Opcode 디스어셈블러와 기계 학습 기반의 트랜잭션 패턴 분석을 이용한 이더리움 스캠 컨트랙트 및 악성 유저 탐지 시스템 설계

이채현 O*, 고경찬*, 우종수**, 홍원기*

*포항공과대학교 컴퓨터공학과

**포항공과대학교 정보통신연구소

{chlee0211, kkc90, woojs, jwkhong}@postech.ac.kr

A Design of Ethereum Scam Contract and Malicious User Detection System using Opcode Disassembler and Machine Learning based Transaction Pattern Analysis

Chaehyeon Lee O*, Kyunchan Ko*, Jongsoo Woo**, James Won-Ki Hong*

*Department of Computer Science and Engineering, POSTECH

**Graduation School of Information Technology, POSTECH

요 약

블록체인의 데이터 불변성, 데이터 투명성은 사람들로 하여금 블록체인상의 데이터를 과도하게 신뢰하도록 하며, 특히 스마트 컨트랙트 코드의 타당성을 일반 사용자가 판단하는 것이 매우 어렵다. 이는 사용자들을 여러 스캠에 취약한 상태로 만들며 실제로 막대한 피해를 초래하는 이더리움 기반의 스캠이 많이 발생하고 있다. 따라서 본 연구에서는 이더리움 네트워크상의 스캠 컨트랙트와 이를 배포하는 악성 유저 (어카운트)를 식별할 수 있는 탐지 시스템을 제안한다. 스마트 컨트랙트의 바이트 코드를 분석하고, 스캠 컨트랙트와 연관된 트랜잭션 패턴 분석을 통해 스캠 컨트랙트 및 악성 유저를 탐지할 수 있는 방법을 소개하고 향후 연구 방안에 대해 기술한다.

I. 서 론

이더리움 [1]은 비트코인 이후 등장한 2 세대 블록체인이다. 비트코인이 화폐의 기능, 즉, 전자 지불 시스템으로써의 역할에 집중한다면, 이더리움은 스마트 컨트랙트 개념을 지원함으로써 거래나 결제 이외에도 다양한 목적으로 애플리케이션을 개발하고, 운영할 수 있게 플랫폼으로써의 기능을 제공한다.

스마트 컨트랙트는 Nick Szabo 에 의해 제안된 것으로, 기존의 서면 계약을 코드로 구현하고 특정 조건이 만족되었을 시 계약이 이행되도록 프로그램화하여 자동화하는 데 목표가 있다. 이더리움은 이러한 스마트 컨트랙트 개념을 블록체인에 도입하여, 신뢰하지 않는 당사자 사이에서도 자동화된 계약을 실행할 수 있게 한다. 따라서 특정 기능을 포함한 디지털 계약을 작성하고자 하는 사용자는 누구든 스마트 컨트랙트를 작성하고, 컴파일한 후 이더리움 네트워크에 배포할 수 있다. 블록체인을 기반으로 함에 따라 누구든 스마트 컨트랙트 실행과 관련한 데이터를 확인할 수 있고, 삭제할 수 없다.

하지만 이러한 데이터 불변성, 데이터 투명성을 제공하는 블록체인 기술의 장점은 사용자로 하여금 블록체인 네트워크의 사용자들은 모두 올바른 방향으로 행동한다고 생각하게 만들어 스마트 컨트랙트를 포함한 블록체인상의 데이터를 과도하게 신뢰하도록 현혹한다. 또한 배포된 스마트 컨트랙트의 코드는 퍼블릭 하게 공개되어 있어 누구나 확인할 수 있지만, 일반 사용자들이 스마트 컨트랙트 코드의 타당성을 판단하는 것이 매우 힘들다. 따라서 스마트 컨트랙트의 안정성을 간과하게 함에 따라 여러 스캠(사기)에 취약한 상태가 된다.

