A study on High Scalable Blockchain and the Application to Data Existence and Integrity Authentication

March ２０１９

Yuefei Gao

A study on High Scalable Blockchain and the Application to Data Existence and Integrity Authentication

Graduate School of Systems and Information Engineering

University of Tsukuba

March ２０１９

Yuefei Gao

Trust in the LORD with all your heart,

And lean not on your own understanding;

In all your ways acknowledge Him,

And He shall direct your paths.

Proverbs 3:5-6

Acknowledgements

The five-year studying experience in University of Tsukuba is a significant, meaningful and unforgettable time in my life. What I have learnt here are not only the research skills but also the abilities that can be used in the future.

I would like to express my great appreciate to Prof. Nobuhara, my research advisor for all the guidance and support he has offered during my graduate studies at University of Tsukuba, Japan. From Prof. Nobuhara, I have learnt the research skills, problem thinking methods as well as interpersonal relationship skills. I would like to thank Prof. Kawai, for the guidance, support and advice that helped my research processes and dissertation preparation. I would like to thank Prof. Koga, Prof. Shibuya and Prof. Takayasu, for the reviews and comments that helped to improve my doctoral dissertation. Also, I would like to gratitude to CMU lab members for a helpful environment during my studying.

I would like to offer my special thanks to my family and my friends, who always support, understand and encourage me. Thank you very much for unconditional understanding, warm support and always praying for me. Also, I wish to acknowledge the Japanese Government Scholarship and Rotary Yoneyama Memorial Foundation, for the scholarships that supported me in my doctoral study.

Finally, thank God, my LORD. Thanks for guiding me in the darkness and being with me in the hard times. Thanks for letting me experience You in the five years. I will keep on trusting You in my future. All the glory belongs to You.

Abstract

This research focus on solving the low scalability problem of the blockchain. The proposed method is a high scalable blockchain protocol that is based on proof of stake and sharding. A decentralized trusted timestamping is created as a blockchain application.

First, a proof of stake and sharding based protocol is proposed as a possible solution for blockchain's low scalability problem. Scalability is that the transaction processing speed grows nearly linearly with the size of the whole blockchain network. Low scalability means that the transaction processing rate does not increase with the network size, and this is a primary limitation of blockchain technology. The basic idea of the proposed method is to implement proof of stake consensus algorithm instead of proof of work consensus. In this way, problems of proof of work, such as high energy consumption and potential security problems, could be solved. Also, sharding protocol is implemented with proof of stake to processing transactions in parallel. Evaluation experiments were conducted in a simulation AWS EC2 network to confirm the scalability of the proposed method.

Second, a decentralized trusted timestamping is created as a blockchain application which could help to authenticate existence and integrity of digital data. Considerable intellectual property is created and shared everyday on the Internet and the number will grow faster in the future. Intellectual property rights protection is required and current methods will be hard to fulfill the requirement. Decentralized trusted timestamping based on the Blockchain was proposed. However, two problems exists. In order to solve the two problems, this study proposes a methodology for storing a maximum of N*20 bytes of data and implements a search function for stored data. Evaluation experiments were performed to confirm the efficiency of the search function and to find that the proposed method implements decentralized trusted timestamping could be finished in an average time of 20 min at a possible cost of 0.24 USD.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Our society structure is changing, from a centralized one to a decentralized one (Figure 1). For example, media has changed from one-to-many to many-to-many. In the 1990s, people didn't have smartphones, so they listened to the radio and watched television to know what was happening. Today, almost everyone has a cellphone and people are more connected. Because of the new technologies such as smartphones and the Internet, our society has changed from a centralized one to a half-decentralized one. In the future, it is predicted that our society will become a completely decentralized one. The technologies that will help this transition are Internet of Things (IoT), Artificial Intelligence (AI) and the blockchain. These three technologies are also said to be the symbols of the fourth industrial revolution [1]. Among them, the blockchain might be the most promising one.

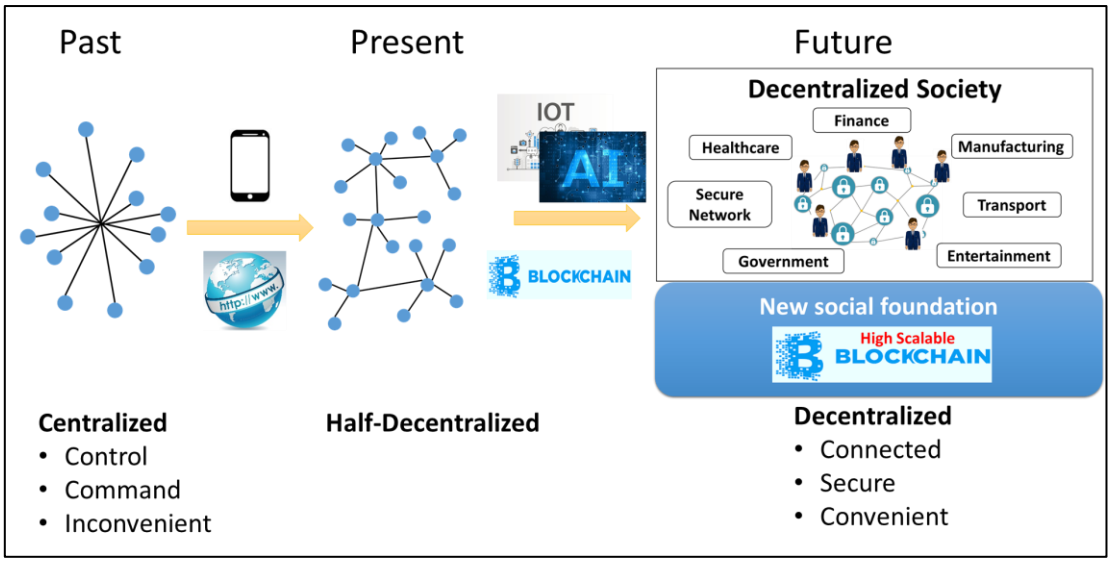


Figure 1.1 Social structure transition

The blockchain has drawn attention all around the world. This technology was discussed in World Economic Forum (WEF)'s annual conference in January, 2018 [2]. 2,500 top leaders including top business leaders, international political leaders and economists were gathered together in this year's conference. Until last year, the topic was AI, but this year the topic was the blockchain. In addition, the market size of the blockchain increases dramatically. In 2016, blockchain's global market size was 200 million dollars and in 2018, it increases almost 3 times as around 550 million dollars. In 2021, the size of the blockchain technology market worldwide is predicted to reach 2,300 million dollars, which

will increase almost 4 times than in 2018 [6]. In Japan, the government has also started considering applying the blockchain technology in many aspects. For example, Financial Services Agency intends to use it in financial transactions, Cabinet Office wants to manage administrative documents with it and Ministry of Internal Affairs and Communications would like to use it in online voting for election [7]. Except the government, Japanese universities have also started researches on this new technology. In July of 2017, the University of Tokyo and Keio University founded the group named BASE Alliance to research the blockchain [8]. Therefore, today no matter in the world or in Japan, the blockchain technology is a very hot topic.

The blockchain technology is promising and it is said to be the hope for the future. This is one of the reasons why it becomes so popular in the world. Our society tends to be more and more decentralized these days and it is predicted that it will be more decentralized in the future. On the way to the future, the blockchain is one of the significant technologies that will help this transformation. Besides, a new social foundation will be needed to serve as the infrastructure of decentralized society. The blockchain technology, which is famous for its decentralization, will contribute to future society [1].

However, becoming the new social foundation means taking the responsibility to support the whole decentralized society, and the blockchain now does not have the ability to meet the requirements and serval limitations need to be solved first. The blockchain has three main problems: 1) the calculation costing too much electricity is wasteful; 2) the blockchain's size is too large to store; 3) the blockchain's scalability is low. With the explosion of blockchain based applications, the third problem might be the most urgent one. Low scalability of the blockchain means that transaction processing speed does not scale with the network size. For example, in the case of Bitcoin's blockchain, the rate of transaction processing is 7 transaction per second (tps) [3] while the network size is around 2.9 to 5.8 million [9]. In contrast, the average processing rate of payment system such as Paypal is 115 tps [4]. In VISA's case, the processing rate can reach a peak of 56,000 tps [5]. These problems need to be solved first so that the blockchain could be capable of being the new social foundation.

Several methods have been developed to solve the high electricity consumption problem and the large blockchain size problem [10], while the scalability problem remains unsolved. Some previous works studied the scalability problem and proposed different solutions. These methods might be effective to increase scalability to some extent, nevertheless they didn't solve the scalability fundamentally. In this thesis, a new blockchain protocol is proposed to address the scalability problem. The scalable blockchain could work as the social foundation for the future decentralized society, where a large number of applications will be built. A decentralized trusted timestamping is one of the possible applications based on blockchain and it is implemented as an example of decentralized society usage scenario. Address the scalability problem of the blockchain effectively is the main motivation behind this research.

1.2 Organization

This study investigates blockchain's scalability problem and implements an example application. To improve blockchain's performance in scalability, proof of stake (PoS) and sharding methods are applied. A scalable blockchain could serve as the new social foundation.

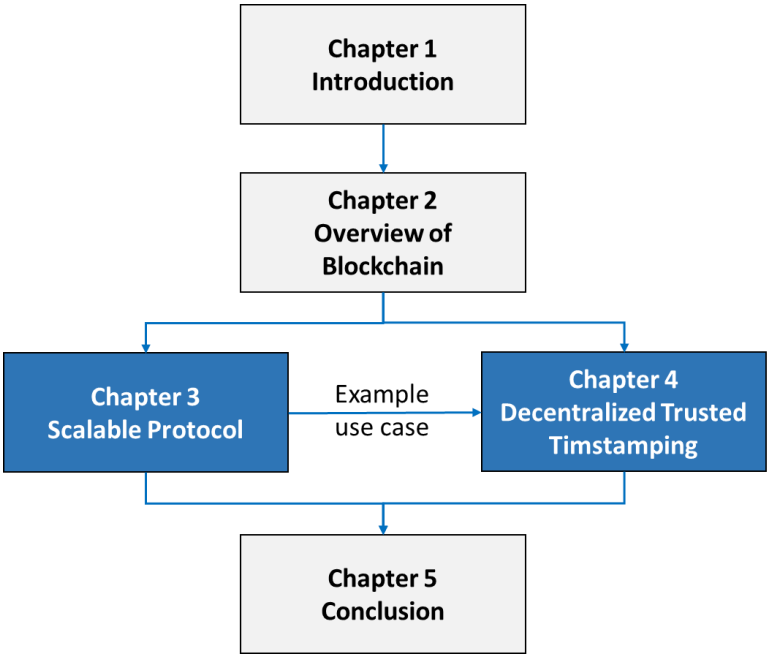


Figure 1.2 Thesis contents flow chart

Figure 1.2 illustrates the structure of his thesis, including five sections. Chapter 1 gives the introduction of the social backgrounds of the research. Chapter 2 presents the overview of the core technology in this thesis – the blockchain. In Chapter 3, the proposed scalable blockchain protocol which is based on PoS and sharding is introduced as a possible solution for blockchain's scalability problem. In Chapter 4, a decentralized trusted timestamping is presented as an example of a usage scenario of blockchain in the decentralized society. Finally, Chapter 6 ends the thesis by summarizing the accomplishments and drawing conclusions.

Chapter 2

Blockchain

2.1 Introduction

The history of the main technologies used in the blockchain can date back to the 1970s, when the first designs of cryptographic hash functions was created [12]. Then in 1977, Rivest - Shamir - Adleman (RSA) public-key cryptosystem was created [13]. The important consensus protocol proof of work (PoW) was invented in 1992 [14]. Several systems such as e-cash [15], Hashcash [16] and b-money [17] were proposed based on these core technologies.

E-cash is an anonymous electronic money that was proposed by David et. al. in 1982. Public key signature schemes were used to ensure E-cash's security. Hashcash was proposed by Adam in 2002 as a PoW system to limit spam emails and denial-of-service attacks. The mechanism of Hashcash is a significant proposal and is applied in a famous system now. To be more specifically, Hashcash increased the cost for senders to reduce spam emails. An amount of CPU calculation is required for email sender and the result is later validated by email receivers. Although senders have to spend a certain amount of time CPU calculation every time before sending an email, receivers who validate the email only need a negligible computational cost. General speaking, this process could be described as 'hard to calculate but easy to confirm'. Another important system is B-money, which was a proposal of Wei in 1998. B-money is an anonymous, distributed electronic cash system and it has influences on other electronic cash systems. For example, the concept of broadcasting the money transferring transactions to all participants is used in later electronic cash systems. These systems founded the development of Bitcoin, which was published by an unknown person or group of people using the name Satoshi Nakamoto in 2008 [18].

There are no banks or governments to issue Bitcoin and it is managed by users in a peer-to-peer network. Bitcoin is a cryptocurrency system that achieves anonymity and decentralization, which were not implemented simultaneously by previous systems [23]. Decentralization of Bitcoin means it is not controlled by any people or institutions. Bitcoin does not exist physically and is actually transaction data. Breakthrough in solving the double-spending problem makes Bitcoin overcome the previous electronic cash systems. Double-spending problem is the main technological difficulty for distributed database in financial systems. This problem means user attempts spending the same currencies twice. In the case of paper currencies, it will not in users' hands after being spent; therefore, double-spending problem does not happen. However, if it is electronic currency, double-spending is possible and may cause security problems. Financial systems are supposed to reject double-spending transactions but it

is difficult to implement [24]. In Bitcoin's blockchain, the double-spending problem is solved in a decentralized way, without the need for a centralized third authority. The double-spending problem occurs when two different transactions spend the same money. Bitcoin prevents this problem by deciding that the first transaction of these two being recorded on the blockchain is the real one [37].

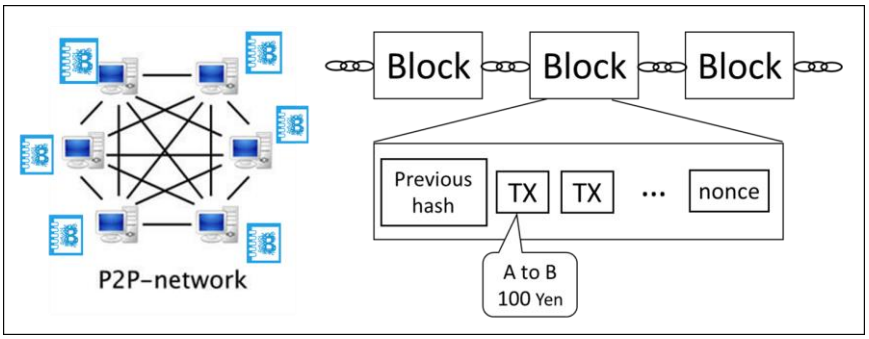


Figure 2.1 The blockchain

The blockchain was invented to serve as a public ledger of Bitcoin at first. As shown in Figure 2.1, the blockchain works in a peer to peer network and it is a note of transaction records. A page in this note is called a 'block'. In a blockchain, many blocks are connected together like a chain. A block mainly contains three information: hash value of the previous block, unconfirmed transactions and a random value called a nonce.

2.2 Important concepts in the blockchain

This section gives an overview of the technologies behind the blockchain. The technologies in the blockchain are not new. This section explains several concepts related to the blockchain.

2.2.1 Cryptography

Cryptography is used to secure the data in the blockchain. In this section, two main cryptographies that are used in Bitcoin are explained. These two cryptographies are public key cryptography and hash function.

