# Scalable Storage Scheme for Blockchain-Enabled IoT Equipped Food Supply Chains

Janitha Pranath Rupasena

(169230G)

Degree of Master of Science

Department of Electronic and Telecommunication Engineering

University of Moratuwa
Moratuwa, Sri Lanka

# Abstract

Blockchain is an innovative technology which enabled new applications for solving numerous problems in distributed environments such as the internet of things (IoT) equipped food supply chains (FSCs). In FSCs, the large volume of IoT data such as audio, video, images, and sensor data will be transferred to ensure the traceability of the food to its source. When blockchain technology is used in an FSC, the storage requirements in the nodes will grow with time, since blockchain only allows information adding, without deleting the already stored information. Therefore, it is evident that offchain storage offers more flexibility than onchain storage in IoT equipped supply chains. This work proposes a scalable storage scheme using offchain storage, where the data from IoT devices in the supply chain will store offchain. To reduce the growth of offchain storage, we exploit the fact that some information regarding a particular food item may not be required after the expiration date. The scalability of the proposed scheme is validated through numerical and experimental results.

**Keywords:**
blockchain, internet of things, food supply chain, offchain storage.

# Declaration

I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

| | |
|---|---|
| Signature | Date |

The above candidate has carried out research for the Masters thesis under my supervision.

| | |
|---|---|
| Signature of the supervisor | Date |

# Acknowledgement

This work was carried out during the years 2019-2021 at the Department of Electronic and Telecommunication Engineering, University of Moratuwa.

I owe my deepest gratitude to my supervisors Dr. Madhusanka Liyanage and Dr. Kasun Hemachandra for their incredible guidance to successfully complete the research. I would like to thank Mr.Tharaka Hewa for his valuable contribution as an instructor. A conference paper have been published based on the research outcomes and I would like to thank all of the coauthors for their immense support.

I would like to thank all the authors and owners of the materials that have been referred in this thesis and finally, I would like to thank all the people who supported to successfully complete the research.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Supply Chains (SCs) play a major role in transferring goods or services from the initial supplier to the end customer [2]. To gain end consumers' trust, SC authorities need to maintain the quality, integrity, and credibility of the entire SC. In a food supply chain, efficient monitoring of the food products in each stage of the SC is crucial to ensure food safety. As consumers and regulatory bodies are becoming increasingly concerned about food quality, the concept of traceability in the SC has become a critical requirement.

The use of internet of things (IoT) devices and technologies in food supply chains (FSCs) has emerged recently, leading to substantial innovation and research towards developing auditable, transparent, reliable tracing systems for FSCs. Conventional IoT-based tracking are built on top of centralized infrastructures, leading to limitations such as data integrity, tampering, and single point of failures [3]. However, in FSCs, to maintain the integrity of the products along the SC, it is essential for the stored records to be tamper-proof. Therefore, it is preferred not to rely on any centralized third-party intermediary. As a potential solution for the issues related to centralized architectures, blockchain technology, which is a peer-to-peer digital ledger that does not depend on centralized services, has been considered in recent works related to SCs [4]. In the case of blockchain, all the records are based on a consensus reached at least by the absolute majority of the peers within the network itself. This decentralized ledger is immutable by design and provides a source of information which auditable as well as transparent. Therefore, blockchain can be effectively incorporated into SCs to maintain transparency and immutability in a distributed manner. Furthermore, from an IoT perspective, the sensors in a blockchain-based FSC would only require stable connection to their closely located peer, eliminating the need for connection with a centralized cloud. Thus, it is evident that blockchain can be effectively used in IoT equipped FSCs to make traceable data available at every step of the SC [5].

When a large number of IoT devices are used in an FSC, the amount of transaction data that will be stored in the blockchain will also be extremely large. This data may

include audio, video, sensor data, and images. Therefore, the blockchain ledger storage will grow continuously, requiring large storage volume on each node, leading to higher costs. This may significantly affect the scalability of the SC. To circumvent this, offchain storage models have been proposed to store the incoming transaction data. Then only the hash of the corresponding data, which is of significantly smaller size compared to the original data, could be stored in the blockchain. Smart contracts as well as distributed hash tables (DHTs) can be used to establish a link between hash found in the chain and the actual physical storage [6]. A distributed file system among the already available nodes is used to maintain the offchain storage [2].

Since IoT data will be stored in an offchain storage, the offchian storage will also grow with time. However, it is well understood that traceable data of a particular food product may not be required after a certain period of time (eg. the expiration date). Therefore, in this thesis, we exploit this fact and propose a technique to reduce the growth rate of the offchain storage in a blockchain enabled IoT equipped FSC. To the best of our knowledge, this is the first work providing an end to end solution for IoT equipped FSCs with a recyclable offchain storage mechanism. The proposed solution uses smart contracts, distributed file systems with distributed hash tables, and distributed database to assure an efficient, secure, and trusted environment. We validate our claims using numerical results and evaluate the proposed system with a prototype implementation using Hyperledger blockchain platform.

The remainder of this thesis is organized as follows. Chapter 2 presents the literature survey and related work. Chapter 3 presents the proposed methodology and chapter 4 presents numerical results that derived in the research. Furthermore, Chapter 5 presents experimental results obtained through extensive implementations to validate the numerical results. Chapter 6 presents the discussion and the future work. Chapter 7 concludes the thesis.

# Publications

[1] J. Rupasena, T. Hewa, K. T. Hemachandra, and M. Liyanage, "Scalable storage scheme for blockchain-enabled iot equipped food supply chains," in *2021 Joint European Conference on Networks and Communications 6G Summit (EuCNC/6G Summit)*, pp. 300–305, 2021.

# Chapter 2

# Literature Survey

## 2.1 Blockchain Technology

### 2.1.1 Defining Blockchain

At the beginning of 2009, the digital currency was not a novel idea, especially within the coding community. But everything changed in January 2009, when Satoshi Nakamoto decided to surprise a chosen group of people with an interesting email. With this single act, Satoshi Nakamoto took the first step in ushering in the paradigm of decentralized cryptocurrency that would change the world of digital currency forever. Before Nakamoto's idea, any value exchange over the internet had to go through a trusted third party. However, his idea of a distributed system removed the need for trust between the involved participants [7].

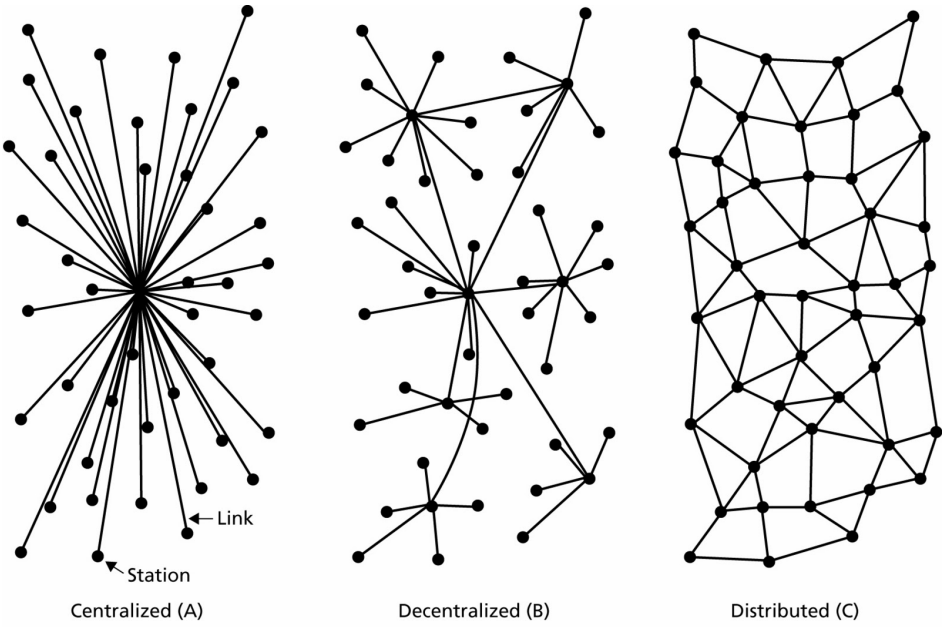


Figure 2.1: Different Kinds of Network Types [1]

There are three different types of networks, as displayed in Figure 2.1. Regular centralized networks are normally controlled by a single central node and hold the sole responsibility of operating the network smoothly while keeping the other members in the network happy. There are several advantages that come with a centralized network. For example, it gives better security, convenient management of resources and creates the basis for trust within the network. The biggest disadvantage with a centralized network is its risk of a failure happening at the central location, which would definitely bring down the whole network [8]. However, in a decentralized network, the control would be spread among several nodes, and the data would not need to go through a central point. Decentralized networks also provide better privacy, and it is difficult to trace data across the network. One of the best examples for a decentralized network would be the internet [9]. In a distributed network, all the information are spread out and shared within the entire network. In such a network, several nodes in the network are used to store segments of the database, and members can easily approach the data by referring to the database at their location. This allows them to use the data without interfering with the works of others. Also, attacks like Distributed Denial of Service (DDoS) are impossible due to the unavailability of a single point of data storage. Although there are multiple ways of carrying out a DDoS attack, the most common way is by targeting a single central point in the network with a large number of data transactions to crash the network. Since there is no central point in storing data in a distributed network, such a DDoS attack would be futile. A major drawback of a distributed network is its lack of trust. As there is no centralized controlling figure, maintaining the trust would be difficult. [8]

In 2009, Satoshi Nakamoto solved this quandary of peer to peer trust. He did it with his invention of Bitcoin currency and especially with the underlying blockchain. Till this invention, people had to place their faith in financial institutions to operate their value transactions, information, asset ledgers and identity verifications. Blockchain is simply a ledger of facts. Transactions data are not stored on only one central network with a common control entity, but it is shared among all the members within the network. Data is kept in blocks within the blockchain network that captures the information. Every block receives a hash value and a timestamp. As each block is linked to the block that came before, it is nearly impossible to make changes to the data as one would have to change all the blocks in the chain [10]. All the members within the network have equal power. They are all commonly known as nodes. Some of the network nodes function as miners whose contribution is adding new transaction data to the blockchain by connecting blocks together and assigning hash values to them. All the network nodes automatically verify any change that happens to any of the other network nodes. The meaning of this is that a transaction is considered as valid only when the particular block that it belongs is added to the blockchain and validated

by the other network nodes [10]. These miners are the workhorses of the blockchain and help to provide a safe and verified network. Mining computers consume a lot of time and energy for their work. For their contribution, they are usually awarded a small amount of cryptocurrency [8]. Figure 2.2 is a graphical explanation of how blockchain works.

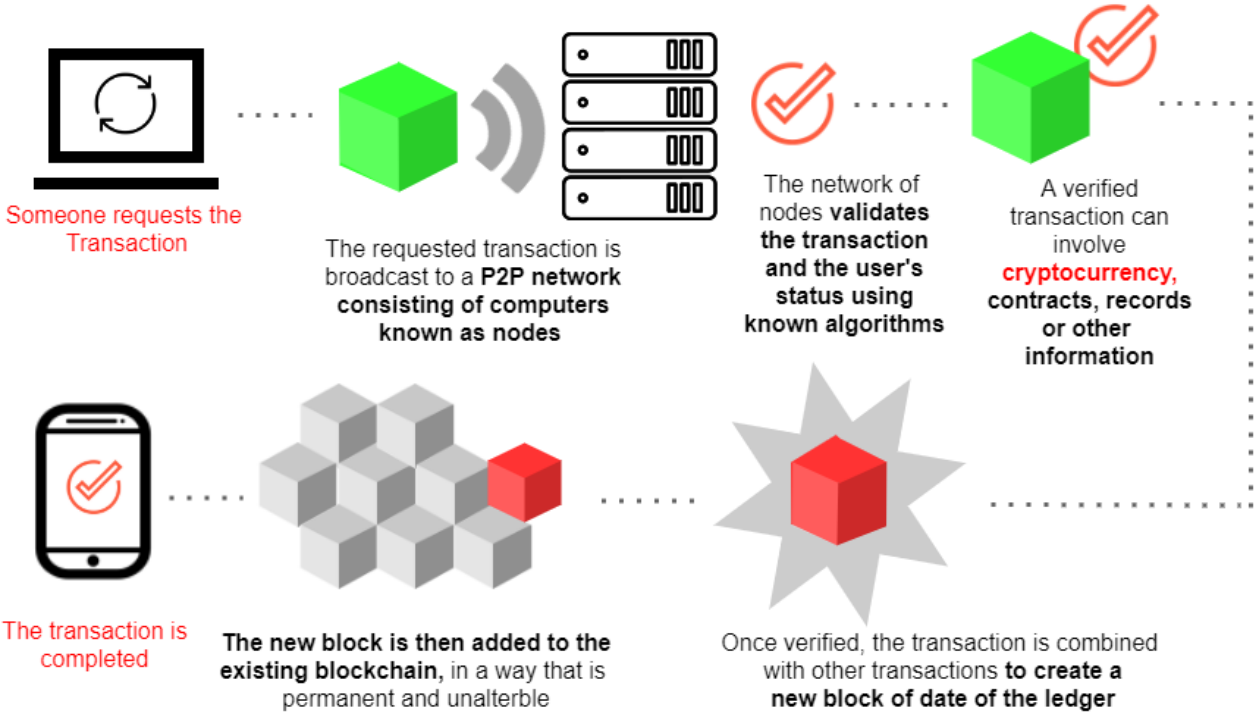


Figure 2.2: A visual explanation of how a blockchain work

### 2.1.2 Key features of Blockchain Technology

Blockchain technology is still practised as a proof of concept in many industries. Therefore it doesn't have an accepted architectural approach. However, the design goal for a vast number of blockchain use cases is based on the core value proposition of achieving immutable, trustless and efficient operations. An overview of the functional components and the key features found in blockchain technology are as follows.

#### Block

In the blockchain, block is the main data storage entity. It stores and preserves the data pertaining to many transactions. These blocks are linked together by storing the hash of the previous block inside the present block, making an enclosed circle of blocks in the public ledger. Typically, there are two main parts in a block, a header and the transactions. The header stores the metadata, which contains all the relevant details about the block [11].

Figure 2.3 shows the underlying structure of a block and shows how blocks are chained together using the block header metadata mentioned as follows.

- Version number: a 4-bytes field to describe the block version number.
- Previous block hash: a 32-bytes field to indicate the previous block hash of the blockchain. It is working as a pointer between the previous block and the current block in the ledger.
- Timestamp: a 4-bytes field to display the actual block generation time.
- Merkle tree: a 32-bytes field to display the combined hash (SHA-256) of all transactions related to the particular block.
- Difficulty target: a 4-bytes field to display the block difficulty target.
- Nonce: a 4-bytes field to create the block and calculate the different hashes.

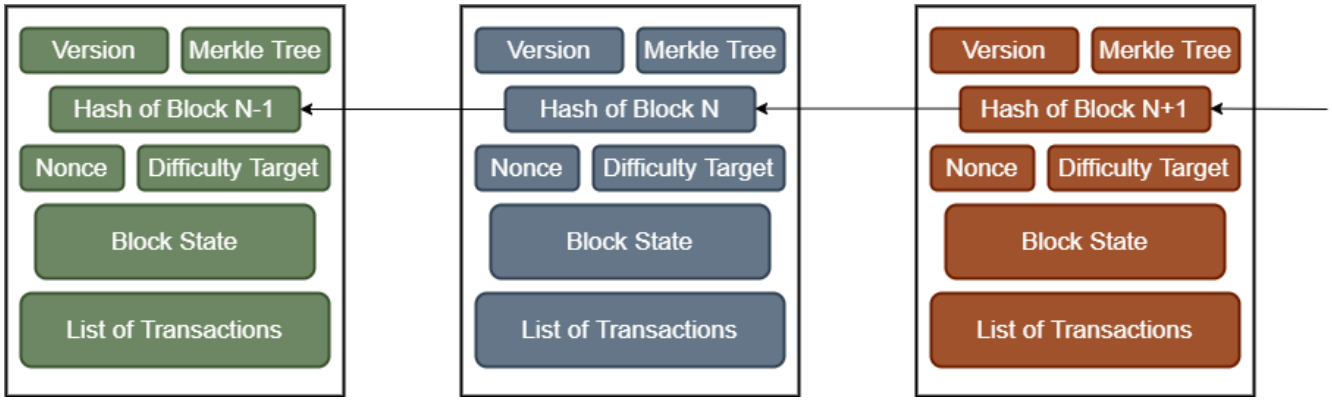


Figure 2.3: Data structure of blocks in a blockchain

If anyone attempts to tamper with the transaction records stored in a block, the hash for that specific block would be changed, and it would no longer point to the previous block that is already verified. This makes the blockchain easy to audit in near real time and makes it difficult to alter the records without the consensus in the network [12].