실제로 불특정 다수를 속이기 위한 신용 사기가 여러 암호화폐를 이용해 빈번하게 발생하고 있고, 폰지 스캠, 피싱 스캠 등 다양한 형태의 스캠 컨트랙트가 무분별하게 배포되어 막대한 피해를 초래하고 있다. 2019 년에는 암호화폐 스캠으로 인한 피해 금액이 40 억 달러에 이르렀으며 암호화폐 스캠으로 인해 이더리움 21,000 개가 도난당한 사례가 있다. 따라서 블록체인 사용자들을 불법 스캠으로부터 보호하기 위한 시스템이 필요하다고 판단된다.

따라서 본 연구에서는 사용자를 스캠으로부터 보호하기 위해 이더리움 네트워크상에 배포된 주요한 스캠 컨트랙트를 식별하고, 스캠 컨트랙트를 배포하는 악의적인 유저 (악성 어카운트)를 탐지할 수 있는 시스템을 설계하였다. 배포된 스마트 컨트랙트의 바이트코드를 Opcode 로 변환해 주는 Opcode 디스어셈블러와 기계 학습을 이용해 스캠 컨트랙트와 연관된 트랜잭션 패턴을 분석하여 특정 스마트 컨트랙트의 스캠 여부와 스캠 컨트랙트를 배포하는 악성 유저를 식별할 수 있는 시스템을 제안한다.

II. Background

스캠 컨트랙트는 스캠의 종류에 따라 정상 컨트랙트와 구별되는 트랜잭션 패턴을 보일 수 있다. 특히, 사용자는 스캠의 종류에 따라 스마트 컨트랙트에 일정 코인을 전송하거나, 스마트 컨트랙트로부터 일정 코인을 전달받는다. 폰지 스캠의 경우, 일정 코인을 보내면 해당 코인보다 더 많은 코인을 되돌려 받는 패턴을 보이게 되는데 그림 1 은 폰지 스캠의 유형 중 하나의 예시를 보여준다. 투자자가 스마트 컨트랙트에게 일정 금액을 보내면, 스마트 컨트랙트는 해당 투자자의 어카운트와 투자 금액을 기록한다. 이후 새로운 투자자가 스마트 컨트랙트에 기록되어 있는 것보다 10% 많은 이더를 보낼 시, 스마트

컨트랙트는 자신이 기억하고 있는 이전 투자자에게 해당 금액을 전송하고 새로운 투자자의 정보로 기록을 업데이트한다. 이 과정이 반복되어 새로운 투자자가 계속해서 스마트 컨트랙트에 이더를 보낸다면 이전 투자자는 자신이 투자한 금액보다 10% 이상의 이익을 얻게 된다. 스마트 컨트랙트는 아무 이더도 지니지 않고 새로운 투자자로부터 이전 투자자에게 이익을 전달하는 역할만을 수행

```
pragma solidity >=0.4.0 < 0.7.0;

contract Ponzi {
   address payable public currentInvestor;
   uint public currentInvestment = 0;

   function () external payable {
      uint minimumInvestment = currentInvestment * 11/10;
      require(msg.value > minimumInvestment);

      address payable previousInvestor = currentInvestor;
      currentInvestor = msg.sender;
      currentInvestment = msg.value;

      previousInvestor.send(msg.value);
   }
}
```

하며, 스마트 컨트랙트의 구현에 따라 다양한 형태의 폰지 스캠을 작성할 수 있다.

그림 1 폰지 스캠 예시

피싱 스캠은 대가를 지불하기 위한 목적으로 코인을 스마트 컨트랙트에 전송하기 때문에 피싱 스캠 컨트랙트는 특정 패턴의 이더를 꾸준히 수신하는 경향이 있다.

III. 이더리움 스캠 컨트랙트 및 악성 유저 탐지 시스템

본 연구에서는 스마트 컨트랙트의 바이트 코드와 트랜잭션 패턴 분석을 통해 이더리움 네트워크상의 스캠 관련 컨트랙트를 식별하고, 악성 유저(악성 이더리움 어카운트)를 탐지하는 시스템을 디자인하였다.

