Proceedings of the 8th International
Scientific and Practical Conference
«Recent Advances in Global Science»
(August 26-28, 2026).
Vilnius, Lithuania

No
307

INFORMATION AND WEB TECHNOLOGIES

Штучний Інтелект у Задачах Виявлення Мережевих Атак

Онацький Олексій Віталійович[1],
Демовський Артем Віталійович[2]

[1] кандидат технічних наук, доцент кафедри кібербезпеки та технічного захисту інформації
Державний університет інтелектуальних технологій і зв'язку, Україна

[2] студент II курсу
Державний університет інтелектуальних технологій і зв'язку, Україна

Мережеві системи виявлення вторгнень (NIDS) виконують функцію безперервного аналізу мережевих подій з метою виявлення ознак несанкціонованої або аномальної активності. Класичні підходи ґрунтуються на сигнатурному аналізі, аналізі стану протоколів і виявленні аномалій. Сигнатурний механізм забезпечує високу точність для відомих атак, але залежить від актуальності бази правил. Аномальний підхід формує модель нормальної або шкідливої поведінки та оцінює відхилення нових спостережень від цієї моделі. Саме ця постановка створила основу для використання статистичних методів, машинного навчання та глибоких нейронних мереж у мережевому моніторингу [2, 10, 11].

Машинне навчання в NIDS використовується для двох основних задач: класифікації мережевого трафіку та виявлення аномалій. У контрольованому навчанні модель будується на розмічених прикладах і встановлює відповідність між сукупністю ознак трафіку та відомими класами. Бінарна постановка розділяє нормальний і шкідливий трафік, багатокласова – визначає категорію атаки. Неконтрольовані методи працюють без повного набору міток і шукають статистично нетипові спостереження. Глибоке навчання дозволяє формувати багаторівневі представлення мережевих даних без ручного конструювання всіх ознак, але підвищує вимоги до обсягу вибірки, часу навчання та обчислювальних ресурсів [1, 7, 8, 9]. Узагальнену схему використання моделі машинного навчання у системі аналізу мережевого трафіку наведено на рисунку 1.

INFORMATION AND WEB TECHNOLOGIES

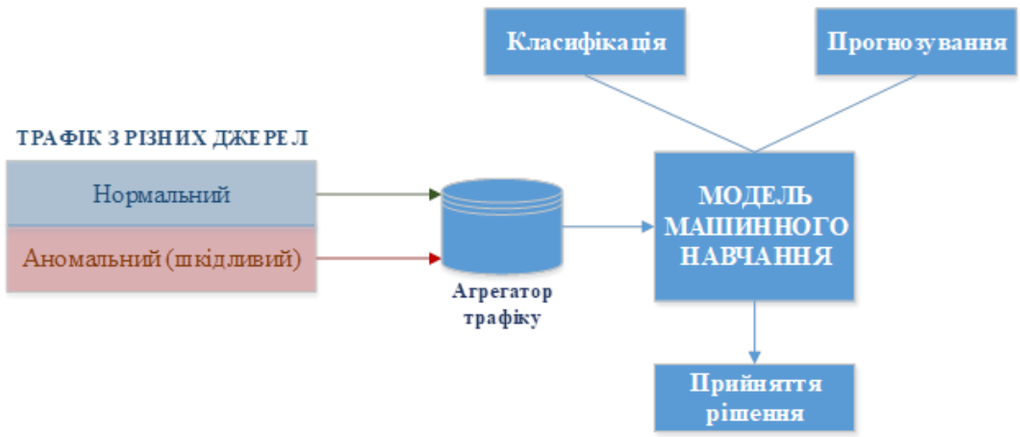

Рисунок 1. **Узагальнена схема застосування машинного навчання для аналізу мережевого трафіку**

На вхід NIDS надходять пакети або мережеві потоки, сформовані різними вузлами та сервісами. Сирий трафік не подається безпосередньо до більшості класичних моделей. Спочатку виконується агрегація та формування ознак: тривалість потоку, кількість пакетів, обсяг переданих даних, міжпакетні інтервали, інтенсивність передачі, тип протоколу, напрямок обміну, TCP-прапорці та статистичні характеристики розподілу пакетів. Після перетворення мережевої події у вектор ознак модель виконує класифікацію або обчислює аномальний бал. Вихід моделі використовується механізмом прийняття рішення для створення сповіщення, передачі події на додатковий аналіз або запуску захисної реакції. Помилки на етапі агрегації чи формування ознак безпосередньо змінюють простір, у якому працює алгоритм, тому якість NIDS визначається не лише вибором класифікатора [1, 2].

