Detecting Malicious Behaviors on Ethereum

Hallee Ray
Computer Science Post Bacc.
University of Colorado - Boulder
Boulder, Colorado, USA
hara7620@colorado.edu

BD Tinsley
Computer Science Post Bacc.
University of Colorado - Boulder
Boulder, Colorado, USA
beti7384@colorado.edu

ABSTRACT

The goal of our project was to elucidate a set of criteria that could potentially indicate fraudulent activity in Ethereum trading. The questions that we sought to answer included identifying trends and expected values for each of the features of the transaction data, with a speculation that outliers could indicate suspicious behavior. However, it is important to note that our dataset did not include any true labels as to whether any of the transactions were proven to be fraudulent. Therefore, all of our results are conjectured.

We utilized K-Means to cluster and characterize the wallets. Interestingly, we identified a single wallet address across several time periods that appeared as a cluster outlier. Looking into the various features present in this cluster, we identified that this wallet was involved in far more transactions than average and executed transactions with higher than average values. Despite not being able to reliably determine if any of the transactions in the dataset were fraudulent, our data exploration and modeling helped us to generate an idea of what kind of activity we would assume to likely be fraudulent.

KEYWORDS

Ethereum blockchain, anomaly detection, data mining, prediction modeling, fraud detection, feature engineering, machine learning

1 INTRODUCTION

The goal of our project was to analyze historical Ethereum transaction data to generate a model that could identify if a transaction was likely fraudulent or not. With the rise of cryptocurrency, there is increasing concern regarding its validity and safety. Many scams and fraudulent activity have been reported and monitored, however, due to the decentralized and unregulated nature of cryptocurrency fraud and scams are common.

The questions that we were interested in answering about this dataset were aimed at providing us with a framework with which to identify fraudulent or suspicious activity in Ethereum transactions. Due to the nature of our dataset, we could not ascertain with any certainty whether a transaction or wallet is fraudulent, however, we were able to identify several outliers that we suspect are likely fraudulent.

The questions that we originally wanted to answer included exploring correlations between the transaction features such as the time the transaction was executed and the value of the transaction. Intuitively, we would suspect that fraudulent activity could occur outside of a wallet's typical active hours, and we wanted to explore if the value of transactions outside of expected hours differ from the wallet's usual transaction values. Our dataset included several sub datasets that included all transactions within a time period. Due to memory restrictions, we could only reliably analyze transactions from about a month period at a time, therefore, exploration into this particular question was not meaningful due to the lack of representation.

What is the average time between a new coin going live and the first purchase? What are the average values of the transactions during the lifetime of Ethereum, how does it map to early-stage hype? Initially, we considered that the answers to these questions would help us to identify pump and dump type of activity, where investors drive up the cost of a coin by generating hype and buying up the resource and selling when the price has been gouged.

For coins with historical fraudulent activity, what are the general statistics for the transactions that were reported as fraudulent? The goal of this question was to generate a baseline understanding of what kinds of behavior are fraudulent, however, the dataset we utilized for the project lacked information regarding the validity of the transactions. Therefore, while this would be an important question to have answered, we were unable to.

Due to the restrictions of our dataset, we pivoted away from these questions to focus on questions that would allow us to discern if a transaction is fraudulent using domain knowledge. For example, exploring statistical metrics for the values that are sent and received helped us to determine if a particular transaction was unusual due to an unusual value. Following this idea, we generated several variables that we used as reference points for determining if a transaction was unusual. We identified several extreme outliers that skewed the mean for many of the features and highlighted the variance in the data points.

2 RELATED WORK

2.1 The Literature of Fraud Data and Identification

Unfortunately, scams and fraudulent activity in cryptocurrency occur frequently and are difficult to monitor due to the anonymous and decentralized nature of cryptocurrency. The Federal Bureau of Investigation (FBI) and the Homeland Defense & Security Information Analysis Center (HDIAC) have made efforts to monitor and predict fraudulent behavior in cryptocurrency.

