Secure Sharing Scheme of Sensitive Data in the Precision Medicine System

Deukhun Kim[1], Heejin Kim[2] and Jin Kwak[3, *]

Abstract: Numerous industries, especially the medical industry, are likely to exhibit significant developments in the future. Ever since the announcement of the precision medicine initiative by the United States in 2015, interest in the field has considerably increased. The techniques of precision medicine are employed to provide optimal treatment and medical services to patients, in addition to the prevention and management of diseases via the collection and analysis of big data related to their individual genetic characteristics, occupation, living environment, and dietary habits. As this involves the accumulation and utilization of sensitive information, such as patient history, DNA, and personal details, its implementation is difficult if the data are inaccurate, exposed, or forged, and there is also a concern for privacy, as massive amount of data are collected; hence, ensuring the security of information is essential. Therefore, it is necessary to develop methods of securely sharing sensitive data for the establishment of a precision medicine system. An authentication and data sharing scheme is presented in this study on the basis of an analysis of sensitive data. The proposed scheme securely shares sensitive data of each entity in the precision medicine system according to its architecture and data flow.

Keywords: Precision medicine, sensitive data, keyless signature infrastructure, data sharing.

1 Introduction

Owing to recent advancements in internet technology, all objects are connected through the Internet of Things (IoT) in various industries, leading to the evolution of increasingly intelligent societies. Numerous industries have emerged, of which the medical domain is of great significance; therefore, various studies are being conducted in this field [Liu, Li, Qu et al. (2017)]. In particular, subsequent to the announcement of the precision medicine initiative (PMI) by the United States in 2015 [Hudson, Lifton and Patrick-Lake (2015)], the global interest in precision medicine has increased considerably. In the medical field, genetic, clinical, and lifestyle data are required to implement precision medicine. Thus, sensitive data, such as patient history, deoxyribonucleic acid (DNA), and personal information, are considered to be key data. The success of efforts to collect a broad

[1] ISAA Lab., Department of Computer Engineering, Ajou University, Suwon, 16499, Korea.

[2] Korea Orphan & Essential Drug Center, Seoul, 04523, Korea.

[3] Department of Cyber Security, Ajou University, Suwon, 16499, Korea.

[*] Corresponding Author: Jin Kwak. Email: security@ajou.ac.kr.

Received: 09 March 2020; Accepted: 02 May 2020.

CMC. doi:10.32604/cmc.2020.010535 www.techscience.com/journal/cmc

spectrum of patient information depends on broad public support and willingness to participate [Sankar and Parker (2017)]. This is recognizes the enormous benefits of data collected and created for research purposes being made available for secondary uses, as open science gains increasing support. However, also challenges relating to the collection, storage, and re-use of research data [Xafis and Labude (2019)]. In the near future, much larger volumes and complex datasets for precision medicine will be generated [Qian, Zhu and Hoshida (2019)]. However, publishing data may divulge individual sensitive data. Currently many existing privacy protection schemes cannot provide the balance of utility and protection. Accordingly, research on measures to protect privacy in various fields is ongoing [Gu, Yang and Yin (2018); He, Zeng, Xie et al. (2017); Min, Yang, Wang et al. (2019); Yin, Shi, Sun et al. (2019)]. This is where the discussion of privacy and techniques often ends in the scientific health literature when internet-related technologies have made privacy a much more complex challenge with broad psychological and clinical implications [Aboujaoude (2019)]. Especially, an individual's genetic data forms the bedrock of precision medicine [Beauvais and Knoppers (2020)]. This is recognized as sensitive for multiple social reasons, raising concerns about privacy and questions about best practices for governance of personal genomics data access [Rubin and Glusman (2019)]. Also, technological advances require collecting and sharing the massive amount of data and thus generate concerns about privacy [Noorbakhsh-Sabet, Zand, Zhang et al. (2019)]. In addition, patient health data are often found spread across various sources. But, precision medicine and personalized care requires access to the complete medical records [Chen, Jiang, Wang et al. (2018)]. Precision medicine data storage requirements are ever increasing and long-term data protection schemes become more complex. The assurance of sensitive data integrity has almost not been discussed yet. Sensitive data needs to be secured against loss and forging [Buchmannm, Geihs, Hamacher et al. (2019)]. So, data protection and privacy law are key determinants in precision medicine's future [Beauvais and Knoppers (2020)]. Cloud computing with protected patient privacy would become more routine analytic practice to fill the gaps within data integration along with the advent of big data. Integration of multitudes of data generated for each individual along with techniques tailored for big data analytics may eventually enable us to achieve precision medicine [Qian, Zhu and Hoshida (2019)]. Thus, to implement precision medicine, studies must be conducted on the development of security techniques to protect privacy and share sensitive data.

