https://doi.org/10.22670/knom.2019.22.3.20

동일한 환경에서 구성된 비트코인과 이더리움의 메모리 풀 자카드 유사도 분석

맹 수 훈◆, 신 혜 영*, 김 대 용*, 주 홍 택°,

Analysis of Memory Pool Jacquard Similarity between Bitcoin and Ethereum in the Same Environment

SooHoon Maeng◆, Hye-yeong Shin*, Daeyong Kim*, Hongtaek Ju°

요 약

블록체인은 분산 원장 기반 기술로 블록체인 네트워크에 참여하는 모든 노드들은 P2P 네트워크로 연결되어 있다. 블록체인 네트워크에서 트랜잭션이 생성되면, 트랜잭션은 블록체인 노드들에 의해 전파되고 유효성을 검증 받는다. 검증된 트랜잭션은 P2P 네트워크를 통해 각 노드와 연결되어있는 Peer들에게 전송되고, Peer들은 트랜잭션을 메모리 풀에 보관한다. P2P 네트워크 특성상 블록체인 노드가 전달하는 트랜잭션의 수와 종류는 각 노드마다 달라 모든 노드들이 동일한 메모리 풀을 갖지 못하는 문제가 발생함에 따라 메모리 풀에 저장되어있는 트랜잭션들은 거래 수수료 조작, 이중 지불 문제, DDos 공격 탐지 등의 문제를 해결하기 위해 연구가 필요하다.

본 논문에서는 거래 수수료 조작, 이중 지불 문제, DDos공격 탐지 등과 같은 문제를 해결하기 앞서 메모리 풀의 트랜잭션들을 분석한다. 따라서 본 연구는 블록체인기술을 기반으로 구현된 암호화폐 시스템인 비트코인과 이더리움의 각 노드 메모리 풀에 저장되어있는 트랜잭션들을 수집하고 얼마만큼의 공통된 트랜잭션들을 가지고 있는지 자카드 유사도를 이용하여 분석한다.

Key Words : Blockchain, Bitcoin, Ethereum, P2P Network, Memory pool, Jaccard index

ABSTRACT

Blockchain is a distributed ledger-based technology where all nodes participating in the blockchain network are connected to the P2P network. When a transaction is created in the blockchain network, the transaction is propagated and validated by the blockchain nodes. The verified transaction is sent to peers connected to each node through P2P network, and the peers keep the transaction in the memory pool. Due to the nature of P2P networks, the number and type of transactions delivered by a blockchain node is different for each node. As a result, all nodes do not have the same memory pool. Research is needed to solve problems such as attack detection.

In this paper, we analyze transactions in the memory pool before solving problems such as transaction fee manipulation, double payment problem, and DDos attack detection. Therefore, this study collects transactions stored in each node memory pool of Bitcoin and Ethereum, a cryptocurrency system based on blockchain technology, and analyzes how much common transactions they have using jacquard similarity.

※본 연구는 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구 (No.2018-0-00539, 블록체인의 트랜잭션 모니터링 및 분석 기술개발)와 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (NRF-2018R1D1A1B07050380
◆ First Author : Keimyung University Department of computer engineering, aodtngns3438@gmail.com
° Corresponding Author : Keimyung University Department of computer engineering, hongtaek.ju@gmail.com
* Keimyung University Department of computer engineering, {oedaphneashin, imdy1207}gmail.com
논문번호 : KNOM2019-03-04, Received October 30, 2019; Revised November 29, 2019; Accepted December 21, 2019

Ⅰ. 서 론

블록체인(Blockchain) 기술은 사토시 나카모토(Satoshi Nakamoto)가 발표한 'Bitcoin: A Peer-to-Peer Electronic Cash System'[1] 논문에서 처음 제안되었다.

블록체인 네트워크에 참여하는 모든 노드(Node)들은 P2P(Peer-to-Peer) 네트워크를 통해 연결된다. 각 노드들은 블록체인 네트워크에서 발생하는 모든 트랜잭션(Transaction)들을 네트워크를 통해 공유한다. 서로 다른 트랜잭션들은 일정 주기로 하나의 블록(Block)에 포함되며, 생성된 블록은 이전 블록에 연결되어 블록체인 시스템을 구성한다[2].

블록체인 네트워크에서 트랜잭션이 생성된 후, 트랜잭션은 여러 노드에 의해 전파된다. 각 노드는 해당 트랜잭션이 유효한지 검증한다. 검증된 트랜잭션은 P2P네트워크를 통해 각 노드와 연결되어있는 Peer들에게 전송되고, Peer들은 수신된 트랜잭션을 메모리 풀(Memory pool)에 보관한다. 메모리 풀에 저장된 트랜잭션은 다음 블록에 포함될때까지 메모리 풀에 존재한다. 즉, 트랜잭션들이 블록에 포함되기 위해서는 반드시 메모리 풀에 보관되어 있어야 한다[3].

