Detecting Fraud Smart Contracts

Russel Shawn Dsouza

August 25, 2019

Abstract

1 Problem Statement

This submission aims to detect Ponzi schemes in Ethereum smart contracts by using semi-supervised training.

Smart contracts are an important part of the Ethereum blockchain and the most use aspect of Ethereum along with tokens. Identifying ponzi schemes is difficult as the blockchain doesn't store sourcecode but only compiled bytecode. Even if the source code is available, the average user is not able to understand solidity, the language of smart contracts.

2 Proposed Solution

The proposed solution involved 3 stages.
In the first stage raw bytecode of the compiled contracts passed through a stacked autoencoder to reduce the feature size from 128*128 to 32*32 and learn clusters through unsupervised learning.

The output of the autoencoder is passes to a Convolutional Neural network with 3 encoding layers and 2 decoding layers and a fully connected layer at the end. The CNN classifies the bytecode into 16 different categories, which are: crowdsale, destroy, remove, lock, mint, own, pause, trade, transfer, upgrade, and withdraw. These categories are obtained from OpenZeppelin, a reputed source for standard Ethereum Smart Contracts. The training of the CNN is supervised i.e. the model recieves bytecode and a one-hot representation of the 16 categories. The one-hot representation is obtained by tokenizing, lemmatizing, stemming and cleaning the source code mined from etherscan.io.

The third stage stage is a simple classifier to classify ponzi schemes. The classifier is trained by freezing all previous layers and by using an open source dataset on Ponzi Schemes containing about 184 such schemes.