In this paper, techniques are presented, based on related works, for securely sharing sensitive data among the entities participating in the precision medicine system. Section 2 presents the definition of precision medicine and an analysis of the infringement threats to sensitive data that can occur in a precision medicine system. The corresponding security requirements are also outlined. This section also describes a technique for applying the keyless signature infrastructure (KSI) that is used to reduce the workload when sharing sensitive data that is considered to be big data. Section 3 presents an analysis of the sensitive data defined in the PMI of the United States. Section 4 details the scheme for securely sharing sensitive data, and Section 5 presents the conclusions of this study.

2 Related works

2.1 Definition of precision medicine

Precision medicine is used to provide optimal treatment and medical services to patients. In broad terms, precision medicine is the customization of treatments at the individual patient level [Matrana and Campbell (2020)]. It is also used to prevent and manage diseases through the collection and analysis of big data related to the genetic characteristics, occupation, living environment, dietary habits, etc., of each patient. This concept was developed from personalized medicine that involves the development of various personalized therapies and preventive measures for each patient. However, in the precision medicine model, individuals are classified on the basis of genetic information, lifestyle, and clinical data; it provides more preemptive healthcare services through a large-scale genomic data analysis. In a 2011 report published by the National Research Council (NRC) of the National Academies of Sciences, Engineering and Medicine (NASEM) of the United States, in addition to the new taxonomy of diseases based on molecular biology, emphasis was laid on the need to use molecular biological information, including new genetic information that is being produced rapidly. The goal of the new classification system was to provide the most effective preventive and therapeutic approach based on an individual's lifestyle and environmental and genetic factors instead of adopting "personalization" that comprises specialized prevention and treatment based on an individual's unique characteristics. Thus, the term "precision medicine" is more appropriate than "personalized medicine" in the technical sense as it encompasses the meanings of the terms "accurate" and "precise" [Desmond-Hellmann and Sawyers (2011)]. This was embodied in the 2015 PMI project by the National Institute of Health (NIH) of the United States. In contrast to a one-size-fits-all approach that is based on an average and involves the application of one solution to all patients for treating and preventing diseases, the aforementioned approach is an innovative medical approach that comprehensively considers individual differences such as those in genetics, environment, and lifestyle. This could mean better prediction of someone's disease risk and more effective diagnosis and treatment if they have a condition [Schaefer, Tai and Sun (2019)]. In 2016, precision medicine was selected as the focus of the national science and technology strategy project in South Korea and defined as the "integration and analysis of genomic, medical treatment, clinical, environment, and lifestyle data for providing medical services personalized to the patients' individual characteristics." This project aimed to set up the foundation for precision medicine techniques by establishing a precision medicine cohort that accumulates genetic, medical treatment, environment, and lifestyle data in real time of at least 100,000 people; it also provides a platform to companies and hospitals for linking and using the accumulated research resources. Although a standard classification of the techniques related to the industry of precision medicine is not clearly established, the data are typically classified as genomic information, omics, and big data, and the analysis of vast and varied data is considered the core technique. Precision medicine is a turning point in the paradigm shift for providing better medical services than those currently available. To accomplish this goal, researchers are actively conducting studies across various fields [Ferryman and Pitcan (2018)]. Data, which are the key to implementing precision medicine, are vast in volume, comprise numerous types, and require advanced techniques, such as artificial intelligence

