Analiza Malicioznih Bitcoin Transakcija

Stefan Nožinić

3. jul 2024.

1 Uvod

Ovaj izveštaj opisuje metodologiju i rezultate analize malicioznih Bitcoin transakcija. Korišćeni podaci sadrže atribute transakcija kao što su vrednost, vreme, visina bloka, veličina, naknada i drugi. Cilj analize je identifikacija karakteristika koje razlikuju maliciozne transakcije od legitimnih.

2 Metodologija

Analiza je sprovedena pomoću Python skripti koje su korišćene za scraping, treniranje modela, i analizu podataka. Skripte su podeljene u sledeće fajlove:

- `scrape.py`: Sadrži kod za prikupljanje podataka o Bitcoin transakcijama.
- `train.py`: Koristi se za treniranje modela za prepoznavanje malicioznih transakcija.
- `analyze.py`: Analizira rezultate i izračunava korelacije između atributa i verovatnoće da je transakcija maliciozna.
- `model.py`: Definiše strukturu modela korišćenog u analizi.

3 Podaci

Podaci korišćeni u analizi sadrže sledeće atribute:

- `value`: Vrednost transakcije
- `time`: Vreme transakcije
- `block_height`: Visina bloka
- `txid`: ID transakcije
- `is_scam`: Oznaka da li je transakcija maliciozna

U datasetu je prisutno ukupno 225364 transakcija, od kojih je 112099 označeno kao maliciozne.

Podaci su preuzeti sa `bitcoin_hacks_2010to2013` baze i servisa `blockchair.info` i sačuvani u formatu `.csv`.

4 Korelacija Atributa sa `is_scam`

Korelacije između različitih atributa i `is_scam` su prikazane u Tabeli 1.

Atribut	**Korelacija sa `is_scam`**
value	-0.081
time	-0.189
block_height	-0.175
sequence	0.004
ver	-0.083
vin_sz	-0.068
vout_sz	0.142
size	0.120
weight	0.120
fee	0.127
tx_index	0.100
block_index	-0.175

Tabela 1: Korelacija atributa sa `is_scam`

5 Rezultati Analize

Sledeći su ključni rezultati analize:

- Prosečna vrednost transakcija:
  - Maliciozne: 1.714e9
  - Legitimne: 2.232e10
- Prosečna naknada transakcija:
  - Maliciozne: 2610942.90
  - Legitimne: 669121.40
- Adresa sa najvećom vrednošću malicioznih transakcija:
  - `output_addr`: 1B8n7yaXZdRShCP75...
  - Vrednost: 2490935946000
  - Vreme: 134

6 Treniranje modela

Model je treniran koristeći Random Forest algoritam sa sledećim parametrima:

- bootstrap: True
- maxDepth: 5
- numTrees: 100
- Ostali parametri su podešeni prema podrazumevanim vrednostima.

Najbolji model je postigao evaluacione rezultate:

Mera	**Vrednost**
Preciznost	93.40%
AUC	94.02%

Tabela 2: Evaluacioni rezultati modela

7 Zaključak

Analiza pokazuje da postoje određene karakteristike koje mogu pomoći u identifikaciji malicioznih Bitcoin transakcija. Prosečna vrednost i naknada transakcija se značajno razlikuju između malicioznih i legitimnih transakcija. Trenirani model pokazuje visoku preciznost u klasifikaciji transakcija.

Dalje unapređenje modela može uključivati analizu dodatnih atributa i optimizaciju hiperparametara.