### Public blockchains and Private blockchains

Bitcoin was designed to be fully decentralized and permissionless, allowing anyone in the network to contribute to the proof of work mining operation and access the copy of the ledger. However, this method of control was not efficient as it did not reach the desired economies of scale in processing. There was a high risk of a single group of miners forming a majority in consensus, which would threaten the integrity of the trustless consensus process.

Private or permissioned blockchains are built with a different value proposition than the aforementioned Bitcoin or public blockchain. Private blockchains can be quite useful in cases where the integrity of the audit trail is not of utmost importance, and there is a requirement to standardize the data sharing across industries for new market structures and efficient operations [12].

In a permissionless or public blockchain, anybody can contribute in any capacity. It allows anybody to approve new blocks during the mining operation, whereas in a permissioned or private blockchain, only selected entities can take part in the network. This restricts the mining or processing to only designated members or organizations in the network. Even though there is the benefit of zero cost verification, the cost of networking is still high as the private blockchain still has to depend on the trusted parties [12].

Among these private and public blockchains, there are federated blockchains, which instead of having a single or no central organization, might have several centralized organizations in control. The writing and reading permissions might be specified only to a certain set of nodes, and this approach is a combination between the public blockchains and the private blockchains in a single organization [13].

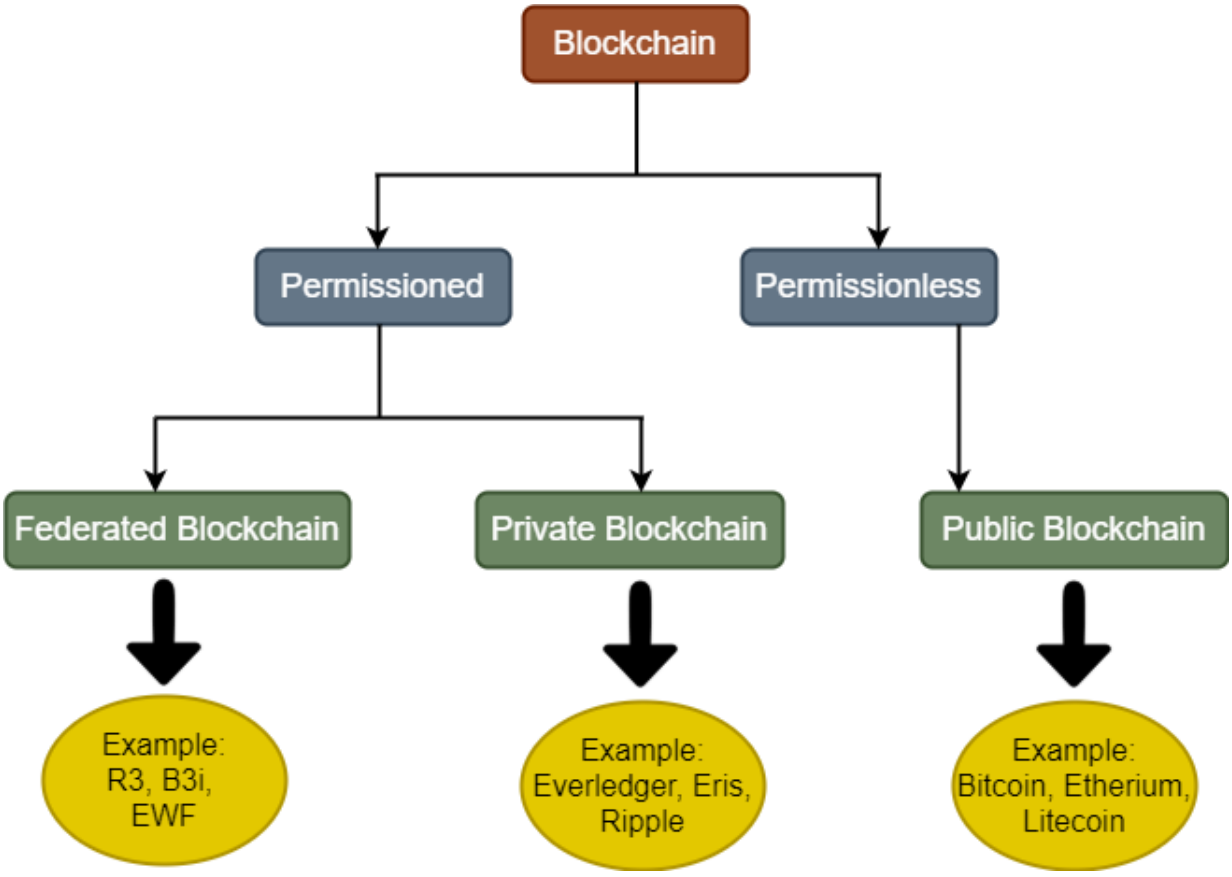


Figure 2.4: Types of blockchains

Every type of blockchain has its own advantages and disadvantages. A public blockchain

has the significant disadvantage of having to draw a lot of computing power to hold and maintain the distributed ledger. It also has the problem of the electricity, which requires to operate each of the transactions that increases with every new network node that joins. Compared to a private blockchain, a public blockchain is more wasteful and slow. However, they are still cost efficient and fast in comparison to the existing financial systems which are used currently. Also, public blockchains are highly secured. It is very difficult to alter the transaction data or double spend the money, and it is completely anonymous. The use of private blockchains is more suited for traditional business and governance models. This means that private blockchains would not revolutionize the world, yet they can make things much better. For instance, governments can work with private blockchain voting polls and at the same time save billions as voting would become fully resistant to corruption. Voting would be truly secure. There are several other examples like this where public blockchains can positively contribute. But it is unlikely that governments and businesses would leap at any system that takes away their decentralized control and security [14].

Table 2.1: Differences between public and private blockchain.

| Item | Public Blockchain | Private Blockchain |
|---|---|---|
| Access | Read and write for anyone | Read and write for a single organization |
| Speed | Slow | Light and Fast |
| Efficiency | Low | High |
| Security | Proof of stake, proof of work and other mechanisms of consensus | Preapproved participants |
| Immutability | Nearly impossible to tamper | Could be tampered |
| Consensus process | Permissionless and anonymous | Permissioned and known identities |
| Asset | Native Asset | Any Asset |
| Network | Decentralized | Partially decentralized |

### Consensus protocol

A major difficulty with blockchains with public ledgers is ensuring the achievement of the consensus protocol by the entire set of peer-to-peer participants [15]. A consensus protocol is put in place to make sure that the participants would follow the network rules and the transactions are validated in the correct order. It is also used to verify the information within a block and that the miners are compensated fairly, and avoid issues like the double spending problem. Arguably, the most used consensus protocol is the proof of work protocol which is used in the Bitcoin network. There are other options, such as the one used by another

cryptocurrency called Blackcoin, which is called proof of stake. All consensus protocols cannot guarantee the maintenance of the right order of transactions within the network. They are threatened by cases of arbitrary concurrent failure (a type of Byzantine fault) of multiple nodes. A Practical Byzantine Fault Tolerance (PBFT) consensus protocol can be implemented to reduce this problem.

- **Hashing**

Some algorithms use cryptographic hash functions in order to perform a consensus protocol. A hash function can be considered as a mathematical operation that takes any data size as input and gives an output of fixed data size. It is irrelevant if the input is a single word or an entire novel, the output will still be the same. But it is almost impossible to get the original input data by reversing the output code. Therefore, it is used to store things like passwords. In some blockchain technologies, like the Proof of Work method, hashing mechanisms are utilized to add new transaction data into the blockchain through the mining process. Each block consists a unique hash value for every transaction, the entire set of transactions within the block bundled together that contains a unique hash value and time stamp into a block, and all the blocks are connected to their parent block's unique hash value [16].

In [17], it explains the whole process of the hashing, where a miner's software would take the hash value of an existing block, containing all the underlying old data, and would combine it with the next unhashed transaction raw data to make a new hash. The same process would take place when the software identifies a new transaction. It would then merge the second hash value with the next transaction data to create a third hash value. This process will be continued as new transactions get identified up by the miner's process, encapsulating all the data of the transactions into a single hash value and simultaneously generating a completed block that would then act as a segment of the blockchain.

If an outsider were to tamper with a block within the chain, the relevant value of the hash would completely be modified and no longer would match with the rest of the blocks in the blockchain. Therefore to go unnoticed, an attacker would have to change all the interconnected blocks within the blockchain. All network participants also have private keys that are used to make transactions and access their wallets. If a record is changed, this private digital signature would be invalidated and all the other members in the network would immediately realize that something nefarious has happened [18].

- **Proof of Work**

Various types of blockchains could use multiple types of protocols with numerous types of algorithms. The major one would be the Proof of Work (PoW) which is used by Bitcoin as well. All the transactions that happen within a period of about ten minutes are gathered by the network, and then they are placed into a block. The miner nodes in the network compete to generate the block by attempting to solve a PoW puzzle. This puzzle is quite costly, complex, time-consuming and requires a lot of computational power to solve successfully. For a valid and accepted solution, it should hash to a value less than the current target, which is generated by the main algorithm. When a node has more computational capacity, it has a higher chance to find a solution to the puzzle and reap the reward of a small amount of cryptocurrency together with any fees paid by the network nodes doing the transaction. When a miner node solves the puzzle, it gets broadcasted to all the other nodes, and they immediately check that all the rules of protocols have been correctly followed and that the consensus is fulfilled. This newly created block is then accepted and added to the blockchain as the current block. Then with the new hash base, this process would now start over again [17]. The PoW consensus mechanism is shown in 2.5.

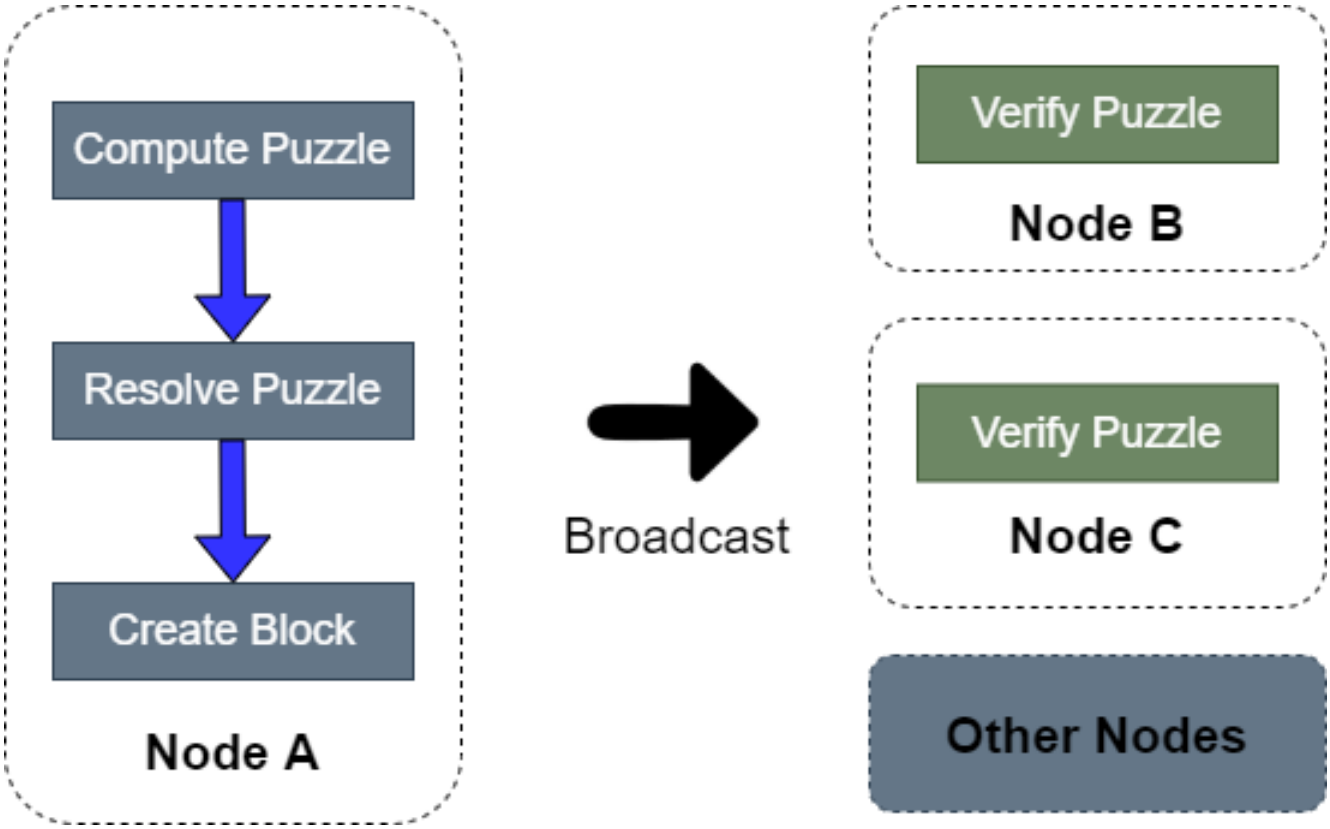


Figure 2.5: PoW consensus mechanism

In some cases, different nodes create blocks simultaneously, which then creates a “fork”

in the blockchain. In such a situation, the block that was generated at the beginning and had the lengthiest chain of earlier blocks receives the credit of being the one true chain, and all the other parts get dismissed from the fork.

PoW is commonly considered as a stable and safe consensus protocol with a robust defence mechanism against regular DDoS attacks and any corruption in the ledger. However, this also contains several drawbacks. The main drawback with this is the consumption of energy. The system intentionally requires more energy with the time-consuming and challenging decryption puzzle. With the increase in the number of miners, this energy consumption issue is getting more worse. The increase in the number of miners also forces the puzzle to be more difficult in order to maintain the block time around the ten minute mark and to limit the amount of cryptocurrency released [19].

With the increase in the amount of miners, miners can form bigger pools of mining, and when large corporate entities are involved in the mining game, it becomes more difficult for the new entrants. The cryptocurrency-based rewards for the miners will start to dwindle as the available amount of cryptocurrency is very limited and is allocated for existing only for a certain period of time meanwhile the miners count rise up. This will increase competition. When the rewards start to reduce with time, nodes would be discouraged to begin mining, thus generating an oligopoly of organizations that runs the entire mining business [19]. In the long run, there will be no cryptocurrency rewards for the miners. The miners will be rewarded only from the transaction fees, which the senders themselves would decide.

As the blockchain concept depends on a distributed consensus approach, it is also vulnerable to a 51% attack. The 51% attack happens when a hashing power of the single miner becomes more than half of the combined hashing power inside the whole network. For example, if such a 51% attack happens, the attacker can revert old transactions and do double spending of coins [20].

- **Proof of Stake**

One of the more suitable alternative approaches to PoW consensus protocol is the Proof of Stake (PoS), which depends on the size of a node's deposit. The bigger the deposit, the bigger the stake. It doesn't relate to a node's computational power. Nodes can add parts or all their deposits inside a "stacking wallet", which help to lock up their cryptocurrency while making the node as a "validator". The validators contribute to generate and vote on blocks. The way it happens rely on the network's type of consensus mechanism. Byzantine Fault Tolerant (BFT) style PoS and chain-based PoS are the two main algorithms that are used. In the BFT style PoS, validators are assigned randomly. They are given the authority to suggest blocks, but the final selection depends on the vote of the other validators. A chain-based PoS would also randomly select a validator within a certain period of time. This

selected validator then receives the authority to generate a block that should still be linked to a previous block. The rewarding schemes for all the PoS mechanisms are the same. The more deposit a node has within its stacking wallet, the bigger the rewards and the chance of getting randomly chosen [21].

PoS has a lot of advantages when compared to PoW. The best one is the saving in the consumption of energy. To validate transactions in PoS, network nodes don't have any requirement in mining currency. Then, there is no requirement for huge computational power. Instead, it is a node's deposit size that matters. As the cost of mining is decreased, it is reasonable to reduce the rewards as well. A second advantage with PoS is the reduction in risk of centralization. PoS also clearly discourage any possible 51% attack as a network node has to posses many cryptocurrency than all of the other network nodes. Most of the people are developing hybrid versions that combine both PoW and PoS to achieve the best of both approaches, but the truth is, it would be challenging to come up with a better consensus algorithm [21].

- **Hyperledger**

The Hyperledger project was started in 2016 by the Linux Foundation. It helps the evolution of blockchain technology across the globe with an open source and a code base distributed ledger. The primary target of this project is creating cross industry collaborations and developing blockchains technologies that help to make a multitude of businesses and industries better. Several different projects based on blockchain are currently happening under the Hyperledger platform. For instance, the Hyperledger Fabric is a permissioned blockchain that runs arbitrary smart contracts and provides strong security features and an open source. Hyperledger was created with the directive of accelerating industry wide cooperation for developing reliable and high performance blockchain and distributed ledger based technology frameworks that could be utilized across a number of industry sectors to enhance the performance, transactions, and efficiency of the various business processes and operations. Hyperledger is a global collaboration that includes prominent businesses and organizations from the fields of Internet of Things (IoT), supply chain management, finance, banking, manufacturing, production and technology [22].

