

MASTERS THESIS

On Defining Regions by Data Clustering for Increasing the TPS Rate of State-Based Blockchain

by

M Mohaiminul Islam

Under the Supervision of

Dr Amit Banerjee

Submitted in partial fulfillment of the requirements
for the award of the degree of
Master of Science in Computer Science

to the



DEPARTMENT OF COMPUTER SCIENCE
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
SOUTH ASIAN UNIVERSITY, NEW DELHI - 110021, INDIA
July, 2021

MASTERS THESIS

On Defining Regions by Data Clustering for Increasing the TPS Rate of State-Based Blockchain

by

M Mohaiminul Islam
SAU/CS(M)/2019/07

Under the Supervision of
Dr. Amit Banerjee

Submitted in partial fulfillment of the requirements for the award of the degree of

Master of Science in Computer Science

to the



**Department of Computer Science
Faculty of Mathematics and Computer Science
South Asian University, New Delhi - 110021, India
July, 2021**

Certificate

This is to certify that the thesis entitled “**On Defining Regions by Data Clustering for Increasing the TPS Rate of State-Based Blockchain**” submitted by **M Mohaiminul Islam (Enrollment No. SAU/CS(M)/2019/07)** to the Department of Computer Science, Faculty of Mathematics and Computer Science, South Asian University, New Delhi, 110021, India in partial fulfillment of the requirements for the awards of the degree of Master of Science in Computer Science, is a record of the bonafide work carried out by him under my supervision and guidance.

The results contained in this thesis have not been submitted in part or full to any other university or institute for the award of any degree or diploma.

Dr. Amit Banerjee

(Supervisor)

Department of Computer Science

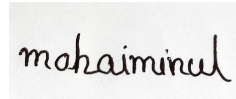
Faculty of Mathematics and Computer Science

South Asian University, New Delhi, India

July, 2021

Declaration

I hereby declare that, the thesis entitled **On Defining Regions by Data Clustering for Increasing the TPS Rate of State-Based Blockchain** being submitted to the Department of Computer Science, Faculty of Mathematics and Computer Science, South Asian University, New Delhi in partial fulfilment of the requirements for the award of the degree of **Master of Science in Computer Science** contains the original work carried out by me under the supervision of **Dr. Amit Banerjee**. The research work reported in this thesis is original and has not been submitted either in part or full to any university or institution for the award of any degree or diploma. Whenever I have used materials (data, theory and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the reference section.

A rectangular box containing a handwritten signature in black ink that reads "mohaiminul".

M Mohaiminul Islam
SAU/CS(M)/2019/07

Acknowledgement

First of all, I would like to express my heartfelt gratitude to my respected supervisor Dr. Amit Banerjee for continuous guidance and support through the whole procedure of the research and motivated me to accomplish the research in the field of Blockchain, and also I would like to thank my respected senior and my mentor for my work Mohd Sameen Chishti, who was always available for the insightful discussions and helping me day-night while I was getting stuck in any part of the thesis, and his continuous guidance and support made me to complete my work successfully. and I would like to express my deep and heartfelt gratitude for all of those who have helped, supported and encouraged me not only through this thesis but also all along my studies and education session till now, and at last, I want to thank my family and friends for their unconditional support.

M Mohaiminul Islam
SAU/CS(M)/2019/07
July, 2021

Abstract

Blockchain is a distributed ledger technology where the miners verify or execute the transactions sequentially. This is one of the major problem of the blockchain network. More specifically, due to sequential mining the transaction per second (TPS) rate of the state-based public blockchain is not adequate enough for applications with high network load, such as supply chain management system or trading. For this, various researchers have proposed various techniques for parallelizing the mining process to increases its TPS. In this thesis, we consider one such technique, that uses region based concept for distributing the transactions to multiple miners for its simultaneous validation. That is, we investigate a framework for improving the TPS rate of current day blockchains, by clustering the transactions into disjoint subsets and mining them in parallel. The idea is to improve the TPS rate of a blockchain. The implementation uses a designated server for clustering the data and for broadcasting the transactions to the miners. Due to data clustering the transactions can be used searched in the blockchain. In our implementation, we use the k-mean for data clustering, Our experimental results demonstrate the benefits of the proposed methodology for the supply chain-based blockchain.

Keywords: Blockchain,Supply Chain,K-Mean algorithm,parallel mining,TPS rate

Contents

Declaration	v
Acknowledgement	vi
Abstract	vii
List of Tables	1
1 Introduction	2
1.1 Introduction	2
1.2 Motivation	5
1.3 Problem Statement	6
1.4 Scope of The Proposed Work	6
1.5 Contribution	7
1.6 Thesis Outline	7
2 Preliminary	8
2.1 Blockchain	8
2.2 Frequently used Terminologies	10

3 Literature Review	12
3.1 Blockchain	12
3.2 Supply Chain	13
3.3 Parallel Execution of Transactions in Blockchain	13
4 Proposed Approach	15
4.1 Overview of the Proposed Framework	15
4.2 Clustering of Transaction	17
4.3 Parallel Mining Concept	19
5 Experiment Setup and Results	21
5.1 Clustering Transactions	21
5.2 Comparison of sequential and parallel mining	22
6 Conclusion and Future Works	24
References	25
Appendices	29
A Demo supply chain based blockchain	30

List of Figures

1.1	Traditional Supply Chain Management [20]	3
1.2	Blockchain Based Supply chain [5]	5
2.1	Blockchain Technology	9
4.1	System Model	16
4.2	Block Transaction	17
4.3	Transaction Clustering	18
5.1	Elbow Method	22
5.2	Clustering The Transaction	22
5.3	Delay in Sequential Mining with Increasing difficulty level	22
5.4	Parallel Mining with Increasing the Number of Transactions	22

List of Tables

1.1	Transaction speed of different blockchain based cryptocurrency. [9]	. . .	4
-----	---	-------	---

Chapter 1

Introduction

1.1 Introduction

The blockchain technology is a disruptive and innovative technology that has the ability to fundamentally alter the way we store and interact with information. According to neutral sources, 10% of global GDP will be kept on blockchain, which means that blockchain databases will store 10% of global GDP, according to the World Economic Forum. Other sources estimate that the value added by blockchain would accelerate to 175 billion by 2025, [17]. Blockchain is not a band-aid solution; it will eventually replace existing processes in a variety of businesses. Blockchain is a distributed ledger, [14], which implies that the ledger may be controlled by numerous nodes situated anywhere in the world. It is truly distributed computation performed by the nodes. It is decentralised in the sense that each node can choose whether or not to create a block.

Essentially, a blockchain is a collection of blocks. Each block contains a cryptographic hash, which means that if we change any small number or digit, the hash will change, which means that if we tamper with anything, the link to the previous block will be lost due to the hash changing, which is why this technology is so powerful. There are a few bad nodes that are actively attempting to bring the network down, and

