Department of Computer Science
University of Hong Kong

Final Year Project

Improvement of Blockchain Consensus Algorithm by Integrating Distributed Machine Learning

Sangwon Park (3035556060)
Chungha Yu (3035553135)
Supervisor: Dr. Chow, Kam Pui

1. Background

Since the introduction of Bitcoin by Satoshi Nakamoto in 2008, blockchain has evolved to be applicable not only in the finance, government, and education industries but also as the pillar of the new web 3.0 protocol. It has been able to expand in such an exponential manner due to its unique trait: blockchain does not require a central authority [1,4]. This trait enables it to solve the double-spending problem prevalent in conventional banks and governments that impose extra fees and taxes [3].

However there is one major downside of the whole system. According to the University of Cambridge Electricity consumption index, it is estimated that blockchain consumes electricity at an annualized rate of 127 terawatt-hours, which contributes 0.3% of global annual carbon emission [7,8]. The reason for high consumption of energy is due to the nature of bitcoin's Proof of Work (PoW) consensus algorithm. To establish a competitive market, PoW forces miners to hold higher computational power than their competitors in order to earn the reward. Consequently, the creation of new blocks consumes more time, energy, and capital [3].

As a result, many novel consensus algorithms have been formed to solve PoW. Among many solutions, Proof of Stake (PoS) introduced by Sunny King in Peercoin gained popularity due to low energy consumption and efficiency in the chain. PoS is based on the amount of ownership in the chain [6]. Stakers with more tokens will be likely to be selected as the next block validator and earn the reward [2].

Such alternative consensus algorithms do solve the problem of PoW; yet, consensus algorithms in the current market do not maintain the major benefit of PoW: formulation of a competitive market. Competitive market is a key factor in blockchain security. In any blockchain, if the attacker holds 51% of the network, the attacker is able to modify information of transactions in the linked blocks. For PoW, an attacker needs to hold 51% of computational power in the network which will require enormous sums of money to outpower all other miners [9]. Whereas for PoS, an attacker needs to control over 51% of total tokens staked in the

network. It will be extremely vulnerable on blockchains where there is a high concentration of coin ownership among a few parties.

The idea of the Decentralized Machine Learning network (DML) is to provide computational power in a decentralized manner. Node providers with higher computational power receive more rewards, similar to the Proof-of-Work (PoW) blockchain. The key difference is that DML utilizes computational power for training datasets, whereas PoW wastes energy by solving mathematical problems solely for the creation of new hashes. DML bridges the efficiency gap that PoW algorithms possess.

With the rise of AI and machine learning, developing insightful information by utilizing large data sets and training models has become all the more significant. As a result, when running machine learning, people focus on three tasks: how long does it take for a process to reach convergence? How good is the solution? How much training data is needed to reach the good solution? According to Google research, the use of GPUs in training data sets is practical but loses efficiency when the model is limited by the memory storage space of the GPU. Therefore, experiments conducted show that a distributed approach to machine learning accelerates the training of modest to large sized models [10].

By having these distributed servers form a network and leverage blockchain to tokenize the transactions of data and parameters, this paper aims to build a decentralized network that connects computing power providers and incentivizes them through tokens to provide machine learning services for customers.

2. Objective

To achieve decentralized machine learning, this project requires 3 main objectives: construction of blockchain with proof of machine learning, construction of a peer-to-peer network for data transfer, and deployment of distributed machine learning nodes.

Objective 1: Blockchain with Proof-of-machine-learning

1.1. To create PoML consensus algorithm

1.2. To create peer-to-peer network

1.3. To create block module including transaction, account, and wallet module

1.4. To provide API interaction module

1.5. To test out block generation time

Objective 2: Distributed Machine Learning Nodes

2.1. To set up AWS servers and ensure communication

2.2. Properly receive data models and training set

2.3. To provide rating after the completion of training data set

2.4. To find methods to measure compute power within the servers

Objective 3: Peer-to-Peer network for large data transfer

3.1. To create peer-to-peer network

3.2. To transfer large dataset at high-speed rate

3.3. To provide fault tolerance network

3. Methodology

To accomplish the objectives, the project requires the completion of 3 main stages. First stage focuses on the development of the blockchain network while the second focuses on setting up distributed nodes. The last stage involves setting up a peer to peer network for high-speed file transfer.

