Exploiting Blockchain Data to Detect Smart Ponzi Schemes on Ethereum by Samuel Idowu

Ethereum is defined as a famous open source blockchain-based distributed platform, and a smart contract is a set of promises and protocols specified in digital form. Studies found that 7 million USD has been gathered during 9/2/2013 to 9/9/2014 by scams in Bitcoin. The article explored Rubixi smart Ponzi scheme, that used smart contracts. The significance of detecting Ponzi schemes includes, (I)Investors and users lacking the knowledge of what Blockchain is, therefore trapped by scam (II) It is Imperative for health and rightfulness of Blockchain market. It is difficult detecting smart Ponzi schemes, a lot of verified samples must be collected, it is inefficient to manually check smart contracts with large influx of Ethereum smart contracts each day and more than 2 million smart contracts running out of which 1% has source code. The article presented the framework adopted for the study, as can be seen in Figure 2. The article explained that the smart contract is run on Ethereum Virtual Machine(EVM), that converts source code to bytecode. Smart contracts are written in high level language such as Solidity. The article furthered explained the rubixi scheme, as it presented a visual representation of what the contract looks like. It mentioned that It is possible to identify a smart Ponzi scheme from its transactions history if source code is not available. The article collected its data from all the open source code smart contract created before 9/7/2017 from the **etherscan.io** and manually check whether they are Ponzi scheme contracts. **200 Ponzi scheme** contracts and **3580 non-Ponzi scheme** contracts are identified. It explored account features, analysis of the rubixi transaction led to contract pays to known account, many participants receive nothing from contract and some participant having more payment transactions than investment transactions. The article also presented the statistics of extracted account features as can be seen in TABLE 1. It also explored code extraction, 64 different opcodes found in 3780 contracts, thus code feature is 64 dimension. Figure 4, is a visual representation of the front fragment of bytecode(left) and operation code(right) of rubixi code. The classification model employed by the article is the adopted bagging based algorithm (Random forest). A Random Forest (RF) is a combination of Decision Trees, trained by the training sets obtained by the bagging method. That is N-training sets **->** N decision trees **-> Result from voting.**

The article also gave the formula used for its evaluation metrics, exploring precisions, recall and F-Score. In addition, it explored ten important code features to detect smart Ponzi, and they can be found in Figure 6, as it has been visually represented in the article. Furthermore, the article also represented the number of detected smart Ponzi schemes detected, with corresponding probability range, and the number of detected smart Ponzi schemes created each month in Figure 7 and 8 respectively. In conclusion, machine learning approach has been proposed to detect smart Ponzi schemes. It has been estimated that over 500 smart Ponzi schemes exist. There will be more efficient classification model in the future.