Репрезентативність навчальних даних є центральним обмеженням інтелектуальних NIDS. Sommer і Paxson описали проблему «закритого світу»: експериментальна вибірка має фіксований склад сервісів, користувачів, атак і часових характеристик, тоді як продуктивна мережа постійно змінюється [3]. Модель може демонструвати високі метрики на тестовій частині того самого набору даних і втрачати точність після перенесення в іншу інфраструктуру. Огляд мережевих датасетів підтверджує значні відмінності між ними за способом збору, набором протоколів, сценаріями атак, балансом класів і правилами маркування [4]. Результат експерименту тому характеризує не алгоритм у відриві від середовища, а пару «алгоритм – дані – процедура оцінювання».

INFORMATION AND WEB TECHNOLOGIES

IoT-мережі посилюють проблему неоднорідності трафіку. Сенсори, камери, шлюзи, контролери та побутові пристрої мають різні профілі активності, використовують різні протоколи та часто працюють з обмеженими ресурсами. Дослідження NIDS для IoT розглядають машинне навчання як засіб розпізнавання DDoS, DoS, reconnaissance, spoofing, brute-force, Mirai та інших типів шкідливої активності [5]. Набір CICIoT2023 був створений для відтворення масштабних атак у середовищі зі 105 IoT-пристроїв і дозволяє досліджувати як відокремлення атак від нормального трафіку, так і класифікацію конкретних категорій мережевої активності [6].

Проблема дисбалансу класів безпосередньо впливає на оцінювання NIDS. У мережевих вибірках окремі атаки можуть становити незначну частку спостережень, тоді як інші класи домінують. За такого розподілу Accuracy здатна залишатися високою навіть при систематичних помилках на рідкісних класах. Оцінювання потребує Precision, Recall, F1-score, Balanced Accuracy, аналізу хибнопозитивних і хибнонегативних спрацювань та матриці помилок. Для захисної системи помилки мають різну вартість: надмірна кількість false positive перевантажує аналітиків, а false negative означає пропущену шкідливу активність [2, 9].

Експлуатаційні характеристики моделі мають таке саме значення, як і її точність. NIDS працює з потоком подій, тому затримка прогнозування, пропускна здатність, використання оперативної пам'яті та процесорного часу визначають можливість розгортання моделі в реальному середовищі. Складні ансамблеві та нейронні моделі можуть забезпечувати високу якість класифікації, але створювати неприйнятну затримку на периферійних вузлах. Легші алгоритми придатні для попередньої фільтрації, тоді як ресурсоємні моделі можуть виконувати поглиблений аналіз лише відібраних подій. Така багаторівнева архітектура відповідає практиці розподілу навантаження між мережевими сенсорами, центральними колекторами та аналітичними сервісами.

Відбір ознак визначає складність моделі та її здатність відокремлювати нормальну поведінку від атак. Мережеві набори часто містять корельовані або слабкоінформативні параметри. Їх безконтрольне використання збільшує розмірність простору ознак, час навчання та ризик перенавчання. Кореляційний аналіз, статистичні критерії, mutual information, важливість ознак у деревах і методи з регуляризацією застосовуються для формування компактнішого представлення трафіку. При цьому процедура відбору повинна виконуватися лише на навчальній

INFORMATION AND WEB TECHNOLOGIES

частині даних; використання інформації з тестової вибірки створює витік даних і завищує оцінку якості моделі.

Мережевий трафік змінюється в часі навіть без дій зловмисника. Оновлення програмного забезпечення, поява нових сервісів, зміна робочих процесів або підключення нових пристроїв змінюють статистичний розподіл ознак. Для моделі це проявляється як drift між навчальним і поточним трафіком. Статична NIDS поступово накопичує помилки, якщо профіль середовища відрізняється від початкової вибірки. Реальне застосування машинного навчання потребує контролю якості після розгортання, накопичення нових перевірених прикладів та періодичного перенавчання.

Ще одна група ризиків пов'язана з тим, що модель машинного навчання сама стає об'єктом впливу. Зловмисник може змінювати характеристики мережевої активності так, щоб шкідливий потік залишався функціонально придатним, але переходив у область ознак, яку модель класифікує як нормальну. Для систем, що автоматично поповнюють навчальні дані, додатковою загрозою є внесення некоректно маркованих або спеціально сформованих прикладів. Стійкість до таких впливів потребує незалежної валідації даних, контролю джерел навчальної вибірки та перевірки моделі на модифікованих сценаріях атак.

