Bachelor thesis

Independent degree project

Computer Engineering

The blockchain based system to guarantee the data integrity of IIoT

Yifei Shen

Mittuniversitetet
MID SWEDEN UNIVERSITY

Campus Härnösand Universitetsbacken 1, SE-871 88. **Campus Sundsvall** Holmgatan 10, SE-851 70 Sundsvall.
Campus Östersund Kunskapens väg 8, SE-831 25 Östersund.
Phone: +46 (0)771 97 50 00, Fax: +46 (0)771 97 50 01.

MID SWEDEN UNIVERSITY
Department of Information Systems and Technology

Examiner: Ulf Jennehag, ulf.jennehag @miun.se
Supervisor: Stefan Forsström, Stefan.Forsstrom@miun.se
Author: Yifei Shen, yish1700@student.miun.se

Main field of study: Computer Engineering
Semester, year: Spring, 2018

Abstract

With the advent of big data era, there is countless data produced from various kinds of machines every second. These data are used to help people to do studies, produce goods, improve the efficiency of the industrial development, and so on. Especially considering the Internet of things which connects all the smart machines together, the importance of data integrity gets unprecedented attention from us. In order to keep the data integrity, blockchain comes to its birth using its own structure to guarantee the data integrity efficiently. This project is focus on the simulation of data system based on the blockchain in the background of industrial internet of things which is shown in a form of a website with coding language jade and the environment is Node.Js. The main task is to measure the time consumed in the process of block mining under different parameters include data length, nonce, difficulty, sensor number in order to find the correlation between block mining time and different parameters. Thus, the rule about the impact of different parameter on the mining time are concluded. According to the rule I found, the evaluation about scalability, efficiency and safety of this system are given and I also summarize two formulas to calculate the efficiency of block mining. Ethical consideration and future work are addressed in the conclusion part.

Keywords: SHA-256, blockchain, data integrity, IoT, IIoT

Acknowledgements

I want to thank to my supervisor Forsström Stefan, he guided my thesis writing from the beginning to the end in a very patient way and positive attitude. He helped me to draw the blue picture about my whole project and gave many useful suggestions to improve my project in detail. He also taught me the structure of the thesis, the key point of each part which helps me save the time during the writing. Every week we had a meeting I sent my thesis to him and he always reviewed my thesis carefully and gave me useful and feasible suggestions on how to improve the writing and what to do next. Actually, my thesis is finished by his help step by step, without his help, it is impossible for me to finish the project and the thesis. I am an exchange student, my English is not good enough but he is always patient to understand what I am talking about and give me encourage when I still have a lot of work to do before the deadline. I think Forsström Stefan is a really good supervisor in my mind.

I also want to thank my supervisor in China, Lu ting. Although I have an exchange studies in Sweden which is so far from China and there is time difference of 6 hours, she always cares about the progress of my thesis through the internet and gives me suggestions on how to improve the thesis, treating me just like my classmates in China which makes me touched.

Furthermore, I want to express my thanks to my dear friend Xinyun Chen, he helped me a lot with my studies and always give me encouragement when I feel sad no matter how far we are.

Last but not the least, I especially love my friends in Mid Sweden university which keeps me company like a family and thanks to my father and my mother who always give me support and tell me never give up. I cannot finish the project without their love and support, thanks a lot.

Table of Contents

Terminology

Abbreviations

IoT	Internet of Things
IIoT	Industrial Internet of Things
SHA-256	Secure Hash Algorithm-256
RFID	Radio Frequency Identification
IaaS	infrastructure as a service
PaaS	platform as a service
SaaS	software as a service
ITU	International Telecommunication Union
ICT	Information and communication technology
T2T	Thing to thing
H2T	Human to thing
H2H	Human to human
WSN	Wireless sensor network
RFID	Radio-frequency identification
FTTX	Fiber to the x FTTX
EPC	Electronic Product Code
AI	Artificial intelligence
QR code	Quick Response Code
AGV	Automated Guided Vehicles
PoW	Proof of work
4G technology	Fourth generation of broadband cellular network

5G	Fifth generation of broadband cellular network technology

1 Introduction

With the high-speed development in economic and scientific technology, numerous data for different proposes is produced by countless smart devices every second. Big Data Era makes the world more efficient not only in the scientific area but also in our daily life, such as medicine, biology, industry, traffic and so on. But every coin has two sides. Data management is absolutely a difficult job for any system. The integrity of data is hard to maintain but it is a key in the data flow. Any problem which broke the data integrity will cause an irreparable loss in accuracy, reliability and money. There is no denying that some feasible solutions pour into scientists' mind to improve the data integrity such as error detection and error protection, but someone deems that the database itself can guarantee the integrity, then it will make the process simple and efficient especially in the industry where a serious of machines need to work step by step. To achieve this goal, a new structure database comes to its birth-- block chain.

1.1 Background and problem motivation

According to a reliable forecast[1] , wireless data traffic will increase by a 1000 fold by 2020 with more than 50 billion devices connected to the Internet. Many of these will be small embedded devices with sensors and actuators, enabling new types of intelligent applications. All these smart devices are connected and they collaborate together form what we call the Internet-of-Things (IoT). Because of the scale of devices and the sheer number of sensors, actuators and the information they produce, we will have inevitable problems in the future if we want to enable secure communication between all these devices connected to the IoT. Especially considering the Industrial IoT, where industries will have all their equipment connected as well. These industries also need to ensure that the integrity of the sensor values is maintained at all times, to avoid tampering attacks and failures.

One emerging technology[2] to solve this is to use block chains. Block chains work like a distributed open database which consists a set of ordered blocks secured from tampering and revision. With block chains, the data calculated by certain sensors or actuators from the industry can be safely stored in the block without any worry about the correctness of data which can guarantee the integrity of whole industrial system.

1.2 Overall aim

The purpose of this project is to implement and evaluate an industrial data integrity system using block chain technologies, in order to ensure that no outside force is able to alter the IIoT data. Hence, the task is to survey, implement, and quantitatively evaluate this system and measure its performance in a typical industrial IoT scenario. The problem I will handle is to determine the benefits, drawbacks, and overall performance of the system, as well as how it effects the scalability, response times, and overall quality and effectiveness.

1.3 Concrete and verifiable goals

Therefore, my goals which I want to achieve on the problem are listed as following:

(1) Survey the area of IOT, IIoT and Blockchain technologies in order to gain the knowledge about the structure of Blockchain, its applications and benefits or risk to use it and be familiar with the background of IIOT.

(2) Determine a typical IIoT scenario where data integrity is required.

(3) Implement a data integrity system for the IIoT.

(4) Perform measurements of block mining time under different parameters and give the analysis in accordance to the established scenario.

(5) Evaluate the end results in fields of scalability, efficiency, safety how it effects the scalability, response times, overall quality and effectiveness. and also give a propose future work.

1.4 Scope

In this thesis I will focus on simulating an industrial data integrity system using block chain technologies which can show a better performance of data integrity than the traditional database. And it will also include a survey about the area of Internet of things and industrial internet of things as the background of the whole system. In the end of the thesis, I will quantitatively evaluate this system and measure its performance according to the data stored in block chain. However, the area of improving a certain encryption or decryption algorithm will be put out of the scope of this work, I will use the encryption algorithm from the library, because the aim of this thesis is to simulate a whole blue picture of a system based on the techniques of blochchain not to improve the performance of encryption or decryption.

1.5 Outline

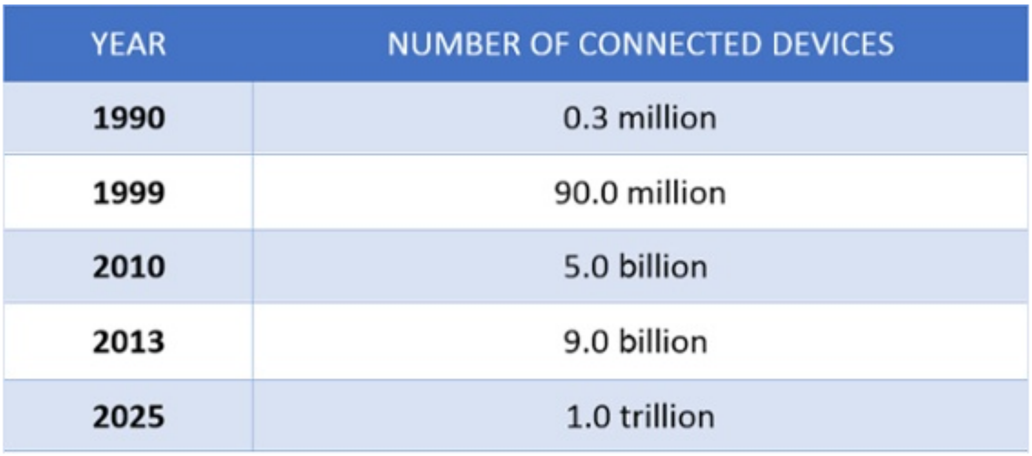
The thesis is outlined as followed. Chapter 2 expounds the related knowledge about Internet of things, Industrial internet of things and blockchain. Chapter 3 describes the feasible methods I used to conquer the predefined five goals respectively. The subject of Chapter 4 is the detail implementation of the simulation of IIOT system based on blockchain. Chapter 5 includes the measurement of the results about block mining time under different parameters which are presented in the form of table and line chart and also includes the screenshots of the whole system and the explanation about different functions of different website pages. In the end, chapter 6 makes a conclusion about this thesis alongside the ethical consideration and future work.

2 Theory

In this chapter, all the knowledge of background and techniques mentioned in the paper are introduced to help have a deep understanding of the problems which require to be solved and give some inspirations about the methodology. Chapter 2.1 covers the development, characteristics, key techniques and applications of internet of things. Chapter 2.2 presents the architecture, goals and challenges, typical scenario of industrial internet of things. Blockchain is the theme of chapter 2.3 which includes the working mechanism, applications and pros and cons of blockchain. In chapter 2.4, some related work about the thesis are described.

2.1 Internet of things

With the quick development of internet, a new generation of network technology has come to birth, Internet of things， which is on the basement of the internet [3] but actually it is an extension and a new trend of future, gaining vast attention from all over the world and various of industries. The true value behind internet of things is that it makes possible that billions of smart devices can be connected together[4] and communicate with each other without or with minimal human intervention. According to the reliable survey by company HP, we can know that the number of connected devices is increasing in a high speed and it will be up to 1.0 trillion in 2025. On one hand, the development of internet of things brings great benefits to our life. For example, more intelligent transportation system, health care becoming more convenient, the high efficiency in manufacturing and so on. On the other hand, [5] the large number of devices connecting to the internet increases the difficulty of controlling and managing the data and the safety problems which brings challenge to the IOT.



YEAR	NUMBER OF CONNECTED DEVICES
1990	0.3 million
1999	90.0 million
2010	5.0 billion
2013	9.0 billion
2025	1.0 trillion

Figure2-1 The number of connected devices in different year

2.1.1 The development of IoT

The term "Internet of Things" was earned by the founders of the original MIT Auto-ID Center (Kevin Ashton in 1999 and David L. Brock in 2001). MIT also established the "Auto-ID" and give the basic definition of the Internet of Things: all the things can be connected through the internet. In the early stage of IoT, it is only based on the Radio Frequency Identification (RFID) technology under the background of logistics system which achieved the intelligent management on the logistic system. With the development of technology and application, various kinds of powerful techniques come into use to support internet of things like embedded intelligence, nanotechnology, robot technology, communication technology and so on.

The report from International Telecommunication Union points out that human being is in a new era of communication. The goal of information and communication technology has developed. They are not only satisfying human to human communication but also realizing the connection between human and things, things and things, and ubiquitous things. The age of networked communications is approaching. The Internet of Things makes it possible that we can gain a new dimension of communication (as shown in Figure 2-2), which expands the range of connection from any time, any place, any person to any object connected. The interconnection of T2T things to things, H2T human to things and H2H human to human formed the internet of things.

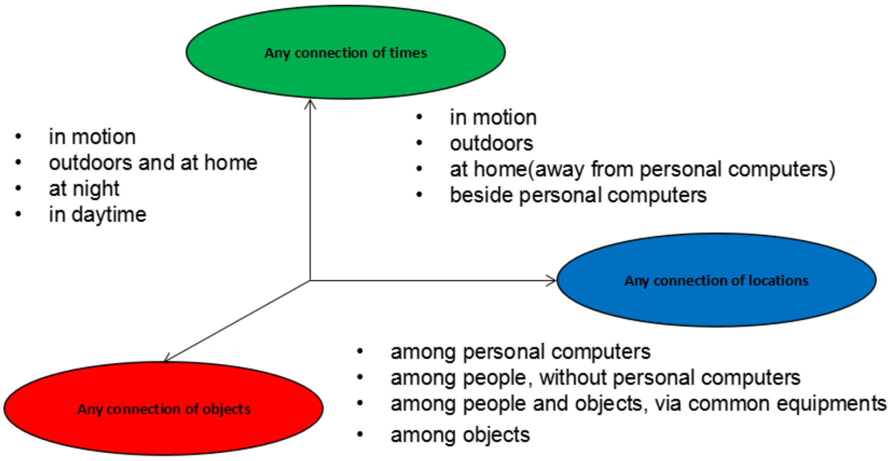

Any connection of times
• in motion
• outdoors and at home
• at night
• in daytime
• in motion
• outdoors
• at home(away from personal computers)
• beside personal computers
Any connection of locations
Any connection of objects
• among personal computers
• among people, without personal computers
• among people and objects, via common equipments
• among objects

Figure2-2 A new dimension of communication under the background of IoT

At present, many developed and developing countries have adopted the Internet of Things as a new industry and have introduced strategies to develop. In the United States, IBM proposed that the framework of Smart Planet which is used to change the way government, companies and people interact in 2009. In Europe, the concept of the Internet of Things has been highly valued and strongly supported by the European Commission which has been formally established as a strategic development plan for European information and communication technologies. There is no doubt that the Internet of Things is changing the world.

2.1.2 Three basic characteristics

The basic characteristics of the Internet of Things can be summarized as comprehensive perception, reliable transmission, and intelligent processing.

(1) Comprehensive perception: With the help of RFID, WSN (Wireless sensor network), sensors and other technologies in order to mine the deep information from the object.

(2) Reliable transmission: Through the telecommunication network and the Internet (like 4G/5G, WiFi, FTTx …) guarantee the successful transmission of object's information and data which can provide the reliance on information exchange and sharing.

(3) Intelligent processing: Use intelligent computing technologies such as cloud computing and fuzzy identification to analyze and process massive amounts of data and information, and implement intelligent control of objects.