1. 전체 아키텍처

그림 2 는 이더리움 스캠 컨트랙트 및 악성 유저 탐지 시스템의 전체 아키텍처를 보여준다. 탐지 시스템은 스캠/정상 스마트 컨트랙트 수집 모듈, 스마트 컨트랙트 Opcode 분석기, 트랜잭션 수집 및 특징 추출 모듈, 기계학습 기반의 악성 유저 판별 모듈, 데이터 베이스로 구성된다. 각 모듈은 별도로 개발되어 독립적으로 운영된다.

일부 공개 사이트에서는 스캠이라고 판별된 스마트 컨트랙트의 바이트 코드 또는 스캠 컨트랙트 어카운트를 공개하고 있다. 스캠/ 정상 스마트 컨트랙트 수집 모듈에서는 공개된 사이트로부터 스캠 컨트랙트와 정상 컨트랙트를 수집한다. 수집된 두 분류의 컨트랙트는 분류가 결정되지 않은 임의의 스마트 컨트랙트의 분류를 결정하기 위해 사용될 수 있다. 수집된 스마트 컨트랙트는 Ⅲ-2 에서 소개할 스마트 컨트랙트 Opcode 분석기의 입력으로 전달된다. Opcode 분석기를 이용하면 사전에 공개되지 않은 스캠 컨트랙트를 식별할 수 있어 데이터 셋을 확장하는데 활용할 수 있다. 스마트 컨트랙트 Opcode 분석기는 Ⅲ-2 에서 자세한 명세를 제공한다.

Ⅱ에서 설명한 바와 같이 스캠 컨트랙트는 스캠의 종류에 따라 특정 트랜잭션 패턴을 보일 수 있다. 스캠 컨트랙트는 스캠 서비스를 이용하는 여러 사용자들과 상호 작용하게 되고, 여러 트랜잭션을 발생시킨다. 동일한 스캠은 일정한 트랜잭션 패턴을 보일 것이라 예상됨에 따라 특정 컨트랙트와 연관된 트랜잭션의 패턴을 분석하여 해당 컨트랙트가 스캠 컨트랙트인지 아닌지, 그리고 임의의 두 컨트랙트가 동일한 유형의 컨트랙트인지를 판별하는데 활용할 수 있다. 따라서 트랜잭션 수집 및 특징 추출 모듈에서는 스마트 컨트랙트와 연관된 트랜잭션을 이더리움 풀 노드로부터 수집하고, 트랜잭션 패턴을 정의하기 위해 트랜잭션과 연관된 특징들을 추출한다. 트랜잭션 패턴 분석에 활용 가능한 특징들로는 스마트 컨트랙트로 코인을 전송한 트랜잭션의 수, 스마트 컨트랙트가 코인을 전달한 트랜잭션의 수, 컨트랙트로 전송된 이더의 양, 컨트랙트로부터 전달받은 이더의 양, 스마트 컨트랙트와 연관된 트랜잭션의 생애 주기, 연관된 두 트랜잭션 사이의 평균 시간 등이 있다.

기계 학습 기반의 악성 유저 판별 모듈에서는 추출된 트랜잭션 특징들을 기계학습을 통해 학습한다. 스캠인지 아닌지를 이진 분류하기 위해 GBM, DRF, XGBoost 등의 기계학습 분류 모델을 활용한다. 최적의 분류 모델을 생성하기 위해 H2O 의 AutoML 기능을 활용하여 자동으로 학습 모델을 생성하고 최상의 성능을 도출하는 모델을 선정한다. 학습된 분류 모델은 서버에 저장되어 특정 컨트랙트 어카운트의 트랜잭션 특징이 입력으로 전달되면 해당 어카운트가 스캠 컨트랙트를 배포한 악성 유저인지, 아닌지를 식별하는데 사용할 수 있다.

2. 스마트 컨트랙트 Opcode 분석기

동일한 유형의 스캠 컨트랙트의 경우, 스마트 컨트랙트의 코드 사이에 유사성이 있을 수 있다. 스마트 컨트랙트는 EVM 이 실행할 수 있는 형태인 바이트 코드로 변환되어 저장되는데 이러한 바이트 코드는 스마트 컨트랙트의 내용을 이해하기에 적합하지 않은 형태로 바이트 코드를 이용해 스마트 컨트랙트 간 유사성을 파악하는 것이 어렵다. 따라서 스마트 컨트랙트 Opcode 분석기를 이용해 분류가 정해지지 않은 스마트 컨트랙트가 스캠 컨트랙트인지 아닌지를 식별하고자 한다.