1. Blockchain Network

To formulate a blockchain network that provides decentralized machine learning, the project requires a new form of consensus algorithm. It requires analysis on blockchain projects that use different consensus algorithms and research their algorithm models to figure out the information needed to implement an accurate method of incentivizing node providers. The project plans to try out various protocols to ensure that a transparent and fair decision is made when choosing which node provider can mine a new block.

2. Data Transfer Network

Due to the size limit of blocks in blockchain, where Bitcoin has a limit of 1MB, transferring training data sets will result in slow and congested blockchain. For instance, it will acquire 10 millions blocks if the training data set is 1 terabyte. Hence, the project will implement another layer of peer-to-peer network that is dedicated for transferring training datasets and return the result of transfer as a transaction on the blockchain. There exist such data transfer models such as BitTorrent protocol or InterPlanetary File System (IPFS). The project will research deeper on the architecture of such protocols and implement the network for demo testing purposes.

3. Distributed Machine Learning Node via Cloud Providers

To effectively and timely build a distributed machine learning network over the course of 6 months, instead of procuring on-premise servers or additional equipment, this project will use Amazon Web Services (AWS) to replicate what it would feel like if different kinds of servers and nodes with different types of compute power were to connect and run on the blockchain. This project will imitate various machine learning nodes through cloud providers and make sure their functionality and operations are similar to real life production scenarios. Through in-depth research of the AWS documentation and tests of different services, this project will implement servers of providers that have the relevancy and cost effective services to run machine learning models.

4. Schedule and Milestones

Objective	Deadline	Details and Learning Hours
Phase 1 Deliverables	October 1 2023	Project Plan and Website (15 hours)
Research and Hypothesis Testing	October 29 2023	Determine consensus algorithm (40 hours)
Preliminary Blockchain Network Setup	November 25 2023	Set up blockchain network along with servers (35 hours)
First Presentation	January 8 2024	Preparation (15 hours)
Phase 2 Deliverables	January 21 2024	Preliminary implementation and detailed interim report (25 hours)
Peer to Peer Network Implementation	February 2024	Development (40 hours)
Testing of Machine Learning Models	March 2024	Testing and running models (40 hours)
Final Presentation	April 15 2024	Preparation (15 hours)
Phase 3 Deliverables	April 23 2024	Finalized tested implementation and final report (20 hours)
Project Exhibition	April 26 2024	Preparation (10 hours)

5. References

[1] N. Shi, “A new proof-of-work mechanism for bitcoin”, *Financ Innov*, vol. 2, no. 1, p. 31, Dec. 2016.

[2] Larimer, D, “Transactions as proof-of-stake”, 909, Nov. 2013.

[3] M. Würsten and C. Cachin, “Filecoin Consensus Performance analysis Master Thesis,” 2022.

[4] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," Manubot, vol. 3991, no. 51521347, 2008.

[5] M. Ball, A. Rosen, M. Sabin, N. Prashant, and Vasudevan, “Proofs of Useful Work,” 2021.

[6] W. Zhao, S. Yang, X. Luo, and J. Zhou, "On peercoin proof of stake for blockchain consensus," 2021 The 3rd International Conference on Blockchain Technology, pp. 129-134, March 2021.

[7] J. W. Kirkwood, "From work to proof of work: Meaning and value after blockchain," Critical Inquiry, vol. 48, no. 2, pp. 360-380, 2022.

[8] "Cambridge Bitcoin Electricity Consumption Index (CBECI)," Cambridge Centre for Alternative Finance, Cambridge, UK, 2023.

[9] F. A. Aponte-Novoa, A. L. S. Orozco, R. Villanueva-Polanco, and P. Wightman, "The 51% Attack on Blockchains: A Mining Behavior Study," IEEE Access, vol. 9, pp. 140549-140564, 2021.

[10] Jeffrey Dean, Greg S. Corrado, Rajat Monga, Kai Chen, Matthieu Devin, Quoc V. Le, Mark Z. Mao, Marc'Aurelio Ranzato, Andrew Senior, Paul Tucker, Ke Yang, and Andrew Y. Ng. 2012. Large scale distributed deep networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'12). Curran Associates Inc., Red Hook, NY, USA, 1223–1231.