"In February 2022, the FBI formed the Virtual Assets Unit (VAU), a specialized team dedicated to investigating cryptocurrency-related crimes" [5]. In the report, the FBI identified common types of malicious behaviors using cryptocurrencies: investment fraud, recovery schemes, and kiosk scams. Investment fraud is described as malicious actors who encourage individuals to make "investments" in cryptocurrency. The FBI found that losses from cryptocurrency-related investment fraud schemes reported to the IC3 rose from $2.57 billion in 2022 to $3.96 billion in 2023, an increase of 53

Recovery schemes involve representatives from fraudulent businesses claiming to provide cryptocurrency tracing and promise an ability to recover lost funds. They request an upfront fee for this service and then immediately cease communication with the sender once the funds are received. Kiosk scams are scams where a criminal actor will direct an individual to withdraw cash from their bank and locate and send money via a kiosk or ATM-like machine. In 2023, the IC3 received more than 5,500 complaints reporting the use of cryptocurrency kiosks, with losses over $189 million.

Knowing the existence of these scams and monitoring statistical metrics regarding cryptocurrency transactions and fraudulent activity is not sufficient to predict and prevent such scams. The Homeland Defense & Security Information Analysis Center (HDIAC) has employed various real-time predictive machine learning models and on- and off-chain monitoring to detect fraudulent activity. The HDIAC describes the purpose of the off-chain module as preventing fraud before it occurs, whereas the on-chain module utilizes real-time surveillance to detect fraud after it has occurred [6].

3 DATA SET

The dataset we used were entire blocks of transaction data on the Ethereum ledger. Our dataset incorporated block 100400 to block 1388733. In total, this dataset included nearly 2.8 million transactions. Depending on the transaction size, any given block could contain between 1 to 149 transactions, with a block containing 3 transactions on average. We were able to pull this data directly from the Ethereum ledger through a paid third party service, Infura [7]. In all, the transactions we obtained covered many days between August 17, 2015 and April 23, 2016.

The attributes the transaction dataset contained were

- `hash`: The 64-bit unique hex identifier for the specific transaction.
- `nonce`: A unique, sequential integer assigned to every transaction sent from a specific Ethereum wallet address.
- `block_hash`: The 64-bit unique hex identifier of the associated block.
- `block_number`: The integer value of the associated block.
- `transaction_index`: The integer index value of this transaction on the associated block.
- `from_address`: The 64-bit unique hex identifier of the wallet account sending Ethereum.
- `to_address`: The 64-bit unique hex identifier of the wallet account receiving Ethereum.
- `value`: The integer value of Ethereum used in the transaction.
- `gas`: The integer value that denotes the computational work required to execute the transaction as set by the network.
- `gas_price`: The integer-valued cost per unit of computational work set by the `from_address` in Ethereum.
- `input`: The hex value data payload of the transaction.
- `block_timestamp`: The integer value that represents the time its associated block was created or mined.
- `max_fee_per_gas`: A newly incorporated feature that specifies the maximum total gas price the `from_address` is willing to spend for this transaction [2].
- `max_priority_fee_per_gas`: Another newly incorporated feature which specifies the maximum tip in Gwei (ETH 10^{-9}) that the `from_address` is willing to send to validators to prioritize this transaction [2].
- `transaction_type`: An integer value to denote the type of transaction this one is, as incorporated in EIP-2718 [3].
- `max_fee_per_blob_gas`: The integer value in Gwei the `from_address` is willing to spend to store larger "blob" type data in storage, as incorporated in EIP-4844 [4].
- `blob_versioned_hashes`: Hex-valued addresses to the data storage blobs, as incorporated in EIP-4844 [4].

Due to the daily limits imposed on the Infura billing model, we were able to pull around 130,000 transactions in a 24-hour period over the course of several days.

4 MAIN TECHNIQUES APPLIED

4.1 Data Cleaning & Preprocessing

For data preprocessing, we needed to clean the data of empty values. We found that, due to the timeframe of our dataset, many of the newer attributes (`max_fee_per_blob_gas`, `blob_versioned_hashes`, `max_fee_per_gas`) were empty because those features were added to Ethereum transactions well after the transactions in our dataset. Furthermore, the default value for `transaction_type`, `0x0`, was present for all transactions. The label `0x0` for transactions implied they were legacy transactions, which did not add helpful information for distinguishing any single transaction from others. Therefore, we removed it.

