УДК 004.8:004.056
DOI https://doi.org/10.32782/2663-5941/2026.1.2/54

Шпита Г.В.
https://orcid.org/0009-0005-0995-1210
Державний університет «Житомирська політехніка»

Вакалюк Т.А.
https://orcid.org/0000-0001-6825-4697
Державний університет «Житомирська політехніка»

СТАН ДОСЛІДЖЕНЬ У ГАЛУЗІ ЗАСТОСУВАННЯ МАШИННОГО НАВЧАННЯ З ПІДКРІПЛЕННЯМ ДЛЯ ТЕСТУВАННЯ ЗАХИЩЕНОСТІ ІНФОРМАЦІЙНИХ СИСТЕМ

Проблема виявлення вразливостей в інформаційних системах набуває все більшої актуальності в умовах постійного розвитку стратегій кібератак. Традиційні методи перевірки на проникнення, хоч і вдаються до використання автоматизованих рішень, все ще потребують значних людських ресурсів, що створює потребу у пошуку нових підходів. Нещодавні дослідження демонструють, що застосування машинного навчання дозволяє певною мірою автоматизувати окремі фази процесу перевірки на проникнення, але дані системи мають і недоліки. Оскільки навчання системи виконується на попередньо підготовленій виборці, агенти зазвичай можуть виявляти лише вже відомі вразливості. Дана проблема свідчить про те, що для ефективного застосування інтелектуальні алгоритми мають адаптуватися до динамічних умов, продовжуючи процес навчання під час практичного застосування. Для вирішення задач з подібними вимогами часто застосовується навчання з підкріпленням та його різновиди. Дослідники виділяють даний підхід як перспективний засіб вирішення проблеми пошуку потенційних вразливостей навіть якщо інформації про них не було у навчальній вибірці. В даній роботі розглянуто сучасні підходи до тренування агентів за допомогою навчання з підкріпленням для виконання перевірки на проникнення. Нещодавні експерименти із застосуванням симуляторів перевірки на проникнення демонструють, що моделювання атак через частково спостережувані марковські процеси прийняття рішень та використання рекурентних ієрархічних підходів та актор-критик алгоритмів підвищує ефективність агентів у складних середовищах з неповною інформацією. Також, систематичні огляди підтверджують існування тенденції переходу від таких класичних алгоритмів як Q-learning і DQN до більш складних моделей, здатних працювати у багатовимірних просторах дій. Використання даних методів має потенціал до суттєвого підвищення ефективності процесу перевірки на проникнення навіть в умовах динамічних середовищ.

Ключові слова: *алгоритм, машинне навчання, методи навчання з підкріпленням, перевірки на проникнення, пошук вразливостей, інформаційна система.*

Постановка проблеми. В умовах зростання залежності суспільства від інформаційних технологій, виявлення вразливостей у комп'ютерних мережах залишаються актуальними проблемами, оскільки із розвитком інформаційних систем, покращуються і засоби подолання їх захисту. Для запобігання потенційних хакерських атак спеціалісти вдаються до процесу перевірки на проникнення – моделювання можливих дій зловмисників з метою завчасного пошуку наявних вразливостей та їх вилучення. Останніми роками питання автоматизації процесів перевірки на проникнення набуває все більшого значення, оскільки масштаби та складність сучасних інформаційних систем, а також збільшення варіативності кіберзагроз роблять виключно ручні підходи малоефективними і економічно нераціональними. Проте, не дивлячись на наявність великого різноманіття інструментів для автоматизованого пошуку відомих вразливостей, їх ефективність і точність результатів також обмежуються експертизою спеціаліста з кібербезпеки, адже навіть автоматичні

звіти потребують фільтрації, оцінювання, перевірки на хибнопозитивні спрацьовування і пріоритизації виявлених проблем. Це формує потребу в розробці інтелектуальних систем, здатних автоматично досліджувати мережі, виявляти слабкі місця та змінювати стратегію дій у залежності від контексту. Тим не менш, хоча використання машинного навчання і виглядає як логічний крок уперед, варто враховувати, що класичні алгоритми машинного навчання потребують використання тренувальних вибірок. Це ставить під сумнів використання традиційних підходів, адже швидкий розвиток методів проникнення створює ризик того, що вибірка, застосована для навчання, може відносно швидко втратити свою актуальність. У звя'зку з цим постає проблема вибору методу машинного навчання, який зміг би задовільнити потребу у адаптивності системи пошуку вразливостей в умовах як постійного розвитку мережевих інфраструктур, так і постійного виникнення нових типів вразливостей.

