Thesis for Master's Degree

MyData Blockchain

Giljun Jung

School of Electrical Engineering and Computer Science

Gwangju Institute of Science and Technology

2019

MyData Blockchain

마이데이터 블록체인

MyData Blockchain

Advisor: Heung-No Lee

by

Giljun Jung

School of Electrical Engineering and Computer Science
Gwangju Institute of Science and Technology

A thesis submitted to the faculty of the Gwangju Institute of Science and Technology in partial fulfillment of the requirements for the degree of Master of Science in the School of Electrical Engineering and Computer Science

Gwangju, Republic of Korea
2019. 06. 10.
Approved by

Professor Heung-No Lee
Committee Chair

MyData Blockchain

Giljun Jung

Accepted in partial fulfillment of the requirements for
the degree of Master of Science

June 10th 2019

Committee Chair ____________
Prof. Heung-No Lee

Committee Member ____________
Prof. JongWon Kim

Committee Member ____________
Prof. Gunoo Kim

MS/EC 20171186 Giljun Jung (정길준). MyData Blockchain (마이데이터 블록체인). School of Electrical Engineering and Computer Science. 2019. 31p. Prof. Heung-No Lee

Abstract

The amount of personal data has grown at a rapid pace recently. Global Internet companies such as Facebook and Google used users' accumulated data to track users' activities and to predict users' spending patterns and preferences. Unlike their intention, privacy concerns have been raised by using users' data for activities which are agreed and not agreed in advance. Furthermore, it has become a big issue that the revenue from utilizing users' data is not given to the data owners at all. As a result, Europe Union defined General Data Protection Regulation to protect that such privacy issues do not arise in data utilization. Complying with these privacy laws, blockchain models on personal data such as MeDShare, MedBlock, and Nebula Genomics have recently emerged.

In this paper, we aim to propose a blockchain model for MyData application. This model allows the creation of a shared data economy among data owners, data providers, and data consumers. The proposed model allows data owners to control the use of their data and generate revenue from their own data while maintaining anonymity. It also enables data consumers to use this model for various applications because they can obtain quality data legitimately. First, anonymized identity data is assigned to each data instead of using de-identified identity data for each person. Second, data relayers support data discovery effectively by querying. Furthermore, forming two types of tokens such as security and utility makes the ecosystem stable continuously. Thus, the proposed model can have three advantages such as anonymity, the effectiveness of search, and stability in the ecosystem compared to the three aforementioned models.

List of Contents

Abstract i

List of Contents ii

List of Tables iii

List of Figures ii

I. INTRODUCTION 1

II. RELATED WORK 3

III. BACKGROUND 7

3. 1. Blockchain 7

3. 2. Cryptographic Primitives 8

3. 2. MyData 10

3. 3. EU GDPR 11

3. 4. Data De-identification 11

IV. PROPOSED MODEL 14

4. 1. Overview of Proposed Model 14

4. 2. Personal Data and Privacy Protection 17

4. 3. Data Storage 20

4. 4. Data Transaction 21

4. 5. Token System 22

V. DISCUSSION 25

VI. CONCLUSION 27

References 28

Acknowledgement 31

List of Tables

Table 1. Properties of the blockchain technology……………………….......………………………….……… 9

Table 2. The contents, definition, and applications of smart contract……..………………….………….. 17

Table 3. The attributes of ANID………………………………..………………………….……………… 19

Table 4. The summary of comparisons between the models……………………………….…………….. 26

List of Figures

Figure 1. Overview of MeDShare………………………………………………………………………… 4

Figure 2. Overview of MedBlock…………………………………………………………………………. 5

Figure 3. Overview of Nebula Genomics………………………………………………………………… 6

Figure 4. Blockchain has a decentralized and distributed network……………………………….……… 7

Figure 5. Block is transferred by hashing and chained in blockchain………………………….…………… 9

Figure 6. MyData in relationship between requesting party, personal data store, and data handback……… 10

Figure 7. The main rights of the GDPR……………………………………………………….………… 11

Figure 8. The procedures of data de-identification………………………………………..………… 13

Figure 9. The structure of the proposed MyData blockchain……………………………………………… 15

Figure 10. Overview of the proposed model... 16

Figure 11. The relationship between PIID, DIID, and ANIDs.. 18

Figure 12. Storing data separately with on-chain and off-chain……………………………………………… 20

Figure 13. The diagram of data transaction…………………………………………………...…………… 22

Figure 14. Token separation to make ecosystem stable……………………………….…………………… 24

I. INTRODUCTION

The amount of data will increase by about 41.4% per year from 2012 to reach 40 ZB by 2020 [1]. In last year, the head of Dell Technology announced that by 2020, an average of 200 PB of data will be produced in cities, and 99% of the data will be produced by machines, not people [2]. Numerous data are coming from Internet of things (IoT) and smart devices through human activities [3]. These smart devices, IoT, and technology such as artificial intelligence (AI), blockchain are the products of the 4th industrial revolution (4IR) [4]. 4IR has aimed to make intelligent world through connectivity, decentralization, sharing, and opening. In order to build this intelligent world, various technologies such as big data, AI, and blockchain have been utilized.

Personal data produced are actively used for big data analysis and pattern recognition [5]. The use of Deep Learning (DL) is extending steadily for many companies to solve optimization problem in various areas. In addition, DL requires a lot of data for learning, and this demand has made the field of big data more interesting [6]. However, privacy problems have occurred in the use of personal data, including the recent leakage of personal information from Facebook [7]. When using personal data, every entity should always be careful about privacy issues. In order to prevent privacy issues of personal data utilization, data de-identification was proposed, but the criteria for the assessment of appropriateness are ambiguous, and it is limited that the re-identification can be made possible through combination with other data in the future [8]. Last year, the Europe Union (EU) announced a general data protection regulation (GDPR) that grants data owners many rights, including data movement and deletion. By providing these multiple rights, the GDPR allows the data owner to have his own data sovereignty [9].

Many Internet technology (IT) companies have made a lot of profit by utilizing accumulated personal data [10]. However, their data utilization strategy is pointed out by users as unfair, asymmetric, and unequal, and the revenue from utilizing personal data never returns to the data owner [11]. Thus, for the first time in Finland, the concept of MyData was presented to give back control of personal data to the data owner [12]. Blockchain allows transparent data movement as all transactions are recorded and cannot be modified [13]. In addition, the goal and purpose of returning the sovereignty of personal data from a centralized entity in the direction of decentralization of data is similar to the GDPR. As a result, some personal data utilization models through blockchain [14], [22-24] are appeared.

In this paper, we aim to propose a blockchain model for MyData application. This model allows the creation of a shared data economy among data owners, data providers, and data consumers. The proposed model allows data owners to control the use of their data and generate revenue from their own data while maintaining anonymity by assigning separate anonymous ID (ANID) to each personal data. It also enables data consumers to use this model for various applications because they can obtain quality data legitimately. The proposed model facilitates data retrieval by data relayers and allows the ecosystem to be more stable by using two types of tokens simultaneously. The remainder of this paper is organized as follows: Section II reviews the related work. Section III introduces the background. Section IV proposes a blockchain model for MyData application. Section V and VI is discussion and conclusion respectively.

II. RELATED WORK

Blockchain technology is still evolving and many applications related to blockchain are appearing extensively. However, blockchain application for personal data management is not well-explored yet. One of the first contributions in this direction is [15], where a protocol was developed to convert blockchain into an automatic access control manager for a decentralized personal data management system. The use of auditable contracts and privacy architecture built on blockchain infrastructures for transparent data access, data owners' personal data sharing and processing was proposed in [16]. Similarly, a framework for integrating online identity and reputation information based on social-dependent network to provide online behavioral ratings was proposed in [17].

Several studies in the healthcare domain explored blockchain technology for the medical data access. An influential and highly relevant contribution is [18], and the authors proposed an architecture based on AI and blockchain technology to allow users to control personal data including medical records. In the similar lines a decentralized record management system is proposed in [19] to handle patient's electronic records using blockchain technology. The research work in [20] proposed a blockchain-based mobile application architecture, enabling patients to easily and safely own, control and share their own data without infringing privacy. When compared to conceptual design and system architecture for human-centric personal data and identity management based on the MyData initiative, that is in compliance with the GDPR using blockchain and smart contracts technology [21].

MeDShare [22] was introduced by Q. Xia *et al.* in 2017. The purpose of MeDShare is making cloud service providers of personal data such as electronic medical records (EMRs) via blockchain in order to share the data securely in a trust-less environment. MeDShare has 6 entities such as user, authenticator, processing and consensus nodes, existing database infrastructure, smart contract center, and blockchain network as shown in Figure 1. When a user requests data access by query, authenticator checks the legitimacy of the request by verifying the signature. The form generated includes a hash of the timestamp and a hash of the requester's ID at the time the request was received. The processing and consensus nodes forward the generated form to the existing database infrastructure. The existing database infrastructure sends the retrieved data to the processing and consensus nodes. They send a request to the smart contract center to append a rule set to the requested data. A generated smart contract is tagged to the form. Results of the processed data are sent to the authenticator and the processing and consensus nodes process a block based on the information of the request and broadcast the block into the blockchain network. The results are finally distributed to each requestor by querying.