These three characteristics also represent three layers of IoT (perception layer, network layer and application layer) respectively (which is shown in figure 2-3). Some people compare theses three characteristics to adult senses, nerves, and brains. People's senses are used to obtain information (comprehensive perception), nerves are used to transmit information (reliable transmission), and the brain is used to process information (intelligent processing). Thus, people have the intelligence to adapt to changes in the external world. With these three basic characteristics, the Internet of Things just works like human beings and this gives the reason why the Internet of Things can also change the world just like human beings.

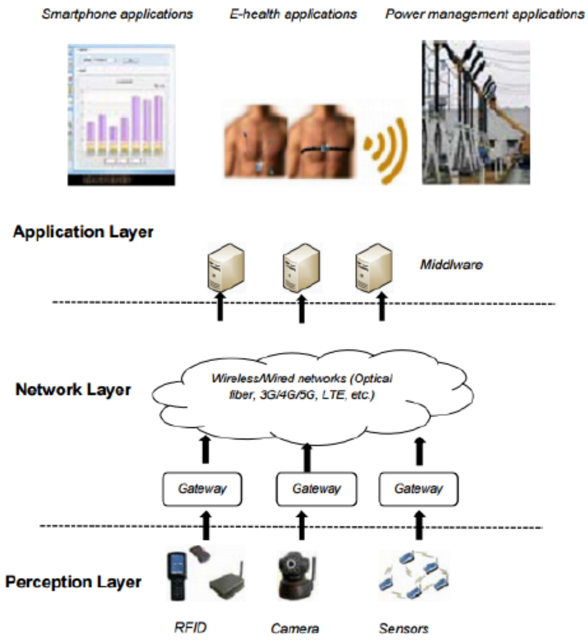


Figure2-3 Three layers of Internet of things

2.1.3 Key techniques

There are five key technologies which plays the important roles in the Internet of Things: RFID technology, sensor technology, wireless network technology, artificial intelligence technology, and cloud computing technology. These technologies emerged at different times and therefore, they still are at the different levels of development. For example, Read Frequency identification technology is mature enough now while the cloud computing technology is still in the state of developing. The following is the detail about these five techniques:

(1) RFID technology: RFID (Read Frequency identification) is a kind of non-contact automatic identification system which identifies items by radio frequency. RFID is composed of EPC (Electronic Product Code), communication induction antenna and the reader. Electronic product code is the unique identity of the physical objects of all over the world just like the function of URL in the Internet. Communication induction antenna plays a central role in the process of realizing data communication because it is responsible for passing the signal from the reader to the tag, and at the same time, passing the feedback signal from the tag to the reader. Reader can not only read the tag information but also overwrite it.Because of the advantages of RFID like its non-contact, time-saving, durability and safety, RFID is used widely in many areas such as food traceability, military equipment, security and anti-theft.

(2) Sensors are primarily responsible for receiving the “speaking content” from the object. It combines both multidisciplinary modern science and engineering technology. The technology includes collecting the information, transforming information and recognizing the key information. Until now, the detection of sensors has not reached a high level in terms of the quantity, stability and reliability, it is still one of the important bottlenecks in the area of Industrial Internet of Things.

(3) Wireless network technology: The object in the Internet of Things need to communicate with people without barriers. It will inevitably be inseparable from high-speed, high-volume data transmission wireless networks. Wireless networks include both global voice and data networks that allow users to establish long-range wireless

connections, as well as close-range Bluetooth, infrared, and Zigbee technologies.

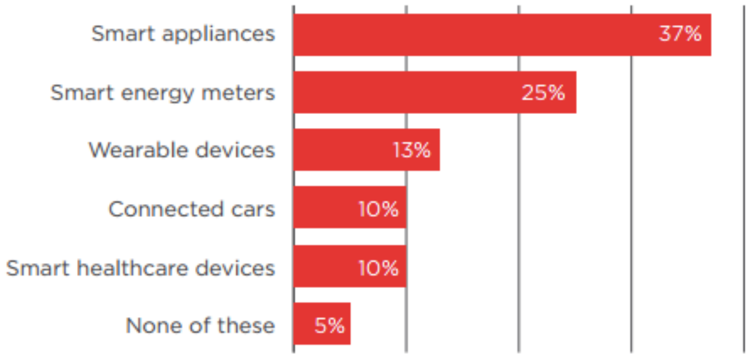
(4) Artificial intelligence technology: Artificial intelligence technology is a new emerging discipline. The main task of artificial intelligence in the current Internet of Things is mainly to identify the content read from the object. It mainly simulates, emulates, and thinks like a human being which achieves automatic computer processing.

(5) Cloud computing technology: The development of the Internet of Things cannot be separated from the support of cloud computing technology. The cloud computing platform can be used as the brain of the Internet of Things to implement storage and calculation of massive data which can solve the trouble of limited computing and storage capabilities of terminals in the Internet of Things.

The main service form of cloud computing is infrastructure-as-a-service(IaaS) represented by Amazon, Platform as a service(PaaS) represented by Saleforce as well as software as a service(SaaS) represented by Microsoft. The local terminal only needs to send a request message over the internet and there will be thousands of computers in the "cloud" to provide you with the necessary resources and feedback the result.

2.1.4 Real world application of IoT

Nowadays various applications of IoT appear in order to make our life become more convenient and efficient. A survey conducted by KRC Research in UK, US, Japan and Germany who are the early adopters of IoT has revealed what kind of device is most popular among the customers. We can find the answer that smart applications which really changed our life style. I list some typical and successful real world applications of the Internet of Things and show their excellent functions.



Source: GSMA Report

Figure2-4 A survey conducted by KRC Research in UK, US, Japan and Germany

(1) Smart house: Smart house makes it possible that all the electronics products, communications products and information appliances in the home can communicate and exchange data which achieves the interconnection of various types of electronic products in the home network, and achieves anytime, anywhere on the smart device control. When we leave home, there is no need for us to switch the electronic products. Temperature and humidity are controlled by the smart house itself which reduces family expenses invisibly and makes our life in an easier and more convenient way.

(2) Safe City Construction: Utilize the sensors deployed in the streets to realize image-sensitive intelligent analysis and interact with the police to realize the linkage between probes and probes, probes and people, probes and alarm systems so as to build a harmonious and safe urban living environment.

(3) Food safety： A two-dimensional code is attached to each animal in the farm. The two-dimensional code will be kept on the meat sold by the supermarket. When consumers want to buy the meat, they can first read the two-dimensional code through the mobile phone and know the history of livestock growth to check if the meat is safe enough to eat in order to keep the food safety. In China, one billion animals already have this quick response code.

2.2 Industrial internet of things

We know the benefits that the Internet of Things has brought to us like increasing efficiency, improving the quality of our life and providing a better experience. But what is Industrial Internet of Things?

IIoT is the abbreviation of Industrial Internet of Things which makes it possible that the communication between machine to machine in the industry without human intervention in order to improve the quality of production and increase the profit, which the core of “Industry 4.0”.

The main difference between IIoT and IoT can be described as two points: one is about the different focus, the main focus point of IIoT is the application of production and services, which aims at producing higher value equipment and assets such as energy, transportation and industrial control. Actually, Industrial IoT involves projects with clear business benefits. At the same time, it has higher requirements for operational safety like data integrity, while IoT pays more attention to the consumption area such as home applications, smartphones, TVs, the difference between Industrial Internet of Things and Consumer Internet of Things is shown in the figure 2-5. The other point is that industrial IoT belongs to IoT, in fact it is a subcategory of the broader Internet of Things. The research direction of IIoT is mainly about industrial applications such as agriculture and manufacturing. There is no doubt that IIoT has a massive potential.

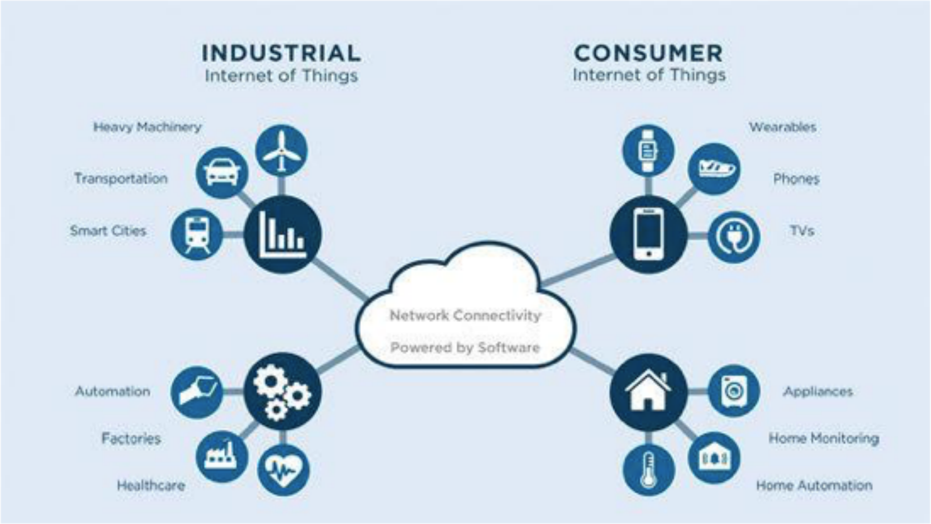


Figure2-5 Different range of Industrial internet of thins and Consumer internet of things

2.2.1 The architecture for Industrial Internet of Things

The architecture of Industrial internet of things can be described in two ways, one is from the hardware and the other is from the software.

From the hardware, [6] it can be divided into four key components, they are physical layer, networks, industrial clouds and intelligent terminals.

(1) Physical layer: It is the most basic component in the whole system which have the direct impact on the industrial production and implementation. Many physical devices compose the physical layer like sensors, monitors, Automated Guided Vehicles, manufacturing equipment and so on. The raw products are transported to a serious of destinations by AGV according to their RFID which shows the key parameter of the products and the data produced during the manufacture. The sensors and monitors are all responsible for recording the product's information to ensure the safety and integrity of the whole process.

(2) Networks: Networks plays an important role in transforming the information from sensors to control system to cloud applications. There are several techniques support the networks such as MCNs (Mobile communication networks), industrial Ethernet,com-Wi-Fi e.g., MOXA-Nport-W2150A, USB-WiFi module IEEE 802.11, 4G,5G and so on. The development of networks speeds up the transforming time in a big scale and provide the stability.

(3) Industrial clouds: [7] [8] The industrial clouds are responsible for computing the algorithm in an optimal way, making the decision and storing the massive data which produced during the computation.

(4) Intelligent terminals: Through the information displayed by intelligent terminals (for example smartphones, LED screen, web pages), workmen, users and manager can interact with their industrial system in an efficient, easy and visual way. Intelligent terminals make the manufacturing process more efficient and easier to manage.

From the respect of software, the Industrial Internet of Things has three layers: physical layer, control layer, and application layer. The physical layer is composed of various devices which provide inter-communication from machine to machine, human to machine. The control layer is like a middleware between physical layer and application layer. The main task of this layer is to collect the required data according to the demand of application layer and transform the data to the application layer through channels, gateways, collectors. The application layer provides a lot of APIs which are designed by the developers in order to implement in some interesting applications. With the help of API, the data produced by hardware can be shared easily so that the cost of the whole manufacture can be reduced and provide an efficiency way to manage the industrial system.

2.2.2 Challenges and goals of IIoT

Although the industrial Internet of things brings great benefits to the industry in many areas, it also faces many challenges in the aspect of security, privacy, safety and so on. Some lawbreakers have extended their claws to the Industrial Internet of Things and they attempt to cause a loss in the country’ s economic, even threaten the lives of the people by destroying it. One of a fearful and influential attacks occurred in the USA in 2003 which made two critical monitoring system of a nuclear planet invalid by the Slammer worm. Similarly, it is hard to imagine that if one day the national grid is attacked by the hacker, the entire city will be unable to supply electricity and it will cause how much loss and it will also bring much panic to citizens to make the government untrusted.

From the above example, we can understand how important the security of the Industrial Internet of Things and privacy are. There are three goals set to the IIoT to guarantee the security and privacy:

(1) Availability: During the producing process, all the components which will be used during the manufacture like sensors, monitors, channels should be available all the time in order to avoid the delay which may cause wrong intermediate calculation results or the loss of productivity.

(2) Integrity: Integrity means data integrity and component integrity. This includes the protection against tamper like someone deliberately tampers with the computing result which may not fulfill the required standard of quality. Another one is component integrity, all the components should be participated together and works in order according to the design plan in the producing process that guarantee the quality of products.

(3) Intelligence: All the components should be labeled invisibly or visibly by machines. The label works like a book to tell the whole life of its owner. According to these labels, we can know the history of the components about the progress of its operation. If errors occur, we can use these labels to trace the entire production process and it is easy for us to find the place where the mistake is. These labels guarantee the security without physically attack.

2.2.3 Typical scenario of IIoT

There are several typical scenarios in industry which has implemented the IIoT in order to improve the quality and productivity. The following are three typical scenarios of IIoT:

(1) Manufacturing supply chain management: Enterprises can use the Internet of Things technology to timely acquire the detail about the raw material procurement, inventory, sales, and other information. Through the big data analysis, it is possible for them to predict the price trend of raw materials, supply and demand relations. Furthermore, it helps to improve and optimize the supply chain management system in order to improve the efficiency of the supply chain and cut costs.

(2) Optimization of the production process: The ubiquitous sensing characteristics of the industrial internet of things improve the ability and level of production line process detection, real-time parameter acquisition, and material consumption monitoring. Through the analysis and processing of data, intelligent monitoring, intelligent control, intelligent diagnosis, and intelligent decision can be realized, intelligent maintenance, increase productivity in order to reduce energy consumption. For example, Iron and steel enterprises use various sensors and communication networks to realize real-time monitoring of the width, thickness and temperature of processed products in the production process which helps to improve product quality, and optimize production processes.

(3) Industrial safety production management: Safety production is the primary consideration in modern industry. The technology of Industrial internet of things can be used to install the monitor in some dangerous production environment like mining equipment, oil and gas to enhance the existing network monitoring platform into a

systematic, open, and diverse integrated network monitoring platform, which can effectively safeguard the safety of industrial production.

2.3 Blockchain

Blockchain is a new emerging word. It is the first time to be proposed by a person named Nakamoto Satoshi, which is considered as a breakthrough technology. Actually, it is a kind of distributed system, unlike the traditional central database, which is not owned by anyone but a public one without intermediaries. The data stored in the current blocks can be searched by anyone which greatly increases its openness and transparency so that the trust between users can be established.

The structure of blockchain is a chain of blocks which is connected with each other in order. Once one block is created, no one can change the block content anymore and the process of block creation is according to a majority consensus mechanism. The creation of the block is a really hard and time-consuming, we call it block mining which needs to calculate a long length of hash according to the stored data. But one coins have two sides, because of the difficulty of block mining, it increases the difficulty of tamper. The hacker is almost impossible to attack the blockchain-based database until now.

