Таблиця 1

Порівняння підходів

Метод	Точність	Zero-Day виявлення	Автоматизація
Regex-фільтри	~70%	Немає	частково
WAF (фаєрволи)	~ 85%	Обмежено	Середня
Нейронна мережа	>90%	Так	Повна

1. Li, X., Wang, Z., Xu, M. (2021). A Neural Network-Based Approach for SQL Injection Detection. *Journal of Network and Computer Applications*, 178, 102988. Hilpisch Y. Financial Theory with Python. A Gentle Introduction. Sebastopol: O`Reilly, 2021. 201 p.
2. Kaggle.com SQL Injection Dataset

Розробка системи аналізу вторгнень на основі логів веб-серверу

УДК 004.056.53 Анастасія Піддубровська
Національний університет «Одеська політехніка»,
9480567@stud.op.edu.ua

За даними Verizon DBIR 2024 веб-додатки залишаються головним «шляхом-у» для зловмисників: торік шаблон Basic Web Application Attacks утворив понад 8 % підтверджених зламів і посів перше місце серед векторів початкового проникнення. У середньому організації витрачають 194 дні на виявлення та ще 64 дні на локалізацію інциденту, тож зловмисник місяцями може залишатися непоміченим.

Паралельно з цим високонавантажені сайти генерують десятки або навіть сотні гігабайт HTTP-логів щодоби, що унеможливлює ручний аналіз і вимагає автоматизованих методів обробки. Незважаючи на стрімкий розвиток таких методів, проблема створення ефективних систем аналізу вторгнень на основі логів серверів на сьогодні залишається відкритою.

Метою роботи є підвищення рівня кібербезпеки веб-серверів шляхом розробки системи аналізу вторгнень на основі логів веб-серверу.

Для повноти розгляду матеріалу коротко наведемо характеристики систем аналізу вторгнень. Сучасні дослідження щодо систем аналізу вторгнень можна умовно згрупувати у три взаємопов’язані напрями.

По-перше, застосування глибинного навчання: моделі типу LogEDL (2024) поєднують Transformer-кодери з evidential loss-функцією та демонструють найвищий F1-показник на наборах HDFS, BGL і Thunderbird; новітня архітектура LogLLM (2024/2025) будує двонапрямну зв’язку BERT ↔ Llama і перевершує попередні DL-підходи на чотирьох публічних датасетах.

По-друге, традиційні та гібридні ML-методи: емпіричне дослідження ICSE 2024 засвідчило, що звичайний k-NN навчається у 1000 разів швидше за CNN і водночас підвищує F1 на 6 п.п., а робота Web Traffic Anomaly Detection Using Isolation Forest (Informatics 2024) досягає Precision 95 % і F1 92 % без складного препроцесингу .

По-третє, передобробка та інженерія ознак: велике дослідження в Empirical Software Engineering (2024) показало, що якість парсингу логів слабо корелює з точністю детектування, а вирішальним є коефіцієнт distinguishability шаблонів .

Така еволюція методів підкреслює прагнення спільноти знайти баланс між обчислювальною вартістю, інформативністю ознак і здатністю моделей адаптуватися до нових типів трафіку.

У цьому контексті активно розвиваються самонавчальні та контрастивні підходи**.** Метод Contrastive Log Learning for Unsupervised Anomaly Detection (CLLAD, 2025) істотно знижує залежність від маркованих даних: автори повідомляють про зростання F1-міри на 7–10 п.п. порівняно з LogBERT, зберігаючи час навчання в межах однієї години на наборі OpenHAB .

Ключовою ідеєю є формування позитивних і негативних пар шаблонів запитів та навчання моделі максимізувати інформаційний розрив між нормальними та аномальними подіями.

Нарешті, індустріальні кейси підтверджують практичну ефективність описаних рішень. Упровадження Elastic 8.13 SIEM у трафіку онлайн-ритейлера з середнім навантаженням 350 000 HTTP-запитів за хвилину дозволило виявити та заблокувати 83 % бот-сканувань ще на рівні периметра, скоротивши середній час реакції SOC з 18 до 4 хвилин без помітного впливу на продуктивність кластера . Цей результат демонструє, що грамотне поєднання потокового парсингу, евристичної фільтрації та ML-моделей робить систему придатною для протидії атакам у реальному часі навіть під високими навантаженнями.