2.2.1.1 Public key cryptography [12, 42]

Public key cryptography was invented in the 1970s and in Bitcoin it is used to create key pairs. Each key pair includes a private key and a public key. A public key is derived from a private key. In message communications, private keys are used to encrypt messages and public keys are used to decrypt the encrypted messages. In Bitcoin transactions, private keys are used to sign payments and public keys

are needed when receiving Bitcoins. A private key could be any picked or random number and it should be kept secret all the time. A public key is generated from a private key by applying Elliptic Curve Cryptography (ECC). In Bitcoin, a specific elliptic curve defined in secp256k1 standard is used. Figure 2.2 is an example of a sepc256k1 standard elliptic curve. Calculation a public key from a private key is not reversible; therefore, it is impossible to find a private key with a public key. This irreversible processes increase the security of Bitcoin network.

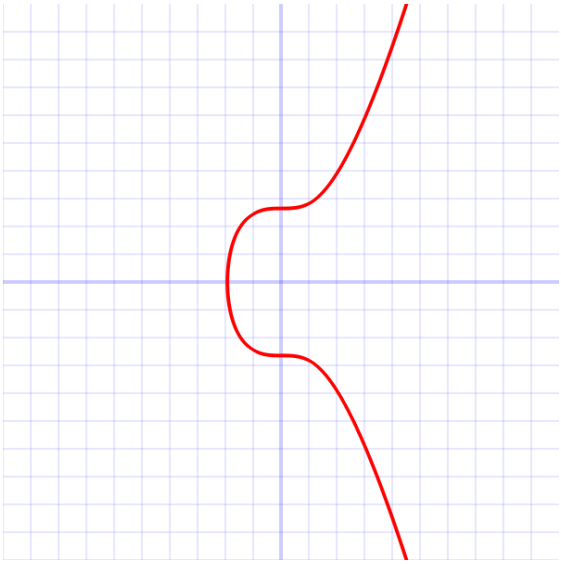


Figure 2.2 An example of a secp256k1's elliptic curve

Except private-public key pairs, public key cryptography is also used in digital signatures. Public keys can be thought as bank accounts, then private keys work as signatures that unlock bank accounts, and digital signatures are the authentication when spending the money. The function of a digital signature is to ensure that a message was created by a signer and it has not be tampered with or forged. The processes of digital signature contains 3 steps: 1) a person A signs a transaction with A's private key to generate a digital signature; 2) A sends the message and the digital signature to another person B; 3) B verifies the digital signature with A's public key.

2.2.1.2 Hash function [37]

A hash function is an algorithm that takes inputs of data and creates a hash value as outputs. The hash function has three important features. First, the length of the output is not related with the input data. For a hash function, the input data could be any type with different lengths but the outputs always have same lengths. Second, the output is always the same for the same input. For a same input, the hash value output will never be different. Third, the hash function is not reversible. This means that it is impossible to calculate the input data from an output hash value.

Bitcoin used two hash functions: SHA256 and RIPEMD160. These two hash functions work in

similar ways and the difference is the length of their outputs. SHA256 produces a 256-bit hash value while RIPEMD160 produces a 160-bit hash value. SHA256 and RIPEMD160 are used to create Bitcoin addresses from public keys.

2.3 Mining

Based on the concepts above, details of mining is explained in detail in this section. The process to create a block is called 'mining' and the computer nodes who do the mining work are known as 'miners'. Mining is the process that people who help to protect Bitcoin's trust receive bitcoin as rewards. In the blockchain's peer-to-peer network, first, miners create block. Then, miners compete to solve a difficult math puzzle which requires trillions of calculation and an average of 10 minutes. Finally, the miner who solve the math puzzle successfully could record the new block on the blockchain and receive an amount of Bitcoin as a reward [25, 26].

More details of mining is explained with the case of Bitcoin. In Bitcoin's peer-to-peer network, when a transaction is broadcasted, it is stored as an unconfirmed transaction in the 'mempool', which is short for memory pool. Each node in the Bitcoin network has a RAM to store these pending transactions. Miners collects the unconfirmed transactions to create new blocks [36]. In order to create a new block, miners have to solve a math puzzle, which is to calculate a hash value. A simple example is shown in Equation 2.1. To solve the math puzzle is to adjust the value of the nonce until the hash value on the left side of the equation fulfills two conditions: 1) the hash value should be smaller than the target value; 2) the hash value should be start with a number of leading zeros, for example, '000000000000000000049493cc46dcda4a3d94dc86afd8b74d8486615027a58d0a' is a block hash value with 17 leading zeros.

$$\text{Hash (Previous Block Hash, Transactions, Nonce)} \leqslant \text{Target} \tag{2.1}$$

More than trillions of hash calculation is needed to produce a block and this amount of calculation cost large amount of electricity. Hash rate is used to measure the power consumption for a cryptocurrency to continue functioning. Figure 2.2 presents the change of the Bitcoin network hash rate since 2009 to 2018. As the time of writing, the hash rate is around 50EH/s, which means 50 quintillion hashes per second [41]. To complete this amount of calculation results in high consumption of electricity. In 2015, the Bitcoin consumes $150,000 worth of electricity in only one day. This amount of electricity could power 31,000 homes or 50,000 electrical cars [35]. Bitcoin's hash rate is decided by the blockchain difficulty.

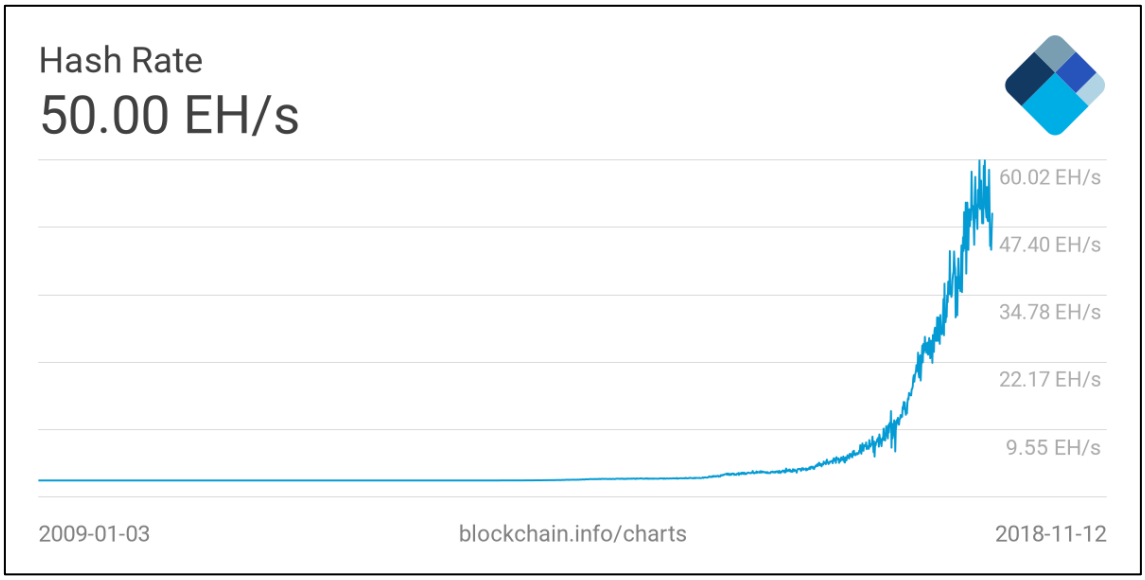


Figure 2.3 Hash rate of the Bitcoin network [40]

The blockchain difficulty is adjusted by increasing or decreasing the number of leading zeros. The difficulty is adjusted every 2,016 blocks in order to make the block interval to be around 10 minutes. Every 10 minutes, the miner who solve the math puzzle successfully could recorded the new block on the blockchain and receive an amount of Bitcoin as a reward [37]. The amount of the reward is decreasing every four years. The total amount of the Bitcoin is around 21 million. Without the issuance of banks, the Bitcoin is produced in mining. New Bitcoins are given as rewards to the miners who create new blocks and recorded them on the blockchain successfully. The amount of this reward decreased to half every four years. In the first four years, the reward for creating each block was 50 Bitcoins and in November of 2012, the reward reduced to 25 Bitcoins per block. In 2016, the reward decreased again, to 12.5 Bitcoins. Finally, in the year 2140, the 21 million Bitcoins will be mined. After that, miners won't receive new issued Bitcoin as rewards and their income will only come from transaction fees [38, 39]. At first, when the Bitcoin was not famous, it is possible to mine Bitcoin with a normal computer. However, with the increasing participation of mining, the competition becomes fierce and special hardware is necessary for successful mining.

In the beginning, a person could mine Bitcoin with a personal computer. Now people participate mining pools to mine together. The mining type changes from one-man mining to pool-mining. In addition, the mining hardware also transited from CPU to ASIC. CPU mining was used in the first phase of Bitcoin mining and its hash rate was around 20MH/s. Then in the second phase of Bitcoin mining, GPUs were used. GPU mining provide a hash rate ranging from 100MH/s to 500MH/s. Next, FPGAs mining were applied and it could offer a hash rate about 1 GH/s. Finally, ASICs chips, which are built for a specific using applications, were used in Bitcoin mining. With ASICs chips, the hash rate could achieve 3TH/s [38, 39].

In Section 2.2 and Section 2.3, the core technologies related to the blockchain are explained in detail. Bitcoin's blockchain is used as an example in the explanation. The blockchain technology was created as the public ledger for Bitcoin; however, it could be used in a much wider field. The next section introduce the transition of the blockchain technology.

2.4 Transition of the blockchain

Until now, the blockchain has being applied in hundreds and thousands of projects. These projects could be put into the three large categories shown in Figure 2.3. The three categories also represent the transition of the blockchain technology. Blockchain 1.0 is cryptocurrencies, Blockchain 2.0 is smart contracts, and Blockchain 3.0 is decentralized application [19-21].

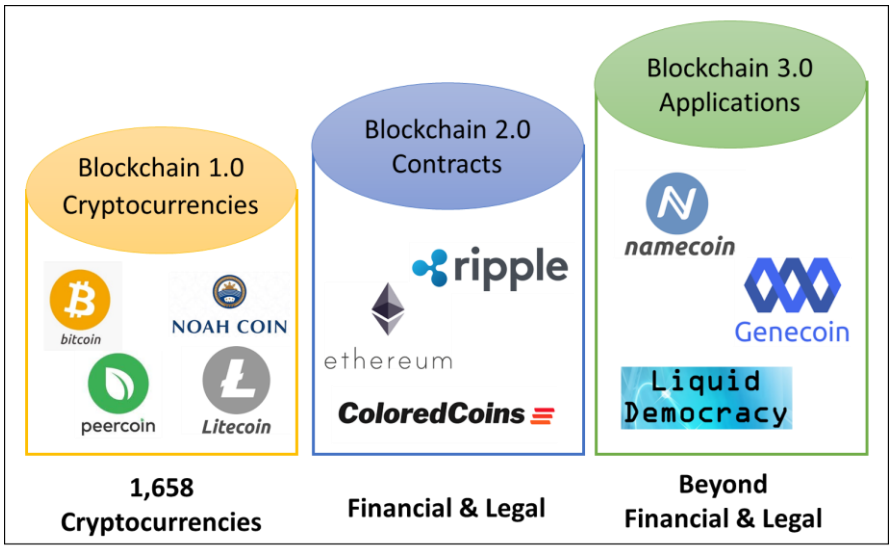


Figure 2.4 Transition of the blockchain technology [19-21]

Blockchain 1.0 includes the cryptocurrency projects. Bitcoin was the first one and now there are more than 1,600 cryptocurrencies. These cryptocurrencies have their own characteristics and differences in design and use cases. For example, Litecoin is similar to Bitcoin in mining processes but has a lower latency and applied different crypto algorithms [27]. Peercoin is the first cryptocurrency project that applied PoS consensus protocol. Although the consensus in Peercoin is not completely PoS but the hybrid PoW and PoS, the implementation of PoS has significant meaning for solving the high energy consumption of PoW [10]. NOAHCOIN project, which is founded in 2016, focuses on intensify the trade and market between Japan and Philippines [28]. The first generation of the blockchain, i.e. cryptocurrencies, contributes to electronic money transfer. Compared with the traditional bank money transfer, cryptocurrencies are much faster and more reliable at lower costs.

Blockchain 2.0 refers to financial and legal services that based on smart contracts. For example, Ethereum is a decentralized platform for cryptocurrencies and it provides Turing-complete programming language for creating smart contracts [29]. Ripple is a smart contract system for payment and exchange. Ripple protocol is based on BFT, and it is open source and distributed [30]. Colored Coins is another example, a platform for digital and physical assets. Colored Coins expends the use case of Bitcoin to assets, for instance, track and register assets [31]. After being used to develop cryptocurrencies, blockchain was then used to design financial and legal projects. With the Blockchain 2.0 projects, financial transactions are extended beyond money transfer. Also, legal related projects contribute to protect legal rights by providing trustless registration and certification platforms.

Blockchain 3.0 is the decentralized application, containing the projects except for financial and legal purposes. For example, Namecoin provides a decentralized domain system. Namecoin is based Bitcoin's blockchain and is the first project to implement merged mining, which means allowing a miner to mine more than one blockchain at the same time [32]. Genecoin is a project related to biology. Genecoin samples the DNA, turns it into data, and stores the DNA data in the Bitcoin's blockchain. In this way, the DNA data could be stored permanently and securely. Genecoin is one of the projects help with human transition into the future [33]. Liquid Democracy is from the project LiquidFeedback, which helps decision making by providing voting platform. The system is fair, transparent and reliable. In Liquid Democracy system, everyone has the same rights, the rules are made publicly, and all data are recorded [34]. Blockchain 3.0 contains the projects beyond finance and legal. After Blockchain 2.0, the blockchain technology is used in much wider fields. This technology could change our lives in many aspects.

In the early years, blockchain was not as famous as Bitcoin. However, the blockchain attracts more and more attention than Bitcoin. Now this technology is not only the public ledger of Bitcoin but a technology that may change the world. The reasons that the blockchain attracting more and more people mainly because of the features of the blockchain.

2.5 Features and challenges

The blockchain technology is the hope for the future decentralized society because of its special features, nevertheless, it also has limitations. The blockchain technology has both advantages and disadvantages [22]. The three main features of the blockchain technology are decentralization, immutability and trustlessness. The blockchain network is a decentralized peer-to-peer network. Once the data is recorded on the blockchain, it is immutable because it will never be modified or deleted. The blockchain is trustless because nobody has to trust anybody else in order for the system to function. However, the blockchain also faces challenges. First, the calculation costing too much electricity is wasteful. Second, the size of the blockchain is becoming larger and difficult to store. Until December

in 2017, the bitcoin's blockchain has reached the size 149 GB. In the recent coming years, the size is predicted to increase exponentially and reach TB level [11]. Third, the scalability is low. When the network size increases, the transaction processing speed does not increase. These challenges may limit the blockchain for applications in a world-wide level.

2.6 Contributions

This research focuses on improving the low scalability problem of the blockchain. The purpose is to create a high scalable Blockchain that can work as the foundation for decentralized society. The contributions of this research are as follows:

1) the proposed protocol introduced a scalable protocol, to reduce the computation and increase the performance of the blockchain;

2) the proposed protocol gives a comparison of the performance of existing two protocol and the scalability of the proposed protocol is confirmed in an ideal simulation network consisting of 100 nodes;

3) a decentralized trusted timestamping is implemented as an example usage scenario for the future decentralized society, and its effectiveness is indicated.

In summary, this research proposed a scalable blockchain protocol based on PoS and sharding to solve the low scalability problem. In the same simulation network, this research compares the proposed protocol with two existing protocols to confirm the performances. This research also implemented a decentralized trusted timestamping service which could help to protect the digital data with high security, low cost, and most importantly, could meet the increasing requirements for digital data protection in the future.