The famous application of blockchain is not doubt the Bitcoin. Actually, the fields where blockchain has been used are not only limited in the business or finance but also extends to government, industry internet of things, music, internet applications and so on. Following text are details about the mechanism of blockchain operation, three main kinds of blockchain and its applications in different area.

2.3.1 How blockchain works

It is not difficult to learn how blockchain works as long as you understand four basic components of the blockchain which is shown in figure 2-6.

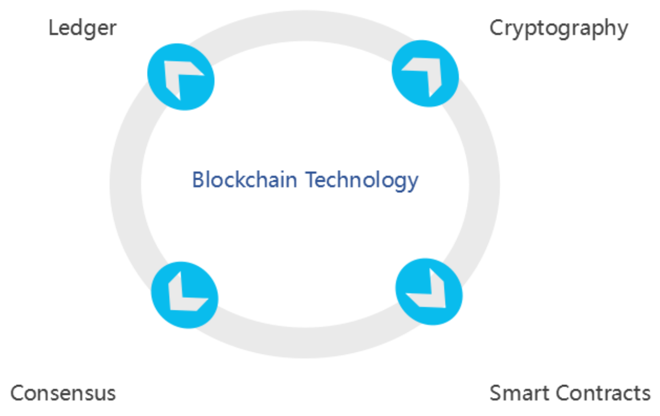


Figure2-6 Four Components of blockchain technology

(1) Network of Nodes: These nodes are given the permission to read the data and send their message to the blockchain. Meanwhile, they are all responsible for checking the feasibility of creating the new blocks. When one node wants to store the message, it must calculate the hash result which need to satisfy a certain degree of difficulty (Proof of work) according to the current data and then the block can be mined after the calculation which will be added to the end of the blochchain later. The basic content of one block is pre hash of the previous block, timestamp about the creation time, data, nonce and the hash result.

(2) Distributed database system: All the nodes in the blockchain have the same copy of the main blockchain without delay, once a new block is added to the chain, all the copy will be renewed at the same time.

(3) Shared record: All the records are renewed at the same time which make the blockchain open and transparent.

(4) Cryptography: the main encryption algorithm used in the blockchain is SHA-256 in order to calculate the hash result during the block mining which is a key point during the mining process. A nonce requires to be calculated according to the difficulty the blockchain set.

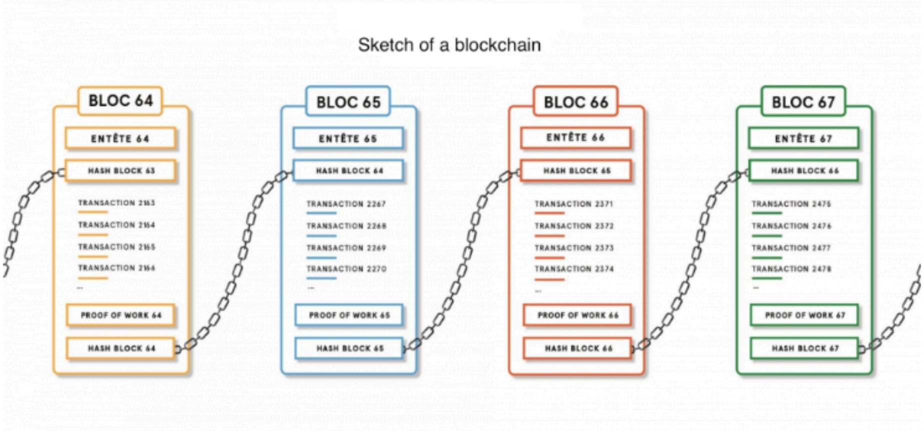


Figure2-7 The structure of the blockchain

According to the different range of participants who can use the blockchain, now there are three different kinds of blockchain.

(1) Public chain: In this chain there is no limitation about opportunities to the interaction, every node can transfer and read the data from the chain without any permission. Because it maximizes openness and transparency, it is the most popular kind of blockchain among the market and media. The typical representative of publich chain are Bitcoin and Ethereum.

(2) Private chain: The private chain is controlled by limit number of people, the information is not public, it has the strictest limitation among three kinds of blockchain.

(3) Alliance chain: The scope of its user's limitations is between the public chain and private chain. The use of the alliance chain must be limited with permission, and relevant information will be protected. The typical representative of alliance chain is supply chain institution and bank union.

2.3.2 Applications of blockchain

There are many applications which is based on the blockchain, they are in a wide range of fields. The following text introduces some typical application which are widely used nowadays.

(1) Bitcoin: It is the first distributed virtual currency which can be used in the whole world. The composition of the bitcoin is a serious of

complicated codes which does not rely on a particular currency institution to issue.

(2) Ethereum: The main function of Ethereum is providing a distributed computing platform featuring smart contract(scripting) functionality. The currency used in the Ethereum is named by Ether which is based on the blockchain technique.

(3) Hyperledger: The Hyperledger is a foundation which focus on the exploration of blockchain techniques. The goal of this organization is to create a new distributed ledger which can be the standard for the blockchain techniques, allowing more applications can be easily built on the blockchain technology.

(4) Government voting: The voting process can be compared as the transaction between voters and candidates while the currency is not the money but the number of votes. The process of voting by using the structure of blockchain is transparent and open which guarantee the fairness of the election.

(5) Internet of things: The blockchain can be used in IOT to improve the interoperability between devices that the messages sent from other machines can be trusted and the data integrity that prevent the data from tamper.

2.3.3 Advantages and disadvantages

Everything has two aspects. The blockchain has many delightful advantages, but it also has its own disadvantages. When we consider whether to choose to use it, we should consider its strengths and weaknesses comprehensively to make a rational and scientific judgment.

Advantages of blockchain can be summarized as:

(1) Immutability. The structure of blockchain guarantees the data integrity in order to prevent the data be attacked by the hackers and improves the safety factor of the system.

(2) Transparency and openness are always kept during the transaction which creates the trust between the users. A wide range of users who can from any countries as long as they are permitted before they use the alliance chain or private chain.

(3) Decentralization. It can still run when some terminals shut down or go wrong because there is no central authority. It is always available.

(4) Automation. The activities can be automatized as long as they have signed the smart contacts.

Disadvantages of blockchain can be summarized as:

(1) The process of the block mining is not only time-consuming but also power-consuming. The current annual consumption of bitcoin mining is estimated to be approximately 48.37 TWh. The time spent on creating bitcoin block is around 10 to 60 mintues. At the same time, a large number of hardware have to be involved in the computing which cost a lot.

(2) A lot of pressure on the storage of the distributed copies of blockchain. According to the survey, the storage of bitcoin is up to 105 Gbytes.

(3) Copies of blockchain provides the transparency but may harm some users' privacy because every node can check the content of each block.

2.4 Related work

This chapter presents some related work about this thesis and describe them in a brief way. Here are three related work focus on blockchain based system in the background of industrial internet of things. The implementation of first paper and the third one are all about the smart city.

In the paper [9] Evaluating the Efficiency of Blockchains in IoT with Simulations, they simulated a blockchain based system and evaluated the efficiency from the aspect of number of nodes, execution time and average power. They used Ethereum as the blockchain implementation and used Raspberry Pi 2 as the platform. The hash algorithm they used is SHA-3. They did the measure of the performance of the system with ABSOLUT toolset which can test the power used in mining. They also took the network latency into account, the result they found network latency does not impact the mining efficiency if it is under 100ms and mining time is liner related to the number of node.

Another related work [10] is A Lightweight Scalable BlockChain for IoT Security and Privacy which implemented the LSB in the smart city, LSB used a time-based consensus algorithm instead of PoW which can guarantee the security and privacy which is the emphasis in their implementation.They only evaluated the efficiency of blockchain in energy consumption.

Manish Lamichhane [11] also did some efforts on implementing Ethereum Blockchain in the smart waste management system which is also an implementation of smart city. In the end he tested the efficiency of the work with the test machine consisted of 4 CPU's and 8 GB RAM. and tested the mining time according to number of threads.

3 Methodology

This chapter aims at presenting the structure of whole work of the thesis. The purpose of this project is to set up an industrial data integrity system using block chain technologies. In order to achieve this goal, I list five problems (which is listed in chapter one) need to be solved. Now, I will give different feasible methods to address these problems respectively.

3.1 Survey on IIoT and blockchain

Goal one is to survey the area of industrial internet of things and blockchain technologies. The way to survey the area of industrial internet of things and Block chain technologies is reading literature through the Internet to find relative paper and journals about Industrial Internet of Things, Industrial Internet of Things and Block chain. The source of the most information is from google scholar and academic database provided by Mid Sweden University like IEEE Explore and ACM digital library. The key words I used in search includes IOT key techniques, IIOT definition, blockchain usages and so on. The total amount of the articles I read is more than thirty.

3.2 Scenario determination

Goal two is to determine a typical Industrial Internet of Things where data integrity is required. According to the theory I learnt in Industrial Internet of Things, I will determine a typical Industrial Internet of Things scenario. It is not difficult to be find a suitable scenario as long as I am familiar with industrial internet of things by reading relative literature. And then I can choose one scenario that data integrity plays an important role in the process of the manufacture. I will make the decision according to the type and the amount of data used in the scenario and takes into account if the integrity of the data is broken in this scenario how maliciously affect the productivity and efficiency of the factory in order to find that if it is necessary to use blockchain techniques in the management of data in this scenario.

3.3 Implementation and measurement

Goal three is to implement a data integrity system for the Industrial Internet of Things. After the choice of certain scenario, I will implement a data integrity system for this Industrial Internet of Things scenario. There will be three parts in the implementation: generation of data, data storing in blockchain-based database and reading data stored in blockchain. Of course, the focus will be on the data storing, I will choose naïve chain to build the dataset and use jade and html as coding language to achieve the blockchain-based dataset. The algorithm of hash used in naïve chain is SHA-256.The system is shown in the form of website with four pages. The system runs on Mac.

Goal four is to perform measurements in accordance to the established scenario. To check how well the system runs, I need to perform a serious of measurements in accordance to the established scenario. The measurement is about the block mining time under different parameters including data length, the interval of nonce, difficulty, number of blocks and number of sensors. Timer will be set in the code of the system to calculate the accurate block mining time. The amount of test data should be as much as possible so that the result will be more accurate. For the purpose of finding the correlation between block mining time and one of the parameters, the other parameters should be fixed in the test. To calculate the efficiency, mathematical operations should be performed on the results. For example, calculating the block mining rate is achieved by dividing the block mining time by the number of blocks. The measurement results will be shown in the form of tables and line charts using the excel.

3.4 Evaluation

Goal five is to evaluate the end results and purpose future work. In the end I will evaluate the system from three respects which are efficiency, scalability and security respectively. For the scalability, the conclusion about the degree of the different parameter impacted on the block mining time should be given according to the measurement. The best interval of each parameter should also be inferred with the correlation between the parameters and block mining time such as the most suitable difficulty of hash, the maximum number of sensors and the number of blocks. For the efficiency, the conclusion can be given by calculating by the average block mining time of one block under different parameters. As for the security, it is hard to evaluate, so I decide to simulate the scenario of tamper the data and calculate the time consumed in the process of making the temper undetected to check if the system is safe or not.

4 Implementation

4.1 Implementation of hash

The page named hash shows the function of encryption using SHA-256(Secure Hash Algorithm) method. SHA-256 is the most common and efficient cryptographic hash algorithm which is widely used in encryption of data in blockchain technique.

Cryptographic hash algorithm is a method that transforms the original data through a special algorithm into another type of searchable, fix-length new data. Through SHA-256 algorithm, the data will be transferred into 256-bit hash and it is suitable for anti-temper because of two key reasons. One is that it is a kind of one-way function which cannot be decrypted back. Actually you cannot find any hint about the original data through the encrypted result. Another reason that cannot be ignored is that any slight change in data will result in a noticeable difference in encryption results. Thus, it is almost impossible for different inputs to produce the same output.

The SHA-256 algorithm can be summarized as the following:

(1) Padding: Before we start the hash computation, we need to make sure that the length of message is multiple of 512 bits. The method is that we need to add '100…000' to the end of the original data and stop adding zero until the length has achieved multiple of 512 bits.

(2) Block decomposition: The padded data will be split into N 512-bit blocks (Q^1, Q^2, Q^3Q^1, Q^2, $Q^3 \ldots Q^n$) depends on its size and each 512-bit block will be divided into 16 32-bit words ($Q_0^i, Q_1^i, Q_2^i \ldots Q_{15}^i$), and then they will be extended into 64 32-bit words which can be labeled as $W_{0,}, W_1 W_2 \ldots W_{63}$according to the rule of SHA-2.

(3) Hash computation: First initialize 8 32-bit buffers ($H_1, H_2, H_3, H_4, H_5, H_6, H_7, H_8$) which can store the intermedia and final results. And then each block will be calculated 64 cycle iterative computation.

The overall implementation is a block-chain based IIOT system to guarantee the data integrity which is displayed in the form of a website. The main coding language is Jade and Html and the runtime environment is Node.js which is an excellent environment based on Chrome V8. There are four pages in this website. It shows the creation of a whole distributed system using blockchain step by step from one block to one blockchain made up of a number of blocks and finally become a distributed system composed by four distributed sensors, the number of sensors can be changes but four is the default.

In this system we can insert the data into blockchain and research the data stored in the blockchain according to the number of block. The advantage of using the blockchain is that if the data is stored into the block, it cannot be changed anymore so that data integrity is guaranteed. If someone tampers the data which has been stored before, it will trigger alarm which can be detected according to the color of the blockchain. Even if people who has tampered the data wants to make tampered content undetected, it will cost him a large amount of time to re-mine the whole blockchain from the tampered block to the last one.

The website consists of four page, which are separately called hash, block, blockchain and distributed. In the following content, I will introduce their function and implementation separately. The structure of the whole system is shown in the figure 4-1.