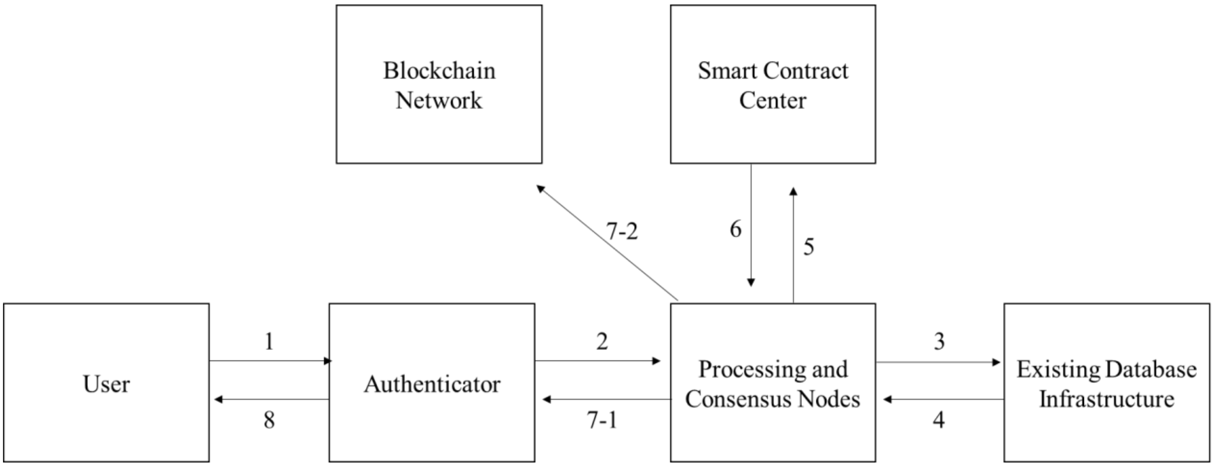


Figure 1. Overview of MeDShare.

MedBlock [23] was introduced by K. Fan *et al.* in 2018. The purpose of MedBlock is sharing personal data such as EMRs via blockchain while maintaining security. MedBlock has 6 modules such as client, endorser, orderer, committer, database, and ledger as shown in Figure 2 and it shows the complete process of MedBlock. First, the client encrypts the users' EMRs with the patients' public key. When the client signs that the information is accurate using the patient's private key and the department's private key, the client sends hash value of the data to the endorser. Second, endorser verifies that the signature is completed. When finished, endorser saves the data to the local database and sends the receipt to the client. After this, the client waits for the receipt of orderer continuously. Third, the endorser sorts all the uploads and packs into blocks according to the upload time. When it's time for the endorser to be the first turn, the endorser will send an offer to add blocks to the orderers. Fourth, the orderers who are in charge of consensus reach a consensus and send the consensus to committers. Fifth, after collecting sufficient receipts, endorser notices the information of uploading successfully to the client. Sixth, committers add the block to the ledger according to the result of the consensus. If the client has not received a confirmation receipt for a long time, the client could select another endorser to restart the request. Once all the blocks to be uploaded have been verified, endorser will broadcast the information throughout the network, noticing the turn of the following endorser. The features of MedBlock are it uses combined access control protocol which can achieve the effect of zero knowledge proof and ring signature, and it gives up real-time data upload for the system efficiency.

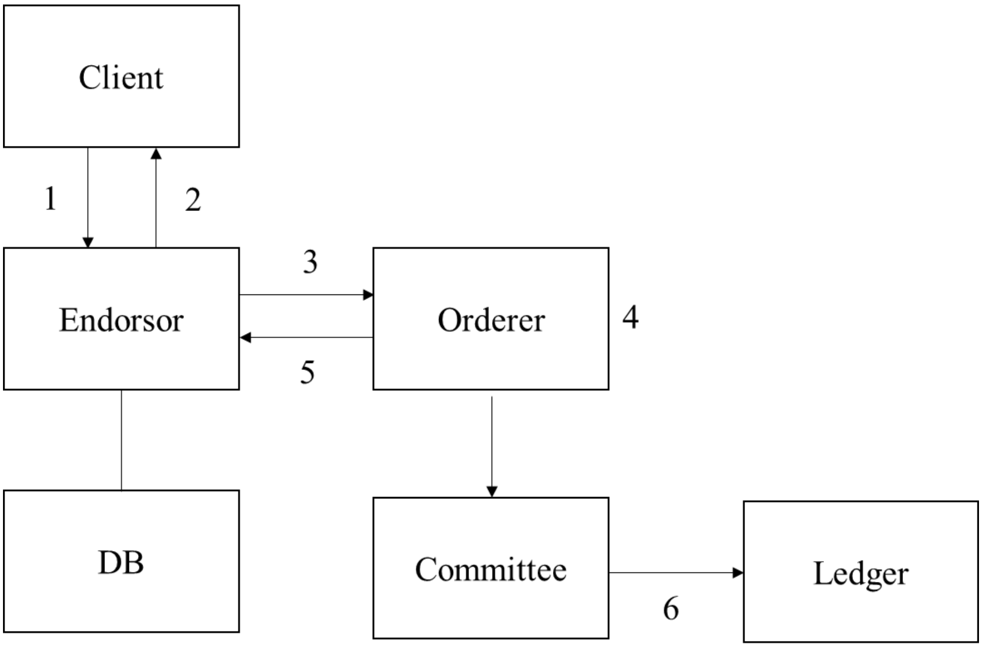


Figure 2. Overview of MedBlock.

Nebula Genomics [24] was introduced by D. Grishin *et al.* in 2018. The purpose of Nebula Genomics is sharing personal data such as genomic data and making analysis platform using Blockstack [25] platform for decentralized applications. The genomic data is made from human genome sequencing. There is also phenotypic data to be used in survey which is conducted by buyers. The network of Nebula Genomics consists of data owner nodes, data buyer nodes, secure compute nodes, and Nebula server as shown in Figure 3. Data owner nodes, which may be individuals or organizations that own genomic databanks, store their data on the private storage such as inter planetary file system (IPFS) [26] or BitTorrent [27]. When data owner nodes want to know their genomic information individually, they can get their genomic data by offering their sequencing order and token from Nebula server. When data buyer nodes pay tokens to buy the data, they can receive the encrypted data after data buyer verification by Nebular server. In addition, if data buyer nodes want to discover the data or do a survey, data owner nodes can receive tokens from data buyer nodes by responding. The secure compute nodes can be operated by data buyers, Nebula server, or any third party. The personal data which was in private storage are sent to the secure compute nodes in order to enable secure computations using Arvados [28] in Intel Secure Guard Extension (SGX) [29]. Arvados is the bioinformatics platform to compute on genomic data, and Intel SGX is a set of instruction codes and allows creation of private memory regions which is called enclaves. The secure compute nodes negotiate encryption key with data owner nodes.

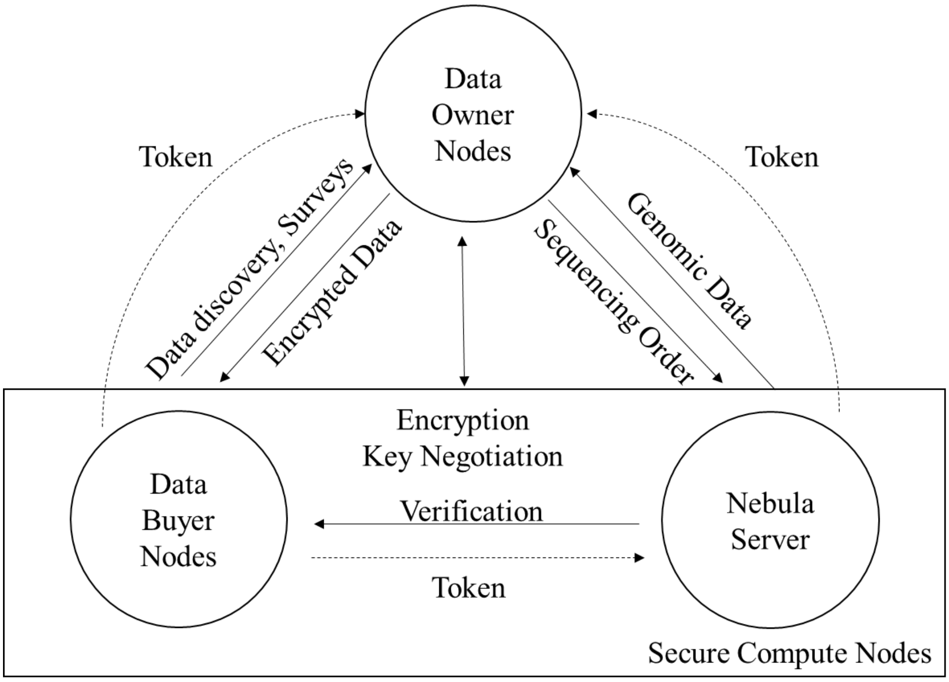


Figure 3. Overview of Nebula Genomics.

The various reviewed literatures in this section such as MeDShare, MedBlock, and Nebula Genomics which are shown the latest trends on dealing with personal data via blockchain have some limitations. In this paper, after some explanations about preliminary knowledge, we aim to propose a secure, effective, and stable blockchain-based model to utilize personal data, that overcomes these limitations. The main advantages of our model are to provide anonymity during transactions, effectiveness during discovering the data, and stability on blockchain ecosystem.

