УДК 004.7 **doi: 10.26906/SUNZ.2024.1.133**

Р. М. Марченко, А. А. Коваленко, В. Г. Знайдюк

Харківський національний університет радіоелектроніки, Харків, Україна

АНАЛІЗ МЕТОДІВ ВИЯВЛЕННЯ АНОМАЛЬНОГО ТРАФІКУ В МЕРЕЖАХ IoT

Анотація: Метою даної роботи є проведення комплексного аналізу методів та підходів до виявлення аномалій в мережах Інтернету речей (IoT). З урахуванням стрімкого розвитку IoT і збільшення кількості підключених пристроїв, проблема виявлення аномального трафіку стає актуальною для забезпечення безпеки та ефективності цих мереж. У роботі розглядаються різні методи та підходи до виявлення аномалій, включаючи статистичний аналіз, мережевий моніторинг, поведінковий аналіз, а також застосування сучасних технологій машинного та глибокого навчання. Кожен із цих методів розглядається з точки зору його застосовності в контексті IoT та оцінюються його переваги та обмеження. Робота також розглядає сучасні виклики і перспективи розвитку у галузі безпеки IoT, з фокусом на захисті від кіберзагроз та посиленні систем виявлення аномалій.

Ключові слова: Інтернет речей, аномальний трафік, машинне навчання, глибинне навчання, статистичний аналіз.

Вступ

Розвиток Інтернету речей (IoT) та пов’язаних технологій супроводжується експоненціальним зростанням кількості пристроїв різних типів, що працюють на основі різноманітних технологій та підключені до мережі, яка об'єднує їх локально та через мережу Інтернет. Характерною рисою IoT є збільшення кількості сенсорів, які збирають дані з навколишнього середовища, а потім аналізують ці дані та впливають на фізичний світ через виконавчі механізми.

Пристрої IoT використовуються у багатьох сферах, включаючи побутову техніку, охоронні системи, медичне обладнання, системи управління та носимі пристрої. За сучасними оцінками, кількість підключених до IoT пристроїв щодня зростає, і до 2025 року їх може бути приблизно 85 мільярдів, що охоплюватиме галузі виробництва (40%), медицину (30%), роздрібну торгівлю та безпеку (20%) [1].

Цей суттєвий розвиток IoT відкриває нові можливості для майбутніх застосувань. У міру збільшення цінності даних, що зберігаються, обробляються та передаються, разом із масштабом зростають і атаки на них [2–5]. Ці прогнози показують, що кількість і рівень загроз і атак на пристрої IoT зростатиме, що потребуватиме більш надійних заходів безпеки. Використання IoT охоплює різноманітні сценарії, від окремих пристроїв до розгортання технологій на крос-платформеному рівні та використання систем реального часу у хмарних обчисленнях [6].

Функціональність в мережі IoT включають три основні завдання: передачу даних, отримання даних та їх обробку. На рівні застосунків, вбудовані інтерфейсні модулі дозволяють пристроям взаємодіяти з основною архітектурою. План управління пристроями визначає джерело та призначення даних для забезпечення операцій введення-виведення у пристроїв. Наприклад, агрегатор об'єднує дані, надані різними пристроями, в єдиний набір. Шар зв'язку виступає проміжним рівнем з мережними компонентами, які встановлюють різні протоколи та стандарти для керування трафіком у системі.

Використання стандартних протоколів дозволяє реалізувати належну комунікацію між пристроями IoT. Для таких систем важливий наявний набір простих правил для ініціалізації та обміну інформацією. Схематично багаторівневу архітектуру IoT можна представити у вигляді рис. 1, де наведено структуровану архітектуру типової системи IoT з відокремленими рівнями апаратного забезпечення, комунікацій та застосунків.

На найнижчому рівні знаходяться фізичні IoT-пристрої, які відповідають за збір даних та взаємодію з фізичним світом. Над ним розташований комунікаційний рівень, який містить протоколи зв'язку для транспортування даних в мережі. Цей рівень також включає агрегаційний шар, що узагальнює дані з різних джерел перед подачею на верхні рівні. Найвищий рівень включає веб-портали, управління API та хмарні/граничні сервіси, що надають різноманітні сервіси для обробки та аналітики подій.