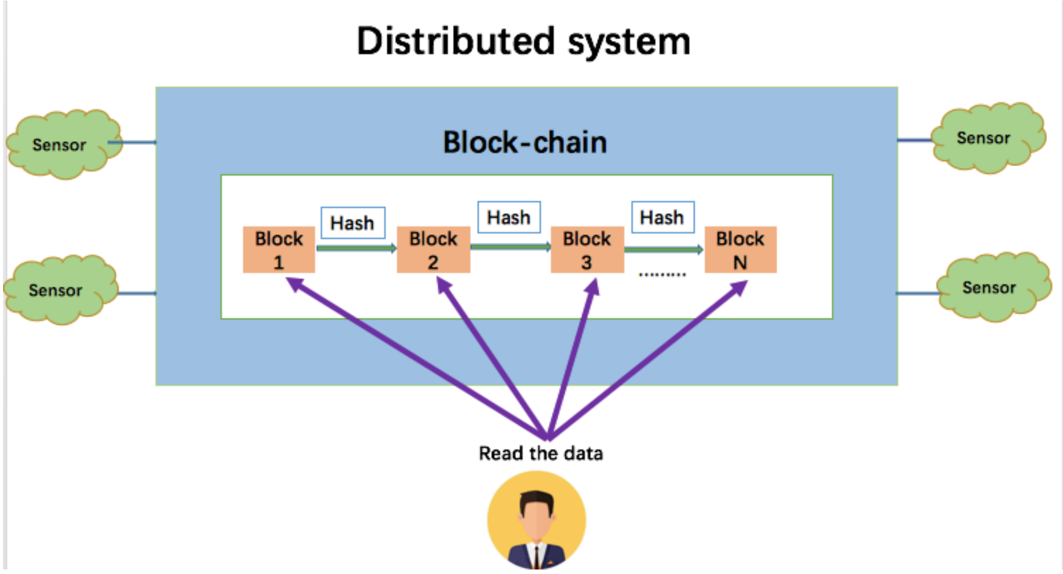


Figure4-1 The structure of the whole distributed system

For the code structure. The file hash.jade, block.jade, blockchain.jade and distributed.jade are respectively the implementation of hash page, block page, blockchain page and distributed page. The main function is written in blockchain.js which is the heart of the whole project. For different purpose, some functions are re-written in the jade file like data_mine(). The correlation between these files are illustrated in figure 4-2.

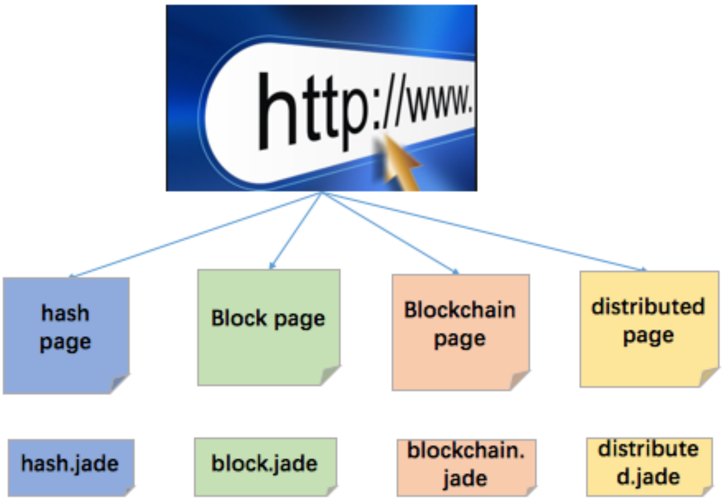


Figure4-2 The structure of the code files

4.2 Structure of the block

As we all know, each blockchain is composed of a large number of blocks according to the length we want to create. So the creation of block is a basic component. On the basis of block, we can create the blockchain easily. In this chapter, I will firstly introduce the structure of single block and then it will be easy to learn how to construct the blockchain. The structure of one block is shown in figure 4-3.

The block I used in the IIoT system is naïve block which guarantees the integrity of data produced from different sensors. The data structure of the block includes Block number, Nonce, Pre point, Hash and Data.

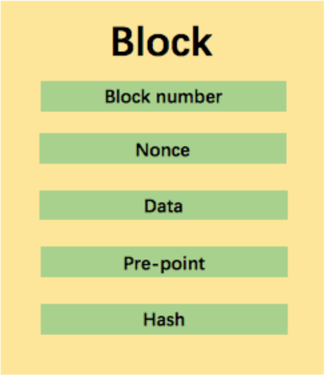


Figure4-3 The structure of one block

Block number: In the head of the block, there is a number which is unique for a certain block. This figure is incremental according to the length of blockchain you want to create. It will start from zero which like the genius block. There is no limit to the length of blockchain, so the maximum of block number is unlimited. But it is clear that the longer length of blockchain you want to create, the more time spent in the computer calculating operation. This thesis also gives the analysis of the relationship between operation time and the block number in the blockchain.

Nonce: The nonce number is used to calculate the hash. In this system, the standard of block mining success is to check if the result of hash calculation has achieved the difficulty you have set or not. The difficulty of hash is according to the number of zero in the head of hash calculation results. For example, if you set the difficulty equals four and the standard of block mining success is the hash calculation result should be like 0000… or you set the difficulty equals seven and then the hash result should be like 0000000…The difficulty is an important fixed constant in the block mining which has a great impact on the block mining time.

Different nonce number intervals will lead to significant difference in the speed of hash calculation, in other words, the block mining speed will be different which is the key parameter we want to test. The nonce interval depends on the difficulty you have set in block mining. If you set the degree of difficulty very high, but the maximum Nonce is not big enough, sometimes it is possible that the operation will fail to find a suitable nonce to achieve the difficulty you set. While if you set the degree of difficult in a low degree, and then you set the nonce in a wide interval, time will be wasted in the process of computer calculating operation because the computer may calculate a very big nonce that suits for the difficulty you have set but actually a small nonce can have the same effect on the hash result. So there is an obvious correlation between difficulty and nonce interval and both of them have a significant impact on the block mining time. In this thesis I will give an analysis about the correlation.

Pre point: Pre point plays a key role in the structure of blockchain which links a large number of blocks orderly and guarantee the data integrity of the whole database. Without its existence, the block only exists alone and do not form a strong blockchain structure which prevents the tampering attacks and unknown failures.

The pre point of one block is the hash result of its previous block which not only provides the link between the previous block and the current one but also link the current block to the next block. Pre point is also a key value to detect if anyone has tampered with data, because if people has changed data which has stored in one block before, for example replace the data with a new one or delete some key parameters produced from the sensors, the hash of the current block will be changed as long as it occurs any change in the data. With the change of the data, the hash result will be changed simultaneously. Under these circumstances, the hash of the current block has changed but the pre point of the next block has not changed yet. It is almost impossible for people who wants to make the tampered data undetected because he/she has to re-mine all the blocks from the tampered block to the last block to make the pre point and the hash of current one equal. Meanwhile, it is very easy for the manager of blockchain to manage the blockchain data, if something goes wrong, it is not difficult for manager to detect the tampered place where the pre point of current block is not the same as the hash of previous hash.

In this system, as long as the pre point is not as the same as the hash of the previous block, the background of the block will return green to red which look likes an alarm to tell the manager about the occurrence that someone has changed the data and the manager should take action at once.

Hash: Hash is also a key point in the structure of block which use SHA-256 to calculate the result. How SHA-256 works has already introduced before. In naïve block the hash is calculated according to the block number, data and nonce. So any change in these three parameters will impact the calculation result of hash a lot.

To check the hash result is a failure or a success depends on the difficulty level you have set. The number of difficulty means the number of zero in hash head. Because there are sixteen possible characters in a hex value, each time we increment the difficulty by one we make the puzzle 16 times harder. So we need to control the difficulty and the interval of nonce. In the test, if the difficulty is 6 and the maximum nonce should be over 5000,000,000. The analysis of the correlation of difficulty and block mining time will be shown in the following chapter.

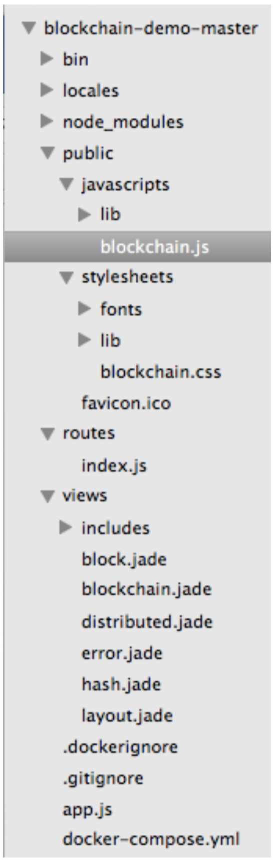
Data: In the block, there is no doubt that data integrity is very important. Under industrial internet of things background, the data which will be inserted into the block is the test results which produced from the sensors. The data in this system is a simulation of temperature and humidity sensors from the factory of different cities where temperature and humidity are key observations affecting the production part, therefore, the data integrity is very important in this factory.

The composition of the data can be any character, for example numbers in different formats like integer, decimal, fraction, and characters in different languages. In this thesis, all the data I used is the combination of English and integer. The length of the data is not limited but there is no doubt that there is correlation between the length of data with the block mining time and it will be analyzed in this thesis.

The creation of blockchain is based on each block. When we know the structure of a single block, it is very easy to learn the constructed mechanism of the blockchain.

4.2.1 The implementation of block mining

The figure 4-4 shows the whole structure of the code, the main function of block mining is in blockchain.js which is located in the lib, it is called by block.jade ,blockchain.jade and distributed.jade to implement the function of block mining.



```
▼ blockchain-demo-master
  ▶ bin
  ▶ locales
  ▶ node_modules
  ▼ public
    ▼ javascripts
      ▶ lib
        blockchain.js
    ▼ stylesheets
      ▶ fonts
      ▶ lib
        blockchain.css
      favicon.ico
  ▼ routes
      index.js
  ▼ views
    ▶ includes
      block.jade
      blockchain.jade
      distributed.jade
      error.jade
      hash.jade
      layout.jade
    .dockerignore
    .gitignore
    app.js
    docker-compose.yml
```

Figure4-4 The structure of code

There are total six function in blockchain.js: function sha256(block, chain), function updateState(block, chain), function updateHash(block, chain), function updateChain(block, chain), function mine(block, chain, isChain).All the function will be called by the block.jade, blockchain.jade, distributed jade, hash.jade. It works as the heart of the whole codes.

To implement the block mining, five functions are used, the main function is mine(block, chain, isChain).

Firstly, the function mineButtonAnimation(block,chain) will be started when mine button is pressed, and then the mine function will call the function getText(block,chain) which is used to read the text of the block including the block number, data and nonce. And then the data will be sent back to the function of mine. According to the text, function SHA256(block,chain) is called to calculate the hash result. After finishing the calculation, one judegement will be made based on the parameter ischain to check if the pre-point of the block equals the hash of previous block. If the equation is established, the hash result will be inserted into the block, else the function updateState will be called to change the background color of the block to show the state of block mining. The flow of block mining is presented in figure 4-5.

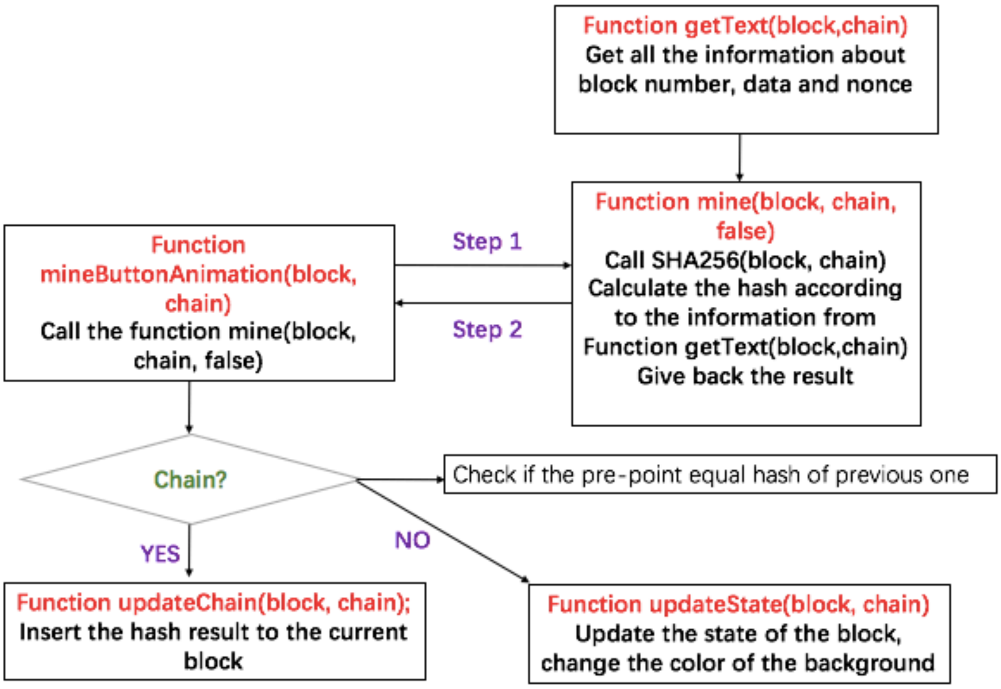

Function getText(block,chain)
Get all the information about block number, data and nonce
Function mine(block, chain, false)
Call SHA256(block, chain)
Calculate the hash according to the information from Function getText(block,chain)
Give back the result
Step 1
Step 2
Function mineButtonAnimation(block, chain)
Call the function mine(block, chain, false)
Chain?
Check if the pre-point equal hash of previous one
YES
NO
Function updateChain(block, chain);
Insert the hash result to the current block
Function updateState(block, chain)
Update the state of the block, change the color of the background

Figure4-5 The flow of block mining

4.3 Creation of the blockchain

As I mentioned before, the link between two blocks is the pre point. With the help of the pre point, a large number of blocks will be bound together and they will be in order which can be easily checked with their block number.

There are three states of the block which are presented through the different background color including grey, red and green representing their own meaning. This is implemented by the function updateState(block, chain).The work flow of this function is shown in the figure 4-6 . The meaning of different background color is shown in the figure 4-7.

When the block is not activated, the background color is grey. When the block is activated but the data has not been inserted into the block yet, the color of background will become red aim to tell that you have created new blocks but some blocks are empty. When the sensor has produced the data and successfully inset the data into block you should press the mine button. After that the computer starts to mine the block. Actually it is the process of calculating the hash result according to the data, nonce and block number which should be suitable for your difficulty set in the code. If block mining has completed, the background will become green to show the success.

