УДК 004.056.53

Д.А. Гончар
В.В. Лукічов

МЕТОДИ АНАЛІЗУ МЕРЕЖЕВИХ ПОДІЙ З ВИКОРИСТАННЯМ МАШИННОГО НАВЧАННЯ

Вінницький національний технічний університет

Анотація
Досліджено способи аналізу мережевих подій шляхом використання машинного навчання .
Ключові слова: машинне навчання, siem-системи.

Abstract
The methods of analyzing network events using machine learning were studied**.**
Keywords: machine learning, siem-systems**.**

У сучасному світі спостерігається зростання мережевих загроз, які включають віруси, фішингові атаки та шпигунське програмне забезпечення. Це підвищує необхідність розробки автоматизованих рішень для аналізу подій у мережі. Традиційні інструменти моніторингу часто не справляються через великий обсяг даних і появу нових типів атак. Тому все більше уваги приділяється методам машинного навчання, які можуть автоматично виявляти аномалії та нові загрози.

Мережеві події містять різноманітну інформацію: журнали доступу, поведінку користувачів, системні логи та дані мережевого трафіку. Основна проблема полягає у великій кількості подій, шумі та необхідності швидкого аналізу, щоб відрізнити легітимну активність від підозрілої [1].

Методи машинного навчання пропонують різні способи виявлення загроз. Класифікатори допомагають розділити події на категорії «нормальні» та «підозрілі», тоді як методи виявлення аномалій можуть знаходити нові, раніше невідомі атаки. Аналіз шаблонів у великих обсягах даних дозволяє виявляти складні атаки, такі як APT (Advanced Persistent Threat).

Для аналізу мережевих подій використовують різні алгоритми. Зокрема, серед популярних класифікаторів – логістична регресія, дерева рішень, Random Forest та SVM. Методи кластеризації, як-от K-means та DBSCAN, допомагають виявляти аномалії, коли події не мають попередньо визначених міток. Глибокі нейронні мережі та рекурентні моделі (LSTM) ефективні для аналізу часових рядів. Також використовуються методи підкріплювального навчання для автоматизації оптимізації захисних механізмів.

Аналіз мережевих подій за допомогою машинного навчання включає кілька етапів. Спершу здійснюється збір даних з мережевих журналів та SIEM-систем. Далі проходить попередня обробка, яка охоплює фільтрацію, нормалізацію, видалення дублікатів та заповнення пропусків. Потім будується та тренується модель на основі обраного алгоритму, після чого її точність оцінюється за допомогою метрик, таких як F1-міра. Після успішної валідації модель інтегрується в реальне середовище для постійного моніторингу, а також періодично перенавчається [2].

Важливими є публічні набори даних, такі як KDD'99, NSL-KDD, CICIDS та UNSW-NB15, які допомагають тестувати моделі та перевіряти їхню ефективність. Машинне навчання можна інтегрувати з сучасними SIEM-системами, такими як Splunk чи ELK, для аналізу великих обсягів даних у реальному часі.

Проте методи машинного навчання мають певні обмеження. Великий обсяг даних може призводити до перевантаження систем, а також до великої кількості хибних спрацьовувань, що ускладнює роботу аналітиків. Моделі потребують регулярного оновлення, адже загрози постійно змінюють свої тактики.

Перспективи розвитку методів машинного навчання для аналізу мережевих подій включають використання гібридних підходів, які поєднують машинне навчання з традиційними правилами безпеки. Все актуальнішим стає федеративне навчання, яке дозволяє тренувати моделі на розподілених вузлах без передачі даних. Розробка самонавчальних систем, які здатні адаптуватися до нових даних у реальному часі, є ще одним перспективним напрямом [3].

