWASP Gen AI Security Project. 2025. URL: https://genai.owasp.org/llmrisk/llm01-prompt-injection/ (дата звернення: 05.05.2026).
2. AIVSS Scoring System For OWASP Agentic AI Core Security Risks v0.8. OWASP Agentic AI Security Working Group. 2025.
3. Liu Y. et al. Formalizing and Benchmarking Prompt Injection Attacks and Defenses. USENIX Security Symposium. 2024.
4. Prompt Injection Attacks on Agentic Coding Assistants: A Systematic Analysis of Vulnerabilities in Skills, Tools, and Protocol Ecosystems. arXiv. 2026.
5. Towards Secure Agent Skills: Architecture, Threat Taxonomy, and Security Analysis. IEEE S&P. 2026.
6. Chen S. et al. StruQ: Defending Against Prompt Injection with Structured Queries. USENIX Security Symposium. 2025.

Least Significant Bit steganography in SVG XML architecture

UDC 004.056.5:004.92

Nataliya Zagorodna[1], Oleh Yarema[2]

Ternopil Ivan Puluj National Technical University,
[1]zagorodna_n@tntu.edu.ua, [2]yarema.oleh.m@gmail.com

Traditional graphical steganography relies on raster images to embed covert data by modifying pixel color values. However, as web systems shift toward scalable, responsive media, raster-based methods face security scrutiny and structural limitations. Steganography is, first of all, the science of hiding a secret message within a non-secret file to avoid triggering suspicion. The Least Significant Bit (LSB) method is the foundational algorithm of this domain. It operates on the principle that digital media files possess data bits that contribute minimally to human perception [1].

In a standard uncompressed image, each pixel is typically represented by 24 bits of data – 8 bits each for the red, green and blue color channels (RGB). The leftmost bits (most significant bits) describe the primary color information, while the rightmost bits (least significant bits) represent very minimal variations in shade [1]. LSB steganography exploits this by replacing these LSBs with the binary stream of an encrypted secret file.

Mainstream implementations of LSB are localized to raster graphics, audio files, and video containers. These mainstream applications suffer from bottlenecks that limit their utility in modern security bases. Because raster pixels follow natural statistical distributions, sequential LSB injection alters the global pixel histograms. Modern steganalysis tools can easily detect these changes [2]. Raster LSB is highly fragile. Everyday web processes, such as converting a PNG to a lossy JPEG, scaling an image down, or applying compression, can destroy the LSB array, corrupting the hidden data. Modern firewalls and automated Deep Packet Inspection (DPI) systems target traditional media attachments on the first priority basis and subject them to algorithmic checks for embedded hidden data.

To bypass some of the limitations of raster media, steganography can use Scalable Vector Graphics (SVG). SVGs are text files written in structured XML code. They do not contain a grid of pixels, instead they contain mathematical instructions how to render an image by the web browser [3].

We propose a framework where LSB principles are mapped directly onto the XML DOM structure. An SVG defines colors textually via hexadecimal strings or standard CSS RGB strings [3]. By targeting the lowest bits of these color palettes, inline fills, and stroke attributes, a secret payload can be distributed across the structural elements of a webpage.

Future scientific investigations should focus on the following unexplored dimensions. Research is needed to develop parsers that dynamically map the XML tree of an SVG, isolate color attributes, and handle the LSB flipping within string data types rather than raw binary matrices. SVGs also rely on precise floating-point coordinate points. A massive avenue for research lies in coordinate LSB manipulation, where the thousandths decimal place of a geometric shape is altered to hold data. Because vector points are highly precise, shifting an object by 0.0001 millimeters is visually non-existent but offers massive data capacity.

Future studies must benchmark SVG LSB against standard defensive tools. Because SVGs are processed as code by firewalls rather than images, research can empirically prove whether vector steganography can bypass mainstream pixel-based steganalysis engines.

LSB steganography must evolve alongside modern web standards. By embedding data into the XML architecture of SVG files, we can open up a lightweight and novel vector for secure data transmission.

1. Aditya, S., Ved, M., Shashikant, K., Samadhan, K., & Shilpa, M. A. (2024). Image steganography using least significant bit. International Journal of Research Publication and Reviews, 5(4), 2505–2508. https://doi.org/10.55248/gengpi.5.0424.0966
2. Madoš, B., Hurtuk, J., Čopjak, M., Hamaš, P., & Ennert, M. (2014). Steganographic algorithm for information hiding using scalable vector graphics images. Acta Electrotechnica et Informatica, 14(4), 42–45. https://doi.org/10.15546/aeei-2014-0040
3. Xu Z. Xu D. Li Z. Zheng X. & Zhang C. (2026). GVIS: Generative vector image steganography. In proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 9384–9393). https://openaccess.thecvf.com/content/CVPR2026/papers/Xu_GVIS_Generative_Vector_Image_Steganography_CVPR_2026_paper.pdf

Метод виявлення ботнет-активності в корпоративній мережі на основі багатокритеріальної оптимізації XGBoost

УДК 004.056:004.85

Владислав Самойленко[1], Сергій Гахов[2]

Державний університет інформаційно-комунікаційних технологій,
[1]v.samoylenko@duikt.edu.ua, [2]gakhovsa@gmail.com

Ботнет-активність залишається однією з найнебезпечніших загроз для корпоративних мереж, оскільки використовується для DDoS-атак, розсилання спаму, крадіжки даних та прихованого віддаленого керування зараженими вузлами. У дослідженні авторів [1] на наборі даних CSE-CIC-IDS2018 було порівняно Random Forest, XGBoost та SVM для виявлення ботнетів. Найкращим базовим рішенням стала XGBoost, яка досягла середнього значення F1-міри 0,99 за крос-валідацією при часу навчання близько 16 с. Це обґрунтувало перехід від простого вибору моделі до її оптимізації з урахуванням умов практичного мережевого моніторингу.

Метою роботи є розроблення методу виявлення ботнет-активності, який одночасно забезпечує високу якість класифікації, контроль рівня хибних спрацювань та достатню пропускну здатність для потокового моніторингу. Для уникнення витоку інформації дані Friday-02-03-2018 було поділено на раннє вікно S1 для навчання й оптимізації та пізніше вікно S2 для фінального тестування, а benign-підмножину Thursday-15-02-2018 використано як незалежний негативний день для калібрування порога рішення за обмеженням FPR ≤ 1 %. Основою експерименту слугував набір даних CSE-CIC-IDS2018 [2].

Наукова новизна дослідження полягає у формуванні відтворюваного протоколу вибору конфігурації XGBoost [3], у якому багатокритеріальна оптимізація поєднується з operationally-oriented перевіркою моделі. Оптимізація виконувалася за двома цілями: середньою Average Precision на часово-коректній forward-валідації та наскрізною пропускною здатністю інференсу. Пошук конфігурацій здійснювався з використанням Optuna [4]. На відміну від підходів, де оцінюється лише якість класифікації або лише час predict, у роботі враховано повний ланцюг інференсу, включаючи побудову DMatrix, а робочий поріг обирається без використання майбутнього тестового вікна S2.

За результатами багатокритеріального пошуку було відібрано три репрезентативні конфігурації Pareto-фронту: Light, Balanced і Heavy. Основні результати порівняння наведено в таблиці 1. Модель Balanced з 40 деревами забезпечила F1 = 0,9963 на майбутньому вікні S2 при FPR = 0,0098 на незалежній benign-підмножині Thursday та медіанній наскрізній пропускній здатності 1 583 532 потоків/с. Для порівняння, конфігурація Heavy з 90 деревами дала лише незначно вищу F1 = 0,9965, але зменшила пропускну здатність до 559 256 потоків/с.

Важливо, що використання часово-коректної схеми оцінювання «навчання на ранньому вікні — тестування на пізнішому» наближує експеримент до реальних умов функціонування засобів мережевого моніторингу. Такий підхід зменшує ризик отримання надто оптимістичних оцінок якості, які можуть виникати під час випадкового перемішування потоків. Додаткове калібрування порога на незалежному benign-дні підвищує надійність практичного використання моделі та дає змогу заздалегідь обмежити інтенсивність хибних спрацювань.

Таблиця 1

Порівняння репрезентативних конфігурацій XGBoost

Конфігурація	Дерев	F1 на S2	Пропускна здатність, потоків/с
Light	20	0,9845	1 901 519
Balanced	40	0,9963	1 583 532
Heavy	90	0,9965	559 256

Отримані результати показують, що для практичних систем виявлення ботнетів максимізація лише метрик якості є недостатньою. Раціональнішим є вибір конфігурації, що забезпечує баланс між точністю, контрольованим обсягом хибних спрацювань та швидкодією обробки трафіку. У цьому дослідженні такою конфігурацією є Balanced: вона зберігає майже ту саму якість, що й складніша Heavy-модель, але має приблизно у 2,8 раза вищу наскрізну пропускну здатність.

Перспективи подальших досліджень полягають у розширенні методу за рахунок локальної інтерпретації ознак, часової агрегації рішень для послідовностей потоків та перевірки стійкості моделі на нових фрагментах мережевого трафіку. Це дасть змогу підвищити пояснюваність результатів для аналітика безпеки та адаптивність системи до змін у поведінці ботнетів.

Практичне значення запропонованого методу полягає у можливості використання XGBoost у системах реального часу для виявлення ботнет-активності в корпоративних мережах без перевантаження аналітиків надмірною кількістю хибних тривог. У підсумку, запропонований підхід може бути використаний як основа для побудови масштабованих засобів мережевого моніторингу, у яких одночасно враховуються точність виявлення, контроль FPR та продуктивність обробки трафіку.

1. Samoilenko V., Gakhov S. Comparative Analysis of Machine Learning Methods for Detecting Botnet Activities in Corporate Networks. SSRN Electronic Journal. 2025. DOI: 10.2139/ssrn.5188775.
2. Sharafaldin I., Habibi Lashkari A., Ghorbani A. A. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. ICISSP 2018. P. 108–116. DOI: 10.5220/0006639801080116.
3. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System. Proc. of the 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining. 2016. P. 785–794. DOI: 10.1145/2939672.2939785.
4. Akiba T., Sano S., Yanase T., Ohta T., Koyama M. Optuna: A Next-generation Hyperparameter Optimization Framework. Proc. of the 25th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining. 2019. P. 2623–2631. DOI: 10.1145/3292500.3330701.

Ключові контролі стандартів інформаційної безпеки для захисту критичної інфраструктури

УДК 004.056.5

Сведенюк Олексій [1], Курій Євгеній [2]

Національний університет "Львівська політехніка",
oleksii.svedeniuk.asp.2025@lpnu.ua [1], yevhenii.o.kurii@lpnu.ua [2]

Актуальність та постановка проблеми. Об'єкти критичної інфраструктури (ОКІ) України перебувають під постійним тиском цілеспрямованих кібератак (APT). Конвергенція IT- та OT-технологій створює нові вектори загроз, де компрометація одного сегмента загрожує катастрофічними наслідками для національної безпеки. Складність традиційних підходів до захисту заважає їх швидкому впровадженню, що зумовлює необхідність виокремлення пріоритетних заходів контролю [1].

Мета роботи — визначити та обґрунтувати набір критичних контролів на основі міжнародних стандартів (NIST CSF, CIS Controls v8, ISO/IEC 27001), адаптованих для захисту активів вітчизняної критичної інфраструктури.

Наукова новизна. У роботі запропоновано адаптивну модель ієрархізації засобів захисту, яка враховує специфіку функціонування АСУ ТП (SCADA) та забезпечує безперервність критичних бізнес-процесів навіть в умовах обмежених ресурсів.

Виклад основного матеріалу. Аналіз фреймворків дозволяє виділити «фундаментальні контролі», що захищають від 85% поширених атак. Для ОКІ критичними є: 1) інвентаризація апаратних і програмних активів; 2) управління

вразливостями; 3) контроль привілейованих облікових записів; 4) сегментація IT/OT мереж; 5) аналіз журналів подій безпеки. Підхід «CIS Implementation Group 1» дозволяє ОКІ сформувати базову лінію захисту [2]. Особлива увага має приділятися механізмам MFA для віддаленого доступу до технологічних сегментів, що запобігає несанкціонованому втручанню в системи керування.

Висновки. Впровадження пріоритетних контролів безпеки стратегічно важливе для стійкості ОКІ, оскільки оптимізує витрати та прискорює реагування на інциденти. Подальші дослідження будуть спрямовані на автоматизацію перевірки відповідності цим контролям у реальному часі.

1. Сіденко В. П., Гнатюк С. О. Метод підвищення рівня захищеності критичних інформаційних систем держави. Кібербезпека: освіта, наука, техніка. – 2024. – Т. 4, № 24. – С. 138–154.
2. CIS Critical Security Controls Version 8. Center for Internet Security, 2021. 82 p.

Дослідження методів побудови постквантових крипто-кодових конструкцій на гіпереліптичних кодах

УДК 621.395.7 (043.2)

Сергій Євсеєв[1], Владислав Сокол[2]

Національний технічний університет «Харківський політехнічний інститут», [1]Serhii.Yevseiev@gmail.com, [2]Vladyslav.sokol@gmail.com

Фундаментальний зсув парадигми інформаційної безпеки у напрямку постквантової криптографії стимулює інтенсивний пошук нових алгебраїчних структур. Дослідження фундаментальних аспектів застосування гіпереліптичних кривих у сучасних системах криптографічного захисту інформації, наведені у [1], показують багатообіцяючі результати. Доведено, що такі багатовимірні алгебраїчні структури здатні забезпечити надзвичайно високий рівень безпеки при використанні значно коротших ключів порівняно з традиційними еліптичними аналогами завдяки більшій розмірності групи класів дивізорів нульового степеня. Але залишилися невирішеними питання, пов'язані з алгоритмічним забезпеченням швидкого та детермінованого підрахунку кількості раціональних точок для кривих довільного роду над довільними полями Галуа. Причиною цього явища виступають суттєві об’єктивні математичні труднощі, безпосередньо пов’язані з неминучим експоненційним зростанням обчислювальної складності в процесі безпосереднього розрахунку порядку Якобіана для складних типів кривих. Специфіка імплементації суворих методів багатофакторної автентифікації на основі модифікованих крипто-кодових систем у фінансовому секторі формалізована у [2]. Встановлено, що глибоко інтегровані механізми захисту, які органічно поєднують завадостійкість та криптографічну конфіденційність, є критично необхідними для безпечного проведення віддалених банківських транзакцій в умовах гібридних кіберзагроз. Але залишилися невирішеними питання, пов'язані з оптимізацією пропускної здатності комунікаційного каналу під час постійної передачі автентифікаційних токенів надто великого розміру. Варіантом

подолання відповідних труднощів може бути повна відмова від традиційних плоских структур та перехід до використання кодів з екстремально високою алгебраїчною щільністю та багатовимірною просторовою геометрією. Саме такий підхід логічно випливає з результатів аналізу, проте практичні методики побудови перевірочних матриць безпосередньо з Якобіанів гіпереліптичних кривих у науковій літературі досі не формалізовані. Все це дозволяє стверджувати, що доцільним є проведення дослідження, присвяченого побудові та аналізу алгебро-геометричних кодів на базі гіпереліптичних кривих над полем Галуа. Фундаментальною основою синтезованих крипто-кодових конструкцій виступає математичний апарат алгебраїчної геометрії [3]. Скінченне поле Галуа $GF(2^m)$ строго задається за допомогою незвідного полінома f(x) та канонічного базису елементів 1, x, x^2, …, x^{m-1}. Для виконання процедур серіалізації бітових векторів у системі застосовується цілочисельне кодування вигляду

$$v = \sum_{i=0}^{m-1} v_i 2^i. \quad (1)$$

Гіпереліптична крива C роду g над полем $GF(2^m)$ задається рівнянням у загальній формі

$$C: y^2 + h(x)y = f(x), \quad (2)$$

де $h(x) \in GF(2^m)[x]$ є поліномом зі степенем $\deg(h) \le g$, а $f(x) \in GF(2^m)[x]$ виступає нормованим поліномом степеня $\deg(f) = 2g+1$ або $\deg(f) = 2g+2$. Критичною вимогою для криптографічного застосування є відсутність сингулярних точок, що означає неможливість існування розв'язків $(x, y) \in \overline{GF\left(2^m\right)} \times \overline{GF\left(2^m\right)}$, які б одночасно задовольняли базовому рівнянню C та системі його часткових похідних [4].

У процесі синтезу архітектури сформовано математичну модель алгебро-геометричного коду на базі гіпереліптичної кривої. У явному вигляді модель детерміновано кортежем параметрів $M = (GF(2^m), C, P_{rat}, E_L, H, G)$. Елемент C задає рівняння кривої (2), а P_{rat} – множину проєктивних точок (4). Матриці E_L, H та G визначають межі кодового простору. Запропонована формалізація адаптує топологічні властивості многовидів до формату дискретних структур даних [1]. Застосована у дослідженні трансформація координат повністю та остаточно розв'язала цю алгоритмічну проблему. Відображення кожної скінченної афінної точки у систему проєктивних координат із введенням додаткової просторової змінної дозволило звести рівняння кривої до гомогенної (однорідної) форми. Здійснено імплементацію згенерованих гіпереліптичних кодів у дві фундаментальні теоретико-кодові архітектури: асиметричну схему Мак-Еліса (McEliece) та симетричну схему Рао-Нама (Rao-Nam). Встановлено, що класична асиметрична парадигма з жорстким табличним декодером вимагає виконання $O(k^2)$ операцій на розв'язання лінійних систем відносно відкритого тексту та провокує експоненційне зростання обсягів оперативної пам'яті. Натомість інтеграція алгебро-геометричного коду в симетричну канальну архітектуру з використанням криптографічного генератора псевдовипадкових чисел для формування маски штучної помилки дозволила повністю обійти етап обернення щільних матриць.

1. Alimoradi, R. (2016). A Study of Hyperelliptic Curves in Cryptography. IJCNIS, 8(8), 67–72. https://doi.org/10.5815/ijcnis.2016.08.08
2. Yevseiev, S., Kots, H., Liekariev, Y. (2016). Developing of multi-factor authentication method based on niederreiter-mceliece modified crypto-code system. Eastern-European Journal of Enterprise Technologies, 6(4(84)), 11–23. https://doi.org/10.15587/1729-4061.2016.86175.
3. Hubrechts, H. (2011). MEMORY EFFICIENT HYPERELLIPTIC CURVE POINT COUNTING. Int. J. Number Theory, 07(01), 203–214. https://doi.org/10.1142/S1793042111004034.
4. Conceição, R. (2020). ON INTEGRAL POINTS ON ISOTRIVIAL ELLIPTIC CURVES OVER FUNCTION FIELDS. Bull. Aust. Math. Soc., 102(2), 177–185. https://doi.org/10.1017/S0004972720000155

Дослідження методів та засобів ідентифікації дезінформативних новин у соціальних мережах

УДК 004.8:004.7

Тарас Труш[1], Марія Стадник[2]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]tarastrush.dev@gmail.com, [2]stadnyk_m@tntu.edu.ua

У період стрімкої цифровізації суспільства соціальні мережі стали ключовим джерелом поширення інформації, але водночас є інструментом для навмисного поширення дезінформації. Досвід України, яка з 2014 року опинилася серед перших держав, що зіткнулися з масштабними дезінформаційними кампаніями в соціальних мережах, демонструє, що ця проблема переросла у фактор загрози національній безпеці. Фейкові профілі, ботоферми, маніпулятивні матеріали та синтетичний медіаконтент активно використовуються як інструменти гібридної війни. Додатковим викликом є швидкий розвиток генеративного штучного інтелекту, який значно ускладнює відокремлення правдивої інформації від неправдивої. У зв'язку з цим зростає потреба у створенні дієвих автоматизованих систем виявлення дезінформації.

Метою дослідження є аналіз сучасних методів автоматизованої детекції дезінформації в соціальних мережах та оцінка ефективності гібридних і мультимодальних підходів для виявлення маніпулятивного контенту. Поширення дезінформації в соціальних мережах та розвиток генеративного штучного інтелекту створюють серйозні загрози інформаційній безпеці. Сучасні методи виявлення фейкової інформації потребують поєднання текстового, візуального та мережевого аналізу, однак україномовний інформаційний простір залишається недостатньо дослідженим.

На основі аналізу публікацій з баз даних Scopus, Web of Science, Google Scholar та arXiv за 2017–2025 роки встановлено, що методи детекції дезінформації поділяються на три основні категорії. Контент-орієнтовані підходи базуються на аналізі текстових та візуальних ознак новин із використанням NLP-технік і трансформерних моделей (BERT, RoBERTa, DeBERTa) [1]. Мультимодальні методи поєднують обробку тексту та зображень

через механізми early fusion, late fusion і cross-modal attention, що дозволяє виявляти deepfake-контент і складні семантичні невідповідності між текстом і зображенням [2]. Користувач-орієнтовані підходи базуються на поведінкових характеристиках акаунтів: зокрема, ECS (Ensemble of Specialized Classifiers) застосовує правило максимуму для виявлення нових типів ботів, невідомих під час навчання [3]. Мережеві підходи аналізують структуру поширення інформації в соціальних мережах через графові нейронні мережі (GCN, GAT, BiGCN), моделюючи часові та структурні закономірності поширення [4]. Гібридні підходи, що поєднують контентний і мережевий аналіз (dEFEND, BERT+LightGBM), демонструють стабільно вищу ефективність порівняно з унімодальними системами [5,6].

Окремо досліджено україномовний контекст: датасет EUvsDisinfo фіксує значне зростання проросійської дезінформації напередодні повномасштабного вторгнення у 2022 році [7], а Shared Task UNLP 2025 представив перший публічний бенчмарк для виявлення маніпуляцій у Telegram-дописах українською мовою (9 557 записів, 10 технік маніпуляції) [8].

За результатами дослідження встановлено, що гібридні підходи стабільно перевершують унімодальні: dEFEND досягає 90.4% точності на PolitiFact [5], BERT+LightGBM - 82.9% на LIAR [6]. Водночас жодна з розглянутих систем не є універсальною через проблему зміщення домену: моделі з точністю ~99% на WELFake демонструють падіння на 30% і більше при тестуванні на нових доменах. Україномовний інформаційний простір залишається критично недослідженим - існуючі бенчмарки (EUvsDisinfo, UNLP 2025) є лише першими кроками. Перспективами подальших досліджень є розробка крос-лінгвальних моделей для україномовного середовища, мультимодальних систем детекції deepfake-контенту, а також методів виявлення дезінформації, згенерованої великими мовними моделями.

1. Sayyadiharikandeh M., Varol O., Yang K.-C. et al. Detection of novel social bots by ensembles of specialized classifiers // Proceedings of the 29th ACM International Conference on Information and Knowledge Management. New York : Association for Computing Machinery, 2020. P. 2725–2732.
2. Lv J., Gao Y., Li L., Shi L., Li S. Multi-modal fake news detection: a comprehensive survey on deep learning technology, advances, and challenges // Journal of King Saud University – Computer and Information Sciences. 2025. Vol. 37. P. 306.
3. Sayyadiharikandeh M., Varol O., Yang K.-C. et al. Detection of novel social bots by ensembles of specialized classifiers // Proceedings of the 29th ACM International Conference on Information and Knowledge Management. New York : Association for Computing Machinery, 2020. P. 2725–2732.
4. Sayyadiharikandeh M., Varol O., Yang K.-C. et al. Detection of novel social bots by ensembles of specialized classifiers // Proceedings of the 29th ACM International Conference on Information and Knowledge Management. New York : Association for Computing Machinery, 2020. P. 2725–2732.
5. Sayyadiharikandeh M., Varol O., Yang K.-C. et al. Detection of novel social bots by ensembles of specialized classifiers // Proceedings of the 29th ACM

International Conference on Information and Knowledge Management. New York : Association for Computing Machinery, 2020. P. 2725–2732.
6. Essa E., Omar K., Alqahtani A. Fake news detection based on a hybrid BERT and LightGBM models // Complex & Intelligent Systems. 2023. Vol. 9. P. 6581–6592.
7. Leite, J. A., Razuvayevskaya, O., Bontcheva, K., Scarton, C. EUvsDisinfo: A dataset for multilingual detection of pro-Kremlin disinformation in news articles. Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. New York: Association for Computing Machinery, 2024, pp. 5380–5384.
8. Kyslyi, R., Romanyshyn, N., Sydorskyi, V. The UNLP 2025 Shared Task on Detecting Social Media Manipulation. Proceedings of the Fourth Ukrainian Natural Language Processing Workshop, 2025, pp. 105–111. URL: https://aclanthology.org/2025.unlp-1.12.pdf.

Середовище для аналізу атак на SDN-орієнтовані системи

УДК 004.056:004.7

Юрій Кльоц[1], Сергій Мостовий[2]

Хмельницький національний університет,
[1]klots@khmnu.edu.ua, [2] serhii.mostovyi@khmnu.edu.ua

Сучасні мережеві інфраструктури дедалі частіше використовують централізовані засоби керування, моніторингу та конфігурування обладнання, що наближає їх до принципів програмно-конфігурованих мереж. Одним із прикладів такої системи є Omada Controller, який забезпечує централізоване керування маршрутизаторами, комутаторами, точками доступу та бездротовими клієнтами. Концентрація керуючих функцій в одному компоненті підвищує зручність адміністрування, однак одночасно формує критичну точку впливу для потенційних атак.

Атаки на SDN-системи можуть бути спрямовані на порушення доступності контролера, зміну або блокування службової взаємодії між контролером і мережевими пристроями, імітацію легітимної активності, перевантаження каналів керування або виявлення вразливих сервісів. Особливу складність становить те, що частина атакувального трафіку за окремими ознаками може бути подібною до нормальних службових процесів, зокрема реєстрації пристроїв, оновлення станів, обміну телеметрією або підключення клієнтів. Тому дослідження таких атак потребує не лише фіксації факту аномальної активності, а й аналізу варіативності нормального трафіку в умовах реальної або наближеної до реальної мережевої інфраструктури [2].

Використання готових наборів пакетів для розроблення методів виявлення та протидії атакам не завжди забезпечує достатню достовірність результатів. Такі набори часто не враховують особливостей конкретного контролера, моделі мережевого обладнання, структури службового обміну, кількості клієнтів, типів підключення та характеру адміністративних дій. Крім того, у відкритих датасетах зазвичай складно встановити точний контекст формування трафіку, умови проведення атаки та відповідність отриманих пакетів реальним

сценаріям експлуатації. Це обмежує можливість коректної оцінки ефективності алгоритмів виявлення аномалій і може призводити до надмірної кількості хибних спрацювань або пропуску специфічних для конкретної системи атак.

На рис. 1 представлено структурну схему експериментальної мережі для дослідження нормального та аномального трафіку, пов'язаного з роботою Omada Controller. Центральним елементом схеми є маршрутизатор TP-Link, до якого підключено вхідний інтернет-канал, машину з установленим Omada Controller, звичайний комутатор, PoE-комутатор, а також окремі машини для збору пакетів. Така побудова дає змогу відтворювати типові умови функціонування мережі з централізованим керуванням та одночасно фіксувати трафік у ключових точках інфраструктури.

У нижній частині схеми показано клієнтську частину мережі. До звичайного комутатора підключено два дротові ПК, а до PoE-комутатора – дві точки доступу, через які працюють декілька Wi-Fi-клієнтів. Один із внутрішніх вузлів позначено як досліджуваний пристрій, трафік якого може окремо дзеркалюватися на машину збору пакетів. Це дозволяє аналізувати поведінку конкретного пристрою в умовах штатної роботи або під час моделювання атакувальної активності.

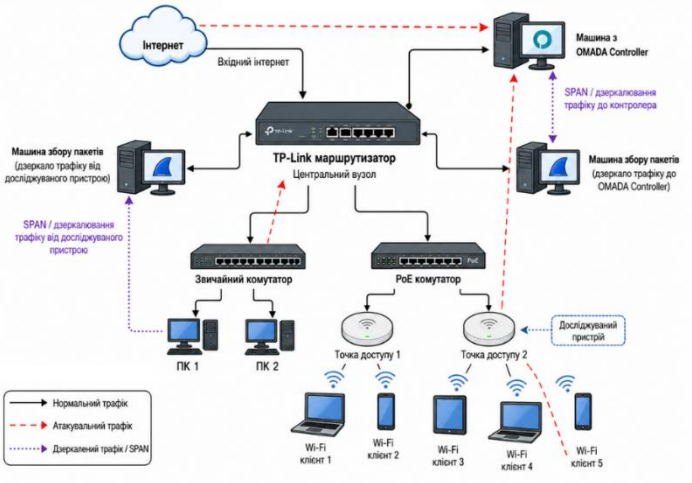


Рис.1. Схема дослідження трафіку та збору пакетів

Окремо на рисунку виділено машини збору пакетів. Одна з них призначена для фіксації дзеркального трафіку, що надходить до Omada Controller, інша – для збору трафіку від досліджуваного пристрою. Для позначення різних типів потоків використано умовні лінії: суцільні стрілки відповідають нормальному мережевому трафіку, пунктирні червоні стрілки – атакувальному трафіку, а фіолетові пунктирні лінії – дзеркальованому трафіку/SPAN. Така схема відображає можливість одночасного дослідження штатної взаємодії між компонентами мережі та виявлення змін у трафіку під час атак на контролер або окремі мережеві пристрої.