Chapter 3

Scalable Blockchain Protocol Based on Proof of Stake and Sharding

3.1 Introduction

Blockchain, the technology behind Bitcoin since 2008, has become the core infrastructure for novel decentralized applications in recent years. Blockchain is a distributed and immutable public database that stores confirmed transactions in chronological order with high security [19]. At present, the blockchain technology has transited from generation 1.0 to 3.0. Blockchain 1.0, the first generation of the blockchain, is mainly related to cryptocurrency. When blockchain first came out, it severed as the database of cryptocurrencies such as Bitcoin. Blockchain 2.0, the second generation of the blockchain, is related to smart contracts. The applications of blockchain starts to be wider. Contracts here refer to financial and legal contracts. For example, bonded contracts, identification, and copyrights [20]. Blockchain 3.0, the third generation of the blockchain, is related to services beyond finance. Blockchain 3.0 provides infrastructure for applications in fields such as political, healthy, and art [21]. The transition of blockchain technology is due to its decentralization, high security, and immutable properties. Although these advantages widen blockchain's application areas, blockchain still faces a major barrier: low scalability.

Scalability in the Blockchain could be defined as the relationship between the network size and the transaction processing speed. A blockchain is scalable if the transaction processing speed scales with the network size. This means that if the network size increases and transaction processing speed also increases linearly, the blockchain is scalable. Otherwise, if the transaction processing speed does not increase with the network size, the blockchain is not scalable or low scalable. Low scalability is a primary limitation of blockchain technology [43, 52, 53]. In the case of Bitcoin's blockchain, the rate of transaction processing is 7 transaction per second (tps) [3]. In contrast, the average processing rate of payment system such as Paypal is 115 tps [4]. In VISA's case, the processing rate can reach a peak of 56,000 tps [5]. Although Bitcoin's network is large, however, only 1 block could be created in an average 10 minutes. This does not change no matter how large the network size is. Besides, the amount of transactions that can be included in a block is limited. These two reasons cause the low scalability of Bitcoin's blockchain. In other words, the processing rate limitation of blockchain is influenced by block size and block interval. Increasing block size could help improve the transaction throughput, but large blocks turn out to take longer propagation time. Reducing the block interval lowers the latency,

which is the time taken for a transaction to be confirmed; however, it raises block duplication rate [44, 52]. The major objective of this study was to develop a scalable blockchain protocol.

Figure 3.1 presents the comparison of two major consensus protocols with our proposed protocol. Figure 1(left) shows the flow of PoW protocol and Figure 1(middle) shows the flow of PoS algorithm. In a peer-to-peer network, the nodes agree on a new Block, including unconfirmed transactions from the transaction pool, by running PoW or PoS protocols. Figure 1(right) presents the overview of our proposed method. The main idea of the protocol is to divide unconfirmed transactions in the network into transaction shards and a peer-to-peer network into multiple network shards. Furthermore, transaction shards can be processed in parallel in the network shards with PoS consensus protocol to become middle blocks. Finally, middle blocks are included in a final BLOCK to be recorded on the blockchain.

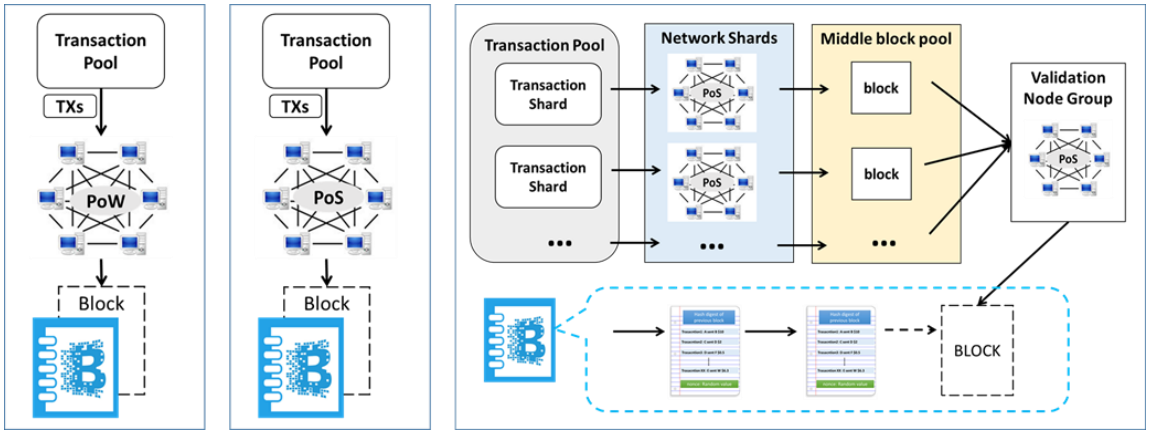


Figure 3.1 Overview of existing consensus protocols: PoW(left), PoS(middle), and the proposed protocol (right)

Table 3.1 Comparison of the two major consensus protocols and the proposed protocol

	Consensus Algorithm	**Scalability**	**Latency**	**Throughput**
Bitcoin's Protocol [10,11]	PoW	×	600 sec	3-7 tps
Peercoin's Protocol [16, 22]	hybrid PoW/PoS	×	600 sec	3 tps
Proposed Protocol	PoS + Sharding	✓	27 sec	36 tps

To evaluate the proposed method, we performed several simulation experiments. Table 3.1 shows the comparison of the proposed protocol with the two major existing protocols. For both Bitcoin's and Peercoin's [45] protocols, the latency is 600 seconds while the throughputs are less than 10 tps each. For our proposed protocol, the latency reduced to 27 seconds and the throughput reached 36 tps in a simulation network comprising 100 AWS EC2 instances. The evaluation experiments show that our proposed method is scalable, as the throughput scales with the size of the simulation network.

The contributions of this research are as follows:

- the proposed protocol ensures scalability;

- the proposed protocol introduced a sharding based PoS protocol, to reduce the computation and increase the performance of the blockchain;
- the proposed protocol also gives a comparison of the performance of existing PoW, PoS, and the proposed sharding based PoS protocols in an ideal simulation network consisting of 100 nodes to confirm the scalability of the proposed protocol.

3.2 Related concepts and works

The scalability problem is related to consensus protocols. This first part of this section explains and contrasts two main consensus protocols PoW and PoS. Then the section part of this section introduces related works.

3.2.1 Related concepts

3.2.1.1 PoW

The PoW is a consensus protocol used to maintain the security of cryptocurrencies. In the Bitcoin's case, miners (computational nodes) continue solving a mathematical puzzle as a competition. The fastest miner records a new block into the blockchain and receives bitcoins as a reward. The result of the competition is decided by the CPU power of the nodes. Theoretically, the larger the CPU power of a node, the higher its probability of recording a block successfully and receiving a corresponding reward. Equation (3.1) shows the mathematical puzzle, which is a hash puzzle. Miners calculate the hash with three parameters, hash of the previous block, new unconfirmed transactions, and nonce (a random number). Target is a hash of 256-bit length, starting with the expected number of leading zeros 46]. To solve this puzzle, miners repeat calculations with different nonce values until it satisfies eq. (3.1). In Bitcoin's blockchain, one-way hash algorithm SHA-256 is used. If an attacker intends to launch attacks on the blockchain, then the attacker would be required to perform as much computations as the rest nodes in the Bitcoin network. Therefore, the attack would not succeed unless the attacker owns more than half of the total CPU power of the entire Bitcoin network. This is known as the 51% attack

$$\text{Hash}\,(\text{Index}, \text{Previous Block Hash}, \text{Transactions}, \text{Timestamp}, \text{Nonce}) \leqslant \text{Target} \qquad (3.1)$$

The PoW protocol provides security protection to cryptocurrencies; however, the cost of hash calculations to produce new blocks cannot be ignored. Six hundred trillion SHA-256 hash computations are performed by Bitcoin network per second, which results in an estimated electricity

and hardware of over one million dollar worth per day without practical use [10, 47].

3.2.1.2 PoS

The PoS is one of the consensus protocols designed to replace PoW so that the future cryptocurrencies would not depend greatly on much energy consumption. Peercoin is the first cryptocurrency implemented PoS. In PoS, as shown in eq. (2), the generation of a new block is proportional to the concept of coin age rather than the CPU power. Coin age is the stake status defined as the amount of coins times holding period. The larger the coin age, the higher the probability of the node to record a block successfully. PoS do not need heavy hash calculations as PoW does. Therefore, PoS is considered much more cost effective than PoW.

$$\text{Hash}\,(\text{Index}, \text{Previous Block Hash}, \text{Transactions}, \text{Timestamp}) \leqslant \text{Coin age} * \text{Target} \qquad (3.2)$$

Since the high security provided in the case of PoW depends on a large number of calculations, people may have concern on PoS's security. However, the probability of 51% attack is lower in the case of PoS for two reasons. First, performing a 51% attack in PoS network is extremely expensive, requiring up to $50 million [48]. Second, the attack may turn out to be ineffective. This gives low incentive for attackers; thus, they may never attempt to attack the network.

Table 3.2 Comparison of PoW and PoS protocols

	PoW	**PoS**
Determiner	CPU power	Coin age
Cost	High	Low
51% attack concern	Potential possible	Almost impossible

Table 3.2 presents the main differences between the two consensus protocols. PoW protocol is based on CPU's power to perform large amount of computations. This consumes much electricity, resulting in high cost. PoS is based on the number of coin age rather than computation speed; therefore, the cost is low. In terms of 51% attack concern, it is proved that PoS is less likely to be attacked than PoW. The cost to perform a 51% attack in a PoS network is much higher than in a PoW network [51].

3.2.2 Related works

Recent years, several researches studied the scalability problem of the blockchain. First, Section 3.2.2.1 introduces some of these works. Then, Section 3.2.2.2 explains one of the significant concepts

that could help with the scalability improvement.

3.2.2.1 Previous researches

There have been several papers research the scalability problem of blockchain. In 2016, Eyal et al. proposed Bitcoin-NG to solve blockchain's scalability problem. Bitcoin-NG is a Byzantine fault tolerance based protocol [52]. Luu et al (2016) designed ELASTICO protocol to improve the scalability problem. ELASTICO is based on PoW, practical Byzantian fault tolerance and sharding [53]. OmniLedger (Eleftherios et al., 2017) implemented a hybrid consensus protocol and sharding was proposed as a possible solution [54]. In 2018, Zamani et al. introduced RapidChain based on Byzantine fault tolerance and sharding [55]. Three of the previous researches applied sharding, which is explained in next section.

3.2.2.2 Sharding

Sharding is a traditional technology used to partition the database. Sharding involves separating large databases into small data shards that can be both fast and easy to manage. Inspired from the database sharding concept, sharding has been pointed out as a technique capable of solving the blockchain's scalability problem. In the current blockchains, nodes are distributed around the world. Each node has to process all transactions and store the states of all the transactions in history. This provides high security but causes scalability concerns. In the case of Bitcoin, the processing rate is in the range of 3-7 tps only. In 2016, Luu et al. presented the ELASTICO protocol for open blockchains. This protocol was implemented using sharding technique. The technique employed in the protocol was to process transaction shards simultaneously in small size mining networks [53]. PoW was applied in creating nodes' identities and the consensus protocol was established on a standard Byzantine agreement protocol and sharding.

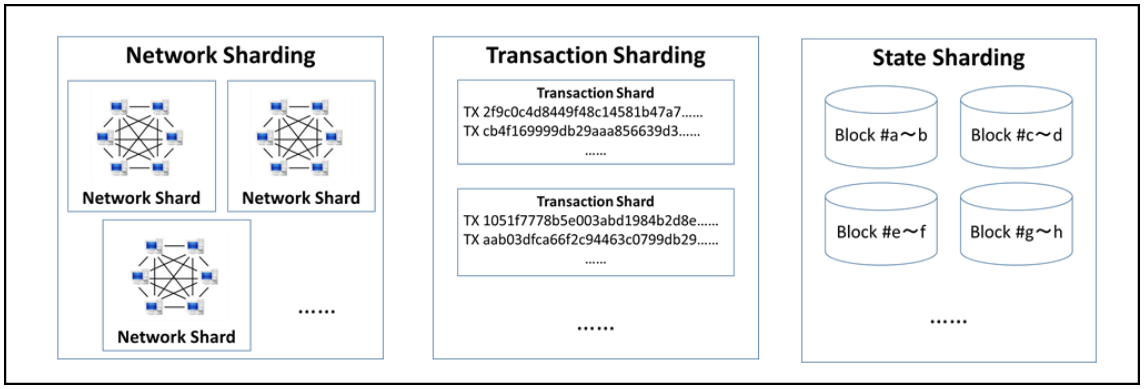


Figure 3.2 Three sharding strategies

Sharding technique could help blockchains become horizontally scalable, which implies that the

transaction processing rate would scale with the network size. There are three main sharding strategies: network sharding (Figure 3.2, left), transaction sharding (Figure 3.2, middle) and state sharding (Figure 3.2, right). Network sharding involves dividing the blockchain network into small network shards so that transactions can be processed in parallel. In transaction sharding, unconfirmed transactions are separated from the confirmed ones and grouped into small transaction shards. State sharding is the process of storing different states (i.e., account information) in different shards to reduce the storage burden of each node. Although state sharding offers a great benefit, need for communications across shards results in high complexity. In addition, since the state data are stored separately, the integrity of the data would be damaged if a shard is attacked. Therefore, state sharding was not considered in this research.

3.3 Proposed Method

In this section, we present the proposed sharding protocol for scalable blockchains. First, we give a detailed explanation of the various sharding strategies and then introduce the proposed protocol.

In network sharding, a mechanism is needed to decide how to separate nodes into shards. From the security perspective, centralization caused by the monopoly of certain nodes should be prevented. The mechanisms we use are 1) random formation and 2) reshuffling. 1) Random formation: In this mechanism, the network is divided into small randomly generated network shards. 2) Reshuffling: After a certain period, the nodes are shuffled and new network shards are formed. Random formation and shuffle prevent attacks by malicious nodes.

For transaction sharding, a mechanism is needed to determine how to create transaction shards to avoid double spending. Double spending can occur if a malicious user creates two transactions from the same input to generate two different outputs in different transaction shards to be processed. There are two main ways to avoid this problem. First, a shard should communicate with every other shard [49]. However, this increases the complexity. Second, transaction shards should be determined by the sender's address [49]. This means that the transactions with the same sender's address are put into the same shard so that double spending transactions can be detected easily. Thus, the complexity does not increase because cross-shard messages are not needed.

3.3.1 Overview of the proposed method

The proposed method employs a combination of sharding protocol and PoS consensus mechanism. Assume there is a network containing cn nodes and it is divided into c groups (i.e., network shards). Each group contains n nodes. In the c groups, $c - 1$ groups are regular groups and one group works as a validation node group. Blocks are of two types: lower-case and upper-case blocks. To distinguish

the two, lower-case ‘block’ represents the middle blocks and upper-case ‘BLOCK’ represents the final blocks. Middle blocks are generated by *c - 1* groups, and then these blocks are processed and combined in the validation group to produce the final BLOCKs, which are recorded on the blockchain. The overview is shown in Figure 3.3.

Middle block and final BLOCKs are produced in epochs. Each epoch includes four steps:

Step 1: Form node groups

Each node works in a group. Leader node is chosen when a node group is formed. The leader node collects the identities of other nodes in the group to create an identity list. This list is broadcast to other group leaders. This process reduces the node communication complexity from $O(n^2)$ to $O(cn)$.

