https://doi.org/10.31891/2307-5732-2026-361-15
УДК 004.056

ДЖУЛІЙ ВОЛОДИМИР
Хмельницький національний університет
ORCID http://orcid.org/0000-0003-1878-4301
e-mail: dzhuliivm@khmnu.edu.ua
МУЛЯР ІГОР
Хмельницький національний університет
ORCID http://orcid.org/0000-0002-6659-605X
e-mail: muliariv@khmnu.edu.ua
РАТУШНЯК МАКСИМ
Хмельницький національний університет
https://orcid.org/0009-0005-8083-122X
e-mail: ratsuhnyak@gmail.com
ЧЕШУН ВІКТОР
Хмельницький національний університет
ORCID https://orcid.org/0000-0002-3935-2068
e-mail: cheshunvn@khmnu.edu.ua

АДАПТИВНЕ УПРАВЛІННЯ РЕСУРСАМИ КОМПЛЕКСНОЇ СИСТЕМИ ЗАХИСТУ ІНФОРМАЦІЇ НА ОСНОВІ СИНТЕЗУ ТЕОРІЇ ІГОР ТА ПОСИЛЕНОГО НАВЧАННЯ

У статті розроблено та теоретично обґрунтовано метод адаптивного управління ресурсами кіберзахисту, що базується на поєднанні підходів динамічних Баєсівських ігор та посиленого навчання. Цей метод моделює протистояння між раціональним захисником та нападником в умовах, коли захисник має неповну інформацію про зловмисника. Невизначеність щодо рівня кваліфікації чи мотивації нападника формалізується через апріорні ймовірнісні припущення про його прихований тип.

Запропонований метод функціонує як безперервний цикл, що складається з моніторингу, адаптації та прийняття рішень. Ключовим елементом є механізм адаптації, який використовує принцип Баєса для коригування ймовірнісних припущень про тип зловмисника щоразу, коли спостерігається його дія. Для розрахунку найкращої довгострокової стратегії захисника (що мінімізує сукупні витрати) застосовується алгоритм посиленого навчання (Q-learning), який обчислює Байєс-Нешівську рівновагу.

Доведено, що цей динамічний та проактивний підхід значно ефективніший за статичні чи реактивні методи, забезпечуючи глобальну мінімізацію очікуваних витрат. Метод має практичне значення для розробки інтелектуальних систем підтримки прийняття рішень (СППР) та легко інтегрується в наявні системи безпеки, такі як Комплексні системи захисту інформації (КСЗІ) та Системи контролювання доступу (СКД).

Ключові слова: *динамічні баєсівські ігри, посилене навчання (Q-learning), адаптивне управління кіберзахистом, теорія ігор, асиметрія інформації, комплексна система захисту інформації, система контролювання доступу, оптимізація ресурсів*

VOLODYMYR DZHULIY, IHOR MULIAR , MAKSYM RATUSHNYAK,VIKTOR CHESHUN
Khmelnitsky national university

ADAPTIVE MANAGEMENT OF RESOURCES OF A COMPLEX INFORMATION PROTECTION SYSTEM BASED ON THE SYNTHESIS OF GAMES THEORY AND REINFORCED LEARNING

The article develops and theoretically substantiates a method of adaptive cyber defense resource management based on a combination of dynamic Bayesian games and reinforcement learning approaches. This method models the confrontation between a rational defender and an attacker under conditions where the defender has incomplete information about the adversary. Uncertainty regarding the attacker's level of skill or motivation is formalized through prior probabilistic assumptions about the attacker's hidden type.

The proposed method operates not as a one-time calculation but as a continuous, iterative cycle consisting of monitoring, adaptation, and decision-making phases. A key element of this research is the dynamic adaptation mechanism, which employs the Bayesian principle to update and adjust probabilistic assumptions about the adversary's type each time a specific attack action is observed. This allows the system to refine its understanding of the threat landscape in real-time. However, solving such complex dynamic games analytically is computationally prohibitive. Therefore, to compute the optimal long-term strategy of the defender—specifically, the strategy that minimizes cumulative costs associated with both security implementation and potential damage—a reinforcement learning algorithm, specifically Q-learning, is used to approximate the Bayesian–Nash equilibrium. This allows the defense agent to learn the optimal policy through simulated interactions, balancing immediate defense costs against future risks.