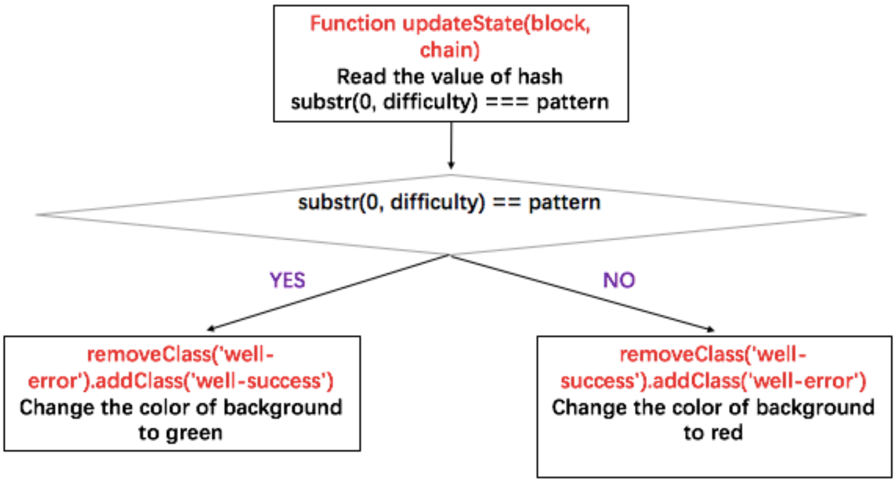
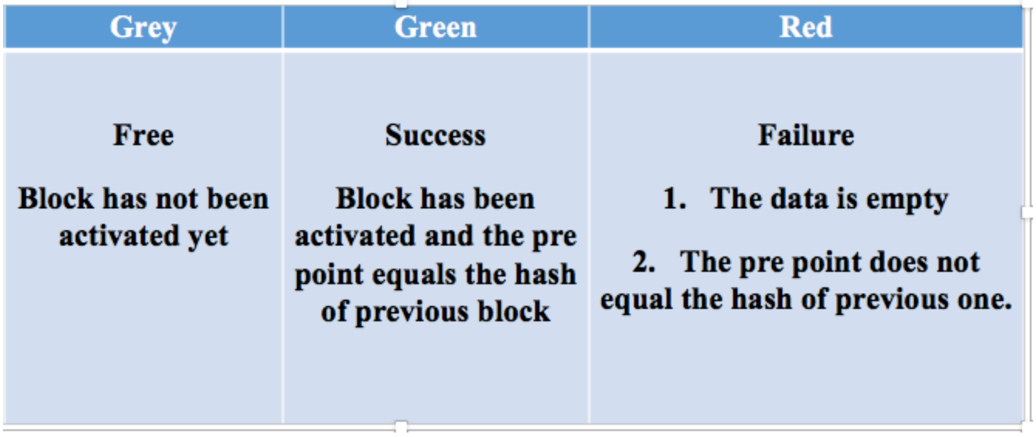

Function updateState(block, chain)
Read the value of hash
substr(0, difficulty) === pattern
substr(0, difficulty) == pattern
YES
NO
removeClass('well-error').addClass('well-success')
Change the color of background to green
removeClass('well-success').addClass('well-error')
Change the color of background to red

Figure4-6 The flow of updating the state of block

The background color also works as an alarm. When someone tampers the data stored in one block, the hash will be changed which will lead to the result that current hash is different from the pre point of the next block, in this condition, the background of block turns into red to trigger the alert on the broken of data integrity.



Grey	Green	Red
Free Block has not been activated yet	Success Block has been activated and the pre point equals the hash of previous block	Failure 1. The data is empty 2. The pre point does not equal the hash of previous one.

Figure4-7 The meaning of different background color

The right steps to create one blockchain in the system are listed as following:

Give the length of blockchain in the text area (i), the number of blocks will be activated. The function is implemented by function setup(i,1)

Insert the data which from the temperature and humidity sensors into the block. The system will show the next number of block which you can use. The function is implemented by function data_insert().

Press the button of mine, the computer will start the calculation of hash to mine the new blocks according to the data from the sensors.The function is implemented by function data_mine ().

If new data has been produced, the cycle will be re-started from the beginning again.

The creation of the blockchain is shown in the following figure 4-8:

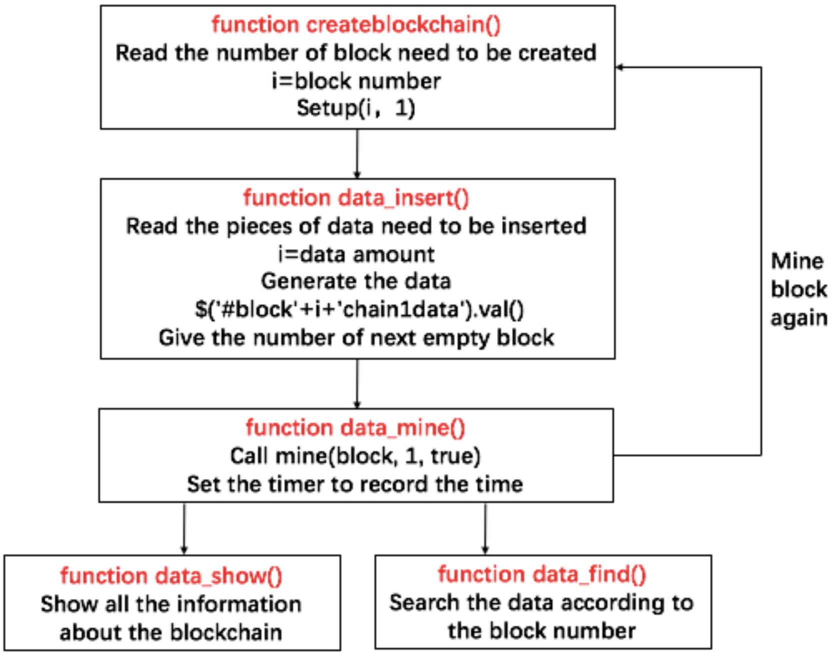

function createblockchain()
Read the number of block need to be created
i=block number
Setup(i, 1)
function data_insert()
Read the pieces of data need to be inserted
i=data amount
Generate the data
$('#block'+i+'chain1data').val()
Give the number of next empty block
function data_mine()
Call mine(block, 1, true)
Set the timer to record the time
Mine block again
function data_show()
Show all the information about the blockchain
function data_find()
Search the data according to the block number

Figure4-8 The flow of creation of blockchain

4.4 Distributed blockchain system

In the distributed page, it is a distributed system which includes four peers, separately represent four different sensors of different factories. It simulates the scenario of IIoT which includes many sensors of course in the real world, and the number of sensors will be far more than four. The simulation can be divided into three parts: generation of data, data storage in blockchain-based dataset and read data stored from the blockchain. In the following content I will introduce each part in detail and The whole flow of distributed system is presented in the figure 4-9.

(1) Data generation: The data are generated from four sensors in the format of the combination of number and text. At the same time the data will be inserted into the block. The four machines are simulated as four sensors which will detect the temperature and humidity of the factory environment in different places. Therefore, the content of each block includes the humidity, temperature and the data source where it is produced. The data length will be changed in order to find the correlation between length of data and the block mining time.

(2) Data storage: After the data generation we should press the button named by "insert the number of data" and the data will be stored into the block but the background of the block is still red because the block has not been minded yet, they are still 'illegal'. In order to make the block legal we should press the 'mine' button. The function behind the button 'mine' is the function data_mine() which calls the function SHA-256() and calculate the hash result which suits the difficulty you set. If the calculation is successful and the background of the block will turn green from red.

(3) Data read: In this part there are totally four functions to check the data integrity.

Get the whole chain data: We can get all data which is stored in the blockchain through the button 'find the data' and then the data stored by four sensors in the blockchain will be listed.

Get the certain block data: We can get the data content from the certain block according to the block number you input and then press the button 'search' the system will give back the content which stored in it.

Check the correctness: This function is used to check whether someone has changed the data which through the function data_check(), the main task is to check whether the pre point equals the hash of previous block. Actually, according to the background of the block we can also check the correctness. If someone change the data the background will change into red to trigger the alarm

Calculate the time: The function is used to calculate the time spent on the creation of the certain number of blocks. It is closely related to the evaluation part which will evaluate the scalability, correctness and correlation between different parameters and the consumed time.

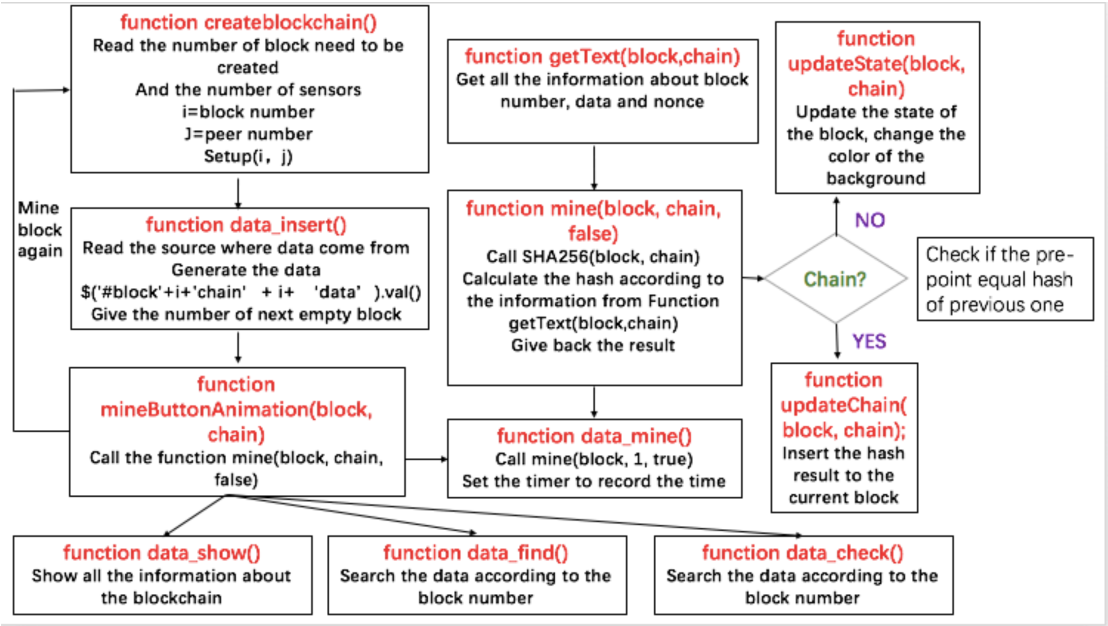

function createblockchain()
Read the number of block need to be created
And the number of sensors
i=block number
J=peer number
Setup(i，j)
Mine block again
function data_insert()
Read the source where data come from
Generate the data
$('#block'+i+'chain' + i+ 'data').val()
Give the number of next empty block
function mineButtonAnimation(block, chain)
Call the function mine(block, chain, false)
function getText(block,chain)
Get all the information about block number, data and nonce
function mine(block, chain, false)
Call SHA256(block, chain)
Calculate the hash according to the information from Function getText(block,chain)
Give back the result
function data_mine()
Call mine(block, 1, true)
Set the timer to record the time
function updateState(block, chain)
Update the state of the block, change the color of the background
NO
Chain?
Check if the pre-point equal hash of previous one
YES
function updateChain(block, chain);
Insert the hash result to the current block
function data_show()
Show all the information about the blockchain
function data_find()
Search the data according to the block number
function data_check()
Search the data according to the block number

Figure4-9 The whole flow of the distributed system

5 Results

In this chapter, the main task is divided into three parts. The first part is to show the main functions of the whole system which is based on the blockchain. The second part shows the results in the form of table and flow charts of measurements about block mining time under different parameters (data length, nonce, difficulty, block number and peer number). The range of the measurement is from one block to one blockchain and from one blockchain to the whole distributed system. A conclusion about the correlation between the block mining time and different parameters and the evaluation about the efficiency and the scalability of the whole system are given in the end of this chapter.

5.1 Main function of the system

In this website, there are total four pages which showing the growth of a blockchain system which are named by hash, block, blockchain and distributed. It is easy for you enter each page according to the page's name in the navigation box.

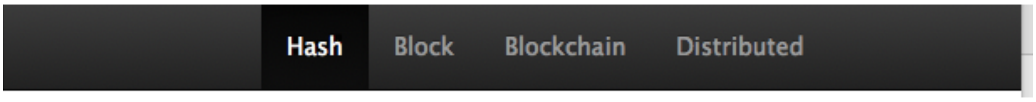
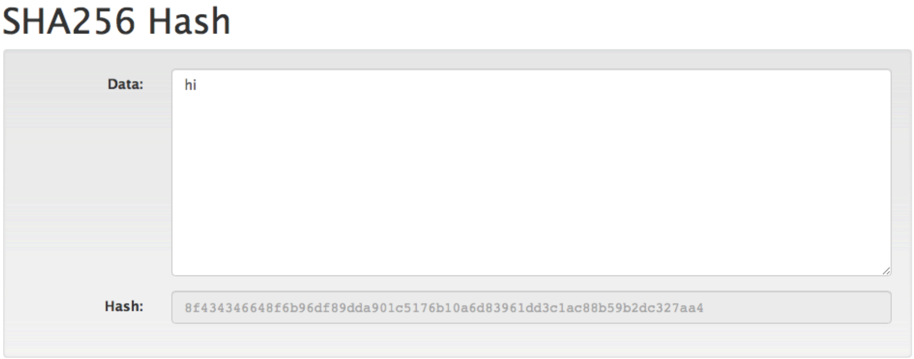


Figure5-1 The navigation box of the system

In the hash page, you can type any word in the blank of data, the hash result will be changed with the data you input according to SHA256 function. The hash result is a key during the process of block mining.

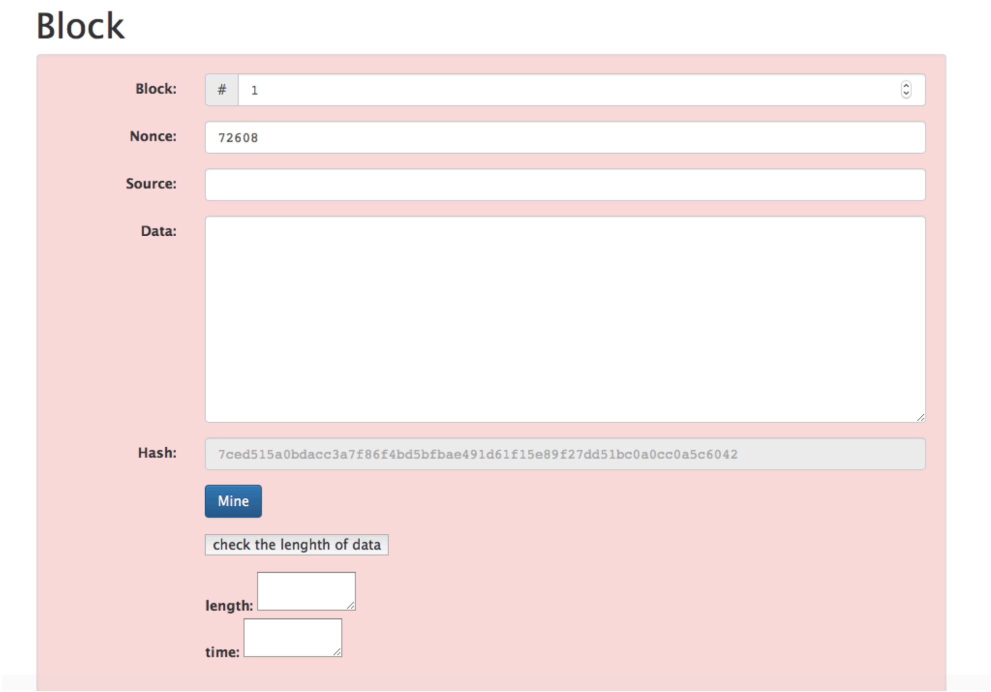

SHA256 Hash

Data: hi

Hash: 8f434346648f6b96df89dda901c5176b10a6d83961dd3c1ac88b59b2dc327aa4

Figure5-2 Hash page

In the block page, one single block is shown which contains the index of the block, nonce, source and data. One single block does not contain the pre-point, this is the only difference between the single block and the one in a blockchain. In this page two functions are achieved, one the mining function, when you input the data and then press the mine button, the calculation is started. When the mining is finished, the time of block mining will be shown in the time textbox. Another function is you can check the length of data before you input it which helps to find the correlation between data length and block mining.