Враховуючи окреслені у попередньому абзаці тенденції — стрімке зростання обсягів HTTP-логів, високу тривалість «невидимості» зловмисників і суперечливі результати новітніх DL- та ML-підходів — постає потреба у практичному рішенні, що поєднує дві риси: масштабованість під реальний трафік і прозорість для фахівців SOC. Саме на цей баланс спрямована розроблена у даній роботі гібридна система аналізу логів веб-серверів. Її ядром слугує комбінація Isolation Forest (швидке безнаглядне виявлення рідкісних патернів) і Decision Tree (інтерпретація причин тривоги у вигляді зрозумілих правил), що дозволяє зберегти високу точність без глибокої залежності від маркованих даних — ключове обмеження для більшості «важких» DL-моделей.

На відміну від чисто трансформерних підходів, система реалізує повний конвеєр «лог → інцидент», де модулі збору (Filebeat/Logstash), парсингу й нормалізації одразу готують дані до машинного аналізу, а блок пояснень на базі Decision Tree видає оператору перелік конкретних аномальних ознак — наприклад, нестандартний HTTP-метод чи вибухове зростання запитів із однієї IP-адреси. Така прозорість, підтверджена внутрішніми тестами, суттєво скорочує час ручної валідації й підвищує довіру до автоматичних сповіщень.

Крім того, модульна архітектура розробленого рішення дозволяє нарощувати пропускну здатність кластера або підмінювати окремі компоненти (наприклад, додати контрастивний препроцесинг чи LLM-класифікатор) без переробки всієї системи. У тестовій інфраструктурі з навантаженням понад 300 000 запитів/хв. рішення зберегло стабільний час реакції, а кількість хибнопозитивів залишилася нижчою, ніж у класичних сигнатурних або чисто частотних методів. Таким чином, запропонована система відповідає сучасним вимогам до реального сектору: швидке розгортання, пояснювані результати й можливість поступово інтегрувати більш складні ML-та DL-модулі у міру розвитку загроз.

1. Verizon. 2024 Data Breach Investigations Report. New York : Verizon, 2024. 114 p.
2. Cost of a Data Breach Report 2024. USA : IBM Security, 2024. 87 p.
3. Barr J. Access Logs for Elastic Load Balancers [Electronic resource]. AWS News Blog. 2014. Access mode: https://aws.amazon.com/blogs/aws/access-logs-for-elastic-load-balancers/ (date of access: 22.04.2025).

Використання NetLogo для моделювання кібератак на IoT системи

УДК 004.2

Юрій Підлісний
Національний університет «Чернігівська політехніка», ypodlesny@ukr.net

Інтернет речей (IoT) є однією з найбільш перспективних технологій сучасності, що знаходить застосування у розумних містах, промислових системах та охороні здоров'я. Проте зростаюча кількість IoT-пристроїв робить їх привабливою ціллю для зловмисників. Одним з ефективних методів аналізу вразливостей IoT-мереж є мультиагентне моделювання, яке дозволяє досліджувати динаміку атак і механізми їхнього виявлення.

У цій статті розглядається використання NetLogo для моделювання кібератак на IoT [1].

Модель у NetLogo створена для симуляції взаємодії між різними агентами в мережі IoT, зокрема пристроями, хакерами та захисниками [2]. Вона включає три основних *типи агентів:*

- *Пристрої (devices)* – елементи IoT-мережі, які можуть бути вразливими до атак.
- *Хакери (hackers)* – атакувальні агенти, що намагаються інфікувати пристрої.
- *Захисники (defenders)* – агенти, які виявляють і нейтралізують загрози.

Агенти мають наступні властивості:

Пристрої (devices):

- infected? – статус пристрою (інфікований чи ні).
- security-level – рівень безпеки (від 0 до 100).
- update-readiness – готовність до оновлень, що покращують захист.