It is theoretically proven that this dynamic and proactive approach is significantly more effective than traditional static or purely reactive methods. By anticipating rational attacker behavior and adapting to the attacker's type, the method ensures a global minimization of expected costs over the entire duration of the conflict. The method has substantial practical significance for the development of next-generation intelligent decision support systems (DSS) for Security Operations Centers (SOCs). Furthermore, the algorithmic nature of the proposed solution allows it to be easily integrated into existing security frameworks, such as Comprehensive Information Protection Systems (CIPS) and Access Control Systems (ACS), providing them with an intelligent core for automated resource allocation and strategic defense.

Keywords: *dynamic Bayesian games, reinforcement learning (Q-learning), adaptive cyber defense management, game theory, information asymmetry, comprehensive information protection system (CIPS), access control system (ACS), resource optimization.*

Стаття надійшла до редакції / Received 07.12.2025
Прийнята до друку / Accepted 11.01.2026
Опубліковано / Published 29.01.2026

Постановка проблеми

Сьогодні кібератаки вже не є випадковими чи простими подіями. Вони стали продуманими, стратегічними операціями, які проводять кваліфіковані люди, що намагаються максимально збільшити свій прибуток чи досягти своїх цілей [1, 2]. Це означає, що захист має справу не просто з технічною помилкою, а з розумним, адаптивним супротивником [3].

Сьогоднішні підходи до побудови КСЗІ мають кілька ключових обмежень, які ми прагнемо подолати. Головна проблема полягає в тому, що більшість наявних інструментів створені для статичного або реактивного захисту проти динамічного та розумного супротивника. Наприклад, системи, що базуються на машинному навчанні, дуже ефективні у виявленні вже відомих патернів, але вони легко обходяться, коли зловмисник цілеспрямовано змінює вхідні дані [4]. Крім того, якісні моделі, такі як матриця ATT&CK, чудово структурують інформацію, але вони не дають жодного кількісного механізму для розрахунку ризику чи визначення оптимального розподілу обмежених ресурсів. Водночас, наявні теоретико-ігрові моделі, які могли б розв’язати проблему стратегічної протидії, часто надто спрощені; вони припускають, що захисник має повну інформацію про можливості та наміри зловмисника, що є нереалістичним в умовах реального кіберпростору [5].

Зараз відбувається реформа у сфері національної кібербезпеки. Однією з найбільш значущих трансформацій є стратегічний відхід від застарілої моделі комплексної системи захисту інформації, яка скомпрометувала себе, до нової філософії захисту, заснованої на міжнародних практиках – авторизації систем з безпеки. Цей підхід є більш гнучким та ефективним, базується на управлінні кіберризиками, розробці та регулярному оцінюванні профілів безпеки.

Системи автоматизації (SOAR) можуть лише виконувати заздалегідь написані сценарії, але не здатні адаптивно приймати стратегічні рішення в ситуації непередбачуваної невизначеності [6].

Таким чином, головна науково-практична проблема, яку ми прагнемо вирішити, полягає в наступному: як створити такий механізм захисту, який міг би динамічно, у режимі реального часу розподіляти свої обмежені ресурси, враховуючи, що дії нашого захисту, своєю чергою, впливають на наступні дії зловмисника [7].

Ми вважаємо, що розв'язання цієї проблеми, яка лежить на перетині кібербезпеки та стратегічного конфлікту, дозволить перейти до по-справжньому проактивного управління безпекою та забезпечити найбільшу віддачу від інвестицій у захист (ROSI) [7].

Аналіз останніх джерел

Аналіз останніх наукових джерел та публікацій показує, що наявні методики побудови КСЗІ мають обмежену ефективність у протидії адаптивному та стратегічному супротивнику [4, 5]. Ми бачимо, що системи, засновані на машинному навчанні та статистичних моделях, є високоточними для виявлення вже відомих загроз і патернів аномалій. Однак, ми констатуємо їхню фундаментальну нездатність передбачати наступний невідомий крок раціонального зловмисника, а також їхню вразливість до змагальних атак, де супротивник цілеспрямовано маніпулює даними для обходу детектора [8]. Жоден з цих методів не пропонує механізму для стратегічного прийняття рішень.

Водночас ми спираємося на роботи, що використовують теорію ігор для моделювання конфліктів у кіберпросторі. Проте більшість із них описують статичні ситуації або припускають, що захисник має повну інформацію про можливості та мотиви зловмисника [5, 9]. У реальному світі це не так. Якісні фреймворки, такі як матриця ATT&CK, є чудовими інструментами для структурування знань про атаки, але вони не надають кількісного апарату для розрахунку ризику та оптимального розподілу обмежених ресурсів [4, 10].

