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.