

ЗАСТОСУВАННЯ МЕТОДУ ГОЛОВНИХ КОМПОНЕНТ ДЛЯ АНАЛІЗУ АНОМАЛІЙ В МЕРЕЖЕВОМУ ТРАФІКУ

Зростання обсягу мережевого трафіку та ускладнення архітектури мереж призводять до виникнення нових викликів у забезпеченні інформаційної безпеки. Аналіз аномалій у мережевому трафіку є критично важливим для забезпечення безпеки та стабільності комп'ютерних мереж. Зі зростанням обсягів даних та складності мережевих інфраструктур, традиційні методи виявлення аномалій стають менш ефективними. У цьому контексті метод головних компонент виступає як потужний інструмент для пониження розмірності даних та виявлення прихованих патернів у великих наборах інформації [1].

Метод головних компонент, або звично PCA, є одним з найбільш поширених підходів для пониження розмірності даних, що використовується у візуалізації чи безпосередньо в алгоритмах машинного навчання. Фахівці з безпеки працюють з наборами даних великої розмірності. Буває корисним виділити у даних більш "значущі" її частини, щоб зрозуміти, як вони поєднуються між собою, виявити приховані зв'язки у деякому латентному просторі, або просто зобразити n -вимірний простір у вигляді дво- чи тривимірного графіку. Попри те, що метод головних компонент активно використовується в аналітиці та побудові різних моделей машинного навчання, досить часто механіка та елегантність, що схована за математикою PCA, є незрозумілою або неінтуїтивною.

Метод головних компонент належить до лінійних методів зниження розмірності, що дозволяє виявляти в даних основні напрямки варіацій, які максимально пояснюють дисперсію вихідного набору даних. У випадку аналізу мережевого трафіку, PCA дозволяє зосередитися на найбільш значущих характеристиках, що полегшує ідентифікацію аномалій, які можуть свідчити про потенційні загрози. PCA проектує багатовимірний простір в меншу кількість осей, дозволяючи виділити приховані патерни та можливі зв'язки між характеристиками мережевого трафіку [2].

Фахівці з безпеки кіберпростору часто взаємодіють з великими обсягами даних із багатьма змінними, що можуть ускладнювати аналіз і пошук аномальних патернів. Метод PCA стає корисним інструментом для таких випадків, адже дозволяє виділити найбільш значущі компоненти, які зберігають максимальну частку інформації з оригінального набору даних. PCA дозволяє звести багатовимірний простір, що містить безліч характеристик мережевого трафіку до двох-трьох вимірів, де відхилення від

норми можна візуалізувати та проаналізувати за допомогою спрощеної моделі.

Для застосування PCA потрібна ретельна попередня підготовка даних. Дані повинні бути нормалізовані, оскільки PCA є лінійним методом і чутливий до масштабування змінних. Якщо змінні мають різні одиниці виміру, метод може неправильно визначити вагомість характеристик. Попри свою ефективність, PCA здатний виявляти лише лінійні залежності між змінними, що може обмежувати застосування цього методу для складніших нелінійних патернів, характерних для деяких типів мережевого трафіку.

Метод PCA має як переваги, так і недоліки, що варто враховувати під час його використання для аналізу аномалій (див. таблиця 1).

Таблиця 1 – Переваги та недоліки методу головних компонент

Переваги	Недоліки
Ефективний для великих наборів даних	Потребує ретельної нормалізації даних
Проста математична основа	Виявляє лише лінійні залежності, що може бути недостатнім для складних мережевих патернів
Дозволяє сконцентрувати увагу на найважливіших характеристиках трафіку	Для побудови надійної моделі необхідні обширні дані нормального трафіку

Для застосування методу PCA у виявленні аномалій у мережевому трафіку необхідно мати достовірний та обширний набір даних нормального трафіку для створення моделі, що визначатиме аномалії шляхом порівняння нових даних із нормою. Постійний моніторинг та актуалізації даних є важливим елементом, оскільки безпека мережі є динамічним процесом, де з часом можуть виникати нові патерни.

Завдяки своїй здатності знижувати розмірність даних та виділяти найбільш значущі компоненти PCA є ефективним інструментом. Його використання дозволяє спростити багатовимірний аналіз, зосередившись на основних характеристиках, що пояснюють відхилення у мережевих даних, що допомагає не тільки зменшити обсяг оброблюваних даних, а й виявити приховані закономірності у трафіку. Проте, для успішного застосування методу необхідно враховувати його обмеження. Оскільки метод визначає лише лінійні залежності, він може бути недостатньо точним для характерних сучасним мережам складніших патернів. До того ж, побудова надійної моделі потребує ретельної підготовки даних і використання широкого набору нормального трафіку.

Метод може слугувати засобом для попереднього аналізу даних у системах виявлення аномалій, надаючи можливість фахівцям з кібербезпеки швидко зосередитися на ключових аспектах мережевого трафіку, що потребують подальшої уваги. Для покращення точності можна розглядати використання РСА у комбінації з іншими методами, що дозволяють обробляти нелінійні взаємозв'язки та підвищують ефективність виявлення аномалій у складних мережах.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Signature vs. Anomaly-Based Detection: Which Is More Effective? URL: <http://surl.li/fioodi> (Дата звернення: 23.10.2024).
2. Математика машинного навчання. Метод головних компонент. Частина 2. URL: <http://surl.li/rtsglb> (Дата звернення: 24.10.2024).