— Simulation Methods and Applications —

Discrete Event Simulation to model Blockchain network and study Mining Performance using Queuing Theory

ISEN 625 – Section 600,700

<u>Team:</u>

Meet Shukla

Manan Shah

Fall 2021

"On my honor as an Aggie, I have neither given nor received unauthorized aid on this academic work."

Executive Summary

Blockchain is an old technology, however it did not see the limelight until its application in Bitcoin which was the first cryptocurrency developed. The immutable and decentralized aspect of blockchain to store and transfer information is of interest in multiple other industries like healthcare and supply chain. This study attempts to study the nature of blockchain technology from the purview of Bitcoin, which is currently the largest cryptocurrency as per market cap. Here, concept of discrete event simulation is used where the memory pool translates as a M/M/1 server and the mining pool as a M/M/c server. A Poisson distribution is fitted to the historical data and random inter-arrival times for transactions are generated, whereas from the sales data of mining machine based on its processing power, a triangular distribution was fit to the various servers in the mining pool. A ‘m’ miner ‘n’ node scenario of Bitcoin system is developed wherein the input distributions are as mentioned before and output statistics are calculated. The output characteristics analysed were, (a) average number of confirmed transactions in a day, (b) average number of blocks mined in a day (c) average number of transactions in a block and (d) the system throughput i.e. number of confirmed transactions per block per second. These parameters are validated using the actual data with error values less than 1% for all. Similarly, the model was used to check the profitability of miners over a 15 day replication with changing difficulty levels in the mining pool. The stochasticity in Bitcoin network was proven as every time a block is generated, there is a different miner who can win by mining the block. The profitability in bitcoin mining is also studied. It is concluded from this analysis, that owing to the high costs of electric consumption it is not profitable to always participate in mining whenever a block is generated. Instead, it is suggested to develop further upon this basic model and build an agent-based simulation which calculates the probabilities of winning and help a miner decide when to participate and how much computing power should be used for mining. Although the proposed model is built using the system of Bitcoin, the same can be used to study other cryptocurrency as well and blockchain in general.

Contents

1. Introduction

Bitcoin is a Peer-to-Peer (P2P) digital currency which was developed around 2008 by an unknown entity referred to as ‘Satoshi Nakamoto.’ The basis of the Bitcoin infrastructure is rooted in the concept of blockchain, which is a single long unalterable chain containing information on various transactions inside a box and shared with every user in the network. This kind of distributed ledger, containing new and all the previous transactions, is validated by all users within the network and only then does it add onto form a new block and attach to chain. The core principle which is consensus based makes such a ledger almost impossible to be tampered with as no old information can be changed, and every new information must be agreed upon by all the users in the network. These attributes of it being immutable, reliable and a decentralized system, has encouraged the application of blockchain in other domains like supply chain, banking, stock trading, finance and healthcare. However, the problem that persists is the amount of computational power and time needed to form and validate these blocks, due to which its applications are still being studied and algorithms modified to come up with different models.

As of November 21, 2021, there are more than 7000 different types of cryptocurrencies in the market. Motivated by the increasing number of users in the blockchain ecosystem and its applications in multiple domains the report aims at simulating a simplified version of the ‘mining pool – miner behavior’ and studying various characteristics associated with mining of each block. The primary incentive in the Bitcoin mining game is that the winner takes it all, hence the goal of every player (miner) is to be the first in correctly guessing the puzzle and mine the block. However, as the computation power of the network increases, so does the difficulty level in solving the puzzle and the probability of individual miners wining reduces. This leads to forming of mining pools wherein individual miners contribute their mining power towards solving the puzzle and the reward is distributed.

In this study, a simple queuing theory concept has been used to model the memory pool (where transactions are collected) as a M/M/1 server and the mining pool (where miners exist) as the M/M/c server. This proposed model can be used to observe the realistic behavior of both – the memory pool and mining pool. The output parameters in proposed model, i.e. block mining time and the average number of transactions in memory pool were validated by comparing them to real output data from Bitcoin database. Other statistics which were calculated are, memory pool count, average number of unconfirmed transactions in a day, and the number of blocks generated.