(AI), IoT, and cloud techniques, for storage, processing, and analysis. They require a system that provides a platform for storing and managing cloud-based genomic analysis data and enables data mining techniques to obtain meaningful results. Such a precision medicine system can be configured through entities including a cohort that is a patient specific group that donates the data, a data controller that collects the data, and a data processor that processes the collected data. The collected or processed data significantly vary between the different entities. They include omics/diagnostic data, clinical data, such as electronic medical record (EMR)/electronic health record (EHR), drug compliance, personal diet, wearable sensor data, environmental data, and information regarding personal preferences. These sensitive data can be classified into four types, namely healthcare, genetic, lifelog, and privacy data. Therefore, in contrast to the existing medical methods, precision medicine involves the collection and analysis of data using Information and Communication Technology (ICT). Sensitive personal information, such as genetic data, entails significant risk and potential ripple effects if disclosed to the public or misused. Moreover, an infringement of these data can result in penalties as stipulated in the patient privacy protection regulations of the Health Insurance Portability and Accountability Act (HIPAA), General Data Protection Regulation (GDPR), or Personal Information Protection Act (PIPA) [Klonoff and Price (2017)]. Owing to the economic value of the personal and medical information generated in the healthcare sector and the high profitability of cyber-attacks against medical institutions, cyber security infringements, such as Distributed Denial of Service (DDoS) attacks and ransomware, are steadily increasing. User anxiety is also increasing owing to the handling of sensitive information such as patients' private data and disease and genetic information in cloud environments. Security can thus be viewed as a key factor in the implementation of precision medicine. Therefore, it is necessary to develop methods for securely sharing sensitive data to establish a precision medicine system.

2.2 Analysis of infringement threats to sensitive data and security requirements for a precision medicine system

This section presents an analysis of the infringement threats that may occur when sharing sensitive data that are collected and processed in a precision medicine system. The security requirements for preventing these threats are also presented.

2.2.1 Sensitive data infringement threats

- Threat 1: Data exposure

In a precision medicine system, when sharing data between entities, data may be exposed owing to data eavesdropping in sniffing attacks. Privacy may be compromised in such cases.

- Threat 2: Data forgery and falsification

Data may be forged or falsified due to a change in data in spoofing attacks or factors such as network errors. In such a situation, the results obtained through big data analysis may vary, and difficulties may be encountered in providing precision medicine services.

- Threat 3: Unauthorized entities

To provide precision medicine services, accurate data based on facts are required. If unknown or unreliable data are collected and processed by an unauthorized entity, the authenticity of the services may be compromised.

- Threat 4: Replay attack

If the same data are repeatedly collected owing to data retransmission, it may become difficult to process the collected data to establish accurate statistics and devise a classification system for the diseases.

- Threat 5: Repudiation

After the data are sent or received in the process of data sharing, repudiation on sending or receiving data may occur. This may result in difficulties in providing precision medicine services.

2.2.2 Security requirements

- Requirement 1: Data confidentiality guarantee

To prevent the exposure of data during data sharing, an encryption technique that facilitates the sharing of data in the form of cipher text must be implemented. During this step, data confidentiality must be guaranteed by configuring a method for encrypted communication with entities that are authenticated.

- Requirement 2: Data integrity verification

To prevent data forgery or falsification, data integrity must be verified—the results obtained before and after sharing the data must be compared using a hash function. Recently, techniques such as blockchain and KSI have been used for data integrity verification.

- Requirement 3: Entity authentication

To prevent the sharing and processing of data by unauthorized entities, a mutual authentication technique must be implemented on the entities that are configured in a precision medicine system.

- Requirement 4: Data validity verification

To prevent data replay attacks, data must be validated by applying techniques such as the use of a sequence number or time stamp on the data transfer protocol.

- Requirement 5: Nonrepudiation

To prevent repudiation of data sharing, a digital signature technique must be applied in the data sharing process.

In addition, as the sensitive data used in a precision medicine system is considered to be big data, a method of securely and efficiently processing these data must be developed. The KSI technique can reduce the data processing load by reducing the cryptographic computations for protecting the data.