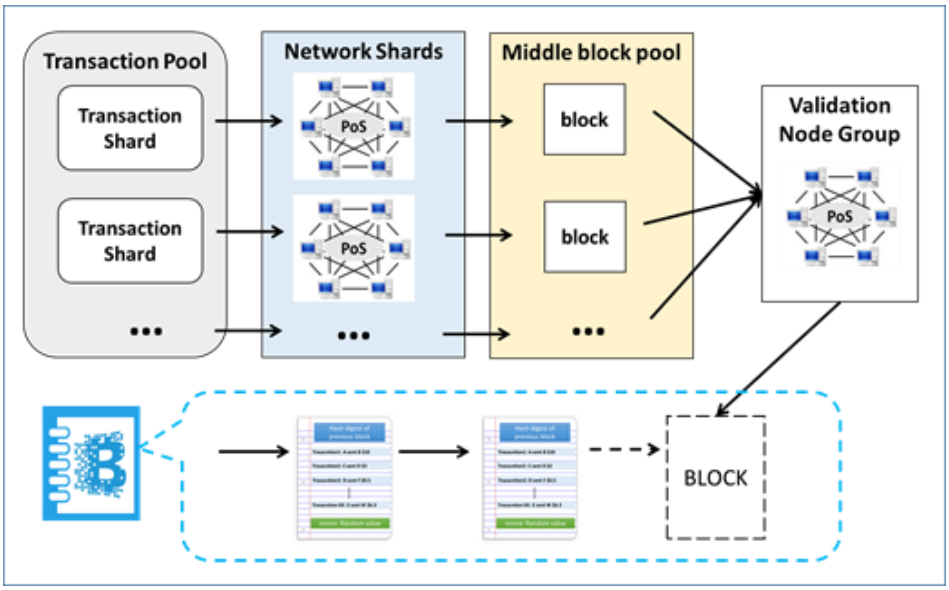


Figure 3.3 Overview of proposed method

Step 2: Create middle blocks

Internal group consensus is run in each node group to produce middle blocks. PoS consensus protocol is applied to determine the amount of coin age owned by each node. The node owning large coin age (i.e., coin amount times holding period) has a higher probability to successfully generate a new middle block.

Step 3: Generate final BLOCK

The middle blocks created by general node groups are collected and combined in the final validation group also by running a PoS consensus protocol, the final BLOCK is produced and broadcast to the whole blockchain network.

Step 4: Reshuffle the nodes

This step is executed every *t* epochs. All of the nodes are reshuffled to form new node groups.

3.3.2 Forming node groups

At first, all the nodes in the network form node groups. Assume there are n nodes in a group. All the nodes in the same node group are supposed to know each other's identities. Each node broadcasts its identity to all the other nodes is a simple way. However, this causes high message complexity of $O(n^2)$. In this study, we used this strategy to make the complexity low. The strategy is presented in Section 5. We assume there were c node groups in total with one group chosen as the final validation group randomly.

3.3.3 Create middle blocks

When the node group formation is completed, transaction shards are assigned to the $c - 1$ regular node groups. When separating transactions into shards, those that have the same sender's address are put in the same shard. This mechanism for transaction sharding has two advantages: 1) it prevents double spending and 2) it helps avoid cross-shard communication. Transaction shards are produced in regular node groups to create middle blocks, which will be processed in the final validation group. A middle block contains an index, previous BLOCK hash, a transaction shard and a timestamp.

3.3.4 Generating final BLOCKs

Middle blocks produced by regular node groups are collected, verified and combined by the final validation node group to generate the final BLOCK. A PoS consensus protocol is run to select a node whose final BLOCK is recorded in the blockchain. A final BLOCK contains an index of previous BLOCK hash, middle transactions and a timestamp.

3.3.5 Reshuffling the nodes

Nodes are reshuffled and new groups are formed every t epochs. Node reshuffling helps to prevent malicious nodes from taking control of a network shard, which reduce the risk of centralization. Thus, reshuffling helps to keep the network secure.

To evaluate the scalability of the proposed method, we conducted simulation experiments under different conditions.

3.4 Implementation and evaluation

We implemented the proposed protocol and conducted experiments to evaluate its scalability and those of the existing protocols in the same simulation network. The two aims of the experiments are as follows: 1) to confirm that the performance of the proposed protocol is in agreement with the theory, and 2) to compare the proposed protocol with existing consensus protocols PoW and PoS.

As introduced in Section 3.1, scalability means that the transaction processing speed scales with the size of the network almost linearly. Therefore, if the transaction processing amount increases linearly with the network size in the evaluation experiments, the proposed protocol could be confirmed to be scalable. Otherwise, the proposed protocol does not have the scalability.

3.4.1 Experiment setup

We implemented the proposed protocol using Node.js and conducted simulation experiments on Amazon EC2 to measure the performance. The size of the simulation network ranged from 20 to 100 t2.micro Amazon EC2 instances. Each instance performed as a single node with 1 vCPU and 1.0 GB of memory. Therefore, when the number of nodes increased, the computation power of the network also increased. The number of nodes n in a group varied among 5, 10 and 20, while the network size cn varied from 20 to 100. We conducted a total of 15 experiments in different settings to measure the scalability of the proposed protocol. To compare the proposed protocol with PoW and PoS protocols, we also conducted experiments on the two existing protocols in the same simulation network with up to 100 nodes.

3.4.2 Evaluation experiment

The experiments contain two parts. The first part is to evaluate the performance of the proposed protocol with regard to throughput and latency in different conditions and compare it with two existing protocols. We fixed the size of a node group as n = 5, 10, 20, and performed five experiments for each node group size for network size cn = 20, 40, 60, 80, 100. Thus, a total of fifteen experiments were performed. The second part is to evaluate the proposed protocol not only with regard to its throughput and latency, but also network shards' size, data size, and detailed latency.

First, for each node group size, we started a simulation network of 20 nodes and increased the network size four times, up to 100 nodes. We measured the latency for 10 BLOCKs and calculated the throughput. Second, we conducted experiments using the same simulation network for total nodes ranging from 20 to 100 for PoW and PoS protocols. We measured the latency for 10 PoW blocks and

10 PoS blocks.

Figure 3.4 compares the throughput and latency between the existing and the proposed protocols. As can be seen from the data bars in Figure 4, the PoW protocol has highest latency of around 60 seconds while PoS protocol has the smallest latency of 12 seconds. The latency of the proposed protocol is around 27 seconds. The lines show the throughput of the three protocols. As shown in Figure 3.4, when the network size increases, neither PoW nor PoS's transaction processing rate increases. However, the proposed method processes more transactions as the network size increases. When the network size was 100 nodes, the throughput of the proposed protocol reached 36 tps, which is higher than the throughput of PoW (3 tps) and PoS (14 tps). This result shows that the throughput of the proposed protocol scales with the network size. In particular, it confirms that our proposed method is scalable.

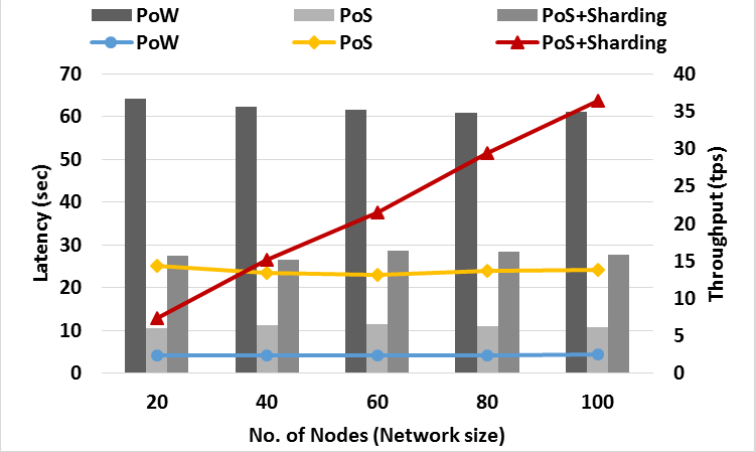


Figure 3.4 Throughput and latency comparison among PoW, PoS, and proposed method (Shard size = 5)

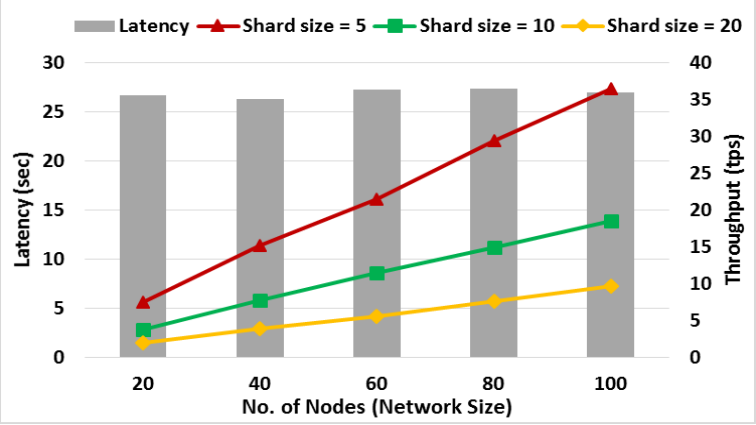


Figure 3.5 Impact of shard size on throughput and latency

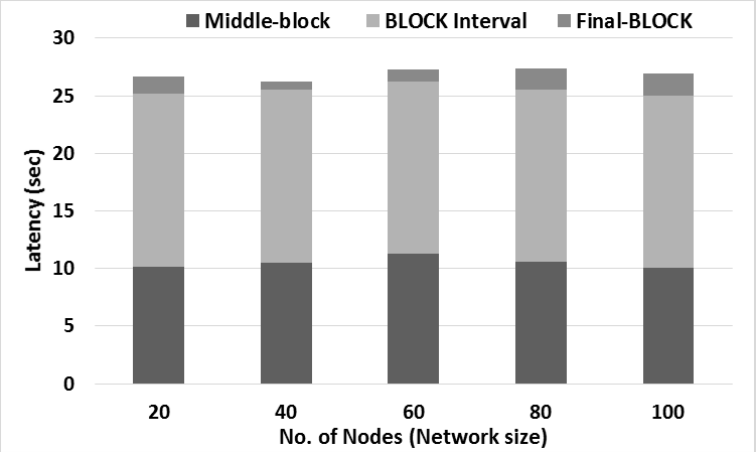


Figure 3.6 BLOCK latency of the proposed protocol

Figure 3.5 shows the impact of different network shards on the throughput and latency. For network shards (i.e., node groups), we used three different sizes of 5, 10 and 20 nodes. When the total network size increased from 20 to 100, the number of network shards varied. Figure 3.6 shows that the latency stayed around 27 seconds for different shard sizes while the throughput increased with the shard size. When the shard size was 5, the throughput increased faster than the other two conditions. Figure 3.6 shows detailed information on BLOCK latency for the proposed protocol. As shown in the figure, latency of proposed protocol mainly consists of three main parts: 1) latency to create middle blocks, 2) interval time between two BLOCKs, and 3) consensus on a final BLOCK. The sum of the three time periods gives the BLOCK latency for the proposed protocol, which is about 27 seconds.

3.5 Discussion

This section discusses the results presented in Section 3.4; analysis of the complexity and security of the proposed protocol.

3.5.1 Results and discussion

Figure 3.4 shows a comparison of the latency and throughput of the three protocols. Although the latency of the proposed method is more than double the PoS's latency, the proposed protocol has a higher throughput than PoS because the proposed protocol employs sharding techniques, which enable it to perform parallel transaction processing. Our proposed method shards the simulation network, the unconfirmed transactions, and process transaction shards parallel in network shards. Therefore, the final throughput is higher than both PoW and PoS. Distribution of transaction processing contributes to the higher throughput and scalability of the proposed protocol.

From the results shown in Figure 3.5, the throughput increased the fastest when the shard size was 5 because when the network size is fixed, the smaller the shard size is, the more is the number of network shard size. More network shards produce more middle blocks. When the block size is fixed, the size of the final BLOCK is decided by the number of network shards. Therefore, the final BLOCK contains more transactions when network shard is small and this contributes to higher throughput. Although in this experiment the throughput only reached 36tps; however, when the network size continues increasing, the throughput would also continue increasing and contribute to a higher scalability. As shown in Figure 3.5, for a certain network size, the throughputs are related to shard sizes. The number of nodes per shard should be decided carefully when the network size is large.

3.5.2 Complexity analysis

Assume there were *cn* nodes in the network. To know other nodes' identities, each node broadcasts its identity to other nodes and receives the identity information of other nodes in the network. This results in $O(n^2)$ message complexity. In the proposed protocol, *c* node groups are formed and each contains *n* nodes. To reduce the message complexity, we propose a mechanism to reduce both inner shard communication and cross shard communication.

In a given node group, if each node broadcasts its identity to other nodes, the message complexity would also be $O(n^2)$. To reduce the inner shard communication, a leader node is randomly selected and the identities of all the other $n - 1$ nodes are sent to the leader node. Therefore, nodes do not need to broadcast its identity in the group nor to the whole network. The identity information is shared among leader nodes. A non-leader node could ask for other nodes' identities from the leader node of its group.

Among the *c* node groups, cross-shard communication is needed when transaction shards are assigned to them. Without the communication between the node groups, two transactions with the same input address may be processed in two different node groups. Cross-shard communication helps prevent the double spending problem; however, it results in high message complexity. Our method to avoid this kind of cross-shard communication is to put transactions with the same input addresses in the same transaction shard. Therefore, the double spending problem could be avoided without relying on cross-shard communication.

Cross-shard communication cannot be avoided since leader nodes need to share identity information about all the nodes. In our proposed protocol, we reduce message complexity by reducing inner group communication and cross shard communication complexities from $O(n^2)$ to $O(cn)$.

3.5.3 Security analysis on 51% attack

51% attack is one of the security concerns of blockchain. In the case of using PoW as the consensus protocol, the node(s) that take control of more than 51% of the total CPU power can launch malicious attacks successfully [3] and an invest of 90 million USD is needed [56]. A 51% attack is related to cost and incentive. PoS is considered more secure in this regard. Larimer D. (2013) confirmed that it is much more costly to perform a 51% attack in a PoS based network than a PoW based network [50]. While 51% attack in PoW based networks require huge cost and large amount of hard ware equipment, in the case of PoS based networks, not only the cost (control of over 51% possessions) is required to launch a successful attack but also but also the amount of coin possessed and the holding period.

Concerning the incentive, even if a 51% attack succeeded, the attacker would not benefit much because of the mechanism in a PoS network. This leaves low incentive for malicious nodes to attack the network.

In the proposed method, we implemented three mechanisms to achieve high security. They are 1) randomness - the formation of node groups and the choice of leader nodes are randomly selected, 2) reshuffle - for every t epoch, all the nodes are reshuffled to form new groups, and 3) coin age limitation - effective coin age is limited and coin age is reset to zero once successfully used. These three mechanisms help to prevent malicious nodes from taking control of the network shard and lower the incentive to launch a 51% attack.