III. BACKGROUND

3. 1. Blockchain

Blockchain is a decentralized distributed database technology that combines encryption with the guaranteed assurance of transaction tampering. By using timestamping of transactions and messages, the blockchain provides universally verifiable evidence of the presence or absence of transactions in a distributed database, and the underlying cryptographic primitives using hash functions and digital signatures are computed in all cases It is guaranteed to be safe and verifiable. Blockchains are distributed so that a large number of independent nodes manage jointly and maintain the consistency of transactions between distributed nodes using decentralized consensus protocols such as Byzantine fault tolerance (BFT) algorithms [30] without requiring a central authority. Blockchain transactions are transparent to all users of the system, while blockchains provide anonymity to users by enabling them to conduct pseudonym anonymity transactions without having to disclose personal information. Due to the destructive and innovative nature of blockchain technology, many distributed applications such as Crypto calls and smart contracts have evolved. Bitcoin was introduced in 2009 with decentralized cryptocurrency based on blockchain technology [13], and Bitcoin is now the largest cryptocurrency with a market capitalization of more than 200 billion USD [31]. In a nutshell, blockchain technology is built around three fundamental concepts: distributed databases, trust protocols, and encryption. We will briefly explain them in the following subsections.

Distributed Database: Blockchain technology, built on the concept of peer-to-peer (P2P) networks and distributed storage [32], can be considered a distributed data repository through primary system replication using P2P protocols, where transactions are atomic changes to block-grouped datastores [18].

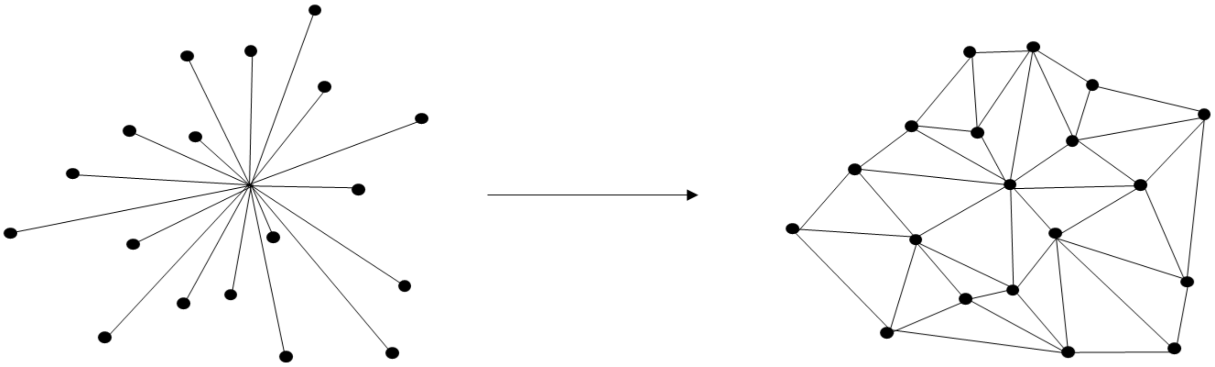


Figure 4. Blockchain has a decentralized and distributed network.

The Trust Protocol: In order not to have a central authority to enable trust in the system, there must be some mechanism between the parties concerned to establish trust, which can be achieved by the distributed consensus of the parties concerned. Blockchain trust is guaranteed through a distributed consensus protocol. The protocols can vary slightly from system to system, but the idea of gaining trust with the consensus that involves the parties remains unchanged. The two most prevalent concepts of this protocol are proof-of-work (PoW) and proof-of-stake (PoS) following a BFT scheme [30].

PoW is the concept that a service requester must solve a cryptographic puzzle in order to participate in a network, initially proposed in hashcash [33] as a countermeasure against a denial of service (DoS) attack using CPU cost functions. Blockchain, particularly in Bitcoin [13], is used as a verification technique to find and append headers suitable for new data blocks to the blockchain. To add a block, the node must find the correct nonce, which becomes a predefined hash format with certain restrictions. At the same time, to avoid 'dishonest' attempts to change the ledger, users can only add blocks to the longest chain which has invested the most work proof.

PoS is another way to verify and add blocks to the blockchain, the next block is selected by the node [34]. Therefore, the node adds and verifies the block depending on how much stake it holds in the system. Thus, ownership will lead actors to behave honestly, and if they behave dishonestly, they will lose their stakes.

3. 2. Cryptographic Primitives

Hash Functions: Hashing is used to ensure the integrity of the data and hash function is an input independent mean linear algorithm that takes a set of variables or data and converts it into a fixed-size hash digest [35]. A successful hash function has the following characteristics such as deterministic, efficient, distributed, preimage-resistance, and collision resistance [14]. In addition, the key concept of a hash function in the blockchain is to organize and link data together. This is done by hashing of variable elements in the block header such as a hash of the previous block, Merkle root of transactions, timestamp, and nonce. The concept of Merkle Tree [36] is to hash the results of each transaction after each transaction to build the tree structure until the top-level node known as Merkle root is acquired. This type of data theorem provides a safe and efficient way to view the contents of the block and summarizes all transactions in the block [37].

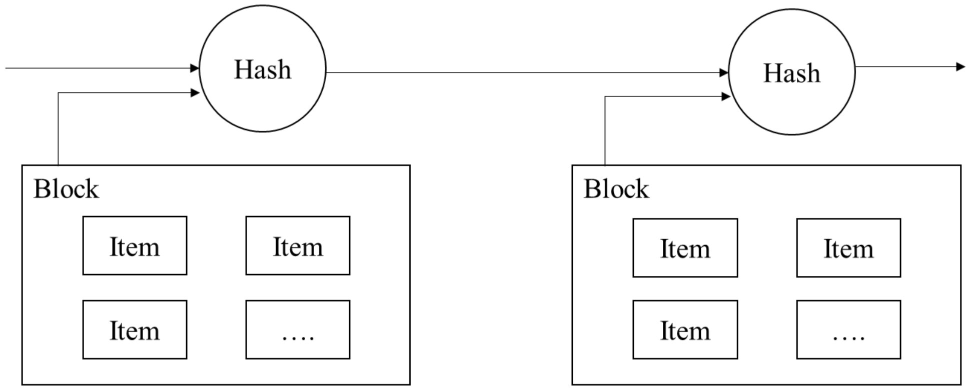


Figure 5. Block is transferred by hashing and chained in blockchain.

Digital Signatures: One of the main objectives of blockchain technology is to be able to verify the authenticity and non-repudiation of data. Digital signatures are a cryptographic system that guarantees two attributes such as authenticity and non-repudiation. Authenticity means the quality of being real or true, and non-repudiation uses key pairs with asymmetric encryption algorithms, such as RSA [38], to ensure that data is not changed. Over the years, a safer version of the digital signature has been developed. Bitcoin, for example, uses an elliptical digital signature algorithm (ECDSA) to generate keys [39].

A summary of the above-mentioned technology leads to the following features as described in Table 1.

Immutable	Data written to database cannot be changed or deleted without consensus leading to data integrity
Decentralized	No single point of failure/control achieved by a decentralized architecture and distributed database
Transparent	All data sent through the blockchain is visible to all network participants
Pseudonymous	The identity of data senders and receivers is unknown
Chronological	Every transaction is time-stamped and can be traced back

Table 1. Properties of the blockchain technology. [14]

The use of blockchain as the tamper-resistant ledger will record the transfer and undoubtedly prove ownership of the asset. This enables smart contracts, which were already conceptualized 22 years ago [40]: the production of computer programs that can safely implement previously concluded contracts. Consequently, the concept of a smart contract is to take a contract clause and convert it into a code that can be implemented on its own. Thus, there is no need for an intermediary responsible for the execution of the contract and instead, a trusted computer program will depend on it. Complex contracts and payment contracts can be included in standardized contracts and then monitored and executed at low transaction costs because they are digitally and immutably managed [41].

3. 2. MyData

MyData[42] is a study commissioned by the Finnish government to manage personal data. This Nordic self-identity model is driven by the concepts of human-centered control, usability, accessibility, and openness. MyData can be used to protect data flows between sectors such as government, health care, and finance. The core of MyData authentication is user-managed access, opening-up ID Single Sign-On and Oauth 2.0. Blockchain manipulation attempts to control access to web APIs can be easily detected, so blockchain is used for distributed control of fraud through the network of entire stakeholders. The study, which joined forces with Sovrin, aims to strengthen digital human rights while also opening up new opportunities to develop innovative personal data services. It also aims to address the EU GDPR [9], a new rule on the control and processing of personal information that went into effect in May 2018. The global trends of MyData in major countries are Smart Disclosure in America, Midata in England, MesInfos in France, and MyData in Finland. The names are different, but the common purpose is to activate the data economy.

