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.

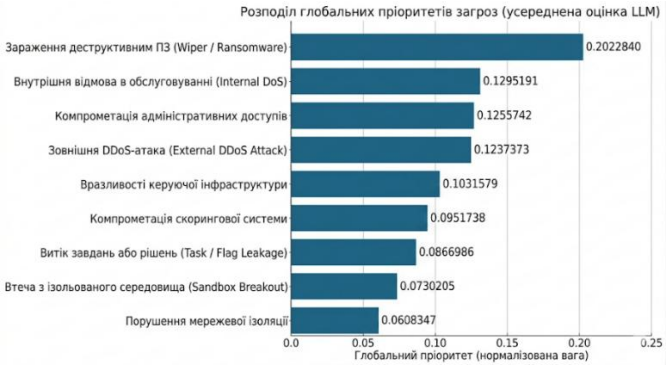

Рис.1. Узагальнена класифікація методів перехоплення інформації у СКК

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.