Отже, дослідження атак на SDN-орієнтовані системи керування мережею є актуальним, оскільки централізація функцій адміністрування, моніторингу та керування трафіком підвищує критичність контролера як об'єкта захисту. На прикладі Omada Controller показано, що використання лише готових наборів пакетів не завжди забезпечує достовірність результатів, адже вони можуть не враховувати особливості конкретної топології, службової взаємодії пристроїв і реальних умов експлуатації. Запропонована експериментальна схема дає змогу формувати власний набір даних, фіксувати нормальний та атакувальний трафік у ключових точках мережі й створює основу для подальшої розробки методів виявлення аномалій та протидії атакам на системи централізованого керування мережею.

1. Mansoor, A., Anbar, M., Bahashwan, A. A., Alabsi, B. A., & Rihan, S. D. A. (2023). Deep Learning-Based Approach for Detecting DDoS Attack on Software-Defined Networking Controller. Systems, 11(6), 296. https://doi.org/10.3390/systems11060296

Дослідження вразливостей протоколів динамічної маршрутизації

УДК 004.056.

Сергій Мостовий[1], Сергій Савченко[2]

Хмельницький національний університет,
[1]serhii.mostovyi@khmnu.edu.ua, [2]sserggii@yahoo.com

Корпоративні мережі постійно піддаються впливу різноманітних загроз, які можуть призвести до порушення конфіденційності, цілісності або доступності мережевих сервісів. У контексті динамічної маршрутизації ці загрози мають критичне значення, оскільки впливають на механізми обміну маршрутною інформацією, що є основою функціонування будь-якої IP-мережі. Завдяки використанню динамічних протоколів маршрутизації, таких як RIP, OSPF, IS-IS, EIGRP, та BGP [1,3], забезпечується автоматичне оновлення таблиць маршрутизації в реальному часі, що дозволяє адаптувати мережу до змін. Однак, ці самі властивості роблять протоколи маршрутизації вразливими до різноманітних атак, що можуть порушити їхню функціональність і безпеку.

Загрози у корпоративних мережах можна умовно класифікувати на кілька груп, кожна з яких стосується певних аспектів функціонування мережі.

Перша група загроз стосується маршрутизаторів та мережевого обладнання. Ці загрози включають атаки, метою яких є виведення маршрутизатора з ладу або отримання доступу до його конфігурації [1].

Друга група загроз - загрози цілісності маршрутної інформації. Зловмисники можуть змінювати або підмінювати маршрутні оголошення, що суттєво впливає на мережу. У межах динамічної маршрутизації подібні дії можуть спричинити перехоплення трафіку, його перенаправлення або повне блокування комунікацій [1].

Третя група загроз - загрози доступності. Атаки типу DoS (Denial of Service) та DDoS (Distributed Denial of Service) можуть спричинити перевантаження маршрутних процесів, що призводить до зниження продуктивності або відмови

протоколів маршрутизації. Для протоколів, чутливих до частоти та своєчасності оновлень, таких як OSPF [1-3], це може мати критичні наслідки.

Четверта група загроз пов'язані з автентифікацією. Відсутність автентифікації дозволяє атакуючому змінювати таблиці маршрутизації, не будучи виявленим, що може призвести до серйозних порушень у роботі мережі [1].

Остання група загроз стосується загрози від внутрішніх порушників. Внутрішні порушники можуть мати доступ до конфіденційної маршрутної інформації, яку вони можуть змінювати або використовувати для несанкціонованого доступу до корпоративних сервісів.

Більшість наведених загроз використовує вразливості протоколів динамічної маршрутизації.

Аналіз вразливостей протоколів динамічної маршрутизації показує, що кожен з протоколів має свої специфічні слабкі місця, які можуть бути використані зловмисниками для атак на мережу. У таблиці 1 представлено порівняння основних вразливостей протоколів маршрутизації RIP, OSPF, EIGRP і BGP, а також типи атак, до яких ці протоколи схильні.

Таблиця 1

Порівняльний аналіз вразливостей протоколів динамічної маршрутизації

Протокол	Вразливість	Типи атак
RIP	Відсутність автентифікації, повільна конвергенція	Spoofing,route injection,DoS
OSPF	Вразливість до впровадження фальшивих LSA-пакетів	Spoofing, replay-атаки, DoS
EIGRP	Відсутність криптографії, маніпуляція оновленнями	Spoofing, route injection, replay-атаки
BGP	Вразливості в AS-PATH, BGP hijacking	BGP hijacking, Route Leak, DoS

Захист мережі від цих загроз потребує застосування різноманітних методів, серед яких автентифікація, шифрування, виявлення аномалій і застосування політик маршрутизації, що дозволяють мінімізувати ризики та забезпечити безпеку даних, що передаються між маршрутизаторами [3].

Відсутність належного захисту маршрутних оголошень може призвести до катастрофічних наслідків, включаючи втрату конфіденційності трафіку, його перенаправлення через вузли зловмисника, блокування сегментів мережі або порушення доступності критичних сервісів. Тому виникає потреба у створенні універсального методу захисту, здатного підвищити стійкість динамічної маршрутизації до атак, не знижуючи продуктивність і не змінюючи логіку роботи протоколів.

1. Manzoor A., Hussain M., Mehrban S. Performance analysis and route optimization: redistribution between EIGRP, OSPF & BGP routing protocols. Computer Standards & Interfaces, 2020, 68: 103391.
2. Cisco. Dynamic Routing Protocols: OSPF, EIGRP, RIPv2, IS-IS, BGP. URL: https://community.cisco.com/t5/networking-knowledge-

base/dynamic-routing-protocols-ospf-eigrp-ripv2-is-is-bgp/ta-p/4511577 (дата звернення: 20.04.2026).
3. Кульчинський І. Аналіз роботи протоколів динамічної маршрутизації. Збірник тез V Всеукраїнської студентської науково-технічної конференції „Природничі та гуманітарні науки. Актуальні питання “, 2012, 1: 67-67.

Analysis of Authentication-Based Attacks in Wireless Networks

UDC 004.7.056

Danylo Matiuk[1], Maryna Derkach[2],
Inna Skarga-Bandurova[3]

Ternopil Ivan Puluj National Technical University,
[1]matiuk.danylo@icloud.com, [2]m_derkach@tntu.edu.ua,
Oxford Brookes University,
[3]iskarga-bandurova@brookes.ac.uk

Modern wireless networks play a crucial role in data exchange and are widely deployed across commercial, governmental, educational, and residential environments. However, they remain among the most vulnerable components of both corporate and household security infrastructures. As their adoption increases, so does the number of potential targets for cyberattacks. Most attacks on wireless networks target authentication mechanisms, as these mechanisms are designed to prevent unauthorized access to access points.

Several widely used protocols provide secure authentication for wireless networks, including WPA, WPA2, and WPA3. These protocols differ in the data required for successful authentication, as well as in their encryption and validation methods. Each protocol supports two main authentication modes: WPA-Personal and WPA-Enterprise. The WPA-Personal mode relies on a pre-shared key (PSK), where all users and devices must know the same password to access the network. In contrast, WPA-Enterprise uses a RADIUS server for centralized authentication, where user credentials are verified individually.

Although WPA is considered obsolete for modern devices, it is still used in practice, particularly in low-power or legacy systems that rely on older algorithms. Similarly, WPA2, introduced in 2006 as the standard for secure Wi-Fi networks, remains widely used today. This is supported by the results of a wireless network scan (Fig. 1) performed using a portable NetScope device as part of ethical hacking activities [1].

SSID	RSSI	Channel	Security	Vendor	BSSID
[redacted]	-59 dBm	11	WPA*	Shenzhen	[redacted]
	-65 dBm	10	WPA2		
	-71 dBm	2	WPA*	TendaTec	
	-80 dBm	1	WPA2	Tp-LinkT	
	-81 dBm	1	WPA*		
	-82 dBm	10	WPA*	TendaTec	
	-83 dBm	6	WPA2		
	-88 dBm	8	WPA*		

Fig.1. Results of a wireless network scan

The compromise of wireless networks is rarely an isolated issue and often serves as an entry point for deeper intrusions into IT environments. By analysing information

broadcast over the air about wireless networks and associated client devices – including hidden SSID/BSSID, channels, signal strength, security mechanisms, and equipment manufacturers – an attacker can assess the level of radio-frequency security within a given environment and determine an appropriate attack strategy. In WPA2-Personal networks, authentication and key establishment are performed using a four-way handshake. Attacks targeting the four-way handshake aim to exploit vulnerabilities in the process of establishing a secure connection between a client device and a wireless access point. During the attack, adversaries attempt to exploit the four-way handshake used for authentication and key establishment by transmitting deauthentication frames, which belong to the IEEE 802.11 control frame subclass. These frames operate at the lower levels of the Wi-Fi protocol stack and are used to terminate connections. The presence of a deauthentication reason code field enables analysis of how control messages are processed by the receiving side and provides a formal description of connection termination events. As the network becomes destabilized, the attacker may deploy a rogue access point that mimics the target SSID. If the PSK is known, an Evil Twin attack can be performed, forcing the victim's traffic to pass through the attacker-controlled node and enabling Man-in-the-Middle scenarios, such as credential interception and content manipulation [2]. The Evil Twin attack is also relevant to WPA2-Enterprise networks. Although WPA2-Enterprise is generally considered more secure than WPA2-Personal, it remains vulnerable to online brute-force attacks. Since WPA2-Enterprise credentials often correspond to domain user accounts, compromised credentials may enable unauthorized access to additional systems within a corporate network.

In January 2018, the Wi-Fi Alliance introduced WPA3 as a successor to WPA2. WPA3 enhances security through the implementation of the Simultaneous Authentication of Equals (SAE) protocol and the Dragonfly key exchange mechanism, which provide stronger protection against password-based attacks. However, in 2019, several vulnerabilities were identified, including downgrade attacks exploiting backward compatibility, where a client can be forced to connect to a rogue WPA2 access point, allowing attackers to capture the handshake.

Based on the analysis of attacks on wireless security protocols, several conclusions can be drawn. WPA3 replaces the vulnerable PSK–based authentication approach, which allows attackers to capture Wi-Fi traffic and perform offline brute-force attacks, with a more secure SAE-based key exchange mechanism. Nevertheless, given the widespread use of WPA2, the following considerations remain important:

1) The Evil Twin attack on WPA2-Enterprise is ineffective against clients that use certificate-based authentication methods (e.g., EAP-TLS or PEAP with EAP-TLS), as no reusable credentials are exposed and server certificate validation is enforced during the initial authentication phase.
2) Protection against dictionary attacks relies on the use of strong and unique passwords. Additionally, the Pairwise Transient Key PTK, which is generated per session, ensures secure data transmission by encrypting communication between the client device and the access point.

1. Matiuk, D. S., & Derkach, M. V. (2025). NetScope: pentestinh bezdrotovykh merezh [NetScope: wireless network pentesting].

Proceedings of the XIV International Scientific and Technical Conference of Young Scientists and Students “Current Issues of Modern Technologies”, 11-12 December 2025, Ternopil, 302-304. PE Palianytsia V.A. [in Ukrainian].

2. Banakh, R., Nyemkova, E., Justice, C., Piskozub, A., & Lakh, Y. Data Mining Approach for Evil Twin Attack Identification in Wi-Fi Networks. Data, 2024, vol. 9(10), p. 119.

Налаштування безпечної мережевої інфраструктури для балансування навантаження та відмовостійкості

УДК 004.7:004.056

Вікторія Ваврічен[1], Тарас Лобур[2]

Тернопільський національний технічний університет імені Івана Пулюя, [1]viktoriavavrichen@gmail.com, [2]lobur_t@tntu.edu.ua

У сучасних комп’ютерних мережах важливим завданням є забезпечення стабільного й безпечного доступу до Інтернету та безперервної передачі даних. Для організацій, які використовують хмарні сервіси, відеозв’язок, віддалене адміністрування та інші онлайн-ресурси, навіть короткочасна втрата з’єднання може спричинити порушення робочих процесів. Тому актуальним завданням є використання кількох незалежних каналів зв’язку для балансування навантаження та автоматичного перемикання на резервний канал.

Під час дослідження було використано VMware Workstation, у якому створено віртуальне тестове середовище. VMware Workstation дозволяє використовувати віртуальні мережеві адаптери, комутатори та окремі віртуальні мережі, що є зручним для тестування мережевих сценаріїв, зокрема пропускної здатності, стабільності та безпеки з’єднання. Основним мережевим пристроєм виступив MikroTik RouterOS CHR.

У тестовому середовищі було змодельовано мережеву інфраструктуру з трьома незалежними каналами зв’язку, зокрема основний WAN-канал, Starlink та LTE як резервне підключення. На початковому етапі було виконано базове налаштування MikroTik RouterOS: призначено IP-адреси інтерфейсам, налаштовано шлюзи, DNS, локальну мережу та правила NAT [1]. NAT використовувались для забезпечення безпечного доступу внутрішньої мережі до зовнішніх ресурсів через кожен з каналів зв’язку. Наступним етапом було застосовано механізм Mangle з використанням алгоритму per-connection-classifier для реалізації балансування навантаження, що дало можливість розподілити трафік на три частини. Далі виконано маркування з’єднання і пакетів відповідно до таблиць маршрутизації, які забезпечують рівномірний розподіл трафіку між усіма трьома каналами, що забезпечує коректну маршрутизацію навіть у випадку відмови одного з них.

Після налаштування балансування навантаження та відмовостійкості на базі MikroTik RouterOS було проведено моніторинг мережевого трафіку. Для цього використовувалися такі інструменти, як Ping, Tracert, SpeedTest, а також Wireshark для глибшого аналізу пропускної здатності, стабільності та безпеки

з'єднання [2]. Було проаналізовано фактичне проходження трафіку через різні інтерфейси, що дозволило оцінити коректність маршрутів, в тому числі їх легітимність, NAT та Mangle-правил, балансування навантаження та відмовостійкості. Під час моніторингу також аналізувалися показники якості з'єднання: затримка, втрата пакетів, перевантаження каналів, помилки конфігурації та пропускна здатність, що дозволило оцінити, який канал працює стабільніше, як поводиться мережа під час навантаження та чи виникають затримки під час перемикання на резервний канал.

Результати тестування підтвердили відсутність втрат пакетів (0%) і стабільну відповідь від усіх трьох каналів зв'язку (рис.1), що свідчить про коректність реалізованих налаштувань інтерфейсів, відсутність втрати пакетів і зміни маршруту, правил Mangle та NAT для кожного активного підключення.

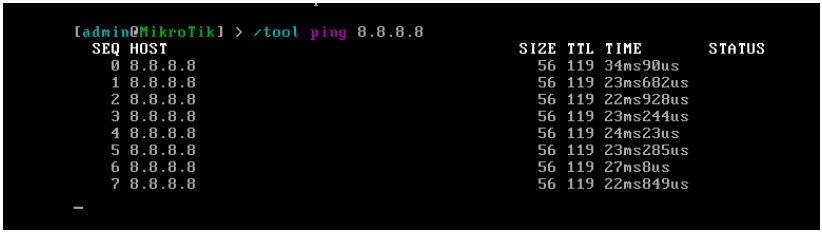


Рис.1. Результати тестування стабільності каналів зв'язку

Для тестування механізму відмовостійкості змодельовано сценарій, коли при відключенні одного з інтерфейсів трафік автоматично перенаправляється через інші канали зв'язку, що забезпечує безперервність з'єднання. У MikroTik RouterOS відмовостійкість використовується для резервування WAN-з'єднання та автоматичного перемикання на інший канал у разі відмови основного (рис.2).

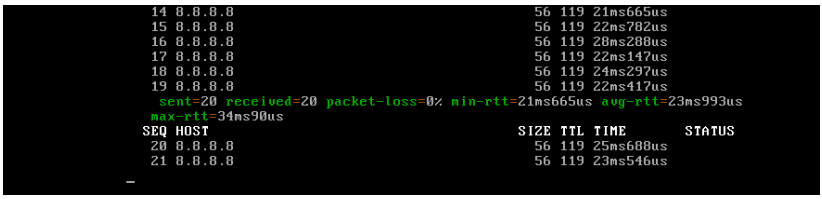


Рис.2. Результати тестування механізму відмовостійкості

Отже, у результаті роботи було налаштовано віртуальну мережеву інфраструктуру на базі MikroTik RouterOS з трьома незалежними каналами зв'язку, реалізовано балансування навантаження та механізм автоматичного перемикання на резервний канал. Результати тестування свідчать, що поєднання балансування навантаження і відмовостійкості дозволяє підвищити стабільність та безпеку мережевої інфраструктури. Балансування навантаження забезпечує ефективніше використання кількох каналів, а резервування дозволяє зберегти доступ до Інтернету у разі відмови одного з підключень. Проведено моніторинг

пропускної здатності, аналізу затримок, оцінено роботу каналів і виявлено особливості проходження пакетів у багатоканальній мережі.

1. O. Mishko, D. Matiuk, M. Derkach, Security of Remote IoT System Management by Integrating Firewall Configuration into Tunneled Traffic, Sci. J. TNTU, 115(3) (2024) 122–129.
2. Lypa, B., Horyn, I., Zagorodna, N., Tymoshchuk, D., Lechachenko T. (2024). Comparison of feature extraction tools for network traffic data. CEUR Workshop Proceedings, 3896, 1-11.

Підготовка фахівців з кібербезпеки в умовах розвитку штучного інтелекту: необхідність посилення фізичного та радіотехнічного компонентів освіти

УДК 004.056:37.091.3

Сергій Семендяй

Національний університет "Чернігівська політехніка", serhii_semendiai@icloud.com

Сучасний розвиток технологій штучного інтелекту суттєво трансформує підходи до підготовки фахівців у сфері інформаційних технологій та кібербезпеки [1]. Значна частина завдань, пов'язаних із програмуванням, автоматизацією аналізу коду, генерацією конфігурацій та навіть пошуком вразливостей, дедалі активніше виконується із використанням інтелектуальних систем. Це призводить до зміни вимог до професійних компетентностей майбутніх фахівців. Водночас існує категорія знань та практичних навичок, які не можуть бути повністю замінені засобами штучного інтелекту, оскільки вони безпосередньо пов'язані з фізичними процесами поширення сигналів, особливостями роботи апаратури та реальними характеристиками середовищ передавання інформації.

Освітні програми спеціальності «Кібербезпека та захист інформації» традиційно містять дисципліни, пов'язані з технічним захистом інформації, безпекою бездротових і мобільних систем, виявленням технічних каналів витоку інформації та фізичними основами технічних засобів розвідки. Саме ці дисципліни формують у студентів розуміння того, що будь-яка інформаційна система функціонує не лише на програмному чи мережевому рівні, а й у реальному фізичному середовищі, де існують електромагнітні поля, паразитні випромінювання, наведення, побічні електромагнітні випромінювання та інші фактори, здатні створювати додаткові загрози безпеці інформації.

У сучасних умовах особливого значення набуває підготовка фахівців, які здатні працювати із вимірювальним обладнанням та спеціалізованими комплексами технічного контролю. Йдеться про використання аналізаторів спектру, SDR-платформ «HackRF» та Ettus Research USRP, нелінійних локаторів, детекторів поля, детекторів бездротових протоколів, а також спеціалізованих пошукових комплексів, зокрема типу «ANDRE» та «Delta». Ефективне використання таких засобів вимагає не лише теоретичних знань, а й

практичного досвіду роботи з реальними сигналами та розуміння фізичних принципів функціонування апаратури.

Крім того, важливим напрямом розвитку сучасного освітнього процесу є створення можливостей для віддаленої роботи студентів із зазначеним обладнанням під час онлайн-навчання. Організація дистанційного доступу до SDR-платформ, засобів аналізу спектру та спеціалізованих вимірювальних комплексів дозволяє забезпечити безперервність практичної підготовки, розширити доступ студентів до лабораторної бази та сформувати навички роботи з реальними системами навіть в умовах змішаного або дистанційного формату навчання (рис. 1).

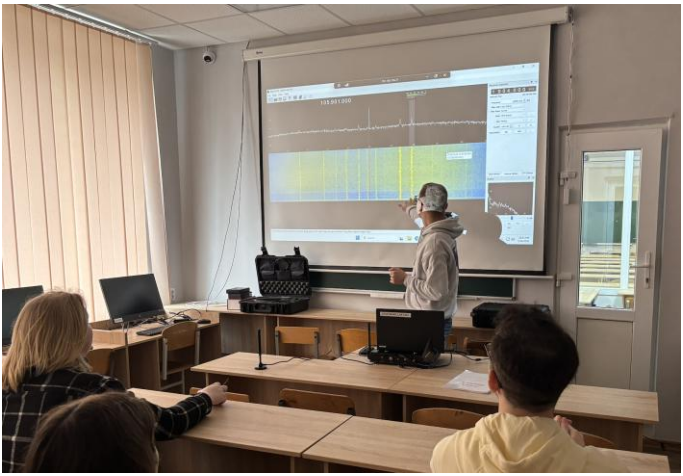


Рис. 1. Практична демонстрація можливостей дистанційного доступу до обладнання лабораторії кібербезпеки для аналізу спектра та дослідження сигналів

Важливим напрямом удосконалення сучасної освіти у сфері кібербезпеки є розширення практичного використання засобів моделювання процесів передавання та аналізу сигналів. Попри достатній рівень підготовки студентів у сфері програмного забезпечення, актуальним залишається формування глибокого розуміння фізичних процесів передавання інформації у дротових та бездротових середовищах, принципів формування спектру сигналів та впливу завад на функціонування систем зв'язку [2]. У цьому контексті перспективним є активніше використання середовищ MATLAB та «GNU Radio» у навчальному процесі.

Використання MATLAB дозволяє реалізовувати імітаційне моделювання каналів передачі інформації (в тому числі й технічних каналів витоку інформації), досліджувати вплив шумів та навмисних завад, аналізувати характеристики модуляції, демодуляції та кодування сигналів. Студенти можуть досліджувати залежність ймовірності бітової помилки від рівня сигнал/шум, моделювати роботу систем із частотним перестроюванням, оцінювати вплив широкосмугових та вузькосмугових завад. Такі підходи

формують фундаментальне розуміння процесів забезпечення стійкості систем передавання інформації, особливо в умовах навмисного завадового впливу (рис.2).

Своєю чергою, «GNU Radio» є ефективним інструментом для практичного опрацювання як аналогової, так і цифрової обробки сигналів та роботи із SDR-платформами [3]. Використання «GNU Radio» дозволяє студентам реалізовувати системи сканування спектру, виявлення та класифікації сигналів, досліджувати особливості різних протоколів бездротового зв'язку та аналізувати характеристики електромагнітного середовища. Особливу цінність має можливість інтеграції «GNU Radio» із SDR-пристроями, що дає змогу поєднати програмне моделювання з роботою у реальному радіоефірі.

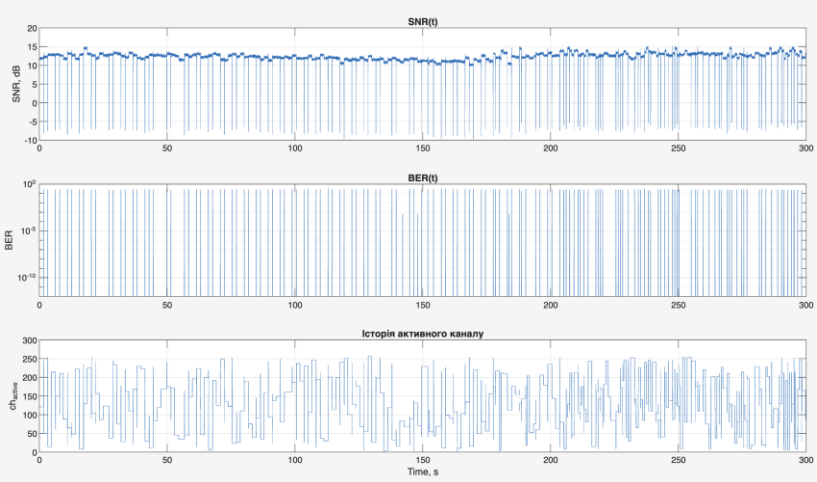


Рис. 2. Моделювання в середовищі MATLAB роботи захищеної системи передавання інформації

Важливим аспектом підготовки сучасного фахівця з кібербезпеки є розуміння принципів роботи засобів технічного контролю та їхніх обмежень. Студенти повинні не лише знати принципи функціонування детекторів поля чи нелінійних локаторів, а й розуміти їхні слабкі місця, можливі способи обходу та методики перевірки ефективності їх роботи. Наприклад, використання прихованих каналів передачі інформації, нестандартних схем модуляції або короткочасних імпульсних передач може суттєво ускладнювати процес виявлення сигналів. Аналогічно, сучасні бездротові пристрої можуть використовувати адаптивні алгоритми зміни частоти та потужності передачі, що також створює додаткові складності для систем моніторингу.

Не менш важливим є формування у студентів навичок аналізу електромагнітної обстановки та практичного виявлення технічних каналів витоку інформації. Фахівець з кібербезпеки повинен розуміти, яким чином можуть виникати побічні випромінювання, як вони поширюються у просторі,

які фактори впливають на дальність їх виявлення та якими методами може бути забезпечений їх контроль. У сучасних умовах ці питання набувають особливого значення у зв'язку із широким використанням бездротових технологій, IoT-пристроїв та програмно-визначених радіосистем.

Таким чином, розвиток технологій штучного інтелекту не зменшує актуальності фізичного та радіотехнічного компонентів підготовки фахівців з кібербезпеки, а навпаки – підвищує їх значення. Автоматизація окремих програмних задач призводить до того, що конкурентною перевагою майбутніх спеціалістів стають саме глибокі фундаментальні знання у сфері фізичних процесів передавання інформації, технічного захисту інформації та роботи зі спеціалізованими вимірювальними комплексами.

У сучасних умовах ефективна підготовка фахівців з кібербезпеки потребує поєднання програмно-аналітичних компетентностей із глибоким розумінням фізичних процесів передавання інформації, практичними навичками аналізу сигналів та роботи зі спеціалізованою вимірювальною апаратурою.

1. Горлинський, В. Освітні пріоритети підготовки фахівців з кібербезпеки в умовах воєнного стану в державі / Віктор Горлинський, Борис Горлинський // Information Technology and Security. – 2024. – Vol. 12, Iss. 2 (23). – Pp. 268-282. – Bibliogr.: 36 ref.
2. Лаптєв О. А., Марченко В. В. Застосування завад для захисту інформації від витоку радіоканалом // Сучасний захист інформації. 2025. № 1 (61). С. 89–97.
3. ПІДХІД ДО ВИЯВЛЕННЯ ТА КЛАСИФІКАЦІЇ РАДІОКЕРОВАНИХ МОДЕЛЕЙ ЗА ЇХ РАДІОСИГНАЛОМ / В. О. МАРТОВИЦЬКИЙ та ін. Вісник Херсонського національного технічного університету. 2025. Т. 2, № 2(93). С. 228–237. URL: https://doi.org/10.35546/kntu2078-4481.2025.2.2.27 (дата звернення: 16.04.2026).

Забезпечення стійкості бездротового каналу зв'язку для дистанційного керування мобільною платформою

УДК 004.056.5

Софія Яворівська[1], Марина Деркач[2], Тарас Лобур[3]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]avorivskasofia@gmail.com, [2]m_derkach@tntu.edu.ua, [3]lobur_t@tntu.edu.ua

У сучасних умовах стрімкого розвитку Інтернету речей (IoT) питання безпечного дистанційного керування пристроями набуває критичного значення [1]. Бездротові канали зв'язку, що використовуються для передачі команд, часто стають об'єктами кіберзагроз, таких як перехоплення сигналу або зловмисна модифікація даних. Для забезпечення надійної роботи подібних систем необхідно поєднувати апаратну стійкість із комплексним управлінням ризиками інформаційної безпеки.

Для реалізації такого підходу було розроблено систему дистанційного керування мобільною платформою. Головним вузлом системи виступає мікроконтролер ESP32-S2-WROVER, який забезпечує отримання, обробку

команд керування та відповідає за мережеву взаємодію через Wi-Fi. Апаратна частина включає підсистему стабілізації напруги на базі AMS1117-3.3 для живлення логічної частини та драйвер двигунів L298N для силового блоку. Використання драйвера обумовлене необхідністю комутації значних струмів для двигунів M1 та M2, якими ESP32 керує через сигнали IN1-IN4 та EnA/EnB. На рисунку 1 представлена принципова електрична схема системи дистанційного керування мобільною платформою на базі мікроконтролера ESP32-S2-WROVER, розроблена у середовищі KiCad.

