УДК 004.056.5
Б. Паперовський, Н. Загородна
(Тернопільський національний технічний університет імені Івана Пулюя)

АНАЛІЗ СУЧАСНИХ МЕТОДІВ ВИЯВЛЕННЯ АНОМАЛІЙ ТА МОЖЛИВИХ ВТОРГНЕНЬ В РОБОТІ КОМП'ЮТЕРНОЇ СИСТЕМИ

В сучасному світі відбувається шалений розвиток комп'ютерних та інтернет-технологій, тому однією з найбільш актуальних проблем суспільства стає інформаційна безпека та її складова – кібербезпека, від якої залежить функціонування всіх сучасних комп'ютерних систем (КС) у промисловості, енергетиці, транспорті, медицині і т.д.

Висока кваліфікація кіберзлочинців дозволяє створювати та використовувати унікальні, ще не відомі ІТ-індустрії шкідливі програми, знаходити нові вразливості в програмних продуктах, роботі КС і використовувати їх для проведення комплексних кібератак. Протистояти постійному зростанню кількості й складності деструктивних впливів на КС можна, зокрема й використовуючи інтелектуальні системи розпізнавання кіберзагроз, які базуються на методах машиного навчання (Data Mining).

Алгоритми Data Mining використовуються для виділення нової важливої інформації з вибірки великих даних. В умовах постійного збільшення об'єму даних, а також зростання важливості результатів аналізу цих даних, питання ідентифікації аномалій в роботі КС стає особливо актуальним. Аномалія – це відхилення поведінки системи від стандартної. Результати аналізу без попереднього виключення аномальних екземплярів даних можуть бути сильно спотворені.

Аномалії можуть бути віднесені до одного із трьох основних типів:

Точкові аномалії виникають в ситуації, коли окремий екземпляр даних може розглядатися як аномальний по відношенню до основних даних.

Контекстуальні аномалії спостерігаються, якщо екземпляр даних є аномалією лише в якомусь певному контексті. Для виявлення цих аномалій основними виділяються контекстуальні і поведінкові атрибути.

Колективні аномалії з'являються, коли послідовність пов'язаних екзеплярів даних є аномалією для цілого набору даних.

В доповіді будуть розглянуті методи машинного навчання без учителя (unsupervised learning). Продемостровано результати застосування методів аналізу "ізольований ліс" (isolation forest) та LOF (Local Outlier Factor) до даних про навантаженість процесора для виявлення аномалій його роботи. Запропоновано удосконалений метод, що дозволяє підвищити точність виявлення аномалій на основі комбінації зазначених вище методів.

Література

1. S. Agrawal, J. Agrawal, "Survey on Anomaly Detection using Data Mining Techniques", Procedia Computer Science, vol. 60, 2015, pp. 708 – 713.

2. C. Chio, D. Freeman. "Machine Learning and Security", O'Reilly Media, Inc., – [Електроний ресурс] – Режим доступу: https://github.com/oreilly-mlsec/book-resources/tree/master/chapter3/datasets/cpu-utilization. – 01.12.2017.