스마트 컨트랙트 Opcode 분석기는 바이트 코드 디스어셈블러, Opcode 분석 모듈, 스캠 컨트랙트 판별 모듈로 구성된다(그림 3). 먼저, 바이트 코드 디스어셈블러를 이용해 사전에 수집된 정상 스마트 컨트랙트와 스캠 컨트랙트의 바이트 코드를 Opcode 로 변환한다. 바이트 코드의 변환을 위해서 공개되어 있는 디컴파일러 또는 디스어셈블러를 활용한다. 사용 가능한 오픈 툴은 Ⅳ-2 에서 소개한다.

변환된 Opcode 는 스마트 컨트랙트 구현에 따라 다수의 Opcode 명령으로 이루어져있고, 스캠의 종류에 따라 특정한 Opcode 패턴을 가질 것이라고 가정한다. 스캠 컨트랙트의 경우, 정상 컨트랙트와 구별할 수 있는 특정한 Opcode 특성을 가질 수 있기 때문에, Opcode 분석 모듈은 변환된 Opcode 의 명령어를 통계적으로 분석하여 컨트랙트를 구성하는 명령어 집합을 구한다. 또한, 이 과정에서는 컨트랙트 별 특성을 파악하기 위해 기계 학습을 활용할 수 있다. 통계 분석 및 기계 학습 결과를 통해

스캠 컨트랙트 판별 모듈에서는 주어진 스마트 컨트랙트가 스캠의 성향을 띠는지 아닌지를 구별할 수 있다.

IV. 관련 연구

1. 관련 연구

[2] 는 폰지 스캠의 유형과 유형별 보안 문제에 대해 소개하고, 공개되어 있는 컨트랙트를 유형별로 통계 분석하였다. 이더리움상의 폰지 스캠과 관련한 트랜잭션의 수를 비교하고 스캠의 생애 주기에 대해 설명함으로써 폰지 스캠이 얼마나 많은 영향을 주고 있는지 분석했다.

[3] 는 스마트 컨트랙트 바이트 코드의 Opcode 를 기반으로 특징을 추출하여 이더리움 네트워크상의 폰지 스캠 컨트랙트를 탐지할 수 있는 모델을 디자인했다. operand 와 suffix 를 제외하고 74 개의 서로 다른 Opcode 를 이용해 스마트 컨트랙트를 표현할 수 있는 Opcode set 을 정의한다. 그리고 Opcode 시퀀스와 set 의 길이 등을 feature 로 하는 모델링 전략을 제시하였다. [4] 는 데이터 마이닝 기술과 기계 학습을 이용해 폰지 스캠 컨트랙트를 탐지할 수 있는 모델을 제안했다. 폰지 스캠의 사기 행위를 식별하기 위해 해당 연구에서는 Ether flow graph 를 이용해 7 가지의 특징을 추출하여 account feature 로 활용하였다. 그리고 Opcode 를 code feature 로 활용하였다. Account feature, opcode feature, Account + Opcode feature 의 세 카테고리로 기계 학습을 이용해 폰지 스캠 컨트랙트를 분류해본 결과, Account + Opcode 를 이용했을 때 0.94 의 Precision 을 얻을 수 있었다. [5] 에서는 스마트 컨트랙트의 바이트 코드를 Opcode 로 변환하고, Opcode 별 빈출 정도 등을 통계적으로 분석하여 폰지 스캠 컨트랙트와 정상 컨트랙트를 구분해보았다.

[6] 에서는 네트워크 임베딩을 통해 이더리움 네트워크의 피싱 스캠을 탐지하는 방법론을 제안한다. 이더리움 트랜잭션 히스토리로부터 특징을 추출하고 네트워크 임베딩 알고리즘을 이용하여 피싱 노드를 식별하였다.

