УДК 004.56

А. В. Притула
Л. М. Куперштейн

МЕТОД ОЦІНЮВАННЯ ТЕОРЕТИЧНИХ МЕЖ ЙМОВІРНОСТІ КОМПРОМЕТАЦІЇ ВЕБ-ЗАСТОСУНКІВ У ДВОАГЕНТНІЙ МАРКОВСЬКІЙ ГРІ

Вінницький національний технічний університет.

Анотація

Розглянуто задачу аналітичного оцінювання теоретичних меж ймовірності компрометації цільового вузла у двоагентній марковській грі протиборства «атака–захист», побудованій на основі MAL-графа веб-застосунку. Запропоновано метод, що поєднує аналітичне виведення верхньої та нижньої меж із емпіричною перевіркою методом Монте-Карло на основі нерівності Хефдінга. Показано, що при наявності вузлів із нульовим оборонним атрибутом нижня межа дорівнює одиниці незалежно від політики захисника, а у разі ненульових атрибутів ймовірність компрометації затухає експоненційно зі зростанням захисного параметра. Отримані аналітичні оцінки для двох навчальних середовищ узгоджуються з емпіричними значеннями, отриманими для статичних політик опонентів, що підтверджує коректність запропонованого підходу. Метод придатний для перевірки політик, синтезованих методами навчання з підкріпленням, та для побудови довірчих інтервалів ймовірності компрометації з гарантованою точністю.

Ключові слова: марковська гра, MAL-граф, ймовірність компрометації, нерівність Хефдінга, метод Монте-Карло, навчання з підкріпленням, кібербезпека, веб-застосунок.

Abstract

The problem of analytical estimation of theoretical bounds of the target node compromise probability in a two-agent attack–defense Markov game built on a MAL graph of a web application is considered. A method is proposed that combines analytical derivation of upper and lower bounds with empirical verification using the Monte Carlo method based on the Hoeffding inequality. It is shown that in the presence of nodes with a zero defense attribute the lower bound equals one regardless of the defender's policy, while in the case of non-zero attributes the compromise probability decays exponentially as the defense parameter grows. The analytical estimates obtained for two instances of the training environment agree with the empirical values for static opponent policies, which confirms the correctness of the proposed approach. The method is suitable for verification of policies synthesized by reinforcement learning methods and for constructing confidence intervals of the compromise probability with guaranteed accuracy.

Keywords: Markov game, MAL graph, compromise probability, Hoeffding inequality, Monte Carlo method, reinforcement learning, cybersecurity, web application.

Вступ

Сучасні методи оцінювання захищеності веб-застосунків дедалі частіше спираються на формальне моделювання сценаріїв атак з використанням теорії марковських процесів прийняття рішень та методів навчання з підкріпленням [1-3]. Такий підхід дає змогу описати протистояння зловмисника й захисника як двоагентну марковську гру, у якій кожен агент будує власну політику дій з метою максимізації відповідної функції винагороди. Особливе значення в межах цієї парадигми мають графи моделювання атак, побудовані за допомогою метамови MAL [4], оскільки вони формалізують структуру можливих шляхів компрометації і дозволяють параметризувати ризик через метрику часу до компрометації (TTC) [5].

Незважаючи на значний прогрес у синтезі політик методами глибокого навчання з підкріпленням [6, 7], залишається відкритою принципова проблема – відсутність аналітичного апарату, який би дозволяв незалежно перевірити коректність отриманих емпіричних оцінок. Без таких меж результати моделювання не можна вважати достовірними, оскільки існує ризик появи артефактів моделі або переоцінки ефективності політик через особливості вибірки. У роботах [6, 8] вказано на необхідність побудови верифікаційних інструментів, які поєднують аналітичні гарантії з емпіричною перевіркою у

формі стохастичних запусків середовища.

Метою цієї роботи є розробка методу аналітичного оцінювання верхньої та нижньої меж ймовірності компрометації цільового вузла в двоагентній марковській грі протиборства «атака–захист», а також процедури їх емпіричної перевірки з гарантованою точністю на основі нерівності Хефдінга.