Chapter 4

Decentralized Trusted Timestamping

4.1 Introduction

Considerable intellectual property is created and shared everyday on the Internet. For example, approximately 300 hours of videos are uploaded on Youtube every minute [57], and on average, 1.83 million photos are uploaded publicly on Flickr every day [58]. Such digital intellectual property can be tampered with or forged relatively easily. One solution to this problem is a technique called "trusted timestamping". This process can track the creation and modification time of digital data to ensure the existence and integrity of the data. Trusted timestamping is issued by a central Time-Stamping Authority (TSA). Users send a digital file to the TSA, where it is signed with the current time digitally [59]. However, this process has security problems. For example, if the TSA's timestamp server is hacked, the timestamp will be unreliable. Recently, decentralized trusted timestamping has been implemented to address this problem [65]. Currently, there are several decentralized timestamping services based on Bitcoin's peer-to-peer digital currency infrastructure, called "blockchain". Essentially, the blockchain is a distributed database that does not rely on central servers, i.e., it is decentralized [37]. The blockchain stores all confirmed Bitcoin transactions [61]. The existence and integrity of transactions are protected, i.e., the transactions cannot be tampered with or forged. Based on such features, web services, such as BTProof [62] and Proof of Existence (PoE) [67], have implemented decentralized timestamping. However, there are two significant problems with such applications: 1) only 40 bytes of data can be stored in a transaction in the blockchain, and 2) searching services using related keywords have not been implemented. This study proposes a methodology for storing a maximum of N*20 bytes of data and implements a search function for stored data. A comparison of the TSA, the current trusted timestamping process, and the proposed method is shown in Figure 4.1.

Existing web-based trusted timestamp services embed a maximum of 40-byte hash in the blockchain. The hash data cannot be reversed, which makes it difficult to determine the creators and other related information. To store such information, we propose a methodology for storing a maximum of N*20 bytes in the blockchain. For creators who prefer not to timestamp their data anonymously, the proposed method can embed related information (e.g., file name, creators' name, and comments) and the hash of the digital data. Note that the related information is decrypted in plain text. In addition to expanding the storage space, we also implement a search service that allows users to search for digital data by keywords. This search service aims to simplify the retrieval of intellectual property.

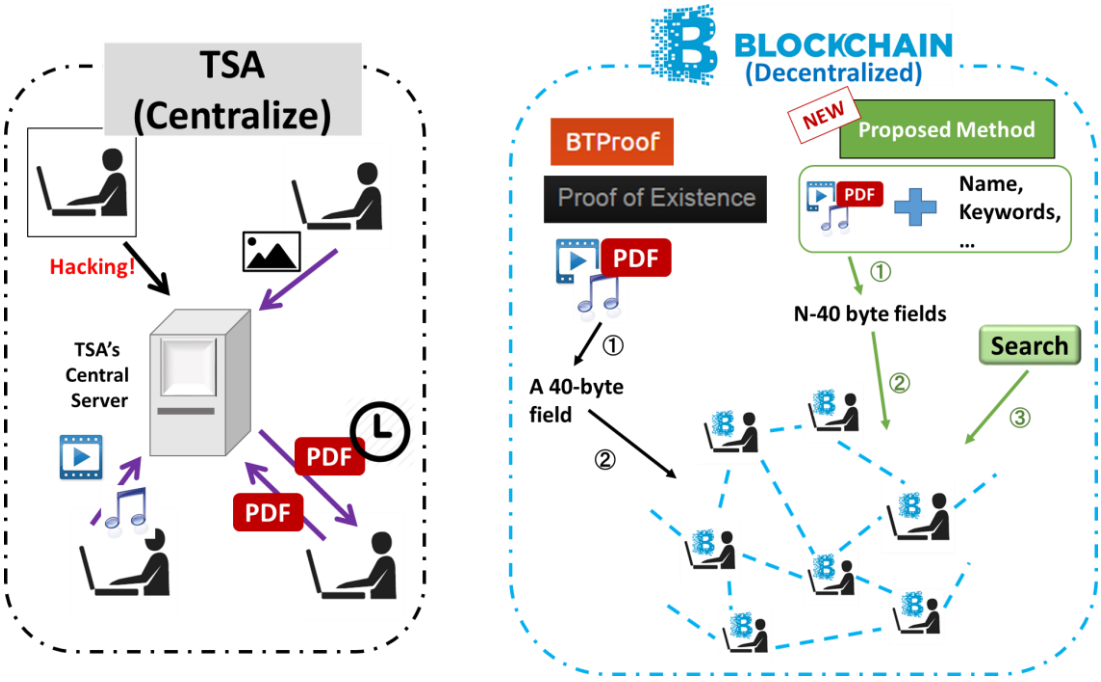


Figure 4.1 Concept of trusted timestamping based on TSA and blockchain

We conducted three experiments to evaluate the proposed method. We set N = 3 and conducted an experiment to evaluate the proposed method and verify the existence and integrity of a digital file. We also evaluated the effectiveness of the search service. In addition, we evaluated the cost and computation time (broadcast time) of the proposed method. The experimental results show that the proposed method can prove the existence of intellectual property at a particular time and that the intellectual property has not been modified since that time (integrity). Results also indicate that the search service can search for intellectual property documents using keywords. We also find that the proposed method implements decentralized trusted timestamping in an average time of 20 min at a possible cost of 0.24 USD.

The remainder of this paper is organized as follows. In Section 2, we provide an overview of Bitcoin and the blockchain, and describe current decentralized trusted timestamping services. In Section 3 the proposed method is described in detail. Experimental results are presented and discussed in Section 4, and conclusions are presented in Section 5.

4.2 Related work

4.2.1 Blockchain's structure

The blockchain is the public ledger of Bitcoin and it has three important features.

1) It is equivalent to a large record book that contains a massive number of entries. All confirmed Bitcoin transactions are recorded in the blockchain. Although nearly all entries describe Bitcoin transactions, it is possible to record other information in the blockchain.

2) No individual or entity controls the blockchain. It is shared among all Bitcoin users; thus, whenever something is written on the blockchain, agreement among many records is required.

3) It is extremely difficult to tamper with or forge records stored in the blockchain. Multiple copies of the blockchain are owned by numerous people around the world; therefore, forging or tampering requires modification of all blockchain records in various locations, which is exceedingly difficult.

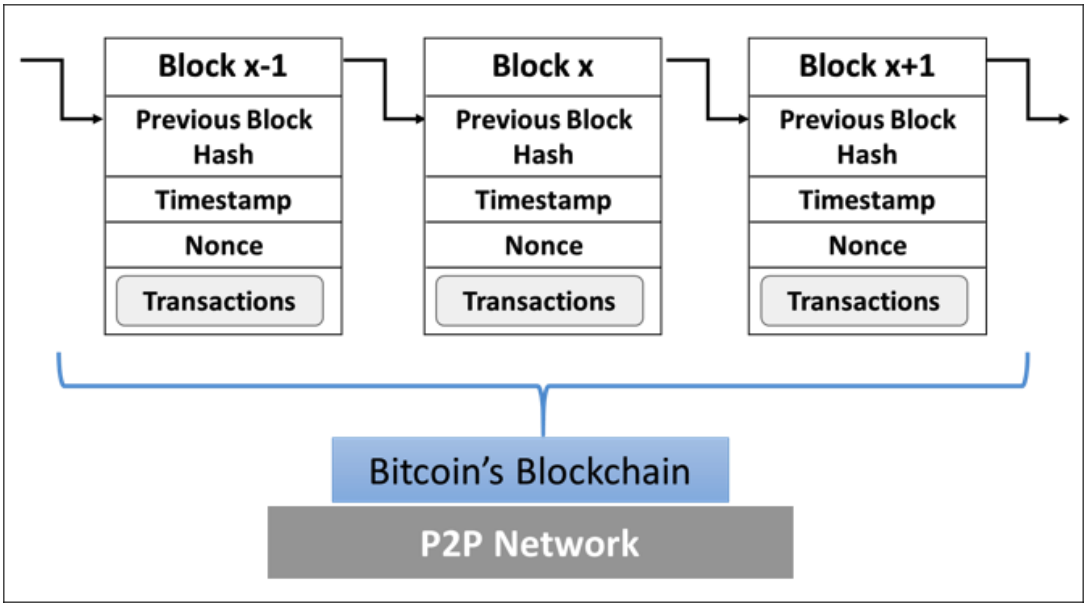


Figure 4.2 The structure of Bitcoin's Blockchain

The existence and integrity of data in the blockchain are guaranteed by the above features. Existence means that the data have actually existed since a certain time. Integrity means that the data have not been altered after a certain time. Transactions are stored in chronological order in the blockchain. Figure 4.2 illustrates the features and data contained in the blockchain. The blockchain consists of nearly 400,000 blocks, and each block primarily contains 1) the previous block's hash, 2) a timestamp, 3) a nonce, and 4) hash strings for approximately 500 Bitcoin transactions.

Figure 4.3 shows the data contained in a Bitcoin transaction. Transaction data primarily include

version, input, output and locktime. The output field contains the number of transferred Bitcoins and the receivers' addresses.

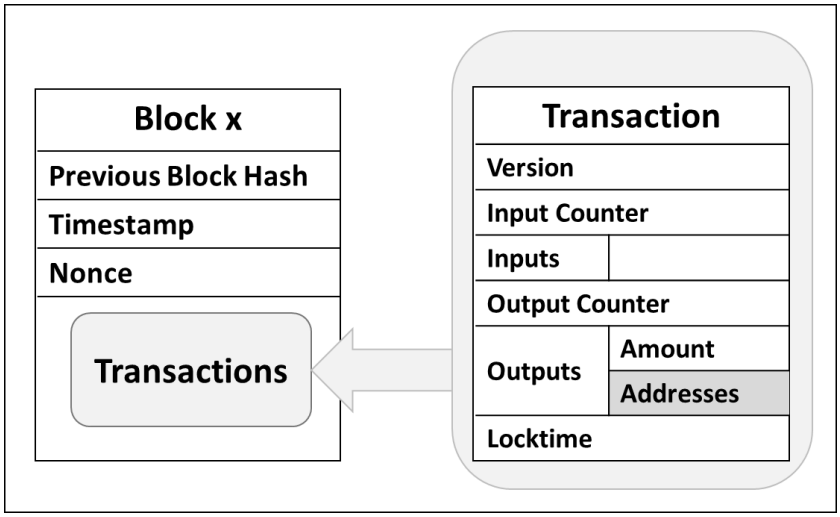


Figure 4.3 Example Bitcoin transaction in the blockchain

Bitcoin has three primary types of transactions, i.e., common, aggregating, and distributing [61].

1) Common transactions are simple payments that have one input and two outputs (output 0 is sent to the receiver and output 1 sends the change back to the sender).

2) Aggregating transactions have multiple inputs and a single output.

3) Distributing transactions are transactions that distribute one input to multiple outputs.

In this study, we use distributing transactions.

4.2.2 Intellectual property

According to the World Intellectual Property Organization (WIPO), intellectual property is a term referring to creations of the mind, including inventions, literary and artistic work, designs, symbols and commercial images, which are protected by law through patents, copyright and trademarks [64]. Currently, compared to traditional paper records, tremendous amounts of digital intellectual properties are created and shared daily across the Internet. Digital files are more convenient for storage and retrieval; however they can be tampered with or forged. The existence and integrity of such digital files should be proven to protect digital intellectual property. One solution is trusted timestamping [59].

4.2.3 Current trusted timestamping services

Trusted timestamps are issued by a central TSA. A user sends a digital file to the TSA, where it is digitally signed with the current time. The protocols of this technology were specified in RFC 3161

[60]. Before we explain how trusted timestamps are issued, we discuss the concept of a hash. Hash functions are cryptographically secure functions [66]. For digital data, the hash value is first calculated using a hash function. Hash functions are one-way functions that are collision-resistant [63], which means that it is impossible to obtain the original digital data from the hash values. Using hash functions increases security when sending data to the TSA. Note that the SHA-256 and RIPEMD-160 hash functions are used in Bitcoin's protocol. A 256-bit hash is calculated using the SHA-256 hash function and a 160-bit hash is calculated using RIPEMD-160 when a shorter hash value is required.

There are several web applications implemented using different methods, e.g., BTProof and PoE, for decentralized timestamps based on the blockchain. They help prove that a digital file existed at a point in time (existence) and that it has not being altered since the given time (integrity). Note that BTProof and PoE require a digital file as input.

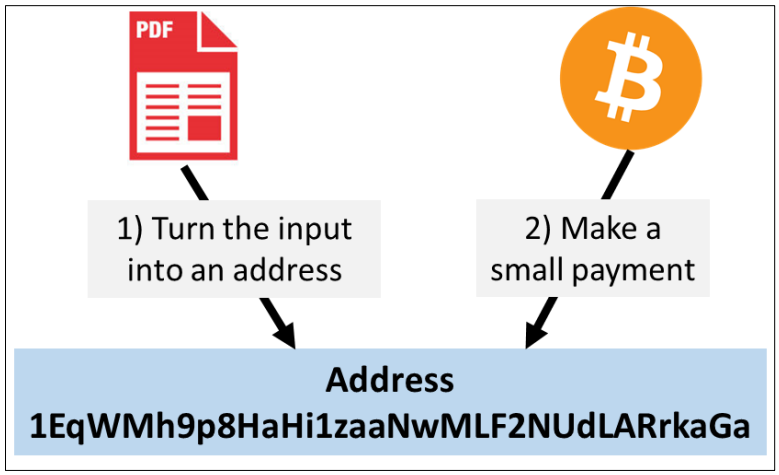


Figure 4.4 BTProof process

In BTProof's trusted timestamping, the input file is first hashed and converted into a Bitcoin address at first. As shown in Figure 4.4, an address is a string of characters and numbers. Then, by making a small payment to the address, the transaction and thus the hash are stored in the blockchain. Transactions in the blockchain are extremely difficult to tamper with or forge; therefore, the digital file is stored securely [62]. However, the hashed digital content cannot be reversed, which means that it is impossible to determine if any special data have been stored in the transaction. If additional information can be stored in part of a transaction and can be converted to plain text, such information can be identified and searched.

PoE implements Bitcoin's *script language* to embed digital files in the blockchain. Bitcoin transactions are validated by executing a script written in a Forth-like scripting language. Currently, most Bitcoin transactions have the form "A pays B". Such transactions are based on scripts referred to as Pay-to-Public-Key-Hash (P2PKH) scripts. However, Bitcoin transactions are not limited to the "A pays B" form. There are five standard types of transaction scripts: P2PKH, public-key, multi-signature

(limited to 15 keys), pay-to-script-hash (P2SH), and data output (OP_RETURN). PoE applies two types of these, i.e., P2PKH and OP_RETURN [61]. Figure 4.5 shows these two types of script.

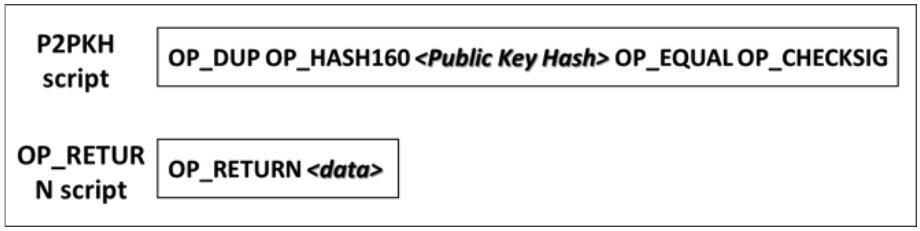


Figure 4.5 P2PKH and OP_RETURN scripts

In PoE, a document is also hashed; however, it is hashed to a 32-byte string rather than a Bitcoin address. Then the 32-byte string is embedded in the scriptPubKey field of the transaction. Thus, a special transaction is constructed. After broadcasting this special transaction, the 32-byte hash is stored in the blockchain with the special transaction, thereby making it difficult to tamper with or forged PoE also use "DOCPROOF" as a marker for their transactions by placing it at the beginning of the 32-byte string, which makes it easier to search for transactions [67]. Figure 4.6 shows the format of the embedded information created by PoE.