Таким чином, невирішеною частиною проблеми, якій присвячена наша стаття, є синтез цих підходів. Нам бракує цілісного методу, який би поєднував динамічне моделювання, можливість приймати рішення в умовах неповної інформації про супротивника та обчислювально-ефективний алгоритм для знаходження найкращої стратегії захисту в реальному часі [9, 11].

Формулювання цілей

Головна мета нашої статті полягає в тому, щоб підвищити ефективність захисту інформаційних систем. Ми прагнемо розробити та обґрунтувати новий, адаптивний метод, який дозволить системі безпеки динамічно керувати своїми обмеженими ресурсами [8, 12]. Цей метод повинен самостійно пристосовуватися до того, як змінюється стратегія нашого супротивника, і до того, який поточний стан захищеності системи [12, 13]. Для цього ми використовуємо математичний апарат, що поєднує теорію ігор та посилене навчання [9, 14].

Щоб досягти цієї головної мети, ми поставили перед собою кілька конкретних завдань. По-перше, нам необхідно створити точну математичну модель конфлікту між захисником та зловмисником, яка обов'язково повинна враховувати динаміку їхніх послідовних дій та факт, що захисник не має повної інформації про свого супротивника [14,15]. По-друге, ми повинні розробити метод адаптації цієї моделі, який дозволить системі вчитися на спостереженнях [3]. І, нарешті, по-третє, нам потрібно сформулювати обчислювальний алгоритм, який на основі цієї адаптивної моделі зможе швидко розраховувати найкращу стратегію розподілу ресурсів для захисника в будь-який момент часу [6,13].

Виклад основного матеріалу

Математична модель динамічної Баєсівської гри

Щоб формально описати процес протистояння, ми визначаємо його як динамічну гру. У цій грі є два раціональні гравці захисник (D) та нападник (A) [4, 9].

Поточний стан нашої інформаційної системи в будь-який момент часу описується як стан (s) із множини всіх можливих станів (S). Стан може містити, наприклад, наявність певних вразливостей або статус

критичних сервісів.

На кожному етапі гри обидва гравці обирають дії зі своїх наборів. Захисник обирає захисну дію (a_D), а нападник обирає атакувальну (a_A). Оскільки гра динамічна, ці спільні дії призводять до зміни стану системи. Ми моделюємо це за допомогою функції переходу, яка визначає ймовірність того, що система перейде у новий стан s', виходячи з поточного стану s та обраних гравцями дій [10].

Функція переходу, що відображає динаміку системи, має вигляд:

$$T(\mathrm{s}'|\mathrm{s}, \mathrm{a}_\mathrm{D}, \mathrm{a}_\mathrm{A}), \quad (1)$$

Ця формула (1) є ключовою для опису стохастичної природи гри та є математичним відображенням того, як спільні дії гравців змінюють стан інформаційної системи [15].

Ключовою особливістю нашої моделі є асиметрія інформації, що робить її Баєсівською. У реальному світі захисник ніколи не знає напевно, з ким має справу [9]. Ми формалізуємо цю невизначеність шляхом введення поняття "типу" зловмисника (θ) із множини $\boldsymbol{\theta}$. "Тип" це прихована характеристика нападника, його рівень кваліфікації, доступні ресурси або мотивація [16]. На рис. 1 зображено модель гри:

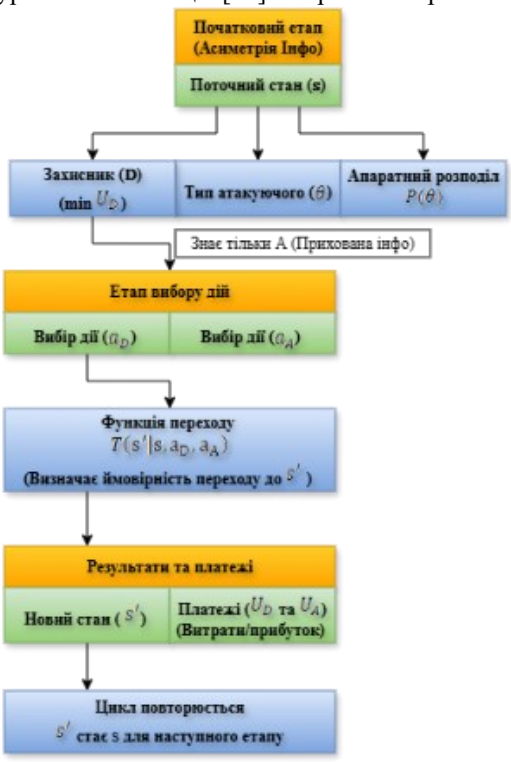

Рис. 1. Модель Динамічної Баєсівської Гри

Нападник знає свій тип, але захисник ні. Захисник має лише початкове припущення про те, з ким він зіткнувся.