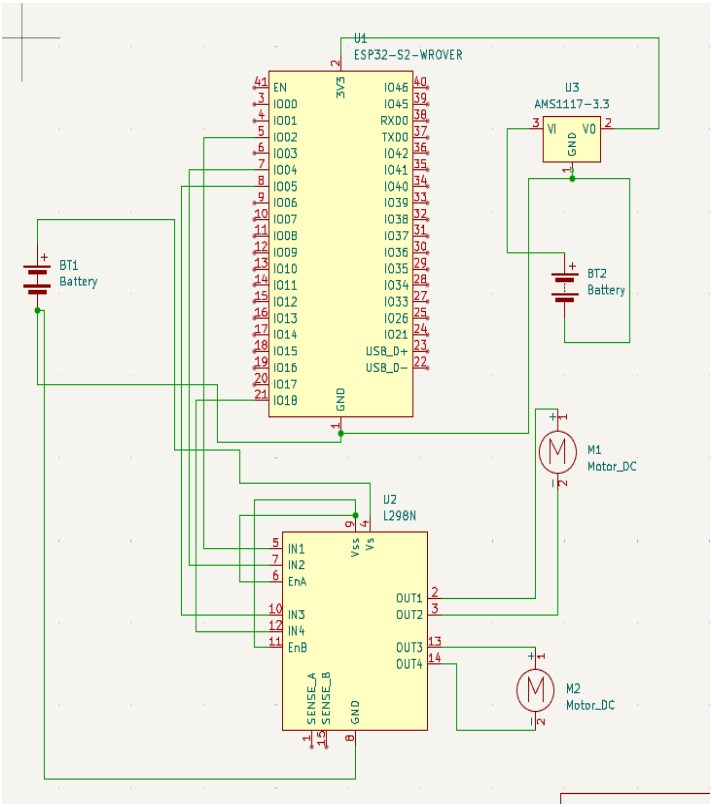


Рис. 1. Принципова електрична схема системи дистанційного керування мобільною платформою

Детальна архітектура взаємозв'язків між мікроконтролером, підсистему живлення та силовими компонентами системи представлена на принциповій схемі. Живлення драйверу двигунів від окремої батареї BT1, що відповідає тільки за мотори, адже їм потрібно більше енергії, та логічної частини від окремої батареї BT2 із загальною точкою заземлення GND дозволяє мінімізувати перешкоди в бездротовому каналі зв'язку.

Проєктування апаратної частини дозволило забезпечити фізичну надійність системи дистанційного керування мобільною платформою, проте робота у відкритому бездротовому середовищі вимагає додаткового рівня захисту на рівні передачі даних.

На основі аналізу розробленої архітектури було встановлено, що відсутність криптографічного захисту та використання відкритих мережевих портів створює умови для реалізації низки кіберзагроз, таких як несанкціоноване перехоплення трафіку (sniffing) або атаки типу DoS, що можуть призвести до повної втрати контролю над мобільною платформою.

Оцінка ризиків дозволила обрати стратегію їх зниження шляхом практичного впровадження механізмів захисту відповідно до стандарту ISO/IEC 27001/2022 [2], який надає систематизований підхід до управління інформаційною безпекою. Ці механізми включають:

- мережевий контроль доступу через фільтрацію MAC-адрес, фактично забороняючи підключення до мережі неавторизованим пристроями, і надаючи його лише заздалегідь внесеним MAC-адресам до списку дозволених;
- обов'язкове використання криптографічних засобів для захисту трафіку, а саме впровадження автентифікації команд, що передбачає механізм перевірки цифрового підпису кожної команди керування, це в свою чергу дозволяє системі ідентифікувати достовірність джерела;
- закриття всіх непотрібних мережевих портів на Wi-Fi-модулі ESP32, аби мінімізувати потенційні точки входу для атак.

Водночас для систематизації виявлених вразливостей було впроваджено ведення журналів логування спроб підключення до Wi-Fi та команд, які надсилаються для керування мобільною платформою, що дозволяє здійснювати моніторинг активності користувачів в мережі та вчасно ідентифікувати аномальні спроби доступу.

У результаті реалізовано захищену систему дистанційного керування мобільною платформою на базі мікроконтролера ESP32-S2-WROVER. Практичне дослідження, що включало серію тестувань, зокрема спроба підключення неавторизованого пристрою, перехоплення трафіку без шифрування, імітація повторної атаки (replay attack) та DoS-навантаження, підтвердило надійність обраних механізмів захисту у забезпеченні стійкості бездротового каналу зв'язку. Фактично несанкціоновані дії не мали успіху, а система дистанційного керування стабільно блокувала підозрілі запити.

1. Malyuta Y., Derkach M., Lobur T. (2025) Modelling Fog Computing Network Architecture for Secure IoT Data Processing. Security of Infocommunication Systems and Internet of Things, vol 3, no 2.
2. ISO/IEC 27001:2022. Information security, cybersecurity and privacy protection — Information security management systems — Requirements.

Контейнеризація як розвиток механізмів ізоляції процесів

004.056 Ситник Маргарита [1]

Харківський національний університет радіоелектроніки,
[1]marharyta.sytnyk@nure.ua

Розвиток сучасних програмних систем супроводжується зростанням вимог до їхньої надійності, масштабованості та переносимості. Початковим етапом у масовому спільному використанні апаратних ресурсів стала віртуалізація, однак перехід до мікросервісної архітектури виявив потребу в більш швидких та компактних рішеннях, що призвело до популяризації контейнеризації. Контейнеризація — це пакування програмного коду разом з усіма необхідними бібліотеками та залежностями. Її результатом є єдиний, легкий виконуваний файл, який стабільно працює на будь-якій інфраструктурі. Поява технології Docker значно спростила створення та управління контейнерами, зробивши цей підхід масовим [1, 2].

1. Ізоляція процесів та її реалізація

В основі контейнеризації лежить ізоляція процесів – здатність ядра ОС обмежувати видимість системних ресурсів для об’єктів та регулювати їхнє споживання. Завдяки цьому можливий запуск кількох застосунків на одній системі без їх взаємного впливу. Технічна реалізація ізоляції в ОС Linux базується на двох компонентах ядра:

- Namespaces: формують персоналізований «світогляд» для кожного процесу, ізолюючи файлову систему, ідентифікатори процесів, мережу, ідентифікатори користувачів тощо.
- Cgroups: розподіляють використання ресурсів і гарантують, що процес не конкуруватиме за пам'ять чи процесорний час, зарезервовані за іншими процесами.

2. Контейнеризація як елемент безпеки

Контекст безпеки контейнеризації тісно пов'язаний з архітектурою спільного ядра ОС. На відміну від повної ізоляції у віртуальних машинах, використання спільного ядра створює ризик втечі з контейнера у разі експлуатації вразливостей хост-системи. Для мінімізації поверхні атаки критично важливим є дотримання принципу найменших привілеїв, зокрема впровадження підходу Rootless Docker, а також регулярне сканування базових образів на наявність вразливостей до моменту їх розгортання в системі.

3. Контейнеризація та Docker

Функціонування контейнеризації на базі Docker забезпечується кількома ключовими компонентами, серед яких основними є образ (image), контейнер (container) та Dockerfile.

Образ являє собою незмінний шаблон, що містить програмний код, бібліотеки та всі необхідні залежності для запуску застосунку. Образи формуються пошарово, що дозволяє ефективно зберігати та повторно використовувати їх компоненти [3].

Контейнер є запущеним екземпляром образу, який функціонує як ізольований процес у системі. Контейнери створюються на основі образів і забезпечують виконання застосунків у відокремленому середовищі.

Dockerfile — це текстовий файл, що містить інструкції для автоматизованого створення образу. У ньому визначається базове середовище, необхідні залежності та команди, які потрібно виконати для підготовки контейнера до роботи.

4. Порівняння контейнерів та віртуальних машин

Контейнеризація та віртуалізація є різними підходами до ізоляції процесів, які мають суттєві відмінності в архітектурі та ефективності використання ресурсів. Контейнери створюють ізольовані середовища, що використовують спільне ядро операційної системи хоста, тоді як віртуальні машини функціонують як повноцінні незалежні системи з власним ядром.

Порівняння основних характеристик контейнерів і віртуальних машин наведено у таблиці 1.

Таблиця 1

Порівняння контейнерів і віртуальних машин

Характеристика	Контейнери	Віртуальні машини
Ядро ОС	Спільне	Окреме
Час запуску	1–3 секунди	20–60 секунд
Витрати пам’яті	50–200 МБ	512 МБ – 2 ГБ
Накладні витрати	2–5%	5–15%
Рівень ізоляції	Частковий	Повний

1. Роль контейнеризації та віртуалізації на рівні операційної системи в розвитку хмарно орієнтованих додатків. Наукова періодика Міжрегіональної Академії управління персоналом. URL: https://journals.maup.com.ua/index.php/it/article/view/4822/5115 (дата звернення: 26.04.2026).
2. Docker Inc. What is Docker?. Docker Documentation. URL: https://docs.docker.com/get-started/docker-overview/ (date of access: 28.04.2026).
3. Docker Inc. What is an image?. Docker Documentation. URL: https://docs.docker.com/get-started/docker-concepts/the-basics/what-is-an-image/ (date of access: 28.04.2026).

Critical Infrastructure Security: Electronic Communications Networks of Electronic Communications Operators

UDK: 004.056:004.7:621.39:351.86

Olena Shelest-Polishchuk [1],
Bohdan Skybun[2]

Kyiv Professional College of Communication,
[1] deksog@ukr.net, [2] skubyn.bogdan@gmail.com

The further development of digital technologies, digitization of information, communication, social, management and production processes together with the growth of levels of virtual cyberspace form a new digital reality and digital space. In turn, cyberspace combines with physical space to form a new cyber-physical space. At the same time, electronic communications ensure the full functioning of the cyber-physical space, and also provide an opportunity to receive, transmit, process, store and protect huge arrays of digital information produced by humanity. Thus, in modern realities, electronic communications act as an important transport system for the transmission of information, data and communication on a global world level. Also, electronic communications act as a catalyst for further development of digital society, digital economy and digital infrastructure.

Currently, the digital infrastructure creates the prerequisites for building a new level of critical infrastructure that is able to function in the modern realities of the growth of cyber threats. At the same time, the level of dependence of critical infrastructure on modern electronic communications is increasing, because computer networks, information and information and communication systems of enterprises, organizations, institutions, companies and corporations are built on their basis.Also, the globalization of management and production processes takes place on the basis of the use of international electronic communications and the global data transmission network, namely, monitoring, management, and security systems are created at the facility, corporate, and sectoral levels of production and management. Thus, "security of critical infrastructure" is characterized as "a state of protection of critical infrastructure, which ensures the functionality, continuity of work, restoreability, integrity and stability of critical infrastructure" [3, Article 1].

Today, stability and stability are quite important factors that characterize the functioning of the infrastructure under the influence of external and internal factors of influence.

All this requires electronic communications operators to build their own infrastructure, which would be resistant to external and internal factors of influence, as well as ensure sustainable functioning under the influence of cyber threats. Thus, "resilience of critical infrastructure" is defined as "the state of critical infrastructure, which ensures its ability to function normally, to adapt to constantly changing conditions, to withstand and quickly recover from threats of any

species" [3, Article 1]. It should be taken into account that in accordance with Government Resolution No. 1109, electronic communications are included in the List of critical infrastructure sectors [1], and therefore electronic communications operators need to ensure a sufficient level of security, stability and stability in relation to their own networks and infrastructure for the possibility of providing electronic communication services and electronic networks for other sectors of critical infrastructure. Thus, an important feature of electronic communications is that, on the one hand, they are a sector of critical infrastructure, and on the other hand, they are part of other sectors of critical infrastructure. Currently, the financial and banking spheres, the energy sphere, the security and defense sphere, the educational and medical spheres are quite sensitive to the stable and sustainable functioning of electronic communications.

In addition, electronic communication services and services based on electronic communications are the basis of the modern development of digital technologies, digitalization of society and economy, namely: e-education, e-banking, mobile banking, e-services, telemedicine, e-government, etc. Also, the development and spread of electronic communications among many countries and broad segments of the population gave impetus to the rapid development (quantitative and qualitative indicators) of mobile applications, various software products, various software products (specialized software), as well as the rapid growth of the number of users of messengers and social networks. The next important step was the transition to the virtual space of organizations, institutions, and enterprises. Yes, today almost all authorities (central, regional, district and local levels have their own websites and communicate with the population online), educational institutions of all levels, medical institutions, financial and banking institutions, the sphere of service provision, etc. All this also requires stable and sustainable functioning for the possibility of providing various services to the population.

At the same time, experts within the framework of the study "Increasing resilience by accelerating the digital transformation of business in Ukraine" consider digitization and digital transformation as an important element "for increasing resilience and facilitating recovery" [2, p.17], which directly depend both on the level of development of electronic communications and the global data transmission network, as well as on access to them by broad sections of the population.

1. Some issues of critical infrastructure objects, Resolution of the CMU dated October 9, 2020 No. 1109. https://zakon.rada.gov.ua/laws/show/1109-2020-%D0%BF#Text
2. Increasing sustainability by accelerating the digital transformation of business in Ukraine. https://surl.li/asaeds
3. On critical infrastructure: Law of Ukraine of November 16, 2021 No. 1882-IX Vedomosti Verkhovna Rada (VVR), 2023, No. 5, Article 13. https://zakon.rada.gov.ua/laws/show/1882-20#Text

Інтерактивні сценарії як інструмент викладання стандартів технічного захисту інформації

УДК 004.94:004.921

Юрій Скоренький[1], Руслан Козак[2],
Наталія Загородна[3], Тетяна Вітенько[4]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]skorenkyy@tntu.edu.ua, [2]zagorodna.n@gmail.com, [3]ruslank@tntu.edu.ua
[4]vitenko@tntu.edu.ua

Стрімка цифровізація вищої освіти зумовила нагальну потребу в зміні парадигми в бік модульних, стійких та інтерактивних педагогічних моделей. У контексті трансформації української вищої освіти інституційне виживання значною мірою залежить від переходу від екстреного дистанційного викладання до сталої цифрової педагогіки. Викладання основ технічного захисту інформації та узгодження зі стандартами (зокрема, ISO/IEC 27001) становить значний

педагогічний виклик. Від здобувачів освіти вимагається опанування комплексних систем прийняття рішень щодо оцінки ризиків та контролю фізичного доступу. Хоча гейміфікація широко застосовується для формування базової обізнаності з питань безпеки (наприклад, антифішингу), глибоке вивчення стандартів вимагає інтеграції складних компетенцій.

Для подолання цих викликів у межах пілотної ініціативи було здійснено перехід від статичних, перевантажених текстом матеріалів до гейміфікованого середовища в системі управління навчанням ATutor. Інтеграція інструментарію H5P дозволила створити нелінійні «сценарії розгалуження» (Branching Scenarios) на основі методології проблемно-орієнтованого навчання (PBL). Виступаючи в ролі аудиторів фізичної безпеки, студенти виявляли потенційні вразливості та приймали рішення, базуючись виключно на протоколах ISO 27001. Такі сценарії надійно занурюють здобувачів у ситуації, де вони повинні зважувати наслідки своїх дій, що стимулює критичне мислення. Емпіричні дані, отримані через навчальну аналітику з ATutor та H5P, продемонстрували важливу закономірність розвитку аналітичних навичок. Студенти, які спочатку обирали помилкові варіанти, були змушені аналізувати та виправляти змодельовані порушення і продемонстрували значно кращі результати у підсумкових комплексних оцінюваннях на критичне мислення порівняно з тими, хто пройшов сценарії з першої спроби.

Успішна реалізація цього проєкту слугує надійною моделлю для розбудови інституційного потенціалу, підтверджуючи, що інтерактивна та доступна цифрова педагогіка [1] може успішно функціонувати навіть в умовах серйозних зовнішніх викликів, ефективно озброюючи студентів критичними навичками, необхідними для майбутнього.

1. Zagorodna N., Skorenkyy Y., Kunanets N., Baran I., Stadnyk M., Augmented Reality-enhanced learning tools development for cybersecurity major. *CEUR Workshop Proceedings.* – 2022. – V. 3309. – p. 25–32.

Високопродуктивне розпізнавання облич на базі CUDA та Dlib у структурі комплексних систем забезпечення кібербезпеки

УДК 004.93

Олексій Смірнов[1], Віктор Заріцкий[2],
Костянтин Буравченко[3], Сергій Смірнов[4]

Центральноукраїнський національний технічний університет,
[1]dr.smirnovoa@gmail.com, [2]viktorzarickiy@gmail.com,
[3]buravchenkok@gmail.com, [4]smirnov.ser.81@gmail.com

Стрімкий прогрес у галузі розпізнавання образів зумовлений глибокою інтеграцією технологій штучного інтелекту в стратегічні сектори: від військової ідентифікації об’єктів БПЛА в умовах російсько-української війни до автоматизації медицини та промисловості. Це дослідження присвячене вдосконаленню систем безпеки, де детекція облич виступає фундаментальним компонентом інфраструктури "розумних міст", банківських установ та інформаційно-телекомунікаційних систем. Ключовим викликом для таких

систем є необхідність синтезу високої точності розпізнавання та обробки відеоданих у режимі реального часу. Ця вимога детермінує перехід від класичних алгоритмів до методів глибокого навчання (Deep Learning), які демонструють високу інваріантність до змін освітлення, ракурсних поворотів та оклюзій. Бібліотека dlib надає інструментарій для реалізації двох різних концепцій: Метод HOG (Histogram of Oriented Gradients): орієнтований на CPU, відрізняється швидкодією для фронтальних зображень, проте має низьку стійкість до нетипових ракурсів; Архітектура MMOD CNN (Maximum-Margin Object Detection): забезпечує прецизійну точність у динамічних сценаріях за рахунок аналізу багаторівневих ознак.

Проблема високої ресурсомісткості CNN-моделей (мільйони операцій на кадр), що створює "пляшкове горлечко" при використанні центральних процесорів, вирішується впровадженням технології NVIDIA CUDA. Паралелізація обчислень на графічних прискорювачах (GPU) дозволяє досягти оптимального балансу між точністю і частотою кадрів (FPS). Наукова новизна роботи полягає у комплексному порівняльному аналізі та кількісній оцінці ефективності переходу від класичного детектора HOG до GPU-прискореного методу MMOD CNN за допомогою розробленої методики бенчмаркінгу

Алгоритми інтелектуального аналізу даних та їх інтеграція з ШІ. Інструментарій dlib базується на двох підходах: класичному детектуванні та методах глибокого навчання. Алгоритм HOG, оптимізований для CPU, демонструє стабільну роботу з фронтальними обличчями, проте втрачає ефективність при зміні ракурсів. На противагу йому, архітектура MMOD CNN забезпечує прецизійну точність у динамічних умовах шляхом аналізу ієрархічних ознак. Висока обчислювальна складність нейромережі нівелюється застосуванням GPU-прискорення, що дозволяє виконувати масивні матричні операції паралельно та з мінімальними затримками.

Технологія CUDA для прискорення обчислень. Використання архітектури NVIDIA CUDA дозволяє перетворити графічний процесор на потужний обчислювальний вузол для виконання паралельних завдань. На відміну від центральних процесорів (CPU), спроектованих для послідовної обробки команд, GPU з тисячами спеціалізованих ядер забезпечує одночасне оперування великими масивами даних. Завдяки інтеграції бібліотеки dlib із CUDA, ресурсомісткі тензорні операції переносяться на відеокарту, що мінімізує завантаження CPU та гарантує роботу нейромережевих моделей у реальному часі. Для підтвердження ефективності цього методу було створено авторське програмне забезпечення мовою Python, яке за допомогою інструментів NumPy та Matplotlib здійснює комплексний бенчмаркінг та статистичну оцінку алгоритмів детекції

Методика експерименту та програмна реалізація. З метою забезпечення високої достовірності вимірювань розроблено інструментальний модуль FaceDetectionBenchmark. Алгоритм випробувань базується на кешуванні вхідних даних у RAM та обов’язковій фазі "warm-up" (прогріву) графічного процесора, що нівелює вплив ініціалізації CUDA-контексту на результат. Застосування прецизійних таймерів у поєднанні зі статистичним аналізом

середньоквадратичного відхилення дало змогу кількісно оцінити стабільність інференсу та виміряти рівень джиттеру в динамічному відео потоці.

Результати роботи методу HOG (CPU). Апробація класичного детектора на базі CPU продемонструвала високу прогнозованість часових показників. Середня латентність обробки кадру склала 25-30 мс (30-40 FPS), що варіюється залежно від вхідної роздільної здатності. Головною перевагою методу HOG визначено економічну доступність та відсутність потреби у спеціалізованих обчислювачах. Водночас виявлено суттєве обмеження точності: алгоритм втрачає працездатність при значних ракурсних відхиленнях обличчя та в умовах динамічної зміни освітленості.

Результати роботи методу CNN (GPU/CUDA). Використання CNN-детектора без апаратного прискорення виявило критичну нестачу продуктивності: затримка інференсу понад 800 мс (~1.2 FPS) робить застосування центрального процесора (CPU) недоцільним для систем реального часу. Впровадження технології NVIDIA CUDA забезпечило радикальний приріст швидкодії, скоротивши час обробки до 15–20 мс, що відповідає частоті понад 50 FPS. Окрім високої продуктивності, нейромережевий підхід продемонстрував вищу повноту детекції (recall), стабільно ідентифікуючи обличчя в складних ракурсах, де класичний метод HOG виявився неефективним

Порівняльна характеристика. Отримані дані підтверджують ефективність залучення GPU-ресурсів: показник інтенсифікації обчислень (speedup) склав 2,28 відносно базової архітектури.

Висновок. У роботі проведено оцінку та оптимізацію інструментів dlib для детекції облич. Доведено, що обмежена продуктивність CPU не дозволяє використовувати нейромережі (CNN) у режимі реального часу. Натомість впровадження технології NVIDIA CUDA забезпечило кратне прискорення: GPU-орієнтований CNN-підхід перевершив класичний метод HOG як за точністю, так і за швидкістю інференсу.

Security vulnerabilities at the Python LLM frameworks boundary

UDC 004.896:004.056.5

Oleksandr Karnaukhov[1], Nataliya Zagorodna[2], Oleh Yarema[3], Oleksandr Revniuk[4]

Ternopil Ivan Puluj National Technical University,
[1]karnaukhov@live.com, [2]zagorodna_n@tntu.edu.ua, [3]yarema.oleh.m@gmail.com, [4]revo0708@gmail.com

To execute complex workflows such as dynamic data analysis, file system management, and database querying, popular Python orchestration frameworks, including LangChain, LlamaIndex, and Ollama, frequently grant LLMs direct access to runtime environments [1]. This architecture often relies on underlying Python utilities like “exec()” or “eval()” to translate model outputs into system actions.

However, this design introduces a critical conflict between traditional deterministic software security and the probabilistic nature of linguistic models. While standard application security relies on strict input sanitization at the system boundary, AI agents inherently process untrusted natural language payloads directly from users.

This problem gives rise to “prompt injection” vulnerabilities, where an attacker makes malicious linguistic instructions designed to override the agent’s core system prompts [2].

The fundamental security flaw is at the structural intersection where natural language instructions transition into machine-executable Python code. If the orchestration framework lacks strict isolation or sanitization layers, a successful prompt injection can easily escape the application context [3]. This boundary effectively transforms a simple text interface into a vector for exploitation, with potential to escalate linguistic manipulation into unauthorized system command execution, file system traversal, or arbitrary Remote Code Execution (RCE) via Python’s subprocess or OS modules. Consequently, parsing errors and context window exploitation within Python ecosystems present a risk to production-grade AI deployments [4].

Table 1

Security architecture comparison

Dimension	Traditional Python Applications	Emerging Python LLM Agents
Input type	Deterministic - structured data, strings, integers	Probabilistic - unstructured natural language
Security boundary	Strict input sanitization	Open text interface directly exposed to user
Execution mechanism	Pre-defined, compiled, or hardcoded logic functions	Dynamic translation of text into code using “exec()” or “eval()”
Primary threat vector	SQL injection, buffer overflows, malicious code inputs	Prompt injection, linguistic manipulation overriding rules
Highest risk impact	Application crashes, data leaks via specific bugs	Remote Code Execution (RCE) via subprocess
Proposed mitigation	Standard input validation and parameterized queries	Abstract Syntax Tree (AST) validation prior to execution

To address these risks, an empirical, sandbox-based experimental framework to quantify the vulnerability threshold of current Python LLM orchestration layers is proposed. By constructing a prototype AI agent, we will subject both a baseline Python implementation and a standardized framework deployment to a curated benchmark suite of distinct prompt injections. These injections will explicitly target the boundary where natural language triggers backend execution, attempting to make the agent perform unauthorized operations. The experiment can evaluate framework resilience by measuring injection success rates and mapping the specific architectural blind spots where linguistic contexts successfully override Python-level security constraints. Experimental findings are expected to reveal deficiencies in native input sanitization, providing concrete data to advocate for deterministic mitigation strategies prior to code execution.

1. Chen, Y.-J., & Madisetti, V. K. (2025). Information security, ethics, and integrity in LLM agent interaction. Journal of Information Security, 16(01), 184–196. https://doi.org/10.4236/jis.2025.161010
2. Lee, D., Tiwari, M., & Miranda, B. (2026). Prompt infection: Llm-to-llm

prompt injection within multi-agent systems. У Lecture notes in computer science (с. 511–520). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-16092-8_28

3. AlSobeh, A., Gwarzo, Z., & Shatnawi, A. (2025). ShadowPlay: Engineering defenses against role-based prompt injection and dependency hallucination in llm-powered development. У 2025 international conference on cybersecurity and ai-based systems (cyber-ai) (с. 317–325). IEEE. https://doi.org/10.1109/cyber-ai66431.2025.11233258
4. Shi, J., Yuan, Z., Tie, G., Zhou, P., Gong, N., & Sun, L. (2026). Prompt injection attack to tool selection in LLM agents. У Network and distributed system security symposium. Internet Society. https://doi.org/10.14722/ndss.2026.230675

Метод попарного порівняння AHP для пріоритезації безпекових контролів SSDF у CI/CD

УДК 004.056:004.4

Тарас Лечаченко[1], Дмитро Войтович[2]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]lechachenko.taras@tntu.edu.ua, [2]voitovuch855@gmail.com

У зв'язку зі зростанням кількості кіберзагроз і складністю CI/CD-інфраструктур особливої актуальності набуває завдання пріоритизації безпекових контролів для DevSecOps та CI/CD-середовищ із подальшим кількісним оцінюванням їхньої ефективності. Існуючі підходи переважно зосереджуються на виявленні загроз або описі окремих механізмів захисту, однак недостатньо уваги приділяється формалізованим методам оцінювання та ранжування безпекових заходів відповідно до рівня критичності загроз.Як зазначають автори роботи [1], за останні п'ять років кількість досліджень у сфері DevSecOps суттєво зросла, проте питання пріоритезації безпекових контролів у контексті конкретних загроз для CI/CD-інфраструктури залишається недостатньо дослідженим.

Одним із перспективних підходів до розв'язання цієї задачі є застосування методу аналізу ієрархій (AHP, Analytic Hierarchy Process) [2], який базується на попарному порівнянні критеріїв та альтернатив. Використання AHP дозволяє формалізувати процес прийняття рішень, визначити вагомість окремих загроз і безпекових контролів, а також забезпечити обґрунтовану пріоритизацію заходів захисту для DevSecOps та CI/CD-середовищ. Для ранжування заходів захисту проти загроз CI/CD за основу взято STRIDE-методологію та визначено, що серед шести категорій STRIDE три в контексті CI/CD є першочерговими: підробка коду та артефактів, підміна сутності та розкриття інформації. Безпекові контролі для ідентифікації засобів захисту конвеєрів CI/CD було взято з NIST Secure Software Development Framework (SSDF) методології [3], оскільки вона забезпечує орієнтовані на життєвий цикл практики, що відповідають DevSecOps. Для забезпечення точності та практичної доцільності пріоритезації з переліку контролів SSDF було обрано підмножину з десяти засобів контролю:

- PO.1 — Визначення вимог безпеки для програмного забезпечення;
- PO.3 — Захист програмного забезпечення від відомих вразливостей;
- PW.1 — Ідентифікація та захист конфіденційних даних;
- PW.2 — Імплементація принципу найменших привілеїв;
- PW.4 — Безпечне зберігання та керування обліковими даними;
- RV.1 — Перевірка архітектури програмного забезпечення з точки зору безпеки;
- RV.2 — Перевірка коду для виявлення вразливостей безпеки;
- RV.3 — Перевірка програмного забезпечення на наявність логування та аудиту;
- RV.4 — Верифікація цілісності програмного забезпечення;
- PO.4 — Безперервний контроль та покращення практик.

Відповідно до представлених альтернатив (контролів) троє експертів із досвідом роботи у кібербезпеці понад 5 років здійснили попарне порівняння пріоритетності застосування контролів відносно трьох критеріїв загроз моделі STRIDE: Tampering, Spoofing, Information disclosure. Локальні пріоритети за критерієм Tampering обчислені за допомогою геометричного середнього згідно з методологією AHP для агрегації думок експертів. Відповідно агреговані значення Tampering представлені у таблиці 1.