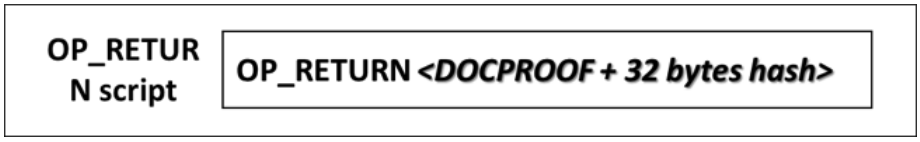


Figure 4.6 PoE script

4.2.4 Problems with current services and the proposed solutions

BTProof and PoE provide decentralized trusted timestamping based on the blockchain. However, PoE's method is limited to 40 bytes of data in a transaction. In addition, neither service can search for digital data using related keywords. To address these problems, we propose a method to store a maximum of N*20 bytes of data in transactions with N data fields. In addition, we implement a search service to enable document retrieval using related information (e.g., file name, creator name, related keywords).

4.3 Proposed Method

Although similar web services (e.g., BTProof and PoE) have been implemented, they are limited by the size of the data recorded in a transaction, and the data cannot be searched using related keywords. The main objective of this study is to implement decentralized trusted timestamping to help protect the existence and integrity of intellectual property using the blockchain by expanding the storage space

from 40 bytes to N*20 bytes. In addition, we provided a related keyword search service for intellectual property.

4.3.1 Timestamping processes

As discussed in Section 4.2, there are three primary types of Bitcoin transactions (common, aggregating, and distributing transaction). We apply distributing transaction in proposed method to distribute a single input into N outputs. We chose the BTProof method to embed digital data in a Bitcoin transaction. We did not choose the PoE method, which embeds digital data using the OP_RETURN script, because multiple outputs with OP_RETURN scripts tend to be rejected by the blockchain. Transactions with multiple OP_RETURN outputs are recognized as "strange transactions", which have a high probability of not being recorded in the blockchain.

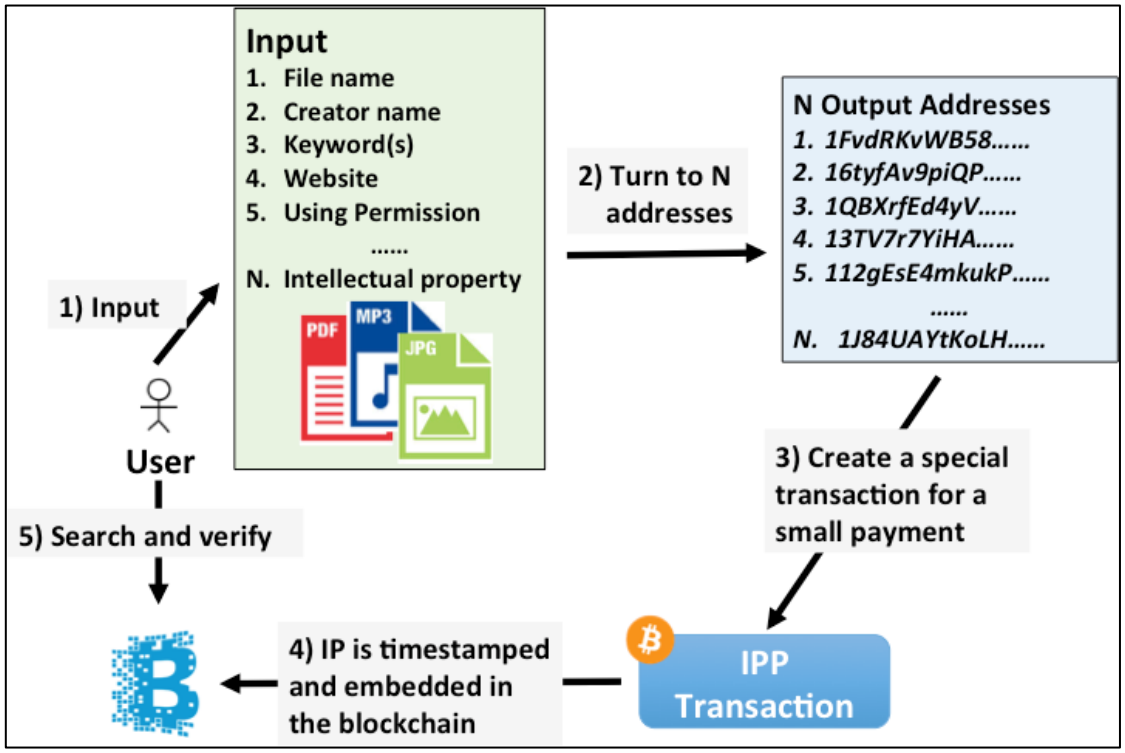


Figure 4.7 Processes of the proposed method

Figure 4.7 illustrates the process of the proposed method. First, the user inputs the related information on an intellectual property and selects the digital file. The N-inputs are then converted to N-Bitcoin addresses. A special transaction is created to make small payments to the N-addresses, which we refer to as an Intellectual Property Proof Transaction (IPP-TX). Once the IPP-TX is recorded, the intellectual property is safely timestamped and embedded in the blockchain. After Step 4 (Figure 4.7) is completed, users can search for and verify the intellectual property.

In Step 2, we apply different methods to 1 ~ N-1 inputs and the Nth input. The 1 ~ N-1 inputs are initially converted to hexadecimal numbers. Then, zeros are added to make all hexadecimal numbers 20 bytes. Finally, N-1 Bitcoin addresses are produced using Bitcoin's Base58 encoding scheme. For the Nth input, the digital file is hashed by the RIPEMD-160 function to obtain a 160-bit (20-byte) string. The Nth address is created by encoding this string using Base58.

4.3.2 Verification process

Verification of the existence and integrity of a digital file is based on a hash function. As discussed in Section 2.3, hash functions are one-way functions. We employ the RIPEMD-160 function. Even a small change in the original file creates a completely different hash value than that of the original file. Using this feature, a digital file's existence and integrity can be proved. Figure 4.8 shows the verification processes.

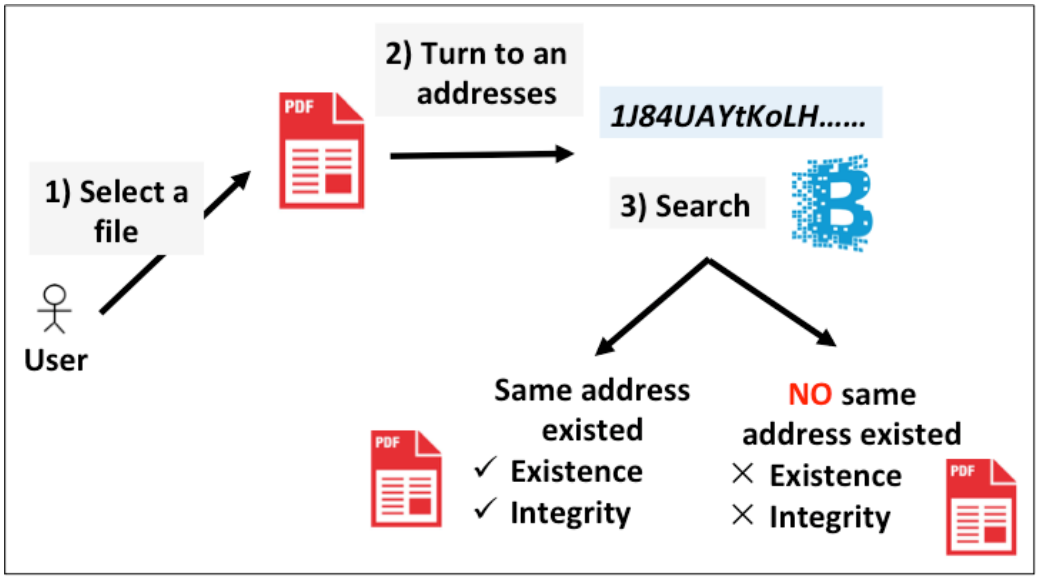


Figure 4.8 Verification process

To verify if an intellectual property has been altered, we repeat the processes to handle the Nth address (Section 3.1) with this file to obtain an address. The address is then searched from the blockchain. If the same address has been recorded, the existence and integrity of the file can be proved; otherwise, there is a high probability that this file has been tampered with or forged.

4.3.3 Search service implementation

Current decentralized trusted timestamping services' transactions cannot be searched without the

origin file because they only include the hash strings, which cannot be reversed. In the proposed method, we include related information, such as file name, creator name, and keywords; thus, intellectual property can be searched for using related keywords. The proposed search service is convenient when searching for specific intellectual property. It may also be helpful when decentralized trusted timestamping is used for patent protection, because it allows searching for patents.

To evaluate our method, we conducted experiments to determine if the timestamping processes (Section 4.3.1), verification processes (Section 4.3.2), and search service are effective. We also performed experiments to evaluate our method relative to time and cost.

4.4 Evaluation Experiments

This section explains our evaluation experiments and discusses the results. The purpose of the first experiment was to evaluate the effectiveness of the proposed method by broadcasting an IPP transaction and determine if it was recorded in the blockchain. The second experiment evaluated the proposed search function, and the third experiment evaluated the proposed method relative to time and cost.

4.4.1 Effectiveness of the proposed trusted timestamping

4.4.1.1 Timestamping

The first experiment was performed to determine if the proposed IPP transaction with extra intellectual property information was recorded in the blockchain successfully.

Here, we set N = 3 and used a PDF file. The three required inputs were 1) file and creator names, 2) keywords, 3) and the digital file (Figure 4.9). The three inputs were converted into three addresses (Section 3.1). Then, a transaction was created to pay small amounts to the three addresses. We set the total cost of the payment to 0.00103 BTC (0.47 USD, 56.71 JPY). Figure 4.9 shows that the transaction was stored in the blockchain successfully.

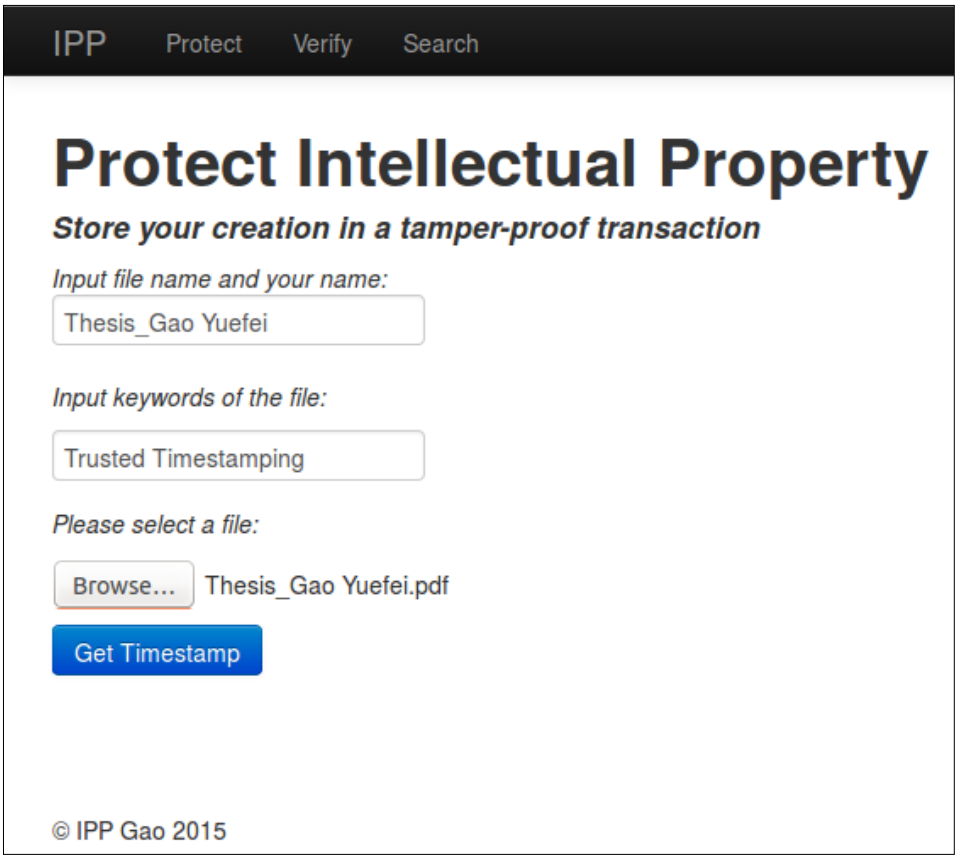


Figure 4.9 Example inputs

In Figure 4.10, 1) is the transaction ID (the hash of the transaction), which can be used for search; 2) shows the three output addresses (including extra information on the intellectual property); and 3) shows the received time (the time at which the transaction was broadcast to the network), the block where this transaction was stored, and the time at which it was included in the blockchain. We consider the time at which the transaction is included in the blockchain as trusted timestamping. Thus, in Figure 10, the trusted timestamp is 2015-12-04 05:29:42. In addition, 4) is the total i, i.e., the cost that can be set. Generally, the higher the cost the faster a transaction can be included in the blockchain.

The results of decrypting the three addresses shown in 2) in Figure 4.10 are shown in Figure 4.11. By comparing with Figure 9, it can be seen that outputs 1 and 2 are decrypted to the original plain text. Output 3 is a messy code because it was created using the one-way hash function. From the decoding results in Figure 11, we know that a person named Gao Yuefei finished a thesis about trusted timestamping before the time point 05:29:42 on December 4, 2015. These results might serve as a trusted proof in the future.

The results of this experiment indicate the effectiveness of the proposed IPP transaction. The information related to a digital file can be stored in the blockchain by converting it to Bitcoin addresses.