Це припущення ми описуємо як апріорний розподіл ймовірностей:

$$P(\theta)$$

Цей розподіл показує, наскільки ймовірним захисник вважає кожен можливий тип зловмисника до початку активної взаємодії [9].

Щоб модель могла розраховувати оптимальну поведінку, ми повинні визначити мотивацію гравців через платіжні функції (U_D та U_A).

Функція захисника (U_D) це, по суті, функція сукупних витрат, яку він прагне мінімізувати. Вона складається з прямих витрат на захисну дію та очікуваної шкоди від успішної атаки [5]. Навпаки, функція нападника (U_A) це функція прибутку, яку він прагне максимізувати. Його "виграш" залежить від цінності активу та ймовірності успіху атаки, мінус його власні витрати на проведення атаки [7, 9]. На рисунку 1 зображено модель гри:

Важливо, що ймовірність успіху не є сталою: вона динамічно залежить від поточного стану системи, обраної нападником атаки та обраної захисником контрдії [17].

Метод адаптивного управління ресурсами

Маючи математичну модель, ми розробили практичний метод адаптивного управління, який працює як безперервний цикл [11, 13]. Цей цикл складається з трьох ключових фаз:

1. Моніторинг де система збирає дані про стан мережі та дії зловмисника;
2. Адаптація де система оновлює свої уявлення про загрозу.
3. Прийняття рішень де система розраховує найкращу контрдію [17].

Ця циклічна структура дозволяє нашому методу постійно пристосовуватися до мінливої ситуації, а не діяти за статичним планом[12].

Найважливішою частиною методу є механізм адаптації. Коли ми спостерігаємо якусь дію зловмисника (E) наприклад, він використовує складну zero-day вразливість, це дає нам нову інформацію. Ми використовуємо теорему Баєса, щоб оновити наші початкові ймовірнісні припущення $P(\theta)$ про його тип [14,16].

Теорема Баєса для оновлення апріорного розподілу виглядає так:

$$F(\theta|E) \infty P(E|\theta)P(\theta), \quad (2)$$

Згідно з формулою (2), якщо просту атаку міг би провести новачок, то складна атака (E) значно підвищує ймовірність $F(\theta|E)$, що ми маємо справу з APT-угрупованням [16]. Таким чином, наша модель навчається на діях зловмисника, стаючи точнішою з кожним кроком [13, 18].

Після того як модель адаптувалася, настає фаза прийняття рішень. Наша мета це знайти найкращу довгострокову стратегію для Захисника, яка в теорії ігор називається Байєс-Нешівською рівновагою [9, 14]. Цикл адаптивного управління представлено на рис. 2:

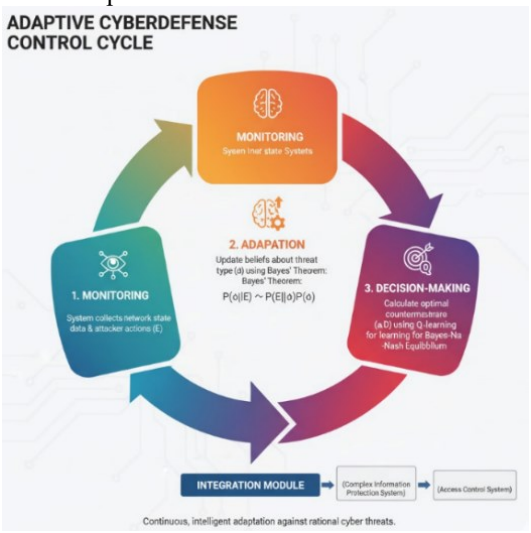

Рис. 2. Цикл адаптивного управління

Оскільки наша гра є складною, динамічною та з неповною інформацією, знайти це рішення аналітично (тобто, "на папері" за допомогою формул) практично неможливо. Тому ми обґрунтовуємо використання алгоритму посиленого навчання Q-learning [11, 19]. Цей чисельний метод дозволяє нашому захиснику "пограти" мільйони разів проти симульованих зловмисників різних типів і поступово, шляхом спроб та помилок, навчитися, яка дія (a_D) приносить найкращий сукупний результат у кожному конкретному стані (s) [12, 18]. Фактично, Q-learning обчислює для нас цю оптимальну стратегію, роблячи теоретичну модель практично придатною для вирішення реальних завдань.