### Smart contracts

A smart contract is defined as a script based program which is kept in the blockchain network. It is a code that is predetermined in which a particular input would always provide the same output and a contract that could encode any list of guidelines and rules performed in the coding language. Also, it can code to be a process of multiple steps and it can hold up fairly difficult contracts. All the entities can examine the contract for themselves and choose

whether to sign it and to trust that the contract will make it through. A smart contract is kept at a specific location in the blockchain and a unique address is given to identify that location. Nodes invoke this smart contract by sending cryptocurrency to that unique address and after that, the consensus protocol takes over and handles the process of the verification [23].

Blockchain technologies keep getting improved every day and the use of smart contracts was implemented in the improved blockchain 2.0. One of the good examples would be Ethereum. It uses the Ethereum Virtual Machine (EVM) to get executed smart contracts by running all the nodes in the blockchain. Although there are several blockchain systems that use smart contracts, Ethereum is the largest one. Ethereum utilizes an open source system that motivates keen developers across the globe to design more advanced smart contracts. Since these smart contracts can communicate with each other, designers can develop and build advanced decentralised applications [20].

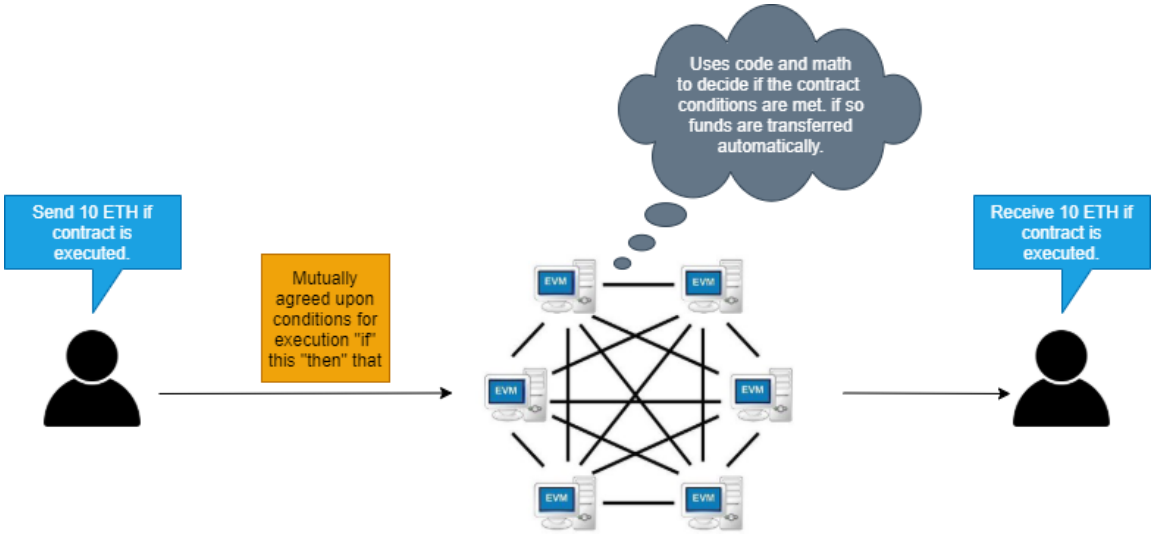


Figure 2.6: An example of how Smart Contract works

### 2.1.3 Characteristics of blockchain

The Blockchain is a decentralized architecture with built-in higher security to enhance the integrity and trust of transactions within the network. The following section gives a summary on main characteristics linked with blockchain, as mentioned in Figure 2.7.

- **Transparency**

Blockchain provides a high level of transparency by distributing transaction details among all the relevant users participated in those transactions. It is not necessary for a third party

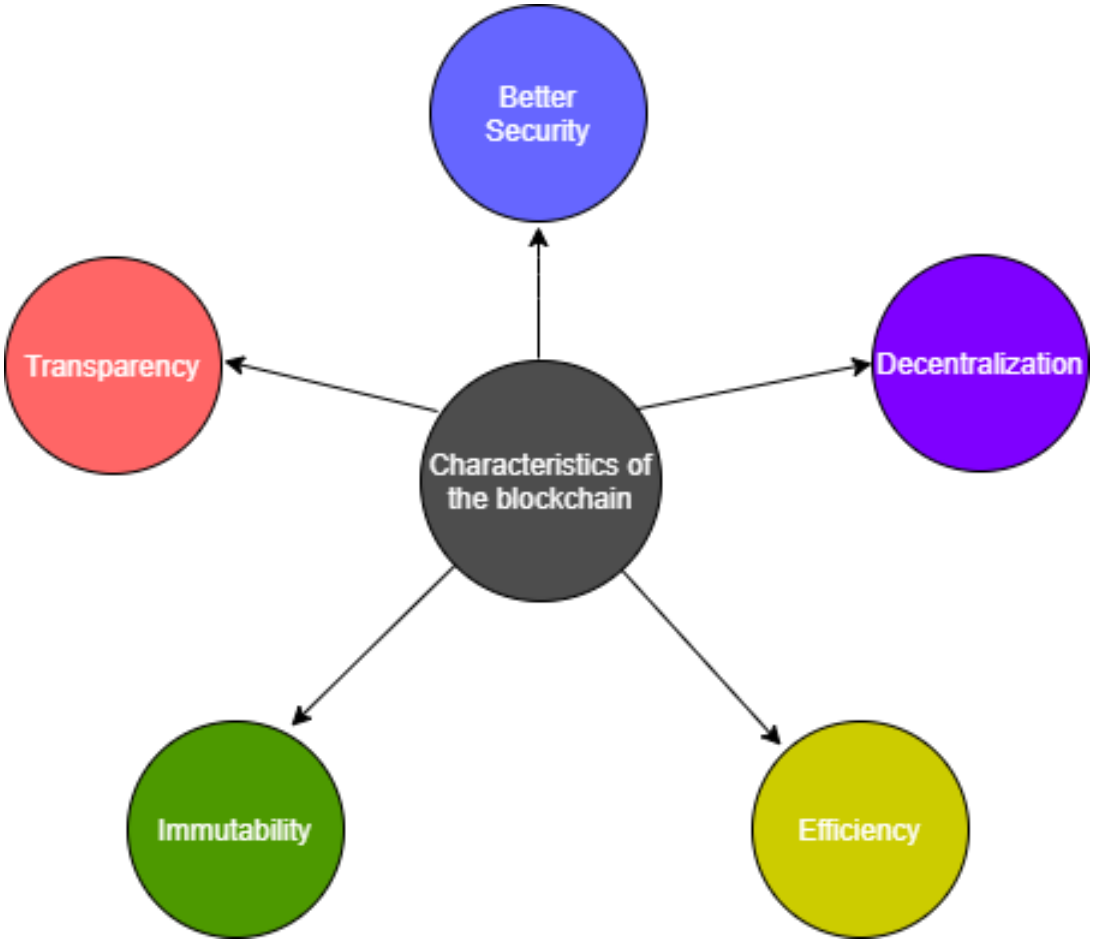


Figure 2.7: Characteristics of the blockchain

which in turn improves business friendliness and guarantees a workflow that can be trusted by everyone in a blockchain environment.

- **Decentralization**

When compared to a centralized architecture with inherent problems, including a single point of failure and scalability, blockchain utilizes a distributed and decentralized ledger to make use of the processing capabilities of all the users that take part in the blockchain network. This factor reduces latency within the network and eliminates the risk of a single point of failure.

- **Immutability**

One of the most important characteristics of the blockchain network is the ability to guarantee the integrity of the transactions by generating and maintaining immutable ledgers. In traditional centralized designs, databases can be manipulated and data can be altered creating the need for third parties to generate trust and guarantee information integrity.

But in blockchain technology, as each block in the distributed ledger links to the previous block creating a chain of blocks, all the blocks are permanently stored and never altered as long as the participating user nodes continue to maintain the whole network [24].

- **Efficiency**

Blockchain enhances upon the traditional centralized architecture by freely distributing the records in the database among the users participating in the blockchain network. It is clearly much more efficient than the traditional centralized design in terms of risk management, settlement speed and cost [25].

- **Security**

Even though security within the system is an important problem for many new technologies, blockchain gives far better security features than most, since it works with a public key infrastructure that guards against nefarious actions to alter the information. The participating users within the blockchain network can wholeheartedly keep their faith in the security features and integrity of the underlying consensus mechanism. Other than that, blockchain gets rid of the risk of a single point of failure that can affect the whole system [26].

## 2.2 Supply Chain and IoT

The main function of a supply chain is moving a service or a product from a manufacturer or a supplier to an end consumer. This process involves a combination of business connections that needs to enhance the efficiency and minimize the unnecessary additional work [27]. To be competitive in the market, organizations need to properly handle their supply chain. By managing the supply chain, organizations can reduce the production cost, enhance and properly coordinate the processes, deliver a better end product to customers and reach their strategic goals. Therefore, most companies monitor their supply chain and modify the flow of human resources, materials, activities and information, which are used in various stages in a process. There are many stages in one supply chain, and the success of one stage does not imply the success of entire supply chain. Therefore, decision alignment has become one of the key components for a successful performance in a supply chain. For a successful decision alignment in supply chain management, real time data has become more important, and IoT can be adopted as it captures real time information through automated identification [28].

IoT can affect the way organizations manage their supply chains between the supplier and the customers. One of the major advantages of the use of IoT is that it can give more specific, accurate, fast and updated information about products and resources. It also can be used to enhance the information sharing, and collaboration capabilities in a supply chain with

the help of IoT devices [29]. IoT technology can also use to enhance integration, efficiency, responsiveness and agility to operations in a supply chain. The most common example for the use of IoT is the Electronic Product Code (EPC), which is an internationally used code system for labelling products. It is applied along with an EPC global network in supply chain applications to assure Radio Frequency Identification (RFID) interoperability. EPC recognizes the items that have EPC tags, and the data stored in these EPC tags are retrieved through the RFID readers that are installed in supply chain settings. Then the members in the supply chain can access and exchange those real time data through the EPC global network. A middle-ware is used to filter the information from the RFID readers as they retrieve much information causing a high load on networks and systems.

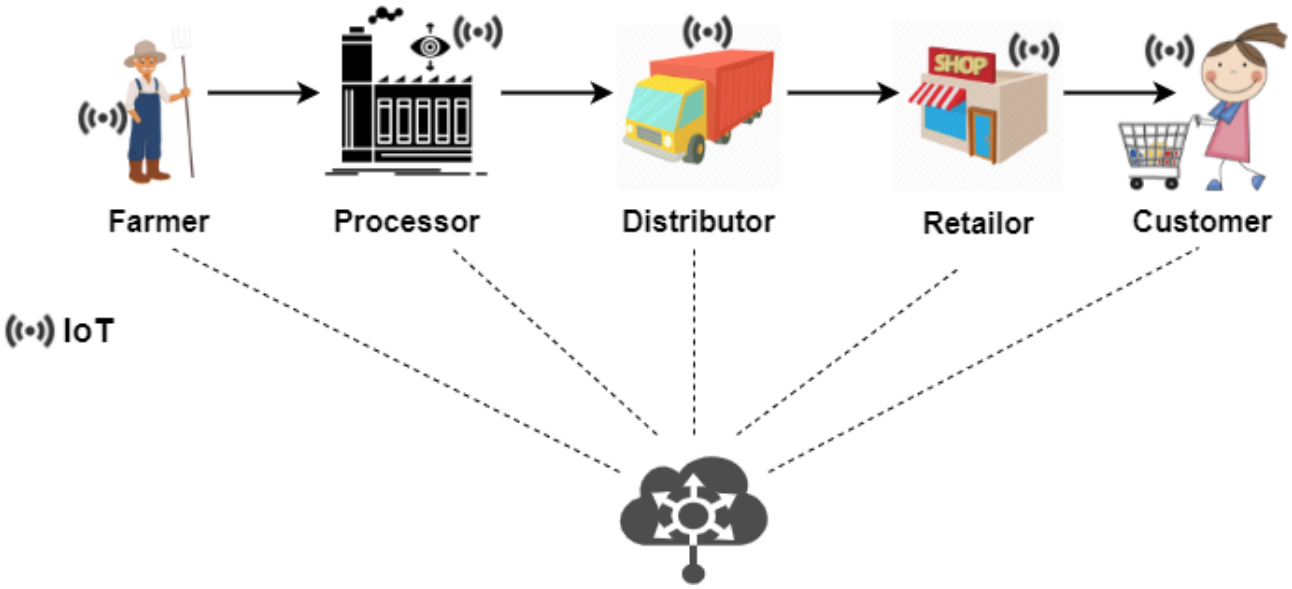


Figure 2.8: Supply chain availability with the use of IoT devices

As mentioned above, RFID retrieves data using a tag, a reader, and radio waves for identification of items in a supply chain [29]. These IoT technologies replace older technologies such as barcodes. Barcodes have been used for nearly 50 years to identify the items. Readability within a short range and unautomated tracking are the disadvantages of barcodes, and these can be eliminated by using RFID tags. At the same time, RFID tags have unique identification codes for various items. The drawback of this technology is that it is more expensive than the barcode system. The cost for a RFID tag mainly comes due to the chip which is used to store data. But the price of this RFID tag is decreasing every year as more and more investments are putting into development on this technology. [30]

IoT has been used for bidirectional communication via sensors in smart supply chains. The embedded sensors gather data and provide them to on-site technicians for decision making, which helps to increase machine time and quality of the customer service. IoT

in the transportation sector can play a key role in the supply chain in future. Visibility of transportation has been increased by the technologies such as RFID and network cloud-based Global Positioning System (GPS), which furnish tracking data such as identification of the location of an item shipped from a seller to a buyer. RFID can also be used in supply chains with Near Field Communication (NFC) to monitor various stages in real-time of a process like purchasing of materials, transportation, production, distribution, storage and sales. It may even be able to gather data which are related to products so that organizations can react instantly to the changes in the market. An advantage of this is that businesses can have zero stock as the response time can reduce to few days [31].

Another application of IoT technology is in the fresh food supply chain, as the demand for fresh food among consumers has grown-up throughout the last few years. The traditional food supply chain is mainly facing three problems at present due to the lack of IoT technology in it. The first issue is that the consumers have lesser tolerance for food degradation. Secondly, the foods are not monitored during the transportation process threatening the health of consumers. Thirdly, the costs and the efficiency of an operation are not monitored in real-time. Research has shown that the real-time information which gathers from IoT devices can use to eliminate many of these challenges [27]. The parameters such as humidity, temperature, location and vibration can be monitored using sensor technology during food transportation. These data will help to evaluate the entire supply chain process with an easy execution process [27].

The traditional architecture is mostly fixed in supply chains and does not deal with variations effectively, and this often creates problems when collaborating. The use of IoT technology can be introduced as an effective solution for this, yet there are many obstacles to this also. Modelling and validating inter-organizational and intra-organizational process interactions are the main concerns in developing IoT technology-based supply chain applications, and insufficient architectural information affect the performance of an IoT system [32].

## 2.3 Blockchains in Supply Chains

Members in a supply chain use different platforms for recording their information. This paves the way to have various types of records about the present state among each party. These records can consist of different types of information and mediums such as emails, paper documents, telephone conversations etc. These records are vulnerable as they can get misplaced and people can get mislead with those data. Blockchain is one of the best solutions for this, as every authorized member in a supply chain can access and verify the same information at the same time. Therefore, all the members in a supply chain can agree on the current state without any misunderstanding or arguments. Blockchain can be applied

in the supply chain with distributed ledgers securing the trust of all the transactions being completed. The record of each transaction is time stamped and is linked to the previous event, offering advantages like visibility and optimisation. The records on the blockchain can only access by the authorised parties, and this implies that the records are secured and can be shared among authorised parties at the same time. Tracking and visibility of products can be assured with the aid of sensors, and each party in the supply chain can access data with reduced risk by linking that information with blockchain technology [33].