2.3 Keyless signature infrastructure

According to a recent survey conducted among the top management teams of medical institutions, approximately 89% of the respondents stated that data integrity is an important issue in precision medicine in terms of decision-making regarding data utilization when analyzing sensitive data on a big-data scale [Safavi and Kalis (2018)].

Thus, the data integrity assurance technique can be considered a key technique in providing precision medicine services. This section presents an analysis of KSI as a data integrity assurance technique [Buldas, Kroonmaa and Laanoja (2013)].

KSI was developed by Guardtime, which is a security company based in Estonia. It can be used as a replacement for the existing public key infrastructure (PKI) signatures. The term "keyless" in KSI does not mean that no cryptographic keys are required during the signature generation process; instead, it means that keys are still required for authentication, but the signatures can be reliably generated and verified without assuming the continued secrecy of keys. Keyless signatures perform signer identification and integrity protection separately and are implemented as multisignatures. Therefore, KSI-based research is being conducted in various fields [Ra and Lee (2018); Mylrea, Gourisetti, Bishop et al. (2018)]. The signing process for the data is detailed below.

Step 1 Hashing:

The data to be signed are hashed, and the hash values are used to represent the data in the rest of the process.

Step 2 Aggregation:

A global temporary per-round hash tree is generated to represent all the data signed during one round.

Step 3 Publication:

The root hash values of the aggregation trees from each round are collected into a perpetual hash tree, known as a hash calendar, and the root hash value of this tree is published as a trust anchor.

An infrastructure is established to implement such signature processes in practice. It consists of a hierarchy of aggregation servers that generate hash trees every round through collaboration. It comprises an aggregation network, a core cluster, and a gateway.

- Aggregation network:

An aggregator is a system component that creates hash trees from the received requests and passes the root hash values to upstream aggregators. Further, upon receiving a response, the aggregator delivers the response to the child aggregators. As each aggregator has its own reserved spot in the hash tree, the servers involved in the creation of a specific signature token can be proved.

- Core cluster:

The core cluster comprises top-level aggregators from each round. It is responsible for creating the hash calendar and propagating and synchronizing it through the aggregation network. To verify the integrity of the root hash values of the calendar, they are archived and distributed to verification servers through the archiving and caching layers. In addition, the roots of the intermediate aggregation trees are stored only in relevant signature tokens. During this process, the gateways copy their calendars from the cache servers using the hypertext transfer protocol, which are then used for signature token verification.

- Gateway:

The gateway operates as a protocol adapter where it accepts the requests of an application and sends them to the designated aggregators. The first level of aggregation occurs at a gateway node. The gateway uses an extender service to validate the signature token. The process of using the KSI is illustrated in Fig. 1.

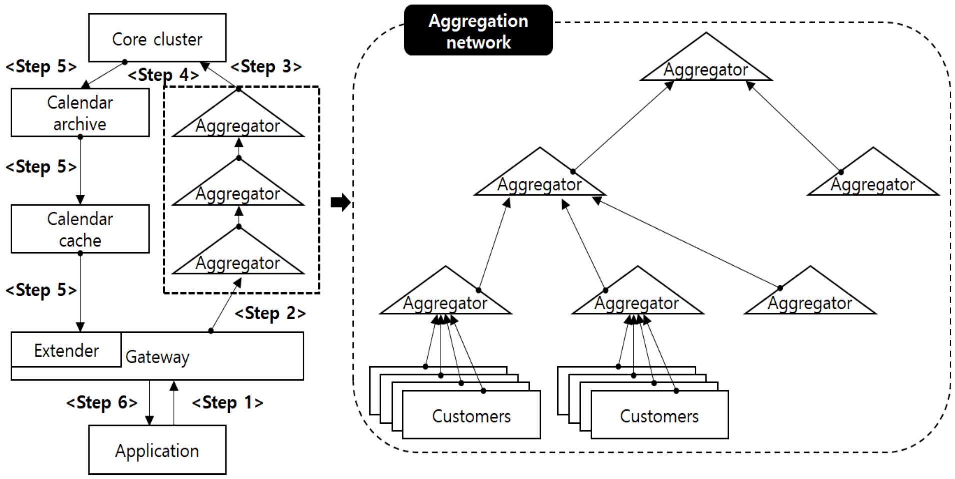

Figure 1: Keyless signature infrastructure

Step 1

The hash of the data to be signed is first computed by the application, and a request is sent to the gateway that provides services to the user.