소개한 관련 연구에서는 온라인상에 공개되어 있는 소스들로부터 스캠 컨트랙트를 수집하고, 자주 등장하는 Opcode 또는 폰지 스캠 컨트랙트와 상호 작용하는 유저의 특징 등을 이용해 폰지 스캠 컨트랙트를 구별하는 방법론을 제안했다. 공개되어 있는 데이터가 매우 한정되어 있어 기계 학습 및 스캠 컨트랙트 식별을 위해 활용 가능한 데이터 셋이 매우 적다고 판단된다. 따라서 본 연구에서는 Opcode 의 특징을 이용해 자체적으로 여러 스캠 컨트랙트를 수집하여 스캠 컨트랙트 식별을 위한 기계 학습의 데이터 셋으로 활용하고자 한다. 또한, 특정 스캠에만 한정되지 않고 여러 종류의 스캠 컨트랙트를 탐지할 수 있는 범용 시스템을 제안한다.

2. 바이트 코드 디스어셈블러

본 논문에서 제안하는 스마트 컨트랙트 Opcode 분석기 내의 바이트 코드 디스어셈블러는 공개되어 있는 오픈 소스 및 툴을 이용할 수 있다.

Etherscan 은 바이트 코드를 Opcode 로 변환하는 온라인 Opcode 디스어셈블러를 제공한다. 웹사이트에 접속하여 변환을 원하는 스마트 컨트랙트의 바이트코드를 입력하면 Opcode 로 변환하여 결과를 출력한다. 그림 4 는 Etherscan 의 디스어셈블러 예시를 보여준다. 바이트 코드를 입력한 결과, 해당 코드를 구성하는 Opcode 의 목록을 반환해주었다.

EVM Bytecode Decompiler 는 node.js 기반의 바이트 코드 디컴파일러 기능을 제공한다. API 를 제공하여 raw 바이트 코드를 가져오거나, 바이트 코드로부터 Opcode, function, event 등을 추출하는 기능을 제공한다.

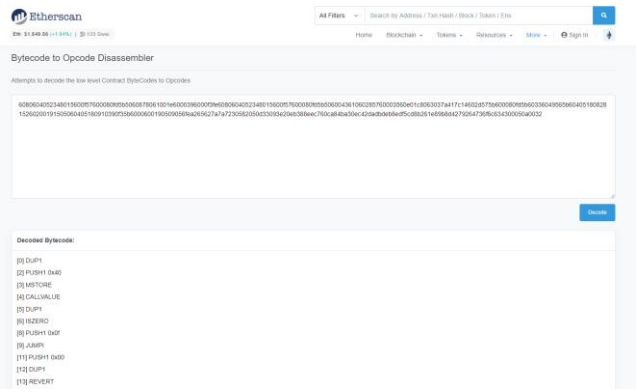

그림 4 Etherscan Bytecode Disassembler 예시 화면

V. 결론 및 향후 연구

본 연구에서는 이더리움 네트워크상의 스캠과 연관된 스마트 컨트랙트를 식별하고, 스캠 활동과 관련 있는 악성 유저를 탐지하기 위한 시스템을 제안한다. 스마트 컨트랙트의 바이트 코드를 Opcode 로 변환한 후 통계 분석하여 스캠 컨트랙트의 특성을 띠는지 확인하고, 트랜잭션 패턴 분석을 통해 스캠 컨트랙트와 연관된 악성 유저 및 악성 행위를 탐지하고자 한다. 본 디자인을 통해 특정 스캠에 한정 짓지 않고 범용 스캠 컨트랙트를 탐지할 수 있을 것이라 기대한다. 향후 연구를 통해 본 연구에서 제안하는 시스템을 구현하고, 시스템을 통해 수집한 실제 데이터 셋을 이용해 탐지 정확도를 측정해 본다.

ACKNOWLEDGMENT

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (IITP-2021-2017-0-01633*, 대학 ICT 연구센터육성지원사업, 2018-0-00749, 인공지능 기반 가상 네트워크 관리기술 개발)

참 고 문 헌

[1] Wood, Gavin. "Ethereum: A secure decentralised generalised transaction ledger." Ethereum project yellow paper 151.2014 (2014): 1-32.