Окремо від основних рівнів зображено план управління пристроями, який забезпечує інтеграцію та координацію всіх IoT-пристроїв у мережі.

Мета статті – провести аналіз методів виявлення аномального трафіку в мережах IoT, виявити основні переваги і проблеми цих методів для подальшого їх дослідження та впровадження, а також вплив факторів на їх ефективність.

Аналіз сучасних методів виявлення аномалій в IoT

Аномалія в контексті IoT – це дані або спостереження, які виходять за межі очікуваної поведінки в системі. Це може бути рідкісна подія або відхилення від типового шаблону в конкретний момент часу або для певного контексту. Аномалії можуть бути спричинені зовнішніми факторами, такими як помилки датчиків або кібератаки. Задача алгоритму виявлення аномалій – виявити ці викиди і, за можливості, визначити їхні причини [7].

Алгоритми виявлення аномалій можна розділити на чотири категорії в залежності від підходу до: вирішення задачі; способу застосування; типу методу; затримки алгоритму. Важливо мати різні підходи для різних застосувань IoT через їхню різноманітну природу та різновиди даних. Наприклад, один підхід може бути кращим для виявлення аномалій у вимірах датчиків, а інший – для виявлення відхилень у мережному трафіку.

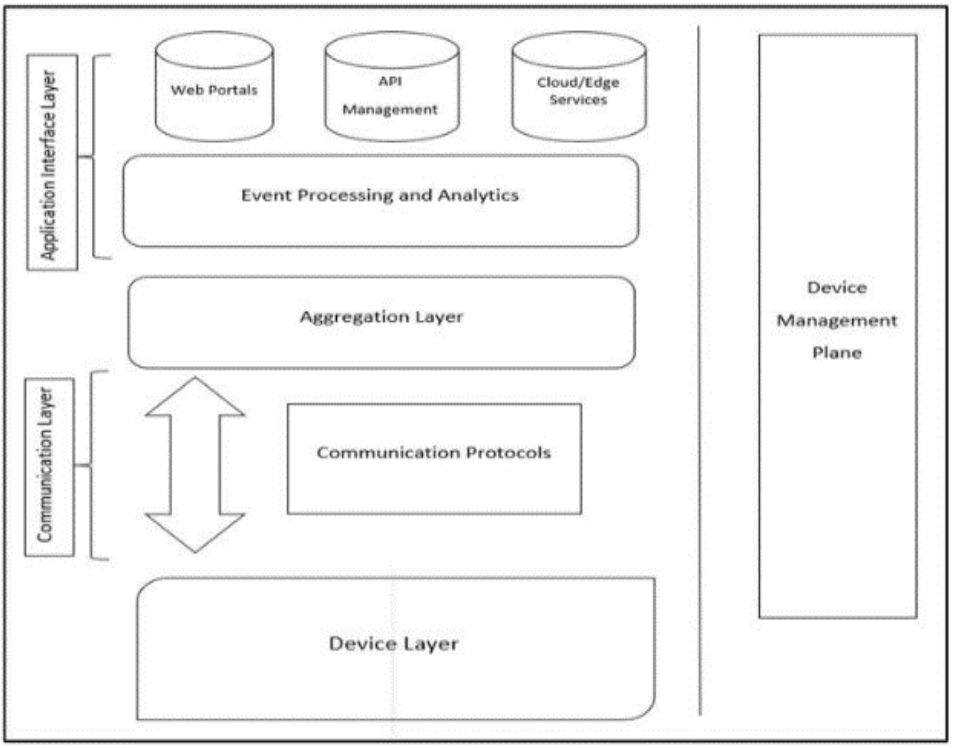

Рис. 1. Багатошарова архітектура IoT

У задачі бінарної класифікації аномалій велике значення має вибір моделі наближення, яка найкраще відображає очікувану поведінку даних. Точність цієї моделі визначає, наскільки ефективно будуть виявлені аномалії. Оскільки IoT включає в себе різноманітні застосунки та типи даних, часто потрібно використовувати різні стратегії для виявлення аномалій, які оптимізовані для конкретних сценаріїв. На рис. 2 показаний приклад однієї з аномалій [7].