Step 2

The received requests are aggregated during the period (round) by the gateway that received the request, and the aggregate request is sent to the upstream aggregation cluster to request the top hash value.

Step 3

The requests are aggregated through multiple layers of aggregators, and a globally unique top hash value is generated by the core cluster.

Step 4

The responses consisting of a verifiable hash tree path are sent back through the aggregation layer.

Step 5

The top hash values for each period are collected into the hash calendar archive layer and distributed through the calendar cache layer to the extender service, which is co-located with the gateway host.

Step 6

The application utilizes an extender service to verify the signatures.

3 Analysis of sensitive data in a precision medicine system

In this section, we analyze the sensitive data collected and processed by each entity of the precision medicine system as regulated by the PMI project of the NIH of the United

States [Hudson, Lifton and Patrick-Lake (2015)]. In the analysis, the data used to provide precision medicine services were classified into core and subgroup data that referred to essential and subsidiary data, respectively.

• Individual demographics and contact information

The individual demographics and contact information included twelve examples such as the participant's name, date of birth, gender, race, occupation, contact information, and income. These data were provided by the research participants and medical service providers and comprised core data that were collected and processed in the precision medicine system.

• Terms of consent and personal preferences for participation in the project

This information included details of the project participation options such as receiving the results of the research. These data were provided by the research participants and were core data.

• Self-reported measures

These comprised self-reported measurement information that included six examples, namely pain scales, disease specific symptoms, functional capabilities, quality of life and well-being measures, gender identity, and family health history. These data were provided by the research participants and comprised core data as well as subgroup data necessary for specific research.

• Behavioral and lifestyle measures

These included information regarding behavior and lifestyle, along with the six examples of diet, physical activity, alternative therapies, alcohol consumption, smoking, and assessment of risk factors. These data were obtained from the participants of prospective or retrospective research studies and medical service providers and comprised core as well as subgroup data.

• Sensor-based observations through phones, wearables, and home-based devices

Sensor-based information obtained through mobile and home-based devices included the four examples of location, activity monitoring, cardiac rate and rhythm, and respiratory rate. These data were obtained using mobile device sensors and commercial bio-monitoring services and comprised core as well as subgroup data.

• Structured clinical data derived from EHRs

Structured clinical data derived from EHRs included the four examples of international classification of diseases (ICD)/current procedural terminology (CPT) billing codes, clinical laboratory values, medication, and problem lists. These were core data obtained from several providers having the information regarding research participants or from direct or institutional management channels for personally uploaded or downloaded information by participants.

• Unstructured and specialized types of clinical data derived from EHRs

These included the three examples of narrative documents, images, and electrocardiogram/electroencephalogram data that were provided by multiple providers and not included in the core dataset. They were obtained by an integrated query and comprised subgroup data necessary for specific research. These types of data were only collected and processed in the precision medicine system.

- PMI baseline health examination

This included information related to three examples, namely vital signs, medication assessments, and past medical history, that was provided by the research participants interacting with the medical service providers, and it comprised core data.

- Healthcare claims data

The healthcare claims data included three examples, namely periods of insurance coverage for the patients participating in the research project and the charges and associated billing codes as received by public and private payers. These data were provided by public and private payers and pharmacy insurance coverage management organizations, and they comprised core data.

- Research-specific observations

These included the four examples of research questionnaires, ecological momentary assessments, physical performance measures, and disease specific monitoring. They were provided by the research participants and research organizations, and they comprised subgroup data that were necessary only for a specific study.

- Biospecimen-derived laboratory data

The biospecimen-derived laboratory data included eight examples such as genomics, proteomics, and cell-free DNA. These data were provided by the research participants, genetic information providers, and outsourced laboratories, and they comprised core data.

- Geospatial and environmental data

Geospatial and environmental data included seven examples, including weather, air quality, and food desserts. They were provided by the statistics of public and private information and comprised core as well as subgroup data.