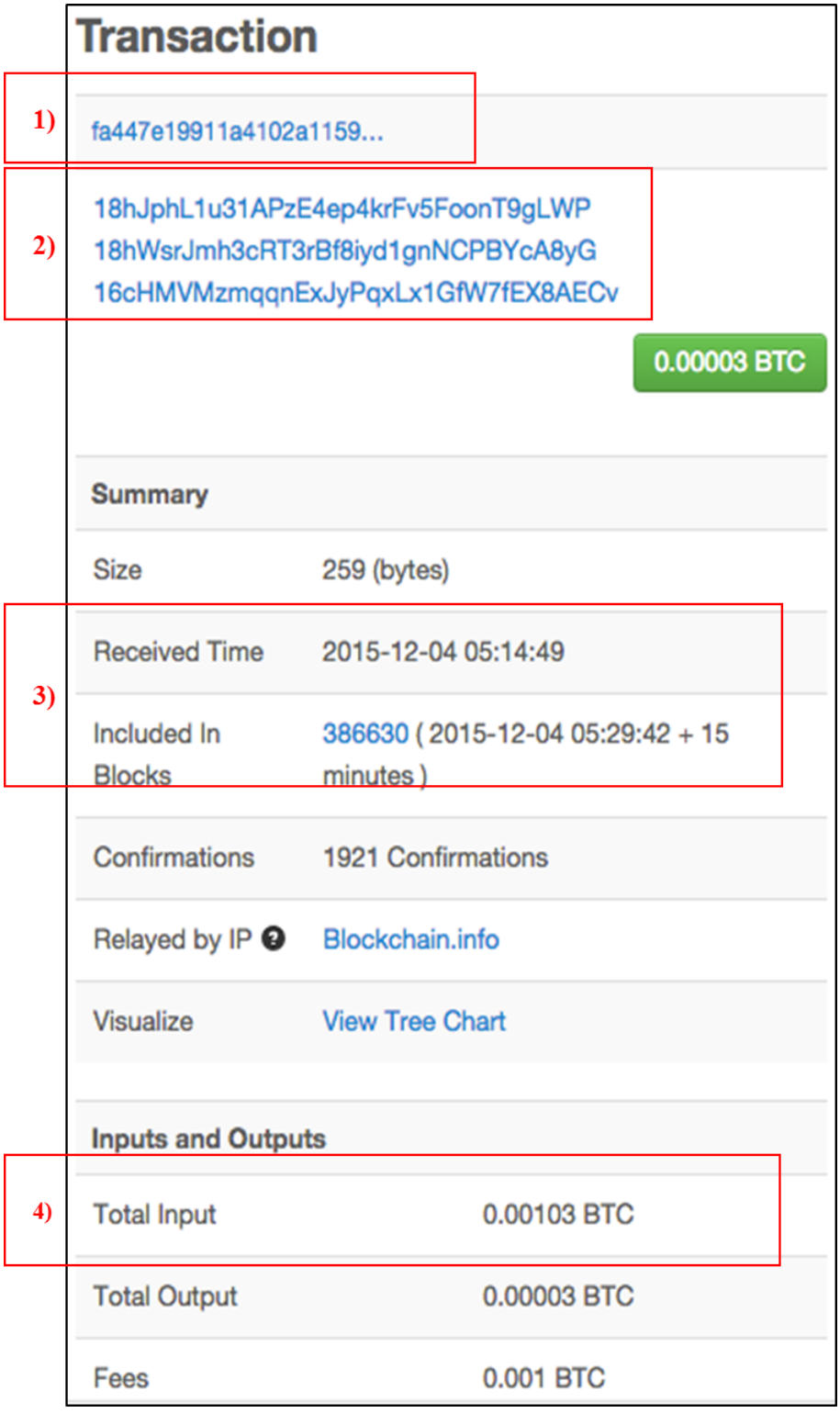


Figure 4.10 Recorded transaction with additional intellectual property data

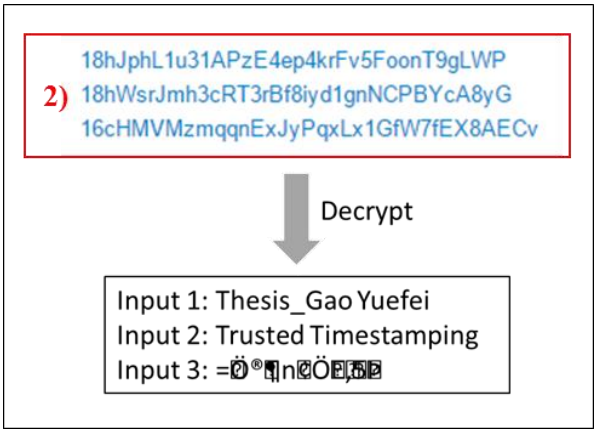


Figure 4.11 Results of the address decryption

4.4.1.2 Verification

This experiment was performed to determine if any modification was made to the intellectual property after it was timestamped. The verification page is shown in Figure 4.12.

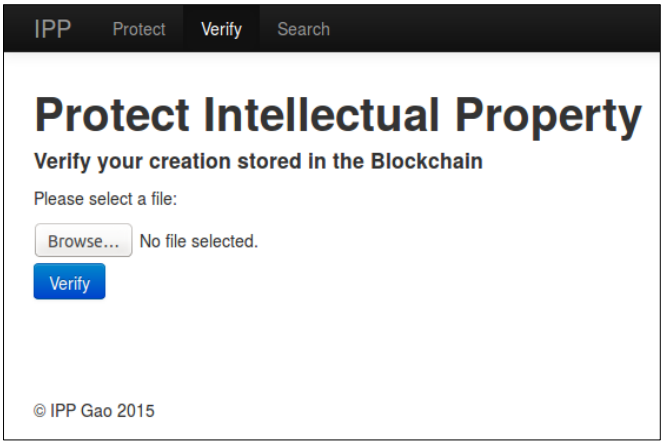


Figure 4.12 Verification page

Two PDF files were used in this experiment. One is the same as the original file and the other has a slight modification to the title (Figure 4.13). First the original file is input and the "Verify" button is clicked. Then, the same process is performed for the modified file. The verification results are shown in Figure 4.14. As can be seen, File 1 is the same as the original file (it has not been altered since the time of the trusted timestamp). However, File 2, generates a different address that is not found in the blockchain. This implies that File 2 may have been tampered with or forged.

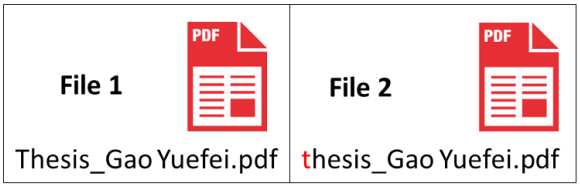


Figure 4.13 Files used for verification

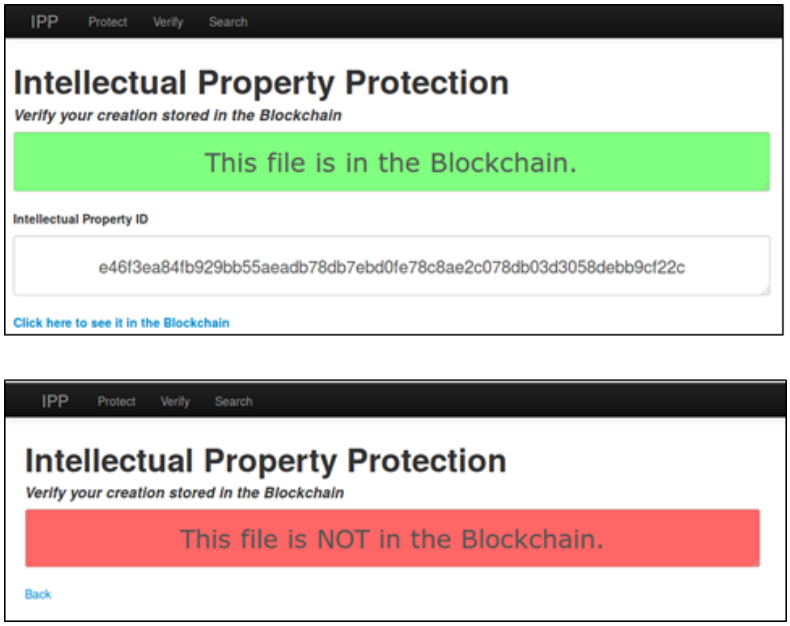


Figure 4.14 Verification results

The results of this experiment indicate that the trusted timestamping is effective for verifying that a digital file has not been modified since a certain time.

4.4.2 Search function evaluation

We implemented the search service to allow the user to search for intellectual property by keywords. The search page is shown in Figure 4.15.

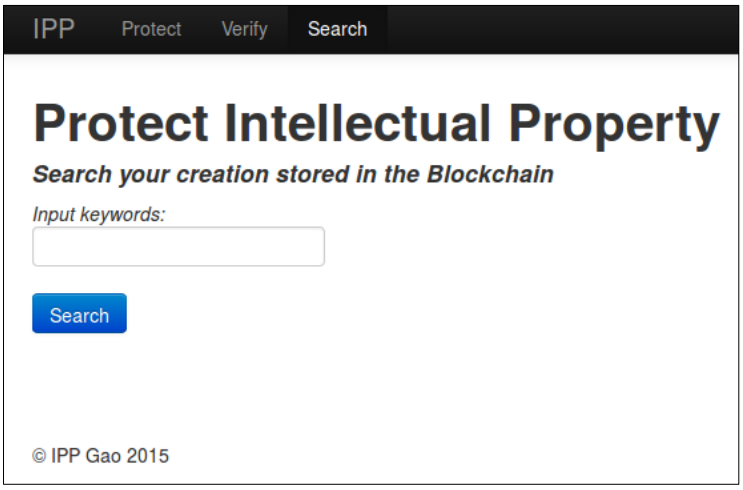


Figure 4.15 Search page

The user inputs a keyword and clicks the "Search" button. The keyword is searched from the blockchain to see if any transactions contain the keyword. The result is then shown to the user. The search service provides an easy way to retrieve the proposed transactions. For both creators and general users, this allows them to search for transactions using a certain keyword.

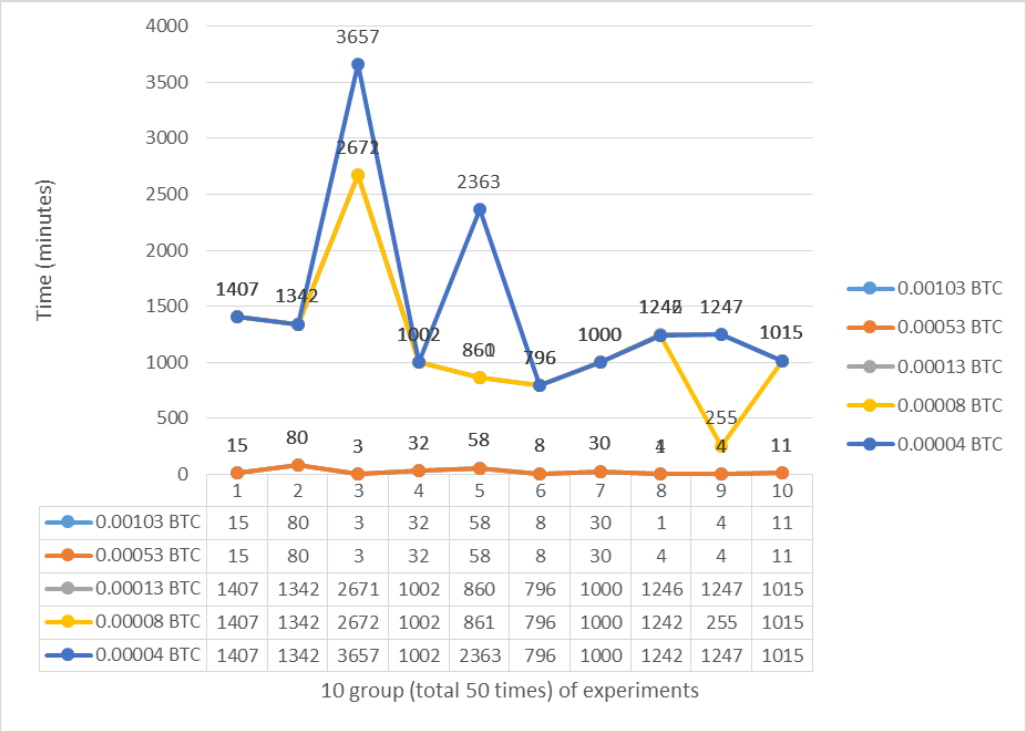
4.4.3 Time and cost evaluation

This experiment was performed to determine the relationship between cost and time. A PDF file was used in this experiment. As shown in Table 4.1, we used 10 groups, each containing five proposed IPP transactions with five different costs. Thus, this experiment was conducted 50 times. Here the costs contains two parts: 1) transfer amount: cost for creating a transaction; 2) transaction fee: cost for the transaction to be recorded into the blockchain. The five transactions in each group were broadcasted to Bitcoin's P2P network approximately simultaneously. This time point is also when the transactions were received by the blockchain (received time). Then we collected another time point for the transactions when they were recorded in the blockchain (included time). We then calculated the period between these two times. The results of the 50 experiments are shown in Figure 4.16.

Table 4.1 Example of a group of transactions

	TX 1	**TX 2**	**TX 3**	**TX 4**	**TX 5**
Cost (BTC)	1.03×10^{-3}	5.3×10^{-4}	1.3×10^{-4}	8×10^{-5}	4×10^{-5}

(TX: Transaction, 1 BTC ≈ 448.19 USD)



	1	2	3	4	5	6	7	8	9	10
0.00103 BTC	15	80	3	32	58	8	30	1	4	11
0.00053 BTC	15	80	3	32	58	8	30	4	4	11
0.00013 BTC	1407	1342	2671	1002	860	796	1000	1246	1247	1015
0.00008 BTC	1407	1342	2672	1002	861	796	1000	1242	255	1015
0.00004 BTC	1407	1342	3657	1002	2363	796	1000	1242	1247	1015

Figure 4.16 Relationship between time and cost

Figure 4.16 shows the relationship between time and cost for the proposed method. As can be seen, transactions with higher costs tend to be recorded in the blockchain in a shorter time. When the costs are set to 0.00103 BTC (0.46 USD) or 0.00053 (0.24 USD), it takes an average of 24 min to obtain a trusted timestamp using the proposed method. When the cost is less (e.g., 0.00013 BTC, 0.00008 BTC, and 0.00004 BTC), the average time increases to greater than 1000 min. The longest time from a transaction being received to being recorded was 3657 min.

The results of this experiment show how cost influences time. Although cost is not the only factor that contributes to time, it is one of the factors that can be set and controlled in the proposed method. By performing experiments relative to time and cost, we attempted to determine an appropriate cost by which transactions can be recorded in an acceptable average time. The proposed method timestamps a digital file in an average time of 20 min with the possible cost of 0.00053 BTC (0.24 USD).

4.4.4 Discussion

Decentralized trusted timestamping services based on the blockchain provide proof of digital data's existence and integrity without relying on a third party (e.g. TSA). We have proposed a new type of

transaction embedded with extra information about a digital file by expanding the storage space from 40 bytes to N*20 bytes. This information is converted to Bitcoin addresses and stored in the blockchain. It is generally known that Bitcoin's transactions are secure in the blockchain, which means that the digital file is also safe.

Our experimental results indicate that the proposed method is effective in proving the existence and integrity of a digital file. In addition, retrieval of an embedded file can be searched using both the transaction ID and keywords. It is possible to obtain a timestamp in an average time of 20 min with a possible cost of approximately 0.24 USD. Note that, in this paper, we only considered cost as a factor related to time. Other factors such as network speed, computer power and difficulty in creating a block are also potential factors. However, such factors were extremely difficult to change in our experiments, which is why the five different transactions in one group were not broadcast precisely simultaneously. The experiment was repeated 10 times to obtain the average time. This may help to exclude the influence of other potential factors.

Chapter 5

Summary

This research focused on developing a protocol to solve the low scalability problem of the blockchain, and this could make the blockchain works better as a decentralized platform for the future society. First, sharding and PoS techniques were applied as main mechanism to implement the proposed high scalable blockchain protocol. Then, a decentralized trusted timestamping was proposed as an application based on the blockchain to protect the intellectual property which are created and shared everyday on the Internet.

In chapter 3, the proposed scalable blockchain protocol was presented as a possible solution to solve the scalability problem. The main idea of the protocol was to divide unconfirmed transactions in the network into transaction shards and a peer-to-peer network into multiple network shards. Furthermore, transaction shards were processed in parallel in the network shards with PoS consensus protocol to reach a higher throughput. Experiments were performed to confirm the scalability. The latency was around 27 seconds and the throughput reached 36 tps in a simulation network which size was 100 AWS EC2 nodes. The main contribution was to develop a sharding based PoS blockchain protocol with better scalability performance.

In Chapter 4, a decentralized trusted timestamping was introduced for intellectual property rights protection. The idea was to change the related information of digital files into Bitcoin addresses and store them in a special transaction on the blockchain. The proposed method expands the storage from only 40 bytes to a maximum of N * 20 bytes and the digital files could be stored in the N fields securely. Also, a search service was implemented to allow searching by using keywords to make the retrieval more convenient. Evaluation experiments were performed to confirm the effectiveness and indicate the possible cost and time for the proposed decentralized trusted timestamping.