메모리 풀에 저장되어있는 트랜잭션들은 거래 수수료 조작[4], 이중 지불 문제[5], DDos 공격 탐지[4][6] 실시간 트랜잭션 분석[7] 등의 연구를 수행하기 위한 중요한 역할을 한다. P2P네트워크 특성상 전파 지연 및 네트워크 상태와 연결된 Peer들에 따라 각 노드가 전파하는 트랜잭션의 수와 종류가 다르기 때문에 모든 노드들이 동일한 트랜잭션을 가지지 않는다[8].

위와 같은 연구들을 진행하기 앞서, 본 논문에서는 블록체인기술을 바탕으로 구현된 대표적인 암호화폐인 비트코인(Bitcoin)과 이더리움(Ethereum)[9]의 각 노드 메모리 풀을 모니터링하여 트랜잭션들을 수집하고 자카드 유사도를[10] 이용하여 각 노드들의 메모리 풀이 얼마만큼의 유사성을 가지고 있는지 분석한다. 분석된 결과는 비트코인과 이더리움 네트워크 개선 필요성과 개선방안 논의에 대한 단초가 된다.

Ⅱ. 관련 연구

Muhammad Anas Imtiaz[11]는 비트코인 네트워크에서 간혹 발생하는 노드들이 간헐적으로 네트워크 연결에서 이탈하는 현상(이하 churn현상)에 대한 논문을 작성했다. 해당 논문에서 저자는 비트코인 네트워크에서 발생하는 churn현상이 노드 간에 전파되는 데이터양에 영향을 주며, 이에 대한 대표적인 결과는 각 노드의 메모리 풀에 있는 트랜잭션이 달라진다는 것이다. 이는 비트코인 네트워크 성능을 감소시키고 이를 완화하기 위해서 메모리 풀의 효율적인 동기화 방법이 필요하다는 결과를 도출했다.

고경찬 외 2명[12]은 비트코인 노드들은 초당 약 2~17개의 새로운 트랜잭션을 주고 받는데 이때, 메모리 풀에 저장되어 있는 트랜잭션들이 각 노드 마다 차이가 있다는 논문을 작성했다. 해당 논문에서 저자는 비트코인 노드들의 메모리 풀에 보관되어있는 트랜잭션들 중 얼마만큼 동일한 트랜잭션이 있는지 알아보는 실험을 진행했다. 실험을 통해서 저자는 각 노드들이 연결되어있는 Peer들이 다르며 이는 메모리 풀의 트랜잭션에 영향을 미친다는 결과를 도출했다.

Ⅲ. 메모리 풀 유사도 분석

1. 실험 배경

블록체인 네트워크에서 트랜잭션이 생성되면, 각 노드들이 연결되어 있는 Peer들에게 트랜잭션을 전파한다. Peer들은 수신된 트랜잭션을 메모리 풀에 보관한다. 하지만, 각 노드들의 메모리 풀에 저장되어 있는 트랜잭션은 항상 같지 않다. P2P 네트워크 특성상 네트워크 상태 및 연결된 Peer에 따라 전송되는 트랜잭션의 수와 종류가 다르기 때문이다[8].

각 노드의 메모리 풀의 동기화가 잘 이루어 지지 않는다면, 블록 체인에서 여러가지 문제[13][14]가 발생할 수 있다. 예를 들어 메모리 풀이 다른 두 노드가 동일한 시점에 블록을 채굴(mining)하면 체인에서는 블록체인 분기(Fork) 현상이 발생한다. 분기 현상이 발생한 경우 체인으로 선택되는 블록은 더 많은 작업 증명(Pow, Proof-of-work)이 수행되어 길이가 더 긴 블록이다. 이때 선택되지 못한 블록을 비트코인에서는 스테일 블록(Stale block), 이더리움에서는 엉클 블록(Uncle block)이라 한다. 비트코인의 스테일 블록과 달리 이더리움의 엉클 블록은 이더리움 메인 체인의 구성요소로 포함된다. 스테일 블록이 생성되면, 블록체인 기술의 여러 가지 취약점[13][14]들 중에서 51% Attack에 노출되게 되며

이중지불을 포함한 체인이 길어질 수 있는 확률이 존재하기 때문에 이중지불을 완벽하게 방지할 수는 없다. [5]에서는 현재 블록체인 네트워크가 이중지불 위협에 대해 노출되어있음을 보여준다. 따라서, 블록 분기가 발생하게 되면 거래 수수료 조작[4], 이중 지불 문제[5], DDos 공격 탐지[4][6]등의 문제가 발생할 수 있다[15]. 위와 같은 이유로 메모리 풀의 동기화가 잘 되고 있는지 분석할 필요가 있다고 여겨진다.

