УДК 004.056.53

Коровайченко Ю.Ю., аспірант
Нікітін А.М., аспірант
Державний університет інформаційно-комунікаційних технологій

МОДЕЛЬ БАГАТОШАРОВОГО АНАЛІЗУ МЕРЕЖЕВОГО ТРАФІКУ ТА РОЛЬ АЛГОРИТМУ RANDOM FOREST У ІЄРАРХІЇ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО ВИЯВЛЕННЯ АНОМАЛІЙ

У сучасних комп'ютерних мережах обсяг і складність трафіку демонструють постійне зростання, що суттєво ускладнює процес виявлення аномалій у таких системах як IPS/IDS (попередження та виявлення вторгнень). Традиційні монолітні методи аналізу виявляються недостатньо ефективними для обробки різнорівневого контексту мережевих даних, який охоплює спектр від сирих пакетів до складних поведінкових патернів. У цьому аспекті багатошаровий підхід аналізу пропонує ієрархічну модель, яка інтегрує сигнатурні, статистичні та методи машинного навчання на пакетному, потоковому й поведінковому рівнях, забезпечуючи комплексне та послідовне виявлення потенційних загроз.

На пакетному рівні аналізу акцент робиться на необроблених даних: прапорцях, значенні часу життя пакета (TTL) та протоколах передачі. Домінуючими тут є сигнатурні методи, призначені для оперативного фільтрування відомих типів атак; водночас застосування методів машинного навчання обмежене високою динамікою обробки та низькою структурованістю інформації. Потоковий рівень передбачає агрегацію даних у форми NetFlow чи IPFIX, з фокусом на таких ознаках, як тривалість з’єднання та обсяг переданих даних, що дозволяє впроваджувати статистичні моделі та базові методи машинного навчання для ефективної редукції шумів. Поведінковий рівень, у свою чергу, орієнтований на вивчення патернів користувацької активності та контекстних відхилень, де методи машинного навчання досягають найвищої продуктивності завдяки здатності моделювати нелінійні залежності в даних.

У загальній ієрархії методів аналізу мережевого трафіку алгоритм Random Forest посідає одне з пріоритетних місць серед ансамблевих підходів у рамках машинного навчання, перевершуючи сигнатурні методи за адаптивністю та евристичні - за стійкістю до варіацій у трафіку. Зокрема, Random Forest виявляється оптимальним для потокового та поведінкового рівнів: він ефективно обробляє

високовимірні набори ознак без ризику перенавчання завдяки механізму баггінгу. Алгоритм характеризується стійкістю до шумів, притаманних мережевим потокам, та забезпечує інтерпретованість через аналіз важливості ознак. На пакетному рівні застосування Random Forest є менш доцільним через обмежену кількість доступних ознак та суворі вимоги до обробки в реальному часі. Алгоритм Random Forest переважає метод опорних векторів (SVM) за швидкістю обчислень на великих масивах даних і градієнтний бустинг - за простотою параметризації, хоча й поступається глибоким нейронним мережам у здатності захоплювати вкрай складні патерни. Таким чином, він забезпечує оптимальний баланс для гібридних систем виявлення вторгнень.

Запропонована багатошарова модель інтегрує алгоритм Random Forest як вершину ієрархії методів, що, за оцінками літературних джерел, демонструє підвищення ефективності на значному рівні. Подальші перспективи розвитку пов'язані з гібридизацією Random Forest та методів глибокого навчання для формування автономних систем кібербезпеки.

Список використаних джерел:

1. Awotunde J. B., Ayo F. E., Panigrahi R., Garg A., Bhoi A. K., Barsocchi P. A Multi-level Random Forest Model-Based Intrusion Detection Using Fuzzy Inference System for Internet of Things Networks // International Journal of Computational Intelligence Systems, 2023. URL: https://link.springer.com/article/10.1007/s44196-023-00205-w

2. Agate V., De Paola A., Ferraro P., Lo Re G. MIDES: A multi-layer Intrusion Detection System using ensemble machine learning // International Journal of Intelligent Networks, 2025. URL: https://www.sciencedirect.com/science/article/pii/S2666603025000156