Аналіз останніх досліджень і публікацій. Проблема автоматизації процесів перевірки на проникнення за допомогою методів машинного навчання з підкріпленням розглядається у низці сучасних робіт. Її актуальність ґрунтується на низький практичній ефективності класичних алгоритмів машинного навчання, натренованих на базах даних вразливостей із відкритого доступу [1]. Серед найновіших досліджень варто відзначити створення тренувальних середовищ для агентів, таких як PenGym [2] а також застосування динамічних мереж при навчанні [3]. Значна увага приділяється моделюванню процесу пошуку вразливостей як задачі частково спостережуваного процесу Маркова [4], ієрархічним методам навчання для обмеження простору дій агента [5, 6], а також автоматизації фаз пост-експлуатації [6]. Узагальнюючі огляди [11, 12] фіксують тенденцію поступового відходу від базових алгоритмів (Q-learning, DQN, DDQN) до використання «актор-критик» підходів (A2C, PPO, SAC), рекурентних ієрархічних моделей та їх комбінацій із графовими стратегіями атак.

Постановка завдання. Метою статті є аналіз сучасних моделей та методів автоматизації тестування на проникнення із застосуванням навчання з підкріпленням, систематизація наявних підходів у даній сфері та визначення їхніх переваг і обмежень.

Виклад основного матеріалу. У порівнянні із традиційними методами автоматизації перевірки на проникнення, застосування машинного навчання виглядає як ефективна альтернатива, адже моделі здатні опрацьовувати значні масиви даних та виявляти неочевидні закономірності, пошук яких не обов'язково наявно передбачався при їх тренуванні. Проте застосування машинного навчання для пошуку вразливостей на практиці все ще має значні обмеження. По-перше для тренування нейромереж необхідні набори даних, які часто бувають неповними та незбалансованими, що може призводити до перетренування і, відповідно, пониженій здатності моделі до розпізнання певних класів вразливостей. Дослідження показали, що моделі, натреновані на стандартних наборах даних із відкритого доступу, можуть мати низьку ефективність у пошуку вразливостей в реальних практичних умовах [1]. По-друге, обмеження навчанням на історичних вибірках загальних вразливостей та експозицій (Common Vulnerabilities and Exposures, CVE) призводить до врахування лише задокументованих векторів атак, у той час, коли подібні набори даних можуть відносно швидко стати застарілими і не репрезентативними в умовах постійного виникнення нових методів проникнення. Таким чином, тренування нейромереж на базах даних із відомими вразливостями не вирішує проблему потреби у врахуванні як постійного розвитку інформаційних систем, так і появи нових векторів атак.

Постає вибір між застосуванням класичних методів машинного навчання, що вимагатимуть періодичного перетренування на оновлених та спеціально збалансованих баз даних CVE, або ж комбінуванням алгоритмів навчання з адаптивними механізмами. Постановка задачі у вигляді представлення деякого середовища, взаємодіючи із яким інтелектуальний агент має адаптуватися та знаходити потенційні засоби проникнення у систему наштовхує на використання навчання з підкріпленням (Reinforcement Learning, RL) як можливого рішення. Представлення побудови стратегії атаки у вигляді марковського процесу вирішування і нагородження агента за успішні вектори атак має потенціал нівелювати потребу у попередньому повному знанні про можливі засоби проникнення, саме тому останнім часом даний підхід привертає все більше уваги.

Імплементацію даного механізму доцільно розглядати як двоетапний процес, що включає створення експериментального середовища для отримання репрезентативних даних та реалізацію агента, що шукає вразливості у середовищі. Виклики першої фази пов'язані з етичною стороною питання, що накладає серйозні обмеження на

методи тестування розроблених систем, а різноманітність наявних конфігурацій інфраструктур і векторів атак суттєво ускладнює порівняльний аналіз створених агентів. Таким чином, постає питання вибору моделі середовища, встановлення стандартів для експериментів та вибору критеріїв ефективності для порівняльного аналізу.

Одним із важливих кроків у цьому напрямку стало створення спеціалізованих симуляторів перевірки на проникнення для тренування інтелектуальних систем, що суттєво спростило підготовку та налаштування середовища для навчання агентів. Так, Нгуєн Г. В. (Nguyen H.V.) та ін. [2] розробили PenGym – платформу, що моделює дії перевірки на проникнення у віртуальному середовищі та сумісна з API OpenAI Gym. Це рішення спрощує інтеграцію з популярними бібліотеками RL і створює можливість стандартизації експериментів. У проведених тестах агенти, натреновані в PenGym, досягали поставлених цілей у середньому за 7,72 кроків, тоді як у традиційних симуляційних середовищах цей показник складав 11,95 кроків [2]. Автори при цьому наголошують на потребі у спільних метриках і орієнтирах (шаблонах), оскільки ізольованість окремих середовищ обмежує відтворюваність та порівнянність результатів.