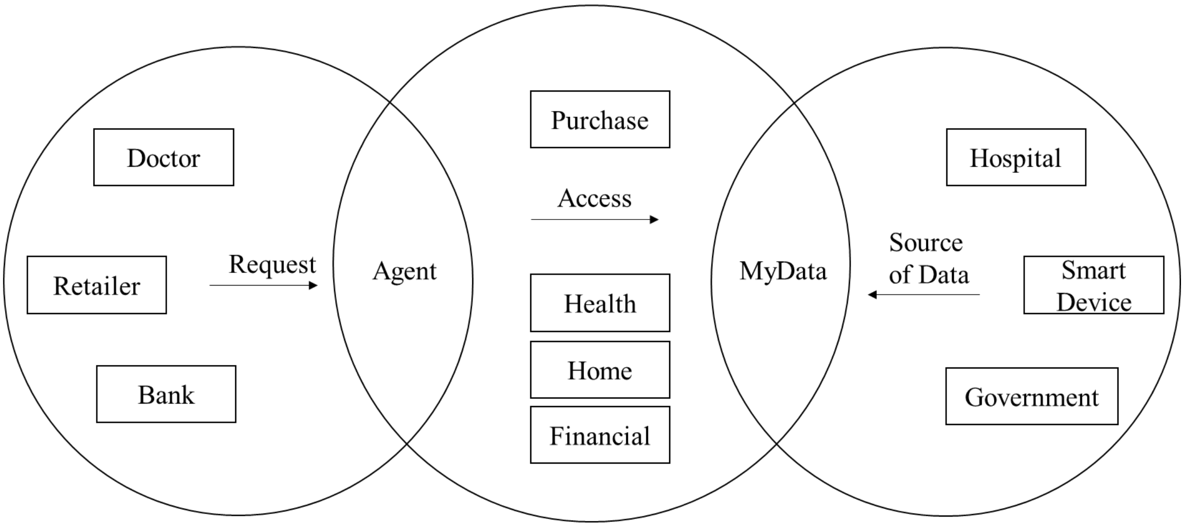
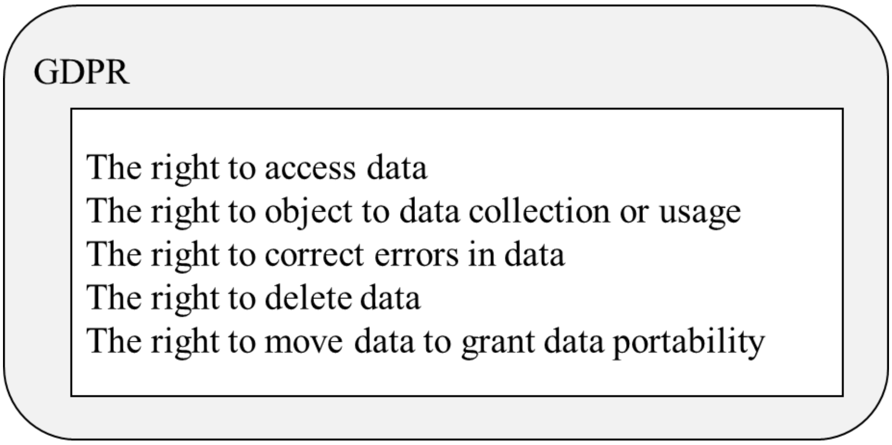


Figure 6. MyData in relationship between requesting party, personal data store, and data handback. [43]

3. 3. EU GDPR

The GDPR [9] is one of the biggest changes in data privacy regulations in recent years and took effect in May 2018 on behalf of data protection directive since 1995. The key goal is to harmonize data privacy laws across Europe, especially strengthening and protecting the privacy rights of EU citizens. One of the most central issues is the question of the user's consent. This regulation states that a service provider must show what the user's consent is and that it is easy for the user to withdraw their consent. In the event of a change in the use of data other than the user's withdrawal of consent or consent, the service provider is required to delete data related to the particular user. Furthermore, as a user's right to access, the service provider must provide an overview of the user's personal information and the purpose of processing at the user's request. The service provider should also provide users with all data in machine-readable formats. Like access rights, data movement rights are also present. Users should be able to extract their personal data from the controller in a readable format and transfer their data to other controllers. Violations of the GDPR may result in a large fine of up to 20 million euros or 4 percent of global sales.



The right to access data
The right to object to data collection or usage
The right to correct errors in data
The right to delete data
The right to move data to grant data portability

Figure 7. The main rights of the GDPR.

3. 4. Data De-identification

Data de-identification measures are carried out in four stages according to the Personal Information de-identification guideline recommended by the Korea Internet Development Agency (KISA) as follows [44]. A preliminary review step to verify that the data to be used contains personal information, a de-identified action step to process the elements that enable personal identification, an appropriateness assessment step to assess whether the de-identification action has taken place properly, and a post-management step to prevent re-identification.

In the stage of the preliminary review, which is the first step, it is possible to check whether each data contains personal information by referring to the personal information protection law, and if it does not include personal information, it can be used immediately for big data analysis. However, if the data contains personal information, you should proceed to the next step.

In the second phase of the de-identification phase, the identifier itself and, since the attribute person may identify a particular individual when combined with other data, action on the identifier and the attribute must precede it. Identifiers included in the data should be deleted by referring to the privacy rules, and the attribute should be deleted as of a general rule. However, identifiers or attributes that must be used after strict de-identification measures. De-identification measures include pseudonymization, aggregation, data reduction, data categorization, data masking, and so on. Each technique can be implemented with various detailed techniques and appropriately selected according to the purpose of using the data.

In the third stage of the adequacy assessment, an evaluation team including external experts is formed to strictly evaluate whether the de-identified data has the possibility of being identified. Use the k-anonymity of the privacy protection model when evaluating the appropriateness and use additional evaluation models if necessary. If improper evaluation results are obtained, the personal information processor should perform additional de-identification measures reflecting the opinions of the evaluation team. The collected information aggregate can be used for big data analysis and so on.

In the post-management stage, which is the fourth step, monitoring and prevention of cases in which data can be identified during the data utilization process even if the data has been evaluated. First, administrative and technical safeguards are needed to ensure that de-identified data is not leaked. Second, non-identified data should be monitored periodically as internal or external factors may change and be re-identifiable. Third, even if the non-identified data is provided to a third party or a contract is made, it is possible to identify again. Finally, re-identified data can be used immediately to stop, destroy or re-use de-identified measures.

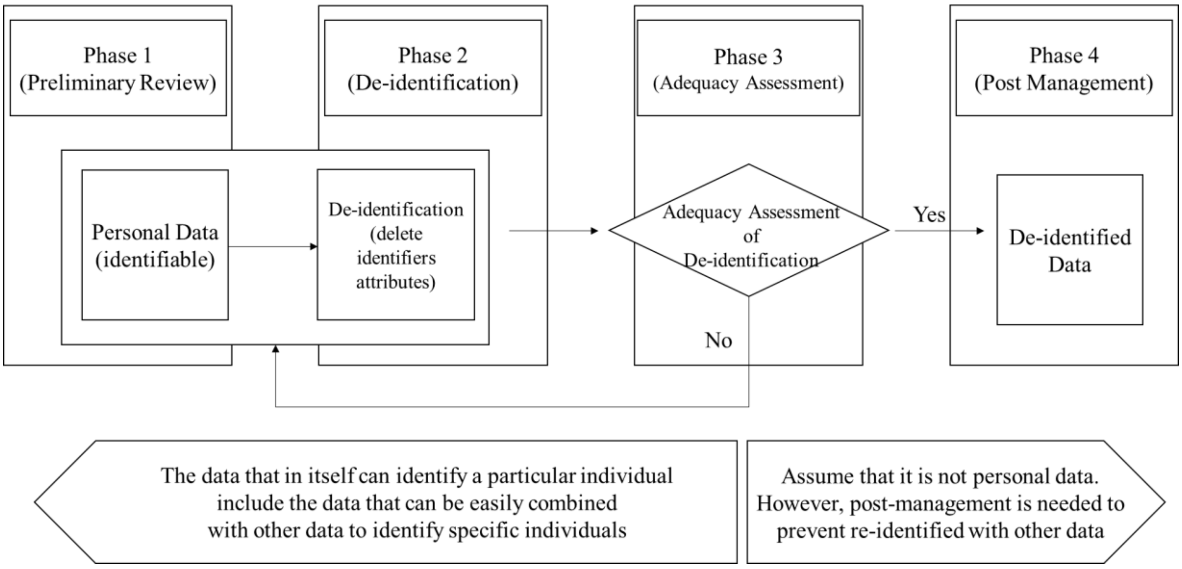

Phase 1
(Preliminary Review)
Phase 2
(De-identification)
Phase 3
(Adequacy Assessment)
Phase 4
(Post Management)
Personal Data
(identifiable)
De-identification
(delete
identifiers
attributes)
Adequacy Assessment
of
De-identification
Yes
No
De-identified
Data
The data that in itself can identify a particular individual
include the data that can be easily combined
with other data to identify specific individuals
Assume that it is not personal data.
However, post-management is needed to
prevent re-identified with other data

Figure 8. The procedures of data de-identification. [44]

IV. PROPOSED MODEL

It is argued that stronger regulation is needed in relation to consumer's data privacy and personal data ownership. This is because of personal data infringement, malicious use of personal data, and collecting data without the explicit consents of the data owner. As the EU GDPR case, global changes take place to protect the privacy and data sovereignty of personal data [9]. In terms of goals and objectives, blockchain and the GDPR share some points. They aimed at decentralization of data control. In doing so, they could return the data sovereignty of the individual from the centralized entity. In addition, blockchain can help data owners more effectively based on agency, data-based applications, the GDPR, and data-related regulations.

There are problems in today's data industry [45]. First, there are data privacy infringements for users. Today's data is collected, utilized, and monetized without the consent or recognition of the owner. Today's data industry is dominated by centralized services that illegally aggregate data and sell it to other companies in order to make a huge profit. Second, collecting and monetizing data legitimately is difficult for applications. Because there is no existing legitimate data market to monetize user data for additional revenue, applications often illegally sell personal data to other services that aggregate more data to sell it to other companies. Third, there is a lack of transparent and legitimate data in quantity and quality for companies. There is no existing data market where companies can purchase high quality, insightful, and legally the data they need at reasonable prices for targeted marketing purposes and business intelligence.

Therefore, we aim to propose a new blockchain model for MyData application. The proposed MyData blockchain model is made by putting together several other materials. For example, the parts of data collection and privacy protection follow the method of [45], the part of data storage follows the method of [46], and the part of token system follows the method of [47].