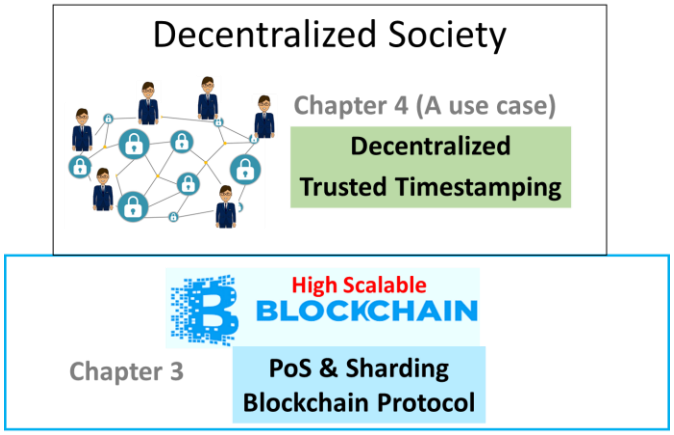


Figure 5.1 The summary of the proposed methods

In summary, Figure 5.1 presents the summary of this dissertation. To solve the low scalability problem of the blockchain, a high scalable blockchain based on PoS and sharding was proposed and implemented. Experiments were performed to confirm its scalability. This high scalable blockchain could help with the social framework and is helpful to create the future decentralized society. Decentralized trusted timestamping is one of the possible usage scenarios in the future society, which can be used to prove data existence and integrity in a shorter time at a lower cost. This research provided a protocol to improve present blockchain and showed a possible blockchain application. In the future, with the significant technologies such as Internet of Things, Artificial Intelligence, and the blockchain, a more connected, secure and convenient society will be built.

Appendix A

Publication List

A. Peer-reviewed international journals

(J1) Y. Gao, and H. Nobuhara, 'A Decentralized Trusted Timestamping Based on Blockchains', IEEJ Journal of Industry Applications, Vol. 6, No. 4, pp. 252-257, (2017).

(J2) Y. Gao, S. Kawai, and H. Nobuhara, 'Scalable Blockchain Protocol Based on Proof of Stake and Sharding', Journal of Advanced Computational Intelligence and Intelligent Informatics, (Submitted).

B. Peer-reviewed international conferences

(C1) Y. Gao, and H. Nobuhara, 'Intellectual Property Protection using Decentralized Trusted Timestamping Based on the Blockchain,' The 31st International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC 2016), pp. 535-538, Okinawa, Japan, Jul. 10-13 (2016)

(C2) Y. Gao, and H. Nobuhara, 'A Proof of Stake Sharding Protocol for Scalable Blockchains', The Asia Pacific Advanced Network (APAN 2017), Vol 44, pp. 13-16, Dalian, China, Aug. 28 (2017)

(C3) Y. Gao, Shin Kawai, and Hajime Nobuhara, 'A study on High Scalable Blockchain', Asia Pacific Advanced Network (APAN 2019), (Accepted).

C. Domestic conferences

(D1) Y. Gao, Q. Zhang, and H. Nobuhara, 'A Study on Copyright Protection Support System for Origami', The19th Research Convention for Origami Science, Math and Education. Nov. 7 (2015).

(D2) Y. Gao, Q. Zhang, and H. Nobuhara, 'Decentralized Timestamping based on Blockchain and the Application to Origami Copyright Protection', The 78th National Convention of Information Processing Society of Japan (IPSJ), Mar. 10-12 (2016).

D. Patent Application

(P1) H. Nobuhara, Q. Zhang, and Y. Gao. Copyright protection support system. Patent Application No. 2015-89317.

E. Awards

(I1) Outstanding Master Thesis Award (March, 2016)

(I2) Monbukagakusho Honors Scholarship AY 2016, Japan (April, 2016 - March, 2017)

(I3) Japanese Government (MEXT) Scholarship AY2017, Super Global University Category (April, 2017 - March, 2018)

(I5) Rotary Yoneyama Memorial Doctoral Course Scholarship（YD） (April, 2018 - March, 2019)

Bibliography

[1] Blockchain Technology Will Usher in the Fourth Industrial Revolution. Retrieved from: https://bitcoinist.com/blockchain-technology-will-usher-in-the-fourth-industrial-revolution/ Accessed Oct. 26, 2018.

[2] World Economic Forum. Retrieved from: https://en.wikipedia.org/wiki/World_Economic_Forum. Accessed Oct. 20, 2018.

[3] Scalability. Bitcoin wiki. Retrieved from: https://en.bitcoin.it/wiki/Scalability/ Accessed Jul. 24, 2018.

[4] PayPal. Retrieved from: https://web.archive.org/web/20141226073503/https://www.paypal-media.com/about/ Accessed Jul. 24, 2018.

[5] VISA. Retrieved from: https://usa.visa.com/dam/VCOM/download/corporate/media/visa-fact-sheet-Jun2015.pdf/ Accessed Jul. 24, 2018.

[6] Why You Should Know the Technology behind Bitcoin. Retrieved from: https://marketrealist.com/2018/01/know-technology-behind-bitcoin/ Accessed Oct. 20, 2018.

[7] Blockchain's Future and the Applications. Retrieved from: http://www.soumu.go.jp/main_content/000550975.pdf/ Accessed Jul. 24, 2018.

[8] Keio University and University of Tokyo founded International Industry-University Cooperation group 'BASE ALLIANCE'. Retrieved from: https://www.nikkei.com/article/DGXLRSP451847_Q7A720C1000000/ Accessed Oct. 20, 2018.

[9] How Many People Use Bitcoin in 2018? Retrieved from: https://www.bitcoinmarketjournal.com/how-many-people-use-bitcoin/ Accessed Oct. 20, 2018.

[10] King, S., & Nadal, S. (2012). PPCoin: peer-to-peer crypto-currency with proof-of-stake. *URL https://peercoin.net/assets/paper/peercoin-paper.pdf. [Online]*.

[11] How to Deal With the Growing Blockchain Ledger Size in Containers. Retrieved from: https://portworx.com/deal-growing-blockchain-ledger-size-containers/ Accessed Oct. 20, 2018.

[12] Preneel, B. (2010, March). The first 30 years of cryptographic hash functions and the NIST SHA-3 competition. In *Cryptographers' track at the RSA conference* (pp. 1-14). Springer, Berlin, Heidelberg.

[13] Rivest, R. L., Shamir, A., & Adleman, L. (1978). A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, *21*(2), 120-126.

[14] Dwork, C., & Naor, M. (1992, August). Pricing via processing or combatting junk mail. In *Annual International Cryptology Conference* (pp. 139-147). Springer, Berlin, Heidelberg.

[15] Chaum, D., Fiat, A., & Naor, M. (1988, August). Untraceable electronic cash. In *Confer*

ence on the Theory and Application of Cryptography (pp. 319-327). Springer, New York, NY.

[16] Back, A. (2002). Hashcash-a denial of service counter-measure.

[17] Dai, W. (1998). B-Money-an anonymous, distributed electronic cash system.

[18] Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.

[19] Swan, M. (2015). "What is the Blockchain," in *Blockchain: Blueprint for a new economy*. "O'Reilly Media, Inc.", pp. x-xi.

[20] Swan, M. (2015). "Blockchain 2.0: Contracts," in *Blockchain: Blueprint for a new economy*. "O'Reilly Media, Inc.", pp. 9-10.

[21] Swan, M. (2015). "Blockchain 3.0: Justice Applications Beyond Currency, Economics, and Markets," in *Blockchain: Blueprint for a new economy*. "O'Reilly Media, Inc.", pp. 29-69.

[22] Blockchain Advantage and Disadvantages. Retrieved from: https://medium.com/nudjed/blockchain-advantage-and-disadvantages-e76dfde3bbc0 Accessed Oct. 24, 2018.

[23] Franco, P. (2015). "The Origins Of Bitcoin", *Understanding Bitcoin: Cryptography, engineering and economics*. John Wiley & Sons. pp. 161-169.

[24] Franco, P. (2015). "Foundations", *Understanding Bitcoin: Cryptography, engineering and economics*. John Wiley & Sons. pp. 3-10.

[25] Blockchain. Software Infrastructure. Retrieved from: https://www.slideshare.net/IBTSMG/blockchain-67620150 Accessed Oct. 24, 2018.

[26] Mechanism of Bitcoin. Retrieved from: https://moblock.jp/articles/17443 Accessed Oct. 24, 2018.

[27] CLARKE, S., CRAIG, I., & WYSZYNSKI, M. Litecoin Cash: The best of all worlds SHA256 Cryptocurrency.

[28] NOAH COIN whitepaper 2018. Retrieved from: https://noahcoin.org/wp-content/uploads/2018/02/Feb_12_New_UpdateLegal_Disclaimer.pdf Accessed Oct. 24, 2018.

[29] Wood, G. (2014). Ethereum: A secure decentralised generalised transaction ledger. *Ethereum project yellow paper*, *151*, 1-32.

[30] Schwartz, D., Youngs, N., & Britto, A. (2014). The Ripple protocol consensus algorithm. *Ripple Labs Inc White Paper*, *5*.

[31] Colored Coins. Retrieved from: https://en.bitcoin.it/wiki/Colored_Coins Accessed Oct. 24, 2018.

[32] Namecoin. Retrieved from: https://namecoin.org/ Accessed Oct. 24, 2018.

[33] Genecoin. Retrieved from: http://genecoin.me/ Accessed Oct. 24, 2018.

[34] Liquid Feedback. Retrieved from: https://liquidfeedback.org/ Accessed Oct. 24, 2018.

[35] Why invest in bitcoin. Retrieved from: https://www.slideshare.net/LuqmanulHakimbinAbd

R/why-invest-in-bitcoin/ Accessed Oct. 24, 2018.

[36] What is the Bitcoin Mempool? Retrieved from: https://99bitcoins.com/what-is-bitcoin-mempool/ Accessed Oct. 24, 2018.

[37] Franco, P. (2015). "The blockchain", *Understanding Bitcoin: Cryptography, engineering and economics*. John Wiley & Sons. pp. 143-158.

[38] Franco, P. (2015). "Mining", *Understanding Bitcoin: Cryptography, engineering and economics*. John Wiley & Sons. pp. 105-113.

[39] Antonopoulos, A. M. (2014). "Mining and Consensus", *Mastering Bitcoin: unlocking digital cryptocurrencies*. O'Reilly Media, Inc. pp. 175-215.

[40] blockchain.info. Retrieved from: https://www.blockchain.com/ja/charts/hash-rate?timespan=all Accessed Nov. 14, 2018.

[41] Sudhir, K. Explaining Hash Rate Or Hash Power In Cryptocurrencies. Retrieved from: https://coinsutra.com/hash-rate-or-hash-power/ Accessed Nov. 14, 2018.

[42] Franco, P. (2015). "Public Key Cryptography", *Understanding Bitcoin: Cryptography, engineering and economics*. John Wiley & Sons. pp. 51-75.

[43] Croman, K., Decker, C., Eyal, I., Gencer, A. E., Juels, A., Kosba, A.. & Song, D. (2016, February). On scaling decentralized blockchains. In *International Conference on Financial Cryptography and Data Security* (pp. 106-125). Springer, Berlin, Heidelberg.

[44] Block intervals. Bitcoin wiki. [Online]. Retrieved from: https://en.bitcoin.it/wiki/Block_intervals/ Accessed Jul. 24, 2018.

[45] Goswami, S. (2017). Scalability analysis of blockchains through blockchain simulation.

[46] Decoding the enigma of Bitcoin Mining. Retrieved from: https://medium.com/all-things-ledger/decoding-the-enigma-of-bit68 coin-mining-f8b2697bc4e2/ Accessed Jul. 24, 2018.

[47] Proof of Stake FAQ. [Online]. Retrieved from: https://github.com/ethereum/wiki/wiki/Proof-of-Stake-FAQ/ Accessed Jul. 24, 2018.

[48] A Proof of Stake Design Philosophy. [Online]. Retrieved from: https://medium.com/@VitalikButerin/a-proof-of-stake-design-philosophy-506585978d51/ Accessed Jul. 24, 2018.

[49] Op Ed: The Many Faces of Sharding for Blockchain Scalability. [Online]. Retrieved from: https://bitcoinmagazine.com/articles/op-ed-many-faces-sharding-blockchain-scalability/ Accessed Jul. 24, 2018.

[50] D. Larimer, Transactions as Proof-of-Stake. [Online]. Retrieved from: https://bravenewcoin.com/assets/Uploads/TransactionsAsProofOfStake10.pdf/ Accessed Aug. 10, 2018.

[51] Franco, P. (2015). "Odds and Ends", *Understanding Bitcoin: Cryptography, engineering and economics*. John Wiley & Sons. pp. 234-236.

[52] Eyal, I., Gencer, A. E., Sirer, E. G., & Van Renesse, R. (2016, March). Bitcoin-NG: A Scalable Blockchain Protocol. In *NSDI*(pp. 45-59).

[53] Luu, L., Narayanan, V., Zheng, C., Baweja, K., Gilbert, S., & Saxena, P. (2016, Octobe r). A secure sharding protocol for open blockchains. In *Proceedings of the 2016 ACM S IGSAC Conference on Computer and Communications Security* (pp. 17-30). ACM.

[54] Kokoris-Kogias, E., Jovanovic, P., Gasser, L., Gailly, N., Syta, E., & Ford, B. (2018, M ay). Omniledger: A secure, scale-out, decentralized ledger via sharding. In *2018 IEEE Sy mposium on Security and Privacy (SP)* (pp. 583-598). IEEE.

[55] Zamani, M., Movahedi, M., & Raykova, M. (2018, October). RapidChain: scaling block chain via full sharding. In *Proceedings of the 2018 ACM SIGSAC Conference on Compu ter and Communications Security* (pp. 931-948). ACM.

[56] Franco, P. (2015). "The Blockchain", *Understanding Bitcoin: Cryptography, engineering a nd economics*. John Wiley & Sons. pp. 113-115.

[57] Robertson, M. R. (2015, November 13). *500 Hours of Video Uploaded To YouTube Eve ry Minute [Forecast]*. Retrieved from: http://www.reelseo.com/hours-minute-uploaded-yout ube/#ixzz3th4DMvS0

[58] Michel, F. (2015). *How many public photos are uploaded to Flickr every day, month, y ear?* Retrieved from: https://www.flickr.com/photos/franckmichel/6855169886/ Accessed Fe b. 25, 2015.

[59] e-timestamp. *Protecting Your Intellectual Property*. Retrieved from: https://www.digistamp. com/about-us/protect-your-intellectual-property Accessed Feb. 25, 2015.

[60] Adams, C., Farrell, S., Kause, T., & Mononen, T. (2005). *Internet X. 509 public key inf rastructure certificate management protocol (CMP)* (No. RFC 4210).

[61] Antonopoulos, A. M. (2014). "Transactions", *Mastering Bitcoin: unlocking digital cryptoc urrencies*. O'Reilly Media, Inc. pp. 111-134.

[62] *BTProof*. Retrieved from: https://www.btproof.com/ Accessed Feb. 3, 2015.

[63] *Descriptions of SHA-256, SHA-384, and SHA-512*. Retrieved from http://csrc.nist.gov/gro ups/STM/cavp/documents/shs/sha256-384-512.pdf Accessed Mar. 18, 2015.

[64] WIPO. *What is Intellectual Property?* Retrieved from: http://www.wipo.int/about-ip/en/ A ccessed Feb. 25, 2015.

[65] Gipp, B., Meuschke, N., Gernandt. A. (2015). *Decentralized Trusted Timestamping using the Crypto Currency Bitcoin (preprint). In Proceedings of the iConference 2015.*

[66] Haver, S., Stornetta, W. *How to Time-Stamp a Digital Document.* Retrieved from: https:/ /www.anf.es/pdf/Haber_Stornetta.pdf Accessed Mar. 18, 2015.

[67] *Proof of Existence.* Retrieved from: https://www.proofofexistence.com/ Accessed Feb. 3, 2015.