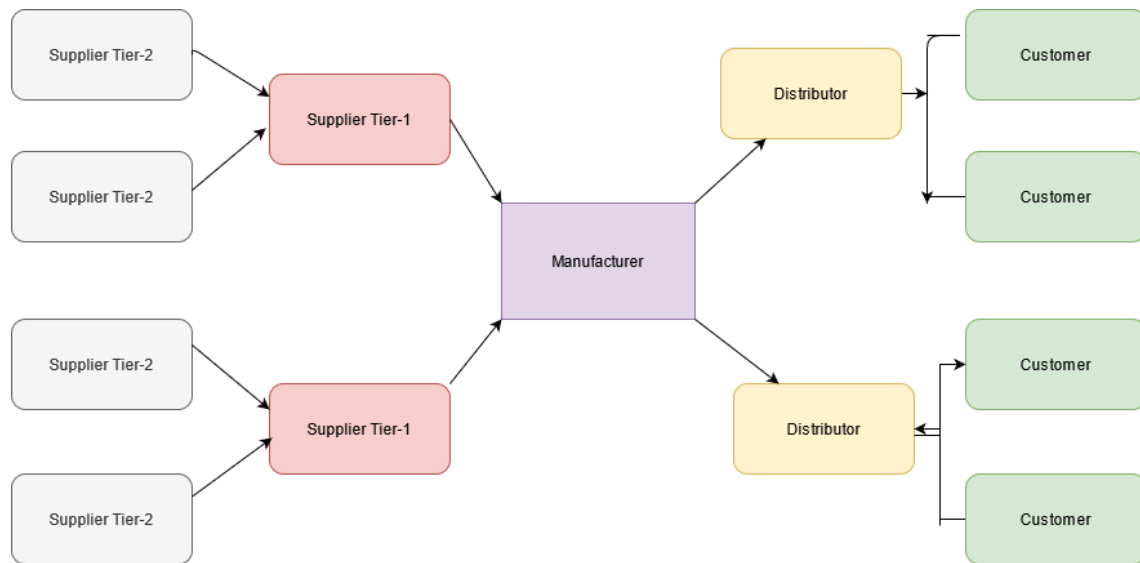


Figure 1.1: Traditional Supply Chain Management [20]

to address this issue, we utilise a consensus mechanism that sits on top and permits the block since it is viewed as the one defining the set of transactions by the majority of nodes. Because the majority of transactions occur between nodes, blockchains appear to be peer-to-peer networks. A node in the network may act as a miner or validator, and they enter into transactions. These transactions are then loaded onto the blockchain, and the miner who cracks the block creates a new block of the transactions that are linked together. This block is actually a series of blocks known as a blockchain. This technology is frequently used in anti-counterfeiting and security applications. If we examine the modern financial system, land system, or supply chain system, a third party is involved to verify the transaction. For example, if one wishes to purchase goods from a market using a credit card, the transaction is verified by a bank. If one wishes to pay with cash, he/she must first withdraw money from the bank. Thus, where transactions are centralised, a third party is involved to validate the transaction. As a result, the likelihood of a single point of failure is extremely high. We can address this issue through the use of blockchain technology, which is a decentralised, immutable ledger system that can be permissioned or permissionless. Client information and transactions are kept anonymous on the blockchain, and each user of the chain receives a copy of the ledger.

Blockchain Technology	TSP Rate	Avg. confirmation time
Bitcoin	3-7	25 min
Ethereum	15-20	2 min
Ripple	1500	4 sec
Bitcoin Cash	61	60 min
Cardano	5-7	3-5 min
Litecoin	26	30 min
Monero	4	30 min
Neo	1000	15-20 sec

Table 1.1: Transaction speed of different blockchain based cryptocurrency. [9]

In a traditional supply chain, there are various issues. For example, in the food supply chain, we don't know how the food is processed, whether it's organic or not, there is a lack of information, data retrieval takes a long time, and the data is unreliable for product tracing. We can fix this issue through the use of blockchain technology. Walmart has previously implemented blockchain technology in their retail stores, but the primary difficulty with the blockchain is its scalability and throughput, [10]. Due to the size of the supply chain market, it will be extremely difficult to maintain this volume of investor and transaction data, and due to the limitations of traditional blockchain technology, the current block size is only 1 MB with a TPS of only 7, and anyone initiating a transaction must typically wait 10 minutes. Thus, this is one of the major blockchain difficulties. Whatever is implemented using blockchain technology has a limited throughput, including bitcoin, Ethereum, and ripple, but visa/credit cards can process thousands of transactions per second.

In this thesis, we try to improve the TPS rate of a state-based blockchain, by clustering the transactions and parallel mining. The TPS of the popular blockchain are shown in Table-1.1. More specifically, we use the region based concept for distributing the transactions to multiple miners for its simultaneous validation. The idea is to improve the TPS rate of a blockchain. The implementation uses a designated server for clustering the data and for broadcasting the transactions to the miners. Due to data clustering the transactions can be used searched in the blockchain.

1.2 Motivation

In the market, supply chain management is a solid rule to follow. However, as is the case with traditional supply chains, several issues arise, such as the lack of information about goods, the length of time required to retrieve data, and the difficulty of tracing a product. For example, a few years ago in China, when Walmart was selling pork [2] in its retail stores, they discovered that one batch of pork was actually infected and unfit for consumption. Thus, they attempt to determine where that batch of pork originated and where it could have gone, but there is no information available due to the supply chain's complexity. The pork might have originated on a farm, been housed in a facility with a refrigerator, and then made its way to a Walmart retail shop. Thus, it had three or four owners and changed hands several times before reaching the Walmart retail locations. As Walmart had no idea where the pork came from or where it went, they were forced to recall all meat in China, which cost them millions of dollars. However, if we employ blockchain technology in the supply chain, all of the information can be stored in a database, allowing us to see where raw materials originate from, who the vendor is, and who purchases it all.

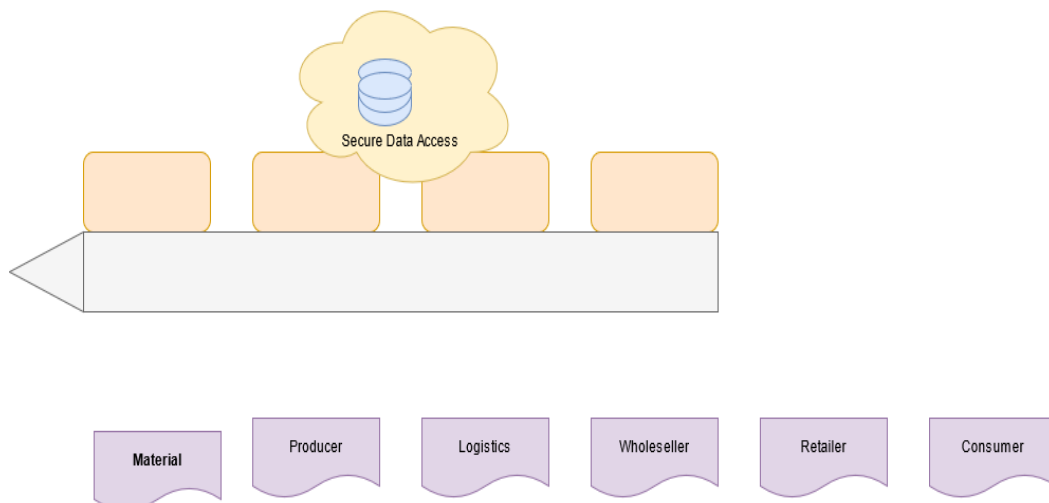


Figure 1.2: Blockchain Based Supply chain [5]

However, today's blockchain technology has a number of limitations. Most notable is its low TPS rate, which has a considerable impact on the system's scalability. Table-

1.1 shows the TPS of a few popular blockchains. As transactions increase daily, the likelihood of a fork increases, resulting in speed decreases. The present block size on the blockchain is just 1 MB, with a 10 – 15 minutes inter-block arrival time. As a result, the transaction per second rate is low. Thus, unless we overcome this issue, businesses will be discouraged from using this technology, and numerous researchers have concluded that this is the primary reason why individuals are unwilling to use this technology.

1.3 Problem Statement

The primary difficulty with the blockchain is its limited throughput, which results from its slow transaction speed. When we integrate a supply chain system with a blockchain system, we observe a throughput issue.[21] discovered that when we increase the number of concurrent jobs in a chaincode, this supply chain-based blockchain takes a long time and its response time increases exponentially. Additionally, data retrieval takes a long time. For example, if we send 20 queries to the supply chain-based blockchain technology, we will receive a response in 0.5 sec. However, if we send 50 inquiries, it will take significantly longer, approximately 1 sec. The reason of low TPS rate is that the transactions are processed and verified serially. The goal of this thesis is to increase a blockchain's TPS rate through the use of the region-based parallel mining concept. This is accomplished by grouping transactions into distinct subsets. The proposed methodology not only increases the TPS rate, but also makes it easier to locate linked transactions in the blockchain.

