УДК 004.8
Ясінський О. - ст. гр. СНнм-61
Тернопільський національний технічний університет імені Івана Пулюя

МЕТОДИ МАШИННОГО НАВЧАННЯ ДЛЯ ВИЯВЛЕННЯ АНОМАЛІЙ У МЕРЕЖЕВОМУ ТРАФІКУ СУЧАСНИХ ІНФОРМАЦІЙНИХ СИСТЕМ

Науковий керівник: к. соц. ком Липак Г.І.

Yasinskyi O.
Ternopil Ivan Puluj National Technical University

MACHINE LEARNING METHODS FOR DETECTING ANOMALIES IN NETWORK TRAFFIC IN MODERN INFORMATION SYSTEMS

Supervisor: Ph.D Lypak H.I.

Ключові слова: машинне навчання, аномалія, мережевий трафік
Keywords: machine learning, anomaly, network traffic

Виявлення аномалій у мережевому трафіку сучасних інформаційних систем є критично важливим завданням, оскільки традиційні методи на основі сигнатур дедалі частіше пасують перед новими типами кібератак та шифрованим трафіком. Основна ідея застосування машинного навчання тут полягає у створенні профілю «нормальної» поведінки мережі, на фоні якого будь-які суттєві відхилення ідентифікуються як потенційні загрози. Це дозволяє системам захисту реагувати на невідомі раніше вразливості, хоча й створює певний ризик хибнопозитивних спрацювань через природну динамічність мережевих процесів [1]. При цьому сучасні рішення мають бути адаптивними, оскільки характер мережевої активності постійно змінюється, що вимагає від алгоритмів здатності до самонавчання у реальному часі без постійного переналаштовування параметрів людиною.

При порівнянні різних підходів варто виділити методи навчання з учителем, такі як дерева рішень або випадкові ліси, які демонструють високу точність у класифікації вже відомих атак, проте вони майже безпорадні, якщо в навчальній вибірці відсутні приклади конкретної аномалії [2]. Натомість методи навчання без учителя, зокрема алгоритми кластеризації типу k-means або DBSCAN, набагато краще підходять для реальних умов, де мітки даних часто відсутні, оскільки вони шукають ізольовані точки або групи даних, що стоять осторонь від основної маси трафіку. Окреме місце посідають методи на основі статистичних моделей та опорних векторів (One-Class SVM), які фокусуються на описі межі нормальності, що робить їх ефективними для виявлення тонких відхилень у структурі пакетів. Особливу складність сьогодні становить аналіз зашифрованих пакетів, де неможливо перевірити вміст, тому методи машинного навчання все частіше орієнтуються на статистичні характеристики потоків, як-от розмір вікна чи інтервали між пакетами, для виявлення прихованих загроз [3].

Сучасні нейронні мережі, особливо автоенкодери, пропонують ще глибший аналіз, оскільки вони здатні самостійно стискати дані та відновлювати їх, при цьому аномальний трафік відновлюється з високою похибкою, що і слугує маркером загрози. Порівняно з простими метричними методами, глибоке навчання потребує значних обчислювальних ресурсів, проте воно значно краще справляється з величезними

обсягами даних у великих корпоративних мережах [4]. Окрім потужності, важливою стає і стійкість самих моделей до маніпуляцій, адже зловмисники можуть намагатися "обманути" алгоритм, поступово підмішуючи шкідливий трафік у навчальну вибірку.

В таблиці 1 подано порівняльний аналіз методів навчання.

Таблиця 1. Порівняння методів навчання

Метод навчання	**Тип вхідних даних**	**Виявлення нових атак**	**Обчислювальна складність**	**Рівень хибних спрацювань**
Навчання з учителем (Random Forest, SVM)	Розмічені дані (мітки атак)	Низький	Середня	Низький
Навчання без учителя (k-means, Isolation Forest)	Нерозмічені дані	Високий	Низька / Середня	Високий
Однокласові методи (One-Class SVM)	Тільки нормальний трафік	Середній	Середня	Середній
Глибоке навчання (Автоенкодери, RNN)	Сирі дані або Потоки	Дуже високий	Висока	Середній

Вибір конкретного методу залежить від специфіки системи: для жорстко контрольованих промислових мереж краще підходять статистичні та однокласові моделі, тоді як для динамічного інтернет-трафіку найбільш перспективним є поєднання методів кластеризації та глибокого навчання. Оптимальним рішенням для сучасних систем стають гібридні моделі, які поєднують швидкість класичних алгоритмів із гнучкістю нейромереж.

Література:

1. Шевченко А. С., Застело Г. І., Шпачинський Є. О. Аналіз застосування методів машинного навчання на основі штучних нейронних мереж для виявлення кіберзагроз. ela.kpi.ua. 2019. URL: https://ela.kpi.ua/items/1da6e657-b91a-4842-85f7-ea72ae928ad2.

2. Shone N., Tran Nguyen N. A Deep Learning Approach to Network Intrusion Detection. ResearchGate. 2018. URL: https://www.researchgate.net/publication/322866638_A_Deep_Learning_Approach_to_Network_Intrusion_Detection.

3. Kwon D., Kim J. A survey of deep learning-based network anomaly detection. ResearchGate. 2019. URL: https://www.researchgate.net/publication/320066760_A_survey_of_deep_learning-based_network_anomaly_detection.

4. Chalapathy R. Deep Learning for Anomaly Detection: A Survey. arxiv. 2019. URL: https://arxiv.org/abs/1901.03407