4. 1. Overview of Proposed Model

First of all, it is needed that transparent data flow between data owners, data providers, and data purchasers. Blockchain technology and smart contact technology are used to build an automated system that enables transparent transactions. Data owners decide the conditions with data providers through a smart contact for their data from using apps. The smart contract includes the consent for storing, selling, and accessing to personal data. All consents agreed by the data owner are stored in the blockchain access layer and can only be modified by the data owner himself. And instead of providing data generated by using the app, the data owner gets the revenue back. The data provider separate ID and payload in the personal data, the ID is stored in the blockchain, and the payload is stored in the off-chain. The contents of personal data are further described in Section 4.2. The contents of data storing each data and the introduction of off-chain are described in Section 4.3. Data purchaser such as data consumer and data processor in Figure 10 presents information about the desired data and the amount to be paid through a smart contact and request access to the data. The data provider grants the data purchaser access to the data within the terms of the smart contract that it has agreed with the data owner. The data purchaser can obtain the desired data with pointer in accordance with the conditions previously agreed by the data owner stored in the blockchain access layer. The data purchaser pays the data provider in exchange for obtaining the data, which is later distributed and delivered to the data owner.

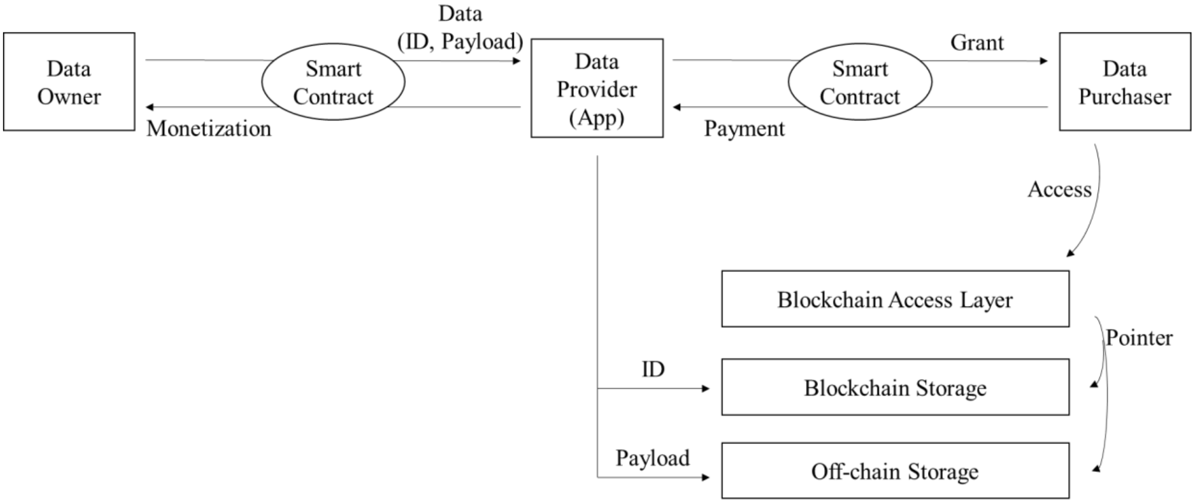


Figure 9. The structure of the proposed MyData blockchain.

In order to maintain this model, there exist 5 functions as follows: data collection, data storing, data discovery, data exchange, and data purchasing as shown in Figure 10. In data collection, data providers collect data on behalf of users through a data collection authorization process which is called DAuth [48]. DAuth helps applications find explicit permission to data owners to collect and monetize the data. In data storing, data is encrypted and stored by the data providers. In data discovery, data consumers can search for desired data, and collect requested data. In data exchange, the data provider accepts the smart contract presented by the data consumer, and the data consumer will be able to access the desired data. In data purchasing, when the exchange is complete, the data consumer pays the data via tokens. Each paid token will be distributed to the data owners and data providers respectively.

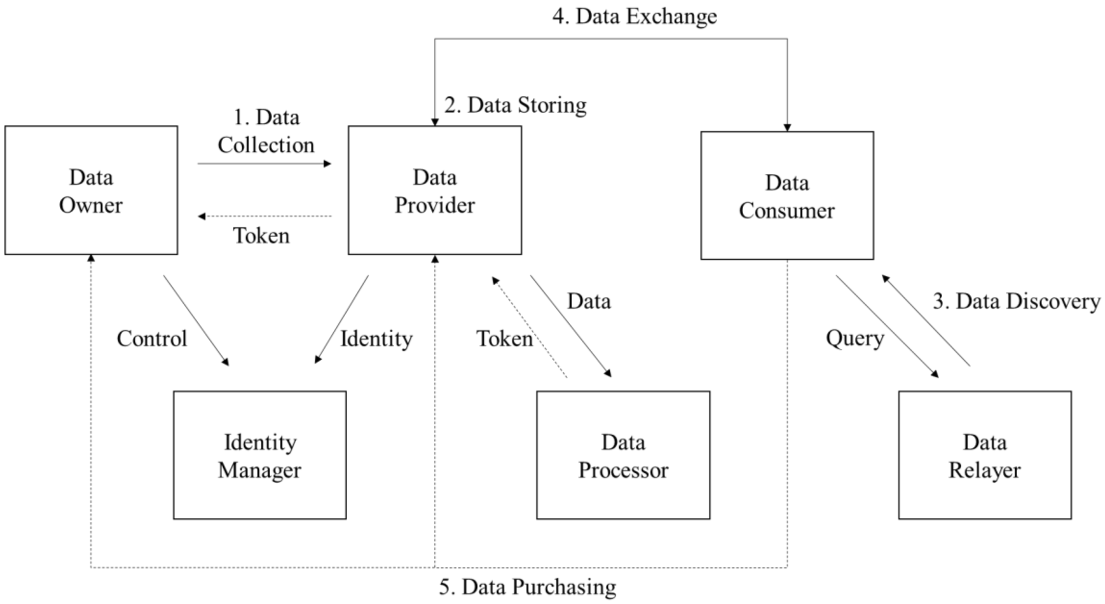


Figure 10. Overview of the proposed model.

For performing the above-mentioned functions efficiently, other entities such as data relayer, data processor, and identity manager are needed. The data relayer aggregates data from multiple data providers and adds filters so that data consumers can find and purchase specific data subsets they want. Data processor is a data analysis service or professional data analysis company that processes raw data into variable data. Anonymized raw data can be provided by the data provider for further processing. Identity manager is authorized by the data owner and instead stores and manages IDs. Identity manager can be compensated by storing the data owners' ID and notifying when the data is sold to the data consumers. There is another entity who is data validator. The role of data validators is verifying the data when it is stored. If the data is verified by data validator, the validator record his signature on it in order to get reward after the data is sold. Data validators which are selected by key stakeholders give more trust to who want to purchase the data. Data purchasers can evaluate data validators because only they know the data is matched to desired data.

In Figure 10, the dotted arrows show the movement of tokens. When a data owner generates the data by using apps, the data provider compensates the data owner for obtaining the data. Data consumer pays tokens in exchange for buying the desired data. The paid tokens are distributed to the data owner and the data provider. Data consumers can purchase data directly from a market or send a query to the data relay in order to buy only specific datasets. Data processor requests the data from data provider in order to analyze the data or to make reports showing consumer goods, the performance of applications, and so on. In addition, data relayer, data processor, and identity manager can receive tokens from the network according to their contributions.

All activities are automatically recorded in the blockchain by a smart contract. It enables transparent and secure transactions, and data owners can track how their data moves. The short definition of a smart contract and what should be included in the smart contract are briefly summarized in Table 2.

Contents	Definition	Application
Smart Contract	Pre-set data access/sharing conditions and automatically establish contracts when conditions are met	Establish Terms and Conditions for identification, data access, collection, storage, and sharing
Identification	Identifying who is the main subject of the data	Save a user's profile to the blockchain
Data Access Privilege	Ensuring that data is accessible when the set conditions are met	Establish conditions such as whether or not a third party has access to data, access conditions, access data types, and ranges
Data Collection/Storing	Collecting and storing personal data in the DB	Encrypt personal metadata and payload data to store in blockchain and to DB in the system
Data Sharing	Sharing data with a third party that an individual wants	Set conditions for third parties to share, purpose of sharing, type of shared data, range, and duration
Trace Transaction	Checking the results like data utilization	Data transaction details are stored in blockchain

Table 2. The contents, definition, and applications of smart contract.

4. 2. Personal Data and Privacy Protection

Personal data is information, when aggregated enough, can infer a person’s characteristics or identity. Today, large amounts of personal data are automatically generated through digital devices such as computers and smartphone, and IoT devices such as smartwatches. The activities of every individual connected to a digital device leave personal data behind. While many of these data are generated through digital devices, information that represents the direct personal interests of an individual can also be considered personal data.

All kinds of personal data are parsed and can be classified into two types: identity data and payload data. Since the capacity of data that can be contained in a block of blockchain is not large, only the hash value of personal metadata will be stored in the on-chain, and the payload will be stored separately as off-chain. Identity data can usually be defined by all kinds of identifiers. Identifiers must be associated with or be identifiable from the owner of the generated data. It is also desirable that an identifier is linked directly to one-on-one relationships with one unique individual. There are three types of identity data: This classification should be followed to protect data owners' privacy by applying clear criteria when utilizing personal data.

Personal Identifiable ID (PIID): An PIID contains all types of identifiers that are directly related to a specific person's identity. Identifiers such as resident registration numbers, email addresses, and contacts can be considered PIIDs. The reason why e-mail or contact is considered a direct relationship is that PIID allows data consumers to contact a specific individual.

De-identified ID (DIID): An DIID is a certain kind of identifier that is not directly linked to a specific individual's identity, but is indirectly linked to an individual's identity that is available for targeted advertisement exposure. DIID is not direct because data consumers cannot communicate with that individual directly through DIID. However, with DIID, data consumers can increase advertisement exposure for digital advertising platforms, such as Google Ads and Facebook, to a particular group of people. In addition, some PIIDs can be converted to DIIDs via data de-identification. If PIID is hashed by a specific algorithm such as sha-256, it becomes DIID which is irreversible back to PIID, but can still be used to point to one unique individual.