1.4 Scope of The Proposed Work

In this, we investigate a framework for improving the TPS rate of current day blockchains, by clustering the transactions into disjoint subsets and mining them in parallel. The proposed system can be used in any system that generates huge number

of transactions and needs to validate the transactions in quickly, such as in supply chains or trading.

1.5 Contribution

The contributions of our work are as follows:

- In this proposed framework, we try to define the concept of regions by clustering the transactions into disjoint subsets to assist parallel mining for improving the TPS rate of a state-based blockchain. For this, we investigate the procedure for clustering by utilising the *data* field of the transactions and a designated entity (referred as manager), for clustering and broadcasting the transactions to the miners of a blockchain network
- We perform a detailed evaluation to analyse the performance of the proposed system.

1.6 Thesis Outline

The rest of the thesis is organized as follows. In Chapter 2 and chapter 3, we explain some important terminology used in our work and discuss the related works, respectively. In Chapter 4 we discuss the proposed methodology of defining the regions and integrating it with the parallel mining concept to improve the TPS rate of a blockchain. In this chapter, we have also give the pseudo-code of the algorithm used for the simulation. In Chapter 5 Presents the experimental evaluation of the proposed model. Finally, we conclude our work in Chapter 6

Chapter 2

Preliminary

2.1 Blockchain

Blockchain technology is a distributed ledger system [14] that enables the storage of records and their traceability. It is capable of tracking a wide variety of data, ranging from financial transactions to supply-chain information. Blockchain technology organises data into distinct groups called blocks that are chronologically linked together. If we wish to alter or counterfeit the data contained in a particular block, we cannot alter or rewrite it. As an example, suppose Nisha and her brother Niloy have been feuding for years over who owns the family's plot of land. Due to the ledger-based nature of blockchain technology, there is an entry in the ledger indicating that Nazrul acquired the property in 1980. When Nazrul sold the property to Nipa in 2000, a new node was created in the distributed database, and each change in ownership of this property is represented by a new node in the ledger, all the way up until Nisha purchased it from their father in 2010, at which point we can see the entire history of the owner in the database. Each block contains some data, including the hash of the previous and current blocks. The data that is stored on the block varies according to the blockchain type. Once a block is created, its hash is computed. Hashes are extremely effective for detecting changes to blocks. Because it is the initial block, it does not include any references to preceding blocks. This block is referred to

as the genesis block in the figure 2.1.

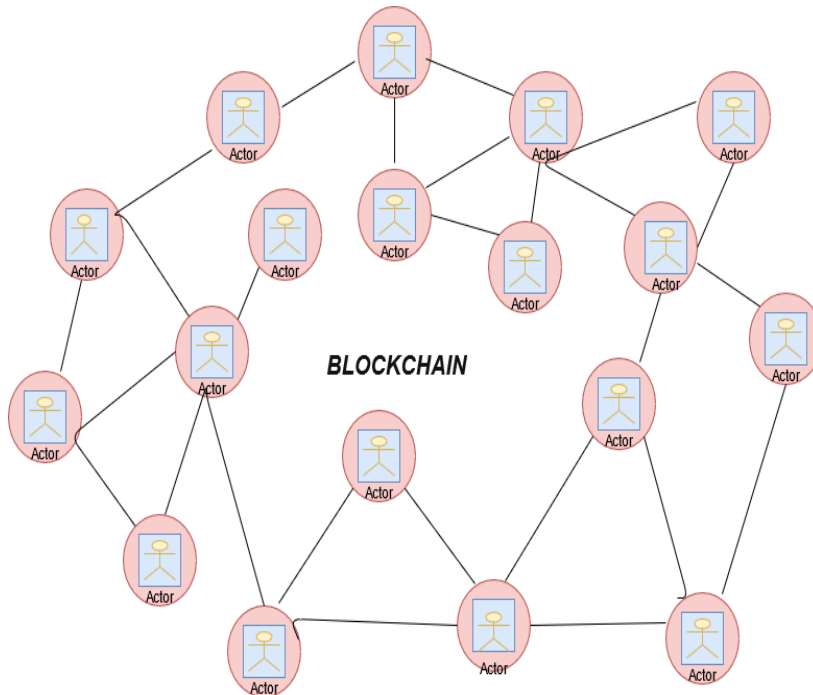


Figure 2.1: Blockchain Technology

In general, blockchains are classified as (a) public, (b) private, or (c) consortium. The public blockchain is the most widely used of the three, with developers porting many applications to it, including supply chain management and bitcoin. The public blockchain is also referred to as the permissionless blockchain, as the data is publicly accessible and does not require permission to view. This allows anyone to view and produce data, while remaining secure and maintaining the user's identity. Additionally, anyone can run a node, which gives 100 percent data openness. It does, however, have some drawbacks, including limited scalability and a low TPS rate. A private blockchain, also known as a permission-based blockchain, on the other hand, is administered by a central institution. As a result, users must obtain authorization to access data stored on a private blockchain. It operates in a closed environment with well-defined participants. Only pre-approved businesses are permitted to run nodes; mining of transactions is not necessary, and nodes are not compensated with tokens due to the high incentive and performance. Hyperledger Fabric, Corda, and Quorum are all examples of private blockchains. Finally, the consortium blockchain is

a cross between private and public blockchains, with a portion of the data preserved on each platform.

2.2 Frequently used Terminologies

1. **Miner:** Anyone with sufficient computing power and a CPU can become a miner. In essence, the miner verifies the transaction in order to create a new block. and this block contains information about the user, the transaction, and so forth. Typically, miners solve a cryptographic puzzle and calculate a hash value in order to create a new block. The first person to solve this challenge and make a block will receive a reward.
2. **Nonce:** Nonce are essentially used to generate a hash, as we know that each block has a unique hash value to prevent data manipulation. Each unique value of a nonce generates a unique hash, and from that created hash, only one hash reaches the destination. If we modify the nonce, the hash will be modified.
3. **Hash:** A hash function is a cryptographic function that accepts numerical data of any length and converts it to numerical data of a fixed length. If we modify the input, the hash value changes as well.
4. **Clustering of Data:** It is a technique for disjointly grouping related or frequent elements. There are various methods that accomplish this, like k-Mean, DBSCAN [16].
5. **Parallel Mining:** Parallel mining is a concept that has been introduced in many researches to increase the TPS of a blockchain by mining the blocks concurrently [1, 4, 11]
6. **Consensus Mechanism:** Numerous consensus mechanisms are used in blockchain technology, including POW, POS, and PBFT to ensure security, trustworthy, and immutable in the blockchain. This technique ensures that each new block

2.2 FREQUENTLY USED TERMINOLOGIES

added to the blockchain is trusted and that the new block becomes part of the blockchain when all nodes in the blockchain agree on it.

7. **Proof of Work (PoW):** A blockchain is made up of numerous nodes or miners. If we wish to add a new block to the network, miners must solve a mathematical problem, and the miner who solves it first is referred to as POW. If fifty percent of the network's nodes or miners validate it, a new block is added to the network. The following are the issues raised in PoW: i) excessive energy use due to the fact that only one miner has solved it, while the rest are attempting but failing due to their energy consumption, which is utterly wasted. ii) If additional miners join and attempt to solve the riddle, a 51 percent attack is possible, which is detrimental to the decentralised system.

Chapter 3

Literature Review

3.1 Blockchain

In [7] authors perform a systematic review of the literature in which they identify several research gaps. According to GDPR articles, data deletion and modification are the most hotly debated topics in the blockchain community. Additionally, the disciplines of IoT and healthcare are the most mentioned study areas. The authors of this study [12] stated that while blockchain technology prevents data manipulation, when huge IoT devices are used, the number of transactions generated increases, potentially increasing latency and cost. The authors employ blockchain-indexed storage (BIS) in the study to store data and information in off-chain storage using BIS, to control the latency and cost.

Similarly, in this document [19], the authors assess the blockchain system's performance, noting that the current system's blockchain traceability system's performance is extremely low. It supports 159 transactions per second and has a convergence time of just 4.71 seconds. This plainly indicates that this system has a throughput problem. They attempt to resolve this issue in this paper. The authors proposed segmenting the network into distinct segments and connecting them together to generate the required throughput.