따라서, 본 논문에서는 메모리 풀이 얼마만큼의 유사성을 띄고 있는지 분석함으로써, 각 노드의 동기화가 잘 진행되고 있는지를 분석한다.

2. 실험 방법

본 논문은 비트코인과 이더리움의 메모리 풀 동기화를 정도를 알아보기 위한 실험으로 각 블록체인 플랫폼 네트워크에 연결된 노드들의 메모리 풀 트랜잭션을 수집하여 진행한다. 실험 환경은 비트코인과 이더리움 4개의 노드를 Bitcoin Core와 Go-ethereum Client를 각각 설치하여 동일한 환경(인텔® 코어™ i5-7500T, 8G RAM, 1TB SSD)에서 진행한다. 비트코인과 이더리움의 메모리 풀 트랜잭션 정보는 Python기반 RPC통신을 이용해 각 노드의 메모리 풀 정보를 수집하는 프로그램을 개발한다. 메모리 풀의 정보를 수집하는 주기는 블록이 포함되는 시간은 비트코인과 이더리움 메모리 풀의 트랜잭션 정보가 갱신되는 시간과 동일한 시간인 메모리 풀의 트랜잭션이 블록에 포함되는 시간을 기준으로 한다. 따라서 비트코인과 이더리움 각 4개 노드의 메모리 풀 정보 수집 시간은 1개의 블록이 생성되는 평균 시간으로 비트코인 10분, 이더리움 20초를 목표로 총 10회 진행 한다[16][17]. 수집된 메모리 풀의 트랜잭션은 자카드 유사도(Jaccard index)[9]를 이용해 합집합에 대한 교집합의 비율로 유사도를 측정한다. 자카드 유사도 식은 (1)과 같다. 각 블록체인 플랫폼에서 측정된 메모리 풀 유사도를 비교하여 비트코인과 이더리움의 메모리 풀 동기화를 확인한다.

$$J(A,B) = \frac{n(A \cap B)}{n(A \cup B)} = \frac{n(A \cap B)}{n(A)+n(B)-n(A \cap B)} \quad (1)$$

3. 실험 과정

그림1은 4개의 Bitcoin Core Client가 설치된 노드에서 메모리 풀에 저장되어있는 트랜잭션의 개수를 노드 별로 나타낸 그래프다. 메모리 풀의 트랜잭션 수집은 각 노드에서 동일한 시점에 메모리 풀 수집 프로그램을 실행시켜 10분 간격으로 데이터를 수집하는 과정을 10번 반복하여 진행한다. 비트코인의 메모리 풀 트랜잭션은 노드 별로 약 4000 ~ 12000개의 정보를 가지고 있고 각 실험 별 노드 트랜잭션의 평균 개수는 약 4700 ~ 10900개를 보여 주고 있다.

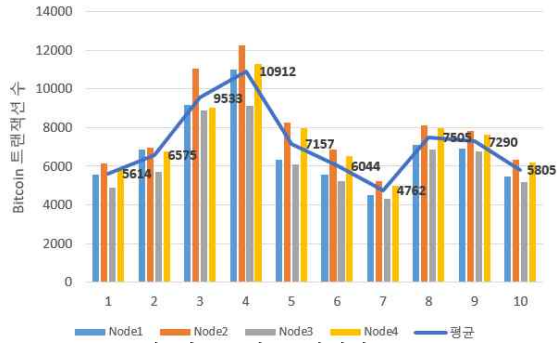

그림 1. 노드 별 비트코인 트랜잭션 수
Fig. 1. Number of bitcoin transaction each node

그림 2는 4개의 Go-Ethereum Client가 설치된 노드에서 메모리 풀에 저장되어있는 트랜잭션의 개수를 노드 별로 나타낸 그래프다.

