Detecting Malicious Behaviors on Ethereum

CU CSPB 4502 Data Mining Fall 2024
Group 7
Sept 16, 2024

Team 7 Members

The "DataBuffs"

Evan Bean	4th semester in program. Expected graduation May '26.
Hallee Ray	Final semester in program. Expected graduation December '24
Evan Embry	4th semester in program. Expected graduation May '26.
BD Tinsley	4th semester in program. Expected graduation May '25

Project Description

Overview

We will mine data from Ethereum blockchain to discover trends in transactions. Through it, we aim to detect fraudulent or scam-like trends involving investment fraud, kiosk scams and other malicious behaviors.

Interesting Questions

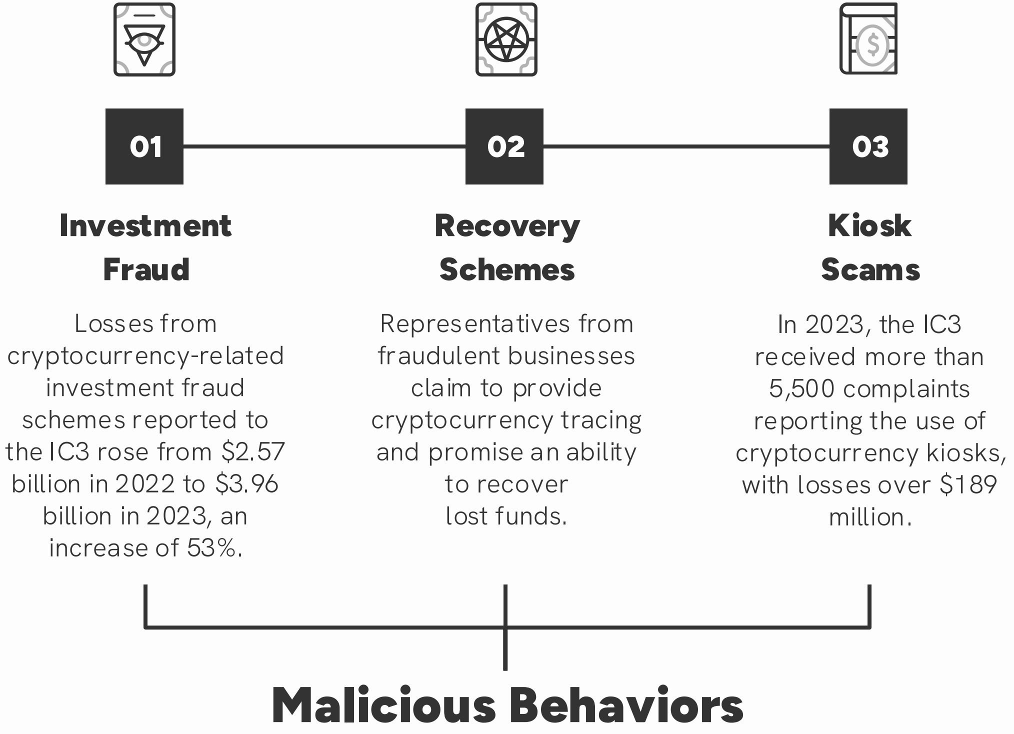
- Is there a correlation between the time of the transaction and the amount?
- What is the average time between a new coin going live and the first purchase?
- What are the average values of the transactions during the lifetime of Ethereum, how does it map to early-stage hype?
- For coins with historical fraudulent activity, what are the general statistics for the transactions that were reported as fraudulent?

Prior Work on This Topic

The FBI reports that....

"In February 2022, the FBI formed the Virtual Assets Unit (VAU), a specialized team dedicated to investigating cryptocurrency-related crimes" (FBI, *2023 Cryptocurrency Fraud Report*)

In it, they have identified common types of malicious behaviors using cryptocurrencies....



01

Investment Fraud

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%.

02

Recovery Schemes

Representatives from fraudulent businesses claim to provide cryptocurrency tracing and promise an ability to recover lost funds.

03

Kiosk Scams

In 2023, the IC3 received more than 5,500 complaints reporting the use of cryptocurrency kiosks, with losses over $189 million.

Malicious Behaviors

Proposed Work

01

EDA

Track transaction statistics, cross validate entries to ensure accuracy

02

Feature Engineering

Generate features based on FBI cryptocurrency fraud report

03

Preprocess Data

Impute missing values, remove duplicates, normalize data

04

Anomaly Detection

Detect outliers using clustering, Z-score analysis, linear regression

Datasets

Ethereum Ledger (link, link)

Real-time transaction data showing transaction type, the sender and recipient wallets and amount via API integration.

Google Trends (link)

Used to map public interest in specific Ethereum-built cryptocurrencies

2023 FBI Fraud Report (link)

Compare our results to FBI findings to determine if we can detect trends before complaints

List of Tools

Python

The backbone language for our project

SciPy/NumPy

To extend Python's mathematical capabilities

pandas

For data analysis

Matplotlib

To visualize our results

GitHub

To house our code

Snowflake

Or other cloud-based data warehousing service to process & share findings

Analysis & Evaluation

We will begin our work by taking transaction history from historical snapshots but will develop a model that can

- Correctly predict malicious behaviors within a historical window
- Predict future malicious behaviors
- Identify behaviors that typically lead to malicious behaviors to mark specific wallets as potentially threatening

CU CSPB 4502 Data Mining Group 7

Detecting Malicious Behaviors on Ethereum

September 16, 2024