Таблиця 1

Локальні пріоритети контролів SSDF за критерієм загрози Tampering

RV.4	PW.2	RV.3	PW.4	PO.3	RV.1	PW.1	PO.4	RV.2	PO.1
0.266	0.184	0.168	0.101	0.071	0.069	0.048	0.041	0.026	0.020

Як і індивідуальні оцінки експертів, так і агрегована матриця Tampering є узгодженою CR=0,05 (для 10 альтернатив RI (random index) = 1.49). Результати пріоритезації за критерієм загрози Tampering показали, що найважливішими контролями є RV.4 — «Верифікація цілісності програмного забезпечення» (0.266), PW.2 — «Імплементація принципу найменших привілеїв» (0.184) та RV.3 — «Перевірка програмного забезпечення на наявність логування та аудиту» (0.168), оскільки саме вони безпосередньо спрямовані на запобігання несанкціонованій модифікації коду та артефактів у CI/CD-конвеєрі. Середній рівень пріоритету отримали PW.4, PO.3 та RV.1, що забезпечують додатковий захист через керування обліковими даними, виявлення вразливостей і аналіз архітектури безпеки. Найменші значення отримали PW.1, PO.4, RV.2 та PO.1, оскільки їх вплив на протидію Tampering є більш опосередкованим.

Наведений приклад пріоритизації є лише фрагментом більш комплексного AHP-розрахунку, виконаного в межах дослідження. Повний аналіз включає розрахунок локальних і глобального пріоритетів засобів контролю SSDF не лише для загрози Tampering, але й для категорій Spoofing та Information Disclosure. Отримані результати підтверджують дієвість запропонованого

підходу в умовах обмежених ресурсів та необхідності визначення найбільш критичних заходів захисту для CI/CD-середовищ.

1. Prates, L., & Pereira, R. (2025). DevSecOps practices and tools. International Journal of Information Security, 24 (1), 11. https://doi.org/10.1007/s10207-024-00914-z
2. Saaty, T. L., & Vargas, L. G. (2012). Models, Methods, Concepts & Applications of the Analytic Hierarchy Process (2nd ed.). Springer. DOI: doi.org/10.1007/978-1-4614-3597-6
3. National Institute of Standards and Technology. (2022). Secure Software Development Framework (SSDF) version 1.1: Recommendations for mitigating the risk of software vulnerabilities (NIST Special Publication 800-218). U.S. Department of Commerce. https://doi.org/10.6028/NIST.SP.800-218

Кібербезпека систем екологічного моніторингу як елемент критичної міської інфраструктури

УДК 004.056:502.3:004.738.5

Андрій Станько[1], Ірина Дідич[2], Артем Гончаренко[3]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]andrjj.stanko@gmail.com, [2]iryna.didych1101@gmail.com
Київський національний університет будівництва і архітектури,
[3]hosting.pat@gmail.com

Системи екологічного моніторингу є важливим елементом цифрової інфраструктури міста, адже забезпечують збирання, передавання, оброблення та візуалізацію даних про якість повітря, стан води, шумове навантаження, радіаційний фон, мікроклімат і екологічні ризики. У smart city вони поєднуються з IoT-пристроями, геоінформаційними платформами, диспетчерськими службами та публічними панелями, тому достовірність і доступність даних впливають на управлінські рішення та реагування служб.

Актуальність проблеми зумовлена тим, що кібератаки на такі системи можуть спричинити втрату даних або формування хибної картини екологічної ситуації. Підміна телеметрії, блокування панелей, компрометація хмарної платформи чи втручання в алгоритми оброблення підвищують ризик запізнілого реагування на аварійні викиди, пожежі та техногенні інциденти. Метою роботи є обґрунтування підходу до кіберзахисту систем екологічного моніторингу як елемента критичної міської інфраструктури.

Відповідно до NIST Cybersecurity Framework 2.0, управління кібербезпекою охоплює ідентифікацію активів і ризиків, захист, виявлення подій, реагування, відновлення та організаційне управління [1]. Для екологічного моніторингу об’єктом захисту є повний ланцюг формування інформації: сенсор, контролер, канал зв’язку, шлюз, edge-вузол, API, хмарна платформа, аналітичний модуль, інтерфейс оператора і публічний вебсервіс.

Наукова новизна підходу полягає в розгляді системи екологічного моніторингу як кіберфізичного ланцюга довіри, де безпека визначається не лише захищеністю пристроїв, а й цілісністю маршруту даних від вимірювання до управлінського рішення. Запропоновано виділяти три рівні кіберзахисту: польовий, мережево-платформний та управлінсько-аналітичний.

Польовий рівень охоплює сенсори, мікроконтролери, автономні станції, джерела живлення та модулі передавання даних. Типові загрози: фізичне втручання, підміна сенсора, стандартні паролі, зміна прошивки або порушення калібрування. Базові вимоги до безпеки IoT-пристроїв передбачають відмову від універсальних паролів, безпечне оновлення, захист конфіденційних параметрів, мінімізацію поверхні атаки та керування вразливостями [2].

Мережево-платформний рівень включає шлюзи, канали зв'язку, VPN-з'єднання, MQTT/HTTP API, edge-вузли, хмарні сервіси, бази даних і механізми автентифікації. Його порушення можуть спричинити перехоплення або втрату телеметрії, DDoS-атаки й компрометацію доступу. На управлінсько-аналітичному рівні ключовими є валідація даних, панелі моніторингу, модулі прогнозування, оповіщення та звітність, оскільки саме тут викривлені дані трансформуються в помилкові рішення

Таблиця 1

Рівні кіберзахисту системи екологічного моніторингу

Рівень системи	Основні компоненти	Типові кіберризики	Захисні заходи
Польовий	сенсори, контролери, станції моніторингу	підміна показників, фізичне втручання, вразливі паролі	автентифікація пристроїв, захист прошивки, контроль калібрування
Мережево-платформний	шлюзи, канали зв'язку, API, хмарні сервіси	перехоплення даних, DDoS, компрометація доступу	шифрування, сегментація, VPN, журналювання подій
Управлінсько-аналітичний	панелі моніторингу, бази даних, модулі прогнозування	викривлення аналітики, недоступність сервісу, помилкові рішення	резервування, валідація даних, контроль доступу, аудит

З урахуванням сучасного ландшафту загроз слід поєднувати технічні й організаційні заходи. За даними ENISA Threat Landscape 2024, актуальними залишаються атаки на доступність, програми-вимагачі, загрози даним, соціальна інженерія та експлуатація вразливостей [4]. Для міського моніторингу це означає резервування каналів, регулярне оновлення ПЗ, централізоване управління доступом, журналювання, перевірку цілісності даних і тестування планів реагування.

Практична реалізація підходу може ґрунтуватися на паспорті кібербезпеки, що містить перелік активів, потоки даних, критичні компоненти, модель доступу, вимоги до резервного копіювання та порядок реагування на інциденти. Отже, кібербезпека систем екологічного моніторингу є умовою надійності

розумного міста, екологічної безпеки та довіри населення. Подальші дослідження доцільно спрямувати на кількісне оцінювання ризиків, інтеграцію з SIEM/SOC-рішеннями та виявлення фальсифікованих або аномальних даних.

1. Bacco M., Delmastro F., Ferro E., Gotta A. Environmental monitoring for smart cities. IEEE Sensors Journal. 2017. Vol. 17(23). P. 7767-7774.
2. Zanella A. et al. Internet of Things for smart cities. IEEE Internet of Things Journal. 2014. Vol. 1(1). P. 22-32.
3. Sicari S. et al. Security, privacy and trust in Internet of Things: The road ahead. Computer Networks. 2015. Vol. 76. P. 146-164.
4. Demertzi V., Demertzis S., Demertzis K. An overview of cyber threats, attacks and countermeasures on the primary domains of smart cities. Applied Sciences. 2023. Vol. 13(2). Article 790.

Застосування методу PERT для оцінки трудомісткості задач у мобільних застосунках управління проєктами

УДК 004.42

Стасюк Сергій[1], Мудрик Іван[2]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]serhii_stasiuk1107@tntu.edu.ua, [2]i1mudryk@tntu.edu.ua

Управління сучасними IT-проєктами стикається з двома критичними викликами: проблемою точної оцінки трудомісткості задач та необхідністю гарантування конфіденційності проєктної інформації. Мобільні застосунки для трекінгу задач обробляють надзвичайно чутливі корпоративні дані (строки, розподіл ресурсів, архітектурні інсайти, бізнес-процеси). Відповідно, інструмент планування має не лише вирішувати проблему зриву дедлайнів, але й відповідати суворим вимогам інформаційної безпеки.

Точна оцінка трудомісткості задач є одним із ключових викликів у сучасному IT-проєктному менеджменті. Сучасні менеджери проєктів все частіше використовують мобільні застосунки для трекінгу задач та управління ресурсами. Однак, однією з найскладніших задач залишається точна оцінка трудомісткості на етапі планування або безпосередньо "на ходу" (наприклад, під час дейлі-мітингів чи зустрічей з клієнтами). Традиційні однопунктові методи систематично ігнорують невизначеність і ризики, що призводить до зривів термінів та перевитрати бюджету. Метод PERT (Program Evaluation and Review Technique) вирішує цю проблему через три сценарії для кожної задачі: оптимістичний (О), реалістичний (Р) та песимістичний (П). Очікуваний час (Т) обчислюється за формулою бета-розподілу:

$$T = \frac{О + 4Р + П}{6}$$

Така зважена оцінка враховує статистичний розподіл можливих результатів і є суттєво точнішою за однопунктове прогнозування [1].

Аналіз популярних інструментів (Trello, Jira, Asana) показує, що вони орієнтовані на хмарну синхронізацію, що розширює поверхню атаки (attack

surface) та створює ризики витоку даних через компрометацію мережевого трафіку або хмарних баз даних.

Аналіз наявних інструментів (Trello, Jira, Asana, Todoist) засвідчує відсутність вбудованої підтримки методу трьох точок як основного механізму планування, а також обмежені можливості захисту локальних даних. Більшість рішень орієнтовані на командну роботу з розгалуженою функціональністю, що робить їх надлишковими для індивідуального розробника чи малої команди [2].

Для вирішення цих прикладних проблем захисту інформації, пропонується децентралізована архітектура мобільного застосунку за патерном MVVM (Model-View-ViewModel), де логіка PERT інкапсульована незалежно від UI. Головною особливістю є "offline-first" підхід: усі три оцінки та метадані проєкту зберігаються виключно в реляційній базі даних на самому пристрої (наприклад, SQLite/Room для Android) без примусової прив'язки до хмарних сервісів. Це радикально мінімізує загрози, пов'язані з безпекою хмарних обчислень.

Зберігання даних локально вимагає надійного захисту кінцевої точки (endpoint security). Для протидії криміналістичному аналізу мобільних пристроїв та несанкціонованому вилученню даних (наприклад, у разі втрати смартфона) застосовуються методи прикладної криптології:

- Шифрування "Data at Rest": Використання бібліотек типу SQLCipher для повного AES-256 шифрування локальної бази даних SQLite.
- Захист ключів доступу: Інтеграція з апаратними сховищами ключів мобільних ОС (Android Keystore / Apple Secure Enclave) для генерації та безпечного зберігання криптографічних ключів.
- Біометричне розмежування доступу: Додатковий рівень аутентифікації на рівні самого застосунку для підтвердження особи користувача перед доступом до проєктних даних.

З точки зору інформаційної безпеки та захисту даних, мобільні застосунки для управління проєктами обробляють чутливі дані про терміни, ресурси та бізнес-процеси організації. Сучасні стандарти кібербезпеки (NIST, ISO/IEC 27001) вимагають шифрування локального сховища, розмежування прав доступу та захисту від несанкціонованого втручання. Інтеграція таких принципів на етапі проєктування застосунку дозволяє мінімізувати ризики витоку корпоративних даних [3].

Бізнес-аналітика на основі PERT-оцінок відкриває додаткові можливості для прийняття управлінських рішень. Накопичені статистичні дані про реальне та прогнозоване виконання задач дозволяють виявляти систематичні відхилення в плануванні, оцінювати продуктивність і коригувати майбутні оцінки. Такий підхід перетворює застосунок з інструменту планування на аналітичну платформу підтримки рішень.

Інтеграція вищезазначених принципів на етапі проєктування («Security by Design») дозволяє створити продукт, що відповідає сучасним стандартам кібербезпеки, зокрема концепціям NIST та ISO/IEC 27001 у частині розмежування прав доступу, захисту від несанкціонованого втручання та забезпечення конфіденційності інформаційних активів.

1. Wysocki, R. K. (2019). Effective Project Management: Traditional, Agile, Extreme, Hybrid. — New Jersey: Wiley. — 696 p.
2. Phillips, J. (2021). PMP Project Management Professional Study Guide. — New York: McGraw-Hill. — 592 p.
3. Ross, R. S., et al. (2020). Security and Privacy Controls for Information Systems and Organizations. — NIST SP 800-53 Rev. 5. — 492 p.
4. Bryk O., Mudryk I., Holubovskyi M., Stoianov Y. Machine learning models and methods aspects of processing unstructured data. Proceedings of the 1st International Workshop on Bioinformatics and Applied Information Technologies (BAIT 2024), Zboriv, Ukraine, 2024. 2024. P. 64–74.

Гібридний метод приховування ЦВЗ у цифрових зображеннях

УДК 004.056.5

Ірина Борисенко[1], Даніїл Стрельченко[2]

Національний університет «Одеська політехніка»,
[1]borisenko.i.i@op.edu.ua, [2]10182258@stud.op.edu.ua

Для побудови стійких систем ЦВЗ використовуються різні підходи перетворення контейнера, наприклад, дискретне перетворення Фур'є (ДПФ) та сингулярний розклад матриць (SVD), які демонструють високу ефективність у задачах спектрального аналізу та приховування даних. ДПФ дозволяє перейти у частотну область, забезпечуючи стійкість ЦВЗ до геометричних атак, SVD виділяє стабільний енергетичний кістяк зображення (сингулярні числа), який є надзвичайно стійким до стиснення алгоритмом JPEG та шумових перешкод.

Актуальність теми полягає в необхідності інтеграції переваг обох математичних апаратів у єдину комбіновану (гібридну) схему, яка забезпечуватиме максимальну стійкість захисту авторських прав до комплексних атак, зберігаючи при цьому високу візуальну якість зображення.

Метою роботи є розробка комбінованого стеганографічного алгоритму вбудовування ЦВЗ, що поєднує частотну декомпозицію на основі ДПФ і SVD перетворень та дослідження його стійкості до стеганоаналітичних атак.

Доведено [1], що перше (найбільше) сингулярне число S_{11} сингулярного розкладу матриці контейнера концентрує в собі основну енергію зображення, що робить його модифікацію надзвичайно стійкою до зовнішніх впливів, цю властивість будемо використовувати в алгоритмі вбудовування ЦВЗ.

Алгоритм вбудовування. Зображення-контейнер розбивається на незалежні блоки розміром 8×8 пікселів, для кожного блоку виконується ДПФ, до коефіцієнтів середньої частоти застосовується кільцева маска з радіусами R_{min} та R_{max}, яка виділяє коефіцієнти, які є найбільш оптимальними для приховування даних, після чого до виділеного кільця застосовується SVD. Значення S_{11} ділиться на коефіцієнт сили вбудовування α – одержуємо S'_{11}. Якщо секретний біт дорівнює 0, значення S'_{11} округлюється до найближчого парного числа, якщо 1 – до непарного. Отримуємо матрицю S_{new}. Далі робимо зворотні перетворення: $US_{new}V^T$ та зворотне ДПФ. Відновлюється повідомлення шляхом зворотних перетворень: ДПФ –> кільце з радіусами R_{min} та R_{max} –> SVD виділеного кільця частот; зчитується перше сингулярне число,

по його парності роблять висновок, який саме біт (0 чи 1) було приховано у даному блоці.

Проведені експерименти показали високу візуальну якість стегоконтейнера (PSNR > 35 dB) та високу стійкість до шумів: гаусівського (при σ = 5 – відсоток відновлення повідомлення до 100%; σ = 15 – середній рівень шуму, відновлення до 80%); мультиплікативного (при низькому рівні імпульсного шуму P < 0,5 %) точність відновлення майже 100 %. Стійкість до компресії JPEG (Q=70, Q=50), завдяки варіативному вибору параметра α, одержували PSNR 45-40дБ.

1. Кобозєва А.А., Хорошко В.О. Аналіз захищеності інформаційних систем. К.: Вид. ДУІКТ, 2010. 309 с.

Вимоги до простежуваності та обґрунтованості результатів вимірювання критичності кіберінцидентів

УДК 004.056:006.91

Ярослав Тарасенко[1], Роман Орлов[2]

[1]Державний науково-дослідний інститут випробувань і сертифікації озброєння та військової техніки, yaroslav.tarasenko93@gmail.com,
[2]Державний університет інформаційно-комунікаційних технологій, romanorlov0110@gmail.com

У сучасних інформаційно-комунікаційних системах важливою умовою обґрунтованого реагування та вибору пріоритетів подальших дій є визначення критичності кіберінцидентів. У задачах реагування на кіберінциденти результат вимірювання критичності є основою для прийняття рішень. У роботі [1] автоматизована класифікація кіберінцидентів за рівнями тяжкості ґрунтується на наборі ознак і моделей їх інтерпретації. Підсумкова оцінка критичності повинна бути простежувана та спиратися на цифрові спостереження.

Перша вимога полягає у простежуваності результату до вихідних даних. Кожна оцінка повинна бути відновлювана на основі журналів подій, телеметрії, засобів виявлення та інших відомостей, використаних для її формування. Друга вимога простежуваності відноситься до обробки інформації та полягає у фіксації послідовності обробки даних і формуванні правил переходу від спостережень до підсумкової оцінки. Такий підхід ґрунтується на охопленні сутності, процесів та користувачів, пов’язаних з історією даних, що представлено у роботі [2]. До третьої вимоги варто віднести структурну обґрунтованість результату. Під структурною обґрунтованістю мається на увазі потребу у врахуванні впливу сукупності змістовних компонентів та контексту події на оцінку критичності. Четверта вимога, пов’язана з пояснюваністю, націлена на можливість демонстрації шляхів формування підсумкового результату критичності за рахунок ознак, вагових коефіцієнтів та порогових значень. Доцільність такої вимоги пояснена у роботі [3], де непрозорість моделей виступає чинником зниження довіри до кібербезпекових рішень, зокрема з використанням штучного інтелекту. П’ятою вимогою є відтворюваність результату. Отже, простежуваність даних і перетворень,

структурна обґрунтованість, пояснюваність та відтворюваність розглядаються як базові вимоги до результатів вимірювання критичності кіберінцидентів.

1. DeCastro-Garcia N., Munoz Castaneda A.L., Fernandez-Rodriguez M. Machine learning for automatic assignment of the severity of cybersecurity events. Computational and Mathematical Methods. 2020. Vol. 2, № 1. URL: https://doi.org/10.1002/cmm4.1072 (дата звернення: 02.05.2026)
2. Pan B., Stakhanova N., Ray S. Data provenance in security and privacy. ACM Computing Surveys. 2023. Vol. 55, Issue 14s. URL: https://doi.org/10.1145/3593294 (дата звернення: 03.05.2026).
3. A survey on explainable artificial intelligence for cybersecurity / G. Rjoub et al. IEEE transactions on network and service management. 2023. Vol. 20, № 4. P.5115-5140.

Відповідальність під час використання штучного інтелекту в судочинстві: теоретичні засади, правові виклики

УДК 347.9:004.8

Віталій Вітів

Тернопільський національний технічний університет імені Івана Пулюя

У сучасному світі стрімке впровадження штучного інтелекту в усі сфери суспільного життя створює значні виклики для судової системи та юриспруденції, зокрема щодо розподілу відповідальності [1, 2].

Алгоритми, що використовуються для аналізу даних, підготовки процесуальних документів, прогнозування судових рішень та автоматизації адміністративних процедур, здатні суттєво підвищити ефективність правосуддя, зменшити навантаження на суддів і сприяти більшій доступності справедливості. Водночас постає фундаментальне питання: хто повинен нести юридичну та моральну відповідальність за помилки, алгоритмічну упередженість, «галюцинації» чи порушення прав учасників процесу, спричинені діями штучного інтелекту. Додатковим ризиком є розголошення конфіденційної інформації та професійної таємниці через завантаження матеріалів до систем штучного інтелекту.

Українська правова система активно розвивається в напрямку цифровізації, однак комплексне регулювання використання штучного інтелекту в судочинстві ще не сформоване. Правові позиції в Україні щодо цієї технології є обережними. Зокрема, стаття 16 Кодексу суддівської етики передбачає, що використання суддею технологій штучного інтелекту є допустимим лише за умови, якщо воно не впливає на незалежність та неупередженість судді, не стосується оцінки доказів, процесу ухвалення рішень і не порушує вимог законодавства [5]. Аналогічний підхід закріплено в Законі України «Про адміністративну процедуру»: адміністративний орган несе відповідальність за акти, прийняті в автоматичному режимі (статті 62–69) [4].

В Україні наразі відсутнє офіційне законодавче визначення штучного інтелекту, хоча стратегічні засади його розвитку визначено в Розпорядженні Кабінету Міністрів України від 2 грудня 2020 року № 1556-р.

У Європейському Союзі чітке визначення міститься в Регламенті (ЄС) 2024/1689 (Artificial Intelligence Act). Згідно зі статтею 3(1), система штучного інтелекту — це машинно-базована система, яка працює з певним рівнем автономності, може адаптуватися після розгортання та генерує виходи, що впливають на фізичне чи віртуальне середовище [3].

Використання штучного інтелекту в судочинстві актуалізує питання співвідношення позитивного і природного права. Згідно з доктриною Томи Аквінського («Сума теології»), відповідальність випливає з раціональної природи людини, її свободи волі та моральної автономії [6, 7]. Штучний інтелект, як продукт людського розуму, не володіє совістю та здатністю до справжньої розсудливості, тому делегування йому остаточної відповідальності суперечить принципам природного права.

На практиці Верховного Суду України матеріали, згенеровані штучним інтелектом, не визнаються повноцінними доказами через відсутність автора, непрозорість алгоритмів («чорна скринька») та ризик недостовірної інформації. Бєлов Д.М. та Бєлова М.В. підкреслюють потенціал штучного інтелекту в обробці даних, але акцентують увагу на ризиках алгоритмічної несправедливості, порушення конфіденційності та проблемах відповідальності [2]. Подібні висновки містяться в дослідженнях Деркача В.Г., Прокопович-Ткаченко Є.Д. та Руденка Є.Г. [9]. Відповідальність у таких випадках розподіляється між користувачем, розробником системи та державою.

Важливий внесок у розуміння еволюції правових позицій зробив Берназюк Я., який наголошує на необхідності розвитку штучного інтелекту в межах «м'якого» права — рекомендацій Ради Європи, ЮНЕСКО та національних етичних кодексів [1]. Міжнародні стандарти, зокрема Рамкова конвенція Ради Європи про штучний інтелект та права людини (2024) та EU AI Act, класифікують застосування штучного інтелекту в правосудді як високоризикове, вимагаючи прозорості, людського нагляду та механізмів оскарження [3].

Суддя Крат В. у своїх роботах справедливо акцентує увагу на необхідності професійних стандартів, подібних до тих, що застосовуються в Австралії та США, де юристи зобов'язані розкривати використання штучного інтелекту та перевіряти достовірність згенерованого контенту [8].

Таким чином, для ефективного регулювання необхідно внести зміни до процесуальних кодексів, розробити спеціальний закон про використання штучного інтелекту в юстиції, запровадити сертифікацію систем, обов'язковий аудит та посилити етичну підготовку суддів і юристів. Відповідальність за використання штучного інтелекту в судочинстві ніколи не може бути повністю передана технології. Вона завжди залишається за людиною як моральним агентом, носієм розуму та совісті. Майбутнє правосуддя полягає в гармонійному поєднанні технологічних інновацій та вічних принципів природного права, де відповідальність є невід'ємною частиною людської гідності.

1. Берназюк Я. ШІ в національному судочинстві: еволюція правових позицій : презентація. Національна школа суддів України, 5 березня 2026.
2. Бєлов Д. М. Штучний інтелект в судочинстві та судових рішеннях, потенціал та ризики / Д. М. Бєлов, М. В. Бєлова // Науковий вісник Ужгородського національного університету. Серія : Право. 2023. Вип. 78. Ч. 2. С. 315–319.
3. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). URL: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng (дата звернення: 16.05.2026).
4. Закон України «Про адміністративну процедуру» від 17.02.2022 № 2073-IX. URL: https://zakon.rada.gov.ua/laws/show/2073-20#Text (дата звернення: 16.05.2026).
5. Кодекс суддівської етики : затверджений XX черговим з'їздом суддів України 18 вересня 2024. URL: https://zakon.rada.gov.ua/rada/show/n0001415-24#Text (дата звернення: 16.05.2026).
6. Aquinas T. Summa Theologica / T. Aquinas. URL: https://www.newadvent.org/summa/ (дата звернення: 16.06.2026).
7. Попов Д. І. Філософсько-правові засади застосування природного права при здійсненні судового угляду : автореф. дис. на здобуття наук. ступеня канд. юрид. наук : спец. 12.00.12 – філософія права / Д. І. Попов ; Львів. держ. ун-т внутр. справ. – Львів : ЛьвДУВС, 2014. – 22 с. – URL: https://dspace.lvduvs.edu.ua/bitstream/1234567890/150/1/popov.pdf (дата звернення: 16.05.2026).
8. Крат В. Штучний інтелект і судочинство : презентація. Верховний Суд, 6 грудня 2024.
9. Деркач В. Г. Використання штучного інтелекту в судовому процесі України: правові, етичні та процесуальні аспекти / В. Г. Деркач, Є. Д. Прокопович-Ткаченко, Є. Г. Руденко // Юридичний науковий електронний журнал. 2025. № 3. С. 460–463.

Розробка безпечного клієнтського інтерфейсу веб-платформи для ігрової спільноти з використанням React.js та TailwindCSS

УДК 004.056

Артем Теклюк

Державний університет інформаційно-комунікаційних технологій
st7890208@stud.duikt.edu.ua

Сучасні веб-платформи для ігрових спільнот повинні забезпечувати не лише зручність використання та швидку взаємодію користувачів із контентом, а й належний рівень інформаційної безпеки. Зростання кількості онлайн-сервісів та інтеграції веб-платформ із зовнішніми ресурсами підвищує ризики

несанкціонованого доступу, витоку даних та атак на клієнтську частину застосунку. У таких умовах важливого значення набуває використання сучасних frontend-технологій із підтримкою механізмів безпечної взаємодії користувачів із системою.

Одними з найбільш популярних інструментів для створення сучасних веб-застосунків є React.js та TailwindCSS. React.js дозволяє реалізовувати компонентний підхід до розробки інтерфейсу та забезпечує ефективне керування станом застосунку, а TailwindCSS використовується для створення адаптивного та оптимізованого інтерфейсу користувача.

Метою роботи є розробка безпечного клієнтського інтерфейсу веб-платформи для ігрової спільноти з використанням React.js та TailwindCSS, що забезпечує ефективну взаємодію користувачів із системою та підвищення рівня захисту даних.

У процесі дослідження було проаналізовано сучасні підходи до побудови frontend-інтерфейсів та визначено основні загрози безпеці клієнтської частини веб-платформи. Реалізовано клієнтський інтерфейс, який забезпечує перегляд новин, турнірів та іншого тематичного контенту, а також підтримує механізми авторизації користувачів і контроль доступу до функціоналу системи.

Для підвищення безпеки застосовано механізми захисту від XSS-атак, безпечного зберігання токенів авторизації та обмеження доступу до окремих компонентів інтерфейсу. Використання компонентної архітектури React.js забезпечило модульність системи та спростило реалізацію захищених елементів взаємодії користувача із платформою.

Отримані результати підтверджують ефективність використання сучасних frontend-технологій для створення безпечних веб-платформ. Використання React.js та TailwindCSS дозволяє забезпечити адаптивність інтерфейсу, покращити продуктивність системи та підвищити рівень захисту клієнтської частини веб-застосунку.

1. React Documentation [Електронний ресурс]. – Режим доступу: https://react.dev/.
2. Tailwind CSS Documentation [Електронний ресурс]. – Режим доступу: https://tailwindcss.com/docs/.
3. Banks A., Porcello E. Learning React. – O'Reilly Media, 2020.

Full cycle of responding to cyber incidents in the public sector

UDC 004.056 Iryna Tegubenko[1], Viktor Kotetunov[2]

State Scientific and Research Institute of Cybersecurity Technologies and Information Protection, [1]wr.irttri@gmail.com [2]v.kotetunov@gmail.com

An important factor in ensuring cyber protection of Ukraine in modern conditions is the availability of civil servants, primarily of higher categories, comprehensive knowledge of the current state of cybersecurity, trends of changes in the industry, and understanding of the continuity of the cyclical process of responding to cyber incidents.

The reality of the current moment is characterized by the acceleration of the evolution of cyber security, in particular due to the influence of artificial intelligence technologies [1], an increase in the speed of computing processes, a rapid increase in the complexity and volume of data that must be processed and taken into account in the processes of analyzing the state of cyber security, the transition to a risk-oriented approach to cyber security, the correlation of domestic approaches with the international practices of CISA [2], NIST [3], MITER ATT&CK[4], etc.

Each of the participants in public administration should be familiar with the processes of organizing the full cycle of response to cyber incidents, its main components, properties, and resources, and understand their place, functions, and acquire the necessary competencies. The complete cyber incident response cycle consists of several interrelated modules, namely: the preparation, the detection and analysis, the deterrence, the elimination, the recovery, and the analysis of the effectiveness of cyber incident response measures. In fact, it is a cyclical process, the main one of which is the preparation stage, in terms of the amount of resources, time, and qualifications required. The stage is performed almost continuously. If immediate cyber incident response actions begin at the time a cyber incident is initiated, the outcome for the institution will be known to be negative.