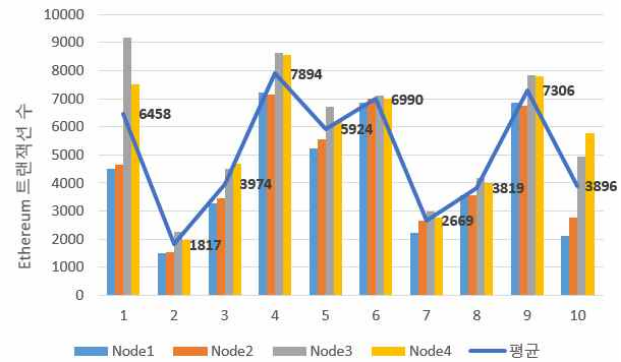

그림 2. 노드 별 이더리움 트랜잭션 수
Fig. 2. Number of ethereum transaction each node

메모리 풀의 트랜잭션 수집은 동일한 시점에 메모리 풀 수집 프로그램을 실행시켜 20초 간격으로 데이터를 수집하는 과정을 10번 반복하여 진행한다. 이더리움의 메모리 풀 트랜잭션은 노드 별로 약 1500 ~ 9000개의 정보를 가지고 있고, 각 실험 별 노드 트랜잭션의 평균 개수는 약 1800 ~ 7800개를 보여 주고 있다. 수집된 메모리 풀의 트랜잭션 정보는 노드 마다 차이가 있는 것을 확인할 수 있다. 따라서 블록체인 플랫폼에서 노드는 연결된 다른 노드들에 따라 수신 받는 정보가 다르다는 것을 보여준다.

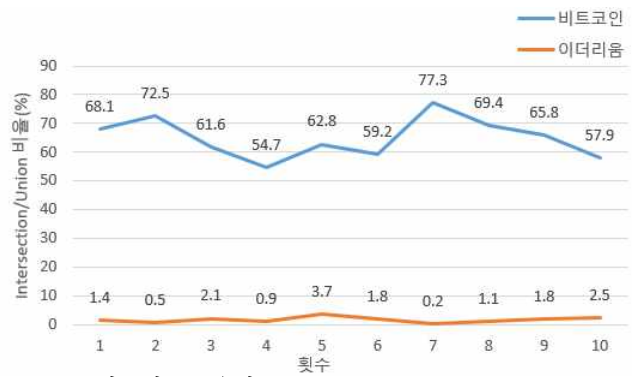

그림 3. 메모리 풀 유사도
Fig. 3. Memory pool similarity

그림 3은 비트코인과 이더리움의 메모리 풀 유사도를 10번의 실험에 따른 결과를 나타낸 그래프다. 또한, 4개의 노드에서 수집된 블록체인 플랫폼 메모리 풀의 트랜잭션을 자카드 지수로 나타낸 결과를 비율로 보여준다. 실험 결과로 비트코인 메모리 풀의 유사도는 약 50% ~ 70%, 이더리움 메모리 풀의 유사도는 5%미만으로 비트코인이 메모리 풀 동기화가 높다는 결과를 실험을 통해 도출했다. 비트코인의 스테일 블록과 다르게 이더리움의 엉클 블록은 이더리움의 메인 블록에 연결되어 보상이 이루어진다. 비트코인의 메모리 풀의 경우 스테일 블록에 대한 보상이 이루어지지 않으므로 비트코인의 메모리 풀의 동기화가 더 높은 수치로 나타났다.
P2P 네트워크의 특성상 비트코인과 이더리움에 연결된 노드의 메모리 풀은 완벽하게 일치하지는 않았지만, 메모리 풀의 동기화 비율은 일정한 범위에 있다는 사실을 확인했다.

Ⅳ. 결 론

비트코인과 이더리움 네트워크에 참여하는 노드들은 P2P 네트워크 기반으로 노드에 연결된 Peer와 네트워크 상태의 차이가 있으므로 메모리 풀의 트랜잭션이 전파되는 시간을 보장하지 않는다.
본 논문에서는 비트코인과 이더리움 노드들의 메모리 풀에 보관되어있는 트랜잭션들이 차이가 있다는 것을 실험을 통해 확인했다. 또한, 비트코인과 이더리움 노드의 메모리 풀에 대한 트랜잭션의 비율을 특정 범위 안에 있다는 사실을 확인했고, 비트코인의 메모리 풀의 유사도가 이더리움 메모리 풀 유사도 보다 높다는 결과를 도출했다.
향후 연구로는 비트코인과 이더리움의 메모리 풀 유사도 분석의 실험 횟수를 증가시켜 유사도를 비교 하는 연구와 각 메모리 풀의 동기화 시간을 비교하는 연구를 진행할 수 있다. 메모리 풀에 보관중인 트랜잭션을 동기화를 통해 블록의 전파 속도를 높일 수 있고 나아가 블록체인 네트워크의 성능을 향상시킬 수 있는 연구를 진행할 수 있다. 또한 블록체인 네트워크 문제로 발생할 수 있는 이중 지불 문제 등 다양한 문제를 해결할 수 있는 방안을 논의할 수 있다. 블록체인 네트워크의 성능이 개선된 후에는 블록체인 네트워크에서 수집된 메모리 풀의 트랜잭션을 통해 불법적인 거래를 탐지 하는 모니터링이 가능할 것이다.

