Keshav Khanna

Professor Shivakumar Mathapathi

165671 Data Analytics Using Python

20 August 2022

Anomaly Detection System for Blockchain Transactions

Introduction:

Firstly, let's introduce the idea of blockchain transactions. Blockchain transactions allow the users to transfer the ownership of digital resources (e.g., cryptocurrency) directly to each other within a decentralized and distributed peer-to-peer network, without the need of a third party (e.g., financial institution). The execution of transactions transfers the ownership of resources, and the global state stored in the blockchain network is modified. Each transaction broadcasted to the blockchain network must be digitally signed by the user and performed by a digital signature algorithm. The usage of digital signatures as one of the main building blocks of blockchain technology is necessary to guarantee the integrity and the non-repudiation of the transactions.

Concerning the results of studies on the usability of digital signatures, there are still many obstacles preventing users from accepting the usage of digital signatures within their everyday usage. These obstacles are related to managing, controlling, and indeed using cryptographic keys. The studies conclude that such tasks are complex and consequently overwhelming for users. Moreover, studies have pointed out that these usability issues also have an impact on users' security, e.g., users are unable to recognize potential intrusions while being involved in a digital signing process.

Most of the aforementioned digital signatures' usability issues are related to the complexity of the digital signing process. Blockchain technology can affect the security of the digital assets owned by users because they are eligible to transfer digital assets only with a correctly and successfully digitally signed transaction. Each time a user wants to transfer a digital asset, a digitally signed transaction needs to be provided, which can be performed only within the aforementioned complex process of digital signing. In the case of frequent transfers over a longer period, the digitally signing process can be performed superficially, thereby affecting the security of the digital assets. In this situation, potential malicious counterparties can use such user behavior as an advantage and attempt to propose to users the signing of transactions that have adverse impacts on their digital resources.

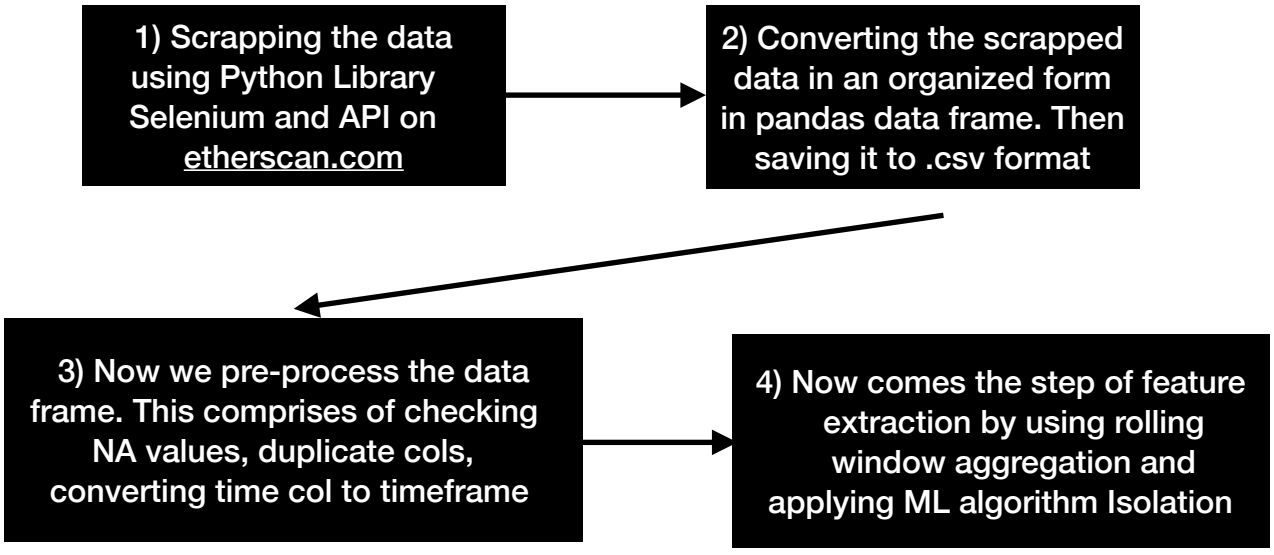
Overview of the project:

This project proposes an anomaly detection mechanism, based on the user's personalized transaction data. By incorporating the proposed method into blockchain dedicated software (e.g., wallet) for managing the digital signature process, the designated transaction, which includes the transfer of cryptocurrency from address A (i.e., sender) to address B (i.e., receiver), is automatically digitally signed on behalf of the sender, except in the case of a potentially anomalous transaction detection, which can potentially produce damage to the sender. In such a scenario, manual approval from the sender to digitally sign such a transaction is required. Properties of the proposed method also help to circumvent existing digital signatures usability issues that can be indirectly present in the usage and, with this, constitute an obstacle to the wide adoption of blockchain technology.