1. Hilpisch Y., Ali M.G., Jasim A.K., Abdulrahman S.A.R., Abu-AlShaeer M.J., Almansoori K.W.N., Tregubenko I. Strategic Technological Integration and National Industrial Resilience: Assessing AI-Driven Efficiency Across Critical Sectors. (2026), 2855 CCIS, pp. 507 - 522, DOI: 10.1007/978-3-032-17023-1_30
2. Federal Government Cybersecurity Incident and Vulnerability Response Playbooks. URL: https://www.cisa.gov/resources-tools/resources/federal-government-cybersecurity-incident-and-vulnerability-response-playbooks (application date 07.05.2026).
3. NIST SP 800-53 Security and Privacy Controls for Information Systems and Organizations. URL: https://csrc.nist.gov/pubs/sp/800/53/r5/upd1/final (application date 07.05.2026).
4. Adversarial Tactics, Techniques, and Common Knowledge (MITRE ATT&CK). URL: https://attack.mitre.org (application date 07.05.2026).

Розробка алгоритму виявлення ШІ-згенерованих зображень на основі машинного навчання

УДК 621.395.7 (043.2)

Фляк Владислав

Національний університет "Одеська політехніка",10328099@stud.op.edu.ua

Сьогодні цифрові технології радикально змінили спосіб сприйняття інформації. Фотографії та відео вже не є надійним доказом реальності події, оскільки штучний інтелект здатний створювати повністю синтетичний контент, який важко відрізнити від справжнього. ШІ-згенеровані зображення стали масовим явищем у соціальних мережах, рекламі, медіа та навіть судовій практиці. Це призводить до кризи довіри до візуальної інформації та створює

серйозні загрози для кібербезпеки, журналістики та суспільства загалом. Актуальність теми обумовлена необхідністю розробки надійних алгоритмів автоматичного виявлення такого контенту на основі машинного навчання.

ШІ-згенеровані зображення – це синтетичні візуальні матеріали, створені алгоритмами штучного інтелекту, переважно методами глибокого навчання. На відміну від традиційних фотографій, які фіксують реальний світловий потік сенсором камери, вони формуються математично – шляхом моделювання розподілу пікселів. Сучасні моделі (GAN, Diffusion Models) здатні відтворювати складні структури: людські обличчя з реалістичною мімікою, текстури шкіри, відблиски, природні ландшафти та архітектурні сцени. Рівень деталізації настільки високий, що для неозброєного ока відмінність між реальним і синтетичним зображенням часто є непомітною.

Однак принцип формування таких зображень суттєво відрізняється від фізичної фотографії. Реальні знімки містять унікальний шум сенсора (PRNU), метадані EXIF та фізично обґрунтовані тіні й відблиски. ШІ-зображення таких природних слідів не мають або імітують їх лише приблизно. Вони генеруються з випадкового шуму або латентного вектора, тому завжди несуть статистичні «відбитки» моделі. Згідно з оглядом Verdoliva (2020), навіть найсучасніші генератори залишають артефакти: асиметрію очей, нереалістичні зуби, неправильні переходи текстур або спектральні аномалії в частотному домені.

Основними методами генерації сьогодні є Generative Adversarial Networks (GAN) та Diffusion Models. GAN працюють за принципом змагання двох мереж: генератор створює зображення, а дискримінатор намагається його розпізнати. Diffusion Models діють інакше: спочатку додають шум до зображення, а потім навчаються його прибирати крок за кроком. Саме вони лежать в основі Stable Diffusion, DALL-E та Midjourney і сьогодні дають найкращу якість та гнучкість. Умовна генерація дозволяє створювати зображення за текстовим описом, що робить контент ще більш різноманітним і небезпечним для виявлення.

Існуючі підходи до виявлення можна розділити на пасивні (форензичні) та активні (на основі машинного навчання). Пасивні методи (PRNU, ELA, частотний аналіз) шукають природні сліди камери, але сучасні генератори їх добре імітують. Активні методи використовують згорткові мережі (ResNet, EfficientNet), проте аналізують лише просторові ознаки і не враховують частотні артефакти. Epstein et al. (2023) та Cozzolino et al. (2024) підтверджують, що універсального рішення поки немає, а головною проблемою є слабка генералізація на нові моделі генерації.

Для подолання зазначених обмежень у даній роботі запропоновано гібридний алгоритм виявлення ШІ-згенерованих зображень, що поєднує аналіз просторових та частотних ознак з механізмом уваги. Алгоритм базується на двох паралельних гілках обробки: просторова гілка використовує попередньо натреновану мережу EfficientNet-B0 для витягнення візуальних ознак (краї, текстури, форми), а частотна гілка обчислює дискретне косинусне перетворення (DCT) та витягає спектральні артефакти через окрему згорткову підмережу. Об'єднання ознак здійснюється через механізм Attention Fusion, який динамічно зважує внесок кожної гілки залежно від вхідного зображення. Це є ключовою перевагою над існуючими методами: замість фіксованого об'єднання модель

адаптивно визначає, чи просторові, чи частотні ознаки є більш інформативними для конкретного зразка. Навчання реалізовано зі стратегією transfer learning (заморожування backbone з наступним fine-tuning), LR warmup, cosine annealing та early stopping. Експериментальна оцінка на датасеті GenImage (зображення від 10+ генераторів: Stable Diffusion, GLIDE, BigGAN, StyleGAN3, VQ-Diffusion та ін.) показала accuracy 94.3%, F1-score 0.929, precision 0.907, recall 0.951 та ROC-AUC 0.976, що підтверджує високу ефективність гібридного підходу та його здатність надійно виявляти зображення від різноманітних генеративних моделей.

1. Verdoliva L. Media Forensics and DeepFakes: an overview. arXiv, 2020. URL: https://arxiv.org/pdf/2001.06564
2. Rafique R., Gantassi R., Amin R. та ін. Deep fake detection and classification using error-level analysis and deep learning. Scientific Reports (Nature), 2023. URL: https://www.nature.com/articles/s41598-023-34629-3
3. Epstein D.C., Jain I., Wang O., Zhang R. Online Detection of AI-Generated Images. ICCV Workshop, 2023. URL:https://openaccess.thecvf.com/content/ICCV2023W/DFAD/papers/Epstein_Online_Detection_of_AI-Generated_Images__ICCVW_2023_paper.pdf
4. Cozzolino D., Poggi G., Corvi R., Nießner M., Verdoliva L. Raising the Bar of AI-generated Image Detection with CLIP. CVPR Workshop, 2024. URL:https://openaccess.thecvf.com/content/CVPR2024W/WMF/papers/Cozzolino_Raising_the_Bar_of_AI-generated_Image_Detection_with_CLIP_CVPRW_2024_paper.pdf

Modern data hiding techniques: adaptivity, artificial intelligence and content synthesis

UDC 004.056:004.8 (043.2)

Artem Frolov[1], Vasyl Rizak[2]

Uzhhorod National University,
[1]artem.frolov@uzhnu.edu.ua, [2]vrizak@uzhnu.edu.ua

Modern steganography has evolved from simple data hiding in least significant bits (LSB) to complex methods that exploit machine learning and adaptive algorithms. The aim of this work is to analyze contemporary data hiding techniques that combine adaptive algorithms, neural network architectures and generative models, and to outline promising directions for further research.

1. Adaptive embedding based on distortion minimization. This is the “gold standard” of modern steganography: instead of embedding data uniformly, the algorithm analyses the content and identifies regions where modifications will be least detectable by steganalyzers. The mechanism relies on additive cost functions: each pixel is assigned a modification cost — pixels on object edges and in textured areas have low cost, while smooth surfaces (e.g. clear sky) have high cost. The key algorithms are: 1) S-UNIWARD, which operates in the spatial and frequency

domains; 2) HILL, which uses high-efficiency filters for cost estimation [1]. These methods remain the most difficult to detect with classical statistical analysis tools (SRM).

2. Generative adversarial networks (GAN-Steganography). This is a next-generation approach in which neural networks compete with one another: one tries to hide data, the other tries to detect it. The system consists of three parts: 1) an encoder that hides the data; 2) a decoder that extracts it; 3) a critic (discriminator) that tries to tell the "stego" apart from the original. During training the encoder learns to deceive the critic, producing visually flawless containers [2]. Unlike classical schemes, the message can be distributed across complex image features that are not described by simple mathematical formulas [3].

3. Coverless (generative) steganography. This method completely abandons modification of an existing file, which resolves the main problem of steganography — the presence of editing artefacts. The secret message is used as a seed for a generative model: the network synthesizes a realistic human face or a landscape, with the generation parameters (eye color, cloud shapes, etc.) themselves representing the encoded bits. Since no original container image ever existed, the analyzer has nothing to compare the result with.

4. Robust steganography for real-world channels (Robust Deep Stego). Most methods break down when the image is uploaded to social platforms (Facebook, Telegram, etc.), since these services compress the files (JPEG compression). During training of the neural network a "noise" layer is added that simulates JPEG compression, resizing or the addition of Gaussian noise. As a result, the network learns to hide data in those image components that are preserved even after aggressive processing by social network algorithms [4].

5. Linguistic steganography. This consists in using large language models (LLMs) to hide data in textual messages. While generating text (for instance, a chatbot response) the algorithm chooses between synonyms or alternative sentence constructions based on a secret key: choosing the word "automobile" instead of "car" may denote bit "0", and vice versa — bit "1" [5]. The resulting text appears entirely natural to a human reader, as it is produced by a modern language model that respects grammatical and stylistic norms [6].

Conclusions. The analysis of current trends in steganography allows us to highlight the following main directions: 1) a shift to intelligent embedding; 2) the dominance of neural network architectures; 3) a conceptual change of approach (coverless steganography); 4) adaptation to real-world transmission conditions; 5) multimodality of hiding. Modern steganography is becoming increasingly adaptive and context-aware: the main vector of development has shifted from the question "how to hide" to the question "how to make the intervention a natural part of a complex environment", where machine learning algorithms play the key role.

1. Holub V., Fridrich J., Denemark T. Universal distortion design for steganography in an arbitrary domain. EURASIP Journal on Information Security. 2014. Vol. 2014:1.
2. Volkhonskiy D., Nazarov I., Borisenko B., Burnaev E. Steganographic generative adversarial networks. Workshop on Adversarial Training,

Neural Information Processing Systems. 2016.
3. Tan S., Li B. Stacked convolutional auto-encoders for steganalysis of digital images. Asia-Pacific Signal and Information Processing Association, 2014 Annual Summit and Conference (APSIPA). IEEE, 2014. P. 1–4.
4. Zhu J., Kaplan R., Johnson J., Fei-Fei L. HiDDeN: Hiding Data with Deep Networks. Computer Vision – ECCV 2018. Springer, 2018.
5. Yang Z.-L., Guo X.-Q., Chen Z.-M., Huang Y.-F., Zhang Y.-J. RNN-Stega: Linguistic steganography based on recurrent neural networks. IEEE Transactions on Information Forensics and Security. 2019. Vol. 14, No. 5. P. 1280–1295.
6. Ma Y., Liu X., Bai S., Wang L.-Y., Liu A., Tao D., Hancock E. Region-wise generative adversarial image inpainting for large missing areas. arXiv preprint arXiv:1909.12507. 2019.

Enhancing facial verification in surveillance systems through super-resolution preprocessing and multi-model embedding concatenation

UDK 004.93 (004.8)

Denys Khanin[1], Viktor Otenko[2]

Lviv Polytechnic National University, [1]denys.o.khanin@lpnu.ua, [1]viktor.i.otenko@lpnu.ua

Facial verification is a critical component of modern security infrastructure, yet its effectiveness in surveillance deployments is limited by two challenges: low-resolution (LR) imagery from cameras [1] and the vulnerability of single-model verification architectures [2]. This paper proposes an integrated pipeline combining super-resolution (SR) preprocessing with multi-model concatenated embedding verification. The approach builds on the experimental work on concatenated embeddings [3, 4] and comparative analysis of SR methods [5].

Experiments on the CFP dataset [6] evaluated verification using embeddings from VGG-Face, Facenet, Facenet512, OpenFace, ArcFace, and SFace in concatenated configurations [3]. The best concatenated configurations consistently outperformed individual models across accuracy and EER. A normalization study [4] revealed that sequential Z-Score followed by L2 normalization is optimal for multi-model concatenation.

Table 1

Verification performance: single models vs. concatenated embeddings on CFP dataset

Configuration	Accuracy, %	EER, %
Facenet512 (best single)	97.45	3.40
Facenet + Facenet512	**97.68**	**2.87**
All 6 models	96.67	3.92

The best pair (Facenet + Facenet512) achieved 97.68% accuracy with EER of 2.87%, outperforming the best single model by 0.23 percentage points, while the all-model concatenation showed lower performance (96.67%) due to noise from weaker

models. A comparative SR study [5] showed Real-ESRGAN [7] excels at real-world degraded images, while FSRNet [8] offers optimal speed-accuracy balance for face-specific tasks.

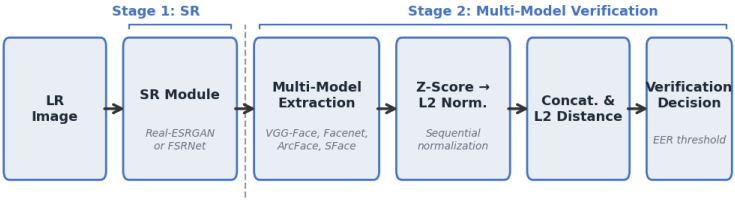


Fig.1. Architecture of the proposed SR-enhanced multi-model concatenated verification pipeline

The pipeline (Fig. 1) has two stages. Stage 1 enhances the LR image via SR (Real-ESRGAN or FSRNet). Stage 2 passes the enhanced image through multiple face recognition models; each embedding undergoes sequential normalization and concatenation before L2 distance verification against an EER-optimized threshold.

$$\hat{e}_i = \frac{(e_i - \mu_i)}{\sigma_i}, \quad then \quad \hat{e} = \frac{\hat{e}}{\|\hat{e}\|_2} \#(1)$$

where e_i is the raw embedding from model i, μ_i and σ_i are its mean and standard deviation, and $\|\cdot\|_2$ denotes the L2 norm. This ensures embeddings from heterogeneous models are harmonized before concatenation.

The pipeline is modular and each component can be updated independently. Multi-model architecture also provides security resilience, as compromising verification requires exploiting vulnerabilities across multiple models simultaneously. The primary trade-off is computational cost, acceptable for security-critical applications. Future work will focus on experimental validation on surveillance-specific datasets and real-time optimization.

1. Zou W.W., Yuen P.C. Very low resolution face recognition problem. IEEE Transactions on Image Processing. 2012. Vol. 21, No. 1. P. 327–340.
2. Goel R., Mehmood I., Ugail H. A Study of Deep Learning-Based Face Recognition Models for Sibling Identification. Sensors. 2021. Vol. 21. 5068.
3. Khanin D., Otenko V., Khoma V. Research on the effectiveness of concatenated embeddings in facial verification. CSDP-2024, Lviv Polytechnic National University, Lviv, Ukraine.
4. Khanin D. Comparative analysis of embedding normalization methods in face recognition systems. SCIFiC-2024, National Aerospace University "Kharkiv Aviation Institute".
5. Khanin D., Otenko V. Comparative analysis of super-resolution methods for improving face recognition accuracy. Computer Systems and Networks. 2025. Vol. 7, No. 1.
6. Sengupta S., Cheng J.C., Castillo C.D. et al. Frontal to Profile Face

Verification in the Wild. IEEE WACV. 2016.
7. Wang X., Xie L., Dong C., Shan Y. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. ICCV Workshops. 2021. P. 1905–1914.
8. Chen Y., Tai Y., Liu X. et al. FSRNet: End-to-End Learning Face Super-Resolution with Facial Priors. CVPR. 2018. P. 2492–2501.

Моделювання мережевих атак на основі аналізу графу мережевих взаємодій

УДК 004.056.55

Дмитро Хіжняк[1], Геннадій Шаповалов[2]

Національний університет «Одеська політехніка»,
[1]9480560@stud.op.edu.ua, [2]shapovalov@op.edu.ua

Сучасні комп'ютерні мережі функціонують в умовах постійного зростання кількості кіберзагроз. Особливо поширеними є атаки типу port scan, DoS/DDoS та brute force, які спрямовані на порушення доступності сервісів або отримання несанкціонованого доступу до систем [1]. Традиційні методи виявлення атак, засновані на сигнатурах, мають обмежену ефективність, особливо щодо нових або модифікованих типів атак [2].

Одним із перспективних підходів є використання графового аналізу, який дозволяє представити мережевий трафік у вигляді орієнтованого графа, де вузли відповідають IP-адресам або хостам, а ребра — мережевим взаємодіям між ними [3]. Такий підхід дає можливість враховувати структуру взаємозв'язків у мережі та виявляти аномальні патерни поведінки.

Метою роботи є моделювання мережевих атак **з** використанням графового аналізу та розробка застосунку для виявляти у мережі аномальних патернів поведінки.

Для оцінки поведінки вузлів використовується інтегральний показник аномальності, що формується на основі нормалізованих графових метрик, зокрема ступеня вершини, інтенсивності трафіку та кількості унікальних з'єднань:

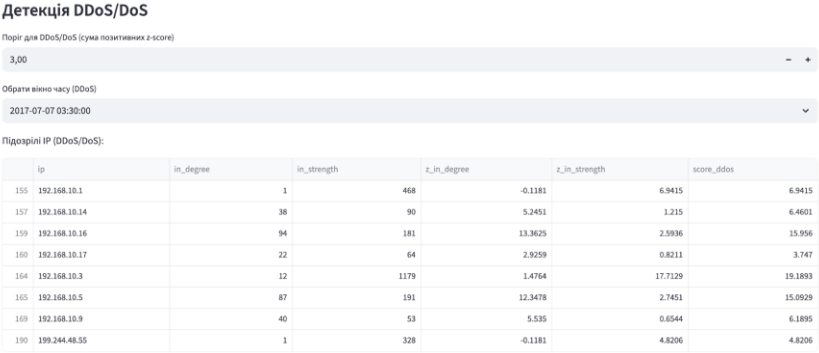
$$z(x_t) = \frac{x - \mu_t}{\sigma_t},$$

де x_t - значення метрики в момент часу t (або у вікні Δt), μ_t та σ_t - середнє і стандартне відхилення метрики у базовому періоді (наприклад, у попередніх k вікнах).

За підходом, що використано в роботі, аналіз структури графа дозволяє виділити характерні ознаки можливих атак. Для port scan типовим є сценарій «один до багатьох», для DDoS — «багато до одного», тоді як brute force характеризується інтенсивними повторюваними з'єднаннями між обмеженою кількістю досліджуваних мережевих вузлів.

Реалізація запропонованого підходу виконана у вигляді програмного застосунку, що здійснює обробку flow-даних, побудову графа мережевих

взаємодій та обчислення відповідних метрик. Для експериментальної перевірки використано набір даних CICIDS2017 [4].



Детекція DDoS/DoS

Поріг для DDoS/DoS (сума позитивних z-score)

3,00

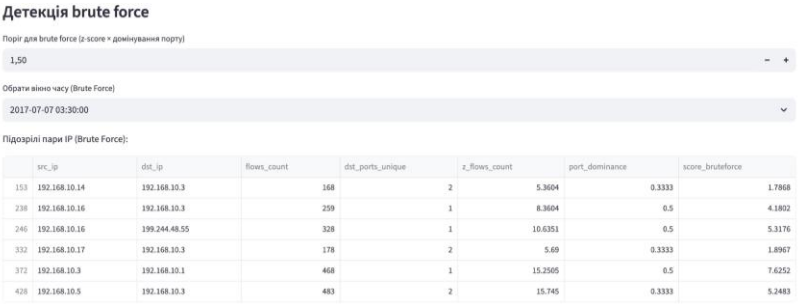
Обрати вікно часу (DDoS)

2017-07-07 03:30:00

Підозрілі IP (DDoS/DoS):

	ip	in_degree	in_strength	z_in_degree	z_in_strength	score_ddos
155	192.168.10.1	1	468	-0.1181	6.9415	6.9415
157	192.168.10.14	38	90	5.2451	1.215	6.4601
159	192.168.10.16	94	181	13.3625	2.5936	15.956
160	192.168.10.17	22	64	2.9259	0.8211	3.747
164	192.168.10.3	12	1179	1.4764	17.7129	19.1893
165	192.168.10.5	87	191	12.3478	2.7451	15.0929
169	192.168.10.9	40	53	5.535	0.6544	6.1895
190	199.244.48.55	1	328	-0.1181	4.8206	4.8206

Рис.1. Результати виявлення підозрілих вузлів для DoS/DDoS у вибраному часовому вікні



Детекція brute force

Поріг для brute force (z-score × домінування порту)

1,50

Обрати вікно часу (Brute Force)

2017-07-07 03:30:00

Підозрілі пари IP (Brute Force):

	src_ip	dst_ip	flows_count	dst_ports_unique	z_flows_count	port_dominance	score_bruteforce
153	192.168.10.14	192.168.10.3	168	2	5.3604	0.3333	1.7868
238	192.168.10.16	192.168.10.3	259	1	8.3604	0.5	4.1802
246	192.168.10.16	199.244.48.55	328	1	10.6351	0.5	5.3176
332	192.168.10.17	192.168.10.3	178	2	5.69	0.3333	1.8967
372	192.168.10.3	192.168.10.1	468	1	15.2505	0.5	7.6252
428	192.168.10.5	192.168.10.3	483	2	15.745	0.3333	5.2483

Рис.2. Результати виявлення brute force для підозрілої пари вузлів

Аналіз адекватності у порівнянні з експериментальними даними свідчить про те, що отримані результати підтверджують доцільність використання графових характеристик, що дозволяє ефективно виявляти типові мережеві атаки та їх поведінкові патерни.

1. Lippmann R. et al. The 1999 DARPA Off-Line Intrusion Detection Evaluation // Computer Networks. — 2000.
2. Sommer R., Paxson V. Outside the Closed World: On Using Machine Learning for Network Intrusion Detection // IEEE Symposium on Security and Privacy. — 2010.
3. Newman M. Networks: An Introduction. — Oxford University Press, 2010.
4. Sharafaldin I., Lashkari A. H., Ghorbani A. A. Toward Generating a New Intrusion Detection Dataset // ICISSP. — 2018.

Developing a secure virtual physical laboratory: addressing VR vulnerabilities in educational environments

UDK 37:004:316.772.5

Yuriy Skorenkyy[1], Oleksandr Parayil[2], Oleksandr Kramar[3]

Ternopil Ivan Puluj National Technical University,
[1]skorenkyy@tntu.edu.ua, [2]sashaparail@gmail.com, [2]kramar_o@tntu.edu.ua

The integration of digital technologies presents new opportunities for organizing the educational process, particularly within the natural sciences. A virtual physical laboratory that combines realistic simulations with interactive tasks is currently under development by the Cyber-Physical Systems Laboratory at TNTU.

While virtual reality (VR) environments offer immense educational benefits, their software implementation introduces significant cybersecurity vectors [1]. Educational applications built on industry-standard platforms like Unity and Unreal Engine face distinct vulnerabilities. Unity applications, for instance, are notoriously susceptible to reverse engineering; without proper obfuscation, attackers can decompile C# assemblies to manipulate core logic or alter Asset Bundles to inject malicious content. Unreal Engine applications similarly risk memory injection attacks and the unauthorized modification of unencrypted Blueprint visual scripting logic. In the specific context of a virtual lab, these engine-specific vulnerabilities directly enable the manipulation of non-player character (NPC) logic — allowing an attacker to alter a virtual assistant's scripts to provide incorrect experimental instructions. Furthermore, insecure network configurations inherent to many standard engine plugins pose a severe risk of data interception, threatening unauthorized access to students' personal information. To mitigate these threats, the application's architecture must be fortified. The proposed solution involves implementing a trusted update mechanism that automatically synchronizes with a secure repository, preventing the execution of tampered assets. Additionally, the integration of strict digital certificate verification guarantees the authenticity of the connection between the client application and the laboratory server.

Deploying virtual laboratories with advanced, secure architectures fosters inclusive education. It ensures equal access to high-quality educational resources, allowing students to master experimental skills in a safe, convenient, and fully protected digital environment [2].

1. 1. Kozak R., Skorenkyy Yu., Kramar O., Brevus V., Zagorodna N., Cybersecurity issues related to incorporation of VR components into Industry 5.0 human-machine interfaces. *Procedia Computer Science*. – 2026. – V. 276. – p. 176-184.
2. 3. Zagorodna N., Skorenkyy Y., Kunanets N., Baran I., Stadnyk M., Augmented Reality-enhanced learning tools development for cybersecurity major. *CEUR Workshop Proceedings.* – 2022. – V. 3309. – p. 25–32.

Реалізація алгоритму недвійкових первинних кодів за допомогою Google Sheets

УДК 004.056

Діана Желізняк[1], Наталія Загородна[2], Кирил Шеханін[3]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]dianajeliznyk@gmail.com, [2]zagorodna_n@tntu.edu.ua,
[3]kyryl.shekhanin@outlook.com

У сучасних системах передачі та зберігання інформації ключову роль відіграють методи кодування, які забезпечують ефективність, завадостійкість та оптимальне використання наявних ресурсів. Традиційно найбільш поширеними є двійкові коди, що базуються на алфавіті з двох символів – «0» та «1». Однак у багатьох випадках застосування недвійкових первинних кодів, які використовують алфавіт потужністю більше двох ($q > 2$), відкриває додаткові можливості для підвищення інформаційної щільності та швидкості обробки даних. Недвійкові первинні коди є системами кодування, де кожен символ може нести більше одного біта інформації, що дозволяє скоротити довжину кодових комбінацій при передаванні того ж обсягу повідомлень.

Теоретичні засади недвійкового кодування базуються на потужності алфавіту q, довжині кодової комбінації n та загальній кількості можливих комбінацій, яка обчислюється за формулою $N_0 = q^n$. Наприклад, для трійкового коду ($q = 3$) або четвіркового ($q = 4$) кількість доступних кодових слів зростає експоненційно порівняно з двійковим представленням тієї ж довжини. Важливою характеристикою є кодова відстань, зокрема відстань Хеммінга, яка визначає здатність коду виявляти та виправляти помилки, що виникають під час передачі даних через канали зв'язку із завадами. Недвійкові первинні коди класифікуються за можливістю виявлення помилок (безнадмірні та надмірні), за структурою (блокові та неперервні), а також за рівномірністю (рівномірні та нерівномірні). Історично передумови виникнення таких кодів сягають ще XIX століття, коли Семюель Морз створив телеграфний код, що використовував три символи (крапку, тире та паузу), а згодом Еміль Бодо розробив 5-бітний код для телеграфії, який дозволяв передавати 32 різні символи.

Галузь використання недвійкових первинних кодів досить широка. Вони активно застосовуються в сучасних телекомунікаційних системах, зокрема в мобільному зв'язку стандартів 4G/5G, де використовуються багаторівневі схеми квадратурної амплітудної модуляції (QAM) з базами 4, 16, 64 та вище. У комп'ютерній техніці недвійкове кодування реалізоване у флеш-пам'яті типу MLC (Multi-Level Cell), де одна комірка зберігає два або більше бітів інформації. У квантових обчисленнях кубіти можуть перебувати в суперпозиції станів, що також потребує недвійкового представлення даних. Крім того, такі коди знаходять застосування в оптичних лініях зв'язку, системах супутникового зв'язку, а також у задачах штучного інтелекту для оптимізації обробки великих масивів даних. Переваги недвійкового кодування полягають у підвищенні інформаційної ємності, зменшенні апаратних витрат при зберіганні даних та можливості створення більш гнучких алгоритмів корекції помилок. Однак існують і недоліки: складність реалізації, підвищена чутливість до шумів у

каналі зв'язку, потреба в точнішій апаратурі для розрізнення багатьох рівнів сигналу, а також труднощі із синхронізацією на високих швидкостях передачі.

Реалізація алгоритму недвійкових первинних кодів може бути легко автоматизована за допомогою табличних процесорів, таких як Google Sheets або Microsoft Excel, де використовуються вбудовані функції BIN2DEC для перетворення двійкових груп у десяткові числа та DEC2BIN для зворотного перетворення. Для практичної демонстрації алгоритму кодування та декодування недвійкових первинних кодів було обрано Google Sheets (рис. 1) та систему числення з основою q=16 (шістнадцяткове кодування), оскільки вона є поширеним компромісом між інформаційною щільністю та простотою технічної реалізації. Як вхідні дані використано довільну 16-бітову двійкову послідовність: 1101011001110101. Процес кодування передбачає розбиття вихідної бітової стрічки на групи по 4 біти, оскільки для бази q=16 один символ несе 4 біти інформації ($\log_2 16 = 4$). В результаті отримано чотири групи: 1101, 0110, 0111, 0101. Кожну таку групу перетворено у десяткове число за правилами двійково-десяткового перетворення: $1101_2 = 13_{10}$, $0110_2 = 6_{10}$, $0111_2 = 7_{10}$, $0101_2 = 5_{10}$. Таким чином, вихідне 16-бітне повідомлення було представлене у вигляді кодового слова, що складається з чотирьох символів шістнадцяткової системи: 13, 6, 7, 5. Це кодове слово є більш компактним порівняно з двійковим оригіналом, оскільки замість 16 двійкових символів використовується лише 4 символи з алфавіту потужністю 16. На етапі декодування виконується зворотне перетворення. Для кожного символу кодового слова (13, 6, 7, 5) за допомогою відповідної таблиці відповідності або обчислювальної формули відновлюється його 4-бітне двійкове представлення: 13 → 1101, 6 → 0110, 7 → 0111, 5 → 0101. Отримані бітові групи конкатенуються у вихідну послідовність 1101011001110101, що повністю збігається з початковими даними, підтверджуючи коректність роботи алгоритму.