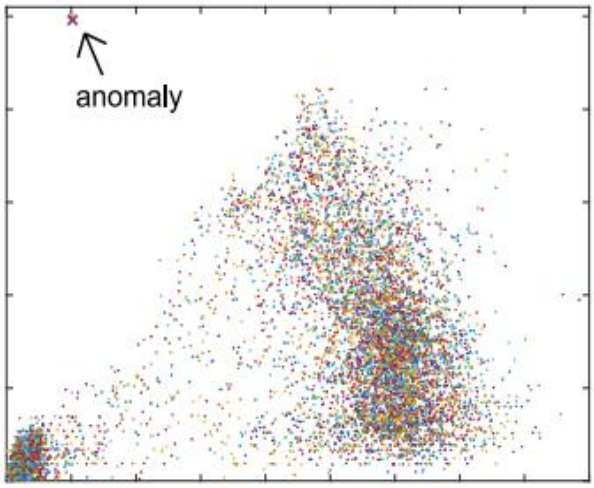

Рис. 2. Діаграма з прикладом аномалії

Методи виявлення аномалій в IoT поділяються на чотири категорії, комбінуючи класифікації з опублікованих результатів досліджень [8] та [9]. Їх класифікують за способом підходу до проблеми, застосуванням, типом методу та затримкою алгоритму. Нижче наведено короткий огляд цих методів та деяких традиційних підходів, що використовуються в IoT:

Класифікація за методом включає:

- геометричні методи: ґрунтуються на ідеї, що при стратегіях, основаних на відстані та щільності даних, очікувані та аномальні дані розділені; зазвичай вони використовують статичний або динамічний поріг для класифікації даних як нормальних або аномальних; декілька прикладів геометричних методів включають в себе методи на основі відстані та щільності;

- статистичні методи: намагаються моделювати нормальні дані за допомогою математичних моделей та розподілів; один із прикладів – метод мінімального об'єму, який намагається створити n-вимірний симплекс навколо заданої області даних;

- методи машинного навчання та глибокого навчання: вибір моделі залежить від характеру наданих даних; наприклад, моделі типу Long Short-Term Memory (LSTM) та трансформатори відповідають за послідовні дані, такі як аудіо, відео та часові ряди; з іншого боку, моделі типу Convolutional Neural Network (CNN) та Autoencoder (AE) підходять для не послідовних даних, таких як зображення.

Класифікація за застосуванням включає:

- конструктивні застосування: спрямовані на позитивну діяльність та надають користь, таку як моніторинг щоденної активності літніх людей для попередження падінь;

- деструктивні застосування: спрямовані на завдання шкоди, такі як атаки на мережу IoT або намагання завдати шкоди даним та застосункам;

- застосування для очищення даних: спрямовані на видалення непотрібних даних або шуму з вхідного сигналу.

Класифікація за типом аномалії включає:

- пунктові аномалії: виникають, коли одна точка даних відхиляється від очікуваної поведінки;

прикладом може бути виявлення шахрайства з банківськими картками;

- контекстуальні аномалії: аномалії, які можуть вважатися такими лише в певному контексті і виявляються, коли розглядаються як контекстуальні, так і поведінкові характеристики;

- колективні аномалії: визначаються на основі всього набору даних та не пов'язані з окремими точками даних.

Класифікація за затримкою включає:

- online алгоритми: обробляють дані під час їх збору і можуть аналізувати одну точку даних або вікно даних без повного доступу до всіх даних;

- offline алгоритми: мають доступ до всіх даних і використовують більш складні обчислювальні методи для розв'язання задачі.

Ця категоризація вказує на різноманітність методів виявлення аномалій в IoT та їх застосування в залежності від конкретного сценарію та потреби.

Основні переваги та проблеми методів виявлення аномалій в IoT, які вимагають подальшого дослідження наведені в табл. 1. Наведена таблиця результатів аналізу надає загальний огляд переваг та недоліків кожної категорії методів виявлення аномалій в Інтернеті речей залежно від різних аспектів їх використання та застосування.

