ISSN 2222-2944. Інформаційні технології: наука, техніка, технологія, освіта, здоров'я. 2023

ДОСЛІДЖЕННЯ МЕТОДІВ КЛАСТЕРІЗАЦІЇ ДАННИХ ДЛЯ ІДЕНТИФІКАЦІЇ ВТОРГНЕНЬ В КОМП'ЮТЕРНІ МЕРЕЖІ

Гавриленко С.Ю., Абдуллін О.Р.
Національний технічний університет «Харківський політехнічний інститут», м. Харків

У сучасному світі Інтернет-безпека має важливе значення через збільшення кількості даних, які передаються мережею. Протокол телеметрії з чергою повідомлень (MQTT) є одним із найпоширеніших стандартів, що використовується для передачі даних. Збільшення кількості доступних пристроїв і протоколів які використовуються, посилює потребу в нових і надійних системах виявлення вторгнень (IDS).

Для вирішення цієї проблеми можна використати методи машинного навчання. Однією із складових машинного навчання є навчання без вчителя. Навчання без вчителя використовується за умови нерозмічених вихідних даних. Складовою навчання без вчителя є кластеризація – процес розподілу об'єктів на кластери, які мають однакові між собою характеристики.

У даній роботі було дослідженню методи кластеризації даних: K-means, DBSCAN та ін. Розроблено програмні моделі з використанням хмарного середовища Google Colab та мови програмування Python.

У якості вихідних даних було використано набір даних MQTT-IoT-IDS2020, який генерується за допомогою змодельованої мережевої архітектури MQTT та складається з дванадцяти датчиків, брокера, імітованої камери та зловмисника.

Для оцінки якості кластеризації використано силуетну оцінку та індекс Девіса Боулдіна. Силуетна оцінка визначає ступінь подібності між об'єктами одного кластеру та відмінність між об'єктами в різних кластерах. Індекс Девіса-Булдіна є відношенням відстаней усередині кластера до відстані між кластерами.

Здійснено попередню обробку даних. Визначено оптимальну кількість кластерів. Виконано налаштування моделей. За результатами дослідження отримано, що модель на основі алгоритму K-means є більш якісною. Силуетна оцінка кластеризації дорівнює 0.95, індекс Девіса Боулдіна дорівнює 0.75.

Отримана модель може бути використана у якості складової системи виявлення вторгнень у мережі.