3.2 Supply Chain

The authors of [21] demonstrate how they are utilising blockchain technology in supply chain data management to address existing supply chain management issues such as a lack of information, product traceability issues, and lengthy delays. They evaluate performance in terms of data submission throughput and data query speed. Similarly, [22] examines the issue of supply chain trust and centralization and proposes a basic framework for utilising the compatibility of blockchain technology to achieve supply chain trust and decentralisation. In [13], the authors presented a trust management framework (referred as Trustchain) for a supply chain management system in order to alleviate the trust issue associated with commodities and data on the blockchain. *Trustchain* tracks supply chain actors via a consortium blockchain. They resolve the trust issue here, but do not benefit from the increased throughput and latency provided by this framework.

Similarly, the author in [23] deploys blockchain technology in supply chain. Because blockchain is a decentralised platform that stores records with an immutable and permanent historical trail, implementing supply chain using this technology will be beneficial. They demonstrate how blockchain may be utilised in the supply chain and address issues associated with traditional supply chain-based blockchains such as privacy, high latency, limited throughput, and speed in this article. Similarly, in this review paper, the authors [15] identify some research gaps in supply chain-based blockchains, such as data storage being prohibitively expensive while failing to establish trust in intermediaries, scalability issues, and size issues in supply chain-based blockchains, in order for future research to address and work on these gaps.

3.3 Parallel Execution of Transactions in Blockchain

In [6], authors address the fact that while smart contracts are run sequentially by miners and subsequently serially re-executed by validators, various throughput issues arise. These issues can be mitigated by allowing miners to execute smart contracts

3.3 PARALLEL EXECUTION OF TRANSACTIONS IN BLOCKCHAIN

in parallel. In this article, the miner performs speculative operations on contracts in order to reach an overall speed of 1.33x for miners and 1.69x for validators. Due to the fact that blockchain technology's transaction verification is extremely sluggish. The authors in [9] suggested a model that is proof of work based on parallel mining and ensures that no more than two miners solve the mathematical challenge, as if all the miners in the chains are engaged solving the riddle, a great deal of energy and time is wasted. As a result, they provide a 34% increase in scalability over the old system.

Similarly, the authors [11] replace the chain data structure with the graph data structure dubbed graphchain since the chain data structure is a single line, which prevents parallel processing. Thus, in this study, they incorporate this graphcha into the model and encourage miners to contest for election, and this model enables several leaders to mine concurrently. Although performance is not enhanced linearly in this work, they attain a higher level of performance than the present model. As researchers recognized the low TPS rate as a serious issue, they developed novel ideas such as parallel algorithms [18], parallel mining and parallel proof of work [9?], region-based parallel mining [4], and frameworks such as *DiPETrans* framework [1], batched chain. Methods such as *PPLFS* (parallel proof of luck and fair share model) [3] can be used to efficiently increase the throughput and scalability of the blockchain network.

Chapter 4

Proposed Approach

From the above discussion and literature review, we see that the significant problem in the supply chain-based blockchain is its low TPS rate. we found the primary two issues of this low TPS rate.

1. The transaction size in the blockchain is very big cause and only one item we can put in each transaction leads to the low system throughput.
2. In supply chain based blockchain transactions verifies one by one, which is another reason for the low TPS rate.

4.1 Overview of the Proposed Framework

Figure 4.1, presents an overview of the proposed system framework. In this model we consider the inclusion of a designated server for clustering the transactions. It is referred to as the Manager in the proposed framework. It is assumed that the transactions of all users, passes through the manager. The manager collects all transactions and can cluster the data into disjoint subsets. Finally it broadcasts each cluster to the miners. The miners can accept any cluster or clusters that it wants to mine and finally upon successful verification the transactions are included in the blockchain.

4.1 OVERVIEW OF THE PROPOSED FRAMEWORK

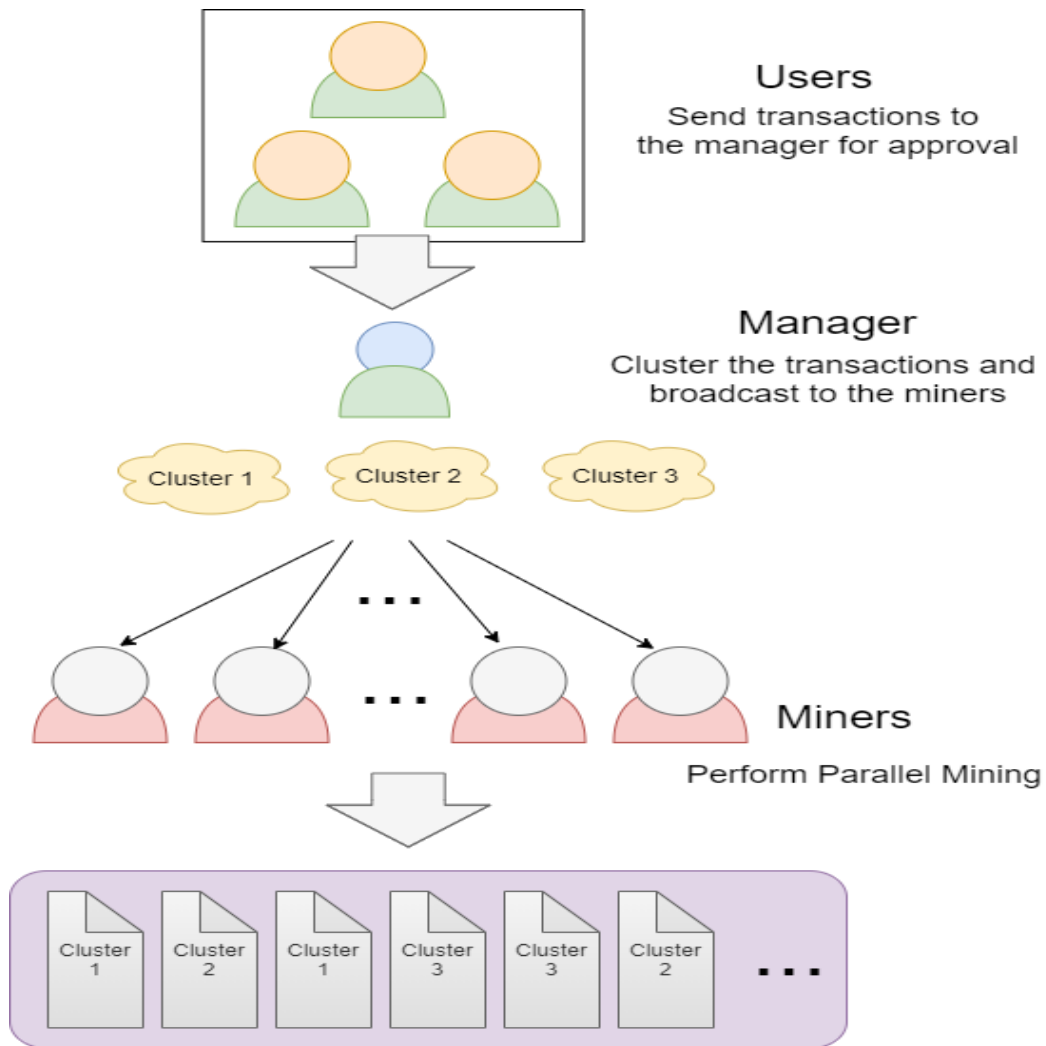


Figure 4.1: System Model

A blockchain transaction consists in diagram 4.2 of various fields, including to, from, amount, nonce, hash, timestamp, mine fee, nonce, gas fee, gas limit and data. The to and from fields are the unique addresses of the initiator and the target entities. The amount field shows the value transferred by the initiator. The timestamp shows when the transaction is initiated and the data field contains additional information about the transaction. In figure 4.1 we show the example of a blockchain transactions. For example the first transaction shows that the U_1 is transferring 10 dollars to U_2 at time TS_1 to buy fruits. The users of the proposed framework can initiate such transactions and forward the same to the manager for clustering.