Block

Block: # 1

Nonce: 72608

Source:

Data:

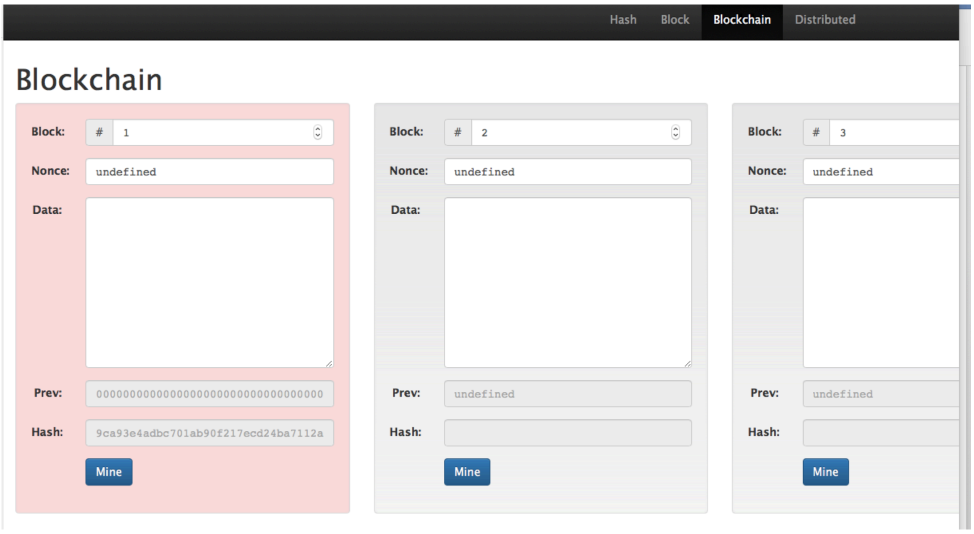
Hash: 7ced515a0bdacc3a7f86f4bd5bfbae491d61f15e89f27dd51bc0a0cc0a5c6042

length:

time:

Figure5-3 Block page

In the blockchain page, you can choose the length of blockchain which fits the system and then the background color of the certain number of blocks will become red according to your demand. The color red tells you that the block is empty without input. Then choose the number of data needed to be created, the data will be inserted into the block, but the block is still “illegal” without proof of work. After the process of mining, the hash results will be shown in each block and all the block update their pre point to make the link with each other. When the background color of the block becomes green which means the new block succeeds in being added to the blockchain and once it is created, no one can change it, otherwise, the background color will become red again to work as an alarm about tamper.



Blockchain

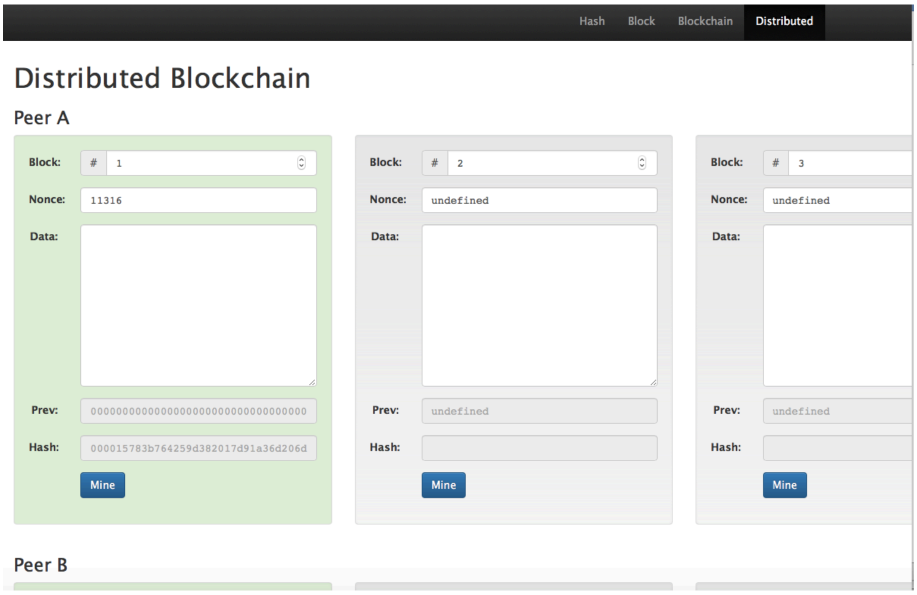
Block: # 1
Nonce: undefined
Data:
Prev: 0000000000000000000000000000000000
Hash: 9ca93e4adbc701ab90f217ecd24ba7112a

Block: # 2
Nonce: undefined
Data:
Prev: undefined
Hash:

Block: # 3
Nonce: undefined
Data:
Prev: undefined
Hash:

Figure5-4 Blockchain page

In the distributed page, it is a combination of the previous pages. There are four sensors in default which can be changed in the code to simulate the different number of sensors. Each sensor has a copy of blockchain which provides the synchronization and fairness of the blockchain system. The data generation, data insert, mining progress are the same as the blockchain page. The difference is in the distributed blockchain, the data has its own source so the content of the data includes not only the result of sensor detection but also the source of it.



Distributed Blockchain

Peer A

Block: # 1
Nonce: 11316
Data:
Prev: 0000000000000000000000000000000000
Hash: 000015783b764259d382017d91a36d206d

Block: # 2
Nonce: undefined
Data:
Prev: undefined
Hash:

Block: # 3
Nonce: undefined
Data:
Prev: undefined
Hash:

Peer B

Figure 5-5 Distributed page

In the distributed page, there are many buttons that achieve different functions which is easier for users to use according to their name. "the length of blockchain" is used to activate the number of blocks you want to use, the button 'choose' is used to given the source of the data, "insert the data" is the number of blocks that the users want to input the data. "Check the length of data" will be given the length of input.

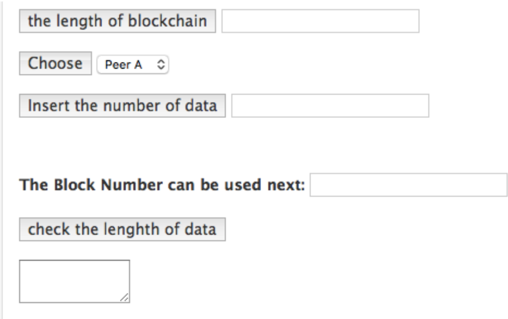

the length of blockchain
Choose
Peer A
Insert the number of data
The Block Number can be used next:
check the lenghth of data

Figure5-6 Data generation buttons in distributed page

After the creation of the blockchain, the system also allows the manager or users to check the correctness of the whole blockchain that is mainly according to whether the pre point hash is the same as the content of previous block's hash. And it can also be checked according to the background color of each blocks. The blockchain should be transparent, the users can search the content of the whole blockchain and it is possible for them to search the data of the certain block according to the block number.

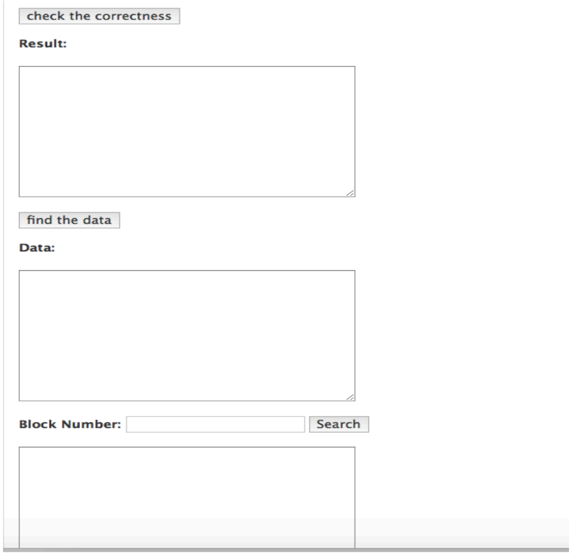


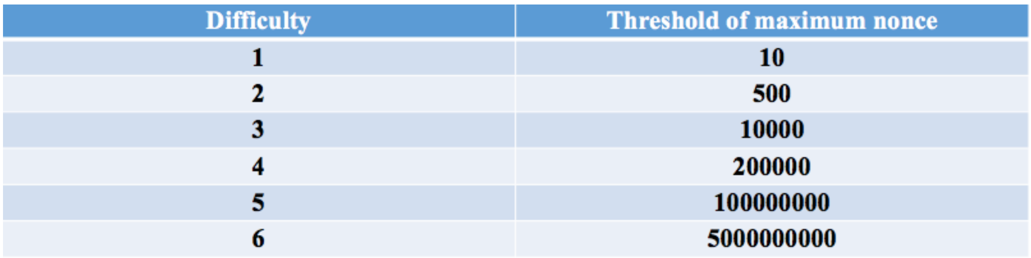
Figure5-7 Search data function in distributed page

5.2 Block mining under different parameters

I measured the block mining time under different parameters by changing the value of the parameters within a certain of range. The tests are conducted from three areas: one block, one blockchain, distributed blockchain.

5.2.1 One block test

In the hash calculation, nonce is used to make the hash results meet the criteria for difficulty so it is very important to give the relatively accurate interval of the nonce, otherwise the calculation of hash may fail. As shown in the figure 5-8, the difficulty means the number of zero in the head of hash and the threshold of maximum nonce is the minimum number of nonce which should be set in the code. From the results, we can find that plus one to difficulty, the number of nonce will be increased at exponential level. It is important to know the quantitative relationship between these two parameters.



Difficulty	Threshold of maximum nonce
1	10
2	500
3	10000
4	200000
5	100000000
6	5000000000

Figure5-8 Difficulty and threshold of maximum nonce in one block

To find the correlation between data length and mining time, I did three measurements in different difficulty which can be shown in the figure 5-9. When the difficulty is three, the mining time is always under 1s regardless of the length of data, but from the lines in the second and third charts, we can find there is a linear positive correlation between the mining time and the length of the block. When the length of data is changed from 1000 to 1500, the excavation time increases significantly.

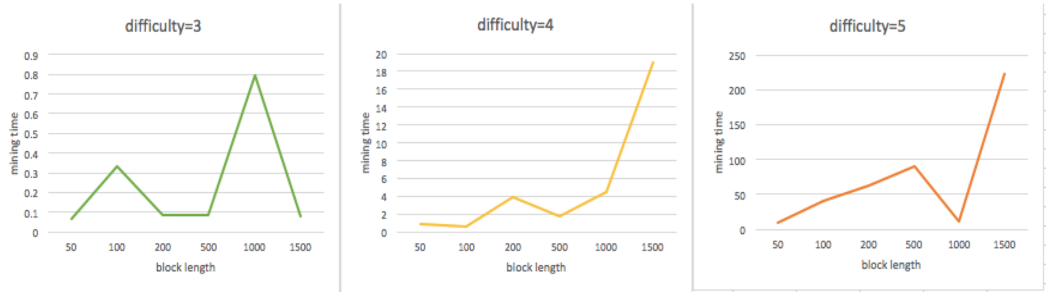


Figure5-9 Mining time and block length under different difficulty in one block

The correlation between mining time and difficulty also use three sets of experiments which is based on different data length 100,500 and 1000. From the figure 5-10 it is clear that mining time rises sharply with the increase of difficulty, the mining time in the figure is use the log calculation, because mining time has an excessive growth. The analysis of the increase multiples of time from difficulty one to difficulty six with three different data lengths, the result is presented in figure 5-11. From the test, we can know the degree of difficulty have a great impact on the mining time.

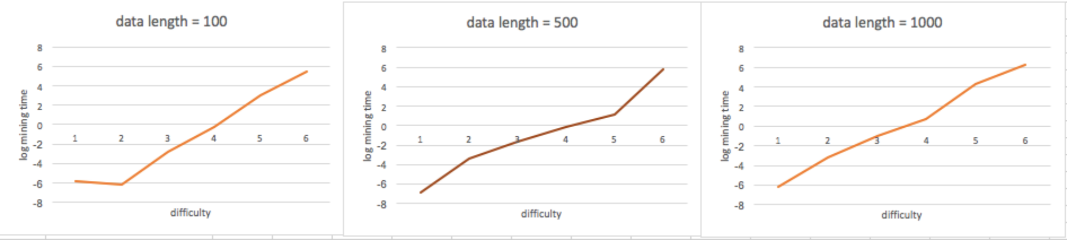
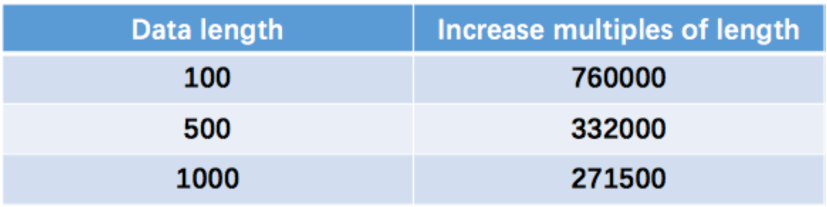


Figure5-10 mining time and difficulty under different data length



Data length	Increase multiples of length
100	760000
500	332000
1000	271500

Figure5-11 Increase multiples of time in one block

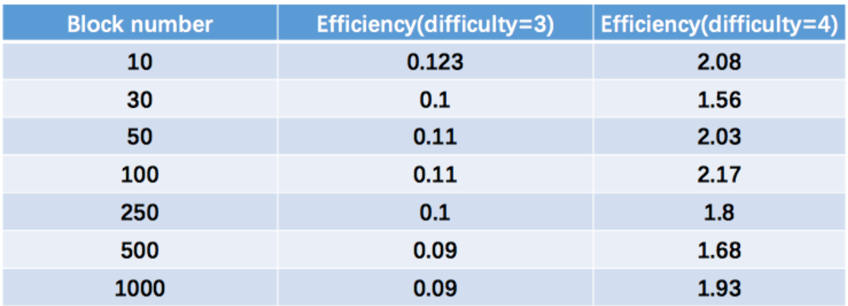
5.2.2 One blockchain test

The correlation between block number and block mining is shown in the figure 5-12, the efficiency in the table is used the following formula

$$Efficiency = \frac{mining\ time(s)}{blocknumber} \tag{5-1}$$

There are two sets of data in different difficulty, one is under difficulty equals three, the other is under difficulty equals four, but two sets of measurements are under the same data length and nonce. From the results, we can find that the efficiency is not changed a lot with the increase of the block number, so I infer that if the difficulty is fixed, the speed of block mining can be calculated with the following formula:

$$Time = Efficiency \times block\ number \tag{5-2}$$



Block number	Efficiency(difficulty=3)	Efficiency(difficulty=4)
10	0.123	2.08
30	0.1	1.56
50	0.11	2.03
100	0.11	2.17
250	0.1	1.8
500	0.09	1.68
1000	0.09	1.93

Figure5-12 Block mining time and block number in one blockchain

The following figure shows the correlation between block mining time and data length under the same block number which is fifty and difficulty which equals three, from the result line we can know that if the data length is between 50 to 300 bits, the mining time just fluctuates around 10 second but once the length exceeds 500, time increases significantly.