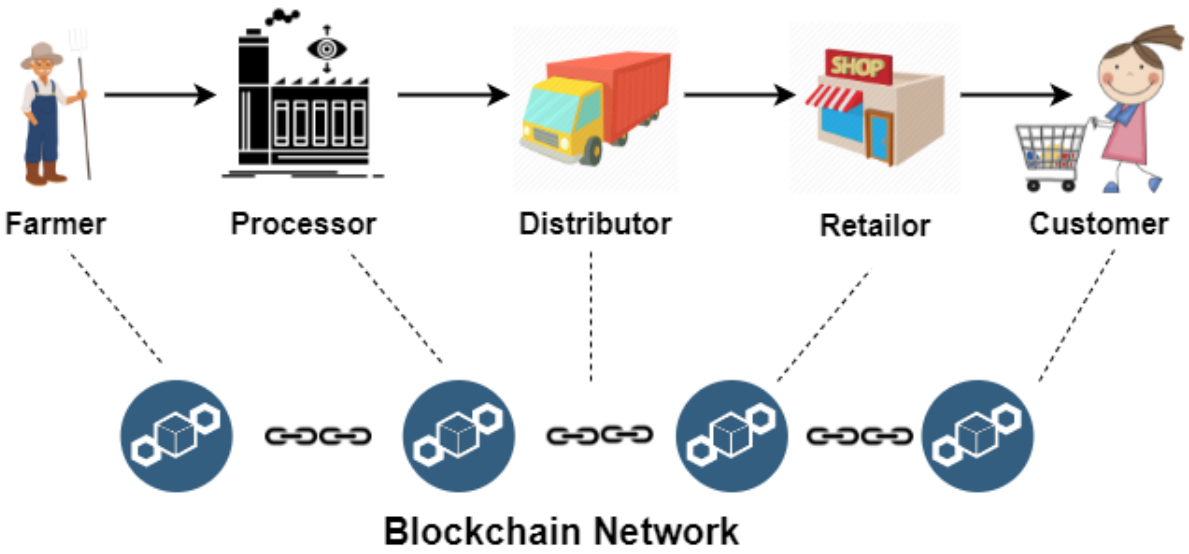


Figure 2.9: Supply chain availability with the use of Blockchain

The disruptive presence and integrity of blockchain with IoT have made it one of the most recent technological trends to be adopted in various supply chain applications, especially in the manufacturing and financial sectors. The combination of technologies allows companies to optimize and improve the efficiencies in the processes leading to strengthening their relationship with business stakeholders, current customers and attracting new customers. For example, it has been found out that businesses in the food and pharmaceutical industries face financial losses due to various supply chain issues. Some of the issues such as fraud, stealing products, gray market and product recall have created a movement of supply chain stakeholders towards more traceable and transparent methods. At this point, blockchain becomes a promising solution as it is immutable, transparent, secure and fault tolerant. [34]

The utilization of blockchain in SCs has been proposed in several works. The authors in [35] proposed a scheme for governance on the drug supply chain via the Gcoin blockchain. The double-spending prevention mechanism in the Gcoin blockchain is exploited to identify counterfeit drugs. Reference [36] proposed a blockchain-based aircraft parts supply chain to improve traceability and to establish the authenticity of spare parts. However, the paper does not discuss the implementation aspects.

## 2.4 Blockchain-Enabled IoT equipped Food Supply Chains

At present, there is a significant increase in the movements of products and information between countries due to the globalization of FSCs and markets [37]. Traditional FSCs are defined by strong vertical integration and coordination among supply chain parties to increase efficiency. As an example, lowering marketing and operating costs and, at the same time, increasing customer needs for food quality and safety [38]. Overall, supply chain transparency and the traceability of food products have become more critical due to the multiple frauds that occur in global FScs. Therefore, FSC parties are under increasing pressure to improve the tracking and tracing capabilities of materials or products from the manufacturer to the consumer, to improve the transparency of their supply chains and to enhance the sharing ability of trusted information among each other [39].

As a solution for the growing food safety issues, many IoT technologies, such as RFID and wireless sensors based hardware and architectures, can be applied to supply chain visibility and traceability. However, one important issue that has not been addressed is that whether the information shared among food supply chain parties in traceable systems can be trusted. This kind of centralized organization can become a vulnerable target for bribery resulting in information asymmetry between the individuals and the organization. For example, the information can be tampered by bribing the administrator and then the whole system will not be able to trust anymore. Further, it has a potential risk for a single point of failure leading the whole system to a failure [40]. The utilization of blockchain is one of the novel technologies that can be adhered to these issues as it removes the reliance on a central entity. In the blockchain, all the data can be saved in a transparent and shared system for all the members in the supply chain instead of storing data in an opaque system. As an upcoming technology, blockchain also has some limitations, and with the increasing number of applications, scalability has become urgent, and a significant concern [41].

The utilization of blockchain and IoT technologies in FSCs has been proposed in several works. In relation to IoT equipped FSCs, in [42], an FSC system is established for China based on RFID and blockchain technology. All aspects of data gathering and information management in each link of the FSC, which enables monitoring, and traceability for the quality and safety of the food "from farm to fork" was discussed. Authors of [43] proposed a scheme for the use of blockchain in SC management for rice. Reference [41] proposed a traceable scheme based on Hazard Analysis and Critical Control Points (HACCP), blockchain and IoT. To solve scalability issues due to storage growth, BigChainDB is used.

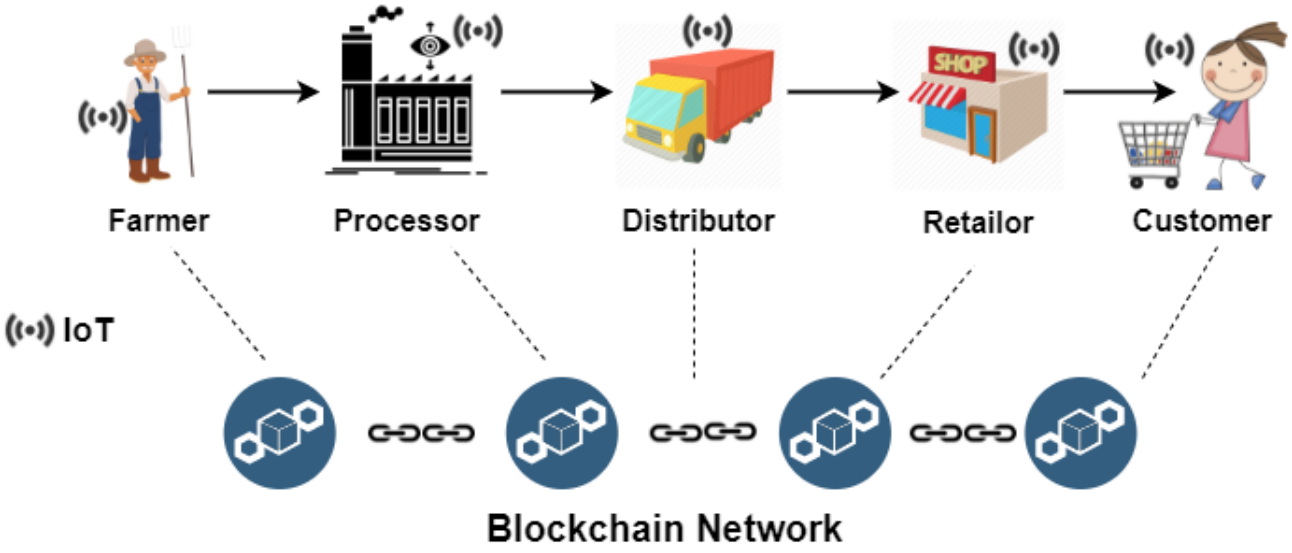


Figure 2.10: Supply chain availability with the use of IoT and Blockchain

## 2.5 Blockchain Storage in Food Supply Chains

When a large number of IoT devices are used in an FSC, the amount of transaction data that will be stored in the blockchain will also be extremely large. This data may include audio, video, sensor data, and images. Therefore, the blockchain ledger storage will grow continuously, requiring a large storage volume on each node, leading to higher costs. This may significantly affect the scalability of the SC. In the blockchain, for storing these transaction details, there are two options to consider as onchain storage (legacy storage mechanism) and offchain storage.

### 2.5.1 Onchain Storage

In the blockchain, it can simply store the information in the chain itself, and this is called onchain storage. For example, it can store binary data as a part of the transaction and then can distribute it within the network together with the rest of the transactions. How much data can be stored onchain and what happens when a piece of data becomes equal or exceeds the given size limit are the questions that come when dealing with onchain storage.

- **Fixed block size**

The key blockchain example for the fixed block size is Bitcoin. Here, it has given more attention to the semi-fixed block size of Bitcoin, that equals one megabyte. As the size of the block in anywhere can be slightly higher than one megabyte as well as can be slightly lesser, the size is semi instead of hard fixed. The segment of the controversy generates from the limited amount of transactions that can store in one block. When this combines with

the present popularity of Bitcoin, it forces users to spend on a high transaction fee. These fees encourage miners to include the particular transaction in their block [44]. The effect of the fixed block size in an onchain storage is clear. Even a small file which is of five megabytes is expensive and impractical for storing. Increased block sizes allow bigger pieces of information to be stored. It eliminates the adverse effects of the allocation of space for general transactions. But there arise several side effects when increasing the block size. An increase of the block size would simultaneously increase the amount of transactions proceed per second. Currently, the throughput of Bitcoin transactions is around seven transactions per second (tps). This is a significantly less value when compares with the throughput of the Visa, where the average throughput is around 2000 tps [45]. Another drawback on increasing the block size is that it would need a full node to spend more on the computation of resources and storage in order to maintain the increased size [46]. A theorized consequence on this is that this may pave the way to a more centralized system as the number of full nodes would decrease. On the other hand, it requires a hard-fork of the chain to change the block size. This means that it would need to update the current tools in the system in order to support the new system. The new version can fail, or the two versions have to continue side by side if all the organizations do not agree on the new fork. Both situations will have negative impacts on the blockchain.

This clearly implies that the use of fixed-size blockchains in onchain storage for FSC is impractical according to the considerations in FSCs. Either the blocks are expensive and small, or the blocks are expensive, large and centralized.

- **Variable block size**

One of the main blockchain examples for the variable block size is Ethereum [47]. The ability of changing the block size has made it a much more attractive option. Specially when it needs to handle large size files, the block size can increase to store the whole file in a single block. One of the advantages of this is that only a single node is needed to store files in a lightweight node to access the information. Regardless of the block size, the hash requires to validate the next block is only 32 bytes without having any negative effect on the remaining part of the chain.

However, it can be found three major issues when storing files in Ethereum. Out of those, two of them are mainly connected with the cost. Firstly the cost of a contract or a transaction is proportional to the block size. The fee for a 256-bit word will be 20,000 gas(tx currency) according to the technical definition in Ethereum. This means that the fee for data of 1 megabyte is 625 million gas [47]. Depending on the prices, this transaction will need thousands of dollars. Secondly, a miner needs to be willing to process the block assuming, that someone is willing to pay for the transaction. Thirdly, while there is no theoretical limit

to the Ethereum block size, it only depends on the demand of the network and votes cast by miners. The block size depends on the block gas limit rather than vice versa. If the miner nodes decide to increase the gas limit on the block, then as a result, the size of the block will be increased. To save large size of files in a single block, the miners should allow such a block size.

### 2.5.2 Offchain Storage

Although onchain storage is not technically or financially practical for the FSC applications, offchain storage methods can be used for such applications. Then the hash of a corresponding data, which is a significantly smaller size, can store in the blockchain, and the incoming transaction data can store in the offchain. Because of the small size of a hash, the corresponding storage cost will also be minimum within the blockchain. There are two main challenges within the offchain storage methods. The first challenge is providing a link between the physical storage location and the hash in the chain. For that, smart contracts and distributed hash tables (DHT) are considered as in [48]. A DHT is a decentralised and distributed key-value pair storing system. It guarantees that any participating node can efficiently retrieve the value associated with a given key using a lookup service. DHT handle how data is distributed and accessed while a smart contract defines what, how, and who.

The second challenge is maintaining a low latency for data storing within the offchain storage. Once the transaction data are received by the nodes, the time taken for the whole process of storing the transaction details in the offchain storage and returning the hash should get an acceptable value. Otherwise, it will be a problem for storing the real time transactions which generates within the blockchain. For the actual file storing, smart contract can use a separate Distributed File System like InterPlanetary File System (IPFS). In [49], an innovative InterPlanetary File System (IPFS) based storage model is proposed for blockchain. However, when the volume of transactions grows, the requirement for IPFS storage also grows significantly.

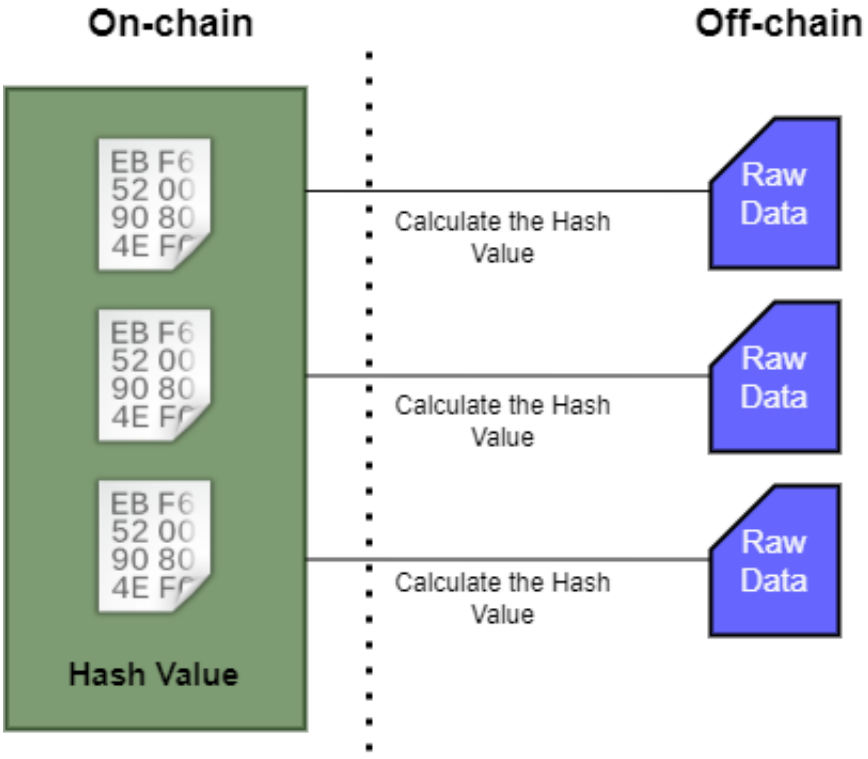


Figure 2.11: How offchain Storage works

# Chapter 3

# Proposed Architecture and Methodology

## 3.1 Problem Formulation

Based on the literature, blockchain is a time-stamped record series of immutable data that can enhance existing FSCs with traceability, anti-counterfeiting, provenance, and ownership of information. Users of this system can access all the transactions in real-time and simultaneously. The main advantage of the blockchain is that information distributed within the network cannot be corrupted, manipulated or hacked in any approach once they are added into the blockchain.

With the rapid evolution of the technology behind IoT devices and electronic chips such as RFID seals, companies have got the opportunity to attach sensors to goods allowing tracking them and detecting potential failures and frauds. Combination of those data from sensors with blockchain can provide traceable, transparent and standardized information to the FSC and help the industries in FSC in ensuring tamper-proof and reliable flows, which were missing in the food industry for a long time.

When a large number of IoT devices are used in an FSC, the amount of transaction data that will be stored in the blockchain will also be extremely large. This data may include audio, video, sensor data, and images. Therefore, the blockchain ledger storage will grow continuously, requiring a large storage volume on each node, leading to higher costs. This may significantly affect the scalability of the SC. Based on the literature, for storing these transaction details, offchain storage is ideal than the onchain storage in the blockchain. Then only the hash of the corresponding data, which is of significantly smaller size compared to the original data, could be stored in the blockchain. Smart contracts, as well as DHTs, can be used to establish a link between the hash found in the chain and the actual physical storage. A distributed file system among the already available nodes is used to maintain the

offchain storage.

But when there are more transactions in the FSCs, the offchain storage will also grow with time, and it requires a large storage volume in the nodes to store the transaction data. This may significantly affect the performance of the whole system. Therefore, there is a high requirement for developing a storage recycling scheme for blockchain-enabled IoT equipped FSCs. Then we can recycle the required offchain storage of the blockchain up to a certain extent. Actually, in here, we consider the fact that traceable data of a particular food product may not be required after a certain period of time. (eg. the expiration date)

## 3.2 Proposed System Model

In this section, we propose a layered architecture to establish a blockchain enabled IoT equipped FSC, with offchain storage recycling feature, and describe the end-to-end system functionality. Fig 3.1 illustrates the system layers and their components.

The data layer (DL) extracts the information related to the interactions among the entities in the FSC, such as trading of raw materials, products, and services, along with a proof of an auditable delivery. In each stage of the FSC, IoT devices are placed to acquire the data, including invoices, videos, audio clips, images, and critical sensor readings, to be transferred to upper layers.

The blockchain layer (BL) handles the transaction data of trading goods and processes. To reduce the amount of onchain storage required, BL only stores the hashes of the data, while the actual data are stored on the storage layer.

The storage layer (SL) is responsible for storing the transaction data of the blockchain in immutable Distributed File System Networks (DFSNs) and immutable Distributed Database System (DDS). All the transaction data including invoices, videos, audio clips, images and sensor data, will be stored in the relevant DFSN, and a hash value corresponding to the respective content will be returned to the BL.

The main functionalities of the proposed system are described in Subsections .

