УДК 621.395

А. І. Рій[1]
М. І. Кирик[1]

ВИЯВЛЕННЯ ТА ІНТЕРПРЕТАЦІЯ АНОМАЛІЙ ЗА ДОПОМОГОЮ МЕТОДІВ ISOLATION FOREST ТА SHAP

[1] Національний Університет «Львівська Політехніка»

Анотація

Запропоновано підхід до виявлення аномалій у мережевому трафіку на основі методу Isolation Forest, який дозволяє ефективно ідентифікувати нетипову поведінку, що може свідчити про наявність кіберзагроз. Розглянуто етапи попередньої обробки трафіку, налаштування гіперпараметрів моделі та оцінено інтерпретованість результатів за допомогою методу SHAP для визначення впливових ознак на рішення моделі.

Ключові слова: виявлення аномалій, Isolation Forest, мережевий трафік, машинне навчання, кібербезпека, мережева безпека, реальний аналіз, алгоритми виявлення аномалій.

Abstract

A method for detecting anomalies in network traffic based on the Isolation Forest approach is proposed, enabling effective identification of atypical behaviour that may indicate potential cybersecurity threats. The study covers the stages of traffic preprocessing, model hyperparameter tuning, and evaluates the interpretability of the results using the SHAP method to identify the most influential features affecting the model's decisions.

Keywords: anomaly detection, Isolation Forest, network traffic, machine learning, cybersecurity, network security, anomaly detection algorithms, SHAP, TreeSHAP, KernelSHAP.

Вступ

Зі зростанням обсягів трафіку та ускладненням атак традиційні методи виявлення, як-от сигнатурні й евристичні, втрачають ефективність, оскільки не здатні розпізнавати нові загрози без попередніх знань. В умовах шифрування даних актуальними стають підходи, здатні виявляти аномалії без маркованих вибірок [1].

Алгоритм Isolation Forest ефективно ізолює нетипові зразки без потреби у великих обсягах анотованих даних [2]. Проте його головним недоліком є низька інтерпретованість результатів, що ускладнює реагування на загрози.

Цю проблему можна вирішити за допомогою методу SHAP, який дозволяє пояснити внесок кожної ознаки в рішення моделі. Метою дослідження є поєднання Isolation Forest із SHAP для підвищення прозорості та точності виявлення аномалій у трафіку.

Результати дослідження

Аномалії в мережевому трафіку — це відхилення від нормальної активності, які можуть сигналізувати про збої або атаки (Port Scanning, Brute Force, DDoS тощо). Вони поділяються на точкові, контекстуальні та колективні. Для виявлення таких відхилень використовується алгоритм Isolation Forest, що ізолює аномальні записи за допомогою дерев, не потребуючи маркованих даних. Аномальність оцінюється на основі середньої довжини шляху в деревах:

$$s(x,n) = 2^{\frac{-E[f(x)]}{c(n)}} \quad (1)$$

де: $f(x)$ — довжина шляху об'єкта x в дереві ізоляції; $E[f(x)]$ — це середня довжина шляху об'єкта в усіх деревах моделі; $c(n)$ - нормувальне значення, яке залежить від розміру вибірки n. Ця нормалізація необхідна, оскільки, якщо кількість об'єктів у дереві збільшується (тобто розмір вибірки більший), довжина шляхів для всіх об'єктів також зростає.

Для пояснення рішень застосовується SHAP, який визначає вплив окремих ознак на результат моделі (через KernelSHAP або TreeSHAP).

Для дослідження використано датасет CICIDS2017, що містить трафік з позначеними атаками. Було виконано очищення, агрегацію на рівні сеансів та стандартизацію 51 ознаки. Після обробки отримано 2 743 912 записів.

Модель налаштовано з параметрами: n_estimators = 100 max_samples = 256 contamination = 0.02 random_state = 42. Результат: AUC-ROC = 0.95, виявлено 1.84% аномалій (~50 тис. сеансів), що свідчить про високу ефективність.

Для з'ясування причин аномалій, виявлених моделлю Isolation Forest, застосовано два підходи SHAP: KernelSHAP і TreeSHAP. Порівняння цих методів наведено в таблиці 1.

Таблиця 1 – Результати порівняння KernelSHAP та TreeSHAP

Критерій	KernelSHAP	TreeSHAP
Швидкість	Повільний (для 500 записів потрібно ~20 хв)	Швидкий (10,000 записів сек)
Інтерпретація	Аналізує аномальний бал	Аналізує середню довжину шляху
Обмеження	Висока обчислювальна складність	Не підтримує пряму інтерпретацію аномального балу

KernelSHAP працює з будь-якими моделями, але повільний на великих наборах даних (до 20 хв на 500 записів) [3]. TreeSHAP, навпаки, оптимізований для дерев, як-от Isolation Forest, і значно швидший (менше 1 с на 10 000 записів). TreeSHAP продемонстрував високу ефективність завдякиприродній сумісності з деревовими алгоритмами [4]. KernelSHAP аналізує аномальний бал, тоді як TreeSHAP — довжину шляху в дереві, але не дозволяє напряму пояснювати сам аномальний бал, що може бути обмеженням.

Висновки

Дослідження показало, що Isolation Forest ефективно виявляє аномалії в мережевому трафіку, зокрема DoS та Brute Force, без потреби в маркуванні даних. Використання SHAP надало прозорість результатам, пояснюючи внесок ознак в аномальний бал. Тестування на CICIDS2017 підтвердило придатність методу для реальних умов, а порівняння KernelSHAP і TreeSHAP виявило баланс між точністю та швидкодією. Комбінація цих підходів підвищує ефективність виявлення загроз у мережах.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Zhang T., Wang Y., Li J. An ensemble framework for network anomaly detection using Isolation Forest and autoencoders // IEEE Transactions on Information Forensics and Security. – 2023. – Vol. 18. – P. 345–359.

2. Kim H., Lee J., Park S. An optimized Isolation Forest-based intrusion detection system for heterogeneous and streaming data // Journal of Cybersecurity and Privacy. – 2024. – Vol. 3, No. 1. – P. 112–129.

3. Patel R., Desai M. Hybrid machine learning models for intrusion detection: A comparative study with Isolation Forest // Future Generation Computer Systems. – 2022. – Vol. 135. – P. 357–372.

4. Yang J. Fast TreeSHAP: Accelerating SHAP Value Computation for Trees // arXiv. – 2021. – P. 2109.09847.

Рій Андрій Ігорович— аспірант кафедри телекомунікацій, Національний університет «Львівська політехніка», Львів, andrii.i.rii@lpnu.ua

Науковий керівник: ***Кирик Мар'ян Іванович*** — д-р техн. наук, професор каф. телекомунікацій, Національний університет «Львівська політехніка», Львів

Riy Adnriy I. — postgraduate student of the Department of Telecommunications, Lviv Polytechnic National University, Lviv, andrii.i.rii@lpnu.ua

Supervisor: ***Kyryk Maryan I.*** — Dr. Sc. professor of the faculty of telecommunications, Lviv Polytechnic National University, Lviv