Таблиця 1 – **Результати аналізу**

Категорія методу	Переваги	Недоліки
За методом		
Геометричні методи	Добре підходять для даних з чітко визначеними структурами.	Можуть бути неефективними для даних із складними структурами або часово залежними даними.
Статистичні методи	Можуть моделювати різноманітні розподіли даних.	Вимагають чіткого розуміння розподілу даних, що моделюються, і можуть бути неефективними для даних зі складними структурами або змінними з часом
Методи машинного навчання та глибокого навчання	Можуть виявляти складні аномалії та залежності між даними.	Вимагають великої кількості даних для тренування. Можуть бути складними для налаштування та оптимізації.
За застосуванням		
Конструктивні застосування	Надають користь та вирішують практичні завдання.	Вимагають розробки специфічних застосунків для кожного випадку.
Деструктивні застосування	Допомагають виявляти та запобігати шкідливим діям та атакам.	Зазвичай потребують додаткових заходів для захисту системи. Можуть призводити до фальсифікації або неправильного реагування.
Застосування для очищення даних	Допомагають видалити непотрібні дані та шум з даних.	Можуть втрачати корисну інформацію. Вимагають заздалегідь відомих шаблонів для очищення.
За типом аномалії		
Пунктові аномалії	Відокремлюють аномалії, які виникають в окремих точках даних.	Можуть пропустити аномалії, які виникають лише в контексті.
Контекстуальні аномалії	Враховують контекст та поведінкові характеристики для виявлення аномалій.	Вимагають складніших аналітичних методів та більше обчислювальних ресурсів.
Колективні аномалії	Визначають аномалії на основі всього набору даних та структури взаємозв'язків між даними.	Можуть бути обчислювально витратними та вимагати великої кількості даних для навчання.
За затримкою		
Online алгоритми	Здатні обробляти дані під час їх збору та аналізувати їх в реальному часі.	Можуть бути обмеженими за ресурсами та вимагати низької затримки.
Offline алгоритми	Мають доступ до всього набору даних і можуть використовувати більш складні обчислювальні методи.	Зазвичай вимагають більше обчислювальних ресурсів та можуть бути повільнішими в роботі.

Обмеження та вимоги до методів виявлення аномалій в IoT

Методи виявлення аномалій включають в себе етап попередньої обробки для визначення нормального діапазону значень, де будь-яке значення в межах визначеного діапазону вважається нормальним. Натомість будь-яке інше значення є винятком. Для потоку даних, залежного від часу, стандартний діапазон значень може змінюватися в залежності від повторюваного циклу, такого як сезон або різні повторювані часові інтервали. Тому правильне визначення повторюваного циклу має вирішальне значення для точності процесу виявлення аномалій.

Отже, слід зазначити наступні важливі обмеження та вразливості:

- визначення довжини повторюваного циклу є найважливішим кроком в аналізі даних IoT; неправильна довжина циклу призводить до невірного виявлення аномалій;

- виявлення аномалій на початку та в кінці кожного циклу є більш складним, оскільки різниця між нормальним станом та аномальним станом є незначною; отже, ймовірність помилки є значною;

- для підтримки точності в стандартних показниках, вимагається постійно перевіряти правильність значень оболонок та їх адаптацію до визначеного циклу і передбачати природні та обґрунтовані зміни в

циклі та відповідні значення, що використовуються для перевірки аномалій з плином часу.[1]

Окрім, того з огляду на результати аналізу, що наведено у попередніх підрозділах, можна сформулювати додаткові обмеження та вимоги до методів виявлення аномалій в IoT. Висока точність: методи виявлення аномалій повинні бути досить точними у виявленні незвичайних подій або аномалій. Особливо важливо виявляти аномалії в реальному часі для запобігання можливим проблемам. Адаптованість до змін: середовище IoT може змінюватися, і методи повинні бути адаптованими до нових умов та типів даних. Вони повинні бути здатними навчатися на нових даних та оновлювати моделі.

Низька обчислювальна складність: оскільки IoT може включати велику кількість пристроїв з обмеженими ресурсами, методи повинні бути ефективними з точки зору обчислень і споживання енергії.

Здатність до роботи в режимі реального часу: деякі випадки виявлення аномалій вимагають негайного реагування. Методи повинні бути здатними працювати в режимі реального часу та виявляти аномалії негайно.

Робота з різними типами даних: IoT може генерувати різноманітні типи даних, від сенсорних даних до великих обсягів текстової інформації. Методи повинні бути придатними для роботи з різними видами даних.

Захист від фальсифікації та атак: методи повинні бути відповідними до заходів з безпеки, оскільки IoT може бути піддається атакам та фальсифікації даних.