[2] Torres, Christof Ferreira, and Mathis Steichen. "The art of the scam: Demystifying honeypots in ethereum smart contracts." 28th {USENIX} Security Symposium ({USENIX} Security 19). 2019.

[3] Peng, Jianxi, and Guijiao Xiao. "Detection of Smart Ponzi Schemes Using Opcode." International Conference on Blockchain and Trustworthy Systems. Springer, Singapore, 2020.

[4] Chen, Weili, et al. "Detecting ponzi schemes on ethereum: Towards healthier blockchain technology." Proceedings of the 2018 World Wide Web Conference. 2018.

[5] Jung, Eunjin, et al. "Data mining-based ethereum fraud detection." 2019 IEEE International Conference on Blockchain (Blockchain). IEEE, 2019.

[6] Wu, Jiajing, et al. "Who are the phishers? phishing scam detection on ethereum via network embedding." IEEE

Transactions on Systems, Man, and Cybernetics: Systems (2020).

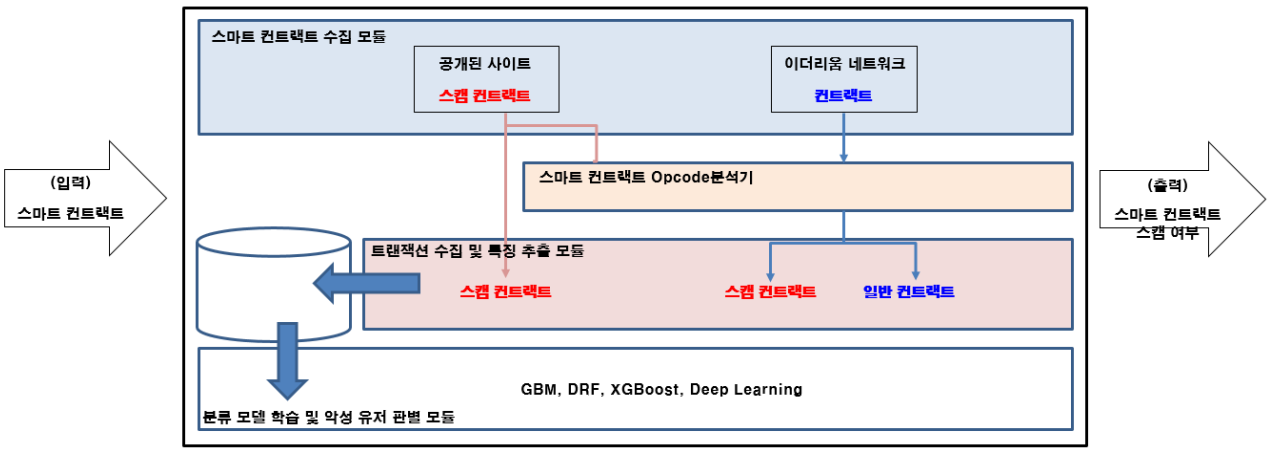

그림 2 이더리움 스캠 컨트랙트 및 악성 유저 (어카운트) 탐지 시스템

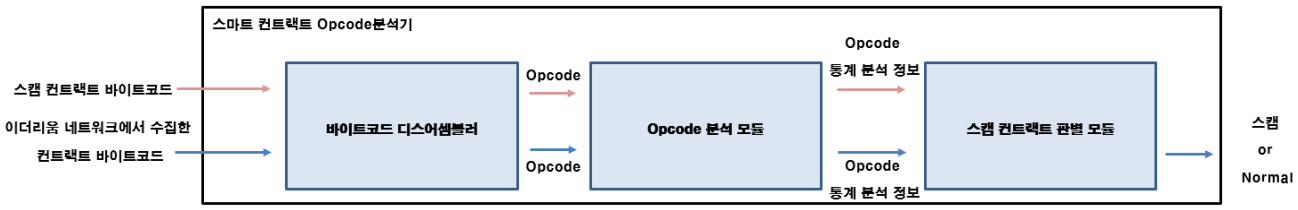

그림 1 스마트 컨트랙트 Opcode 분석기