Вхідні дані	Групи бітів	Символ у базі q=16
1101011001110101	1101	13
	0110	6
	0111	7
	0101	5

Рис. 1. Приклад алгоритму недвійкових первинних кодів завдяки Google Sheets

Такий підхід забезпечує наочність і дозволяє швидко обробляти великі обсяги даних без потреби у спеціалізованому програмному забезпеченні. Узагальнюючи, недвійкові первинні коди, зокрема на базі системи числення з основою 16, є потужним інструментом для підвищення ефективності передавання та зберігання інформації, а їх практична реалізація не потребує надмірних обчислювальних ресурсів.

Подальші дослідження можуть бути спрямовані на порівняння завадостійкості недвійкових кодів з двійковими в каналах з різним рівнем шуму, а також на розробку адаптивних алгоритмів кодування, які динамічно змінюють основу системи числення залежно від умов передачі.

Виявлення мережевих атак засобами машинного та глибокого навчання на основі набору даних UNSW-NB15

УДК 004.056.5:004.85

Марина Ксеніта[1], Марія Стадник[2], Володимир Данилюк[3]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]ksenita.marina@gmail.com, [2]maria.stadnyk@gmail.com, [3]vdanilyuk06@gmail.com

Зростання кількості кібератак, ускладнення мережевої інфраструктури та активне використання автоматизованих засобів сканування, експлуатації вразливостей і приховування шкідливої активності зумовлюють потребу в ефективних системах виявлення вторгнень. Традиційні сигнатурні механізми залишаються корисними для відомих загроз, однак вони недостатньо гнучкі для виявлення нових або модифікованих атак. Тому актуальним є застосування методів машинного та глибокого навчання, здатних аналізувати багатовимірні характеристики мережевих потоків і класифікувати трафік як нормальний або шкідливий.

Метою дослідження є розроблення та експериментальна перевірка програмного підходу до бінарної класифікації мережевого трафіку на основі набору даних UNSW-NB15 із використанням алгоритму XGBoost та рекурентних нейронних мереж RNN, LSTM і GRU. Завдання дослідження охоплюють попередню обробку даних, кодування категоріальних ознак, нормалізацію числових характеристик, балансування навчальної вибірки, побудову послідовностей фіксованої довжини для нейронних моделей та порівняння результатів.

Вхідними файлами є навчальна та тестова частини UNSW_NB15_training-set.csv і UNSW_NB15_testing-set.csv. На етапі попередньої обробки вилучаються поля id та attack_cat, категоріальні ознаки proto, service і state перетворюються за допомогою LabelEncoder, пропущені значення видаляються, а всі ознаки масштабуються методом MinMaxScaler. Для зменшення впливу дисбалансу класів навчальна вибірка додатково балансується алгоритмом ADASYN.

Першим базовим класифікатором обрано XGBoost - ансамблевий метод градієнтного бустингу дерев рішень, який поєднує високу точність, регуляризацію та ефективну роботу з табличними даними. Для перевірки стабільності моделі застосовано стратифіковану п'ятикратну крос-валідацію, після чого модель навчається на збалансованій вибірці та тестується на відкладеному наборі. Другий блок експерименту формують нейронні архітектури SimpleRNN, LSTM і GRU, які отримують на вхід послідовності з 10 записів і використовують сигмоїдний вихідний шар для бінарної класифікації. Таке подання дає змогу оцінити, наскільки рекурентні моделі здатні враховувати локальні залежності між послідовними мережевими спостереженнями.

Набір UNSW-NB15 створено в Cyber Range Lab UNSW Canberra із використанням IXIA PerfectStorm для формування поєднання сучасної нормальної активності та синтетичних атак; він містить дев'ять типів атак:

Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode і Worms. У межах поданого програмного рішення ці категорії агрегуються до бінарної цільової змінної label, що відповідає практичній задачі первинного виявлення факту атаки. Такий підхід є доцільним для початкового рівня системи NIDS, коли найважливішим є швидке відокремлення потенційно шкідливого трафіку від нормального.

Аналіз ROC-кривих показав (рис. 1) високу ефективність усіх досліджуваних підходів. Найкращий результат продемонструвала модель XGBoost із значенням ROC-AUC ≈ 0,98, що свідчить про її високу здатність розрізняти нормальний мережевий трафік та атаки. Рекурентні нейронні мережі також показали високі результати: RNN досягла ROC-AUC ≈ 0,97, а моделі LSTM та GRU – близько 0,96.

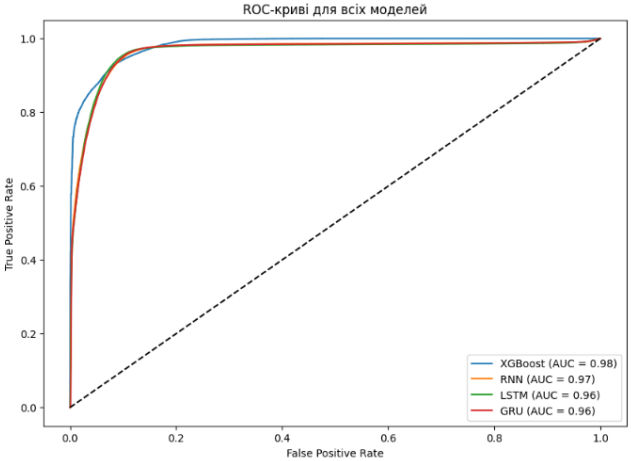


Рис. 1. ROC криві досліджуваних моделей

Отримані значення Precision = 0,9646, Recall = 0,8986, F1-score = 0,9304 та ROC-AUC = 0,9637 підтверджують високу якість класифікації та збалансованість моделі щодо виявлення атак і мінімізації хибних спрацювань. Водночас XGBoost продемонстрував дещо кращі результати порівняно з нейронними мережами при меншій обчислювальній складності.

Проведене дослідження показує, що поєднання класичних ансамблевих алгоритмів і рекурентних нейронних мереж є перспективним напрямом для побудови систем виявлення вторгнень. XGBoost доцільно використовувати як сильний базовий класифікатор для табличних ознак мережевих потоків, тоді як LSTM і GRU можуть бути корисними для моделювання послідовного контексту трафіку. Водночас якість висновків залежить від коректності кодування категоріальних ознак, репрезентативності тестової вибірки, параметрів балансування та кількості епох навчання нейронних мереж. Перспективами подальших досліджень є оптимізація гіперпараметрів, використання attention-механізмів, побудова багатокласової класифікації за типами атак і перевірка моделей на реальному потоковому трафіку.

Захищений клієнт-серверний застосунок OffGrid із E2E-шифруванням і контрольованим файлообміном

УДК 004.056.55

Катерина Холодова[1]

Національний університет «Одеська політехніка», [1]9480541@stud.op.edu.ua

Системи миттєвого обміну повідомленнями є одним із ключових каналів передавання персональних і службових даних. Навіть за наявності наскрізного шифрування зберігаються ризики витоку метаданих, небезпечного обміну вкладеннями, підміни ключів і компрометації кінцевих пристроїв [1]. Тому актуальним є проєктування месенджера, у якому сервер не вважається повністю довіреним, а критичні криптографічні операції виконуються на клієнті.

Метою роботи є розроблення клієнт-серверного застосунку OffGrid для захищеного обміну повідомленнями та файлами в умовах недовіреної серверної інфраструктури. Для цього сформульовано модель загроз, спроєктовано архітектуру «тонкого ретранслятора», реалізовано механізми керування ключами і проведено експериментальну перевірку. Науково-практична новизна полягає у поєднанні E2E-шифрування, двоконтурного передавання вкладень, локального контролю довіри до пристроїв і очищення чутливих даних після secure-сесій.

Архітектура OffGrid складається з клієнтського застосунку, TCP-сервера, HTTP-файлового контуру та підсистем зберігання. TCP-канал використовується для автентифікації, маршрутизації подій і видачі короткоживучих токенів доступу, а передавання великих вкладень винесено в окремий HTTP-контур. Сервер зберігає лише шифротекст і мінімально необхідні метадані, тоді як приватні ключі, стани сесій і операції шифрування залишаються на клієнті.

Керування ключами організовано на рівні пристроїв. Кожен клієнт має стабільний device_id і власні пари ключів: Ed25519 використовується для цифрового підпису, а X25519 - для узгодження спільного секрету. Автентичність ключів співрозмовника контролюється через TOFU з локальним закріпленням: зміна ключового матеріалу для відомого пристрою блокує взаємодію до явного підтвердження користувачем.

Для отримання ключового матеріалу з узгодженого секрету використано HKDF, що забезпечує доменне розділення ключів між різними протокольними задачами [2]:

$$K = HKDF(z; salt, info, 32)\,, \tag{1}$$

де K – майстер-ключ; z – спільний секрет, отриманий через X25519;

salt – сіль (за потреби); info – параметри контексту й доменного розділення; 32 – довжина ключа в байтах.

Для шифрування E2E-повідомлень застосовано ChaCha20-Poly1305 як AEAD-примітив:

$$\begin{aligned}(C, T) &= AEAD\_Enc(K, N, AAD, P),\\ P &= AEAD\_Dec(K, N, AAD, C, T),\end{aligned} \tag{2}$$

де C – шифротекст; T – тег автентичності; K – симетричний ключ; N – одноразове випадкове значення (нонс); AAD – додаткові автентифіковані дані; P – відкритий текст.

Додаткові автентифіковані дані прив'язують шифротекст до контексту сесії, епохи, напряму передавання та ідентифікатора повідомлення; перенесення шифротексту в інший контекст призводить до помилки автентифікації [3].

У файловій підсистемі використано модель «ticket → bearer-token → HTTP». Клієнт через TCP-канал запитує дозвіл, сервер виконує ACL-перевірку і видає підписаний bearer-token з обмеженим строком дії. Далі клієнт завантажує або скачує файл через HTTP, передаючи токен у заголовку Authorization. Файловий сервер перевіряє підпис, часові межі й тип операції, але не має доступу до відкритого вмісту, оскільки файл шифрується на клієнті до передавання.

Захист облікових і користувацьких секретів реалізовано через bcrypt для паролів та PBKDF2-HMAC-SHA256 для секретів відновлення і локальних механізмів доступу. Приватні ключі й токени інтеграцій зберігаються через системні сховища, зокрема keyring або DPAPI, що зменшує ризик їх потрапляння у відкриті конфігураційні файли.

Програмну реалізацію виконано на Python із розділенням відповідальності між сервером, мережевим шаром, графічним інтерфейсом і криптографічними модулями. Така декомпозиція дозволила окремо перевірити маршрутизацію, роботу сесій, доступ до вкладень і локальні процедури захисту секретів.

Експериментальна верифікація підтвердила ключові інваріанти безпеки. Тест зміни identity-ключа показав, що TOFU/pinning виявляє зміну ключового матеріалу для відомого device_id і блокує взаємодію до відновлення довіри. Негативне тестування авторизації засвідчило, що сторонній користувач не може виконувати операції з приватними групами та пов'язаними вкладеннями, а сервер не розкриває факт існування приватного ресурсу. Тест очищення RAM у secure-режимі підтвердив очищення внутрішнього secure-сховища і занулення змінюваних буферів.

Отримані результати показують, що запропонована архітектура забезпечує конфіденційність і цілісність E2E-вмісту, контроль автентичності ключів і розмежування доступу до вкладень без надання серверу відкритих даних або приватного ключового матеріалу. Рішення може бути використане як основа для систем обміну чутливою інформацією, де важливими є мінімізація довіри до інфраструктури, контроль метаданих і захист кінцевої точки.

1. Alatawi M., Saxena N. Sok: An analysis of end-to-end encryption and authentication ceremonies in secure messaging systems. Proceedings of the 16th ACM Conference on Security and Privacy in Wireless and Mobile Networks. 2023.
2. Serrano R. et al. ChaCha20–Poly1305 authenticated encryption with additional data for transport layer security 1.3. Cryptography. 2022. Vol. 6, No. 2. P. 30.
3. Onik A. R. et al. A systematic literature review of secure instant messaging applications from a digital forensics perspective. ACM Computing Surveys. 2025. Vol. 57, No. 9. P. 1–36.

Концептуальна модель проєктування CTF-завдань та методика її застосування для формування компетентностей з мережевої безпеки

УДК 621.395.7 (043.2)

Олександр Черепов[1], Богдан Неймет[2]

Ужгородський національний університет,
[1]oleksandr.cherepov@uzhnu.edu.ua, [2]bohdan.neimet@uzhnu.edu.ua

Війна змусила переоцінити, наскільки готові українські випускники зі сфери інформаційної безпеки до викликів сучасності. CERT-UA фіксує нові інциденти щодня; класичні лабораторні роботи в університеті за ними не встигають. Capture the Flag — формат, який давно зарекомендував себе як ефективний для практичної підготовки [1, 2]. Біда в іншому. У ЗВО CTF-задачі здебільшого з'являються як ініціатива окремого викладача. Вони робляться під конкретну лекцію або набір практичних задач, без явного зв'язку з компетентностями, що має закрити освітня програма. У свою чергу можливість скласти послідовний курс CTF орієнтованих задач, що можна було б відтворювати в іншому університеті чи навіть у наступному році, виявляється важкою задачею.

Мета роботи — побудувати модель проєктування CTF-задач з мережевої безпеки, яка прив'язує таксономію MITRE ATT&CK [1] до фахових компетентностей чинного стандарту вищої освіти спеціальності 125 (F5) «Кібербезпека та захист інформації» та до міжнародних рамок ENISA ECSF [3] і NIST NICE; а також запропонувати методику, як цю модель використовувати на практиці.

Серед наявних рішень виокремлюються три гілки. Платформи на кшталт CTFd чи kCTF дають інфраструктуру для розгортання задач — і нічого більше. Кіберполігони з декларативними мовами опису сценаріїв (KYPO [2], CyRIS, CRACK з мовою VSDL [4]) теж зосереджені на розгортанні. Фреймворк URSID [5], свіжіший і дотичніший за духом, переводить технічний сценарій ATT&CK у множину процедурних варіантів. Проте і він не передбачає зв'язку з освітніми компетентностями. Узагальнені моделі знань (наприклад, A4CKGE [6]) служать радше для аналізу багатокрокових атак, ніж для проєктування навчальних задач.

Запропонована модель забезпечує те, чого бракувало: можливість простежити від рядка освітньої програми до конкретної техніки ATT&CK, з якою працює студент за клавіатурою. Саме у цьому ми вбачаємо наукову новизну, що відрізняє наш підхід від наявних CTF-платформ і від додаткових практик окремих викладачів.

Модель складається з чотирьох шарів. Компетентнісний шар виконує роль точки входу: проєктувальник обирає фахову компетентність з освітньої програми, та додатково зіставляє її з ECSF або NICE, потім розкладає на елементи KSA і фіксує очікуваний рівень за НРК. Онтологічний шар працює нижче: тут MITRE ATT&CK, CWE і CAPEC зведено у спільну мережу знань. Для кожної компетентності у цій мережі видно, які техніки атак вона має закривати і які слабкості становлять її предмет. Сценарний шар бере зв'язки з онтології і перетворює їх у шаблон CTF-задачі — з точками альтернатив (де технік кілька, обираємо одну) і точками параметризації (де налаштування

варіюються між варіантами). Останній шар, інфраструктурний, відповідає за матеріальне втілення: Docker, Vagrant, Terraform або інший фреймворк. Перехід між шарами реалізовано як типізовані функції відображення; саме завдяки їм з'являється той зв'язок «компетентність — техніка — задача», заради якого все це вибудовується.

Методика застосування моделі складається з п'яти послідовних кроків: 1) декомпозиція цільової фахової компетентності з освітньої програми на KSA-елементи з урахуванням рівня НРК; 2) картування KSA на множину технік MITRE ATT&CK; 3) вибір сценарного шаблону з бібліотеки або проєктування нового; 4) генерація варіантів задачі за допомогою параметризації вхідних точок; 5) оцінювання навчальних результатів, а саме: успішність, час виконання, кількість спроб.

Модель було перевірено на групах студентів ДВНЗ «Ужгородський національний університет» освітньої програми «Кібербезпека та захист інформації». Цільова компетентність - здатність виконувати тестування на проникнення (близька за змістом до ECSF Penetration Tester). Акцент був зосереджений саме на активній розвідці. Зіставлення з ATT&CK дало техніки T1595, T1046, T1018 і T1083. За шаблон взяли трирівневу мережу: DMZ, внутрішній і серверний сегменти. У внутрішньому сегменті було прихований цільовий ресурс, доступ до якого можливий лише через проміжний хост. Альтернативні точки у шаблоні (Apache або Nginx, MySQL або PostgreSQL) дозволили породити більше 10 дидактично рівноцінних варіантів задачі. Кожен зі студентів отримав свій варіант. Результатом є те, що час підготовки набору задач скоротився приблизно вдвічі порівняно з ручним проєктуванням, а можливість обміну розв'язками між студентами мінімізувалась через варіативність.

Висновки. Запропонована модель і методика дають викладачеві інструмент, що допомагає зв'язати рядки фахових компетентностей освітньої програми «Кібербезпека та захист інформації» (F5) з конкретними технічними задачами, які студент розв'язує у віртуальному середовищі. Це робить процес осяжним, а навчальний результат повторюваним між дисциплінами та між закладами. Подальша робота включає у себе формальний розвиток сценарного шару (композиційна алгебра і метрика складності варіантів), формування корпусу TTP за матеріалами CERT-UA та відкритих звітів про кіберінциденти проти інформаційної інфраструктури України 2022–2026 років, та експериментальну верифікацію методики у форматі контрольної й експериментальної груп серед студентів освітньої програми.

1. Strom B. E., Applebaum A., Miller D. P. та ін. MITRE ATT&CK: Design and philosophy. MITRE Technical Report MP180360. Bedford, MA : The MITRE Corporation, 2020. p. URL : https://attack.mitre.org/docs/ATTACK_Design_and_Philosophy_March_2020.pdf.
2. Vykopal J., Vizváry M., Ošlejšek R. та ін. Lessons learned from complex hands-on defence exercises in a cyber range. Proceedings of the 2017 IEEE Frontiers in Education Conference (FIE). – Indianapolis : IEEE, 2017. – P.

1–8. DOI: 10.1109/FIE.2017.8190713.
3. European Union Agency for Cybersecurity (ENISA). European Cybersecurity Skills Framework (ECSF). – Athens : ENISA, 2022. URL: https://www.enisa.europa.eu/topics/skills-framework/ecsf.
4. Russo E., Costa G., Armando A. Building next generation cyber ranges with CRACK. Computers & Security. – 2020. – Vol. 95. – Article 101837. DOI: 10.1016/j.cose.2020.101837.
5. Besson P.-V., Viet Triem Tong V., Guette G. та ін. URSID: Automatically refining a single attack scenario into multiple cyber range architectures. Foundations and Practice of Security (FPS 2023). Lecture Notes in Computer Science. Cham : Springer, 2024. DOI: 10.1007/978-3-031-57537-2_8.
6. Xiang X., Ma C., Zeng L. та ін. Uncovering multi-step attacks with threat knowledge graph reasoning. Security and Safety. – 2025. – Vol. 4. – Article 2024019. DOI: 10.1051/sands/2024019.

Удосконалення процедур цифрової криміналістики в системах реагування на інциденти кібербезпеки

УДК 004.056.55

Мар'яна Мельник[1], Віктор Чешун[2], Дмитро Чешун[3]

[1,2]Хмельницький національний університет, [3]Хмельницький фаховий економіко-технологічний коледж Університету економіки і підприємництва
[1]melnyk.masia@gmail.com, [2]cheshunvn@khmnu.edu.ua,
[3]dmytro.cheshun@gmail.com

Сучасний етап розвитку цифрового суспільства характеризується тотальною інтеграцією інформаційних технологій у процеси державного управління та функціонування критичної інфраструктури, що водночас створює безпрецедентні ризики для національної безпеки.

Аналіз поточної ситуації [1-3] дозволяє виявити суттєві недоліки в існуючих підходах до реагування, серед яких найгострішими є низький рівень автоматизації моніторингу подій безпеки та відсутність єдиної методології збору цифрових доказів. Більшість установ сьогодні стикаються з проблемою фрагментарності журналів подій та невідповідністю процедур вилучення артефактів міжнародним стандартам, що часто унеможливлює проведення глибокого технічного аналізу та встановлення реальних причин інцидентів. Досвід масштабних атак останніх років підтверджує, що без впровадження проактивних методів виявлення та структурованих алгоритмів розслідування об'єкти критичної інфраструктури залишаються вразливими до тривалої прихованої присутності зловмисників у їхніх внутрішніх мережах.

Мета дослідження полягає у розробці та практичному обґрунтуванні системного підходу до розслідування кіберінцидентів, який би забезпечував цілісність процесу криміналістики від моменту первинної фіксації аномалії до формування підсумкової аналітичної звітності. Автори поставили за ціль створити гнучку модель, що поєднує сучасні технічні засоби автоматизованого аналізу телеметрії з чіткими організаційними регламентами взаємодії між

різними суб'єктами кібербезпеки. Важливим аспектом дослідження є адаптація вимог міжнародних стандартів ISO/IEC 27035 [4] та рекомендацій NIST [5,6] до специфічних умов функціонування українських державних інформаційних систем. Для досягнення поставленої мети розв'язано низку наукових і практичних завдань, зокрема моделювання процесів збереження цифрових артефактів, визначення оптимального стеку програмних інструментів із відкритим кодом та побудова тестового середовища для верифікації запропонованих підходів у реальних сценаріях складних кібератак.

Отримані результати дослідження базуються на розробці оригінального інтегрованого методу, що об'єднує кілька технологічних платформ у єдину логічну систему реагування та аналізу. Оригінальність полягає у формуванні двох взаємодоповнюючих моделей, де перша регламентує збереження та відновлення цифрових доказів за принципом суворого дотримання ланцюга зберігання, а друга визначає чіткі алгоритми взаємодії учасників розслідування на локальному, галузевому та національному рівнях.

Основними деталізованим етапами реалізації методу є реагування на інциденти, робота з доказами і взаємодія учасників. На етапі реагування впроваджено ієрархічний підхід, де кожен інцидент проходить фази ідентифікації, локалізації, аналізу та «навчання». Акцент зроблено на безперервності – результати розслідування обов'язково мають оновлювати бази індикаторів компрометації (IoC). Для роботи з доказами регламентовано порядок збору волатильних даних (RAM) та створення посекторних копій дисків; використання хеш-функцій (SHA-256) гарантує недоторканність доказів для можливого подальшого судового розгляду. Взаємодія передбачає 4-рівневу систему обміну інформацією (локальний–галузевий–національний–міжнародний), що дозволяє оперативно попереджати інші об'єкти про нові загрози через платформу MISP.

Практична реалізація методу була успішно здійснена з використанням стеку технологій та відповідних їм інструментів: Wazuh для моніторингу кінцевих точок та виявлення аномалій у реальному часі; ELK Stack (Elasticsearch, Logstash, Kibana) для централізованого збору та візуалізації терабайтів логів; TheHive для управління процесами розслідування інцидентів; MISP для автоматизованого оперативного обміну індикаторами загроз серед спільноти фахівців; криміналістичний софт включно з Autopsy (аналіз файлових систем), Volatility (аналіз оперативної пам'яті), Wireshark (мережева форензика).

Пропонований підхід може стати корисним для створення внутрішніх регламентів CSIRT-підрозділів, забезпечуючи швидке відновлення систем та недопущення повторних атак. Автоматизація через Wazuh дозволила скоротити час реакції на типові вектори атак (фішинг, несанкціонований доступ) на 30-40%. Алгоритм фіксації артефактів забезпечує повну реконструкцію дій атакуючого (dwell time, використані скрипти, канали витоку даних). Використання інструментів із відкритим кодом робить метод економічним і доступним для державних установ з обмеженим фінансуванням.

1. ENISA Threat Landscape 2025. European Union Agency for Cybersecurity (ENISA). 2025. URL: https://urli.info/1uHEw (date of access: 8.05.2026).

2. Річний звіт 2025: системи виявлення вразливостей і реагування на кіберінциденти та кібератаки. Оперативний центр реагування на кіберінциденти Державного центру кіберзахисту Державної служби спеціального зв'язку та захисту інформації України. URL: https://urli.info/1uHEt (дата звернення: 8.05.2026).
3. Огляд кіберзагроз та стратегій захисту в 2025 році: досвід CERT-UA. URL: https://urli.info/1pmEF (дата звернення: 8.05.2026).
4. ISO/IEC 27035-1:2023 Information technology – Information security incident management – Part 1: Principles and process. International Organization for Standardization. 2023. URL: https://urli.info/1uHEW (date of access: 12.12.2025).
5. NIST SP 800-61 Revision 2, Computer Security Incident Handling Guide. URL: https://urli.info/1uHEA (date of access: 8.05.2026).
6. NIST SP 800-61r3. Incident Response Recommendations and Considerations for Cybersecurity Risk Management – A CSF 2.0 Community Profile / Alex Nelson et al. National Institute of Standards and Technology. https://urli.info/1pmEm (date of access: 8.05.2026).

Огляд підходів використання DGA алгоритмів

УДК 004.056

Петро Венгерський[1], Юрій Шпак[2]

Львівський національний університет імені Івана Франка,
[1]petro.venherskyy@lnu.edu.ua, [2]Yurii.Shpak@lnu.edu.ua

Сучасні кібератаки дедалі частіше використовують алгоритми генерації доменів (Domain Generation Algorithms, DGA) для забезпечення стійкого зв’язку шкідливого програмного забезпечення з серверами управління (C&C) [1]. Традиційні методи захисту, зокрема «чорні списки» (blacklisting) або статичні IP-адреси, є малоефективними проти DGA, оскільки зловмисники можуть генерувати сотні або тисячі нових доменів щодня, роблячи блокування неактуальним ще до його впровадження.

У зв’язку зі швидкою еволюцією кіберзагроз, зокрема переходом від випадкових наборів символів до словникових DGA, глибоке розуміння того, як і де застосовуються ці алгоритми, є важливим для розробки проактивних систем виявлення індикаторів компрометації [1].

В роботі проводиться системний огляд основних підходів до використання DGA-алгоритмів, проаналізувати механізми їхньої роботи («як вони використовуються») та ідентифікувати ключові вектори їх застосування зловмисниками («де саме»)[2].

На відміну від більшості робіт, які фокусуються переважно на методах виявлення (за допомогою машинного навчання чи аналізі DNS-трафіку), дане дослідження систематизує самі підходи до використання DGA з точки зору архітектури атаки. Проводиться аналіз еволюції вибору початкового "зерна" (seed) та генерації доменів залежно від типу загрози, що дозволяє краще зрозуміти тактичні цілі зловмисників.

У ході огляду визначено основні підходи до того, як використовуються DGA:

- Псевдовипадкові генератори (PRNG): Використовують математичні функції та динамічне "зерно" (наприклад, поточну дату або публічні параметри) для створення великого обсягу доменів, які виглядають як випадковий набір літер та цифр.
- Словникові DGA (Dictionary-based): Комбінують легітимні слова з вбудованих словників, щоб імітувати звичайний трафік і обходити системи лексичного аналізу.
- Механізм відвернення уваги: Програма генерує тисячі DNS-запитів, але справжнім сервером управління виявляється лише один із них, що перевантажує системи моніторингу захисників.
- Щодо того, де саме вони використовуються:
- Управління ботнетами та троянами (C&C): Як основний або резервний канал зв'язку.
- Програми-вимагачі (Ransomware): Для передачі ключів шифрування на сервери зловмисників.
- Фішингові кампанії (наприклад, через SMS): Для швидкої генерації нових посилань з метою уникнення спам-фільтрів [2].

Використання алгоритмів генерації доменів стало стандартом для сучасного шкідливого ПЗ завдяки здатності забезпечувати стійкість ворожої інфраструктури. Аналіз DGA демонструє еволюцію від простої псевдовипадкової генерації до складних словникових та адаптивних алгоритмів. Розуміння специфіки їх використання, особливо в системах C&C та фішингових кампаніях, є важливою основою для створення нового покоління систем кіберзахисту.