1.2 Real-World Block Chain Process

Since Bitcoin is a P2P system, there is absence of a centralized body or any hierarchy which monitor and store transactions, instead every node (here Bitcoin user) in the ecosystem is considered equal and copies of transactions are shared with everyone. Thus, everyone is able consume (by raising transactions) as well as provide (by mining block) service through accessing this information. Every transaction is made using a specific signature and broadcasted to everyone in the network. It is still not validated and waits in a place called the “memory pool” After a particular time interval or size, anyone from the network can pick all those transactions and make it into a block. In order to test its validity, the particular miner has to solve a random puzzle associated with that block by “guessing” a particular series of number (called nonce); once completed solving the puzzle, this valid block with all valid transactions are shared with everyone along with the nonce for others to access this block. All other nodes in with whom the block is shared will verify the transactions. If correct, they confirm and add it to their network and the miner will be rewarded with certain number of Bitcoins. This is what completes the mining process for one block.

Following is the general process flow

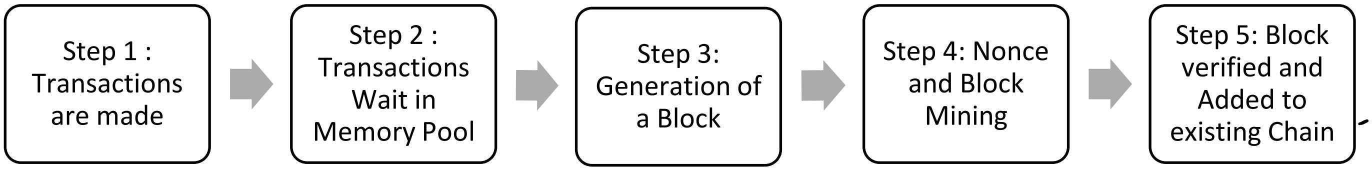

Step 1 : Transactions are made
Step 2 : Transactions Wait in Memory Pool
Step 3: Generation of a Block
Step 4: Nonce and Block Mining
Step 5: Block verified and Added to existing Chain

Figure 1 Blockchain Process from Start to End

Step 1: Every Bitcoin user can initiate a digital transaction to any other user in the same network. As per current data, there are 300000 transactions being made every day

Step 2: All the transactions wait in the memory pool to be picked up by miners and made into a block. Until then, these remain unconfirmed and termed as ‘Unconfirmed Transaction’.

Step 3: At specific time intervals (generally 10 minutes), miners access the memory pool and group transactions up to a particular size (~1 MB) and combine them in a block.

Step 4: To mine a block wherein all the above transactions can be grouped, the miner must solve a puzzle by guessing the value of a hash function (SHA256 used by Bitcoin) also called nonce. Since there are multiple miners, all of them will be trying to guess this nonce at the same time, and only one can be the quickest. The ability to perform these guesses/second depends on the computing power. Once guessed, the block is added to the chain of that miner, they are rewarded with crypto coins (Bitcoins in this case), and its data along with the nonce is broadcasted to all other nodes in the network for verification. The current reward for bitcoin mining is 6.25 bitcoins which is equivalent to approximate $300000.

Step 5: Since the block has been discovered along with its nonce, others will stop solving this puzzle, and instead verify the nonce. If all the previous transactions are true and verifiable, they also add the particular block to their blockchain.

1.3 Research Questions of Study

Following are the main research questions:

- Can the Bitcoin ecosystem be simplified to study discrete event simulation and basic queuing theory?
- What would be the characteristic of memory pool and mining pool in terms of their service rate?

- What will be the effect on queue length based on stochasticity between transaction's arrival rates and the service times based on?
- Under what conditions is mining more profitable, and does an increase in one's computing power guarantee higher rewards in the Bitcoin environment?

2. Methodology

2.1 Proposed Model

There are two main stations for transactions in block-chain; one is the memory pool where these transactions wait to be picked by the miners and the other is the mining pool, where they are converted to a block. To reduce the complexity and study the different output parameters in blockchain system, we develop a two-server model with Memory Pool as a M/M/1 server and Mining Pool as a M/M/c server. Memory pool is an infinite server which shall keep on getting transactions from different users in network. After a specific time or after it collects a particular number of transactions, whichever is earlier, the available transactions are grouped together, and a block is generated. The block is then passed onto the mining pool where it is shared with all the available nodes. Based on the service rates of these nodes in the pool (reflects the computing power in hashrate/second) the block is mined which is defined as its mining time. There can only be one block at a time inside the mining pool, whereas the transactions will keep on coming inside the memory pool. At the end of day, the transactions still in the memory pool are termed as "unconfirmed transactions" and they shall be carried onto the next day

Two different scenarios have been studied. In scenario 1, there is a single miner with "n" nodes (i.e. n number of computers) and the output statistics are calculated on the basis of total computing power in "n" nodes. This setup is used to model and validate the output parameters to justify similarities to real world scenarios. In scenario 2, a group of "m" miners with "n" nodes each, compete in the bitcoin mining game to maximize their individual profits. There are associated costs with Bitcoin mining, such as the base cost of purchasing the equipment followed by the continuous electricity cost of mining. Hence, this scenario confirms the stochasticity in Bitcoin's algorithms as to the winner of each round in mining and whether it is profitable for all miners to participate everytime in this game.