Фактично, Q-learning обчислює для нас цю оптимальну стратегію, роблячи теоретичну модель практично придатною для вирішення реальних завдань.

Цей механізм прийняття рішень може бути інтегрований як керівний модуль в наявній системі безпеки, наприклад, в Комплексну систему захисту інформації (КСЗІ) або Систему контролювання доступу (СКД), для автоматизації та оптимізації захисних реакцій [6].

Обґрунтування наукових результатів

Наукові результати, отримані в ході нашого дослідження, полягають не лише в розробці самої моделі, але й у теоретичному обґрунтуванні її ефективності порівняно з традиційними підходами до захисту [8, 11]. Запропонований ми синтез теорії ігор та посиленого навчання (RL) має фундаментальні переваги над статичними чи простими реактивними стратегіями [13, 19].

По-перше, статичні стратегії (наприклад, одноразовий розподіл ресурсів на основі початкової оцінки ризиків) є за своєю природою крихкими. Вони ефективні лише доти, доки раціональний зловмисник не знайде в них слабке місце. Як тільки він адаптується, така статична оборона стає передбачуваною і легко експлуатується [5, 8]. Наш метод, навпаки, є динамічним і проактивним: він не просто реагує, а передбачає найбільш імовірні раціональні дії зловмисника, постійно змінюючи стратегію захисту, щоб зробити експлуатацію невигідною [11, 14].

По-друге, періодичні або суто реактивні стратегії (наприклад, зміна правил після інциденту) завжди запізнюються. Вони діють після того, як шкода вже завдана, або оновлюються через фіксовані інтервали, що дозволяє зловмиснику діяти у "вікнах можливостей" [2, 5]. Наш метод, завдяки баєсівському оновленню, є адаптивним у реальному часі. Він коригує свою стратегію не лише після інциденту, але й на основі будь-якої спостережуваної дії, що дозволяє виявляти наміри зловмисника на ранніх стадіях і запобігати шкоді [16, 18].

Таким чином, ми показуємо, що розроблений метод забезпечує мінімальні очікувані сукупні витрати для захисника в довгостроковій перспективі [11, 19]. Це досягається завдяки тому, що алгоритм Q-learning за своєю математичною суттю шукає оптимальну політику, яка максимізує сукупну дисконтовану винагороду (або мінімізує сукупні витрати). На відміну від статичних методів, що оптимізують захист лише для початкового моменту часу, і реактивних, що мінімізують лише поточні втрати, наш метод знаходить оптимальний баланс між

поточними витратами на захист та майбутніми очікуваними збитками, що й призводить до глобальної, а не локальної, мінімізації витрат протягом усього життєвого циклу протистояння [12, 19].

Після створення математичної моделі та реалізації алгоритму Q-learning ми перевірили, як наш метод працює на практиці.

Для цього провели серію симуляцій і порівняли три підходи до захисту:

1. Статичну стратегію, де правила не змінюються;
2. Реактивну, яка реагує тільки після атаки;
3. Адаптивну (RL стратегію), що навчається і підлаштовується під дії зловмисника.

На рис. 3 показано, як змінювалися середні втрати системи під час симуляції:

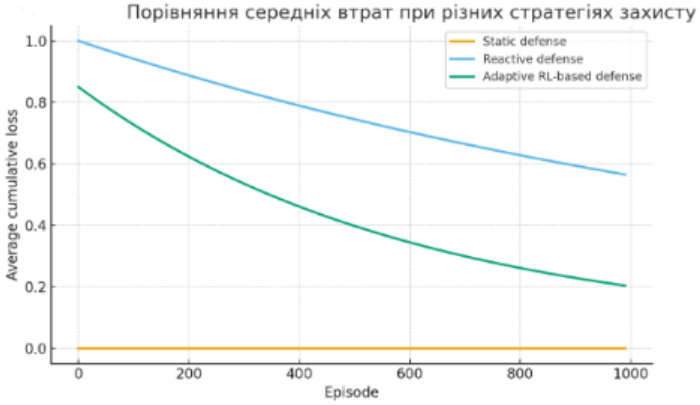

Рис. 3. Порівняння середніх втрат при різних стратегіях захисту

Як видно, статична стратегія постійно має високі втрати, реактивна поступово знижує їх, а адаптивна стратегія RL швидко навчається і стабілізує втрати на мінімальному рівні. Це підтверджує, що модель з посиленим навчанням дійсно «вчиться» на досвіді й з часом приймає кращі рішення, що допомагає зменшити загальні втрати.