Результати дослідження

Розглядається двоагентна марковська гра $\mathcal{G} = (S, A^A, A^D, P, R)$ [9], де S – множина станів системи, A_A – множина дій агента зловмисника, A_D – множина дій захисного агента, $P: S \times A^A \times A^D \to [0,1]$ – функція переходів, яка задає ймовірність переходу системи з поточного стану в наступний за умови виконання обома агентами своїх дій, $R = (R_A, R_D)$ – двобічна функція винагороди, що визначає виграш зловмисника й захисника на кожному кроці взаємодії. Гра розгортається у межах графа інфраструктури $G = (N, E)$, де N – множина вузлів, що репрезентують компоненти веб-застосунку (бази даних, сервери-посередники, міжмережеві екрани, сервери прикладного рівня, сервери кешування та інші елементи інформаційно-комунікаційної системи), а E – множина ребер, які задають топологічні зв'язки між цими компонентами.

Кожен вузол N_i описується парою станозалежних векторів $\left(S_i^A, S_i^D\right)$. Вектор атаки $S_i^A = \left(a_i^1, \dots, a_i^k\right)$ кодує накопичену силу k типів атак (відмова в обслуговуванні, міжсайтове виконання сценаріїв, SQL-ін'єкція, фішинг тощо). Оборонний вектор S_i^D містить силу відповідних оборонних механізмів та додаткову компоненту b_i^d – здатність вузла виявляти спроби компрометації. Усі компоненти беруть значення в дискретному діапазоні $\{0, 1, \dots, b_{max}\}$. Зловмисник у кожному раунді обирає одну з двох дій над довільним видимим вузлом: розвідку, що відкриває компоненти оборонного вектора суміжних вузлів, або атаку типу j, яка інкрементує відповідний атакувальний атрибут на одиницю. Атака вважається успішною, якщо накопичена сила атаки досягає або перевищує рівень оборони. У разі невдалого кроку спробу буде виявлено з ймовірністю, пропорційною здатності вузла до виявлення, що моделюється бернуллівським випробуванням. Захисник у свою чергу обирає або моніторинг, який інкрементує здатність до виявлення, або посилення певного оборонного атрибута.

Гра завершується перемогою зловмисника, якщо скомпрометовано цільовий вузол N_{data}, або перемогою захисника, якщо вторгнення виявлено. Ключовими метриками для подальшого аналізу є ймовірність компрометації цільового вузла до моменту T, позначувана P_{comp}, і очікуване значення TTC [8].

Для аналітичного оцінювання ймовірності компрометації будується спільний розподіл часу, потрібного для зловмисника, аби скомпрометувати кожен вузол на шляху до цільового, за умови оптимальної політики. Припускається, що момент виявлення зловмисника під час спроби атаки типу j на вузол N_i описується геометричним розподілом з параметром b_i^d. Таке припущення відображає дискретну природу процесу, у якому кожна спроба атаки розглядається як незалежна подія з ймовірністю виявлення, пропорційною поточному рівню оборонного атрибута. Зворотний рух зловмисника шляхами вже пройдених вузлів виключається, що відповідає монотонному характеру накопичення компрометації.

За цих припущень ймовірність успішної компрометації одного вузла набуває вигляду:

$$P_{break}(b) = \exp(-c \cdot b), \quad c > 0, \tag{1}$$

де b – поточне значення відповідного оборонного атрибута, а c – коефіцієнт масштабування, визначений вибором геометричного розподілу. З виразу (1) випливає експоненційний характер залежності ймовірності зламу від рівня захисту. Кожна додаткова одиниця оборонного атрибута зменшує ймовірність успішної атаки в сталу кількість разів e^c, що вздовж шляху дає геометричне згасання сукупної ймовірності компрометації..

Верхня межа ймовірності компрометації відповідає найоптимістичнішому сценарію для зловмисника. Вона визначається як максимум добутку локальних ймовірностей успіху вздовж усіх допустимих шляхів від стартового вузла N_{start} до цільового N_{data}:

$$P_{max} = \max_{P \in \mathcal{P}} \prod_{e \in P} p_e \tag{2}$$

де $\mathcal{P}$ – множина всіх допустимих шляхів від стартового вузла N_{start} до цільового N_{data}, а p_e – ймовірність успішного виконання локальної атакувальної дії, асоційованої з ребром e. У разі введення

механізмів повторних спроб або затримок верхня межа уточнюється додатковими множниками $(1 - p_{detect})^k$, які враховують ймовірність невиявлення.