1. Block : <Block number>
2. Transaction Hash : <Unique transaction hash>
3. Timestamp: <Confirmation time of transaction>
4. From: <Initiator's address >
5. To: <Target address>
6. Value: <Amount to be transferred>
7. Fee: <Miner's fee >
8. Gas Price: <Price offered per unit of gas>
9. Gas Limit: <Max amount of gas to be utilized>
10. Nonce: <Position of transaction >
11. Data: <i>Fruit : Guava ; Vegetable : Tomato</i>

Figure 4.2: Block Transaction

In this figure The manager periodically checks this incoming transactions and can use any clustering algorithm like K-Means for grouping the transaction into disjoint subsets. In figure 4.3 we show three cluster where the algorithm uses the first data field as a parameter for clustering. In the figure shows three clusters of meat,veg and fruits. The manager finally broadcast the transactions of each cluster for parallel mining. The details of the proposed framework is given below.

4.2 Clustering of Transaction

Since data of various transactions are distributing throughout the ledger, we need to cluster these to form a similar related transaction for a block. In our implementation, we apply the K-Means clustering algorithm to group these similar transactions into different clusters. In the following, we discuss the k-means algorithm [8], which is used in the proposed implementation for partitioning the transactions into k clusters. The algorithm as described by[16] starts with a random set of k center-points (μ). During each update step, all observations x are assigned to their nearest center-point

4.2 CLUSTERING OF TRANSACTION

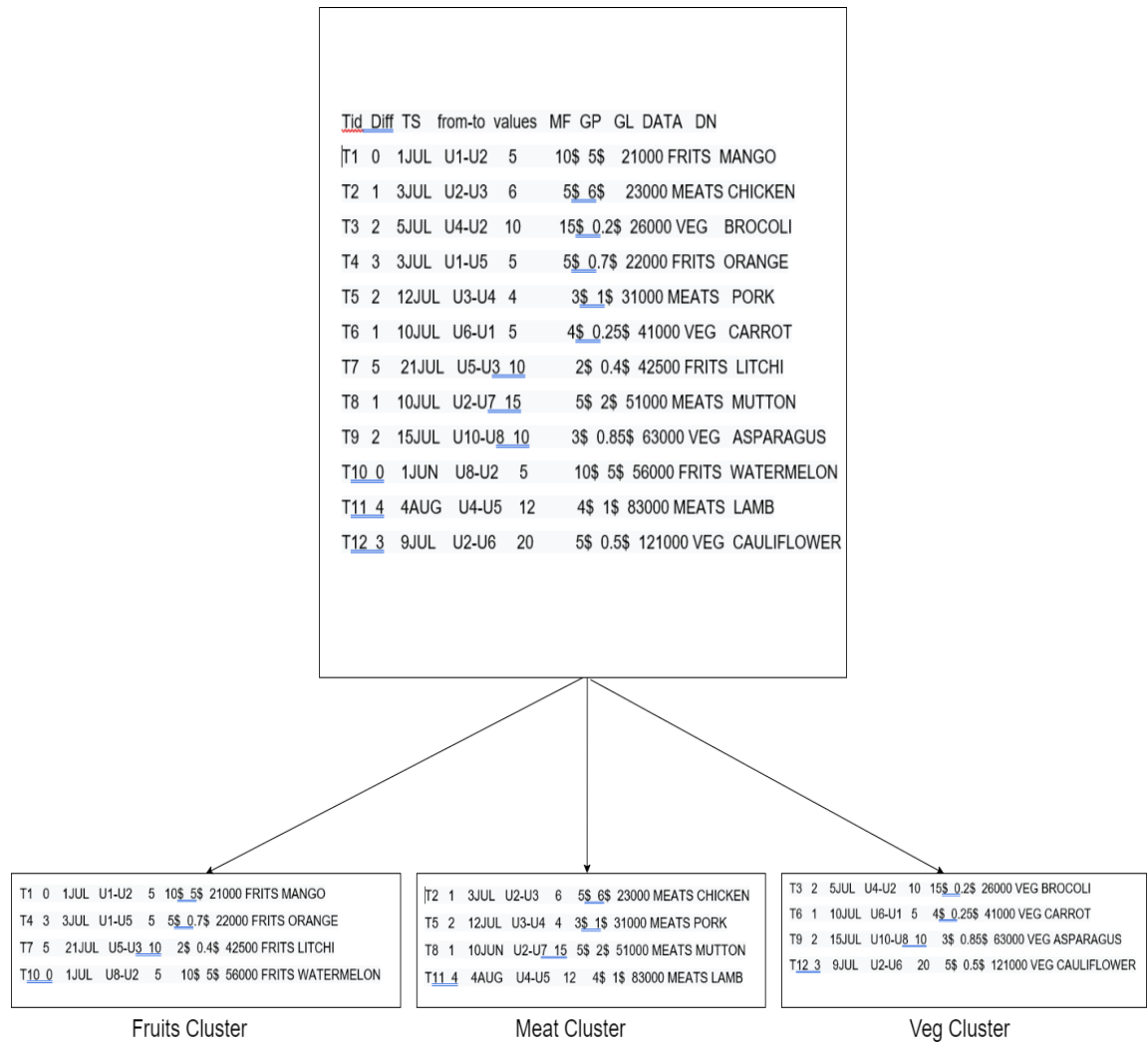


Figure 4.3: Transaction Clustering

(see equation 4.1). In the standard algorithm, only one assignment to one center is possible. If multiple centers have the same distance to the observation, a random one would be chosen.

$$S_i^{(t)} = \{x_p : \|x_p - \mu_i^{(t)}\|^2 \leq \|x_p - \mu_j^{(t)}\|^2 \forall j, 1 \leq j \leq k\} \quad (4.1)$$

Afterwards, the center-points are re-positioned by calculating the mean of the

assigned observations to the respective center-points.

$$\mu_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j \quad (4.2)$$

K-means clustering algorithm:

0. Start with initial guesses itemsets for clustering the centers (centroids)
1. For each data point, find closest cluster center (partitioning step)
2. Replace each centroid by average of data points in its partition
3. Iterate 1+2 until convergence

4.3 Parallel Mining Concept

After clustering, the manager broadcast this cluster to the miners and then miners mined it region based parallel mining. we can achieve the better TPS rate by using parallel mining [4] rather than solo mining. for that we have to divide the blockchain network into n different segments called as regions and cluster this regions. in this method a miner can only mine for a specific region.

The pseudocode for the cluster based parallel mining technique strategy is 1, which is a variation of the algorithm described by the authors in [4]. It is comprised of two functions: *ClusterMining()* and *Validation()*. Miners use these routines to mine blocks in a cluster and to validate the blockchain. We assume in *ClusterMining()* that miners are evenly distributed throughout the clusters, i.e., any cluster can be chosen by any miners. Another supposition made in the pseudocode is that the number of clusters is pre-determined. Take note that these variables can vary according to the amount of transactions in a region, the total transaction costs, and the competition between the miners. *PBH* is the preceding block hash in the function, which is appended to the block to prevent data tampering.

Algorithm 1: Cluster Based Parallel Mining [4]

```

1 Function clustermining ( $C_k, n$ ):
   | // the miners receives cluster  $C_k$  and number of transactions
   |  $n$ 
2   | if Miner wants to mine  $C_k$  then
3   |   | Initialize block  $B_k$  for cluster  $C_k$ 
4   |   | foreach all  $n$  transaction  $T_i : U_i \rightarrow U_j \in C_k$  do
5   |   |   | Assign  $T_i \rightarrow B_k$  for mining
6   |   | end
7   | else

8 Function mineblock ( $B_k$ ):
9   | Broadcast_Block( $B_k$ )

10 Function Validation ( $B_k$ ):
11   | foreach  $k = 1$  to  $n$  do
12   |   | Validation( $B_k$ );
13   | end
14   | set PBH =  $H(B_1|B_2|B_n)$ 

```

Chapter 5

Experiment Setup and Results

In the following we evaluate the performance of the proposed framework. we first compute the delay and then compare the sequential and parallel mining technique.