Anonymous ID (ANID): An ANID is a unique ID given to each data of data owners. This can be used to distinguish unique data owners in the middle of data processing. Because there is no way to communicate with the ANID, it is not possible to guess to the data owner only with the ANID. ANID means one unique individual, but multiple ANIDs can also be generated per single individual and distributed to different stakeholders respectively in the ecosystem.

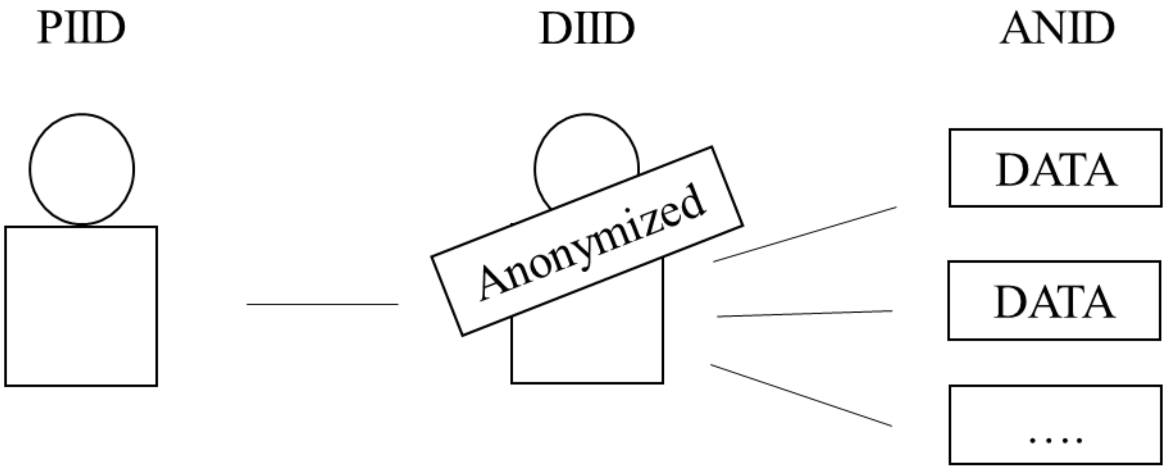


Figure 11. The relationship between PIID, DIID, and ANIDs.

Although the DIID is indirectly linked to an individual's identity, it is likely that it is already connected to the PIID through user profiling. Therefore, to further protect the privacy of the data owner, the DIID is masked with ANID and only the ANID is used in all procedures in data exchange. However, in the end, the ANID must be able to connect to the root DIID or PIID. This is because data consumers can actually purchase if ANIDs is converted to DIID or PIID. However, in order to protect privacy and prevent sensitive identity information such as PIID and DIID from being leaked, ID matching should be handled privately. This allows the requestor to validate and match PIIDs or DIIDs without disclosing critical IDs through the zero-knowledge proof (ZKP) [49], which can only confirm the identity of the identifier, rather than receiving the actual identifier to determine whether the requestor has a valid identifier. The management of connecting and converting these IDs is performed by the Identity Manager.

To protect the privacy of the data owners, identity masking which is the process of replacing PIID and DIID with ANID, and decoupling and anonymizing personal data which is the procedure of separating the storage and management of PIID and DIID from ANID and payload data. In identity masking, when identity masking is complete, ANID is coupled to the payload data, which must have the attributes as shown in Table 3. In decoupling and anonymizing personal data, once the data provider registers the data in the network through smart contact, identity data such as PIID and DIID, and ANID and payload data respectively are stored and managed. Only identity manager nodes can store and manage PIIDs and DIIDs. If a data consumer wants to purchase more suitable quality data, they can merge several datasets through data relayer because identity data and payload data are stored separately.

Anonymity	Except for the ID Manager node, all stakeholders cannot have information about the original identity of the individual, such as PIID or DIID
In-exchangeability	Different stakeholders should have different ANIDs even though they are connected to the same PIID or DIID
Pseudo-identifiability	Within the same stakeholder, ANID should remain the same across one unique person and ANID should be distinguishable between different unique people
Verifiability	Identity manager nodes should be able to verify that such ANID are indeed connected to a certain PIID or DIID

Table 3. The attributes of ANID. [45]

Payload data includes the actual contents of the data such as application installation list, shopping cart history, device usage history, personal interests, and preferences. The storage and management of payload data is the role of data provider. Data consumers can only try e-mail or call with these payload data and cannot contact data owner directly. In addition, data consumers cannot expose payload data to a targeted advertisement. Payload data can be utilized in part for non-commercial purposes such as research and commercial purposes such as business intelligence, without PIID or DIID.

The proposed model can be viewed as a kind of personal information management system defined in [9]. Individuals will have ownership and control of the data as data owners, and data providers will manage and monetize the data within explicit consent. Data owners can manage where their data is stored, and can take it out or delete it if necessary. In this model, data owners are given the right to access data, the right to object to data collection or usage, the right to correct errors in data, the right to delete data, and the right to move data to grant data portability to ensure compliance with the GDPR.

4. 3. Data Storage

In the previous subsection, we mentioned that identity data such as PIID and DIID, payload data and ANID, should be stored separately for data owner's privacy. Blockchain has the property of immutability. It means that if the data is stored in blockchain, it cannot be deleted. However, it is not GDPR compliance. In addition, data stored in the blockchain are opened and verified by all participants. Therefore, the data stored in the blockchain should not contain data that can specify an individual's identity. Thus, blockchain contains only ID data which cannot be used for identifying. The data which does not contain identifiable information is separately stored in off-chain storage. The off-chain means it is not in blockchain, and it can be a cloud or not online repository.

Off-chain is essential for another reason in storing data. The size of the data that can be included in the blockchain is not large. Blockchain network will be stuck in traffic jam by trying to store large data and the tremendous number of data at once. Therefore, off-chain for storing big data is necessary even to maintain a smooth blockchain network. There can be several off-chain at the same time. The off-chain is linked with on-chain which stores the hash value of the data, and all transactions and smart contracts are recorded in the blockchain to provide a transparent and safe data economy.

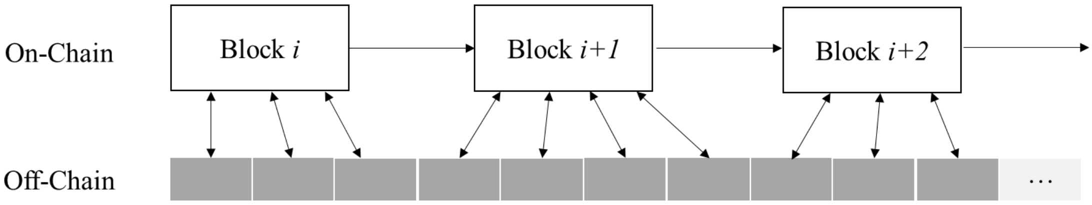


Figure 12. Storing data separately with on-chain and off-chain.

4. 4. Data Transaction

Data Collection: The data provider obtains prior consent from the data owner through the data collection authorization process called DAuth for data collection and monetization. Data providers must obtain consent from individual users through at least once DAuth for each type of data to collect data. Data owners can directly select which data to be collected through the DAuth. Through the DAuth, the kinds of data that users disagree with is not stored and discarded immediately. A process of DAuth is as follows. First, the application displays the DAuth interface to the data owner. Second, the user agrees or rejects the collection of each data individually. Third, the data owner enters an ID to receive compensation. Fourth, the data owner's consent is recorded in blockchain. When data providers collect data, for standardization of data types, data providers are encouraged to use common schemas in standardized data schema registry. Of course, data providers can register another data schema, but incentives are introduced to encourage the use of standardized data schema registry.

Data Discovery: Data consumers should be aware of the identity of the data they want to purchase before purchasing the data. The procedure is called data discovery. Data consumers can search for data via public data market, through data relayers, and from data providers. First, data providers can organize datasets using the collected data and register them on the public market by setting the desired selling price. This method makes it easier for data consumers to discover data. However, here is the downside that data consumers must purchase by organized datasets. Second, data relayer sends identity data to the data consumer and can configure data profiles by query. Thus, data consumers can discover data through the data relayers through the desired specific conditions or filtering. This method allows data consumers to purchase specific data with specific conditions or filters rather than buying the entire dataset from a public data market. Third, data consumer can receive identity data directly from the data provider. This can be done with a second-party data exchange. Companies or data providers that do not want to trade their data in the public data market can exchange data with the data consumer through a mutual agreement.

Data Transaction: Once the data consumer receives IDs of the data, the data can be purchased through mutual agreement. The procedure of data exchange is shown in Figure 13. First, the data consumer suggests the price of the data transaction to the data provider through a smart contract. Second, the data provider may accept or reject the proposed smart contract. Third, if the data provider accepts the smart contract, the data provider should authorize the data consumer so that data consumer can be granted access to the data they want to purchase Fourth, tokens are paid by the data consumer. When the transaction is complete, a record of the data exchange is stored in the blockchain and the data flow is notified to the data owner.