Що стосується самого алгоритму дій системи, питання адаптації RL-агентів до динамічних умов розглянули Лі Ц. (Li C.) та ін. [3], які у своїй роботі запропонували метод DynPen. Його основна особливість полягає у врахуванні змін у топології симульованої мережі, як, наприклад, поява чи зникнення вузлів і зміна зв'язків між ними. Для цього RL-агент було інтегровано разом з графовими моделями атак із застосуванням збереження інформації про попередні стани. У симуляціях кібератак на мережі розміром до десяти вузлів підхід продемонстрував швидшу збіжність та вищу адаптивність у порівнянні з моделями, що не враховують динаміку, але автори підкреслюють, що масштабування цього рішення на великі інфраструктури залишається відкритою проблемою.

Оскільки агент не повинен мати повного доступу до інформації про систему під час перевірки на проникнення, дослідники інколи моделюють завдання як частково спостережуваний марковський процес (Partially Observable Markov Decision Process, POMDP). Лю С. (Liu X.) та ін. [4] показали, що інтеграція рекурентних нейронних мереж (Recurrent Neural Network, RNN) у RL-модель дозволяє агенту враховувати послідовність дій і тим самим приймати більш обґрунтовані рішення в умовах невизначеності. У їхніх експериментах використання RNN забезпечило більшу частину успішних проникнень і скорочення середньої кількості кроків до мети.

Також слід відмітити підхід спрямований на зменшення простору можливих дій агента. Лю Г. (Liu H.) та ін. [5] запропонували ієрархічну архітектуру, де верхній рівень відповідає за навігацію мережею, а нижній – за прийняття локальних рішень. Завдяки механізму маскування дій їм вдалося усувати недопустимі дії із таблиці рішень, що пришвидшило збіжність і знизило кількість невдалих спроб у порівнянні з тестами без використання маскування. Схожий напрямок розвивали Ганем М. Ц. (Ghanem M.C.) та ін. [6], які продемонстрували ефективність ієрархічного навчання з підкріпленням (Hierarchical Reinforcement Learning, HRL) у середовищах, де агент має неповне знання топології системи, на яку виконується атака. Їхні результати підтверджують, що розбиття завдання на підзадачі дозволяє скоротити обчислювальні витрати та масштабувати експерименти на більші та складніші мережі.

Наведені методи покриваються питання тестів на вторгнення у мережу, що є лише частиною процесу пошуку вразливостей. Автоматизація фаз пост-експлуатації стала предметом дослідження Фам В.-Г. (Pham V.-H.) та ін. [7], які представили систему Raijū. У даній системі було інтегровано RL-агента разом з фреймворком перевірки на проникнення Metasploit, що дало змогу відпрацювати дії на кшталт ескалації привілеїв, латерального переміщення та захоплення нових вузлів. Для навчання агента було використано алгоритм «актор-критик із оцінкою переваги» (Advantage Actor-Critic, A2C) і проксимальну оптимізацію політики (Proximal Policy Optimization, PPO), що забезпечило достатню адаптивність політик для різних конфігурацій захисту. Однак автори визнають, що перевірка у масштабних та різнорідних середовищах є необхідною умовою подальшої валідності результатів. У дослідженні Аль-Маджалі А. (AlMajali A.) та ін. [8] сфокусувалися на автоматизації експлуатації вразливостей із застосуванням глибинного навчання з підкріпленням. Запропонований ними підхід скоротив середній час і кількість кроків до успішної атаки та зменшив кількість помилкових дій порівняно з базовими стратегіями. На відміну від рішень, орієнтованих на окремі випадки, метод виявився відносно універсальним і придатним до різних типів вразливостей.

Крім цього, деякі дослідження поєднують RL із графовими підходами до моделювання атак. Так, Ї Ц. (Yi J.) та Лю С. (Liu X.) [9] продемонстрували підхід із інтеграцією графів MulVAL із Double DQN, що дозволило отримувати більш ефективні траєкторії проникнення та прискорити збіжність навчання у порівнянні з базовим методом. Аналогічно до цього, Чжан Й. (Zhang Y.) та співавтори [10] застосували глибоку рекурентну Q-мережу для моделювання у POMDP, завдяки чому агенти з рекурентною пам'яттю досягали вищого відсотка успішних атак, ніж їхні аналоги без такої компоненти.