2.2 Assumptions in Modelling

Bitcoin mining is a very complex process involving high amount of data, transactions, and miners from all over the world with multiple nodes. To simplify and for the purpose of to study the blockchain network from queuing theory point of view, following assumptions are made:

1. There are many threads running parallelly in the block mining process, however in our mining pool we do not consider these individual processes/threads, but consider that to be a black box and instead use the overall service times
2. The size of the block is limited by the "size" of each transaction, which is generally 1MB. Based on bitcoin real-life data, this comes to between 2000 and 3000 transactions. In our model, instead of the 1MB size, we use the constraint of 2100 transactions per block
3. In Scenario 2 it is assumed that all the miners always participate in mining every time a block is generated, and each of them have the same number of nodes for easier

comparison. In practice, every miner can choose, based on the reward policies and number of other miners, whether they would like to participate or not and with how much power. However, this would be in the purview of “Agent-Based” Simulation and outside the scope of current study

4. The input distributions for inter-arrival times and service times (Input Analysis)
5. In calculating the profitability of mining operation, the study ignores purchasing cost of equipment as it is a one-time cost

2.3 Python Modelling

Entire simulation was designed, and all the analysis were performed using python. Figure 2 shows the flow diagram approach that was taken towards developing the model.

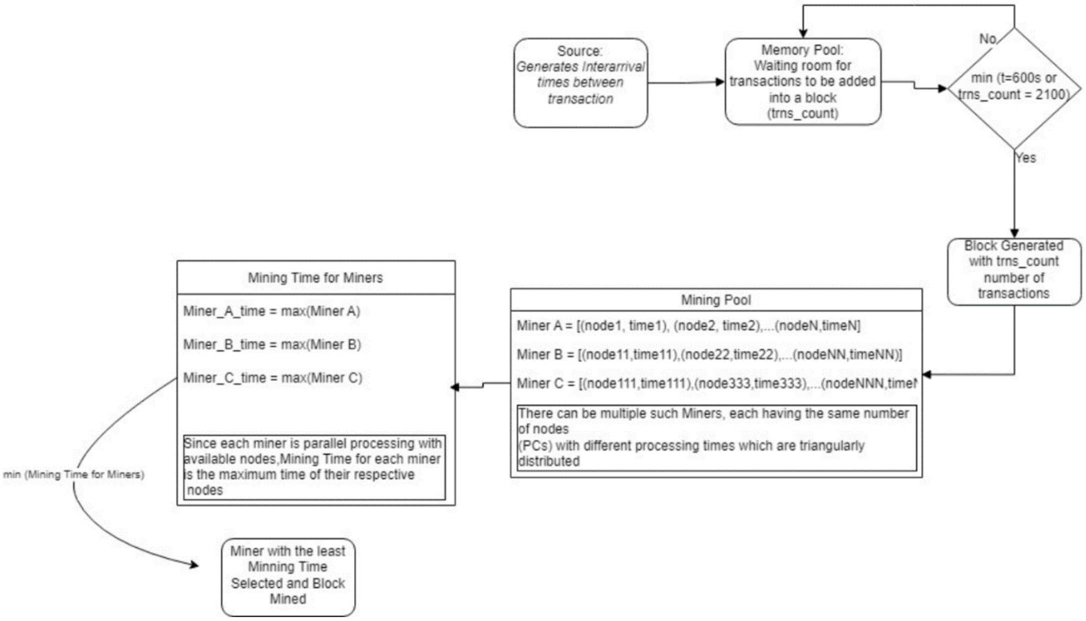

Source:
Generates Interarrival times between transaction
Memory Pool:
Waiting room for transactions to be added into a block (trns_count)
No
min (t=600s or trns_count = 2100)
Yes
Block Generated with trns_count number of transactions
Mining Pool
Miner A = [(node1, time1), (node2, time2),...(nodeN,timeN]
Miner B = [(node11,time11),(node22,time22),...(nodeNN,timeNN)]
Miner C = [(node111,time111),(node333,time333),...(nodeNNN,timeN
There can be multiple such Miners, each having the same number of nodes
(PCs) with different processing times which are triangularly distributed
Mining Time for Miners
Miner_A_time = max(Miner A)
Miner_B_time = max(Miner B)
Miner_C_time = max(Miner C)
Since each miner is parallel processing with available nodes,Mining Time for each miner is the maximum time of their respective nodes
min (Mining Time for Miners)
Miner with the least Minning Time Selected and Block Mined