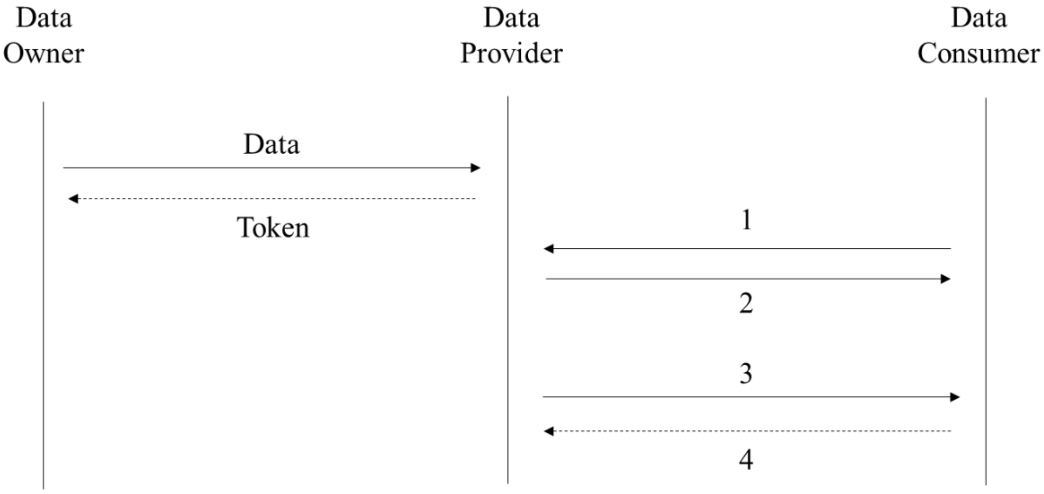


Figure 13. The diagram of data transaction.

4. 5. Token System

Use contributions in order to provide incentives to network participants. Contribution is measured through the level of activity and fame of network participants. Contribution can be expressed as a multiplication of participation and fame. Fame means the network effectiveness and indirect confidence in participants' activities, and contribution means the amount of work that participants have performed on the network over a specific predefined period, contribution period. Therefore, no matter how high the frame is, the contributions can be compensated only if the work on the network is done actually and steadily. Unlike contribution, fame is permanently maintained by participants. Fame is measured including the amount of token held by the participant, the amount of token collateral held by the data provider or identity manager, the number of identities provided, and the actual number of uses of the data schema provided.

Participation increases as participants perform the work of the network and participate in the network. However, participation activities do not necessarily have to be direct. Participation increases if participants participate in the process of facilitating data purchases even if participants do not have direct purchasing activities. The work that can increase participation is limited to cryptographically verifiable tasks. For example, if participants performed activities such as paying fees through data transactions or performing PoW mining, it can be recognized as participating because it is provable cryptographically. Initialize participation after the contribution period so that participants participate continuously in the network and receive contribution reward.

Contribution from participants is calculated during the contribution period and at the end of the contribution period, contribution reward paid to the participants. All fees and mined reward of a block accumulated in the fee pool during the contribution period, and contribution rewards are allocated proportionately to each participant's contribution. Contribution reward is allocated as follows:

$$\text{Contribution Reward} = \text{Mined Reward} * \frac{\text{User}'\text{s Contribution}}{\text{All Contributions}}$$

Tokens are largely divided into two types. One is security token and the other is utility token. A security token is a negotiable ERC20 token, used primarily as a means of data payment or as a means for network participation, such as to qualify to turn a node. Utility tokens are virtual tokens that cannot be traded, which can only be converted to security tokens for use in that model, and are not reversible in order to avoid the practice of buying fame on the network at the cost of payment. Used primarily as a means of providing rewords in the network. When the data consumer finishes paying the data with a security token, the security token is converted to a matching utility token and paid as a reward to the data provider.

Security Token: Used as a means of data payment or as a means of commission payment. It can be purchased on the exchange or converted to matching utility tokens. The following may be used: If you want to purchase data, you can purchase the data by paying the token. In order to participate in the network as a data provider, a data verifier node, or as an identity manager, they must hold a certain amount of security tokens as collateral for that.

Utility Token: A virtual token paid as compensation for productive activities taking place within the proposed model. The amount of data consumers' payments or security tokens issued as rewards, are converted to utility tokens to provide data by users, or paid as compensation whenever a node sells the data. Utility tokens are attributed to individuals and cannot be sent to others. Therefore, it can be used to assess an individual's fame and contribution.

Utility tokens exist separately to control the supply of security tokens. Participants with high fame on the network may be rewarded with more contributions through activities, and consumers prefer to purchase data with high fame. Thus, participants can immediately exchange for security tokens to hold utility tokens rather than leaving the ecosystem so that the ecosystem remains sustainable and efficient.

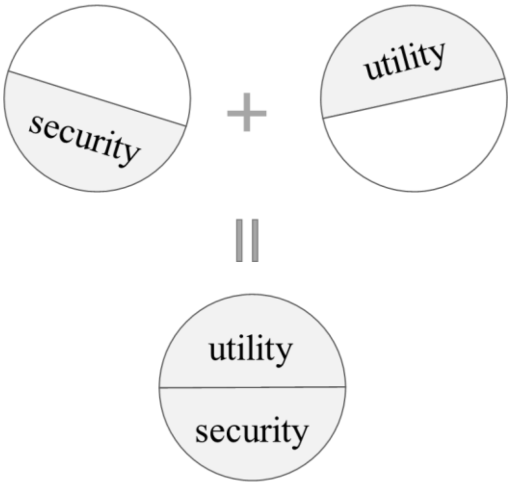


Figure 14. Token separation to make ecosystem stable.

V. DISCUSSION

In this section, we will discuss why our blockchain model is more proper to utilize personal data while the model has anonymity, the effectiveness of search, and stability in the ecosystem. The comparison is summarized in Table 4.

First, we focus on the view of anonymity. In MedBlock, the client encrypts the summary of collected users' EMRs with the patients' public key. It means that MedBlock uses only DIID during data transactions because the summary is encrypted by the key of each patient. In Nebula Genomics, data owners store their genomes as arrays of sequencing hashes. However, the network address from the Ethereum blockchain is dealt in data transaction, it means only DIID is used. Therefore, we can refer that the anonymity of Nebula Genomics is similar to MedBlock. In the case of MeDShare, during the data transaction, it is made that a package which includes a data ID, payload data, and smart contract. Since single package is encrypted at the same time, it cannot be considered that ANID is used to encrypt for each data. However, anonymity can be slightly more guaranteed, since it is not a single DIID for the individual who owns the data. It can be referred that MeDShare has a little more anonymity than MedBlock and Nebula Genomics. If a package has a single payload data, MeDShare has good anonymity. In the proposed model, we use each ANID, not DIID while data transactions. DIID is a hash value of individual, but ANID is hash value of each data. It means that an individual can have many ANIDs, so guessing individual of ANIDs is almost impossible. In other words, if individuals have more data, it is harder to identify them. Therefore, we can refer the proposed model has more anonymity than MeDShare, MedBlock, and Nebula Genomics.

Second, we discuss the view of the effectiveness of search. MeDShare, MedBlock, and Nebular Genomics have the function of the query to discover the data. MeDShare has triggers who have the role of translation the requested query into an understandable structure. However, MedBlock and Nebular Genomics have the same limitation. The person who wants the data has to complete the query himself. It will be difficult for a person with no expertise to write a query. Moreover, query on a blockchain is not the same as normal SQL because blockchain is a decentralized database. The proposed model has a data relayer because of these difficulties. In the proposed model, if the person wants the data himself, it is also possible to complete a query. If data consumers have no expertise, they can pay token and ask data relayer for a query. At the same time, data relayer has an incentive to remain in the ecosystem by getting tokens. This can make it easier to discover in the proposed model than MedBlock and Nebula Genomics.

Third, we look at the stability in the ecosystem. There is no mention of any reward such as a token in MeDShare. Therefore, members of MeDShare have no incentive to maintain an ecosystem. It means that the ecosystem cannot be guaranteed to remain stable. MedBlock also has no mention of any reward such as a token. Therefore, it will be difficult for MedBlock to maintain its ecosystem. Nebula Genomics has a reward system by a token. However, it is not enough to maintain the ecosystem continuously. There are two types of tokens: security token and utility token, and the one-token problem will arise if the model has only one type of token. Utility token is usually priced stable and used for payment. If a token is inclined to utility token, the attraction to receive it decreases. Security token contains the value of the ecosystem. If security token is highly inclined, the ecosystem cannot produce a large amount of distribution, which slows down the function as a token for payment, causing the ecosystem to stall. The proposed model can avoid this one-token problem by having both types of tokens. It keeps the ecosystem more stable than MeDShare, MedBlock, and Nebula Genomics.

	MeDShare	MedBlock	Nebular Genomics	Proposed Model
Anonymity	Conditionally Good	Normal	Normal	Good
The Effectiveness of Search	Good	Normal	Normal	Good
Stability in the Ecosystem	Not Good	Not Good	Not Good	Good

Table 4. The summary of comparisons between the models.

We remain some work for the future. In future, we implement the proposed model to measure how good the advantages we've discussed are over the existing models. In this paper we model roughly conceptual design for MyData blockchain with transparency and the GDPR rights, but in the future, we are planning to do more detailed modeling which can be commercialized in real world.

VI. CONCLUSION

As the number of personal data increase, the amount of data explosively increases. It meets the demand for pattern recognition and big data analysis. However, personal data should be handled with care because of privacy issues. As part of that, the EU has established the GDPR, which is a guideline for many countries and businesses in data utilization. The GDPR ensures that the data owner has various rights. Typically, ownership of the data is transferred from businesses or organizations to the data owner, and the data owner takes ownership back, enabling self-activity and being distributed revenue from the data economy. A model of personal data utilization through blockchain continues to emerge, as blockchain, which decentralizes the centralism and has self-determination of the data, has many similarities.

