Додаток А

Графічний матеріал

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ РАДІОЕЛЕКТРОНІКИ

КАФЕДРА АПОТ

Кваліфікаційна робота магістра

Метод виявлення програм-вимагачів з використанням машинного навчання

Виконала:
студентка групи СКСм-20-1
Ейхман Тетяна Ігорівна

Керівник:
доц. кафедри АПОТ
Адамов Олександр Семенович

Харків 2021

1

Актуальність теми

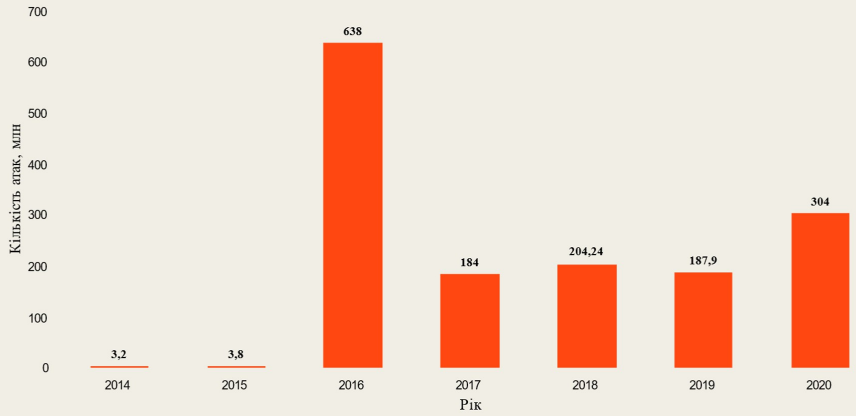
З розвитком технологій збільшується кількість сфер нашого життя, що стають автоматизованими. На комп'ютерні системи з кожним роком покладається все більше і більше відповідальності.

Жертвами хакерських атак може будь хто, а збитки від них сягають мільярдів доларів. Від програм-вимагачів страждають виробництво, торговий сектор, урядові та освітні організації, заклади охорони здоров’я, комунальне господарство та ін.

Зловмисники активно застосовують засоби автоматизації кібератак, використовуючи штучний інтелект для оптимізації своєї діяльності.

Для того щоб впоратися зі зростаючим обсягом атак, необхідно розпочати активне впровадження технологій штучного інтелекту, машинного і глибокого навчання для виявлення і прогнозування кіберзагроз, а також реагування на них в режимі реального часу.

Актуальність теми



Кількість атак з використанням програм-вимагачів у світі за 2014-2020 рр.

Мета та задачі атестаційної роботи

Метою кваліфікаційної роботи є практичне застосування обраного методу машинного навчання для детектування програм-вимагачів.

Для цього поставлені наступні задачі:

- Вивчення предметної області
- Дослідження ролі машинного навчання вкібербезпеці
- Аналіз програм-вимагачів
- Практична реалізація обраного методу - алгоритму кластеризації Affinity Propagation
- Оцінка якості реалізованого алгоритму

Інструменти та засоби реалізації

- мова Python версії 3.9
- Jupyter Notebook – інтерактивне середовище розробки

Бібліотеки для роботи

Програма-вимагач (ransomware) - це

тип шкідливого програмного забезпечення, який надає злочинцям можливість віддалено заблокувати комп'ютер чи інший пристрій або зашифрувати дані на ньому. Після цього програма відображає спливаюче вікно з повідомленням, що доступ до пристрою чи даних буде повернуто після сплати викупу.

В останні роки до простого блокування пристрою чи шифрування даних додались викрадення даних та загроза оприлюднення чи передачі таких даних третім особам.

Машинне навчання може бути корисним у напрямках:

- класифікації даних на основі заданих параметрів;
- кластеризації даних на основі їх подібності або аномальності, якщо вони не відповідають заданим параметрам;
- рекомендації підходів та рішень на основі минулих випадків;
- генерування варіантів дій, що можна застосувати до нововиявлених даних;
- прогнозування на основі вже існуючих наборів даних та результатів.

Вибірка даних для роботи

Для реалізації алгоритму в якості датасета було використано журнал подій на хмарної інфраструктури Google Cloud Platform (GCP) після моделювання атаки на неї.

Журнал подій для кожного об’єкта-події, містить одинадцять ознак, з усіх ознак було обрано шість ключових: назва об’єкту, користувач, дата, назва події, тип об’єкту, IP-адреса.

Affinity Propagation

Для реалізації було обрано алгоритм кластеризації Affinity Propagation (поширення схожості).

Він заснований на концепції попарного «обміну (поширення) повідомленнями» між точками – об’єктамивибірки, що кластеризується.

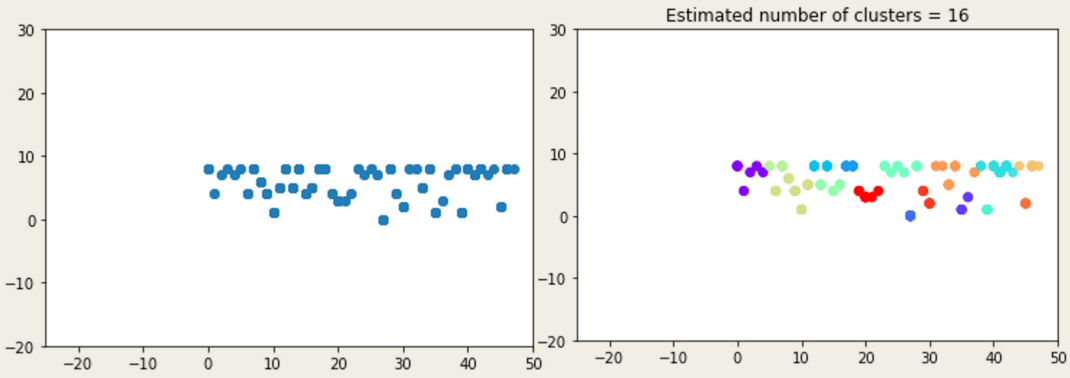
Affinity Propagation

Алгоритм самостійно визначає кількість кластерів і зразкові елементи серед усіх об’єктів вибірки даних – точок даних і формує кластери з даних навколо них. Припускаючи, що кожен з об’єктів вибірки даних може бути центром кластеру, Affinity Propagation уникає проблем інших алгоритмів, пов’язаних з невдалою ініціалізацією

Affinity Propagation

- Алгоритм виконується шляхом чергування двох етапів передачі повідомлень, про відповідальність (наскільки певний об’єкт вибірки підходить для того, щоб служити зразком в порівнянні з іншими об’єктами-кандидатами) та доступність елемента (наскільки для об’єкта доречно обрати одного кандидата як зразок відносно інших кандидатів).
- Отже, кожен з об’єктів вибірки відправляє повідомлення, яким інформує інші об’єкти про свою доступність і про готовність брати відповідальність за інші об’єкти, а також приймає такі повідомлення від інших об’єктів.

Результат роботи алгоритму



Оцінка роботи алгоритму

Оцінку якості роботи алгоритму Affinity Propagation було проведено внутрішніми методами, тобто такими, що засновані на оцінці якості по відношенню до деяких наших уявлень про те, якою повинна бути хороша кластеризація.

- Індекс Силуету = 0.762
- Індекс Девіса-Болдуіна = 0.479
- Індекс Calinski-Harabasz = 1238.884

Висновки

- Проведено аналіз програм-вимагачів, визначено застосування машинного навчання в кібербезпеці;
- Виконано практичну реалізацію алгоритму кластеризації Affinity Propagation;
- Протестовано та оцінено якість роботи реалізованого алгоритму.