### 3.2.1 Product Data Collection

Similar to a conventional IoT equipped FSC, transaction data related to the traceability of the products will be collected and recorded. In the proposed scheme, we pay special attention to the unique product identifier (PID), product expiry date (PED) and the dispute reason field (DRF). The DRF is used to indicate whether the product faced issues such as legal action, large quantities remained unsold, or any other incident that may compromise the quality of the product.

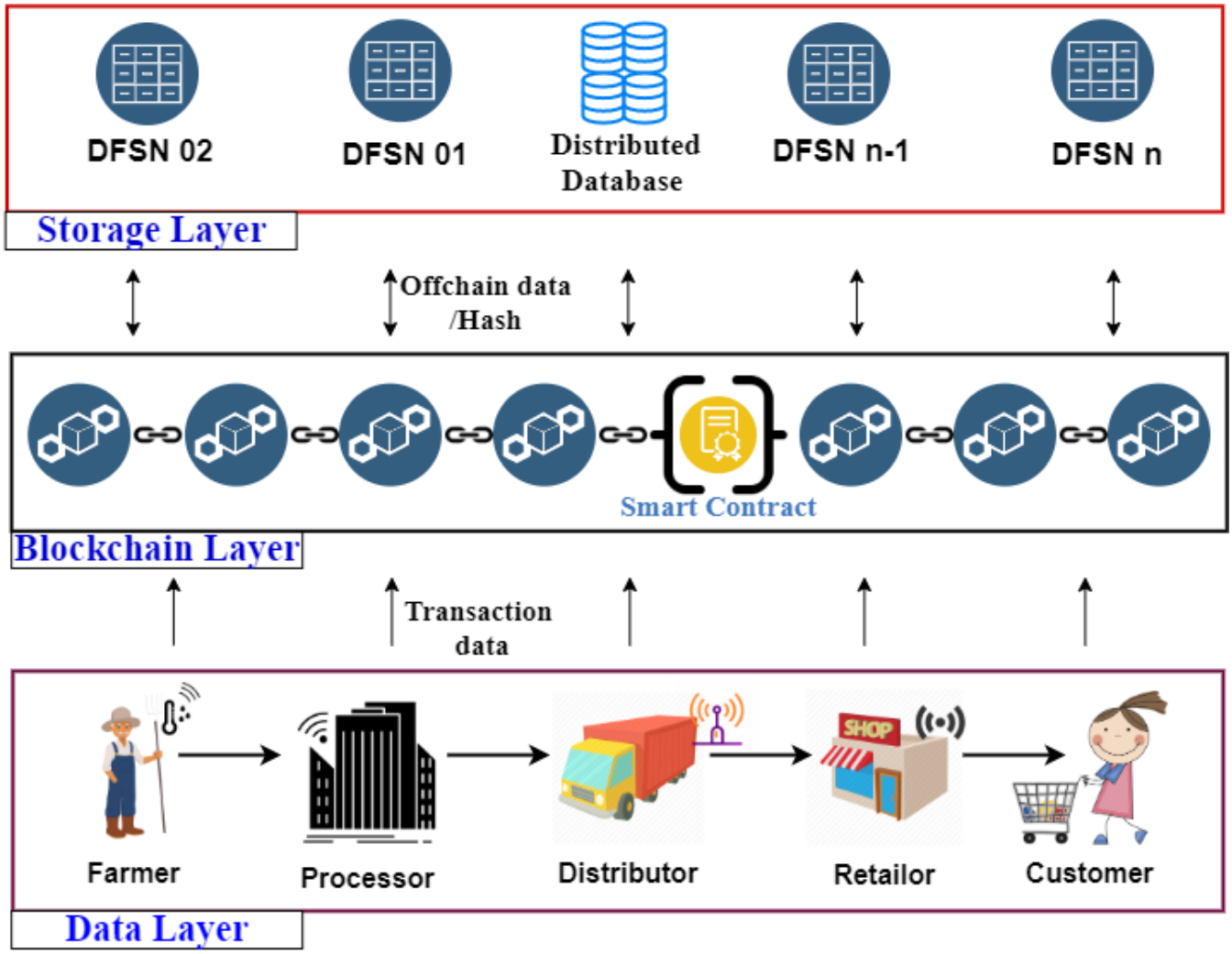


Figure 3.1: Blockchain-based end to end solution for food supply chain

### 3.2.2 File System Creation

Consider a time span of $T$ as the data retention period (DRP) of the system. The DRP is selected based on the life cycle of the products in the FSC. The DRP can be divided in to $n$ slots of equal duration $T/n$. The slot duration (SD) determines the frequency of storage recycling within DRP. To store the data, $n+1$ DFSNs are pre-configure across all the nodes, and unique identifiers Id0 to Id$n$ will be assigned to them. For the remainder of the paper, we set DRP to 1 year, and perform storage recycling weekly. Therefore, $n$ is set to 52. Each week will be mapped to a separate DFSN, and a "special week id" will be mapped to DFSN Id0. Then DDS consisting of two tables, namely, T1 and T2, will be configured across all nodes. T1 is used to keep the mapping of the week identifier (id) (assigned based on the system initiation date) and the corresponding DFSN name, while T2 is used for storing every transaction record in the system. The table structures of the T1 and T2 in the DDS is shown in Fig 3.2.

DDS T1

| Week Id | DFSN Name |
|---|---|
| | |

DDS T2

| Timestamp | Hash | Product Id | Expiry Week Id | DFSN Name | Dispute Reason |
|---|---|---|---|---|---|
| | | | | | |

Figure 3.2: DDS Table Structure

### 3.2.3 Storing Data in DFSN and DDS

When the miners get the transaction data from the DL, they first validate the transactions, check the product expiry date and calculate the expiry week Id according to the system initiation date by executing smart contracts. Then, the corresponding DFSN name is obtained by querying the DDS T1, and the data will be stored in the DFSN corresponding to the identified week, and the hash will be returned. Thereafter, these data will be saved in a T2 entry, containing timestamp, hash value, product id, expiry week id, DFSN name and dispute reason (if any), by executing smart contracts. When generating the next block, each miner stores the returned DFS hashes of the verified transactions in the new block. Then it calculates the Merkle root and the block hash. The flow diagram of the Data storing process in DFSN and DDS is shown in Fig 3.3. The DFSN based storage model for blockchain is shown in Fig 3.4.

### 3.2.4 Building the blockchain

As the Fig 3.4, in the case of miner A successfully calculating the block hash that meets the difficulty, that block will be broadcasted to miners B, C, and D. Miners B, C and D will need to verify the transactions and the block hash. In fact, during the mining process, miners B, C, D would have received mostly the same transactions as miner A would have. Some differences between the local transaction pools of the miners can be cause by network

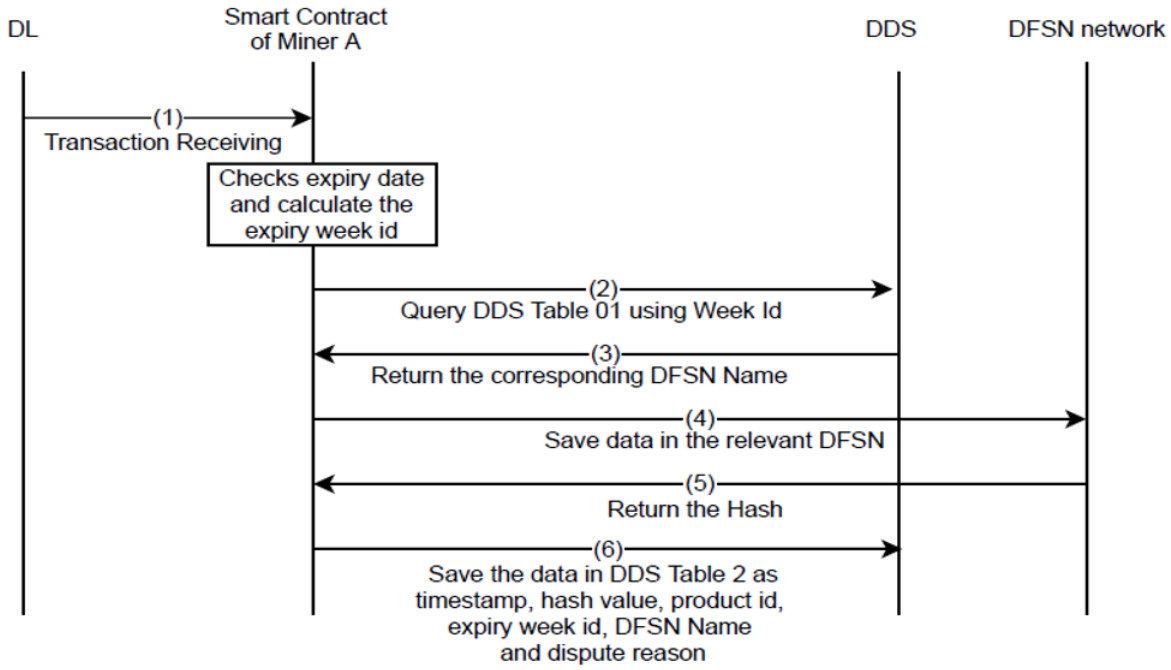


Figure 3.3: Flow diagram of the Data storing process in DFSN and DDS

transmission delays. Hence, most of the transaction DFS hashes in the newly received block are the same as the hashes in the local pool of transactions in miner B, C and D. In the case where the DFS hash of a particular transaction in the block sent by miner A is available in the local transaction pool, miners B, C and D can determine it as a transaction confirmed by them before. Then there will be no need to download it from DFSN [49]. For the rest of the transactions, the data needs to be requested from the DFSN through their corresponding DFS hashes. Before requesting the data from the DFSN, each miner needs to know in which DFSN the corresponding hash has been saved. For that, it needs to query the DDS T2 via the hash and get the DFSN name and then download the data through their corresponding hashes. Then the validity of transactions and the block would be confirmed. Afterwards, the new block can be appended to the blockchain ledger. The flow diagram of the process of new block adding is shown in Fig 3.5.

### 3.2.5 Storage Recycling

At the end of each week, a smart contract will be triggered to identify the products whose expiration date falls within that week. A new DFSN will be created for the corresponding week in the subsequent year, and the data deletion process will be initiated. The smart contract obtains the DFSN name corresponding to the current week by querying DDS T1. Furthermore, it checks T2 to find out whether there are any disputes regarding the transactions related to any of the products expire in that particular week. If there are no disputes,

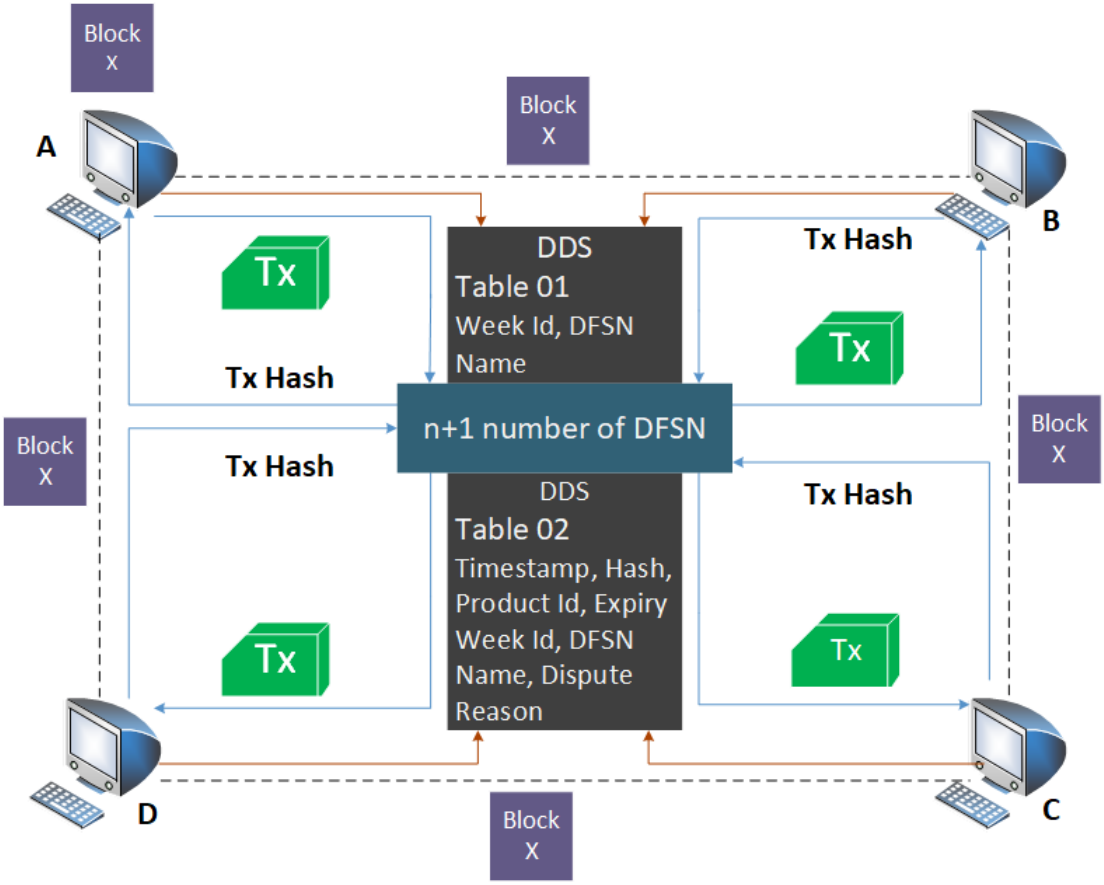


Figure 3.4: DFSN based Storage Model

the smart contract deletes the relevant DFSN from the nodes. Then the deleted transaction data record will be saved in DFSN Id0, which corresponds to the special week id.

If there are any disputes regarding any product expiring during the week, the smart contract will not delete the DFSN until the the dispute is resolved. This is required since we are using immutable DFSNs, which do not allow to change the entries. Therefore, it is necessary to delete the entire DFSN at once rather than deleting specific entries. This process is summarized in Algorithm 1. It is important to note that only the registered users are allowed to perform the specific transaction. The flow diagram of the process of storage recycling is shown in Fig 3.5.

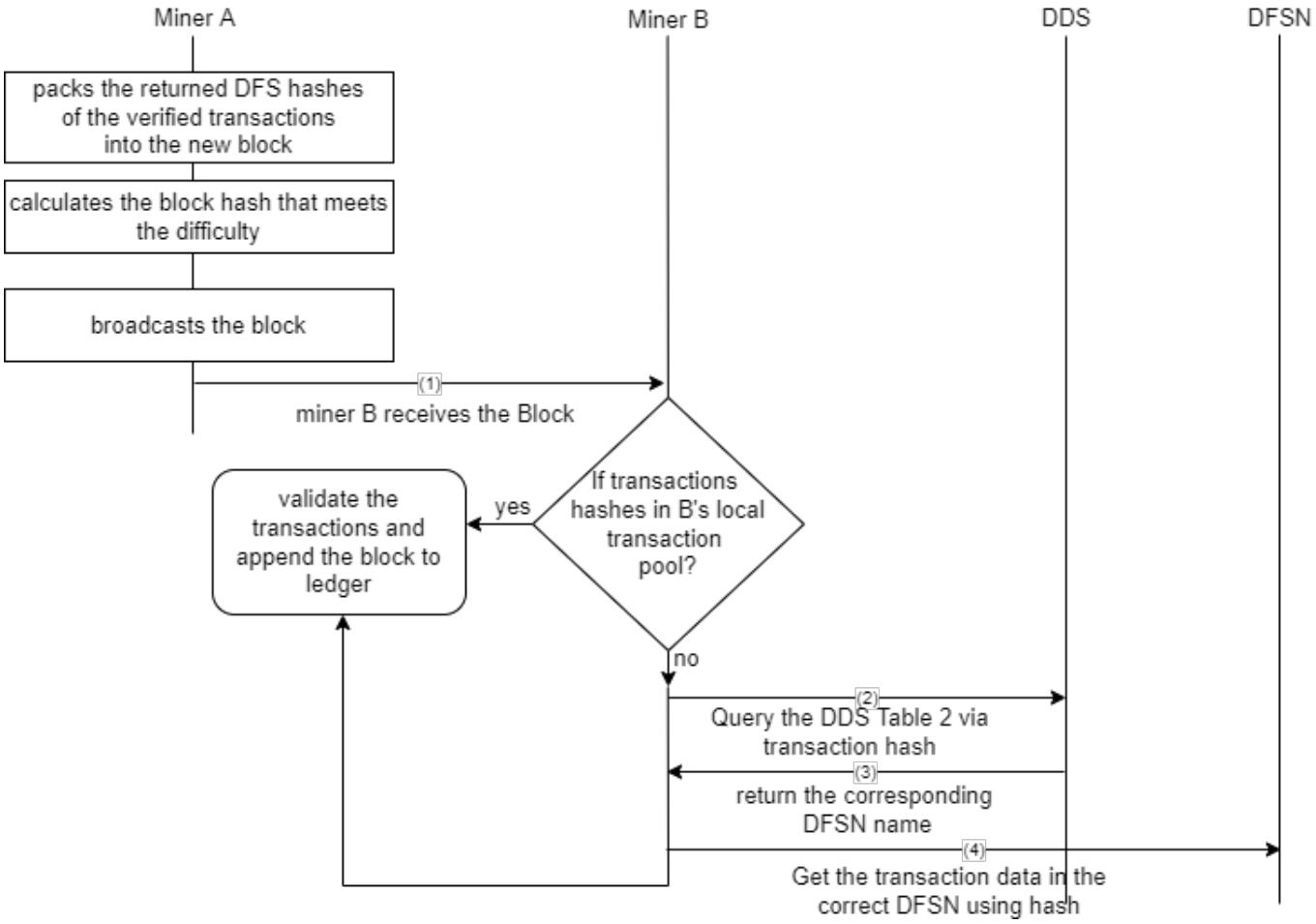


Figure 3.5: How new block is getting added to the Blockchain

**Algorithm 1** Storage Recycling

**Require:** $currentWeekId$

1: **for** each deletion request
2: $dfsnName \leftarrow$queryDDST1(getDFSN($currentWeekId$))
3: $drfValue \leftarrow$queryDDST2(getDRF($dfsnName$))
4: **if** ($drfValue == null$ ) **then**
5: $result \leftarrow$deleteDFSN($dfsnName$)
**do**

The total onchain and offchain storage size for proposed model after $X$ years with DRP set to one year can be calculated as follows.