Таблиця 1 – порівняння методів машинного навчання

Метод	Тип алгоритму	Сильні сторони	Слабкі сторони	Приклади використання
Логістична регресія	Класифікація	Проста реалізація, висока швидкість	Не працює з нелінійними даними	Класифікація нормальних та аномальних подій
Random Forest	Класифікація / Регресія	Стійкий до перевчення, добре працює з різними даними	Вимагає значних ресурсів для тренування	Виявлення загроз за кількома ознаками
SVM (Підтримка векторів)	Класифікація	Висока точність на невеликих наборах даних	Важко масштабувати для великих даних	Виявлення складних патернів у мережевих логах
K-means	Кластеризація	Проста реалізація, швидкий розрахунок	Потребує визначення кількості кластерів	Виявлення груп схожих подій
DBSCAN	Кластеризація	Добре працює з шумними даними та аномаліями	Погано підходить для даних високої розмірності	Виявлення аномалій у потоках трафіку
LSTM (рекурентна нейронна мережа)	Глибоке навчання	Аналіз часових рядів, прогнозування на основі послідовностей	Високі вимоги до обчислень	Аналіз поведінкових патернів у мережевому трафіку
Autoencoder	Ненаглядна класифікація / Виявлення аномалій	Виявлення аномалій без міток, добре працює з великими даними	Складність налаштування та тренування	Виявлення аномалій у логах та трафіку
Градієнтний бустинг (XGBoost)	Класифікація / Регресія	Висока точність, гарна робота з пропущеними значеннями	Тренування може бути повільним	Класифікація атак у реальному часі
Методи підкріплювального навчання	Оптимізація / Адаптація	Адаптація до нових загроз, автоматичне навчання	Складність реалізації та налаштування	Налаштування динамічних правил захисту

Застосування машинного навчання для аналізу мережевих подій має потенціал значно підвищити точність і швидкість виявлення загроз та реагування на них. Завдяки здатності алгоритмів машинного навчання до автоматизованого аналізу великого обсягу мережевих даних, такі системи можуть виявляти складні атаки та аномалії, що залишаються непоміченими традиційними методами захисту.

Однак, щоб досягти максимальної ефективності, моделі машинного навчання повинні бути ретельно налаштовані під конкретні особливості реального мережевого середовища. Зокрема, для успішної роботи моделі потрібно враховувати унікальні шаблони трафіку, типові види взаємодій між користувачами та додатками, а також можливі типи загроз, які характерні для конкретної інфраструктури. Це налаштування охоплює вибір відповідних алгоритмів, створення та підготовку якісного набору даних, а також постійне вдосконалення моделі, враховуючи нові кіберзагрози.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Давиденко Я. Машинне Навчання: Методи Виявлення Аномалій Для Виявлення Незвичайних Моделей, Які Не Відповідають Очікуваній Поведінці. Наука і техніка сьогодні. 2024. № 9(37). URL: https://doi.org/10.52058/2786-6025-2024-9(37)-625-638

2. Theyazn H.H Aldhyani. A review of network traffic analysis and prediction techniques [Електронний ресурс]. - Режим доступу: https://www.researchgate.net/publication/339927785_A_review_of_network_traffic_analysis_and_prediction_techniques

3. An Overview of Machine Learning Methods. Network Anomaly Detection. 2013. P. 83–146. URL: https://doi.org/10.1201/b15088-11

4. Гончар Д. А. Способи аналізу мережевих подій в SIEM-системах [Електронний ресурс] / Д. А. Гончар, В. В. Лукічов // Матеріали Всеукраїнської науково-практичної інтернет-конференції "Молодь в науці: дослідження, проблеми, перспективи (МН-2023)", Вінниця, 22-23 червня 2023 р. – Електрон. текст. дані. – 2023. – Режим доступу: https://conferences.vntu.edu.ua/index.php/mn/mn2023/paper/view/18029

Гончар Данило Андрійович – студент групи 1БС-23м, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: danya.gonchar.2017@gmail.com

Лукічов Віталій Володимирович – к. т. н. , доцент кафедри захисту інформації, Вінницький національний технічний університет, м. Вінниця, lukichov.vitalyi@vntu.edu.ua

Danylo Gonchar – *student of group 1BS-23m, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: danya.gonchar.2017@gmail.com*

Vitaliy Lukichov – *p. h. d., Associate Professor of the Department of Information Protection, Vinnytsia National Technical University, Vinnytsia, lukichov.vitalyi@vntu.edu.ua*