Нижня межа ймовірності компрометації характеризує найгірший для зловмисника сценарій і реалізується за умови застосування захисником оптимальної політики – мінімального захисника, який щораунду підсилює найслабший оборонний атрибут. Оскільки за один раунд захисник змінює лише один атрибут, а дії обох агентів виконуються одночасно, наявність досяжного вузла з нульовим оборонним атрибутом ($b_i^j = 0$ для деякого j), який зловмисник може атакувати раніше, ніж захисник устигне його підсилити, дає ймовірність компрометації, рівну одиниці. За виразом (1) така атака має успіх із певністю, оскільки $\exp(0) = 1$. Цей граничний випадок інтерпретується як «точка відмови» й не усувається жодною політикою захисту, доки непокритий вузол лишається досяжним для атаки в межах одного раунду.

У протилежному сценарії, коли всі оборонні атрибути додатні, нижня межа описується експоненційним згасанням і асимптотично прямує до нуля при зростанні захисних параметрів, що формалізує наближення системи до стану практичної безпечності.

Обидві межі обчислюються підстановкою локальних імовірностей (1) у вираз (2) для репрезентативних значень оборонних атрибутів. Для базового середовища з однорідним розподілом захисту кожен проміжний вузол має значення атрибута, за якого згідно з (1) ймовірність зламу становить ≈ 0.6. Оскільки оптимальний шлях до цільового вузла містить два проміжні вузли, верхня межа дорівнює $Pmax \approx 0.6 \cdot 0.6 = 0.36$. У розширеному середовищі стартовий вузол обирається стохастично, тому довжина оптимального шляху змінюється: найвіддаленіший старт зберігає двовузловий шлях із $P_{max} \approx 0.36$, тоді як найближчий старт відповідає одновузловому шляху з $P_{max} \approx 0.56$. Сукупність допустимих стартів задає діапазон значень верхньої межі від 0.36 до 0.56.

Аналітично виведені межі потребують емпіричної перевірки в реальних запусках середовища під заданими політиками опонентів. Для цього застосовується метод Монте-Карло [10] з гарантованою точністю на основі нерівності Хефдінга [11]. Нехай $X_m \in \{0, 1\}$ – індикаторна змінна, що дорівнює одиниці в разі успішної компрометації цільового вузла в епізоді m, і нулю в іншому разі. Тоді емпірична оцінка ймовірності компрометації визначається як середнє арифметичне:

$$\widehat{\mathrm{P}}_{\mathrm{comp}} = \left(\frac{1}{\mathrm{M}}\right) \cdot \sum_{m=1}^{M} \mathrm{X_m} \tag{3}$$

де M – кількість незалежних епізодів. За законом великих чисел значення $\hat{P}_{comp}$ збігається до істинної ймовірності P_{comp} при необмеженому зростанні M, проте для гарантованої похибки не більше ε з довірчою ймовірністю не нижче $1 - \delta$ необхідне число епізодів визначається з нерівності Хефдінга:

$$M \geq \left(\frac{1}{2\varepsilon^2}\right) \cdot \ln\left(\frac{2}{\delta}\right). \tag{4}$$

Наприклад, для $\varepsilon = 0.02$ і $\delta = 0.05$ отримуємо $M \geq 4612$. Таке обмеження забезпечує відтворюваність експерименту і дозволяє інтерпретувати отримане значення $\hat{P}_{comp}$ як кількісно обґрунтовану оцінку, не залежну від випадкових коливань вибірки.

Для побудови довірчого інтервалу ймовірності компрометації як біноміальної частки використовується метод Клоппера–Пірсона [12], що ґрунтується на бета-розподілі та забезпечує точне покриття за малих значень розміру вибірки, або метод Вілсона [13], який дає вужчі інтервали при помірних обсягах. Перевагою цих процедур над класичним нормальним наближенням є коректне поводження поблизу меж $\{0, 1\}$, що особливо важливо для сценаріїв з низькою або високою ймовірністю компрометації.

У стохастичному режимі тестування дії агентів вибираються з відповідних розподілів політики, тоді як у детермінованому режимі застосовується точка максимуму функції цінності. Зіставлення емпіричних оцінок у цих двох режимах дозволяє визначити характер навченої політики: збіг значень свідчить про детермінованість і стійкість стратегії, тоді як істотні розбіжності вказують на її стохастичний характер або на недостатню збіжність процесу навчання.