Систематизацію підходів здійснили Чень З. (Chen Z.) та ін. [11], які у своєму огляді виокремили ключові проблеми галузі: відсутність масштабованих рішень, нестача стандартизованих метрик, обмежена відтворюваність результатів і потреба у безпечних навчальних середовищах. Ці виклики перегукуються з результатами Морено А.Ц. (Moreno A.C.) та співавторів [12], які показали перспективність інтеграції RL із системами рекомендацій, за допомогою якої агент може отримувати підказки, що враховують як контекст, так і власну історію дій, тим самим поєднуючи переваги різних підходів.

Сучасні дослідження демонструють, що навчання з підкріпленням поступово переходить від теоретичних моделей до прикладних рішень, здатних автоматизувати окремі етапи пошуку вразливостей. Тренувальні середовища, як PenGym, створюють основу для підвищення відтворюваності експериментів і уніфікації підходів. Методи, орієнтовані на динамічні мережі (DynPen) чи пост-експлуатацію (Raijū), демонструють гнучкість і можливість адаптації політик, проте їхні результати здебільшого перевірено на невеликих та контрольованих топологіях. Ієрархічні архітектури, моделювання через POMDP та використання рекурентних компонентів підтверджують свою ефективність у зменшенні простору дій агента та роботі з частковою спостережуваністю, однак наразі брак узгоджених тестів на продуктивність і незалежних відтворювань результатів експериментів не дозволяє робити остаточні висновки про їхню універсальність.

Висновки. Проаналізовані дослідження у даній сфері свідчать про наявність тенденції переходу від теоретичних концепцій до прикладних прототипів, що вже наразі демонструють технічну можливість автоматизації процесу пошуку вразливостей за допомогою навчання з підкріпленням. Результати показують, що застосування даного підходу покращує збіжність навчання у контрольованих середовищах у порівнянні із класичними методами, проте більшість робіт обмежена невеликими тестовими топологіями, що залишає відкритим питання ефективності застосування подібних систем у масштабних прикладних гетерогенних інфраструктурах. Такі підходи як застосування ієрархічних архітектур із маскуванням дій та комбінування RL із графовими моделями атак підтверджують здатність алгоритмів до зменшення простору дій для подальшого підвищення точності та швидкості пошуку вразливостей. Також, перспективними є гібридні рішення, що поєднують автоматичну розвідку і ML-модулі для фільтрації разом із RL-компонентами для планування складних сценаріїв проникнення. Підтвердження універсальності і продуктивності даних підходів в умовах практичного застосування потребує додаткових тестів на більш різнорідних наборах динамічних сценаріїв, що й формує перспективи подальших досліджень.

Список літератури:

1. Guo Yuejun, Bettaieb Seifeddine, Casino Fran. A comprehensive analysis on software vulnerability detection datasets: trends, challenges, and road ahead. *International Journal of Information Security*. 2024. vol. 23(5). pp. 3311–3327. DOI: https://doi.org/10.1007/s10207-024-00888-y

2. Nguyen Huynh Phuong Thanh, Hasegawa Kento, Fukushima Kazuhide, Beuran Razvan. PenGym: a realistic training environment for reinforcement learning agents in automated penetration testing. *Computers & Security*. 2025. vol. 148, Art. no. 104140. DOI: https://doi.org/10.1016/j.cose.2024.104140

3. Li Qianyu, Wang Ruipeng, Li Dong, Shi Feng, Zhang Min, Chattopadhyay Anupam, Shen Ying, Li Yang. DynPen: Automated Penetration Testing in Dynamic Networks. *IEEE Transactions on Information Forensics and Security*. 2024. vol. 19, pp. 8966–8981. DOI: https://doi.org/10.1109/TIFS.2024.3461950

4. Liu Xiaojian, Zhang Yangyang, Li Wenpeng, Gu Wen. Modelling and Intelligent Decision of Partially Observable Penetration Testing for System Security Verification. *Systems*, 2024. vol. 12(12). Art. no. 546. DOI: https://doi.org/10.3390/systems12120546

5. Liu Hongri, Liu Chuhan, Wu Xiansheng, Qu Yun, Liu Hongmei. An Automated Penetration Testing Framework Based on Hierarchical Reinforcement Learning. Electronics. 2024. vol. 13(21). Art. no. 4311. DOI: https://doi.org/10.3390/electronics13214311