- The expected data volume in the blockchain after $X$ years, $V_{onchain}$, is given by

$$V_{onchain} = H + N \times iHash \tag{3.1}$$

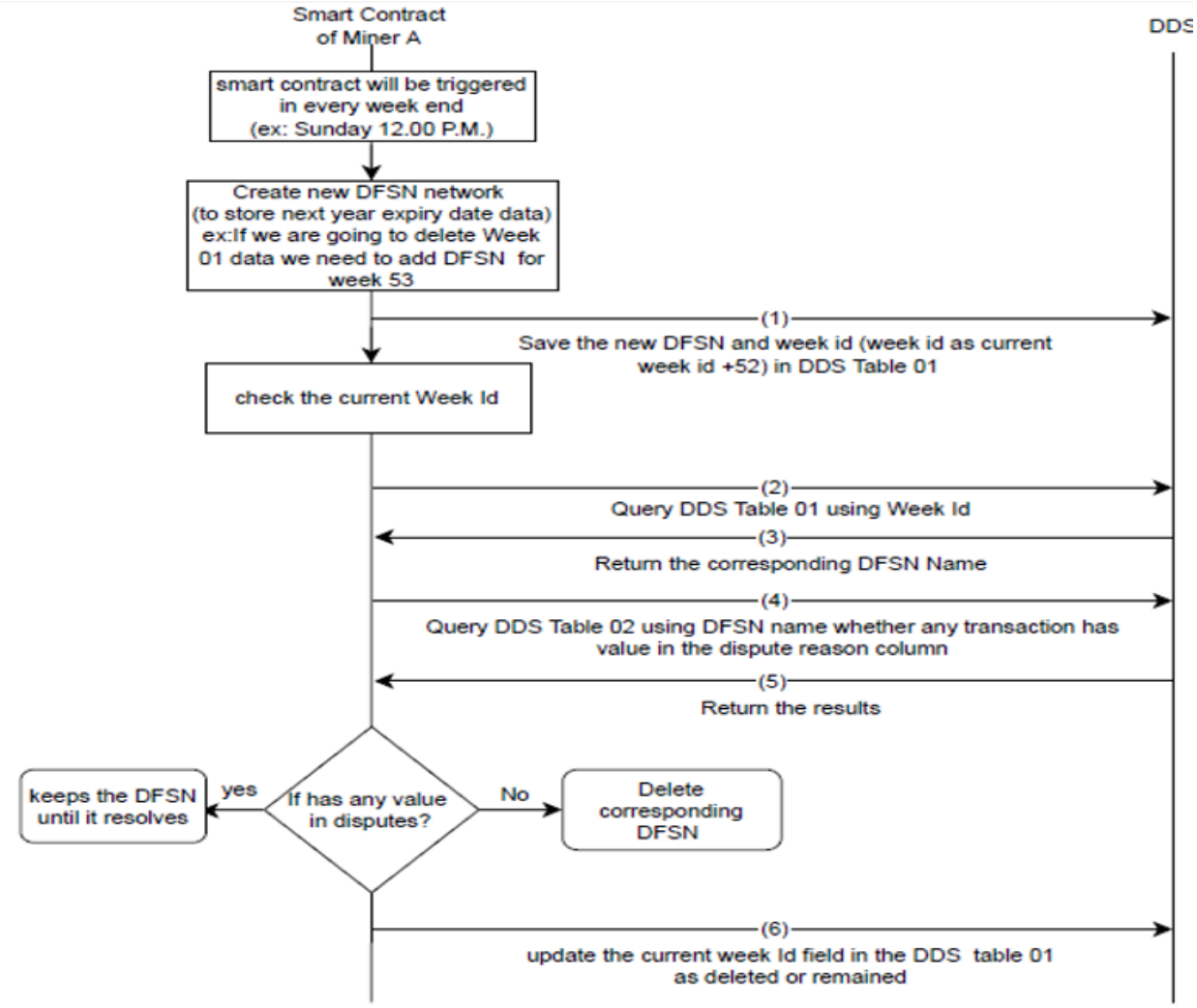


Figure 3.6: The Process of Storage Recycling

- The expected offchain data volume after $X$ years, $V_{offchain}$, is given by

$$V_{offchain} = \sum_{i=1}^{M} T_i \tag{3.2}$$

The expected offchain data volume after $X$ years in a conventional system, $V_{legacy}$, can be obtained as

$$V_{legacy} = \sum_{i=1}^{N} T_i. \tag{3.3}$$

The symbols are defined in Table 3.1

Table 3.1: The Implication of Symbols in formula

| Symbol | Definition |
|---|---|
| $N$ | The number of all transactions in the blockchain after $X$ years |
| $M$ | The number of transactions in the blockchain for $X^{\text{th}}$ year |
| $iHash$ | The size of the DFS hash corresponding to each transaction |
| $H$ | The total data volume of all block header for $X$ years |
| $T_i$ | The original data volume of the $i^{\text{th}}$ transaction |

# Chapter 4

# Numerical Results

In this chapter, to compare the storage growth of the proposed scheme with the legacy storage mechanism in blockchain, we need to use a food transaction database at least in the past five years, including product expiry dates. However, finding actual blockchain data related to the FSCs is difficult. Therefore, in this section, the storage volumes are calculated using actual bitcoin transaction statistics, namely, the average block size [50], the average number of transactions per block [51], and the average transaction count [52], from 2015 January to 2020 December, and assume these data volumes have occurred from the food supply chains. Here we evaluate the offchain storage growth in a single node for comparing the proposed scheme and the legacy storage mechanism.

By using Fig 4.1, Fig 4.2 and Fig 4.3 data, We find the Average Transaction Size, Average Full Volume (MB) and Average Block Count in the particular day of the system as the below equations.

$$Average\ Transaction\ Size = \frac{Average\ Block\ Size}{Average\ Number\ of\ Transactions\ per\ Block} \tag{4.1}$$

$$Average\ Full\ Volume = Average\ Transaction\ Size * Average\ Transaction\ Count \tag{4.2}$$

$$Average\ Number\ of\ Blocks = \frac{Average\ Full\ Volume}{Average\ Block\ Size} \tag{4.3}$$

We use these calculated data from the above equations to map our FSC system with a recyclable offchain storage mechanism by considering the fact that the average full volume in a particular day will be recycled within the day itself from the system if DRP is set to one day. Here, we assume the bitcoin transaction stored dates as product expiry dates.

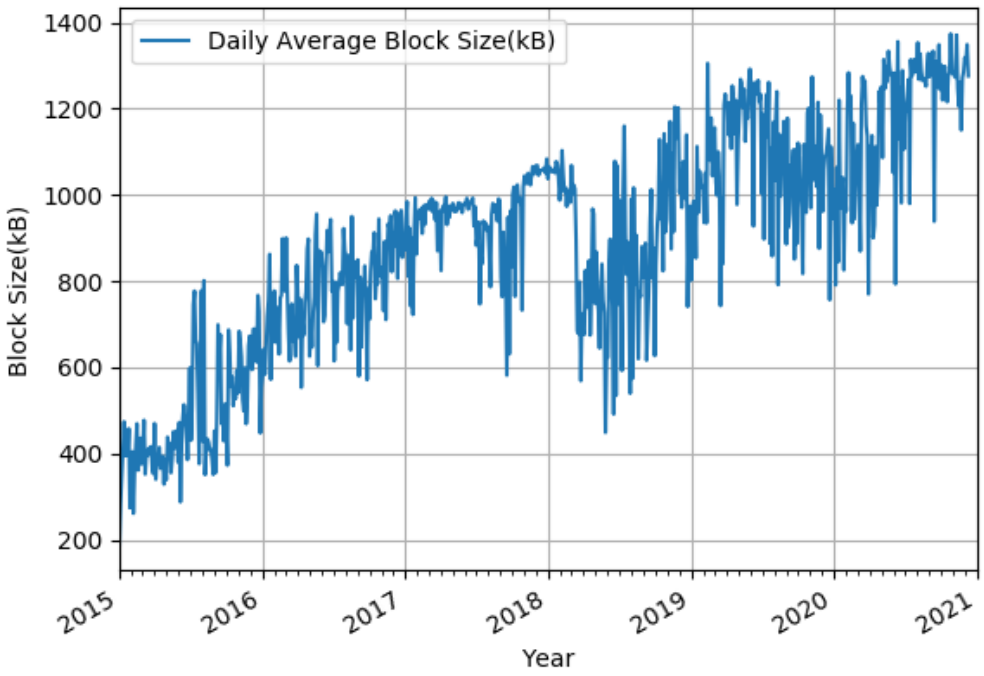


Figure 4.1: Average Block Size

Furthermore, we assume that all records are free of disputes and are eligible for recycling, and IPFS is used to implement the DFSN. To get the weekly full volume, we calculated the cumulative volume of 7 days in each week and considered the seventh day as the product expiry date (transaction stored date).

To calculate the onchain data volume, we multiply the average daily transaction count by the IPFS hash size of 46 bytes. Then we add the block header of 80 bytes by multiplying with the daily average number of blocks, as given in equation (3.1). To use the bitcoin data for the proposed system, first, we need to select the DRP for the system. Based on the chosen DRP, the cumulative data volume of the system will vary. We evaluate the storage volume of the proposed scheme when the DRP is set to one year, one month and one day.

The cumulative data storage comparison between the legacy offchain solution and the proposed solution with DRP set to one year, is shown in Fig. 4.7. In addition, onchain storage volume is also shown. It is clear that the storage volume grows at a very slow rate with the proposed scheme, while the legacy scheme shows a linear growth with time. This behaviour is observed since the proposed scheme recycle the storage every week, and keeps only the records from the previous year in the system, while the legacy offchain storage solution has accumulated all the records over the years. Therefore, Fig. 4.7 clearly shows

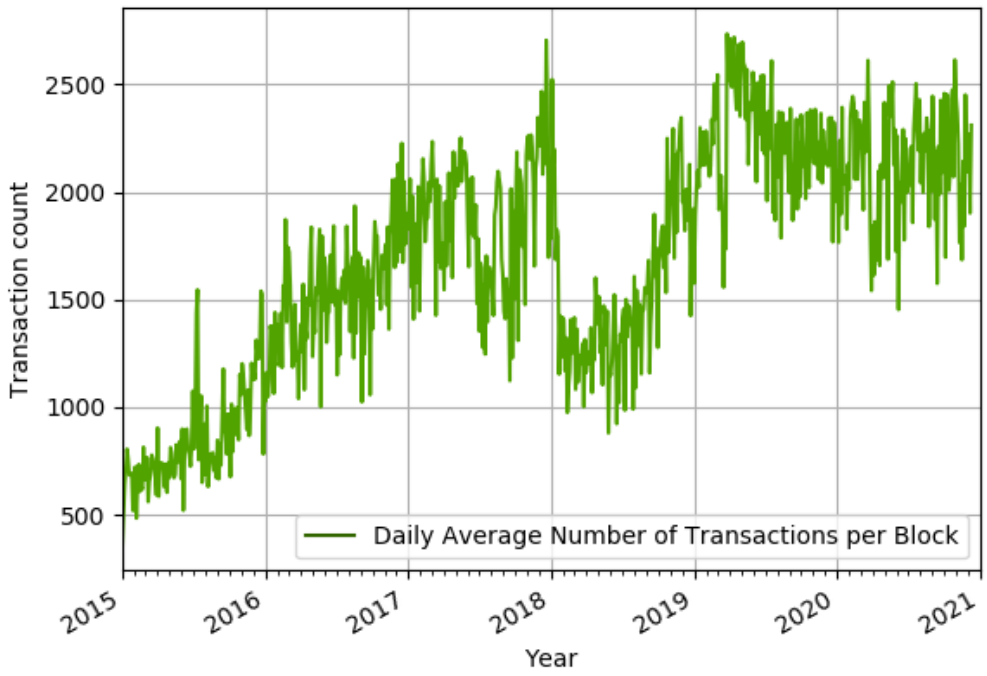


Figure 4.2: Average Number of Transactions per Block

the storage cost savings achievable with our scheme.

Fig. 4.8 shows the percentage of storage saving of the proposed solution compared to the legacy offchain solution. Since we set DRP to one year, in first year both systems accumulate equal data volume. However, after the first year, proposed solution results in significant storage savings compared to the legacy system. It is important to note that percentage saving increases with time due to the storage recycling operation of the proposed scheme. From the numerical results, one can observe that, after 5 years, the proposed scheme results about 78% of storage space compared to the legacy solution.

Fig. 4.9 shows the offchain storage volume growth of the proposed solution with different DRPs when SD is set to one day. The amount of data we keep offchain depends on the DRP. For an example, if we set DRP as one year, system will only keep previous year's records. Similarly, if we choose the DRP as one month, system will only keep the previous month's records. Therefore, shorter DRP leads to a smaller storage volume. However, since we are recycling the storage only after the product expiration date, DRP must be selected based on the life cycle of the products in the SC.

From numerical results, it is evident that the proposed storage recycling scheme yields significant storage requirement reduction and improves the scalability of blockchain enabled

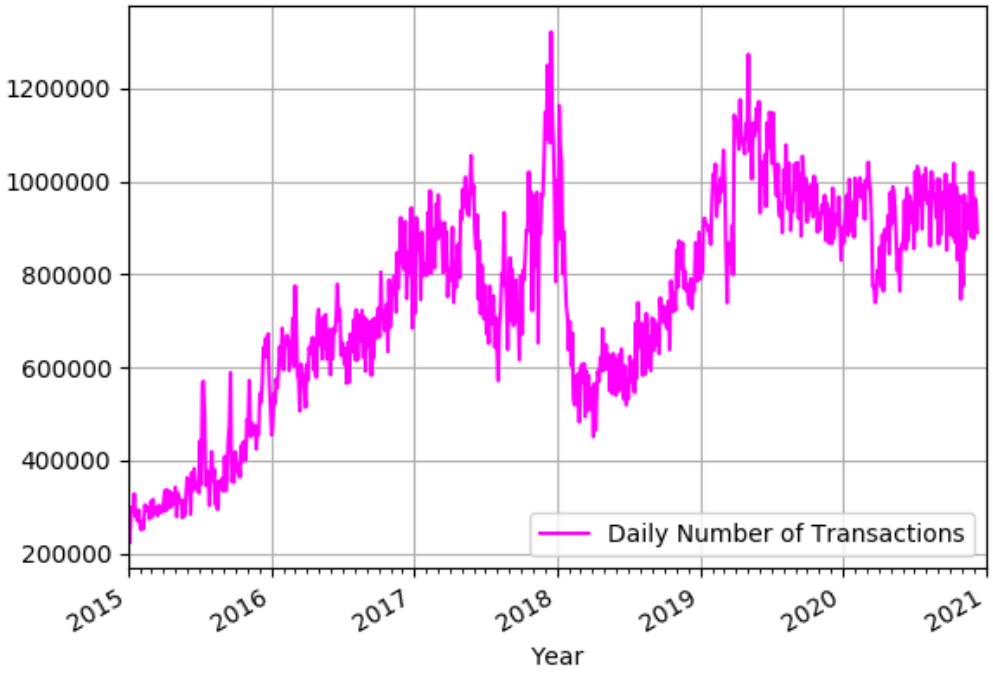


Figure 4.3: Average Transaction Count

IoT equipped FSCs.

