ДОСЛІДЖЕННЯ МЕТОДІВ ВИЯВЛЕННЯ АНОМАЛІЙ У API ЖУРНАЛАХ ДЛЯ ЗАБЕЗПЕЧЕННЯ БЕЗПЕКИ ТА НАДІЙНОСТІ ПРОГРАМНИХ СИСТЕМ

Свиридов А.С., Сєвєрінов О.В.
Харківський національний університет радіоелектроніки, Харків, Україна

На сьогоднішній день захист програмних інтерфейсів (API) застосунків є серйозною задачею. Більшість сучасних компаній надають доступ до своїх послуг через вебзастосунки, де клієнти регулярно обмінюються конфіденційною інформацією між клієнтською і серверною частиною. Це вимагає забезпечення високого рівня захисту таких API-інтерфейсів [1]. Кібернетичний вплив на такі системи може призвести до значних фінансових та технічних збитків.

Традиційний ручний аналіз log-файлів є дуже трудомістким та ресурсозатратним процесом. Автоматизувати процес аналізу великих обсягів даних дозволяють методи штучного інтелекту, зокрема неконтрольованого машинного навчання, оскільки вони здатні виявляти шаблони аномальної активності без попередньої розмітки даних [2].

Окрім цього, для специфічних галузей (наприклад, FinTech) активно досліджуються й інші комбіновані підходи на базі генетичних алгоритмів [3].

Метою доповіді є розробка та оцінка ефективності методів виявлення аномалій у API журналах з метою підвищення безпеки та надійності програмних систем.

Дослідження процесу виявлення аномалій включає в себе чотири основні етапи: підготовка даних, безпосереднє виявлення аномалій, отримання та оцінка результатів.

Підготовча фаза включає збір журналів, вилучення ознак та їхню нормалізацію, що була виконана в цій роботі для підвищення точності передбачення.

Для процесу неконтрольованого виявлення аномалій було обрано чотири моделі: алгоритми кластеризації K-means та GMM, ансамблеву модель пошуку викидів Isolation Forest, а також метод OCSVM (One-Class SVM).

Для зменшення часу обчислень моделі OCSVM через високу розмірність набору даних було використано метод аналізу головних компонентів (PCA).

Експерименти проводилися на наборі даних HTTP Dataset CSIC [4], що містить 61 000 автоматично згенерованих вебзапитів (36 000 нормальних та 25 000 аномальних), включаючи SQL-ін’єкції та XSS.

Аналіз результатів показав, що модель GMM досягла найбільш збалансованого результату з показником F1-міри 0.65 та площею під кривою AUC на рівні 0.78 при швидкому часі навчання.

Модель Isolation Forest продемонструвала найвищу точність (Precision = 0.78) та високу швидкість обробки даних (0.18 с), проте через низьку повноту (Recall = 0.27) вона пропускає значну кількість реальних аномалій.

Модель K-means показала слабку роздільну здатність, оскільки її показник AUC (0.56) наближений до випадкового вгадування.

Алгоритм OCSVM забезпечив прийнятний баланс точності (0.55) та повноти (0.66), проте вимагає значно більших обчислювальних ресурсів та часу на навчання.

Окрім класичних методів, вагомого значення набуває моніторинг динамічних середовищ, де використовується метод виявлення аномалій у Kubernetes-кластерах з використанням LSTM Autoencoder [5].

Завдяки архіте-ктурі рекурентних мереж, цей підхід дозволяє ефективно аналізувати часові послідовності журналів та ідентифікувати приховані залежності через оцінку помилки реконструкції даних. Інтеграція таких нейромережевих моделей із методами неконтрольованого навчання забезпечує масштабовану систему захисту як на рівні окремих API-запитів, так і на рівні всієї хмарної інфраструктури. Крім того, використання механізмів автоматичного навчання дозволяє системі динамічно адаптуватися до змін у структурі трафіку без необхідності ручного перенавчання моделей при кожному оновленні мікросервісів.

В роботі досліджено методи виявлення аномалій у API журналах. Встановлено, що використання методів неконтрольованого навчання є ефективним інструментом для підвищення надійності та безпеки програмних систем.

Зокрема, доведено, що для захисту сучасних хмарних сервісів критично важливим є поєднання базових алгоритмів аналізу з нейромережевими підходами на основі LSTM-автоенкодерів для моніторингу Kubernetes-кластерів.

Подальші дослідження будуть спрямовані на оптимізацію алгоритмів для роботи в реальному часі та інтеграцію з іншими системами кібербезпеки.

Список літератури

1. OWASP, OWASP API Security Project. URL: https://owasp.org/www-project-api-security/ (дата звернення: 25.02.2026).

2. Catherine A. et al. Enhancing API Security in FinTech with Genetic Algorithm-Based Machine Learning Models. 2025.

3. Aharon U., Dubin R., Dvir A., Hajaj C. A classification-by-retrieval framework for few-shot anomaly detection to detect API injection. Computers & Security. 2025. Vol. 150. 104249. DOI: https://doi.org/10.1016/j.cose.2024.104249

4. Sánchez-Zas C. et al. Design and Evaluation of Unsupervised Machine Learning Models for Anomaly Detection in Streaming Cybersecurity Logs. Mathematics. 2022. Vol. 10, No. 21. P. 4043. DOI: https://doi.org/10.3390/math10214043

5. Свиридов, А., Сєвєрінов, О. (2026). Метод виявлення аномалій у kubernetes-кластерах з використанням LSTM Autoencoder. Herald of Khmelnytskyi National University. Technical Sciences, 361(1), 501-509. https://doi.org/10.31891/2307-5732-2026-361-70