У роботі систематизовано основні підходи до застосування алгоритмів генерації доменів у сучасних кібератаках, охарактеризовано псевдовипадкові та словникові DGA, а також механізми відвернення уваги захисників. Показано, що ефективність DGA визначається не лише кількістю згенерованих доменів, а й узгодженістю з тактикою атаки: від вибору «зерна» до імітації легітимної DNS-поведінки. Отримані результати підкреслюють необхідність поєднання технічних засобів виявлення з аналізом контексту застосування алгоритмів у ланцюгу компрометації. Подальші дослідження доцільно спрямувати на вдосконалення класифікації DGA-варіантів та інтеграцію таких знань у проактивні системи кіберзахисту.

1. "Large Language Models for Effective Detection of Algorithmically Generated Domains: A Comprehensive Review," *Computer Modeling in Engineering & Sciences (CMES)*, Tech Science Press, 2024. https://www.techscience.com/CMES/v144n2/63716/html
2. "An end-to-end framework for private DGA detection as a service," PubMed Central (PMC), 2024. https://pmc.ncbi.nlm.nih.gov/articles/PMC11355532/

Багаторівневі підходи щодо безпеки веб-орієнтованих систем

УДК 004.056

Александрос Фотинос[1], Лариса Шумова[2]

Східноукраїнський національний університет імені Володимира Даля,
[1]photinosaleksandros12@gmail.com, [2]shumova@snu.edu.ua

Одним із критично важливих факторів, що визначають успішність веб-системи, є її безпека. Механізми безпеки сучасних веб-орієнтованих систем реалізовані на основі багаторівневого підходу, що охоплює всі компоненти архітектури: клієнтську частину, серверну частину та рівень зберігання даних.

Схема на рисунку 1 демонструє три рівні захисту: клієнтська частина (React 18), серверна частина (NestJS) та рівень зберігання даних (PostgreSQL). Кожен рівень реалізує власні механізми безпеки, що забезпечують комплексний захист від основних векторів атак.

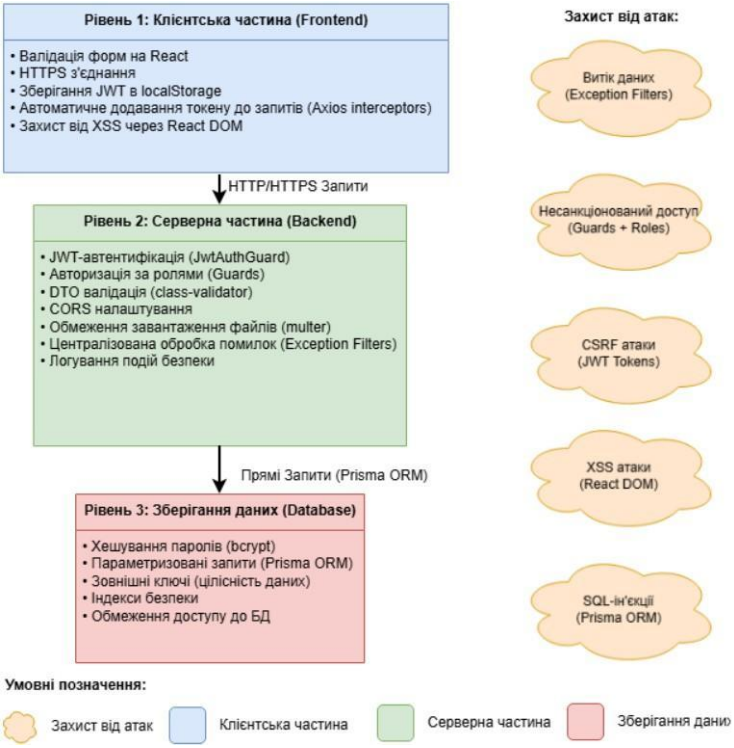


Рис.1. Багаторівневий підхід до безпеки веб-орієнтованої системи

Основним механізмом автентифікації є JWT-токени, які генеруються сервером після успішної перевірки облікових даних користувача. Токен містить ідентифікатор користувача, електронну пошту та роль, має обмежений час

життя двадцять чотири години та автоматично додається до всіх захищених запитів через інтерсептори HTTP-клієнта Axios. Такий підхід забезпечує stateless-автентифікацію, що спрощує масштабування системи та усуває необхідність зберігання стану сесій на сервері.

Ще одним рівнем безпеки є авторизація та розмежування прав доступу, що реалізовано за допомогою механізму Guards у фреймворку NestJS. Guard JwtAuthGuard перевіряє наявність та валідність JWT-токена для всіх захищених маршрутів, а додаткові перевірки ролі користувача обмежують доступ до функціоналу відповідно до призначення облікового запису. Розмежування реалізовано на рівні серверної логіки, що унеможливлює обхід обмежень через маніпуляції з клієнтської сторони.

Окремим рівнем безпеки є захист паролів користувачів, що забезпечується за допомогою алгоритму bcrypt. Паролі не зберігаються у відкритому вигляді в базі даних, а перед записом проходять процедуру хешування. Під час автентифікації введений пароль порівнюється зі збереженим хешем, що гарантує неможливість відновлення оригінального пароля навіть у випадку компрометації бази даних.

Захист від ін'єкційних атак забезпечується за рахунок використання Prisma ORM для взаємодії з базою даних PostgreSQL. Усі запити до бази даних формуються через типобезпечний API Prisma, що автоматично екранує вхідні параметри та використовує параметризовані запити. Це повністю усуває можливість SQL-ін'єкцій, оскільки ручне формування SQL-рядків у коді відсутнє.

Механізм CORS налаштований на серверній стороні для обмеження джерел, з яких дозволені запити до API. У конфігурації вказано конкретні дозволені origin, що запобігає міжсайтовим атакам та несанкціонованому використанню ендпоінтів з сторонніх доменів. Завантаження файлів реалізовано через бібліотеку multer з обмеженням максимального розміру файлу та збереженням у спеціальній директорії з унікальними іменами, що запобігає перезапису існуючих файлів та виконанню шкідливого коду через завантажені ресурси.

Обробка помилок реалізована централізовано через Exception Filters фреймворку NestJS. Користувачу не передаються деталі внутрішніх помилок сервера, що запобігає витоку інформації про структуру системи або версії використаних бібліотек. Помилки логуються на сервері для подальшого аналізу розробником, але клієнт отримує лише узагальнені повідомлення без технічного контексту. Такий підхід відповідає принципам безпеки через невідомість та мінімізує поверхню атаки.

Усі описані заходи багаторівневого підходу щодо безпеки веб-орієнтованої системи реалізовано й перевірено в процесі функціонального та інтеграційного тестування. Комплексний підхід до захисту даних, автентифікації, авторизації та валідації вхідних даних забезпечує відповідність системи сучасним вимогам до веб-додатків та гарантує безпеку веб-орієнтованої системи в онлайн-середовищі.

Аналіз засобів виявлення та протидії атакам типу container escape у середовищах Linux-контейнерів

УДК 004.056.5:004.75
Вікторія Шумська[1], Юрій Дорофєєв[2], Ірина Назарова[3]

Національний університет «Одеська політехніка»,
[1]marfonya.999@stud.op.edu.ua, [2]dym@op.edu.ua, [3]nazarova.i.v@op.edu.ua

У сучасних умовах переходу інформаційних систем до хмарних платформ питання безпеки хмарної інфраструктури набуває особливої актуальності. Однією з базових технологій таких середовищ є контейнеризація, що забезпечує ізоляцію застосунків, спрощує їх розгортання, масштабування та перенесення між різними середовищами. Водночас атаки, пов'язані з порушенням контейнерної ізоляції, виникають дедалі частіше та можуть призводити до компрометації хостової системи. Метою роботи є аналіз основних векторів атак типу container escape у Linux-контейнерах та підходів до їх запобігання.

Контейнеризація є формою віртуалізації на рівні операційної системи, за якої кілька ізольованих середовищ виконання використовують спільне ядро хостової системи. Основними механізмами ядра Linux, що застосовуються для реалізації контейнеризації, є простори імен (namespaces), контрольні групи (cgroups), привілеї (capabilities) та фільтрація системних викликів (seccomp) [1].

Однією з найбільш небезпечних загроз для контейнерної ізоляції є атака типу container escape, тобто вихід процесу за межі ізольованого середовища контейнера з подальшим отриманням доступу до ресурсів хостової системи або інших контейнерів. У дослідженні [2] відокремлено три основні джерела атак типу container escape: небезпечні конфігурації, уразливості компонентів контейнерної інфраструктури та вразливості ядра Linux.

Небезпечні конфігурації послаблюють фактичну ізоляцію контейнера та можуть створювати умови для несанкціонованого доступу до ресурсів хоста або підвищення привілеїв. Для їх виявлення використовують засоби перевірки декларативних конфігурацій контейнерної інфраструктури, зокрема Docker Bench for Security, kube-bench, Checkov і Conftest. Зниження ризику базується на застосуванні принципу найменших привілеїв: відмові від privileged-режиму, мінімізації capabilities, застосуванні user namespace, обмеженні host mounts, runtime-сокетів і ресурсів через cgroups.

Уразливості компонентів контейнерної інфраструктури виникають через помилки в коді runtime-компонентів або використання вразливих залежностей в образах. Показовим прикладом є вразливість CVE-2024-21626 у runc, яка могла дозволити процесу контейнера отримати доступ до файлової системи хоста [3]. Ризики, що надходять з рівня контейнерних образів, можуть створювати умови для подальшої компрометації хостової системи, використовуючи застарілі пакети, вбудовані секрети, недовірені базові образи, скомпрометовані або шкідливі образи в реєстрах, а також небезпечні інструкції збірки. Для протидії атакам, що базуються на уразливостях компонентів контейнерної інфраструктури, застосовують регулярне оновлення runtime-компонентів, використання довірених реєстрів і базових образів, підписування

та перевірку цілісності образів, SBOM, а також сканування за допомогою Trivy, Clair, Anchore, Grype та Snyk Container.

Уразливості ядра Linux є особливо небезпечними для контейнерних середовищ через спільне використання ядра контейнерами та хостовою системою. Прикладом є вразливість CVE-2022-0847 Dirty Pipe [4], яка дозволяла непривілейованому локальному користувачу записувати дані у сторінки page cache, пов'язані з read-only файлами, і таким чином підвищувати привілеї в системі. Основним засобом захисту від подібних атак є своєчасне оновлення ядра. Додатковим рівнем захисту може бути runtime-моніторинг системних викликів і подій ядра за допомогою Falco, Tracee та Tetragon.

Небезпечні конфігурації, уразливості компонентів контейнерної інфраструктури та уразливості ядра Linux можуть взаємно посилювати наслідки одне одного, створюючи умови для порушення меж ізоляції. Наприклад, у кампанії, описаній Aqua Security, атака поєднувала використання зловмисного контейнерного образу з небезпечною конфігурацією контейнера. Така комбінація дозволила шкідливому скрипту всередині контейнера використати механізм cgroup release_agent для виконання коду на хості [5].

Проведений аналіз показав, що зниження ризику атак типу container escape потребує комплексного підходу. Поєднання визначення фактичної ізоляції контейнера на рівні механізмів ядра Linux із runtime-моніторингом системних викликів і подій ядра на основі eBPF в єдиному програмному рішенні є перспективним напрямом формування комплексної оцінки поточного рівня захищеності контейнеризованого середовища. Такий підхід дозволяє враховувати не лише формальні параметри запуску контейнера, а й його фактичний стан під час виконання: належність процесів до просторів імен, обмеження cgroups, набір capabilities, режим seccomp, а також поведінкові ознаки, що проявляються через системні виклики та події ядра.

1. Sultan S., Ahmad I., Dimitriou T. Containers' Security: Issues, Challenges, and Road Ahead. IEEE Access. 2019. Vol. 7. P. 52976–52996. DOI: 10.1109/ACCESS.2019.2911732.
2. Chen K., Zhao Y., Guo J., Gu Z., Han L., Tang K. A Container Escape Detection Method Based on a Dependency Graph. Electronics. 2024. Vol. 13, № 23. Article 4773. DOI: 10.3390/electronics13234773.
3. National Vulnerability Database. CVE-2024-21626 Detail. URL: https://nvd.nist.gov/vuln/detail/CVE-2024-21626 (дата звернення: 04.05.2026).
4. National Vulnerability Database. CVE-2022-0847 Detail. URL: https://nvd.nist.gov/vuln/detail/CVE-2022-0847 (дата звернення: 04.05.2026).
5. Eitani A. Threat Actors Using release_agent Container Escape. Aqua Security. 03.11.2021. URL: https://www.aquasec.com/blog/threat-alert-container-escape/ (дата звернення: 04.05.2026).

Автоматизація Vivado через Jupyter Notebook для вдосконалення проєктів на ПЛІС

УДК 004.274:004.896

Іван Яблоков

ДВНЗ «Донецький національний технічний університет», ivan.yablokov@donntu.edu.ua

За останні кілька років використання штучного інтелекту (ШІ) значно підвищилось у всіх сферах високотехнологічного виробництва. Області застосування варіюються від генерації складних мультимедійних об'єктів до автоматизації розробки програмних та апаратних систем. Зокрема, розробники систем автоматизованого проєктування (EDA – Electronic Design Automation), таких як AMD/Xilinx Vivado, активно впроваджують алгоритми ШІ для переходу від трудомісткого ручного налаштування параметрів ПЛІС (FPGA) до інтелектуальних, повністю автоматизованих процесів. Це значно полегшить роботу інженерів при створенні нових проєктів та оптимізації існуючих архітектур, призначених для захисту рішень у галузі IoT та кіберфізичних систем.

Одним із найбільш перспективних напрямків є використання генеративно-змагальних мереж (GAN) для автоматичної генерації RTL-коду або топологій розміщення компонентів. Однак при генерації великої кількості варіантів дизайну перевірка кожного з них на життєздатність стає критичним «вузьким місцем».

Кожен згенерований нейронною мережею варіант необхідно піддати процедурам синтезу, розміщення (placement) та трасування (routing). Оскільки тривалість цих процесів у Vivado варіюється від декількох хвилин до декількох годин для складних проєктів, загальний час навчання нейронної мережі стає дуже великим. Традиційне використання графічного інтерфейсу (GUI) Vivado не дозволяє ефективно масштабувати цей процес, оскільки вимагає ручного керування та споживає значні ресурси системи на візуалізацію.

Для подолання описаної проблеми пропонується використання середовища Jupyter Notebook для керування усіма процесами. На відміну від стандартного GUI, Jupyter дозволяє реалізувати програмне керування додатком Vivado у режимі batch mode за допомогою TCL-скриптів. Jupyter Notebook у цьому контексті виступає не просто як редактор коду, а як інтерактивне середовище для швидкого прототипування. Його головна перевага полягає у можливості розділити процес на окремі блоки (cells), що дозволяє зберігати проміжні стани обчислень без необхідності повторного запуску всього скрипту.

У сучасних обчислювальних системах для EDA-завдань спостерігається гостра диспропорція між кількістю обчислювальних ядер процесора та доступним об'ємом швидкої оперативної пам'яті. Якщо процесорні потужності у 2026 році дозволяють легко оперувати 64 ядрами та 128 потоками на одну робочу станцію, то оперативна пам'ять перетворилася на основний обмежувач масштабованості.

При розпаралелюванні процесів навчання ШІ споживання RAM зростає лінійно відносно кількості потоків. Кожен процес синтезу для сучасних

складних систем на кристалі (SoC) потребує від 16 до 48 ГБ RAM. Таким чином, для повного завантаження потужного процесора система повинна мати об’єм пам’яті від 215 ГБ до 512 ГБ, що виводить обладнання з розряду персональних комп’ютерів у розряд серверів високої щільності.

Якщо під час піку хоча б одному процесу забракне фізичної пам'яті, ОС активує Swap-файл (підкачку на диск). Оскільки швидкість навіть найсучасніших NVMe-накопичувачів у 2026 році все ще на порядки нижча за швидкість DDR4/DDR5, продуктивність усієї системи навчання миттєво деградує. Це явище отримало назву "Thrashing" — стан, коли система витрачає 90% часу на переміщення даних між диском і пам’яттю, а не на реальні обчислення.

Економічний фактор та дефіцит 2026 року

Важливим аспектом є вартість заліза. Ціни на оперативну пам’ять з набранням популярності ШІ дуже сильно підвищились, оскільки дуже багато компаній залучено у процес розробки різноманітних моделей (LLM, GAN, Diffusion). Виробники пам'яті пріоритезують випуск HBM-пам'яті для прискорювачів, що створює дефіцит звичайної серверної RAM.

Jupyter Notebook дозволяє реалізувати інтелектуальні алгоритми управління пам'яттю:

1) Ресурсний моніторинг: перед запуском чергового пакету (batch) даних, скрипт перевіряє доступну RAM через бібліотеку psutil. Якщо вільний об'єм менше 20%, запуск нових процесів призупиняється.
2) Force Cleanup: після завершення кожної ітерації Jupyter примусово завершує дочірні процеси Vivado та очищає системний кеш.

Висновки. Впровадження паралельного запуску через TCL-скрипти дозволяє досягти майже лінійного прискорення процесу навчання у порівнянні з послідовною перевіркою. Jupyter Notebook дозволяє інтегрувати аналітику безпосередньо в процес навчання. При завершенні циклу навчання система виводить результати у вигляді графіків, гістограм та теплових карт розміщення компонентів. Що значно спрощує аналіз результатів для користувача.

Використання Jupyter Notebook у поєднанні з TCL-автоматизацією Vivado є ефективним рішенням для навчання сучасних генеративних моделей у галузі проєктування ПЛІС. Це не тільки зменшує час роботи за рахунок розпаралелювання обчислень, але й забезпечує повну автоматизацію циклу "генерація-перевірка-корекція", що є критично важливим для створення інтелектуальних систем EDA майбутнього. Запропонований підхід робить процес розробки гнучким, масштабованим та мінімізує вплив людського фактора на етапі ітеративної перевірки дизайну.

1. Vivado Design Suite User Guide: Tcl Command Reference Guide (UG835). – San Jose: AMD/Xilinx, 2025. – 1340 p.
2. Goodfellow I., Bengio Y., Courville A. Generative Adversarial Networks: Deep Learning Series. – MIT Press, 2020. – 800 p.
3. Smith J. The Economics of Semiconductor Memory in the AI Era. – Tech Economic Review, 2026. – 45-52 c.

Системне вдосконалення підходів до забезпечення кібербезпеки об'єктів критичної інфраструктури

УДК 621.395.7 (043.2)

Юрій Якименко[1]

Державний університет інформаційно-комунікаційних технологій,
[1]yakum14@ukr.net

Системне вдосконалення кібербезпеки об'єктів критичної інфраструктури (ОКІ) в Україні є стратегічним пріоритетом в умовах постійних кібератак. Спрямованість кібератак здійснюються в основному на об'єкти енергетики,транспорту, фінансів, зв'язок (телеком), банківська система та державного управління (урядові портали), що є критично важливою складовою сучасних організацій, порушує їх функціонування і створює загрозу державі та суспільству. Кібератаки інтегровано поєднуються з військовими діями, а кіберзагрози стали системними та стратегічними, а не лише залишились як технічними. За сучасними дослідженнями по результатам комплексного аналізу кіберзагроз ОКІ в 2025 році та практикою кіберконфлікту в Україні основними ключовими загрозами визначені: цілеспрямовані атаки (apt); атаки на SCADA/ICS системи; DDOS-атаки на державні ресурси; шкідливе ПЗ (WIPER, RANSOMWARE); соціальна інженерія; AI-підсилені атаки. Як причина показано, що критична інфраструктура є особливо вразливою через застарілі системи та складність модернізації . [1] Класичні підходи щодо підвищення кібербезпеки ОКІ (з периметрового захисту- firewall, IDS/IPS; контролю доступу і антивірсного захисту) вже стали недостатньо ефективнми проти сучасних атак. Тому у 2025 році з'явилися інші- сучасні підходи: щодо перевірки кожного запиту (Zero Trust Architecture), щодо управління ризиками і пріоритезацією загроз (Risk-based Security), щодо моніторингу в 24/7 і реагування на інциденти SOC (Security Operations Center), щодо обміну інформацією про загрози (Threat Intelligence), щодо кіберстійкості, які спрямовані не тільки на захист, но і на відновлення ресурсів (Cyber Resilience). [1,2]

ІТ-інфраструктура об'єктів організацій включає в себе сервери, мережеве обладнання, програмне забезпечення, бази даних, хмарні сервіси та користувацькі пристрої. Її захист базується на класичній тріаді інформаційної безпеки — конфіденційність, цілісність та доступність. Саме закон України «Про основні засади забезпечення кібербезпеки України» встановлює загальні принципи побудови системи кіберзахисту та ролі її суб'єктів. Інший закон України «Про захист інформації в інформаційно-комунікаційних системах» визначає вимоги до захисту інформації і в державних інформаційних ресурсах. Відповідно до вимог цих законів захист ІТ-інфраструктури практично реалізується за принципом багаторівневого захисту. [2]

В 2025 році постановою КМУ затверджений оновлений перехід кіберзахисту ОКІ на ризик-орієнтовану модель, з урахуванням більш глибокого аналізу власних ризиків: шляхом проведення заходів з кіберзахисту та урахуванням отриманих результатів управління ризиками кібербезпеки в організації повинна бути побудована адаптивна система безпеки. Реалізація

нового підходу дозволить скоротити час реагування та відновлення функціонування ОКІ після кібератак, зменшити кількість значних кіберінцидентів, а також підвищити рівень кіберзахисту об'єкту. [3]

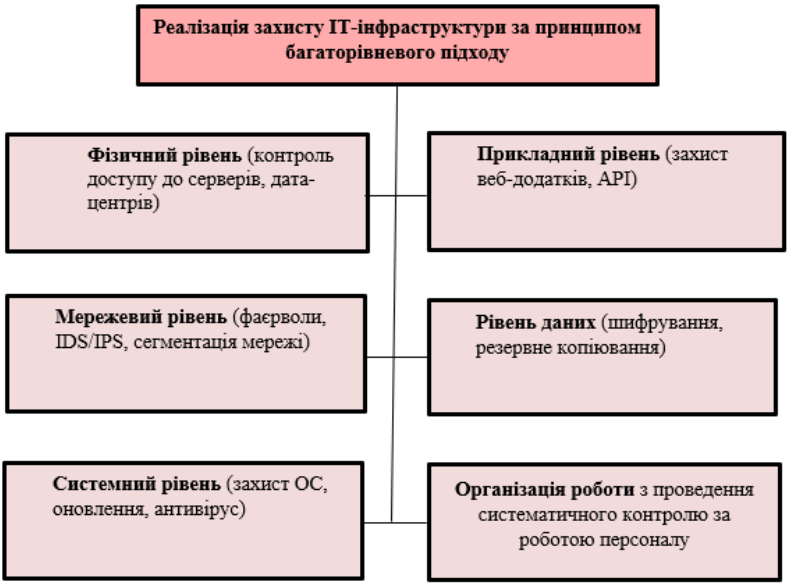


Рис.1. Реалізація захисту IT-інфраструктури за принципом багаторівневого підходу

В той же час відповідно до вимог міжнародного стандарту ISO/IEC 27001, організація, яка визначена як ОКІ, повинна впроваджувати систему управління інформаційною безпекою (ISMS), з основними функціональними завданнями оцінки ризиків, контролю доступу, управління інцидентами та аудитом. Саме вимоги цього стандарту і інших в сфері безпеки забезпечують системність, уніфікацію приємливих підходів в сучасних умовах діяльності і проведення оцінки рівней безпеки організацій. Завдяки функціонуванню ISMS повинно бути забезпечена безперервна готовність ОКІ до виконання своїх задач, визначених в документах політики безпеки.

Таким чином, системне вдосконалення кібербезпеки ОКІ повинно бути спрямовано на комплексне використання всіх можливостей ISMS в організаційному і технічному напрямках забезпечення багаторівневого захисту інформаційних ресурсів та в цілому - високого рівня ефективної діяльності організації. Безперервне вдосконалення треба проводити в послідовності дій: виявлення загроз, аналіз ризиків інформаційної безпеки, забезпечення захисту інформації, моніторингу процесів управління безпекою, реагування на інциденти інформаційної безпеки і відновлення до нормального стану діяльності ОКІ. Результати вдосконалення треба впроваджувати в побудовану

адаптивної системи безпеки організації. Більше можливостей демонструє впровадження інноваційних рішень, якими є використання штучного інтелекту у напрямах: виявлення аномалій, щоб відстежувати поведінку мережі та автоматично реагувати на загрози; прогнозування кібератак і використання сучасних технологій протидії різним кібератакам. Активно використовуються сучасні інноваційні технології: SOAR для автоматизації реагування, XDR для розширеного виявлення загроз, Digital Twins для оцінки інфраструктури, Cyber Range для використання як кіберполігонів у дослідженнях.

1. Звіт ДЦКЗ Держспецзв'язку про роботу Системи виявлення вразливостей і реагування на кіберінциденти та кібератаки (СВВ) за 2025 рік. URL:
2. https://cip.gov.ua/services/cm/api/attachment/download?id=73033
3. Ільєнко А.В, Телющенко В.А., Дубчак О.В. Сучасні кіберзагрози критичної інфраструктури України та світу № 3 (27), 2025 . DOI 10.28925/2663-4023.2025.27.719
4. Постанова КМУ від 13.11.2025 р. № 1470. Загальні вимоги до кіберзахисту об'єктів критичної інфраструктури (ОКІ) в новій редакції.

Трасування безпекових вимог у системах предиктивної аналітики

УДК 004.89:339.138 (043.2)

Дмитро Яценко[1], Володимир Садовенко[2]

Державний університет інформаційно-комунікаційних технологій,
[1]d.yatsenko@stud.duikt.edu.ua, [2]v.sadovenko@duikt.edu.ua

Системи предиктивної аналітики (СПА) у цифровому маркетингу обробляють масивні поведінкові, демографічні та транзакційні дані користувачів, що формують специфічну поверхню атаки. Поряд із класичними загрозами інформаційній безпеці такі системи зазнають впливу атак, специфічних для машинного навчання (МН): data poisoning, evasion, model extraction, membership inference, model inversion [1, 2, 3]. Чинні стандарти управління ризиками штучного інтелекту, зокрема ISO/IEC 23894:2023 [4] та NIST AI 100-2 E2025 [1], формулюють принципи високого рівня, проте не пропонують архітектурного інструментарію проєктування. Безпекові механізми впроваджуються на пізніх етапах життєвого циклу системи, що знижує стійкість і ускладнює верифікацію її властивостей.

Мета роботи — підвищення стійкості СПА до атак на МН шляхом розроблення методу трасування безпекових вимог до архітектурних компонентів, механізмів контролю та метрик верифікації, що враховує особливості маркетингових даних — відкритість каналів збору поведінкових сигналів та схильність навчальних вибірок до забруднення через клік-фрод.

Наукова новизна. Уперше для класу СПА у галузі цифрового маркетингу запропоновано метод трасування безпекових вимог за схемою «вимога — вектор загрози — архітектурний компонент — механізм контролю — метрика верифікації». На відміну від універсальних стандартів управління ризиками AI

[1, 4], метод враховує специфіку маркетингових даних: їхню поведінкову природу, чутливість до приватності та доступність каналів збору для зловмисних впливів. На відміну від каталогів загроз [2, 3], трасування інтегрується безпосередньо в етап архітектурного проєктування.

Формальне подання методу. Введемо скінченні множини: R — безпекових вимог до СПА; C — архітектурних компонентів СПА; T — векторів загроз за NIST AI 100-2 [1], MITRE ATLAS [3], OWASP ML Top 10 [2]; M — механізмів контролю; V — метрик верифікації. Метод трасування визначається як п'ятимісне відношення:

$$\Phi \subseteq R \times C \times T \times M \times V, \quad (1)$$

де кортеж $(r, c, t, m, v) \in \Phi$ задає трасований ланцюжок «вимога r локалізована на компоненті c, протистоїть загрозі t через механізм m із вимірюванням ефективності метрикою v. Властивість повноти трасування формулюється як:

$$\forall r \in R\ \exists (c, t, m, v): (r, c, t, m, v) \in \Phi, \quad (2)$$

Тобто, кожна вимога має хоча б один ланцюжок до метрики верифікації, що слугує критерієм верифікованості архітектури СПА. Кількісним показником якості трасування виступає коефіцієнт покриття загроз $Cov(T) = \frac{|\pi_T(\Phi)|}{|T|}$, де π_T є проєкцією відношення на множину загроз. Введене формальне подання дозволяє верифікувати архітектуру СПА через перевірку умови (2) та обчислення коефіцієнта покриття на множині референсних загроз.

Запропонований підхід. Сформовано матрицю трасування (табл. 1), що пов'язує безпекові вимоги до СПА з цільовими архітектурними компонентами, релевантними векторами загроз за NIST AI 100-2 E2025 [1], MITRE ATLAS [3] та OWASP ML Top 10 [2], типовими механізмами контролю та метриками верифікації. Табл. 1 подає скінченну реалізацію відношення (1) для базового набору вимог потужністю |R| = 6.

