ВИКОРИСТАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ВИЯВЛЕННЯ АТАК НА БЛОКЧЕЙН СИСТЕМИ

Просолов В.В.
Харківський національний університет радіоелектроніки, Харків, Україна

Блокчейн став всюдисущим поняттям, який охоплює соціальні перспективи та нову технологію[1].

Спочатку запропонований як рішення для системи ведення реєстрів криптовалюти Bitcoin, блокчейни тепер використовуються для зберігання записів для багатьох типів програм. Блокчейн асоціюється для більшості людей із чимось більшим, ніж просто світ криптовалют.

Обіцянка, яку багато хто асоціює з блокчейн-додатками, полягає в тому, що вони зруйнують централізовані системи. Використовуючи централізовані системи, люди повинні довіряти контрагенту, але самі не мають ресурсів, щоб зробити це самостійно [2].

Таким чином, простий спосіб визначити місце, де застосовується технологія блокчейн, - це пошук областей, у яких потрібен посередник для сприяння довірі.

Довіра має важливе значення для таких речей, як переказ грошей, голосування, земельні записи, права інтелектуальної власності та ідентифікація. Програмне забезпечення блокчейна використовується для того, щоб замінити посередників, ставши надійним гарантом для системи ведення записів [3].

Мета доповіді полягає в тому, щоб розглянути метод виявлення атак на блокчейн системи з використанням методів машинного навчання. Для цього розглядається та аналізується система блокчейна з точки зору її вразливості щодо шкідливого впливу.

У даній роботі представлено метрики оцінки для цього дослідження, порівняння результатів та обговорення нашого експерименту. Метрики оцінки є дуже важливою частиною виміру якості та продуктивності нашого рішення. Оцінки метрики були зроблені з використанням інтерфейсу розробки Python і відповідних бібліотек Python.

Ми використовували набір даних для 10000 записі. Потім цей набір даних з файлу .csv було поділено з використанням бібліотек Pandas і Numpy.

Окремо ми виділили бінарні змінні, не враховуючи цільову. Після цього проведено попередню обробку даних з метою класифікації, використовуючи алгоритм, розглянутий у попередньому розділі.

У бібліотеці scikit-learn є багато алгоритмів класифікації, які ми могли б використовувати для побудови моделі. У цьому прикладі ми використовуватимемо класифікатор на основі модернізованого методу k-NN, який легко інтерпретувати.

Побудова цієї запропонованої моделі полягає лише у запам'ятовуванні навчального набору. Для того, щоб зробити прогноз для нової точки даних,

алгоритм обирає точку в навчальному наборі, яка знаходиться найближче до нової точки. Потім він надає мітку, що належить цій точці навчального набору, новій точці даних.

У scikit-learn всі моделі машинного навчання реалізовані у власних класах, які називають класами Estimator.

Алгоритм класифікації на основі методу k-NN реалізований у класифікаторі KNeighborsClassifier модуля neighbors. Оскільки цей класифікатор працює за шістнадцятьма класами, то цілком доцільно вибрати змінну числа сусідів n_neighbor=5.

Був проведений ряд експериментів із розбиттям на тестовий та навчальний набір даних. Слід зазначити, що якість моделі, яка визначає правильність (accuracy), це сильно впливає. Зокрема, за всіх заданих умов accuracy не перевищувала 0.89, що фактично визначає вірогідність правильного прогнозу.

Подальший аналіз проведено з використанням методики роботи [4]. Алгоритми k-NN, Random Forest і Decision Tree порівнювали з використанням таких показників оцінки: accuracy, recall, precision та F1. Зокрема, для алгоритму k-NN результати accuracy для 10 000 даних дорівнюють 75.0%. Для поданої моделі - 75.0 %.

Оскільки запропонована модель була тестова, і призначена для відпрацювання методики двоетапного експерименту. Тим не менш, якщо ми цікавимося шкідливими атаками на систему блокований, то часто немає необхідності проводити повномасштабне дослідження даних. Тим більше, що ніхто не дозволить використовувати конфіденційну статистику. Якщо виділити бінарні змінні і за ними провести класифікацію, то остаточний результат такої урізаної моделі показує результат не гірше, ніж повномасштабне дослідження.

Методи аналізу даних широко використовувалися в галузі кібербезпеки та захисту інформації, а недавнє поширення передових методів машинного навчання дозволило точно ідентифікувати кібератаки та виявляти загрози як у режимі реального часу, так і під час аналізу після інциденту.

Список літератури

1. Laurence T. *Introduction to blockchain technology*. Amersfoort – NL: Van Haren Publishing, 2019. 250 p.

2. Ye, C., Li, G., Cai, H., Gu, Y., & Fukuda, A. (2018, September). Analysis of security in blockchain: Case study in 51%-attack detecting. In *2018 5th International conference on dependable systems and their applications (DSA)* IEEE, 2018. p. 15-24.

3. Khan S. N., Loukil F., Ghedira-Guegan C., Benkhelifa E., & Bani-Hani A. Blockchain smart contracts: Applications, challenges, and future trends. *Peer-to-peer Networking and Applications*. 2021. Vol.14. P. 2901-2925.

4. Sanda, O., Pavlidis, M., Seraj, S., & Polatidis, N. Long-Range attack detection on permissionless blockchains using Deep Learning. *Expert Systems with Applications*. 2023. Vol. 218. P. 119606.