6. Ghanem M. C., Chen T. M., Nepomuceno E. G. Hierarchical reinforcement learning for efficient and effective automated penetration testing of large networks. *Journal of Intelligent Information Systems.* 2023. vol. 60(2). pp. 281–303. DOI: https://doi.org/10.1007/s10844-022-00738-0

7. Pham Van-Hau, Hoang Hien Do, Trung Phan Thanh, Quoc Van Dinh, To Trong-Nghia, Duy Phan The. Raijū: Reinforcement Learning-Guided Post-Exploitation for Automating Security Assessment of Network Systems. *Computer Networks.* 2024. vol. 253. Art. no. 110706. DOI: https://doi.org/10.1016/j.comnet.2024.110706

8. AlMajali Anas, Al-Abed Loiy, Ahmad Yousef Khalil M., Mohd Bassam J., Samamah Zaid, Abu Shhadeh Anas. Automated Vulnerability Exploitation Using Deep Reinforcement Learning. *Applied Sciences.* 2024. vol. 14(20). Art. no. 9331. DOI: https://doi.org/10.3390/app14209331

9. Yi Jiadong, Liu Xiaoyan. Deep Reinforcement Learning for Intelligent Penetration Testing Path Design. *Applied Sciences*. 2023. vol. 13(16). Art. no. 9467. DOI: https://doi.org/10.3390/app13169467

10. Zhang Yue, Liu Jingju, Zhou Shicheng, Hou Dongdong, Zhong Xiaofeng, Lu Canju. Improved Deep Recurrent Q-Network of POMDPs for Automated Penetration Testing. *Applied Sciences.* 2022. vol. 12(20). Art. no. 10339. DOI: https://doi.org/10.3390/app122010339

11. Chen Ziyang, Kang Fei, Xiong Xiaobing, Shu Hui. A Survey on Penetration Path Planning in Automated Penetration Testing. *Applied Sciences*. 2024. vol. 14(18). Art. no. 8355. DOI: https://doi.org/10.3390/app14188355

12. Moreno Ariadna Claudia, Hernandez-Suarez Aldo, Sanchez-Perez Gabriel, Toscano-Medina Linda Karina, Perez-Meana Hector, Portillo-Portillo Jose, Olivares-Mercado Jesus, García Villalba Luis Javier. Analysis of Autonomous Penetration Testing Through Reinforcement Learning and Recommender Systems. *Sensors*. 2025. vol. 25(1). Art. no. 211. DOI: https://doi.org/10.3390/s25010211

Shpyta H.V., Vakaliuk T.A. STATE OF RESEARCH IN THE APPLICATION OF REINFORCEMENT LEARNING FOR TESTING THE SECURITY OF INFORMATION SYSTEMS

The problem of detecting vulnerabilities in information systems is becoming increasingly relevant amid the continuous evolution of cyberattack strategies. Traditional penetration testing methods, although they employ automated solutions, still require significant human resources, which creates a need for new approaches. Recent studies show that the application of machine learning makes it possible to partially automate certain phases of the penetration testing process; however, these systems also have limitations. Since system training is performed on a pre-prepared dataset, agents are typically capable of identifying only previously known vulnerabilities. This issue indicates that, for effective application, intelligent algorithms must adapt to dynamic conditions by continuing the learning process during real-world operation. To address problems with such requirements, reinforcement learning and its variants are often employed. Researchers identify this approach as a promising means of solving the problem of discovering potential vulnerabilities even when no information about them was present in the training dataset.

This paper examines modern approaches to training agents using reinforcement learning for conducting penetration testing. Recent experiments involving penetration testing simulators demonstrate that modeling attacks as partially observable Markov decision processes and employing recurrent and hierarchical approaches as well as actor–critic algorithms increases agent effectiveness in complex environments with incomplete information. Furthermore, systematic reviews confirm a trend toward a transition from classical algorithms such as Q-learning and DQN to more sophisticated models capable of operating in high-dimensional action spaces. The use of these methods has the potential to significantly improve the effectiveness of the penetration testing process even in dynamic environments.

Keywords: *algorithm, machine learning, reinforcement learning methods, penetration testing, vulnerability detection, information system.*

Дата першого надходження статті до видання: 14.01.2026
Дата прийняття статті до друку після рецензування: 09.02.2026
Дата публікації (оприлюднення) статті: 08.04.2026