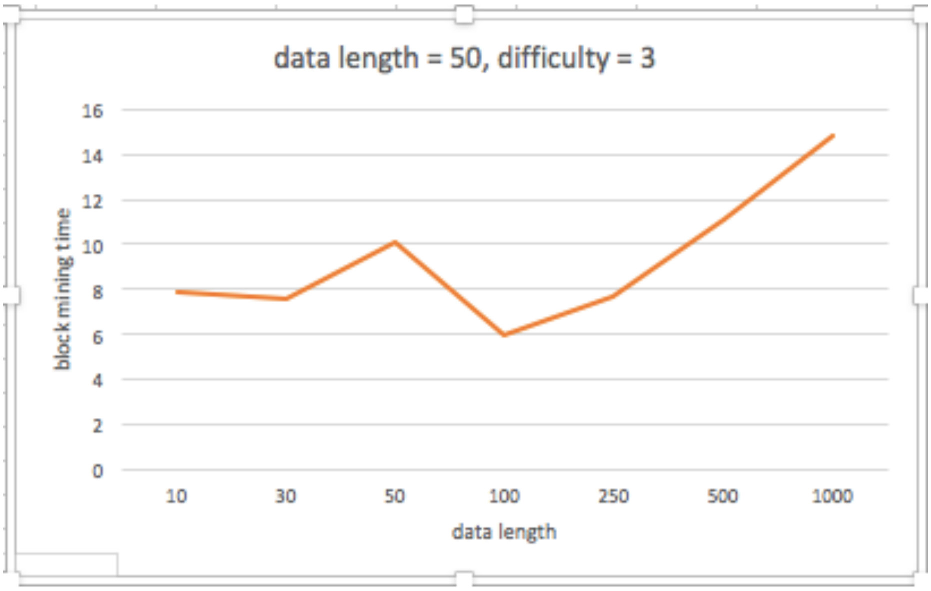


Figure5-13 Mining time and data length in one blockchain

The correlation between difficulty and mining time are just like the result in one block, the mining time grows violently with the difficulty. The test result is shown in the Figure 5-14. In this figure the block mining time is calculated with log function in order to reduce the value change interval. The specific experimental values are shown in Figure 5-15. But it is worth noting that when the difficulty increases from 1 to 3, the block mining does not increase so much. When the difficulty equals four, the block mining time is increased by more than ten times. In the test, I also try to add the difficulty to six or seven but the calculation time is so long more than thirty minutes. The reason behind it is about the performance of computer, so I cannot add the difficulty to a bigger number. The interval of difficulty I have test is from one to five.

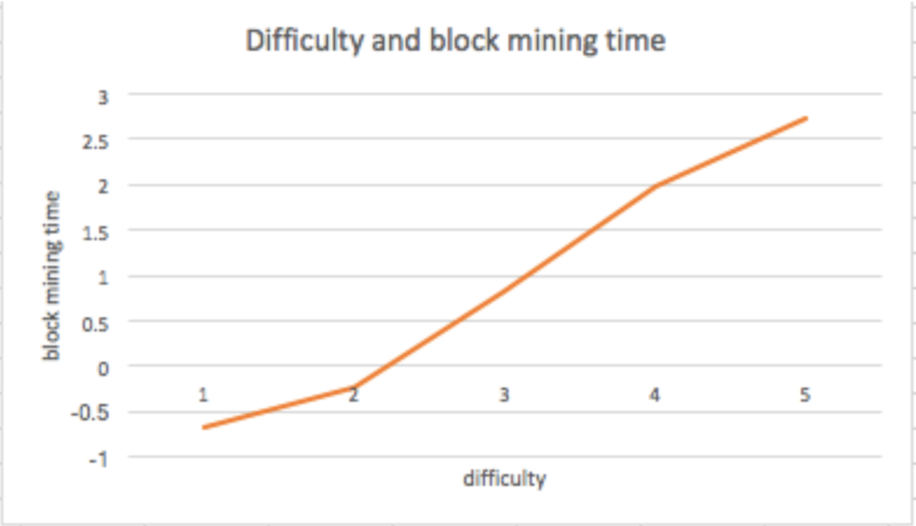
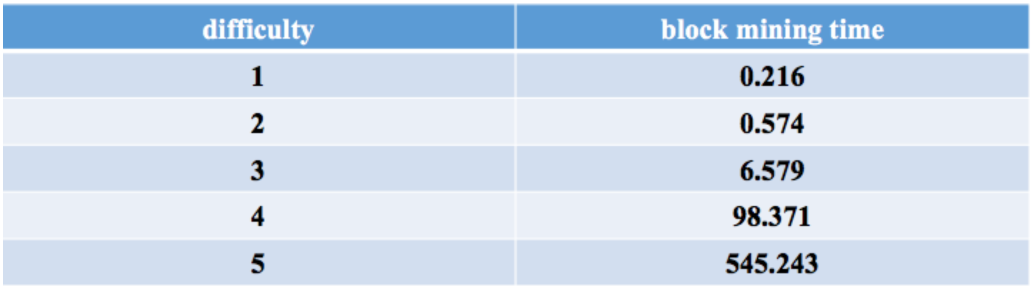


Figure5-14 Mining time and difficulty in one blockchain



difficulty	block mining time
1	0.216
2	0.574
3	6.579
4	98.371
5	545.243

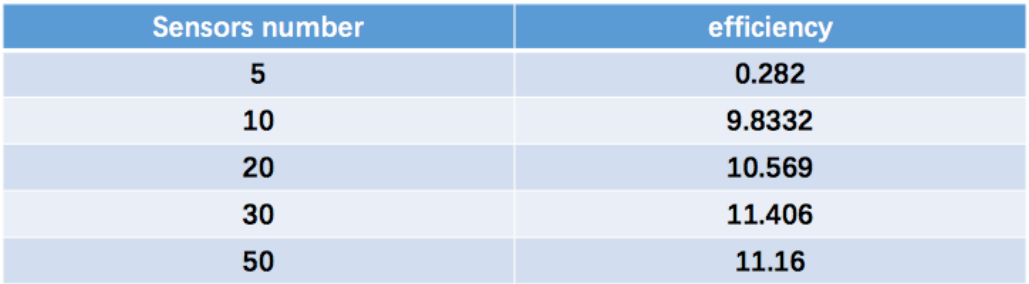
Figure5-15 Mining time and difficulty in one blockchain

5.2.3 Distributed blockchain test

In the distributed system, the number of distributed sensors is very important. I measure the number of sensors which ranges from 5 to 50 with the same difficulty (three is in default) and data length (thirteen bits) to check the efficiency of the block mining, the efficiency is calculated in the following formula:

$$Efficiency = \frac{mining\ time(s)}{sensors\ number} \tag{5-3}$$

In this formula, the calculating result is the time of creating one blockchain for one sensor. It is easy to understand the formula and find the rule that if the calculation of efficiency is a big number, this shows that the speed of block mining is low. With this formula, it is easy for us to compare the efficiency under different parameters. For example, from the measurement results in the table, the efficiency of five sensors (0.282) is higher than ten sensors (9.8332). In the figure 5-16 the trend of the system is the more sensors that operates together in the system, the lower efficiency it will be, which is the same as in many situations in real life. Although the efficiency shows a downward trend, it does not decrease in a high speed, in contract, the number increases slowly. When the increase of sensors number from 10 to 50, the number of efficiency only plus two.



Sensors number	efficiency
5	0.282
10	9.8332
20	10.569
30	11.406
50	11.16

Figure5-16 Mining efficiency and sensors number in distributed system

The results of data length and block mining time are almost the same as the test in one block and one blockchain, when the data length is between 50-500 with the same difficulty (three is in default) and block numbers(fifty), the mining time just fluctuates in a small range around 50 seconds. So we can infer that 500 bits is a threshold. It is better to control the data length not longer than 500 bits.

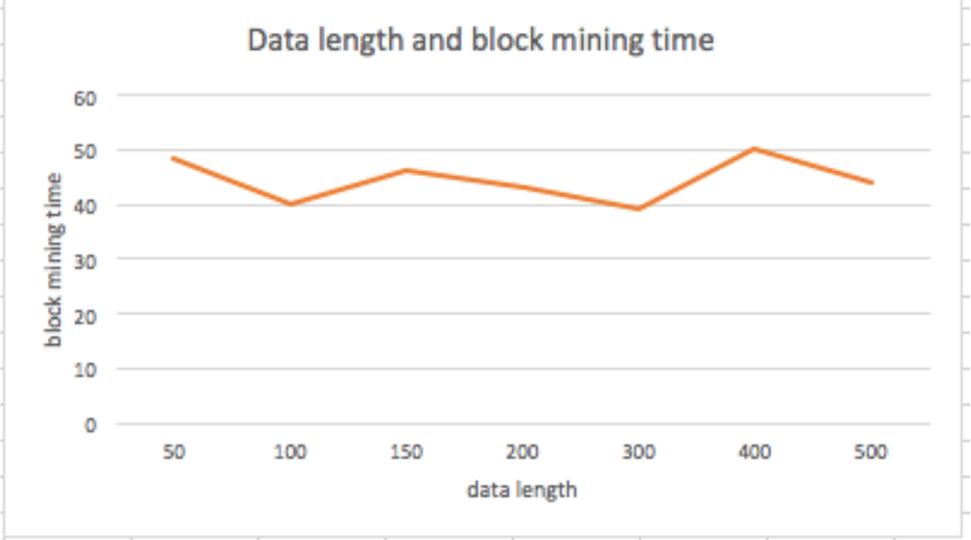


Figure5-17 Mining time and data length in distributed system

The test of difficulty is from one to five. The results of difficulty and block mining time are also almost the same as the test in one block and one blockchain, the mining time in the figure 1 is calculated by log function, because its large increase. In general, the block mining time is greatly affected by difficulty especially when the difficulty exceeds three. The specific experimental values are shown in figure 5-19.

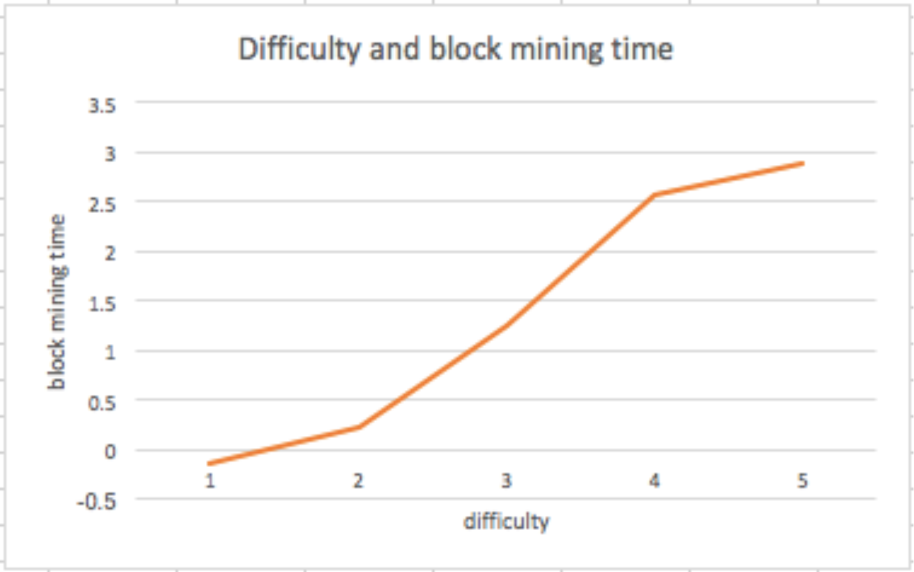
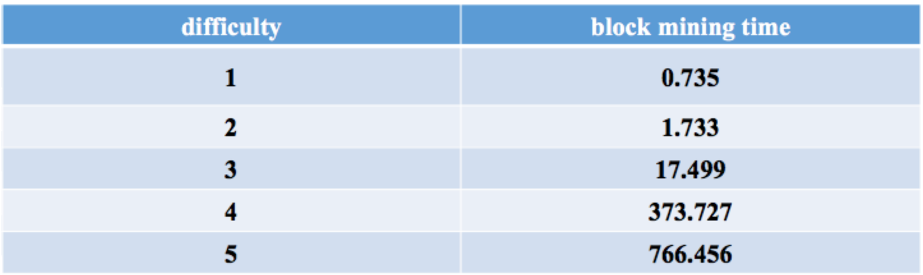


Figure5-18 Mining time and difficulty in distributed system



difficulty	block mining time
1	0.735
2	1.733
3	17.499
4	373.727
5	766.456

Figure5-19 Mining time and difficulty in distributed system

5.3 Evaluation

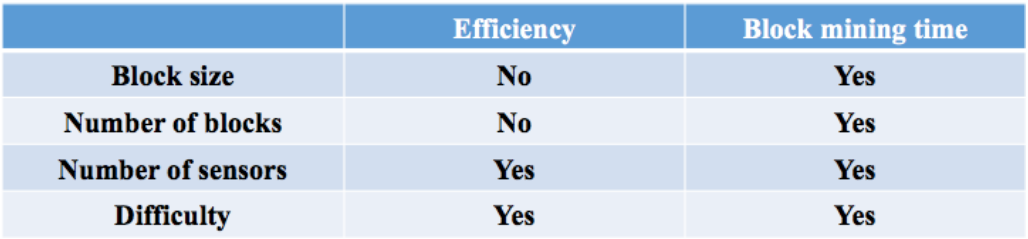
In the evaluation part, I will evaluate the system from three aspects: scalability, efficiency and security. Through the evaluation, the advantage and disadvantages of this system can be found and the improvement should be considered in the meanwhile.