Узгодження теоретичних та емпіричних результатів реалізується через перевірку нерівності $P_{min} \leq \hat{P}_{comp} \leq P_{max}$ та входження довірчого інтервалу в аналітично визначений діапазон. Якщо емпірична оцінка опиняється нижче P_{min} це свідчить про некоректність моделі або помилку реалізації, оскільки порушуються необхідні умови досягнення цільового стану. Якщо ж $\hat{P}_{comp}$ перевищує P_{max}, то

ймовірно йдеться про артефакти моделювання – наприклад, про неконтрольовані витоки інформації між агентом і середовищем.

Для базового навчального середовища під політиками мінімального захисника та максимального зловмисника аналітично отримана верхня межа ≈ 0.36 узгоджується з емпіричною оцінкою $\hat{P}_{comp} = 0.358 \pm 0.014$ (за $M = 5000$, рівень довіри 95%). Для розширеного середовища з випадковим стартом отримано $\hat{P}_{comp} = 0.471 \pm 0.015$, що також лежить у межах теоретичного діапазону приблизно $[0.36; 0.56]$. Узгодженість у обох випадках підтверджує коректність побудованого аналітичного апарату й придатність процедури для подальшого використання як еталонного інструменту перевірки політик, синтезованих методами навчання з підкріпленням.

Висновки

Запропоновано метод оцінювання теоретичних меж ймовірності компрометації цільового вузла в двоагентній марковській грі протиборства «атака–захист». Метод поєднує аналітичне виведення верхньої та нижньої меж на основі геометричного розподілу моменту виявлення з емпіричною перевіркою методом Монте-Карло, що ґрунтується на нерівності Хефдінга та довірчих інтервалах Клоппера–Пірсона або Вілсона.

Показано, що наявність вузла з нульовим оборонним атрибутом призводить до нижньої межі, рівної одиниці, тобто гарантованої компрометації незалежно від політики захисника. У разі ненульових атрибутів ймовірність компрометації затухає експоненційно зі зростанням захисних параметрів. Отримані числові оцінки для базового та розширеного навчальних середовищ узгоджуються з результатами стохастичних запусків під статичними політиками опонентів.

Запропонований апарат придатний для перевірки політик, синтезованих методами навчання з підкріпленням: політика, що демонструє результат нижче нижньої межі, свідчить про некоректність моделі, тоді як перевищення верхньої межі вказує на артефакти моделювання. Інтервал між нижньою та верхньою межами визначає допустимий простір навчання, а положення оцінки всередині нього слугує надійним кількісним показником ефективності.

Подальші дослідження доцільно спрямувати на узагальнення методу для багатоагентних ігор з частковою спостережуваністю, інтеграцію з методами байєсівського навчання з підкріпленням [14] та адаптацію до сценаріїв із динамічними оборонними атрибутами, що змінюються в процесі взаємодії агентів із середовищем.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Sutton R. S. Reinforcement Learning: An Introduction / R. S. Sutton, A. G. Barto. – 2nd ed. – Cambridge, MA : MIT Press, 2018. – 552 p.
2. Притула А. Аналіз підходів тестування на проникнення з використанням машинного навчання з підкріпленням [Електронний ресурс] / А. Притула, Л. Куперштейн // Кібербезпека: освіта, наука, техніка. – 2025. – Т. 4, № 28. – С. 259–271. – Режим доступу: https://doi.org/10.28925/2663-4023.2025.28.789 (дата звернення: 13.06.2026). – Назва з екрана.
3. Клейн О. Формальні моделі комп'ютерних атак в корпоративних мережах [Електронний ресурс] / О. Клейн // Вимірювальна та обчислювальна техніка в технологічних процесах. – 2025. – Т. 84, № 4. – С. 81–88. – Режим доступу: https://doi.org/10.31891/2219-9365-2025-84-9 (дата звернення: 13.06.2026). – Назва з екрана.
4. Johnson P. A Meta Language for Threat Modeling and Attack Simulations [Електронний ресурс] / P. Johnson, R. Lagerström, M. Ekstedt // Proceedings of the 13th International Conference on Availability, Reliability and Security. – [Б. м.] : ACM, 2018. – P. 1–8. – Режим доступу: https://doi.org/10.1145/3230833.3232799 (дата звернення: 13.06.2026). – Назва з екрана.
5. Leversage D. J. Estimating a System's Mean Time-to-Compromise [Електронний ресурс] / D. J. Leversage, E. J. Byres // IEEE Security & Privacy. – 2008. – Vol. 6, No. 1. – P. 52–60. – Режим доступу: https://doi.org/10.1109/MSP.2008.9 (дата звернення: 13–15.06.2026). – Назва з екрана.
6. Kim B.-S. Optimal Cyber Attack Strategy Using Reinforcement Learning Based on Common Vulnerability Scoring System [Електронний ресурс] / B.-S. Kim, H.-W. Suk, Y.-H. Choi, D.-S. Moon, M.-S. Kim // Computer Modeling in Engineering & Sciences. – 2024. – Vol. 141, No. 2. – P. 1551–1574. – Режим доступу: https://doi.org/10.32604/cmes.2024.052375 (дата звернення: 13–15.06.2026). – Назва з екрана.
7. Yousefi M. A Reinforcement Learning Approach for Attack Graph Analysis [Електронний ресурс] / M. Yousefi, N. Mtetwa, Y. Zhang, H. Tianfield // 17th IEEE Conference on Trust, Security and Privacy in Computing and Communications. – [Б. м.] : IEEE, 2018. – P. 212–217. – Режим доступу: https://doi.org/10.1109/TrustCom/BigDataSE.2018.00041 (дата звернення: 14.06.2026). – Назва з екрана.
8. Притула А. Моделювання сценаріїв кібератак як марковського процесу із семантично обмеженим простором дій [Електронний ресурс] / А. Притула, Л. Куперштейн // Кібербезпека: освіта, наука, техніка. – 2025. – Препринт. – Режим доступу: https://doi.org/10.13140/RG.2.2.24871.92328 (дата звернення: 14.06.2026). – Назва з екрана.