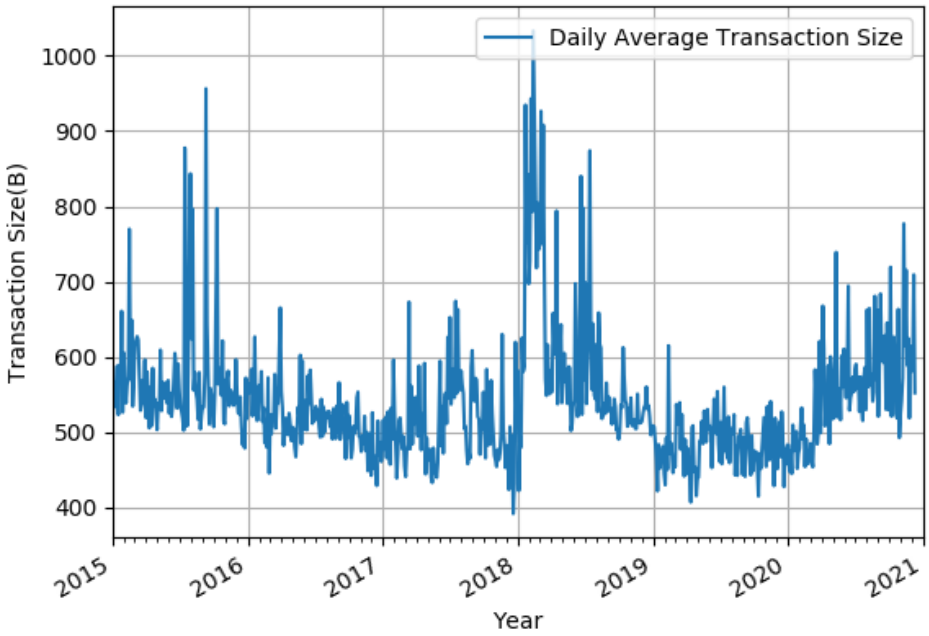


Figure 4.4: Average Transaction Size

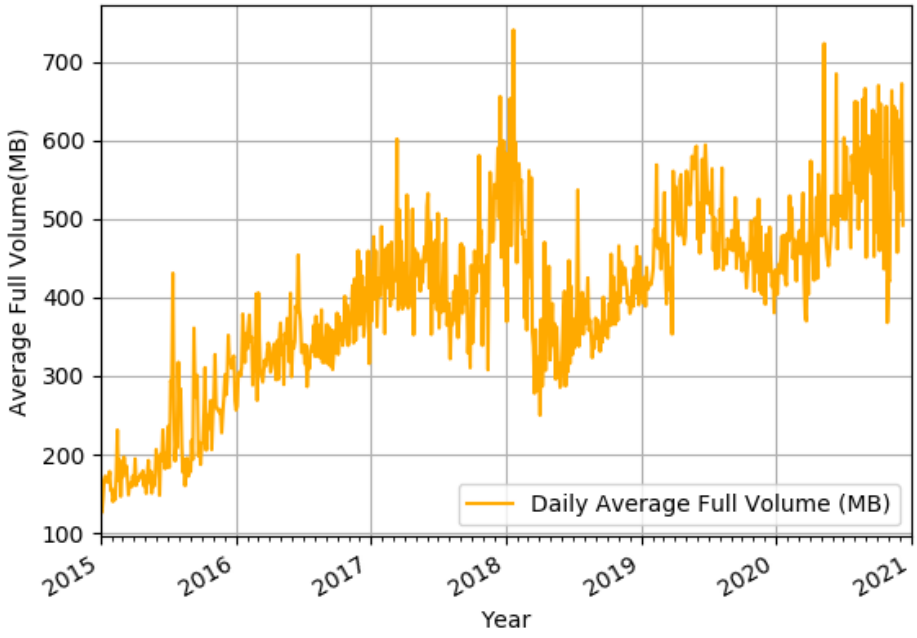


Figure 4.5: Average Full Volume

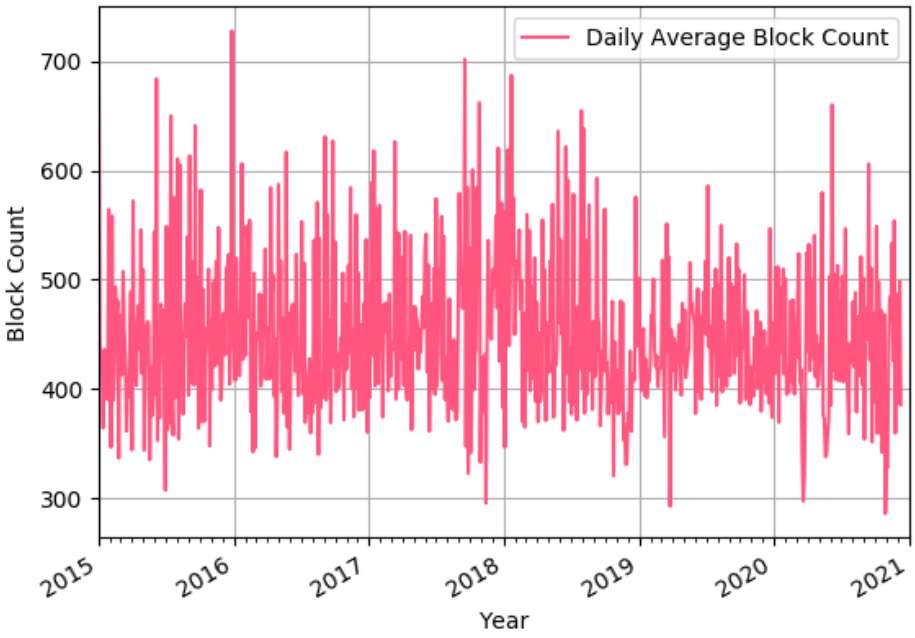


Figure 4.6: Average Block Count

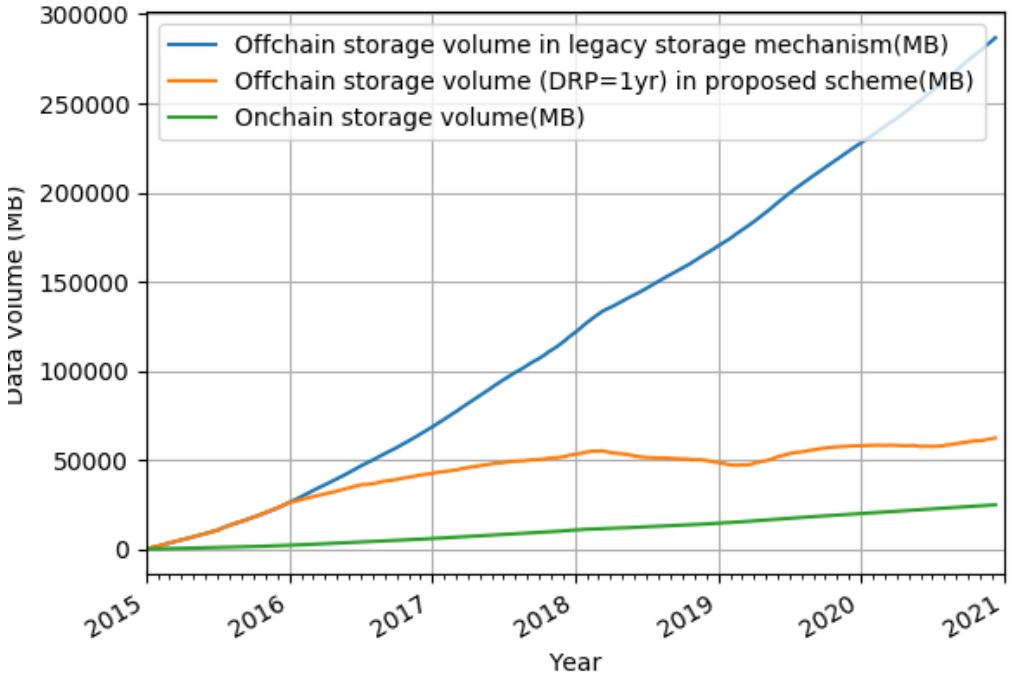


Figure 4.7: Data Storage Comparison

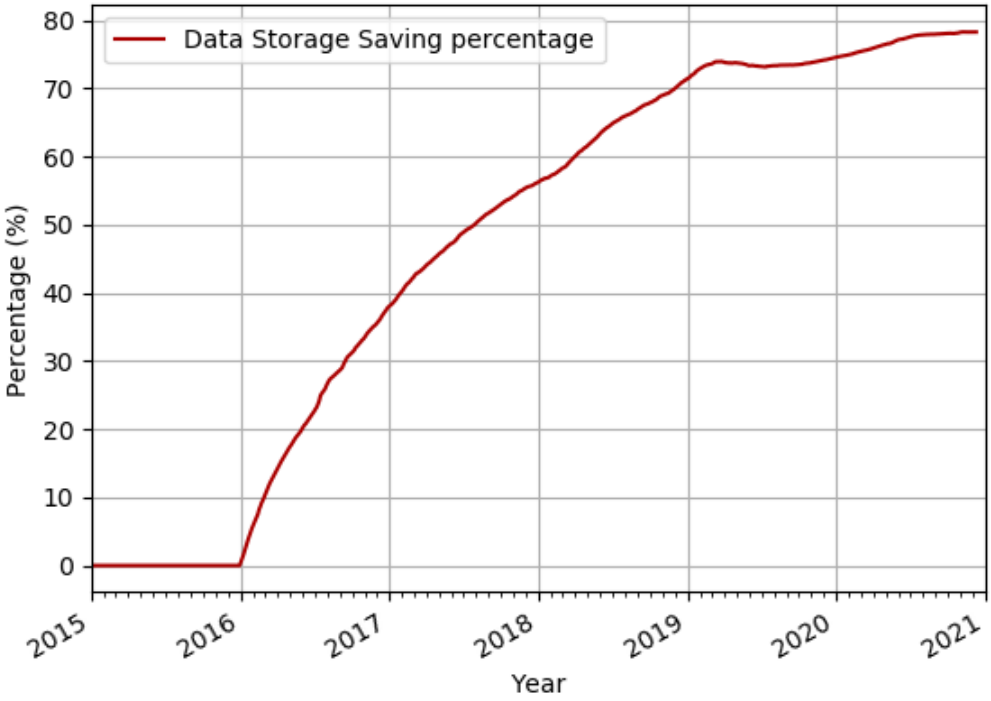


Figure 4.8: Reduction of storage volume compared to the legacy scheme

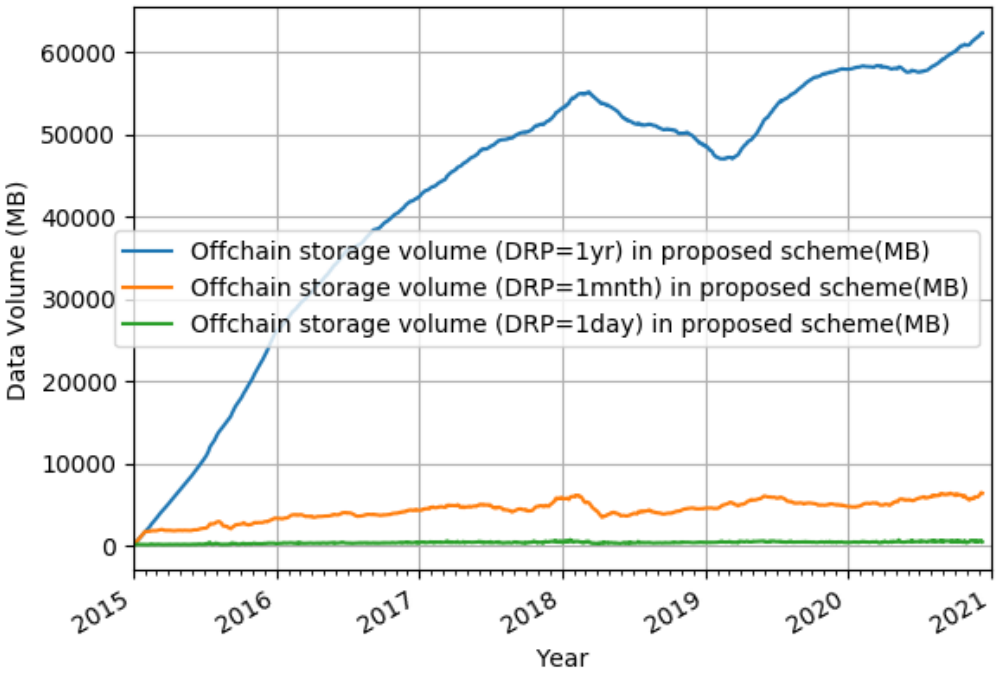


Figure 4.9: Offchain storage growth with different DRPs

# Chapter 5

# Experimental Results

Since the proposed scheme uses immutable DFSNs, it is important to show that the data deletion process can be implemented within an acceptable latency. Therefore, to demonstrate the technical feasibility of the proposed scheme, we implement it on Hyperledger blockchain platform.

## 5.1 Experimental setup

The computing infrastructure used for the implementation consists of a virtual machine (VM) and one host machine. The VM runs Ubuntu 20.04 64-bit with 8GB RAM and four processor cores with 2.40GHz speed. The host machine consists of Intel(R) Core i5 -7200 CPU with dual cores. Fig 5.1 illustrates the implementation setup. The DFSN is implemented using IPFS, and it is configured and implemented on top of the docker containers.

The Hyperledger Fabric blockchain platform (HFBP) is connected to the IPFS as the extension of storage. The IPFS is connected to smart contracts via REST API to store the objects in the distributed storage. The HFBP is connected to an external web client via the REST API to submit and retrieve transactions. The smart contracts are deployed to intervene for different steps of the blockchain. Within the implementation process, we test the solution using one blockchain node and two different IPFS networks, which are setup inside the docker containers. The simulation of transaction traffic is performed by the software running on the external web client.

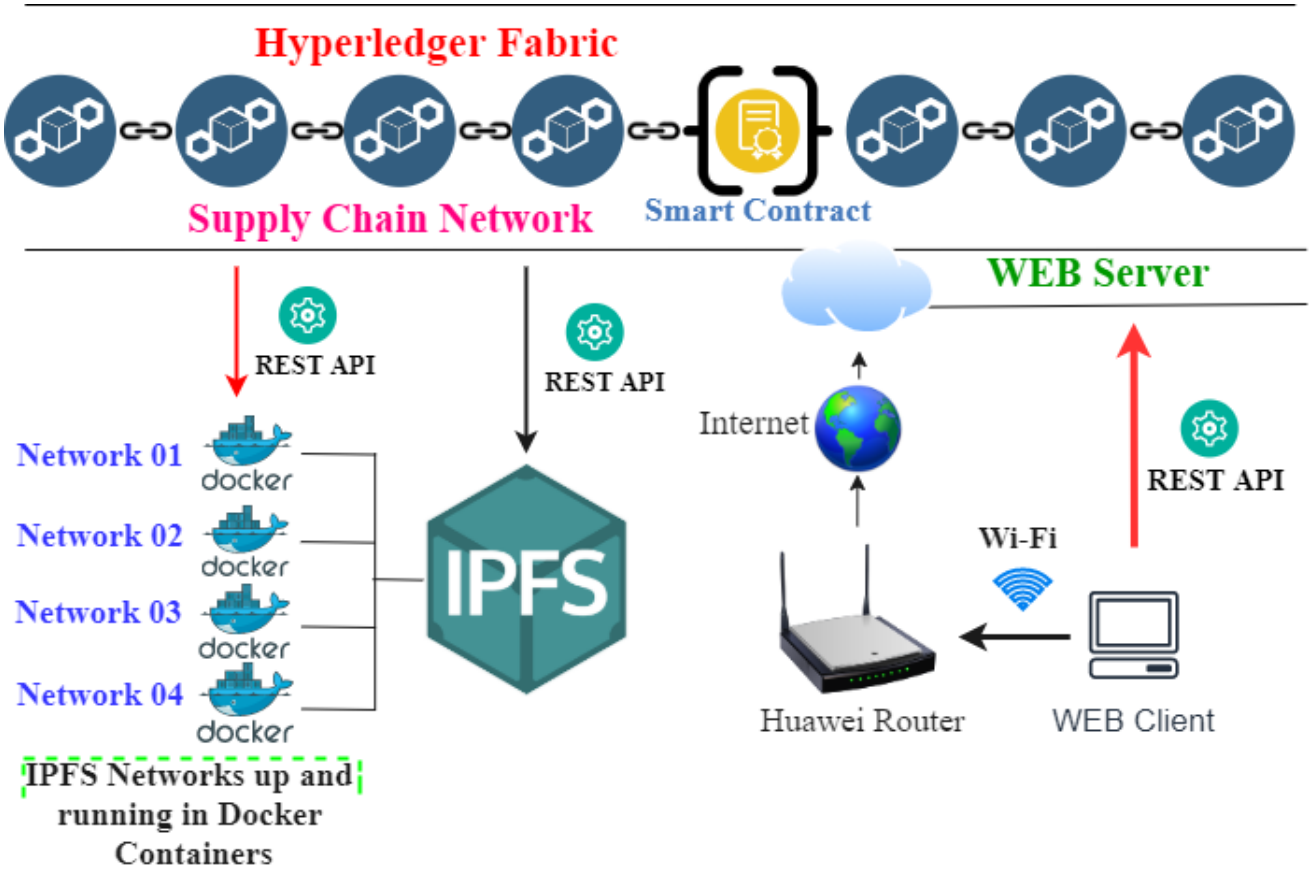


Figure 5.1: DFSN based Storage Model

## 5.2 Latency Calculation

### 5.2.1 Data Storing Latency

We emulate an FSC environment and evaluate the food item data storage latency through the IPFS network to the blockchain. Here, we set up the IPFS network on top of the docker containers. For latency calculations, we considered the activities of transaction request initiation via the external web client to the smart contract, storing the transaction details in the IPFS network via the REST API call from the smart contract, getting the hash of the transaction data from IPFS, storing the hash in the distributed ledger and the time it takes to mine a block. The latency is calculated for block generation times of 500ms, 1s, 2s, 4s and 8s. For each block generation time, the latency values are obtained by running the experiment for 100 instances, and taking the average latency value with a 95% confidence interval. The results are tabulated in Table 5.1.