Інтерпретованість рішення має практичне значення для роботи центру моніторингу безпеки. Сам факт мітки «attack» не пояснює причину спрацювання і ускладнює перевірку інциденту. Для дерев рішень і частини ансамблевих моделей можна оцінити внесок окремих характеристик, тоді як складні нейронні архітектури потребують окремих методів пояснення. Інтерпретація ознак допомагає відрізнити реальну закономірність від артефакту набору даних, виявити залежність від технічних параметрів конкретного стенда та підвищити довіру до автоматизованого рішення.

Практична архітектура NIDS на основі штучного інтелекту рідко обмежується однією моделлю. Сигнатурні правила забезпечують надійне розпізнавання відомих шаблонів, поведінкові механізми фіксують відхилення, а моделі машинного навчання виконують багатофакторну оцінку трафіку. Поєднання цих механізмів з кореляцією подій, журналами автентифікації, телеметрією вузлів і контекстом активів зменшує залежність рішення від одного джерела даних. Штучний інтелект у такій системі виконує роль аналітичного компонента, а не автономного замінника IDS/IPS.

INFORMATION AND WEB TECHNOLOGIES

Список використаних джерел:

[1] Buczak A. L., Guven E. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. IEEE Communications Surveys & Tutorials. 2016. Vol. 18, No. 2. P. 1153-1176. DOI: 10.1109/COMST.2015.2494502.

[2] Khraisat A., Gondal I., Vamplew P., Kamruzzaman J. Survey of Intrusion Detection Systems: Techniques, Datasets and Challenges. Cybersecurity. 2019. Vol. 2. Article 20. DOI: 10.1186/s42400-019-0038-7.

[3] Sommer R., Paxson V. Outside the Closed World: On Using Machine Learning for Network Intrusion Detection. 2010 IEEE Symposium on Security and Privacy. 2010. P. 305-316. DOI: 10.1109/SP.2010.25.

[4] Ring M., Wunderlich S., Scheuring D., Landes D., Hotho A. A Survey of Network-Based Intrusion Detection Data Sets. Computers & Security. 2019. Vol. 86. P. 147-167. DOI: 10.1016/j.cose.2019.06.005.

[5] Chaabouni N., Mosbah M., Zemmari A., Sauvignac C., Faruki P. Network Intrusion Detection for IoT Security Based on Learning Techniques. IEEE Communications Surveys & Tutorials. 2019. Vol. 21, No. 3. P. 2671-2701. DOI: 10.1109/COMST.2019.2896380.

[6] Neto E. C. P., Dadkhah S., Ferreira R., Zohourian A., Lu R., Ghorbani A. A. CICIoT2023: A Real-Time Dataset and Benchmark for Large-Scale Attacks in IoT Environment. Sensors. 2023. Vol. 23, No. 13. Article 5941. DOI: 10.3390/s23135941.

[7] Ferrag M. A., Maglaras L., Moschoyiannis S., Janicke H. Deep Learning for Cyber Security Intrusion Detection: Approaches, Datasets, and Comparative Study. Journal of Information Security and Applications. 2020. Vol. 50. Article 102419. DOI: 10.1016/j.jisa.2019.102419.

[8] Shone N., Ngoc T. N., Phai V. D., Shi Q. A Deep Learning Approach to Network Intrusion Detection. IEEE Transactions on Emerging Topics in Computational Intelligence. 2018. Vol. 2, No. 1. P. 41-50. DOI: 10.1109/TETCI.2017.2772792.

[9] Chandola V., Banerjee A., Kumar V. Anomaly Detection: A Survey. ACM Computing Surveys. 2009. Vol. 41, No. 3. Article 15. DOI: 10.1145/1541880.1541882.

[10] García-Teodoro P., Díaz-Verdejo J., Maciá-Fernández G., Vázquez E. Anomaly-based Network Intrusion Detection: Techniques, Systems and Challenges. Computers & Security. 2009. Vol. 28, No. 1-2. P. 18-28. DOI: 10.1016/j.cose.2008.08.003.

[11] Liao H.-J., Lin C.-H. R., Lin Y.-C., Tung K.-Y. Intrusion Detection System: A Comprehensive Review. Journal of Network and Computer Applications. 2013. Vol. 36, No. 1. P. 16-24. DOI: 10.1016/j.jnca.2012.09.004.