Figure 2 : Flow Diagram of Python Models

For verification, multiple run-throughs of scenarios were done and print statements were given to various arrays for logical truth value. Error statements and incorrect values were intentionally given to check if flags are raised in the output if deviating from above logic

2.4 Data Collection and Validation

In this study, real-world data has been used as input and for validation of the output of the model. Table 1 in Appendix shows the various parameters and the source of the same. A lot of data exists about Bitcoins, hence the output parameters of the proposed model’s were compared to the real world data. The output parameters calculated are average number of confirmed transactions in a day, system throughput (number of confirmed transactions per second per block), average number of blocks mined in a day, and average number of transactions in a block

2.5 Input Data Analysis

Input data which the model takes is inter-arrival time between transactions and the processing power of the nodes available with each miner. Figure 3 shows the plot of raw values for transaction per second in bitcoin environment for the past 60 days. It can be seen that the transactions have achieved a steady state at 3.2 transactions per second. For this reason, we use "Stationary Poisson Process" to generate the inter-arrival rates with lambda 3.2 transactions/sec.

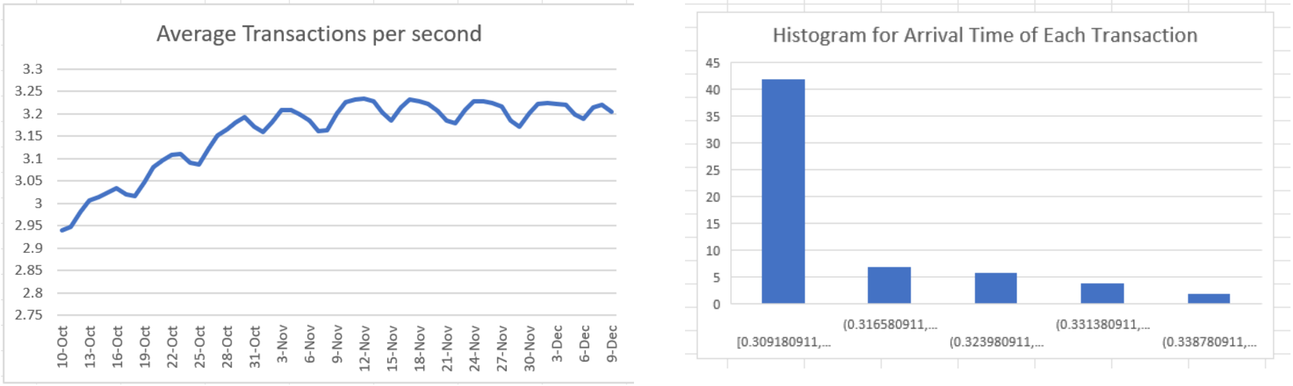


Figure 3: Transaction Inter-arrival rate and Arrival Rate Between Each Transaction over 60 days of data

The data is also imported to python and the arrival times between transactions are fitted to an exponential distribution. K-S test is used to find the goodness of fit of our distribution and it is accepted with aa p-value of 0.6972

The processing power of every node is measured in Terahashrate/second (Th/s). Since definitive values could not be obtained, sales data of GPU and ASICs, which are machines used for Bitcoin mining, was used. Based on their sales as per the processing power, a triangular distribution is used at each node. In Scenario 3, the profitability of mining is also calculated based on the high electricity cost involved with it. This cost is found out to be 1.75 x 10^{-6} \$/Th.

In bitcoin, to maintain the stochasticity, and not let the system be skewed by increasing mining power, there is a difficulty level which relates to how difficult is it to guess the nonce required during mining. This "Difficulty Level" keeps on updating every 14 days and is directly proportional to the total mining power in the pool. It is given by the below equation