Problem statements:

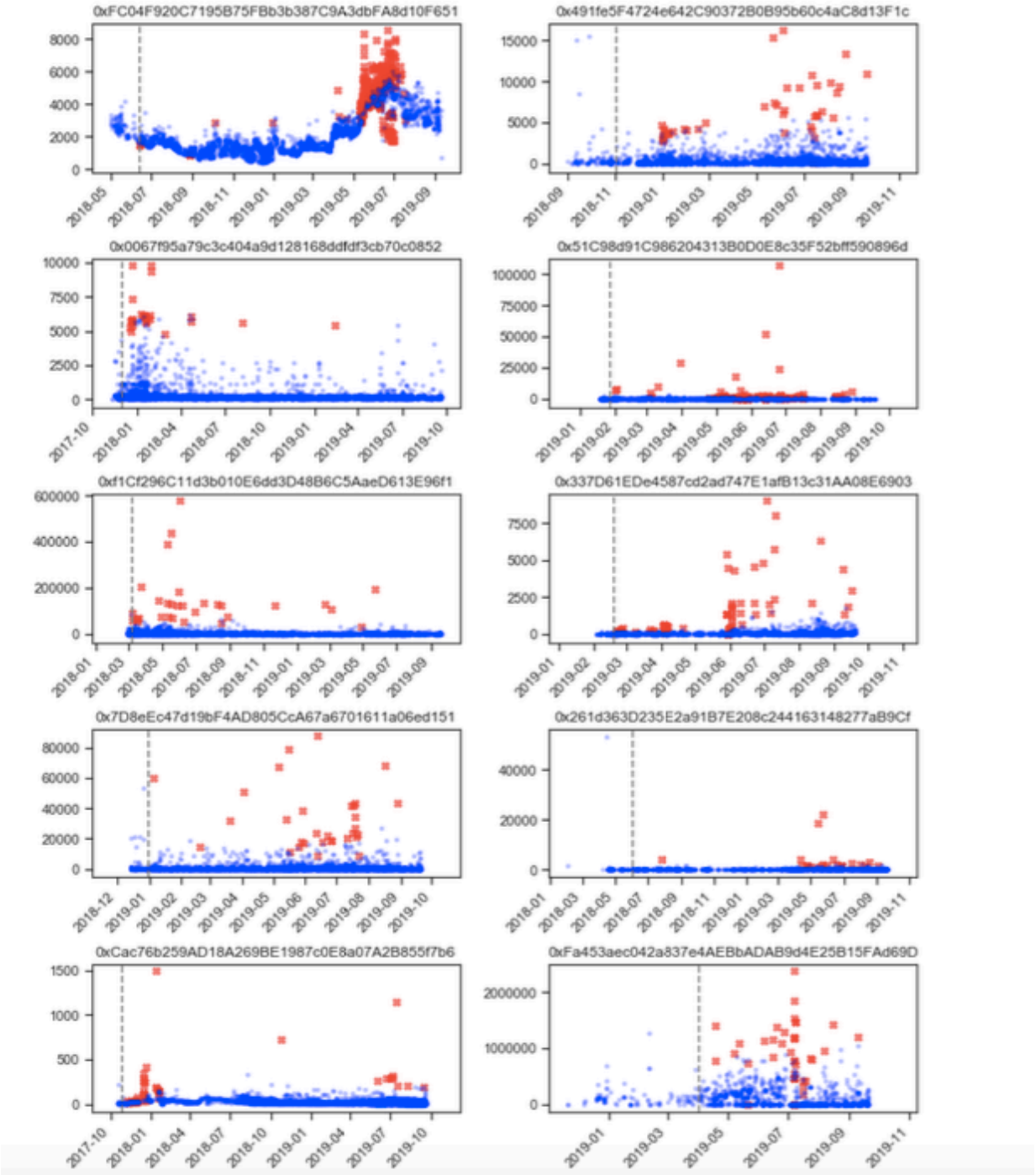
As of today, there have been related works in the field of fraud detection on blockchain networks. But there are no existing solutions whose purpose is to introduce machine learning methods into the blockchain transaction digital signing process, intending to enable automatic digital signing of transactions, while additionally actively detecting potentially fraudulent transactions which are subject to the digital signing performed by the user. Although the work on detecting anomalies among the transactions has already been done, none of the reviewed methods adapt to the address' transaction patterns, and none reviewed the approach of using user data labeled (the anomalies are not known). The novelty of the proposed approach is that it can be used with any address that has a sufficient history of the transaction, regardless of the transaction patterns, and identified anomalous transactions.

Block Diagram:



Data Analytics:

After we would have performed our machine learning algorithm, we would get the following results. The results of anomalous transaction detection are presented in a time series chart form in the figure below. Anomalous transactions are denoted as red crosses, and normal transactions are denoted as blue dots. The first 100 transactions were only used as the starting training data, thus separating them with the dashed line. Moreover, here a sample of 10 such addresses is shown. But in our model we utilize close to 1000 addresses.



Machine Learning Algorithm:

We would use an Isolation Forest ML algorithm consisting of 100 individual decision trees with a contamination factor of 0.01. It's an unsupervised learning algorithm that identifies anomalies by isolating outliers in the data. It isolates the outliers by randomly selecting a feature from the given set of features and then randomly selecting a split value between the max and min values of that feature. This random partitioning of features will produce shorter paths in trees for the anomalous data points, thus distinguishing them from the rest of the data. The Random Forest classification algorithm would be used to determine the feature importances and was also constructed out of 100 individual decision trees, built with Gini impurity split criteria. The experiment would be implemented using the sci-kit-learn v0.21.3 Python package. All of the other settings are left at their default values.

Summary:

In this project, we are experimenting with real-life transactions from the Ethereum public main network to determine malicious transactions. Although there is no scientifically valid metric to determine the quality of unsupervised anomalous transaction detection, the proposed method returns promising results.

Works Cited

Podgorelec, Blaž. Turkanović, Muhamed. Karakatič, Sašo. “A Machine Learning-Based Method for Automated Blockchain Transaction Signing Including Personalized Anomaly Detection,”25 December 2019, https://www.mdpi.com/1424-8220/20/1/147.