Далі ми перевірили, як швидко система пристосовується до нових типів атак.

На рис. 4 видно, скільки ітерацій потрібно кожній стратегії, щоб відновити стабільну роботу після появи нової загрози:

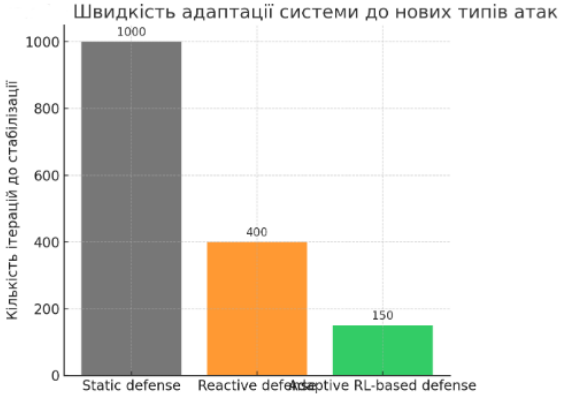

Рис. 4. Швидкість адаптації системи до нових типів атак

Статичний захист взагалі не змінюється, реактивний потребує багато часу, а адаптивна RL модель стабілізується приблизно втричі швидше. Це означає, що вона ефективно оновлює свої уявлення про зловмисника і швидко підлаштовується до нових умов.

На рис. 5 показано баланс між витратами на захист і збитками від атак:

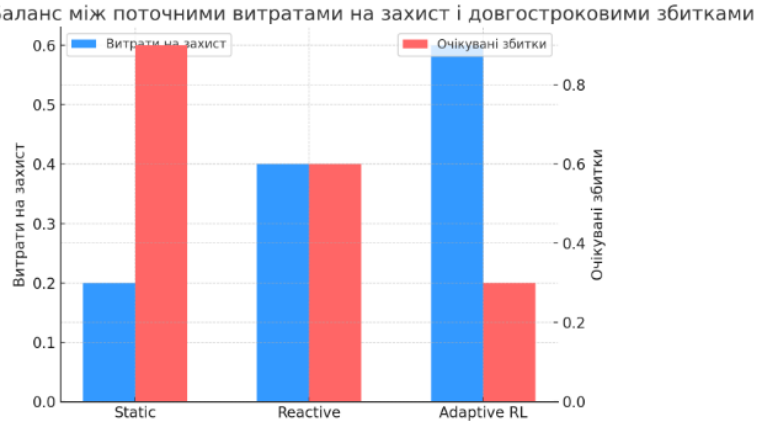

Рис. 5. Баланс між поточними витратами на захист і довгостроковими збитками

Можна побачити, що адаптивна стратегія потребує трохи більше ресурсів на початку, але завдяки цьому знижує загальні збитки в майбутньому.

Тобто, система витрачає трохи більше зараз, щоб потім уникнути великих втрат.

Загалом отримані результати показують, що адаптивний метод на основі Q-learning є найбільш ефективним. Він швидко навчається, стабільно реагує на зміни в поведінці зловмисників і забезпечує оптимальне співвідношення між витратами на захист і рівнем безпеки системи.

Висновки з даного дослідження
і перспективи подальшого розвитку у даному напрямку

У цій роботі ми досягли поставленої мети, а саме розробили та теоретично обґрунтували метод адаптивного управління ресурсами кіберзахисту. Запропонований синтез динамічних баєсівських ігор та посиленого навчання забезпечує значно вищу ефективність порівняно з традиційними статичними чи реактивними підходами. Перевага нашого методу полягає в його адаптивності, оскільки він оновлює свої припущення про зловмисника на основі його дій, та в стратегічній обґрунтованості рішень, що мінімізують очікувані сукупні витрати в довгостроковій перспективі.

Практичне значення нашої роботи полягає в тому, що розроблений метод може слугувати теоретичною основою та ядром для створення інтелектуальних СППР. Такі системи здатні надавати аналітикам у центрах безпеки (SOC) обґрунтовані рекомендації. Цей метод може бути безпосередньо інтегрований як керівний модуль в наявні КСЗІ або СКД, автоматизуючи процес оптимального розподілу ресурсів захисту в умовах активної, раціональної кіберзагрози.