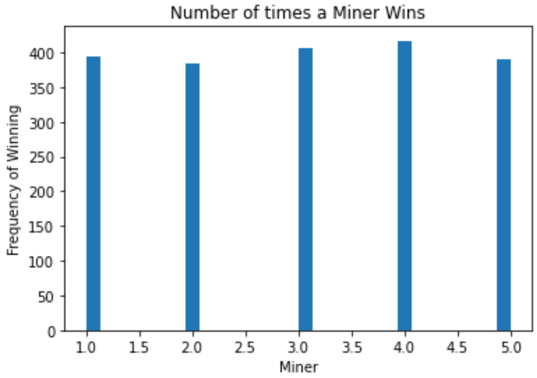
$D_{new} = D_{old} * t / (b_{14} * t_{av})$, [where D = difficulty level; t = total time to mine all the blocks in last 14 days; b_{14} = blocks mined in last 14 days; t_{av}= average block mining time]

3. Results and Discussions

Table 2 in Appendix shows the input parameters that have been used to run Scenario 1 and 2. Scenario 1 is run over a single day to check whether the parameter values are comparable to actual data before beginning with Scenario 2. Table 1 shows the summary results of the runs in both the scenarios.

Output Parameter	Actual Data	Scenario 1	Scenario 2	
			Point Estimate	95%Confidence Interval
Number of confirmed transactions in a day	277,672	279565	277222	(277428.161, 279899.172)
Number of blocks mined in a day	128	133	130.561	(125.504, 133.563)
Number of transactions in a block	2100	2101.99	2161.673	(2084.711,2239.035)
System Throughput (Number of confirmed transactions per block per second	3.17	3.396	3.286	(0.753 , 3.42)

Table 1 : Summary Results of Simulation Run



4. Limitation of Study

Although the study is successful in terms of the proposed model being comparable to the real-world situation based on output data, there are certain limitations to it. These limitations are mainly due to lack of domain knowledge and scope to which the study is based.

Firstly, it is not completely as to which are the multiple other processes which run in mining of block, alongside guessing the nonce. Since, the specificity of that process is assumed to be a black box, any impact on the output that changes in these processes could cause, cannot be identified. Considering this aspect, one of the other assumptions made in the model was that irrespective of the computing power of any node (available with a miner), the “job” of mining or guessing the nonce would be equally divided. In practice, the optimum decision would be that node with higher computing power takes more load as compared to one with lower computing power, to minimize block mining time and maximize the probability of winning. Lastly, it can be seen in the proposed model that even though the algorithm ensures stochasticity in terms of who wins and mining at that instant is profitable, overall, all the five miners went in loss. The reason for this being that the model made them forcefully participate in the mining of every block generated. To overcome this, a better model can be devised using agent-based simulation technique wherein based on certain criteria (difficulty level, number of other miners in the pool, reward levels etc.) a miner can check their probability of winning and then decide whether to participate or no.

5. Conclusion and Future Scope

The study is successful in developing a simulation model based on queuing theory, with the memory pool serving as M/M/1 server and mining pool being the M/M/c server. Even though simplified, the output results were comparable to real-world data validating the logic of our model and the algorithm developed. Based on the actual data, a Poisson distribution was considered for inter-arrival time between transactions, and it was verified by performing a K-S goodness of fit test. A single day run on scenario 1 showed the behavior of transactions and service rates in the block chain network. The second 'm' miner and 'n' node system of the mining pool was developed to run for 15 days and the output statistics with confidence intervals were calculated. The values resembled the trend with actual data with very low error percentage. **In terms of the queue length and unconfirmed transactions, it was observed that the rate at which bitcoins are being mined is a little bit lower than the rate at which they are arriving**. In the mining game between multiple miners, it was observed that the inherent stochasticity in solving the nonce leads to anyone winning the game however the overall profitability depends entirely on the decision of the miner on whether they participate or no. In conclusion, it is not recommended that someone keeps their mining machine on for the entire day.

Going ahead, this proposed model can be used to do what-if analysis as to how the mining scenario and output statistics change when every miner has different number of nodes and the load of jobs in mining pool are divided based on computing power and not equally. Consequently, in the mining game it has been shown that individual miners compete with each other, however in practice, such group of miners' form teams and compete with different teams in the bitcoin network. Each of such teams have a mining manager who makes decision on the size of group, division of rewards and power sharing policies within the team. Such a model can be used to help the mining mangers help make this decision to maximize profits for the entire team. Outside the purview of the course, agent-based simulation can also be employed to study miner behavior and introduce probabilities that would help a miner make decision whether to participate and if yes, how much computing power to use.

References

[1] Memon, R. A., Li, J. P., & Ahmed, J. (2019). Simulation model for blockchain systems using queuing theory. Electronics, 8(2), 234.

[2] Li, K., Liu, Y., Wan, H., & Huang, Y. (2021). A discrete-event simulation model for the Bitcoin blockchain network with strategic miners and mining pool managers. *Computers & Operations Research*, *134*, 105365.