5.1 Clustering Transactions

For clustering the transaction we create our own dataset in which the data field of the transaction are labelled as shown in figure 4.2. For the experiment, we consider 70 user where making random transactions for purchasing different product items. We use python 3.0 for implementing the K-Mean algorithm and show the result of our dataset.

In This figure 5.1 **Elbow** method shows how many cluster will be good for this transaction dataset. Here we take 3 clusters: the fruits cluster and the other two are veg cluster and meat cluster. And this figure 5.2 we cluster our transaction dataset and The yellow color shows the fruits cluster means this algorithm cluster all the fruits related item and blue color shows the meat cluster and green color indicates the veg cluster where all the meat and veg related items are clustered. now we can put this related transaction item into the blockchain and after mining them parallelly so that we can achieve better TPS rate.

5.2 COMPARISON OF SEQUENTIAL AND PARALLEL MINING

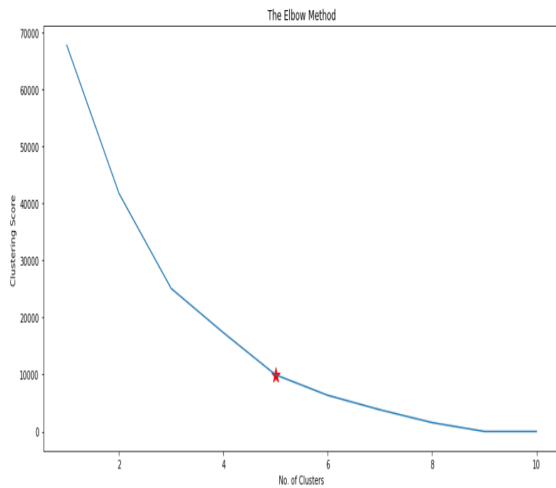


Figure 5.1: Elbow Method

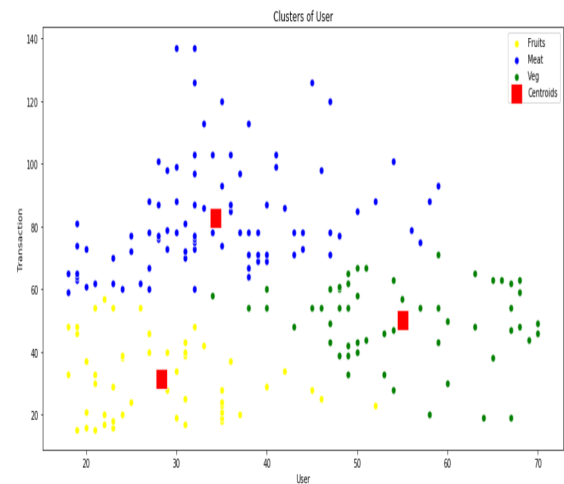


Figure 5.2: Clustering The Transaction

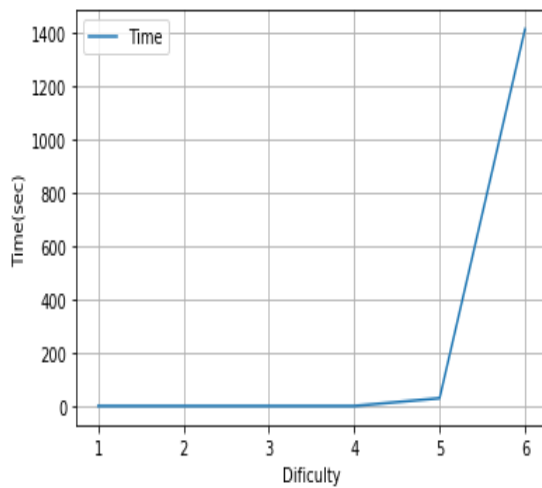


Figure 5.3: Delay in Sequential Mining with Increasing difficulty level

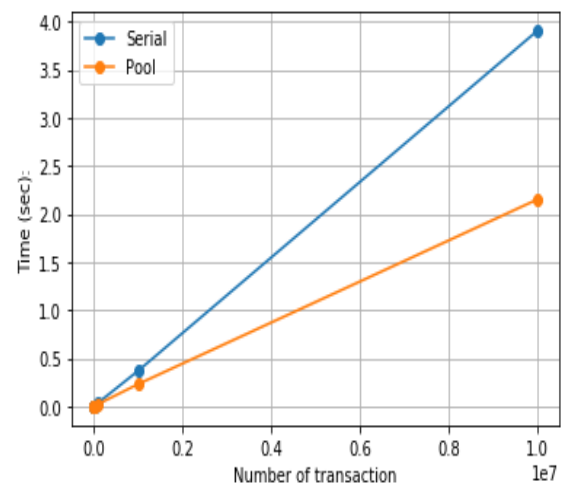


Figure 5.4: Parallel Mining with Increasing the Number of Transactions

5.2 Comparison of sequential and parallel mining

In a traditional blockchain, it verifies the transaction one by one means serially, so mined a block is very time-consuming. In this graph, we see that with the transaction that if the transactions are increased, it will take lots of time to validate the transaction. We can resolve this issue through parallel mining. Here we took a dummy datasets and apply 4 pool in python and saw that we get better result through parallel mining.

The figure 5.3 illustrates the time necessary for sequential mining to validate

5.2 COMPARISON OF SEQUENTIAL AND PARALLEL MINING

transactions of increasing difficulty. The difficulty of mining a block in a blockchain is a metric that indicates how tough it is to do so. A high difficulty indicates that verifying transactions on a blockchain requires greater processing resources. The difficulty is critical because it can help protect the blockchain network from malicious attacks. The figure 5.4, shows a comparison between sequential and parallel mining techniques for increasing number number of transactions.

Chapter 6

Conclusion and Future Works

In a blockchain system, it is clear that low TPS rate is the major issue that affects its acceptance. The goal of this work is to investigate a framework for increasing the TPS rate of modern blockchains by grouping transactions into disjoint groups and mining them concurrently. A dedicated server is used to cluster the data and disseminate the transactions to the miners. Transactions can be utilised to search the blockchain due to data clustering. We implement the proposed approach using k-mean algorithms for data clustering. Our experimental results indicate the proposed methodology improves the TPS rate of the blockchains and group related transactions together for faster search. In the future, we want to do experimental research on the system using an actual supply chain.

References

- [1] Shrey Baheti, Parwat Singh Anjana, Sathya Peri, and Yogesh Simmhan. Dipetrans: A framework for distributed parallel execution of transactions of blocks in blockchain. *arXiv preprint arXiv:1906.11721*, 2019. 10, 14
- [2] Boston379. Bringing blockchain to wal-mart chinaâs supply chain. <https://digital.hbs.edu/platform-rctom/submission/digital-pork-bringing-blockchain-to-wal-mart-chinas-supply-chain/>, November 2017. 5
- [3] K Lino Fathima Chinnarani et al. Pplfs: A high performance consensus method for improving throughput and scalability in blockchain network. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(11):3705–3714, 2021. 14
- [4] Mohd Sameen Chishti and Amit Banerjee. Increasing tps rate of state-based blockchains by parallel mining. *Internet Technology Letters*, 4(2):e220, 2021. 10, 14, 19, 20
- [5] Stuart Corner. Blockchain based supply chain management. <https://www.iothub.com.au/news/combining-iot-and-blockchain-for-supply-chain-tracking-492501/>, 2018. x, 5