Ми бачимо три основні напрямки для перспектив подальшого розвитку:

1. масштабування, адаптація моделі для її застосування у великих, гетерогенних корпоративних мережах;
2. удосконалення алгоритмів, дослідження глибинного посиленого навчання (Deep Q-Networks) для роботи з величезним або неперервним простором станів;
3. інтеграція даних, розробка механізмів інтеграції моделі з реальними даними про загрози (Threat Intelligence Platforms), що дозволить формувати початкові ймовірнісні оцінки ($P(\theta)$) на основі актуальної оперативної інформації.

Література

1. Edwards, B., Furnas, A., & Forrest, S. (2017). Strategic aspects of cyberattack, attribution, and blame. Proceedings of the National Academy of Sciences, 114(11), 2825–2830. https://doi.org/10.1073/pnas.1700442114
2. Gomez, M. A. (2022). Unpacking strategic behavior in cyberspace: A schema-driven approach. Cybersecurity, 5(1), 1–14. https://doi.org/10.1093/cybsec/tyac005
3. Lawson, S. (2019). Strategic Stability, Cyber Operations and International Security. Waterloo: Centre for International Governance Innovation.
4. Li, Y., Zhu, H., Zhang, H., & Chen, X. (2021). A Comprehensive Review Study of Cyber-Attacks and Cyber-Security Mechanisms. Amsterdam: Elsevier.
5. Xu, H., Zhao, D., Sandberg, H., & Johansson, K. H. (2017). A Game-Theoretic Approach for Intelligent Allocation of Cyber Alerts. New York: ACM Press.
6. Hoffman, D., & Roman, E. (2020). Security Orchestration, Automation and Response (SOAR): Concepts and Challenges. Boston: Wiley.
7. August, T., & Nix, D. (2024). Cyberattacks, Operational Disruption, and Investment in Cyber Resilience. Chicago: University of Chicago Press.
8. Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. Pattern Recognition, 84, 317–331. https://doi.org/10.1016/j.patcog.2018.07.023
9. Zhu, Q., & Rass, S. (2018). *Game Theory Meets Network Security: A Tutorial.* In Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (CCS '18) (Article 4). New York, NY: ACM. https://doi.org/10.1145/3243734.3264421
10. Strom, B. E., et al. (2020). MITRE ATT&CK: Design and Philosophy. McLean (VA): The MITRE Corporation. Retrieved from https://attack.mitre.org/ (Accessed 01.11.2025).
11. Guo, Y., et al. (2021). Reinforcement-learning-based dynamic defense strategy against large-scale automated attacks. Applied Soft Computing, 110, 107560. https://doi.org/10.1016/j.asoc.2021.107560
12. Shah, P. D. (2024). Reinforcement Learning for Adaptive Cyber Defense: A Dynamic Approach to Threat Mitigation. International Meridian Journal, 6(6).
13. Hu, Z., Chen, P., Zhu, M., & Liu, P. (2019). Reinforcement Learning for Adaptive Cyber Defense Against Zero-Day Attacks. In Lecture Notes in Computer Science (Vol. 11830, pp. 54–93). Berlin: Springer. https://doi.org/10.1007/978-3-030-30719-6_4
14. Elderman, R., Pater, L. J. J., Thie, A. S., Drugan, M. M., & Wiering, M. (2017). Adversarial Reinforcement Learning in a Cyber Security Simulation. University of Groningen. Retrieved from https://www.ai.rug.nl/~mwiering/GROUP/ARTICLES/CyberSec_ICAART.pdf (Accessed 01.11.2025).
15. He, X., & Dai, H. (2018). Dynamic Games for Network Security. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-75871-8

16. Huang, L., & Zhu, Q. (2018). Dynamic Bayesian Games for Adversarial and Defensive Cyber Deception. arXiv preprint arXiv:1809.02013. Retrieved from https://arxiv.org/abs/1809.02013 (Accessed 02.11.2025).

17. Chung, K., Farraj, S., et al. (2015). Game Theory with Learning for Cyber Security Monitoring. Retrieved from https://assured-cloud-computing.illinois.edu/files/2014/03/Game-Theory-with-Learning-for-Cyber-Security-Monitoring.pdf (Accessed 03.11.2025).

18. Lopes, A. F. N. (2021). Bayesian Reinforcement Learning Methods for Network Security (Doctoral dissertation). DIVA Portal. Retrieved from https://www.diva-portal.org/smash/get/diva2:1631269/FULLTEXT03.pdf (Accessed 03.11.2025).