[3] Bot, Anonymous. (2021). *Blockchain.com Explorer: BTC: ETH: BCH*. Blockchain Explorer Search the Blockchain | BTC | ETH | BCH. Retrieved December 9, 2021, from https://www.blockchain.com/explorer.

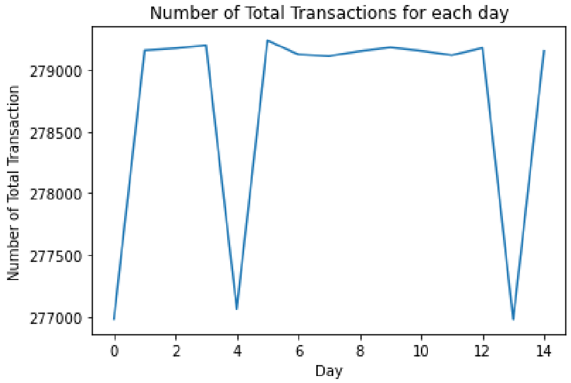
Appendix

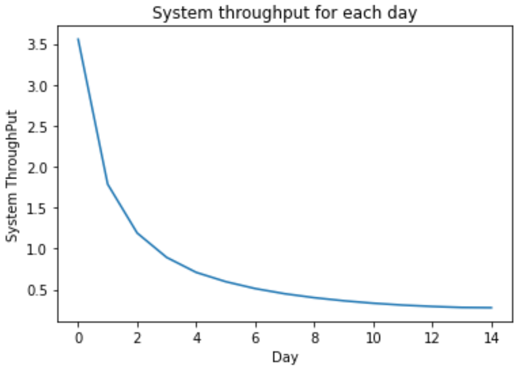
Table 2 : Data Sources and Purpose in the model

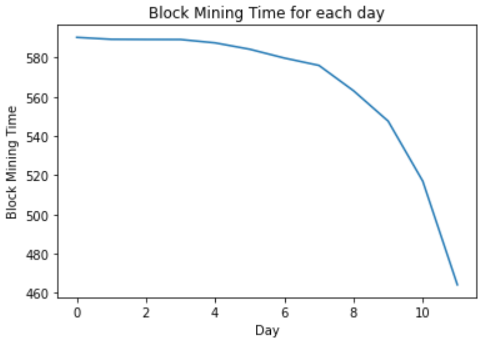
Data Type	Use in Model	Source
Transaction Inter-Arrival Rate	Fit a distribution to randomly generate as input data	Blockchain.com
Service Time at Node	Fit a distribution to randomly generate processing rate at each node	Sales data of ASIC/GPUs
Number of Transactions in a Day	Execution Parameter in the model	Blockchain.com
Average Blocks mined in one day		
Average number of transactions in one block		

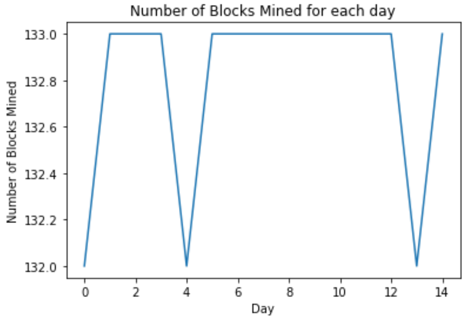
Station	Parameter Description	Values	
Memory Pool		Scenario 1	Scenario 2
	Arrival Rate of each transaction	Exponential Distribution (0.3125 transaction/sec)	Exponential Distribution (0.3125 transaction/sec)
	Queue Length	infinite	infinite
	Number of transactions	min (2100 transaction or 600 seconds)	min (2100 transaction or 600 seconds)
Mining Pool	Number of miners	1	5
	Number of nodes/miner	6	6
	Total number of nodes in pool	6	30
	Processing rate	Triangular Distribution (min = 140175299, mode = 170117315, max = 176960669)	Triangular Distribution (min = 140175299, mode = 170117315, max = 176960669)
Other Input Data	Electricity Cost	Not Required	1.76E-07 $/Th
	Bitcoin Reward in $	Not Required	$300,000
	Number of replications	1 day	15 days

Table 3: Data parameters for Simulation mode


Number of Total Transactions for each day
Number of Total Transaction
Day


System throughput for each day
System ThroughPut
Day


Block Mining Time for each day
Block Mining Time
Day


Number of Blocks Mined for each day
Number of Blocks Mined
Day