Таблиця 1

Матриця трасування безпекових вимог СПА

Вимога	Компонент	Вектор загрози	Механізм контролю	Метрика
Цілісність навчальних даних	Шар прийому даних	Data Poisoning (NIST; ATLAS AML.T0020; OWASP ML02)	Детекція аномалій, санітизація даних	Частка виявлених забруднених записів
Конфіденційність персональних даних	Сховище даних	Membership Inference (NIST; OWASP ML04)	Диференційна приватність, керування доступом	ε-бюджет приватності
Цілісність моделі	Репозиторій моделей	Model Poisoning (OWASP ML10)	Криптографічне підписування артефактів	Повнота перевірки підпису
Контрольованість прогнозів	Сервіс прогнозування	Evasion / Output Integrity (NIST; OWASP ML01, ML09)	Моніторинг дрейфу, валідація входу	Відхилення розподілу прогнозів

Вимога	Компонент	Вектор загрози	Механізм контролю	Метрика
Доступність сервісу прогнозування	Сервіс прогнозування	Availability / Energy-latency Attacks (NIST)	Обмеження частоти, тайм-аути обчислень	SLA доступності, p95 latency
Авторизованість запитів	Шар інтеграції	Model Extraction (NIST; OWASP ML05)	Автентифікація API, rate limiting	Частка авторизованих запитів

Висновки. Запропонований метод дозволяє розглядати безпеку СПА як архітектурно інтегровану нефункціональну характеристику, що підвищує верифіковність архітектурних рішень. Подальші дослідження передбачають експериментальну валідацію методу на прототипі СПА для задач прогнозування відтоку клієнтів та динамічного ціноутворення.

1. Vassilev A., Oprea A., Fordyce A., Anderson H., Davies X., Hamin M. Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations. NIST AI 100-2 E2025. Gaithersburg: NIST, 2025. DOI: 10.6028/NIST.AI.100-2e2025.
2. OWASP Machine Learning Security Top 10. 2023 edition. URL: https://mltop10.info/ (дата звернення: 06.05.2026).
3. MITRE ATLAS: Adversarial Threat Landscape for Artificial-Intelligence Systems. URL: https://atlas.mitre.org/ (дата звернення: 06.05.2026).
4. ISO/IEC 23894:2023. Information technology — Artificial intelligence — Guidance on risk management. Geneva: ISO, 2023. 28 p.

Оцінювання рівня інформаційної безпеки державних інформаційних ресурсів

УДК 004.056

Валентина Ящук[1], Діана Рівняк[2]

Львівський державний університет безпеки життєдіяльності,
[1]valentina.lender@gmail.com, [2]dianarivnak@gmail.com

У сучасних умовах цифровізації державного управління та зростання інтенсивності кіберзагроз проблема забезпечення належного рівня інформаційної безпеки державних інформаційних ресурсів набуває стратегічного значення. Державні інформаційні ресурси є основою функціонування електронного урядування, а їх компрометація може призвести до значних соціально-економічних і політичних наслідків. Це обумовлює необхідність розроблення ефективних методів оцінювання рівня їх захищеності. Метою роботи є розроблення математичної моделі оцінювання рівня інформаційної безпеки державних інформаційних ресурсів на основі інтегрального підходу з використанням системи індикаторів та вагових коефіцієнтів.

Аналіз існуючих підходів до оцінювання інформаційної безпеки показує, що більшість із них базується на якісних або експертних оцінках, що знижує об’єктивність результатів. Перспективним є застосування кількісних моделей,

які дозволяють формалізувати процес оцінювання та забезпечити порівнюваність результатів.

Нами запропоновано методологію оцінювання, яка передбачає формування системи індикаторів безпеки, що охоплюють такі складові, як рівень захищеності інформаційних систем, ефективність механізмів контролю доступу, стійкість до кіберзагроз та здатність до реагування на інциденти. Для кожного індикатора визначаються відповідні метрики, що дозволяють здійснювати кількісну оцінку стану безпеки. Такий підхід передбачає представлення рівня інформаційної безпеки у вигляді інтегрального показника:

$$I_{sec} = \sum_{i=1}^{n} w_i * S_i \quad (1),$$

де I_{sec} — інтегральний показник рівня інформаційної безпеки;

S_i — значення і-го індикатора безпеки;

w_i — ваговий коефіцієнт важливості відповідного індикатора;

n — кількість індикаторів.

Система індикаторів формується за основними складовими інформаційної безпеки, зокрема конфіденційність інформації, цілісність даних; доступність ресурсів, автентичність та контроль доступу, стійкість до кіберінцидентів. Кожен індикатор оцінюється за шкалою від 0 до 1, де 0 відповідає критичному рівню вразливості, а 1 — повній відповідності вимогам безпеки. Значення індикаторів визначаються на основі аналізу технічних параметрів систем, результатів аудиту безпеки та даних моніторингу.

Для врахування ризиків інформаційної безпеки пропонуємо використовувати коригуючий коефіцієнт ризику.

$$R = \sum_{j=1}^{m} p_j * d_j \quad (2),$$

де R — інтегральний ризик;

p_j — ймовірність реалізації j-ої загрози;

d_j — потенційні збитки від реалізації загрози;

m — кількість загроз.

Відтак з урахуванням ризику інтегральний показник безпеки набуває такого вигляду.

$$I_{sec}^{*} = I_{sec} * (1 - R) \quad (3)$$

Отже, інтегральну оцінку рівня інформаційної безпеки пропонується визначати на основі агрегування часткових показників із використанням вагових коефіцієнтів, що враховують критичність окремих компонентів системи. Такий підхід забезпечує можливість отримання узагальненого показника, який відображає поточний стан захищеності державних інформаційних ресурсів та потенційний вплив загроз на інформаційні ресурси.

Наукова новизна роботи полягає у розробленні інтегрованої моделі оцінювання рівня інформаційної безпеки, яка поєднує індикаторний підхід із ризик-орієнтованим аналізом, що забезпечує підвищення точності та

об'єктивності оцінювання. Практичне значення полягає у можливості застосування запропонованої моделі для проведення аудиту інформаційної безпеки державних інформаційних систем; підтримки прийняття управлінських рішень; визначення пріоритетних напрямів підвищення рівня захищеності; моніторингу змін стану інформаційної безпеки у динаміці.

Результати дослідження свідчать, що використання інтегрального показника дозволяє отримати узагальнену оцінку стану інформаційної безпеки та своєчасно виявляти критичні вразливості. Запропонована модель є гнучкою та може бути адаптована до специфіки конкретних державних інформаційних систем.

У роботі запропоновано модель оцінювання рівня інформаційної безпеки державних інформаційних ресурсів на основі інтегрального показника та ризикоорієнтованого підходу. Використання системи індикаторів та вагових коефіцієнтів дозволяє здійснювати кількісну оцінку стану захищеності інформаційних систем. Запропонований підхід забезпечує підвищення об'єктивності оцінювання та може бути використаний у практиці управління інформаційною безпекою державного сектору.

1. Ящук В., Балацька В. Підвищення кіберстійкості критичної інформаційної інфраструктури держави через вдосконалення процесів реагування на кіберінциденти // Цивільний захист в умовах війни : збірник тез доповідей II Міжнародної науково-практичної конференції, м. Львів, 15 квітня 2026 року. Львів : ЛДУБЖД, 2026. С. 92–95.
2. Венгерський П.С., Вишневська Н.С., Хохлачова Ю.Є., Хорошко В.О., Чобаль О.І. Кількісна оцінка кіберзахищеності інформації. Захист інформації. 2023. Т. 25, №2. С. 53–61.

Проблеми інтервального моніторингу цілісності інформаційного стану корпоративних кіберфізичних систем

УДК 004.056:681.5

Павло Матусяк[1], Ярослав Тарасенко[2]

[1]Державний університет інформаційно-комунікаційних технологій, Pavelmatusyak@gmail.com,
[2]Державний науково-дослідний інститут випробувань і сертифікації озброєння та військової техніки, yaroslav.tarasenko93@gmail.com

В умовах цифровізації виробничих і сервісних процесів корпоративні кіберфізичні системи виступають у ролі середовища з особливими ознаками порушення безпеки. Вторгнення у такому середовищі супроводжується поступовим спотворенням сукупності контрольованих параметрів на відміну від миттєвих відмов, спричинених порушенням цілісності інформаційного стану. Під цілісністю інформаційного стану мається на увазі збереження узгодженості, допустимості та відсутності спотворення поточних оцінок параметрів у часі. Для корпоративних кіберфізичних систем таке формулювання набуває особливого значення через можливість накопичення

спотворень на рівні сенсорів, шлюзів, сервісів обробки та під час формування керуючих рішень. У таких умовах інтервальний моніторинг передбачає здійснення контролю в межах допустимого інтервалу змін, який залежить від режиму функціонування, шумів, затримок та невизначеностей моделі. Підтвердження цьому висвітлено в роботі [1], де розглядаються інтервальні спостерігачі, стійкі до атак у ролі засобу оцінювання стану. Такий стан розглядається за умов прихованих атак, які спричиняють маскування порушення цілісності в допустимих межах фіксованого порогу.

Аналіз роботи [2], де представлено удосконалений метод інтервального оцінювання дозволив виявити та сформувати три основні взаємопов'язані проблеми інтервального моніторингу: складність відокремлення початкової фази вторгнень від штатних коливань системи, швидка втрата інформативності фіксованих меж контролю у змінних умовах функціонування, збільшення кількості хибних спрацювань за умов підвищення чутливості моніторингу.

Отже, перспективним напрямком вирішення зазначених проблем є поєднання інтервального моніторингу з інтелектуальним уточненням меж штатного функціонування системи. Важливим є інтелектуальне коригування інтервалів за поточним профілем функціонування системи. Доцільно формалізувати правила коригування, вибір ознак порушення цілісності та узгодження моніторингу з допустимим рівнем хибних спрацювань тривог.

1. Degue K.H., Ny J.L., Efimov D. Stealthy attacks and attack-resilient interval observers. *Automatica*. 2022. Vol. 146. URL: https://doi.org/10.1016/j.automatica.2022.110558 (дата звернення: 04.05.2026).
2. Fan J., Huang J., Zhao. X. Improved interval estimation method for cyber-physical systems under stealthy deception attacks. *IEEE Transactions on Signal and Information Processing Over Networks*. 2022. Vol. 8. P. 1-11.

Алгоритм аудіостеганографії без внесення змін у файл-контейнер

УДК 004.056.5

Костянтин Фріга[1], Юрій Дорофєєв[2], Ірина Назарова[3]

Національний університет «Одеська політехніка»,
[1]10252733@stud.op.edu.ua, [2]dym@op.edu.ua, [3]nazarova.i.v@op.edu.ua

Переважна більшість методів аудіостеганографії передбачає вбудовування корисного повідомлення безпосередньо в аудіоконтейнер, причому одним з основних показників якості методу є степень непомітності внесених змін з точки зору слухової системи людини [1], [2].

В [3], [4] розглянуто спосіб прихованого передавання інформації за допомогою монохромних, кольорових графічних файлів та аудіофайлів без зміни контейнера на основі методу Zero Distortion Technique (метод нульового спотворення - МНС). На відміну від традиційних методів аудіостеганографії, де повідомлення безпосередньо вбудовується в сигнал і може змінювати його структуру, у МНС контейнер використовується як готова бітова послідовність.

Повідомлення відновлюється за координатами позицій, у яких у контейнері вже наявні потрібні бітові фрагменти.

У роботі [4] розглянуто застосування методу нульового спотворення саме в аудіостеганографії. Автори описують підхід, за якого аудіоконтейнер не змінюється, приховане повідомлення передається побітово (один біт повідомлення на один семпл аудіоконтейнера), відновлюється за матрицею індексів. Для захисту цієї матриці використовується Indexed Based Chaotic Sequence, тобто хаотична перестановка координат [3].

Метою роботи є суттєве зменшення обсягу координатних даних при використанні методу МНС для прихованого передавання текстових даних. У запропонованому варіанті використано 6-8-бітове контейнерозалежне кодування, за якого одна координата відповідає не окремому біту, а цілому 6-8-бітовому коду символу. Для цього програма аналізує конкретний аудіофайл, визначає наявні бітові комбінації у MSB-вікні та формує адаптивний алфавіт, залежний від структури контейнера. Додатково передбачено перевірку достатності координат, заборону повторного використання позицій і хаотичну перестановку матриці координат на основі логістичного відображення.

Під час декодування використовується той самий контейнер і відповідний масив координат.

Якщо аудіофайл або індексний масив було змінено, зв'язок між координатами та бітовими фрагментами може порушуватися, тому повідомлення може відновлюватися неправильно. Результати тестування показали, що запропонований підхід при збереженні головної властивості МНС дозволяє отримати зменшення розміру службового індексного файлу не менше, ніж у k разів, де k – розрядність використаної кодировки символів, порівняно з побітовим варіантом та використанням матриці координат.

Окремо розглянуто вплив типових атак на бітову структуру аудіосигналу: заміна молодших бітів не має впливу на повідомлення, якщо робочі коди формуються з MSB-області, тоді як адитивна шумова атака є більш небезпечною через можливий вплив на старші біти при переповненні молодших.

Отже, запропоноване удосконалення зменшує службові витрати МНС і зберігає незмінність контейнера, але потребує суворої відповідності отриманого масиву індексів аудіоконтейнеру.

1. Joshi R., Trivedi M.C., Goyal V., Bhati D. Recent Trends for Practicing Steganography Using Audio as Carrier: A Study. Advances in Data and Information Sciences. Singapore: Springer, 2023. Vol. 522. P. 549-555. URL: https://doi.org/10.1007/978-981-19-5292-0_52 (дата звернення: 15.04.2026).
2. AlSabhany A.A., Ali A.H., Ridzuan F., Azni A.H., Mokhtar M.R. Digital audio steganography: Systematic review, classification, and analysis of the current state of the art. Computer Science Review. 2020. Vol. 38. Article 100316. URL: https://doi.org/10.1016/j.cosrev.2020.100316 (дата звернення: 15.04.2026).
3. Shivani, Yadav V.K., Batham S. Zero Distortion Technique: An approach

to image steganography on color images using strength of Chaotic Sequence. Proceedings of the 2014 International Conference on Information and Communication Technology for Competitive Strategies, 2014.New York: ACM, 2014. URL:
4. https://doi.org/10.1145/2677855.2677905 (дата звернення: 04.05.2026).
5. Sharma S., Yadav V.K., Trivedi M.C., Gupta A. Audio Steganography using ZDT: Encryption using Indexed Based Chaotic Sequence. Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies, 2016. New York: ACM, 2016. URL: https://doi.org/10.1145/2905055.2905272 (дата звернення: 04.05.2026).

Інтеграція OIDC-провайдера в енергетичну систему для забезпечення контролю доступу

УДК 004.056.5

Андрій Волощук[1], Іван Бородій[2], Галина Осухівська[3]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]andrii_voloshchuk3969@tntu.edu.ua, [2]ivanborodii@tntu.edu.ua,
[3]osukhivska@tntu.edu.ua

Енергетична система, що поєднує прилади обліку, датчики параметрів мережі, релейні контролери, диспетчерські сервіси та засоби моніторингу, потребує єдиної процедури ідентифікації з метою обміну даними. Шифрування каналів MQTT/TLS, CoAP/DTLS або HTTPS/TLS захищає передавання даних від перехоплення, однак не гарантує безпеки системи. Тому інтеграція OIDC-провайдера в енергетичну систему є важливою для централізованого керування доступом, дозволяє здійснювати аудит подій та запобігати несанкціонованому підключенню. Це особливо важливо для об'єктів, де обладнання працює тривалий час, а ручне переналаштування доступу може впливати на продуктивність роботи.

OIDC-провайдер у такій архітектурі виконує роль надійної компоненти в енергетичній системі. Він видає підписані JWT-токени, а служба маршрутизації повідомлень перевіряє їх локально за відкритим ключем провайдера. Це дає змогу реалізувати модель нульової довіри, за якої кожна дія перевіряється незалежно від розташування компонента в мережі.

Узагальнену архітектуру енергетичної інфраструктури з використанням OIDC-провайдера наведено на рисунку 1.

На першому етапі роботи системи здійснюється реєстрація та підтвердження нового IoT-пристрою. Для цього доцільно використовувати сценарій Device Authorization Flow, в якому пристрій ініціює запит, отримує службові коди, а оператор підтверджує підключення через SCADA-консоль або інший захищений інтерфейс. Після підтвердження OIDC-провайдер видає підписаний токен доступу, де введення обладнання потребує участі відповідального персоналу [1].

Другий етап полягає у передаванні токена до служби маршрутизації повідомлень під час підключення. JWT-токен є не лише доказом автентифікації,

а й носієм правил доступу. Поле allowed_topics визначає дозволені канали MQTT або ресурси CoAP, permitted_protocols задає допустимі протоколи обміну, roles і scope описують роль та дозволені операції, а security_level розмежовує обладнання передавання та керування [2].

Третій етап передбачає локальне застосування політик доступу. Служба маршрутизації перевіряє підпис JWT, строк дії, призначення, дозволені канали, протоколи та ролі без окремого звернення до сервера авторизації для кожного повідомлення. Усі відмови, спроби звернення до заборонених ресурсів і наближення завершення строку дії токена фіксуються в журналах безпеки, що спрощує аудит інцидентів.

Четвертий етап пов'язаний з вибором протоколу обміну. Аналітичний модуль може враховувати затримку, втрати пакетів, пропускну здатність і навантаження служби маршрутизації, однак він має рекомендувати лише ті протоколи, які зазначені у permitted_protocols конкретного пристрою. Отже, адаптивне перемикання між MQTT, CoAP і HTTPS не обходить політику безпеки, а діє в межах наперед визначених дозволів [3].

П'ятий етап стосується стійкості до відмов і кібератак. Якщо OIDC-провайдер тимчасово недоступний, обладнання з чинними токенами продовжує передавання даних до завершення строку їх дії, тоді як нові або не підтверджені підключення блокуються.

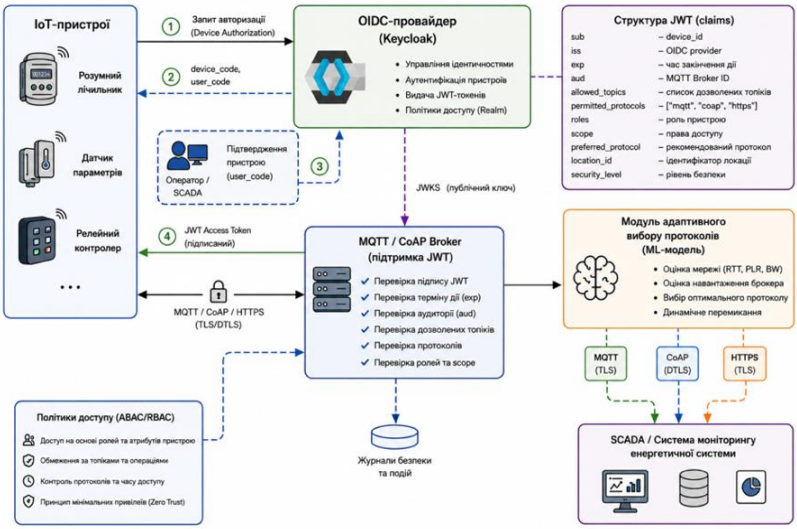


Рис. 1. Архітектура системи IoT-вузлів енергетичної інфраструктури на основі OIDC-провайдера

Така інтеграція забезпечує безперервність роботи компонентів, централізований контроль прав доступу та застосування політик у системі енергетичної інфраструктури.

1. Denniss W., Bradley J. RFC 8628: OAuth 2.0 Device Authorization Grant. IETF, 2019. URL: https://datatracker.ietf.org/doc/html/rfc8628.
2. Jones M., Bradley J., Sakimura N. RFC 7519: JSON Web Token (JWT). IETF, 2015. DOI: 10.17487/RFC7519.
3. Voloshchuk A. & Osukhivska H. Adaptive multi-protocol communication for energy systems. Scientific Journal of the Ternopil National Technical University. 2025. Vol. 119, No. 3. P. 97-106.

Ампліфікація інтегрованої системи управління інформаційною безпекою

УДК 004(056.53+413.4)::001.51

Володимир Мохор[1], Олександр Бакалинський[1], Ярослав Дорогий[2], Василь Цуркан[1,3]

[1]ІПМЕ ім. Г.Є. Пухова НАН України, v.mokhor@gmail.com, baov@meta.ua

[2]ДонНТУ, yaroslav.dorohyi@donntu.edu.ua
[3]КПІ ім. Ігоря Сікорського, v.v.tsurkan@gmail.com

Діяльність будь-якої організації орієнтована на задоволення потреб і очікувань зацікавлених сторін [1]. Серед них виокремлюється зберігання властивостей інформації. Насамперед конфіденційності (приватності), цілісності та доступності [2]. Таке виділення пов'язується з тим, що інформація є цінністю для організацій і тлумачиться як актив. До того ж з використанням продуктів, послуг на основі штучного інтелекту [3]. Це спонукає до забезпечення їх відповідального розроблення, упровадження, використання та, як наслідок, призводить до виникнення емерджентних ризиків. Тож ампліфікування інтегрованої системи управління інформаційною безпекою є актуальним.

Розробляння інтегрованої системи управління інформаційною безпекою на прикладі сфери енергетики було запропоновано в [2]. Її складники визначено з урахуванням внутрішніх і зовнішніх обставин діяльності організацій. З огляду на це інтегровано системи управління інформаційною безпекою, кібербезпекою і приватністю. Завдяки отриманому рішенню можливе забезпечення непорушності властивостей конфіденційності, приватності, цілісності, доступності інформаційних активів. У даному випадку базовим складним виокремлено систему управління інформаційною безпекою. Попри це, впровадженістю продуктів, послуг на основі штучного інтелекту зумовлено необхідність гарантування належності обробляння відповідних емерджентних ризиків. Зокрема, протидіяння негативним проявам, наприклад [3], змінювання способів розроблення, упровадження, використання і поведінки. Тому системою управління штучним інтелектом [3] пропонується розширити запропоноване в [2] інтегроване рішення [1].

Отже, ампліфікація інтегрованої системи управління інформаційною безпекою дозволить забезпечити застосовність продуктів, послуг на основі штучного інтелекту відповідно до потреб і очікувань зацікавлених сторін. І, як

наслідок, гарантувати належне обробляння неприйнятних емерджентних ризиків.

1. International Organization for Standardization. Integrated management systems. A practical guide. 2026 URL: https://www.iso.org/files/live/sites/isoorg/files/store/en/PUB100435_preview.pdf (accessed on: 26.04.2026).
2. Мохор В. В., Цуркан В. В. Інтегрована система управління інформаційною безпекою об'єктів критичної інфраструктури сфери енергетики. *Кібербезпека енергетики* : матеріали науково-практичної конференції (Київ, 27 травня 2022 р.). Київ : ІПМЕ ім. Г.Є. Пухова НАН України, 2022. С. 123–125.
3. ISO/IEC 42001:2023. Information technology. Artificial intelligence. Management system. [From 2023-12-18]. URL: https://www.iso.org/standard/ 42001 (accessed on: 26.04.2026).

Синтез сигналів управління складної форми для захищеного каналу зв'язку БПЛА

УДК 004.056: 621.39

Назарій Когут[1], Орест Синявський[2]

Національний університет "Львівська політехніка",
[1]nazarii.m.kohut@lpnu.ua, [2]orest.y.syniavskyi@lpnu.ua

Вступ. Сучасний етап розвитку безпілотних літальних апаратів (БПЛА) характеризується їх масовим застосуванням у військових та цивільних сферах. Головною умовою успішного виконання місій БПЛА є надійне та безперервне функціонування каналів управління та телеметрії. Проте радіоканали БПЛА є вразливими до навмисних завад (РЕБ), перехоплення даних та підміни сигналів управління (GPS/сигнального супфінгу). Традиційні методи захисту, такі як криптографічне шифрування, не захищають фізичний рівень зв'язку від придушення шумовими або прицільними завадами. Тому розробка методів синтезу сигналів управління складної форми, які мають високу прихованість та завадозахищеність, є актуальним науково-технічним завданням.

Аналіз останніх досліджень і публікацій. Питанням побудови завадостійких систем зв'язку присвячено роботи багатьох вітчизняних та закордонних вчених. Найчастіше для захисту каналів БПЛА використовують технології розширення спектра: псевдовипадкове переналаштування робочої частоти (ППРЧ) та прямого розширення спектра послідовністю (ПРСП) [1– 3].

Проте за умов застосування інтелектуального радіоелектронного придушення (Smart Jamming), традиційні закони формування сигналів стають прогнозованими для завадових систем противника. Потребують вдосконалення математичні моделі синтезу сигналів, які б адаптивно змінювали свою структуру у реальному часі.

Мета роботи. Підвищення завадозахищеності та імітостійкості каналу зв'язку БПЛА шляхом синтезу фазоманіпульованих та частотно-маніпульованих сигналів складної форми на основі нелінійних динамічних систем.

Виклад основного матеріалу. Для досягнення поставленої мети у дослідженні запропоновано комплексний підхід до синтезу сигналів управління, який базується на використанні хаотичних динамічних систем (зокрема, системи Лоренца, Росслера або дискретних відображень Чебишева).

Процес синтезу сигналів складної форми складається з таких етапів:

Генерація псевдовипадкових послідовностей (ПВП): замість класичних кодів Голда чи М-послідовностей застосовуються траєкторії нелінійних хаотичних відображень. Це забезпечує експоненційну чутливість до початкових умов (ключів) та збільшує ансамбль доступних сигналів.

Формування складної структури сигналу: синтез сигналів із багатопозиційною фазовою (MSK, QPSK) або частотною маніпуляцією, де параметри сигналу (фаза, частота, тривалість чипу) змінюються за нелінійним законом.

Адаптація до завадової обстановки: розроблено алгоритм динамічної зміни форми сигналу залежно від спектрального аналізу завад, що фіксуються приймачем БПЛА.

Математична модель синтезованого сигналу s(t) у загальному вигляді описується виразом:

$$s(t) = A(t)\, cos\, cos\, (2\pi\, f_0 t + \phi(t, \vec{x}) + \theta_m(t))$$

де $A(t)$ – закон амплітудної модуляції, f_0 – несуча частота, $\theta_m(t)$– інформаційна фазова маніпуляція команд управління, а $\phi(t, \vec{x})$ – додатковий складний фазовий зсув, що визначається вектором станів $\vec{x}$ хаотичної системи.

Перевагою використання таких сигналів є їхній спектр, який за своїми характеристиками наближається до білого шуму. Це забезпечує високий рівень енергетичної та структурної прихованості (LPI/LPD – Low Probability of Intercept / Low Probability of Detection). Завадові станції супротивника не можуть синхронізуватися з таким сигналом, що нівелює ефективність прицільних та імітаційних завад.

Результати моделювання. Для оцінки ефективності розроблених методів було проведено комп'ютерне моделювання у середовищі MATLAB/Simulink. Моделювався канал зв’язку БПЛА в умовах впливу флуктуаційного шуму та навмисних прицільних завад за частотою.

Результати показали, що використання синтезованих хаотичних сигналів складної форми дозволяє:

- Знизити ймовірність бітової помилки (BER) на 2.5 – 3.0 дБ при однаковому відношенні сигнал/завада порівняно з класичними системами ППРЧ.
- Збільшити структурну скритність сигналу, оскільки функція взаємної кореляції між перехопленим сигналом та копією завади не має чітко виражених піків.
- Забезпечити швидке відновлення синхронізації приймача на борту БПЛА після короткочасного повного блокування каналу зв'язку.

Висновки. У роботі запропоновано та обґрунтовано синтез сигналів управління складної форми для безпілотних літальних апаратів. Використання хаотичних динамічних систем для формування структури сигналів дозволяє

значно підвищити захищеність ліній зв'язку від засобів радіоелектронної боротьби. Складна форма сигналів унеможливлює їх прогнозування та регенерацію завадовими комплексами ворога. Запропоновані рішення можуть бути впроваджені при модернізації чинних радіоліній керування тактичних БПЛА та розробці перспективних завадозахищених систем зв'язку.

1. Кутень Р.Б., Синявський О.Ю.. Методи і засоби забезпечення стабільності та захисту радіозв’язку в умовах складної електромагнітної обстановки. *Комп’ютерні системи та мережі*. – 2024. – №1(6). – С. 99-107.
2. Костяк М.Ю., Синявський О.Ю. Синтез сигналу управління каналу зв’язку БПЛА методом динамічного програмування. *Системи управління, навігації та зв'язку.* – 2026, Том 1, № 83, – С.198-201.
3. Синявський О.Ю. Метод розрахунку чутливості каналу зв'язку БПЛА як критерію синтезу сигналів управління. *Сучасний захист інформації.* Київ. – 2026. – № 1. – С.163-168.

ITSec-2026

Кафедра кібербезпеки ТНТУ

НАУКОВЕ ВИДАННЯ

МАТЕРІАЛИ
XV Міжнародної науково-технічної конференції
«ITSec: Безпека інформаційних технологій»

27-29 травня 2026 року
м. Тернопіль (Україна)

Організаційний комітет конференції не несе відповідальності за науковий зміст, достовірність та коректність викладеної інформації (у тому числі класифікаційного індексу УДК).

Неінформативний текст матеріалів доповіді міг бути скорочений або вилучений на розсуд Оргкомітету конференції.

Оригінал-макет підготовлено на кафедрі кібербезпеки
Тернопільського національного технічного університету
імені Івана Пулюя