- [6] Thomas Dickerson, Paul Gazzillo, Maurice Herlihy, and Eric Koskinen. Adding concurrency to smart contracts. *Distributed Computing*, pages 1–17, 2019. 13
- [7] AKM Bahalul Haque, AKM Najmul Islam, Sami Hyrynsalmi, Bilal Naqvi, and Kari Smolander. Gdpr compliant blockchains—a systematic literature review. *IEEE Access*, 2021. 12
- [8] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1):100–108, 1979. 17
- [9] Shihab Shahriar Hazari and Qusay H Mahmoud. A parallel proof of work to improve transaction speed and scalability in blockchain systems. In *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 0916–0921. IEEE, 2019. 1, 4, 14
- [10] Reshma Kamath. Food traceability on blockchain: Walmart’s pork and mango pilots with ibm. *The Journal of the British Blockchain Association*, 1(1):3712, 2018. 4
- [11] Jia Kan, Shangzhe Chen, and Xin Huang. Improve blockchain performance using graph data structure and parallel mining. In *2018 1st IEEE International Conference on Hot Information-Centric Networking (HotICN)*, pages 173–178. IEEE, 2018. 10, 14
- [12] Shi-Syun Kuo and Wei-Tsung Su. A blockchain-indexed storage supporting scalable data integrity in supply chain traceability. In *2020 IEEE International Conference on Smart Internet of Things (SmartIoT)*, pages 348–349. IEEE, 2020. 12
- [13] Sidra Malik, Volkan Dedeoglu, Salil S Kanhere, and Raja Jurdak. Trustchain: Trust management in blockchain and iot supported supply chains. In *2019 IEEE International Conference on Blockchain (Blockchain)*, pages 184–193. IEEE, 2019. 13

- [14] Satoshi Nakamoto. Bitcoin whitepaper. URL: <https://bitcoin.org/bitcoin.pdf> (ÐÐ°ÑÐ° ÐŸÐ±µ : 17.07.2019), 2008. 2, 8
- [15] Vineet Paliwal, Shalini Chandra, and Suneel Sharma. Blockchain technology for sustainable supply chain management: A systematic literature review and a classification framework. *Sustainability*, 12(18):7638, 2020. 13
- [16] Stephen J Redmond and Conor Heneghan. A method for initialising the k-means clustering algorithm using kd-trees. *Pattern recognition letters*, 28(8):965–973, 2007. 10, 17
- [17] PwC series. Growth of blockchain. <https://www.pwc.com/gx/en/news-room/press-releases/2020/blockchain-boost-global-economy-track-trace-trust.html>, May 2016. 2
- [18] Takahiko Shintani and Masaru Kitsuregawa. Parallel mining algorithms for generalized association rules with classification hierarchy. In *Proceedings of the 1998 ACM SIGMOD international conference on Management of data*, pages 25–36, 1998. 14
- [19] Tobias Sund and Claes Löf. Performance evaluation of a blockchain-based traceability system: A case study at ikea, 2019. 12
- [20] supply chain management diploma. supply chain management. <https://aims.education/study-online/what-is-supply-chain-management-definition/>, May 2015. x, 3
- [21] Hanqing Wu, Jiannong Cao, Yanni Yang, Cheung Leong Tung, Shan Jiang, Bin Tang, Yang Liu, Xiaoqing Wang, and Yuming Deng. Data management in supply chain using blockchain: Challenges and a case study. In *2019 28th International Conference on Computer Communication and Networks (ICCCN)*, pages 1–8. IEEE, 2019. 6, 13
- [22] Soha Yousuf and Davor Svetinovic. Blockchain technology in supply chain management: Preliminary study. In *2019 Sixth International Conference on*

Internet of Things: Systems, Management and Security (IOTSMS), pages 537–538. IEEE, 2019. 13

- [23] Jian Zhang, C Thomas, P FragaLamas, and TM Fernández-Caramés. Deploying blockchain technology in the supply chain. In *Computer Security Threats*. IntechOpen, 2019. 13

Appendices

Appendix A

Demo supply chain based blockchain

Here we build a demo supply chain based blockchain in python 3.0 to show how total supply chain management system works on the blockchain.

```
Enter the number of manufacturers: 2
```

```
The manufacturer keys have been generated.
```

```
Enter the number of stakeholders: 1
```

```
The stakeholder keys have been generated.
```

```
The genesis block is being created.
```

```
The required hash value is:
```

```
000101ecae504c7bf16d341f7c582806b61246b5686308a4ebceb560a8293c00
```

```
The PoW number is: 863908
```

```
The total time taken is: 1.6568260192871094
```

```
Welcome to the supply blockchain.
```

```
The following options are available to the user:
```

1. blockchain view .
2. Enter a tx.
3. View the UTXO.

4. Mine a block.
5. Verify the blockchain.
6. Generate RSA keys.
7. Track the items.
8. Exit.

Enter your choice: 2

Select type of key (M/O) for supplier: M

There are a total of 2 users. Enter your selection: 1

Select type of key (M/O) for receiver: 0

There are a total of 1 users. Enter your selection: 1

Enter the ID of the tracked item: mango

The following options are available to the user:

1. blockchain view .
2. Enter a tx.
3. View the UTXO.
4. Mine a block.
5. Verify the blockchain.
6. Generate RSA keys.
7. Track the items.
8. Exit.

Enter your choice: 4

The number of selected transactions for the block is: 0

No transactions have been selected and therefore no block has been added!

The following options are available to the user:

1. blockchain view .

2. Enter a tx.
3. View the UTXO.
4. Mine a block.
5. Verify the blockchain.
6. Generate RSA keys.
7. Track the items.
8. Exit.

Enter your choice: 3

The list of UTXO are:

---BEGIN PUBLIC KEY---

MIGfMAOGCSqGSIb3DQEBAQUAA4GNADCBiQKBgQC34h184QK3VbozvZMXidLOyeAL
 pC8OpSOWzOHYYKuBES7U05eibptd+ntsTWent+50n131TyexcwzxGgwyqyHe6QKB
 aLuj8k+i9kGPPVDmSgMUbdQQ9zuMVk+atKfZEn5uhTV6wG6f3p2/wB2mBvvBcOw1
 cu8MQ9hcuYHksRCOgwIDAQAB

---END PUBLIC KEY---

---BEGIN PUBLIC KEY---

MIGfMAOGCSqGSIb3DQEBAQUAA4GNADCBiQKBgQCcv+ueOWlSkSN9d0kftkyFP2KL
 T5GR0rWyKGlxMWOr2Ci/Pg/K/uR9OHSeE3GeTvuI3jvUtSRDsZS6JTLd7gRLjLed
 G5YqdrcZKwcvGHW5sFBsS0uurxbcBsGdqXG2/W9kLbGwcXli9ZwKxUy4Fan3SGbN
 quCt6+AOWHLJ6YTpBQIDAQAB

---END PUBLIC KEY---

mango

2021-07-08 02:52:37.922955

b'\textbackslash{}x0c\textbackslash{}x1ero\textbackslash{}x16>\textbackslash{}xf1\textbackslash{}t
 \textbackslash{}xc08@\textbackslash{}xd2\textbackslash{}xa4\textbackslash{}xe7[\textbackslash{}t
 \textbackslash{}xc0\textbackslash{}xed\textbackslash{}x19\textbackslash{}xfd\textba

93\textbackslash{}xd3\textbackslash{}x02\textbackslash{}x17\textbackslash{}xa9>Hi\t
4P\&q\textbackslash{}x10u\textbackslash{}xa7=\textbackslash{}xe0Yj\textbackslash{}x

The following options are available to the user:

1. blockchain view .
2. Enter a tx.
3. View the UTXO.
4. Mine a block.
5. Verify the blockchain.
6. Generate RSA keys.
7. Track the items.
8. Exit.

Enter your choice: 4

The number of selected transactions for the block is: 1

The sign verification for transaction \#1 was true!

The item code was not found on blockchain. Checking for manufacturer
credentials.

The new item has been added under the manufacturer.

The required hash value is:

000b8c2a853e41580b1c8cdfb751a678ec1e3f3b20b0bbb290a318aceeec5f00

The PoW number is: 912226

The total time taken is: 0.08472442626953125

The following options are available to the user:

1. blockchain view .

2. Enter a tx.
3. View the UTXO.
4. Mine a block.
5. Verify the blockchain.
6. Generate RSA keys.
7. Track the items.
8. Exit.