Due to the high verification standards of the data on the Ethereum blockchain, we found that no other columns contained data requiring further processing or imputation. Aside from a few small quirks, such as several transactions with a `to_address` of `0x` followed by 40 hexadecimal zeroes, colloquially known as the *genesis block*, the dataset presented a coherent story of transactions that we could attempt to analyze [1].

4.2 Exploratory Data Analysis & Feature Engineering

Once the data was cleaned, we explored the statistical metrics of several of the features we considered to be highly likely to indicate fraud. We generated graphs of these metrics to identify any outliers and visualize the skew of the values. Additionally, we generated and analyzed the correlation matrix of the features to identify interesting relationships between the features.

The main goal of our feature engineering included gaining and understanding about the behaviors of specific wallets. We included features that allowed us to characterize the wallets as senders or receivers as well as gain insights into typical transaction and gas values for the wallets.

4.3 Anomaly Detection

Our anomaly detection was highly reliant on comparisons between individual transactions and the expected behaviors identified by our statistical analysis. We compared several of our engineered features for individual wallets with the mean transaction values, gas used, and gas prices. We plotted these values with lines showing

the value of the mean and the value of the upper bound for these values. We calculated our upper bound by adding the value of the third quartile to 1.5 times the interquartile range. Unfortunately, due to the presence of extreme outliers, the upper bound often fell below the value of the mean, however, these visualizations provided useful insights into a relative count of how many wallets fell outside of our generated expected values.

4.4 Prediction Modeling

We wished to expand on our anomaly detection by creating prediction models to generate an idea of whether a transaction is potentially fraudulent. We utilized K-Means to generate clusters based on our prior engineered features that give us insights into wallet behaviors. To visualize the separation of the clusters created by our trained K-Means models, we utilized PCA to reduce the dimensions to 2 and generate a scatter plot of the datapoints in the clusters.

Due to the nature of our problem in identifying transactions as fraudulent or not, we wanted to first consider a K-Means model with only 2 clusters with the hopes that it would accurately delineate between 'likely fraudulent' and 'likely not fraudulent'. To get an idea of the relative accuracy and quality of the clusters for this model, we utilized the silhouette score. The silhouette score for our model with two clusters was very high, indicating that the clusters were well defined and separated from each other. We analyzed the values of the centroids for each feature to gain insights into the meanings of each cluster.

To optimize our model and generate the highest silhouette score, we followed the elbow method and generated a graph of the inertia of various K-Means models with different numbers of clusters. The inertia of the K-Means model is a measure of the distance between each data point and the cluster centroid, indicating how well the dataset is clustered. Following this elbow method, we identified that the lowest inertia, therefore best clustering, occurred in K-Means models with 5 to 8 clusters. We were curious about the separation of these model's clusters, so we generated visualizations for models with 5, 6 and 8 clusters.

5 KEY RESULTS

This process allowed us to discover general trends in Ethereum transactions for each of the time points we considered. To understand characteristics and behaviors of the data points, we utilized K-Means clustering. When analyzing the best hyperparameters for optimizing our clustering model, we performed the elbow method to reveal the optimal number of clusters for our dataset. Initially, we thought that two clusters would be sufficient such that we could delineate between likely fraudulent and likely not fraudulent. However, the elbow method indicated that the optimal number of clusters was between 5 to 8. To visualize the separation of these clusters, we performed dimensionality reduction with PCA and then plotted the generated labels. To analyze the relative quality of the clusters in each of our models with different numbers of clusters, we utilized the silhouette score. This score is used to determine the quality of the clusters in terms of their separation and how well each data point is clustered. All our models had very high silhouette scores, indicating that the models successfully characterized the activity of the data points in each cluster. By considering the centroid values for each feature in each cluster, we could generate a criterion of defining behaviors for each cluster. For example, our model with two clusters appeared to separate the data into wallets that executed a huge number of sending transactions for small values, whereas the other cluster had comparable numbers of sent and received transactions for higher values.