Масштабованість: методи повинні бути придатними для роботи в масштабах, що відповідають IoT, де кількість пристроїв і обсяги даних можуть бути дуже великими.

Висновки

У даній статті розглянуто основні методи виявлення аномалій в мережах IoT, проаналізувавши їх ключові переваги та недоліки. Розглянуто методи та підходи, включаючи геометричний, статистичний методи та методи машинного та глибокого навчання, зі спеціальним акцентом на їх застосовність в контексті IoT.

Напрямком подальших досліджень є пошук способу оптимізації існуючих методів виявлення аномалій в мережах IoT. Це включає пошук нових стратегій для підвищення точності та надійності методів виявлення, а також зниження впливу помилкових спрацьовувань.

СПИСОК ЛІТЕРАТУРИ

1. Parimala, V. K. (Ed.). (2024). Anomaly Detection - Recent Advances, AI and ML Perspectives and Applications.IntechOpen.DOI: 10.5772/intechopen.110988. ISBN: 978-1-83769-027-5.
2. Коваленко А. А., Кучук Г. А. Методи синтезу інформаційної та технічної структур системи управління об’єктом критичного застосування. *Сучасні інформаційні системи*. 2018. Т. 2, № 1. С. 22–27. DOI: https://doi.org/10.20998/2522-9052.2018.1.04
3. Ruban, I.V., Martovytskyi, V.O., Kovalenko, A.A. and Lukova-Chuiko, N.V. (2019), “Identification in Informative Systems on the Basis of Users' Behaviour”, Proceedings of the International Conference on Advanced Optoelectronics and Lasers, CAOL 2019-September,9019446, pp. 574-577, DOI: https://doi.org/10.1109/CAOL46282.2019.9019446
4. Kovalenko, A., Kuchuk, H., Kuchuk, N. and Kostolny, J. (2021), “Horizontal scaling method for a hyperconverged network”, *2021 International Conference on Information and Digital Technologies* (IDT), Zilina, Slovakia, doi: https://doi.org/10.1109/IDT52577.2021.9497534
5. Kuchuk, N., Kovalenko, A., Ruban, I., Shyshatskyi, A., Zakovorotnyi, O. And Sheviakov, I. (2023), “Traffic Modeling for the Industrial Internet of NanoThings”, *2023 IEEE 4th KhPI Week on Advanced Technology, KhPI Week 2023 - Conference Proceedings*, 194480, doi: http://dx.doi.org/10.1109/KhPIWeek61412.2023.10312856
6. Li, H., Boulanger, P. A Survey of Heart Anomaly Detection Using Ambulatory Electrocardiogram (ECG). Sensors. 2020; 20(5): 1461. DOI: 10.3390/s20051461.
7. Cook, A. A., Mısırlı, G., & Fan, Z. (2020). Anomaly Detection for IoT Time-Series Data: A Survey. IEEE Internet of Things Journal, 7(7), 6481–6494.
8. M. Fahim, A. Sillitti, Anomaly detection, analysis and prediction techniques in IoT environment: A systematic literature review, IEEE Access 7 (2019) 81664–81681
9. Chatterjee, A., & Ahmed, B. S. (2022). IoT Anomaly Detection Methods and Applications: A Survey. Internet of Things, 100568. Elsevier BV.

Received (Надійшла) 29.11.2023
Accepted for publication (Прийнята до друку) 31.01.2024

Analysis of Methods for Detecting Anomalous Traffic in IoT Networks

Roman Marchenko, Andriy Kovalenko, Vasyl Znaidiuk

Abstract: The aim of this work is to conduct a comprehensive analysis of methods and approaches for anomaly detection in Internet of Things (IoT) networks. Considering the rapid development of IoT and the increasing number of connected devices, the problem of detecting anomalous traffic becomes crucial for ensuring the security and efficiency of these networks. This study examines various methods and approaches to anomaly detection, including statistical analysis, network monitoring, behavioral analysis, as well as the application of modern machine learning and deep learning technologies. Each of these methods is considered from the perspective of its applicability in the context of IoT and its advantages and limitations are evaluated. The work also explores current challenges and future prospects in the field of IoT security, with a focus on protection against cyber threats and the enhancement of anomaly detection systems.

Keywords: Internet of Things, anomalous traffic, machine learning, deep learning, statistical analysis.