9. Littman M. L. Markov games as a framework for multi-agent reinforcement learning [Електронний ресурс] / M. L. Littman // Proceedings of the 11th International Conference on Machine Learning. – [Б. м.] : Morgan Kaufmann, 1994. – P. 157–163. – Режим доступу: https://doi.org/10.1016/B978-1-55860-335-6.50027-1 (дата звернення: 13–15.06.2026). – Назва з екрана.
10. Карташов М. В. Імовірність, процеси, статистика : посібник / М. В. Карташов. – Київ : ВПЦ «Київський університет», 2008. – 494 с.
11. Hoeffding W. Probability Inequalities for Sums of Bounded Random Variables [Електронний ресурс] / W. Hoeffding // Journal of the American Statistical Association. – 1963. – Vol. 58, No. 301. – P. 13–30. – Режим доступу: https://doi.org/10.1080/01621459.1963.10500830 (дата звернення: 15.06.2026). – Назва з екрана.
12. Clopper C. J. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial [Електронний ресурс] / C. J. Clopper, E. S. Pearson // Biometrika. – 1934. – Vol. 26, No. 4. – P. 404–413. – Режим доступу: https://doi.org/10.1093/biomet/26.4.404 (дата звернення: 13–15.06.2026). – Назва з екрана.
13. Wilson E. B. Probable Inference, the Law of Succession, and Statistical Inference [Електронний ресурс] / E. B. Wilson // Journal of the American Statistical Association. – 1927. – Vol. 22, No. 158. – P. 209–212. – Режим доступу: https://doi.org/10.1080/01621459.1927.10502953 (дата звернення: 15.06.2026). – Назва з екрана.
14. Ghavamzadeh M. Bayesian Reinforcement Learning: A Survey [Електронний ресурс] / M. Ghavamzadeh, S. Mannor, J. Pineau, A. Tamar // Foundations and Trends in Machine Learning. – 2015. – Vol. 8, No. 5–6. – P. 359–483. – Режим доступу: https://doi.org/10.1561/2200000049 (дата звернення: 15.06.2026). – Назва з екрана.

Притула Андрій Вікторович – студент групи 125-23а, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: andrik.pritula@gmail.com.
Купершейн Леонід Михайлович – к.т.н., доцент кафедри захисту інформації, Вінницький національний технічний університет, м. Вінниця email: kupershtein.lm@gmail.com

Prytula Andrii V. – Student of Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, e-mail: andrik.pritula@gmail.com.
Kupershtein Leonid M. – PhD, Associated Professor of Information Protection Chair, Vinnytsia National Technical University, Vinnytsia, email: kupershtein.lm@gmail.com