Enter your choice: 5

For the block \#1:

The item ID is mango and the associated timestamp is 2021-07-08 02:52:37.922955

The hash values have been verified.

The PoW number is 912226 and the associated hash is

000b8c2a853e41580b1c8cdfb751a678ec1e3f3b20b0bbb290a318aceeec5f00

The following options are available to the user:

1. blockchain view .
2. Enter a tx.
3. View the UTXO.
4. Mine a block.
5. Verify the blockchain.
6. Generate RSA keys.
7. Track the items.
8. Exit.

Enter your choice: 6

Enter the number of manufacturers: 2

The manufacturer keys have been generated.

Enter the number of stakeholders: 1

The stakeholder keys have been generated.

The following options are available to the user:

1. blockchain view .
2. Enter a tx.
3. View the UTXO.
4. Mine a block.
5. Verify the blockchain.
6. Generate RSA keys.
7. Track the items.
8. Exit.

Enter your choice: 7

Enter the item code: mango

The item (mango) has been found and the tracking details are:

Manufacturer \#1 transferred the asset to Stakeholder \#1 at 2021-07-08
02:52:37.922955

The following options are available to the user:

1. blockchain view .
2. Enter a tx.
3. View the UTXO.
4. Mine a block.
5. Verify the blockchain.
6. Generate RSA keys.
7. Track the items.
8. Exit.

Enter your choice: 1

The list of blocks are:

```
-----  
-----  
0  
2021-07-08 02:52:03.218542  
GENESIS BLOCK  
863908  
b033d79a398267d1887ec2481d8f9a504d742fc59a59d9004bff7348cf7fcef1  
0
```

```
-----  
-----  
1  
2021-07-08 02:53:25.027106  
[<\\\_main\\\_\\.Transaction object at 0x000002610C1DE9C8>]  
912226  
cc6f2b26f47b477a04d7ed204fa168d9885c036ebcf9a56722cecbbcb93bdea4c  
b033d79a398267d1887ec2481d8f9a504d742fc59a59d9004bff7348cf7fcef1  
-----  
-----
```

The following options are available to the user:

1. blockchain view .
2. Enter a tx.
3. View the UTXO.
4. Mine a block.
5. Verify the blockchain.

6. Generate RSA keys.

7. Track the items.

8. Exit.

Enter your choice: 6

Enter the number of manufacturers: 3

The manufacturer keys have been generated.

Enter the number of stakeholders: 4

The stakeholder keys have been generated.

The following options are available to the user:

1. blockchain view .

2. Enter a tx.

3. View the UTXO.

4. Mine a block.

5. Verify the blockchain.

6. Generate RSA keys.

7. Track the items.

8. Exit.

Enter your choice: 2

Select type of key (M/O) for supplier: M

There are a total of 7 users. Enter your selection: 3

Select type of key (M/O) for receiver: 0

There are a total of 6 users. Enter your selection: 3

Enter the ID of the tracked item: litchi

The following options are available to the user:

1. blockchain view .
2. Enter a tx.
3. View the UTXO.
4. Mine a block.
5. Verify the blockchain.
6. Generate RSA keys.
7. Track the items.
8. Exit.

Enter your choice: 3

The list of UTXO are:

```
-----  
-----  
---BEGIN PUBLIC KEY---  
MIGfMA0GCSqGSIb3DQEBAQUAA4GNADCBiQKBgQDLs1Iauvv7YW0mXoIb7BphFnbk  
/kB9LymvcSbTDkrjmJdvDNULWAh7x1ZNjpVSnhoQREvdYnQPSBN2dnkMyOUqyA8P  
8qFGCDAqaANlgo8IyalZtn0JFUfX4XyCZ6NoIhyn0tUjON8ykvivst8f0giiGIQU  
yd96yKYMJFmJ7jVzdQIDAQAB  
---END PUBLIC KEY---  
---BEGIN PUBLIC KEY---  
MIGfMA0GCSqGSIb3DQEBAQUAA4GNADCBiQKBgQCnyz2yVJfLtuF5rDuwJeGr0l/n  
Pp1cWjdHYguoeTD94eqXJXsNGjGT+MMTIdERfYmvxXHhrvJBTJzRuG90dHXAa86b  
1rBH+LAYussgoB3aCSksXBmsZaj/1JkmENj3v0tE9ajMdaC1eDnqBFYWFAPwb6tK  
nMngOAehxIrFjCbh+QIDAQAB  
---END PUBLIC KEY---  
litchi
```


2021-07-08 02:55:32.525176

```
b'@K\textbackslash{x86\textbackslash{x90\textbackslash{xf3x\textbackslash{x9f!\  
;;\textbackslash{xbd|\textbackslash{x98\textbackslash{xd3\textbackslash{xc5\tex  
e\textbackslash{x91\textbackslash{x8bqt\textbackslash{x98\textbackslash{xa9Ek\t  
t\textbackslash{xb3m?M\textbackslash{x05\textbackslash{xfa\textbackslash{xa5"+\  
e\textbackslash{xbe!wq\textbackslash{xd6\textbackslash{xed\textbackslash{xd2'
```

```
-----  
-----
```

The following options are available to the user:

1. blockchain view .
2. Enter a tx.
3. View the UTXO.
4. Mine a block.
5. Verify the blockchain.
6. Generate RSA keys.
7. Track the items.
8. Exit.

Enter your choice: 4

The number of selected transactions for the block is: 1

The sign verification for transaction \#1 was true!

The item code was not found on blockchain. Checking for manufacturer credentials.

The new item has been added under the manufacturer.

The required hash value is:

0003cd640a4b3060a06f7bacb2813e0fa8dea8237abf0852a7fa5d960a319900

The PoW number is: 2084098

The total time taken is: 2.41471791267395

The following options are available to the user:

1. blockchain view .
2. Enter a tx.
3. View the UTXO.
4. Mine a block.
5. Verify the blockchain.
6. Generate RSA keys.
7. Track the items.
8. Exit.

Enter your choice: 5

For the block \#1:

The item ID is mango and the associated timestamp is 2021-07-08 02:52:37.922955

The hash values have been verified.

The PoW number is 912226 and the associated hash is

000b8c2a853e41580b1c8cdfb751a678ec1e3f3b20b0bbb290a318aceeec5f00

For the block \#1:

The item ID is litchi and the associated timestamp is 2021-07-08 02:55:32.525176

The PoW number is 2084098 and the associated hash is

0003cd640a4b3060a06f7bacb2813e0fa8dea8237abf0852a7fa5d960a319900

The following options are available to the user:

1. blockchain view .
2. Enter a tx.
3. View the UTXO.

4. Mine a block.
5. Verify the blockchain.
6. Generate RSA keys.
7. Track the items.
8. Exit.

Enter your choice: 7

Enter the item code: litchi

The item (litchi) has been found and the tracking details are:

Manufacturer \#3 transferred the asset to Stakeholder \#3 at 2021-07-08
02:55:32.525176

The following options are available to the user:

1. blockchain view .
2. Enter a tx.
3. View the UTXO.
4. Mine a block.
5. Verify the blockchain.
6. Generate RSA keys.
7. Track the items.
8. Exit.

Enter your choice: 1

The list of blocks are:

0

2021-07-08 02:52:03.218542

GENESIS BLOCK

863908

b033d79a398267d1887ec2481d8f9a504d742fc59a59d9004bff7348cf7fcef1

0

1

2021-07-08 02:53:25.027106

[<__main__.Transaction object at 0x000002610C1DE9C8>]

912226

cc6f2b26f47b477a04d7ed204fa168d9885c036ebcf9a56722cecbcb93bdea4c

b033d79a398267d1887ec2481d8f9a504d742fc59a59d9004bff7348cf7fcef1

2

2021-07-08 02:56:08.364508

[<__main__.Transaction object at 0x000002610C21EAC8>]

2084098

fe2f016f286d11675d683c005072399a096e67ec6ca585c9490218c1b65caf46

cc6f2b26f47b477a04d7ed204fa168d9885c036ebcf9a56722cecbcb93bdea4c