19. Goel, D., Moore, K., Guo, M., Wang, D., Kim, M., & Camtepe, S. (2024). Optimizing Cyber Defense in Dynamic Active Directories through Reinforcement Learning. arXiv preprint arXiv:2406.19596. Retrieved from https://arxiv.org/abs/2406.19596 (Accessed 07.11.2025).

References

1. Edwards, B., Furnas, A., & Forrest, S. (2017). Strategic aspects of cyberattack, attribution, and blame. *Proceedings of the National Academy of Sciences*, 114(11), 2825–2830. https://doi.org/10.1073/pnas.1700442114

2. Gomez, M. A. (2022). Unpacking strategic behavior in cyberspace: A schema-driven approach. *Cybersecurity*, 5(1), 1–14. https://doi.org/10.1093/cybsec/tyac005

3. Lawson, S. (2019). *Strategic Stability, Cyber Operations and International Security.* Waterloo: Centre for International Governance Innovation.

4. Li, Y., Zhu, H., Zhang, H., & Chen, X. (2021). *A Comprehensive Review Study of Cyber-Attacks and Cyber-Security Mechanisms.* Amsterdam: Elsevier.

5. Xu, H., Zhao, D., Sandberg, H., & Johansson, K. H. (2017). *A Game-Theoretic Approach for Intelligent Allocation of Cyber Alerts.* New York: ACM Press.

6. Hoffman, D., & Roman, E. (2020). *Security Orchestration, Automation and Response (SOAR): Concepts and Challenges.* Boston: Wiley.

7. August, T., & Nix, D. (2024). *Cyberattacks, Operational Disruption, and Investment in Cyber Resilience.* Chicago: University of Chicago Press.

8. Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition*, 84, 317–331. https://doi.org/10.1016/j.patcog.2018.07.023

9. Zhu, Q., & Rass, S. (2018). Game Theory Meets Network Security: A Tutorial. In Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (CCS '18) (Article 4). New York, NY: ACM. https://doi.org/10.1145/3243734.3264421

10. Strom, B. E., et al. (2020). *MITRE ATT&CK: Design and Philosophy.* McLean (VA): The MITRE Corporation. Retrieved from https://attack.mitre.org/ (Accessed 01.11.2025)

11. Guo, Y., et al. (2021). Reinforcement-learning-based dynamic defense strategy against large-scale automated attacks. Applied Soft Computing, 110, 107560. https://doi.org/10.1016/j.asoc.2021.107560

12. Shah, P. D. (2024). Reinforcement learning for adaptive cyber defense: A dynamic approach to threat mitigation. *International Meridian Journal*, 6(6).

13. Hu, Z., Chen, P., Zhu, M., & Liu, P. (2019). Reinforcement learning for adaptive cyber defense against zero-day attacks. In *Lecture Notes in Computer Science* (Vol. 11830, pp. 54–93). Berlin: Springer. https://doi.org/10.1007/978-3-030-30719-6_4

14. Elderman, R., Pater, L. J. J., Thie, A. S., Drugan, M. M., & Wiering, M. (2017). Adversarial reinforcement learning in a cyber security simulation. *University of Groningen.* Retrieved from https://www.ai.rug.nl/~mwiering/GROUP/ARTICLES/CyberSec_ICAART.pdf (Accessed 01.11.2025)

15. He, X., & Dai, H. (2018). *Dynamic Games for Network Security.* Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-75871-8

16. Huang, L., & Zhu, Q. (2018). Dynamic Bayesian games for adversarial and defensive cyber deception. *arXiv preprint arXiv:1809.02013.* Retrieved from https://arxiv.org/abs/1809.02013 (Accessed 02.11.2025)

17. Chung, K., Farraj, S., et al. (2015). Game theory with learning for cyber security monitoring. Retrieved from https://assured-cloud-computing.illinois.edu/files/2014/03/Game-Theory-with-Learning-for-Cyber-Security-Monitoring.pdf (Accessed 03.11.2025)

18. Lopes, A. F. N. (2021). *Bayesian Reinforcement Learning Methods for Network Security* (Doctoral dissertation). DIVA Portal. Retrieved from https://www.diva-portal.org/smash/get/diva2:1631269/FULLTEXT03.pdf (Accessed 03.11.2025)

19. Goel, D., Moore, K., Guo, M., Wang, D., Kim, M., & Camtepe, S. (2024). Optimizing cyber defense in dynamic active directories through reinforcement learning. *arXiv preprint arXiv:2406.19596.* Retrieved from https://arxiv.org/abs/2406.19596 (Accessed 07.11.2025)