The results show that the data storing latency varies with block generation time. However, if we split the time taken for each activity, and ignore the mining delay in each test scenario (assuming the mining delay is identical to the block generation time), the average

Table 5.1: Data Storing Latency measurements for different block generation times

| Block Generation Time | Data Storage Delay (ms) | Data Storing Latency in IPFS (ms) |
|---|---|---|
| 500ms | 1314.22±12.3 | 814.22±12.3 |
| 1s | 1841.02±15.1 | 841.02±15.1 |
| 2s | 2824.75±14.9 | 824.75±14.9 |
| 4s | 4825.24±13 | 825.24±13 |
| 8s | 8865.37±14.8 | 865.37±14.8 |

time for the storing the transaction details in the IPFS network and returning the hash can be found as in Table 5.1. The average time is around 835 milliseconds for all block generation times. This confirms that data storing and hash acquisition from the IPFS network, which run on top of the docker container can be completed with an acceptable level of latency.

### 5.2.2 Storage Recycling Latency

In the experiment, we emulate the storage recycling process by triggering a deletion event hourly. Here, we consider a single, pre-configured IPFS network. From an automated external application, transaction deletion request is initiated to the smart contract with the IPFS network name. Then, the record is stored in the distributed ledger. Once it is stored, the smart contract calls an external REST API to create the new IPFS network and deletes the existing IPFS network. The deleted transaction record is stored in the distributed ledger. Once a block is mined, we calculated the time taken for the entire process to complete as the latency of the food item data storage deletion process. The latency is calculated for block generation times of 500ms, 1s, 2s, 4s and 8s. For each block generation time, the latency values are obtained by running the experiment for 100 instances, and taking the average latency value with a 95% confidence interval. The results are tabulated in Table 5.2.

Table 5.2: Storage Recycling Latency measurements for different block generation times

| Block Generation Time | Data Storage Recycle Delay (ms) | Storage Recycling Latency in IPFS (ms) |
|---|---|---|
| 500ms | 4448.83±34.14 | 3448.83±34.14 |
| 1s | 5441.87±22.85 | 3441.87±22.85 |
| 2s | 7425.07±25.84 | 3425.07±25.84 |
| 4s | 11424.05±28.36 | 3424.05±28.36 |
| 8s | 19411.62±27.61 | 3411.62±27.61 |

The results show that the storage recycling latency varies with block generation time. Similar to the previous experiment, if we ignore the mining delay for each activity, the average time it takes for the IPFS network creation and deletion activity can be found as in Table 5.2. The average time is around 3430 milliseconds for all block generation times. This affirms the recycling can be implemented with an acceptable level of latency. To scale-up the implementation setup, it is possible to configure several VMs as blockchain nodes, pre configure several IPFS networks and can follow the same procedure as above in each node. With this approach, it is possible to recycle the offchain storage, maintaining the accuracy and the integrity of the FSC.

# Chapter 6

# Discussion and Future Work

When a large number of IoT devices are used in an FSC, the amount of transaction data that will be stored in the blockchain will also be extremely large. This data may include audio, video, sensor data, and images. Therefore, the blockchain ledger storage will grow continuously, requiring a large storage volume on each node, leading to higher costs. This may significantly affect the scalability of the SC. To circumvent this, offchain storage models have been proposed to store the incoming transaction data, and only the hash of the corresponding data, which is of significantly smaller size compared to the original data, could be store in the blockchain. Since IoT data will be stored in an offchain storage, the offchian storage will also grow with time.

Therefore, the key outcome of our research is developing an offchain storage recycling scheme for blockchain-enabled IoT equipped FSCs. This model was developed based on the fact that traceable data of a particular food product may not be required after a certain period of time (eg. the expiration date). The proposed solution used smart contracts, distributed file systems with distributed hash tables, and distributed database to assure an efficient, secure, and trusted environment. To demonstrate the technical feasibility of the proposed scheme, we implemented it on the Hyperledger blockchain platform. Within the implementation process, we tested the solution using one blockchain node and two different IPFS networks, which are set up inside the docker containers. From the experiment results, we proved that the proposed scheme can be implemented with an acceptable level of latency, while maintaining all the necessary features of blockchains.

In all the blockchain-enabled systems, we can observe the data storage problem due to the existence of a large amount of transaction data. Though most of the systems are using offchain storage solutions, there is a high requirement of an end to end solution for a recyclable offchain storage mechanism. Therefore, our future studies are focused on developing the proposed solution for blockchain-enabled systems other than the FSCs.

# Chapter 7

# Conclusion

This thesis proposed a scheme to reduce the storage volume growth rate of blockchain-enabled IoT equipped FSCs by recycling the offchain storage. This is achieved by periodically deleting the DFSNs related to the food items, which are no longer required to be stored in the blockchain, due to their expiration. The proposed scheme can be implemented using immutable DFSN and DDS to maintain the authenticity and integrity of the transaction data. Numerical results demonstrated that the proposed scheme leads to 78% saving in storage volume compared to the legacy solution after 5 years, which can result in significant cost savings and improves the scalability of blockchain equipped FSCs. Furthermore, the proposed scheme was implemented on the Hyperledger blockchain platform to verify the technical feasibility. The experimental results confirmed that the proposed scheme can be implemented with an acceptable level of latency, while maintaining all the necessary features of blockchains.

# References


[1] "Distributed," [Online]. Available at https://wiki.installgentoo.com/index.php/Distributed, Accessed: 05.05.2021.

[2] N. Kawaguchi, "Application of blockchain to supply chain: Flexible blockchain technology," *Procedia Computer Science*, vol. 164, pp. 143–148, 2019.

[3] A. Manzoor, A. Braeken, S. S. Kanhere, M. Ylianttila, and M. Liyanage, "Proxy re-encryption enabled secure and anonymous iot data sharing platform based on blockchain," *Journal of Network and Computer Applications*, vol. 176, p. 102917, 2021.

[4] M. P. Caro, M. S. Ali, M. Vecchio, and R. Giaffreda, "Blockchain-based traceability in agri-food supply chain management: A practical implementation," in *2018 IEEE IoT Vertical and Topical Summit on Agriculture-Tuscany (IOT Tuscany)*, 2018, pp. 1–4.

[5] T. M. Hewa, Y. Hu, M. Liyanage, S. Kanhare, and M. Ylianttila, "Survey on blockchain based smart contracts: Technical aspects and future research," *IEEE Access*, 2021.

[6] R. A. Mishra, A. Kalla, A. Braeken, and M. Liyanage, "Privacy protected blockchain based architecture and implementation for sharing of students' credentials," *Information Processing & Management*, vol. 58, no. 3, p. 102512, 2021.

[7] J. Fry, "Booms, busts and heavy-tails: The story of bitcoin and cryptocurrency markets?" *Economics Letters*, vol. 171, pp. 225–229, 2018.

[8] Z. Xi, "The comparison of decentralized and centralized structure of network communication in different application fields," 01 2020.

[9] Gerring T. (n.d.)., "Centralization vs Decentralization: Centralization & Decentralization." [Online]. Available at https://blockchain.wtf/what-the-faq/centralization-vs-decentralization/, Accessed: 07.11.2021.

[10] P. Tasca, T. Aste, L. Pelizzon, and N. Perony, *Banking beyond banks and money*. Springer, 2016.

[11] M. A. Khan and K. Salah, "Iot security: Review, blockchain solutions, and open challenges," *Future generation computer systems*, vol. 82, pp. 395–411, 2018.

[12] C. Catalini and J. S. Gans, "Some simple economics of the blockchain," *Communications of the ACM*, vol. 63, no. 7, pp. 80–90, 2020.

[13] Buterin V. (2015)., "On Public and Private Blockchains." [Online]. Available at https://blog.ethereum.org/2015/08/07/on-public-and-private-blockchains/, Accessed: 07.11.2021.

[14] Thompson C. (2016)., "The difference between a Private, Public & Consortium Blockchain: A Simple Explanation for Dummies." [Online]. Available at https://www.blockchaindailynews.com/The-difference-between-a-Private-Public-Consortium-Blockchain_a24681.html, Accessed: 07.11.2021.

[15] D. Kraft, "Difficulty control for blockchain-based consensus systems," *Peer-to-peer Networking and Applications*, vol. 9, no. 2, pp. 397–413, 2016.

[16] Faife C. (2017)., "Bitcoin Hash Functions Explained." [Online]. Available at https://www.coindesk.com/bitcoin-hash-functions-explained/, Accessed: 07.11.2021.

[17] P. Vigna and M. Casey, *The Age of Cryptocurrency: How Bitcoin and Digital Money Are Challenging the Global Economic Order*. St. Martin's Publishing Group, 2015. [Online]. Available: https://books.google.lk/books?id=Kv8CBAAAQBAJ

[18] Miles C. (2017)., "Blockchain security: What keeps your transaction data safe? Security by the blocks." [Online]. Available at https://www.ibm.com/blogs/blockchain/2017/12/blockchain-security-what-keeps-your-transaction-data-safe/, Accessed: 07.11.2021.

[19] Tayo A. (2017)., "Proof of work, or proof of waste?" [Online]. Available at https://hackernoon.com/proof-of-work-or-proof-of-waste-9c1710b7f025, Accessed: 07.11.2021.

[20] X. Li, P. Jiang, T. Chen, X. Luo, and Q. Wen, "A survey on the security of blockchain systems," *Future Generation Computer Systems*, vol. 107, pp. 841–853, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X17318332

[21] Buterin V. (2018)., "Proof of Stake FAQ." [Online]. Available at https://github.com/ethereum/wiki/wiki/Proof-of-Stake-FAQ, Accessed: 07.11.2021.

[22] Jake F. (2021)., "Blockchain Technology - Hyperledger." [Online]. Available at https://www.investopedia.com/terms/h/hyperledger.asp, Accessed: 07.11.2021.

[23] L. Luu, D.-H. Chu, H. Olickel, P. Saxena, and A. Hobor, "Making smart contracts smarter," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 254–269. [Online]. Available: https://doi.org/10.1145/2976749.2978309

[24] J. J. Sikorski, J. Haughton, and M. Kraft, "Blockchain technology in the chemical industry: Machine-to-machine electricity market," *Applied Energy*, vol. 195, pp. 234–246, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261917302672

[25] K. Christidis and M. Devetsikiotis, "Blockchains and smart contracts for the internet of things," *IEEE Access*, vol. 4, pp. 2292–2303, 2016.

[26] K. Sultan, U. Ruhi, and R. Lakhani, "Conceptualizing blockchains: characteristics & applications," *arXiv preprint arXiv:1806.03693*, 2018.

[27] Z. Yan, P. Zhang, and A. V. Vasilakos, "A survey on trust management for internet of things," *Journal of network and computer applications*, vol. 42, pp. 120–134, 2014.

[28] M. Rezaei, M. A. Shirazi, and B. Karimi, "Iot-based framework for performance measurement: A real-time supply chain decision alignment," *Industrial Management & Data Systems*, 2017.

[29] I. Lee and K. Lee, "The internet of things (iot): Applications, investments, and challenges for enterprises," *Business Horizons*, vol. 58, no. 4, pp. 431–440, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0007681315000373

[30] S. Preradovic and N. Karmakar, *Multiresonator-Based Chipless RFID - Barcode of the Future*, 03 2014.

[31] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[32] M. Tu, M. Lim, and M.-F. Yang, "Iot-based production logistics and supply chain system - part 2: Iot-based cyber-physical system: a framework and evaluation," *Ind. Manag. Data Syst.*, vol. 118, pp. 96–125, 2018.

[33] IBM 2017 a., "One truth across the supply chain network." [Online]. Available at https://www.ibm.com/blockchain/supply-chain/, Accessed: 07.11.2021.

[34] S. Jabbar, H. Lloyd, M. Hammoudeh, B. Adebisi, and U. Raza, "Blockchain-enabled supply chain: analysis, challenges, and future directions," *Multimedia Systems*, pp. 1–20, 2020.

[35] J.-H. Tseng, Y.-C. Liao, B. Chong, and S.-w. Liao, "Governance on the Drug Supply Chain via gCoin Blockchain," *International journal of environmental research and public health*, vol. 15, no. 6, p. 1055, 2018.

[36] Y. Madhwal and P. B. Panfilov, "Blockchain and supply chain management: Aircrafts'parts'business case." *Annals of DAAAM & Proceedings*, vol. 28, 2017.

[37] K. Behnke and M. Janssen, "Boundary conditions for traceability in food supply chains using blockchain technology," *International Journal of Information Management*, vol. 52, p. 101969, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0268401219303536

[38] G. Ortmann and R. King, "Research on agri-food supply chains in southern africa involving small-scale farmers: Current status and future possibilities," *Agrekon*, vol. 49, no. 4, pp. 397–417, 2010. [Online]. Available: https://doi.org/10.1080/03031853.2010.526428

[39] A. Kumar, R. Liu, and Z. Shan, "Is blockchain a silver bullet for supply chain management? technical challenges and research opportunities," *Decision Sciences*, vol. 51, no. 1, pp. 8–37, 2020.

[40] T. McConaghy, R. Marques, A. Müller, D. De Jonghe, T. McConaghy, G. McMullen, R. Henderson, S. Bellemare, and A. Granzotto, "Bigchaindb: a scalable blockchain database," *white paper, BigChainDB*, 2016.

[41] F. Tian, "A supply chain traceability system for food safety based on haccp, blockchain & internet of things," in *2017 IEEE International conference on service systems and service management*, 2017, pp. 1–6.

[42] Tian, "An agri-food supply chain traceability system for china based on rfid & blockchain technology," in *2016 13th international conference on service systems and service management (ICSSSM)*. IEEE, 2016, pp. 1–6.

[43] M. V. Kumar, N. Iyengar *et al.*, "A framework for blockchain technology in rice supply chain management," *Adv. Sci. Technol. Lett*, vol. 146, pp. 125–130, 2017.

[44] A. Antonopoulos, *Mastering Bitcoin: Programming the Open Blockchain*. O'Reilly Media, 2017. [Online]. Available: https://books.google.lk/books?id=tponDwAAQBAJ

[45] Prakhar Sharma 2017., "The Economics of Bitcoin Block Size." [Online]. Available at https://blog.blockonomics.co/the-economics-of-bitcoin-block-size-e9575272c3ee, Accessed: 07.11.2021.

[46] I. Eyal, A. E. Gencer, E. G. Sirer, and R. Van Renesse, "Bitcoin-ng: A scalable blockchain protocol," in *13th {USENIX} symposium on networked systems design and implementation ({NSDI} 16)*, 2016, pp. 45–59.

[47] G. Wood *et al.*, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum project yellow paper*, vol. 151, no. 2014, pp. 1–32, 2014.

[48] H. Zhang, Y. Wen, H. Xie, and N. Yu, *Distributed hash table: Theory, platforms and applications*. Springer, 2013.

[49] Q. Zheng, Y. Li, P. Chen, and X. Dong, "An innovative ipfs-based storage model for blockchain," in *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 2018, pp. 704–708.

[50] "avg-block-size," [Online]. Available at https://www.blockchain.com/charts/avg-block-size, Accessed: 10.12.2020.

[51] "n-transactions-per-block," [Online]. Available at https://www.blockchain.com/charts/n-transactions-per-block, Accessed: 10.12.2020.

[52] "n-transactions-total," [Online]. Available at https://www.blockchain.com/charts/n-transactions-total, Accessed: 10.12.2020.