References

[1] [1] Stoshi Nakamoto, “Bitcoin: A peer-to-peer electronic cash system”, 2008
[2] Vallois, Valentin, and Fouad Amine Guenane. ″Bitcoin transaction: From the creation to validation, a protocol overview.″ 2017 1st Cyber Security in Networking Conference (CSNet). IEEE, 2017
[3] Block. https://en.bitcoin.it/wiki/Block, 2016
[4] Muhammad Saad, Jeffrey Spaulding, Laurent Njilla, Charles Kamhoua, Sachin Shetty, DaeHun Nyang, and Aziz Mohaisen, ”Exploring the Attack Surface of Blockchain: A Systematic Overview”, arXiv:1904.03487v1, Apr 2019
[5] 김민호, 김수진, 최훈 ”블록체인의 이중지불 탐지 알고리즘”, 정보과학회논문지 제45권 제8호, 2018
[6] Muhammad Saad, Laurent Njilla, Charles Kamhoua, Joongheon Kim, DaeHun Nyang, Aziz Mohaisen, “Mempool Optimization for Defending Against DDoS Attacks in PoW-based Blockchain Systems”, ICBC 2019
[7] 백의준, 신무곤, 지세현, 박지태, 김명섭, “비트코인 네트워크 트랜잭션 이상 탐지를 위한 특징 선택 방법”, KNOM Review ’18-02’ Vol.20, 2018
[8] Mohamed Rahouti, Kaiqi Xiong and Nasir Ghani, “Bitcoin Concepts, Threats, and Machine-Learning Security Solutions”, IEEE Access 2018
[9] Vitalik Buterin, “A Next-Generation Smart Contract and Decentralized Application Platform”, 2013
[10] “Jaccard index” , Available Online, http://en.wikipedia.org/wiki/Jaccard_index

[11] Imtiaz, Muhammad Anas, et al. "Churn in the Bitcoin Network: Characterization and Impact.", 2018
[12] 고경찬, 이채현, 홍원기, "비트코인 노드 메모리 풀 유사도 분석", KNOM Conference(2019)
[13] Bitcoin wiki, "Weaknesses," Available Online https://en.bitcoin.it/wiki/Weaknesses
[14] M. Resenfeld, "Analysis of hashrate-based doublespending," arXiv preprint arXiv:1402.2009, 2014
[15] Wüst Karl, "Security of Blockchain Technologies", 2016
[16] "Bitcoin Block Time historical chart", Available Online,https://bitinfocharts.com/comparison/bitcoin-confirmationtime.html
[17] "Ethereum Block Time historical chart", Available Online,https://bitinfocharts.com/comparison/ethereum-confirmationtime.html

맹 수 훈 (SooHoon Maeng)

2018년 8월: 계명대학교 컴퓨터공학과 졸업
2019년 9월 - 현재 계명대학교 컴퓨터공학과 석사과정

<관심분야> 네트워크 관리 및 보안, 블록체인 모니터링 및 분석, 이더리움

신 혜 영 (Hye-yeong Shin)

2019년 2월: 계명대학교 컴퓨터공학과 졸업
2019년 3월 - 현재 계명대학교 컴퓨터공학과 석사과정

<관심분야> 네트워크 관리 및 보안, 블록체인 모니터링 및 분석, 비트코인

김 대 용 (Daeyong Kim)

2019년 2월: 계명대학교 컴퓨터공학과 졸업
2019년 3월 - 현재 계명대학교 컴퓨터공학과 석사과정

<관심분야> 네트워크 관리 및 보안, 블록체인 모니터링 및 분석

주 홍 택 (Hongtaek Ju)

1989년 8월: 한국과학기술원 전자계산학과 학사
1991년 8월: 포항공과대학교 컴퓨터공학과 석사
1997년 8월: 대우통신종합연구소 선임연구원
2002년 2월: 포항공과대학교 컴퓨터공학과 박사
2002년 9월 ~ 현재: 계명대학교 컴퓨터공학부 교수

<관심분야> 네트워크 및 시스템 관리, IoT 관리, SDN 네트워크 관리, 블록체인 모니터링 및 분석