The results of our clustering identified one key wallet that appeared as an outlier across several different time ranges. This wallet in particular had an extremely high number of transactions that it received. This wallet received the most transactions of any of the wallets with an average value received that was greater than the third quartile, indicating an unusual number of transactions for the time period. Additionally, despite sending fewer transactions, the values that this wallet sent far exceeded the third quartile. The average gas used for this wallet was also higher than the population mean. Intuitively, these findings allowed us to characterize the behavior of this wallet. The high gas could indicate that the wallet wanted to prioritize these transactions in the block chain, which could point to potentially fraudulent activity due to increased pressures for transaction success. The huge number of transactions that this wallet received could indicate that the wallet processes transactions on behalf of a business. Additionally, the transactions this wallet received had higher than average values, which could support the idea that this could possibly belong to a business or a successful scam wallet. Analyzing this wallet in terms of distinct counterparties indicated that this wallet executed transactions with the highest number of other wallets. These findings highlighted interesting behaviors that we considered to be likely fraudulent. We could then consider these characteristics when analyzing feature values of other wallets.

While we were unable to determine with any certainty whether any transactions were fraudulent, we did gain valuable insights into expected behaviors for wallets. We also considered the time frame for each dataset that we used and referenced historical data in order to conceptualize the values of transactions in US dollars. When we were generating statistical metrics for the dataset, we discovered that the max average value that a single wallet sent was 935800.0ETH. The price of ETH during between 8/17/2015-9/13/2015 was between 0.87 and 1.35USD, therefore, 935800.0ETH during this time was 814,146 - 1,263,330 USD (source : (https://coinmarketcap.com/c
data/)). This wallet could have only one sent transaction that was the cause of this average, however, this is an outlier. Additionally, 75% of the data had average sent values of 126.652167ETH or less. This value far exceeds the interquartile range and upper boundary, and likely would be considered as a red flag for suspicious behavior.

Other interesting discoveries that we made during our exploration included the identification of several wallets with interesting wallet addresses that could potentially represent "placeholder" or testing wallets. The first of these wallets only received transactions, with the average value of each transaction it received being about 2 ETH or somewhere between about 1.74 to 2.70 USD. The other two wallet addresses processed fewer transactions with even smaller amounts, supporting the idea that these wallets could represent testing wallets. During our exploratory data analysis, we uncovered that the outliers we detected skewed the data so far that the

majority of the data for many of our engineered features fell way below the mean.

6 APPLICATIONS

The knowledge that we gained during our data exploration and modeling provided us with insights with which we could apply to other datasets and time periods to hypothesize whether any transactions are likely fraudulent. With additional information regarding known fraudulent activity, we could extend our modeling to include supervised models that could classify a transaction as fraudulent or not, instead of describing the transaction in terms of activity and behavior.

7 VISUALIZATIONS

REFERENCES

[1] Dev Ranjan via Ethereum Stackexchange. 2018. Who has access to ethereum 0x0000000000000000000000000000000000000000 address? https://ethereum.stackexchange.com/questions/52908/who-has-access-to-ethereum-0x0000000000000000000000000000000000000000-address. Accessed: 2024-12-04.
[2] Ethereum Improvement Proposals. 2019. EIP-1559: Fee market change for ETH 1.0 chain. https://eips.ethereum.org/EIPS/eip-1559. Accessed: 2024-12-04.
[3] Ethereum Improvement Proposals. 2020. EIP-2718: Typed Transaction Envelope. https://eips.ethereum.org/EIPS/eip-2718. Accessed: 2024-12-04.
[4] Ethereum Improvement Proposals. 2022. EIP-4844: Shard Blob Transactions. https://eips.ethereum.org/EIPS/eip-4844. Accessed: 2024-12-04.
[5] Federal Bureau of Investigation. 2023. FBI Cryptocurrency Report. https://www.fbi.gov/services/laboratory/cryptocurrency-report. Available at: https://www.fbi.gov/services/laboratory/cryptocurrency-report.
[6] Homeland Defense & Security Information Analysis Center. 2023. Real-Time Cryptocurrencies Monitoring for Criminal Activity Detection: A Comprehensive System. https://www.hdiac.org/cryptocurrency-fraud-detection. Available at: https://www.hdiac.org/cryptocurrency-fraud-detection.
[7] Infura, a Consensys Formation. 2024. Infura Homepage. https://www.infura.io/.