5.3.1 Scalability

The scalability of the system is a kind of system's ability of dealing with the high-density work pressure. In my scenario, the work pressure is from the increasing number of sensors and the number of block which should be mined. From the result of the measurements, we can find that with the increasing of sensors, the efficiency becomes lower, but the speed is still in a high level that the systems can finish mining 50 blocks in 500 seconds with 50 sensors. But I do not measure if the number of sensors is larger than 50, what will happen. In general, it can conclude that the system is scalable in a not big factory with the small number of sensors or machines in the manufacture.

5.3.2 Efficiency

The thesis gives two formulas to calculate the efficiency, we can find that the efficiency of block mining is directly related to the difficulty of hash and the number of sensors but is not correlated with the block number. As to the block length, 500 bits is a threshold, if the data length is bigger than 500 bits, the efficiency will decrease sharply. In my scenario the number of sensors is four and the data length is about 30 bits, so the efficiency can be guaranteed. If the data should be increased, the suggestion is no more than 500 bits. The figure 5-20 gives a conclusion about whether the factors has the correlation with efficiency and block mining time or not.



	Efficiency	Block mining time
Block size	No	Yes
Number of blocks	No	Yes
Number of sensors	Yes	Yes
Difficulty	Yes	Yes

Figure5-20 A conclusion about different factors with efficiency and mining time

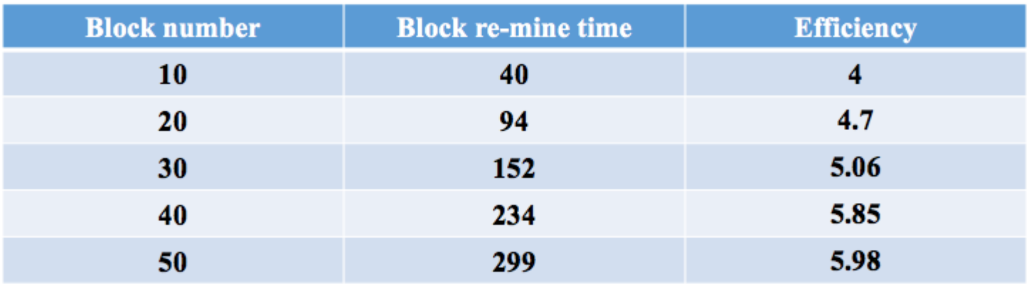
From the test, we can also infer that the parameter that has the greatest impact on the blockchain mining time is no doubt the difficulty which causes exponential growth in block mining time. The second one are number of sensors and number of blocks that increase the time of block mining but not to a big extent. The data length has a relatively small impact on the blockchain mining time. As long as the data length is controlled within a certain range (in the test, the threshold of data size is 500 bits), it has almost no impact on the blockchain mining time.

5.3.3 Safety

In this thesis, the safety of the blockchain-based system is to guarantee the data integrity which is really important in the real world factory. To evaluate the safety of the system, I simulated the scenario that someone tampers the different length of blockchain and analysis the time consumed in re-mining the whole blockchain. The result is shown in the figure 5-21. The efficiency is calculated by the following formula:

$$Efficiency = \frac{mining\ time(s)}{sensors\ number} \tag{5-4}$$

From the results of measure, it is clear that the efficiency number is increasing during the process of re-mining which indicates that the speed of block mining slows down. The re-mine efficiency value is higher than the normal mining efficiency value which can compare with the results in figure 5-16 and figure 5-19. From this respect, the safety is guaranteed.



Block number	Block re-mine time	Efficiency
10	40	4
20	94	4.7
30	152	5.06
40	234	5.85
50	299	5.98

Figure5-21 Block re-mining time and block length in distributed system

Actually, the process of re-mining is hard to proceed successfully, it is almost impossible for the tampered places undetected. There are two important reasons behind it. One is that the background color of each block works as the alarm, if some changes are made, the color will change from green into red immediately which makes it easier for the manager to detect something strange happen and to change it. The other reason is that in the system it is makes it possible for the users to check the correction of the blockchain through the pre point as long as they press the button set in the system and then the system will generate the report about the correctness of the whole blockchian. The users or the manager can check the correctness from the report, if something goes wrong, the report also points out the wrong block number.

6 Conclusions

The Internet of Things makes our life in a more efficient and intelligent way, as a subcategory of the Internet of Things, Industrial Internet of Things not only helps to increase production efficiency and corporate competitiveness but also promotes the coordination of people, society and nature. However, the security of industrial IoT has never been solved especially about the data integrity. Fortunately, with the birth of a new type of decentralized database: blockchain, people have found a turning point. In this chapter, a brief conclusion is made about the work of the whole thesis and the achievement of five goals which is listed in the chapter one.

Goal one: Survey the area of IoT, industrial IoT and Blockchain technologies in order to gain the knowledge about the structure of Blockchain, its applications and benefits or risk to use it and be familiar with the background of IIoT.

By reading the articles and journeys related to IoT, I learn the origin and development of IoT which has already connected things to things, human to things and human to human. Three basic characteristics which are intelligent processing, reliable transmission and comprehensive perception respectively, makes Internet of things have the intelligence to work as human. I also learned four key internet of things techniques: RFID technology, sensor technology, wireless network technology, artificial intelligence technology and cloud computing technology and its application of real world represented by smart city, safe city construction and food safety.

On the basis of Internet of Things, I have a further learning about Industrial Internet of Things. The main difference between IoT and IIoT is IIoT focus on the industry with clear business benefits. The structure of the IIoT composes of physical layer, networks, industrial clouds and intelligent terminals. I also learned the challenges facing to the IIoT and the goal of availability, integrity and intelligence. From these three goals, the integrity of data is what the thesis mainly focuses on. There are three typical scenarios of IIoT that give me idea about a typical IIoT scenario I can implement.

In the study of blockchain, I have learned four components of blockchain. They are shared record, network of nodes, distributed database system and cryptography which can be used to guarantee the data integrity of IIoT. I also learn five applications based on blockchain technology ranging from finance to government which are presented by Bitcoin, Ethereum, Hyperledger, government voting and internet of things. I also have a deep understanding in the blochchain's advantages like immutability, transparency and openness, decentralization and automation. At the same time blockchain also has its problems which are mainly caused by pressure on the storage, time-consuming, power-consuming and privacy. Knowing the pros and cons of blockchain helps me to evaluate it in a more rational way.

Goal two: Determine a typical industrial internet of things scenario where data integrity is required. My typical Industrial Internet of Things scenario is a factory which has many sensors during its manufacture. The sensors are used to detect the temperature and humidity of the manufacture environment. The location of these sensors is not fixed, they can even not be in one city or one country as long as they are all connected to this system. Data transmitted from sensors to blockchain is about 30 bits. The detected data should be kept in order to evaluate the quality of the products to guarantee that the both temperature and humidity are always within the specified range. In this scenario, data integrity is in need to guarantee the quality of the products. If someone tamper the data which is detected by the sensors, it may cause the final product cannot meet the production standards and eventually leads to a decline in the quality of the factory's products and loss of competitiveness among peers.

Goal three: Implement a data integrity system for the IIoT. The implementation of this system is a website with four different pages coded in jade and html. The runtime environment of this system is Node.js. Before running the code, npm should be installed and be started after the installation. The blockchain technology I chose is naïve chain which is not difficult to implement but it has high efficiency in the protection of data integrity. The hash algorithm I chose is SHA-256 which is widely used in blockchain.

I succeed in simulating the factory scene which can be divided into three steps: data generation, block mining and reading the data from the block. The data is generated by the computer and the default length of data is about 30 bits and after the data generation, the mining work should be started which is actually a process of hash calculation. The timer is set to calculate the block mining time. The statue of block mining is shown by the background color of blocks: grey, red and green which means inactivated, failure and success. The users can read the data through the different buttons of website, they can read the whole blockchain, read the data of certain block and also check the correctness of data integrity.

Goal four: Perform measurements of block mining time under different parameters in accordance to the established scenario. The total measurement I made in the test is close to one hundred sets of data which is not an easy work. There are totally five parameters in the measurement: data length, difficulty, block number, sensor number and block mining time. I found the correlation between block mining time and the other four parameters. In the measurement I always keep two parameters fixed in order to explore the rule between the other two one. The results of measurement are shown in the tables and line charts with the tool of excel. Some results have been processed in mathematical operations that simplify the process of rule finding. The conclusion of the correlation between block mining time and five parameters is that difficulty is the most important factor affecting the time of blockchain mining, followed by the number of blocks and the number of sensors. The minimum impact on blockchain mining time is the data length.

During the measurement I also summarize two formulas which can calculate the efficiency of the block mining. The larger the calculated value, the lower the actual mining efficiency. All the formulas, tables, line charts and the conclusions are presented in Chapter five.

Goal five: Evaluate the end results in fields of scalability, efficiency, safety and propose future work. I evaluate the performance of this system from three respects, which are scalability, efficiency and safety. The evaluation is according to the measurement and the rule I found during the measurement.

In terms of the scalability, the speed of block mining reduces with the increase of sensors but not in a big extent and the maximum of sensor number in the test is fifty, the scalability is still unknown when the number of sensors is far more than fifty. But it can sure that the system can work well with the number of sensors about fifty.

Considering the efficiency, the block mining time is an important metrics. In chapter five I concluded that the efficiency is directly related to difficulty, number of sensors and data length. The difficulty I used most frequently in test is three which can mine the block at a not slow not quick speed so the default difficulty of the system is three. The threshold of data length is 500 bit. To maintain system efficiency, the length of data stored in each block should be controlled under 500 bits.

When it comes to safety, I simulated the scenario that the process of re-mining blocks when someone has tampered with the data and want to make it undetected, I recorded the time consumed during the mining and have a calculation of the block mining efficiency. Through the comparison between re-mining efficiency and original efficiency, I find that it takes more long time to re-mine the blocks so the safety is guaranteed to a certain degree. In addition to the efficiency, this system provides two ways to detect the tampering with data. One is through the background color the other one is the function implemented in the system which can be used to check the correctness of the blockchain through the pre-point.

6.1 Ethical consideration

With the development of internet of things, it changes our daily life in all aspects like living habits, the way of entertainment, health care and so on. But some ethical issues also should be toke into consideration.

Everything has two aspects, when something begins to change our life in a subversive way, a deeper consideration should be made by us. Of course, we should enjoy the convenience and high-efficiency about any connection of objects, any connection of times, any connection of locations. However, the risk and challenges behind of it cannot be ignored. The main issue is about the safety. How to prevent the security of our personal information, how to make sure that all the inter-connection between things to things is in a good aim not be misused by illegal organizations to commit a crime and how to recover the losses if the inter-connection has the failure in the process.

In fact, these problems must be solved because hackers are also eyeing for the technique flaws of the Internet of Things, industrial Internet of Things and blockchain such as the attack of bitcoin. Fortunately, many organizations and scholars are currently working on these issues and propose solutions. I think our life will become not only in an efficient but also safe way.

6.2 Future work

The blockchain-based system has finished in three parts which are data generation, data storing and reading the data, but in fact there is still a lot of future work to do to improve the functionality and performance of this system.

Firstly, during the whole process of project, the blockchain-based system is only run by my computer. In the process of my survey, I learnt that the performance of computer decides the efficiency of complicated hash calculation. Thus, the block mining time changes with the computer performance. Testing the system on multiple computers will be the future work in order to judge the efficiency of system more comprehensively.

Furthermore, the system is based on naïve chain which only contains block index, nonce, data, pre-point and hash. There is no doubt that naïve chain is simple but efficient, but in the real environment of factories, these five parameters are not enough for the actual demand. It should be added some parameters according to the real need of the factory. The data part can be divided into more detail parts like the source, the accurate value of temperature and humidity, the state of the value about whether it meets the standard of quality in order to make the data more specific. If the system aims at putting into practice in the real world, the survey of factory requirements must be implemented in advance.

Last but not the least, if the data stored in blockchain can be read by more terminals, the use of the system will be more extensive. The range of the terminal can be wide. For example, the android app of mobile, the real-time equipment like Raspberry Pi, LED numerical control board and so on. If time permits, I will add more real-time equipment to this system

If I have more time, I will improve the system in the above three aspects to make the system more efficient, scalable and safe.

References

[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016 – 2021 White Paper

[2] A. Pouraghily, M. N. Islam, S. Kundu and T. Wolf, "Poster Abstract: Privacy in Blockchain-Enabled IoT Devices," 2018 IEEE/ACM Third International Conference on Internet-of-Things Design and Implementation (IoTDI), Orlando, FL, USA, 2018, pp. 292-293.

[3] In Lee, Kyoochun Lee. The Internet of Things (IoT): Applications, investments, and challenges for enterprises. Business Horizons, 2015, 58:431-440

[4] Roberto Minerva, Abyi Biru, Domenico Rotondi，Towards a definition of the Internet of Things (IoT), IEEE Tech. 2015[DB/OL]

[5] Abdmeziem M.R., Tandjaoui D., Romdhani I. (2016) Architecting the Internet of Things: State of the Art. In: Koubaa A., Shakshuki E. (eds) Robots and Sensor Clouds. Studies in Systems, Decision and Control, vol 36. Springer, Cham

[6] Software-Defined Industrial Internet of Things in the Context of Industry 4.0 Jiafu Wan, Member, IEEE, Shenglong Tang, Zhaogang Shu, Di Li, Shiyong Wang, Muhammad Imran, and Athanasios V. Vasilakos .

[7] Y. Zhang, D. D. Zhang, M. M. Hassan, A. Alamri, and L. Peng. CADRE: Cloud-assisted drug recommendation service for online pharmacies. Mobile Netw. Appl, 2015, 20(3): 348–355

[8] M. Chen, Y. Zhang, Y. Li, S. Mao, and V. Leung. EMC: Emotion-aware mobile cloud computing in 5G. IEEE Netw, 2015, 29(2): 32–38

[9] Jari Kreku, Visa Vallivaara, Kimmo Halunen, Jani Suomalainen. Evaluating the Efficiency of Blockchains in IoT with Simulations[DB/OL]

[10] Ali Dorri, Salil S. Kanhere, Raja Jurdak, and Praveen Gauravaram. LSB: A Lightweight Scalable BlockChain for IoT Security and Privacy

[11] Manish Lamichhane. A smart waste management system using IoT and blockchain technology[DB/OL]