УДК 004.7

А.Л. Столяр
Національний авіаційний університет, Київ

ПРОБЛЕМИ ВИЯВЛЕННЯ АНОМАЛІЙ В КОМП'ЮТЕРНИХ МЕРЕЖАХ У РЕЖИМІ РЕАЛЬНОГО ЧАСУ

Аномалії – це несподівані або незвичайні події, які відбуваються в мережі, а виявлення аномалій – це процес ідентифікації цих подій. У комп'ютерних мережах аномалії можуть бути спричинені різними факторами, такими як апаратні та програмні збої або кібератаки. Виявлення аномалій у режимі реального часу передбачає процес, який ефективно й точно визначає аномалії шляхом одночасного збору й обробки даних із високою обчислювальною ефективністю. Однак виявлення аномалій у реальному часі пов'язане з низкою проблем, які необхідно подолати. Далі розглядаються деякі з ключових проблем, пов'язаних із виявленням аномалій у комп'ютерних мережах.

Виявлення аномалій у реальному часі може бути складним через високу швидкість отримання даних і великий обсяг даних, які необхідно обробити. Обсяги даних можуть варіюватися від гігабайт до терабайт і більше, залежно від розміру і складності мережі. Додатково, мережа може обробляти від тисяч до мільйонів точок даних в секунду. Одним із рішень цієї проблеми є використання розподілених обчислень. Дані можуть оброблятися на декількох серверах, здатних впоратися з великими обсягами і швидкістю даних. Розподілені обчислювальні системи, такі як Hadoop і Spark, використовуються для обробки даних в реальному часі. Іншим рішенням є використання вибірок і фільтрації для зменшення обсягу даних, які необхідно проаналізувати.

Системи виявлення аномалій у реальному часі повинні обробляти різні типи даних, такі як мережевий трафік, системні журнали, метрики додатків і події безпеки. Оскільки кожен тип даних має власну структуру, формат і характеристики, може бути важко співвідносити та аналізувати різні джерела даних. Щоб подолати цю проблему, одним із рішень є використання методів нормалізації даних, щоб забезпечити узгодженість форматів даних. Ще один підхід – конструювання ознак, який представляє собою процес ідентифікації та вибору вагомих атрибутів даних. Це може

допомогти зменшити вплив різноманітності даних і спрямувати увагу системи на найважливіші характеристики.

Якість даних є ще однією проблемою при виявленні аномалій у реальному часі. Дані можуть бути неповними, зашумленими або містити помилки. Проблеми з якістю даних можуть бути спричинені кількома факторами, такими як помилки датчиків, перевантаження мережі або помилки при передачі даних. Система виявлення аномалій у реальному часі повинна вміти виявляти і фільтрувати такі дані. Система також повинна вміти обробляти неповні або відсутні дані та обчислювати або інтерполювати відсутні значення для проведення точного аналізу. Цього можна досягти, використовуючи фільтрування та нормалізацію – це дві процедури попередньої обробки, які можуть допомогти усунути шум і непотрібні дані з вхідного потоку.

Системи виявлення аномалій у реальному часі повинні досягати високої точності виявлення аномалій, мінімізуючи кількість хибнопозитивних і хибнонегативних результатів. Хибнопозитивні спрацьовування виникають, коли нормальна поведінка класифікується як аномальна, тоді як хибнонегативні результати виникають, коли аномальна поведінка класифікується як нормальна. Одним із способів вирішення цієї проблеми є використання машинного навчання. А саме, навчання алгоритму на великих наборах даних для підвищення його точності. Іншим рішенням є використання ансамблевого навчання. Це об'єднання декількох алгоритмів для підвищення їхньої точності.

Системи виявлення аномалій у реальному часі повинні адаптуватися до змін у поведінці мережі та виявляти нові типи аномалій. Мережі є динамічними і можуть змінюватися з часом під впливом різних факторів. Системи виявлення аномалій в реальному часі повинні мати можливість адаптуватися до цих змін і виявляти нові типи аномалій без необхідності конфігурації або ручного втручання. Система виявлення аномалій в реальному часі також повинна вміти вчитися на попередніх аномаліях і вдосконалювати моделі та алгоритми виявлення з часом. Одним із способів подолання цієї проблеми є використання онлайн-навчання. Це дозволить оновлювати алгоритм в режимі реального часу на основі нових даних. Іншим рішенням є використання динамічних порогів.

Це дозволяє коригувати поріг для виявлення аномалій на основі поточних даних.

Системи виявлення аномалій в реальному часі вимагають великих обчислювальних ресурсів для обробки великих обсягів даних в режимі реального часу і виявлення аномалій. Система виявлення аномалій в реальному часі повинна мати можливість горизонтального і вертикального масштабування, щоб бути здатною обробляти зростаючі обсяги даних і трафіку. Крім того, системи виявлення аномалій в реальному часі повинні оптимізувати алгоритми і структури даних для зменшення обчислювальних витрат і досягнення високої продуктивності. Одним з рішень для вирішення цієї проблеми є використання хмарних обчислень. Для обробки даних використовуються віддалені сервери, що дозволяє зменшити навантаження на локальну мережу. Іншим рішенням є використання апаратного прискорення. Для цього використовується спеціалізоване обладнання, таке як графічні процесори для більш ефективної обробки даних.

Усі ці заходи разом забезпечують ефективний моніторинг як зовнішніх, так і внутрішніх ризиків, одночасно підвищуючи загальний рівень продуктивності

У цьому дослідженні було проаналізовано ключові проблеми, які виникають при виявленні аномалій в реальному часі та запропоновано рекомендації, які дозволять перебороти ці проблеми та покращити ефективність роботи систем виявлення аномалій, які працюють у реальному часі.

ВИКОРИСТАНІ ДЖЕРЕЛА

1. Limthong K. Real-Time Computer Network Anomaly Detection Using Machine Learning Techniques. Journal of Advances in Computer Networks. – 2013. – P. 1-5.

2. Chuadhry M.A., Gauthama R.M.R., Aditya M. Challenges in Machine Learning based approaches for Real-Time Anomaly Detection in Industrial Control Systems. Proceedings of the 6th ACM on Cyber-Physical System Security Workshop. – Association for Computing Machinery, New York, USA, 2020.– P. 23-29.

3. Hadoop vs. Spark: What's the Difference? URL: https://www.ibm.com/cloud/blog/hadoop-vs-spark (дата звернення: 27.03.2023).