In this paper, we propose a blockchain model focused on providing transparency and control over users' personal data utilization. Based on blockchain and smart contract technology, the proposed model provides high trust, security, and efficiency. Compared to the blockchain-based model described earlier, the proposed model has more anonymity during data transactions, the effectiveness of search, and stability in the ecosystem. The proposed blockchain model facilitates the ability to transfer the control of personal data to users in a transparent manner and to create and to destroy agreements to access and sell their data to companies that want to purchase personal data.

References

[1] John Gantz and David Reinsel, “THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East”, *IDC iVIEW: IDC Analyze the future*, 2012.

[2] M. Dell, Dell Technologies World 2018, May 2018.

[3] M. Mohammadi *et al*., “Deep Learning for IoT Big Data and Streaming Analytics: A Survey,” *IEEE Communications Surveys & Tutorials*, Jun. 2018.

[4] K. Schwab, “The Fourth Industrial Revolution,” Portfolio Penguin, 2017.

[5] C. M. Bishop, “Pattern Recognition and Machine Learning,” SPRINGER, 2007.

[6] M. Najafabadi *et al.*, "Deep Learning Applications and Challenges in Big Data Analytics”, *Journal of Big Data*, 2015.

[7] C. Newton, “Facebook suspended Donald Trump’s data operations team for misusing people’s personal information,” *THE VERGE*, [Online], 16 Mar, 2018. Available: https://www.theverge.com/2018/3/16/17132172/facebookcambridge-analytica-suspended-donald-trump-strategiccommunication-laboratories

[8] G. Jung and H.-N. Lee, “Data De-identification Issues and Technology Status,” *Summer Annual Conference of IEIE*, 2019.

[9] “General Data Protection Regulation,” 2018. [Online]. Available: https://gdpr-info.eu/

[10] “List of largest Internet Companies.” [Online]. Available: https://en.wikipedia.org/wiki/List_of_largest_Internet_companies

[11] “2018 Data Industry White Paper,” *Kdata*, 2018.

[12] I. Li, A. K. Dey, and J. Forlizzi, “Understanding My Data, Myself: Supporting Self-Reflection with Ubicomp Technologies,” *Ubiquitous Computing, 13th International Conference*, Sep. 2011.

[13] Satoshi Nakamoto, “Bitcoin: A peer-to-peer electronic cash system”, 2008.

[14] B. Faber *et al*., “BPDIMS: A Blockchain-based Personal Data and Identity Management System,” *Hawaii International Conference on System Sciences*, Jan. 2019.

[15] G. Zyskind, O. Nathan *et al*., “Decentralizing Privacy: Using Blockchain to Protect Personal Data,” in *Security and PrivacyWorkshops (SPW), 2015 IEEE*, pp. 180–184, IEEE, 2015.

[16] N. Kaaniche and M. Laurent, “A Blockchain-based Data Usage Auditing Architecture with Enhanced Privacy and Availability,” in *Network Computing and Applications (NCA), 2017 IEEE 16th International*

Symposium on, pp. 1–5, IEEE, 2017.

[17] A. Yasin and L. Liu, "An Online Identity and Smart Contract Management System," in *Computer Software and Applications Conference (COMPSAC), 2016 IEEE 40th Annual*, vol. 2, pp. 192–198, IEEE, 2016.

[18] P. Mamoshina *et al*., "Converging blockchain and next-generation artificial intelligence technologies to decentralize and accelerate biomedical research and healthcare," *Oncotarget*, vol. 9, no. 5, p. 5665, 2018.

[19] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, "MedRec: Using Blockchain for Medical Data Access and Permission Management," in *Open and Big Data (OBD), International Conference on*, 2016.

[20] X. Yue, H. Wang, D. Jin, M. Li, and W. Jiang, "Healthcare Data Gateways: Found Healthcare Intelligence on Blockchain with Novel Privacy Risk Control," *Journal of medical systems*, vol. 40, no. 10, p. 218, 2016.

[21] G. D. P. Regulation, "REGULATION (EU) 2016/679 – Directive 95/46," *Official Journal of the European Union (OJ)*, vol. 59, pp. 1–88, 2016.

[22] Q. Xia *et al*., "MeDShare: Trust-Less Medical Data Sharing Among Cloud Service Providers via Blockchain," *IEEE Access*, Jul. 2017.

[23] K. Fan *et al*., "MedBlock: Efficient and Secure Medical Data Sharing Via Blockchain," *Journal of Medical Systems*, Aug. 2018.

[24] D. Grishin *et al.*, "Blockchain-enabled Genomic Data Sharing and Analysis Platform," *Nebula Genomics*, Jan. 2018. [Online]. Available: https://nebula.org/

[25] M. Ali and J. Nelson, "Blockstack: A Global Naming and Storage System Secured by Blockchains," in *the Proceedings of the 2016 USENIX Annual Technical Conference (USENIX ATC '16)*, Jun. 2016.

[26] J. Benet, "IPFS - Content Addressed, Versioned, P2P File System," Jul. 2014. arXiv:1407.3561v1 [cs.NI]

[27] D. Qiu and R. Srikant, "Modeling and Performance Analysis of BitTorrent-Like Peer-to-Peer Networks ," *Computer Communication Review*, 2004.

[28] "Open Source Big Data Processing and Bioinformatics," *Arvados*. [Online]. Available: https://arvados.org/

[29] V. Costan and S. Devadas, "Intel SGX Explained." [Online]. Available: https://eprint.iacr.org/2016/086.pdf

[30] L. Lamport, R. Shostak, and M. Pease, "The Byzantine Generals Problem," *ACM Transactions on Programming Languages and Systems (TOPLAS)*, vol. 4, no. 3, pp. 382–401, 1982.

[31] "Bitcoin Core Charts." [Online]. Available: https://charts.bitcoin.com/

[32] L. Xu, "Highly available distributed storage systems," PhD dissertations, California Institute of Technology, 1999.

[33] A. Back, “Hashcash - A Denial of Service Counter-Measure.” [Online]. Available: http://www.hashcash.org/ papers/hashcash.pdf, 2002.

[34] BitFury Group, “Proof of Stake versus Proof of Work.” [Online]. Available: http://bitfury.com/content/5-white-papers-research/pos-vs-pow-1.0.2.pdf, 2015.

[35] J. Carter and M. N. Wegman, “Universal Classes of Hash Functions,” *Journal of Computer and System Sciences*, vol. 18, no. 2, pp. 143 – 154, 1979.

[36] R. C. Merkle, “Protocols for Public Key Cryptosystems,” in *Security and Privacy, 1980 IEEE Symposium on*, pp. 122–122, IEEE, 1980.

[37] A. M. Antonopoulos, “Mastering Bitcoin: Unlocking Digital Cryptocurrencies,” O’Reilly Media, Inc., 2014.

[38] R. L. Rivest, A. Shamir, and L. Adleman, “A Method for Obtaining Digital Signatures and Public-Key Cryptosystems,” *Communications of the ACM*, vol. 21, no. 2, pp. 120–126, 1978.

[39] D. Johnson, A. Menezes, and S. Vanstone, “The Elliptic Curve Digital Signature Algorithm (ECDSA),” *International journal of information security*, vol. 1, no. 1, pp. 36–63, 2001.

[40] N. Szabo, “Formalizing and Securing Relationships on Public Networks,” First Monday, vol. 2, Sep. 1997.

[41] M. Swan, “Blockchain: Blueprint for a New Economy,” O’Reilly Media, Inc., 2015.

[42] A. Poikola, K. Kuikkaniemi, and H. Honko, “MyData – A Nordic Model for human-centered personal data management and processing,” *Finnish Ministry of Transport and Communications*, 2015.

[43] “Unlocking the Value of Personal Data: From Collection to Usage”, *World Economic Forum*, 2013.

[44] “Guideline on Personal Data De-identification“, *KISA*, 2018. [Online]. Available: https://www.kisa.or.kr/jsp/common/downloadAction.jsp?bno=282&dno=3&fseq=1

[45] “Airbloc Protocol,” *Airbloc*, Oct. 2018. [Online]. Available: https://airbloc.org/

[46] J. Poon and V. Buterin, “Plasma: Scalable Autonomous Smart Contracts,” Aug. 2017. [Online]. Available: https://plasma.io/

[47] “Delio, Value Your Values,” *Delio*, 2018. [Online]. Available: https://www.delio.io/

[48] J. Schiffman, X. Zhang, and S. Gibbs, “DAuth: Fine-grained Authorization Delegation for Distributed Web Application Consumers,” *2010 IEEE International Symposium on Policies for Distributed Systems and Networks*, 2010.

[49] U. Feige, A. Fiat, and A. Shamir, “Zero-Knowledge Proofs of Identity,” *Journal of Cryptology*, 1988.

Acknowledgement

GIST에 입학한 지 얼마 지나지 않은 것 같은데 벌써 졸업을 앞두게 되었습니다. 대학원생으로서 연구실 생활을 시작하면서 저는 지금까지 해왔던 공부와 앞으로 하여야 할 연구는 다르다는 것을 깨달았습니다. 2년간 INFONET 연구실에서 교수님과 연구실 선후배들과 동고동락하며 소중한 추억을 쌓을 수 있었고, 좋은 글쓰기는 생각만큼 쉽지 않다는 것도 깨닫게 되었습니다. 제가 졸업논문까지 작성하기에 있어, 여러 분야에 대해 토론과 상담을 아끼지 않았던 연구실 선후배, 졸업 논문에 귀한 코멘트를 주신 심사위원 교수님들, 논문과 학업에 있어 아낌없이 지도해주신 지도교수님, 그리고 주변에서 항상 힘이 되어준 친구들과 부모님께 진심으로 감사드립니다.