- Other data

Other data included information obtained through social networks that were based on the statistics of public and private information, and it comprised subgroup data.

4 Proposed scheme

In this section, we develop a scheme for securely sharing sensitive data in a precision medicine system. To define the system and sensitive data to which the data sharing scheme is applied, we first propose a system architecture and data flow to provide the precision medicine service. The sensitive data, required to be restructured according to this scheme, are collected and processed. Next, the data structure and contents of the PMI project are restructured into the four major categories defined in this study, namely healthcare, genetic, lifelog, and privacy data, and mapped to the flow of the established precision medicine system. We then propose a secure data sharing scheme using the KSI-based technique.

4.1 Precision medicine system architecture and data flow process

In this section, the precision medicine system is established, and a process of data flow is proposed, according to which, the restructured sensitive data are collected and processed.

Government has an important role in helping to fund primary research in precision medicine and precision public health, defining and optimizing measures of health care quality and security, and ensuring data privacy standards and protections, interoperability, and integration with surveillance systems. Government partnership and collaboration with the non-profit and private sectors can optimize precision medicine and precision public health for the benefit of global population [Whitsel, Wilbanks, Huffman et al. (2019)]. The proposed system consists of healthcare, genetic, lifelog, and privacy data, depending on the type of sensitive data used, based on the centralized precision medicine data center. Core techniques were applied for data management, processing, and security in the precision medicine system environment. The detailed definitions of the entities that comprise the precision medicine system are as follows:

- Cloud-based precision medicine data center (C-PMDC)

A cloud-based centralized data center collects and processes the sensitive data from entities that constitute the precision medicine system. The analyzed data obtained from the entities can be used to provide precision medicine services.

- Healthcare data area

This area includes the cohort participating in the precision medicine projects and entities that provide and demand healthcare data.

- Cohort

This comprises a group that shares characteristics related to a specific subject investigated in the precision medicine projects. The cohort receives medical treatment and prescriptions from healthcare service providers and fills out self-reported and behavioral data. It creates lifelog data and provides them to the entities included in the lifelog data area.

- Healthcare service providers (HSPs)

These include institutions such as hospitals and pharmacies that offer healthcare services.

They provide clinical data obtained from the cohort through treatment and prescriptions (EHR_{HSP}, $BEHR_{HSP}$) and the collected healthcare data such as self-reported and behavioral data ($SELF\_BH_{HSP}$) to the cloud-based cohort data center.

- Cloud-based cohort data center (C-CDC)

This is a cloud-based data center that collects and manages data from the HSPs. Healthcare data related to cohorts provided by the HSPs are categorized according to specific subjects and the characteristics of each cohort. Therefore, for each cohort, the data center demands and uses clinical data (EHR_{HSP}, $BEHR_{HSP}$), along with self-reported and behavioral data ($SELF\_BH_{HSP}$) from the HSPs.

Accordingly, the provided healthcare data are shared with the C-PMDC. In the United States, the NIH may assume this role.

- National institution curation resources (NICR)

It provides the collected healthcare data, including health examination data ($PBHE_{NICR}$) related to the cohort, to the C-PMDC. In the United States, the Food and Drug Administration (FDA) may assume this role.

- Insurance Institution (II)

It provides the collected healthcare data, including claims data (CLA_{II}) related to the cohort, to the C-PMDC. In the United States, the Center for Medicare and Medicaid Innovation (CMMI) and private insurance companies may adopt this role.

- Genetic data area

This information consists of entities that provide genetic information.

- National institution genetic resources (NIGR)

It provides genetic data (GEN_{NIGR}) related to the cohort to the precision medicine data center. In the United States, the National Cancer Institute (NCI) may assume this role.

- Lifelog data area

This area comprises an entity that provides the lifelog data generated by cohorts and environmental factors.

- Healthcare information technology (IT) companies

These provide lifelog data, including sensor information (SEN_{IT}) generated from mobiles and wearable devices of individuals in the cohorts, to the C-PMDC.

- Geo-spatial data center (GDC)

This provides lifelog data, including geo-spatial data (GEO_{GDC}) generated by environmental factors, to the C-PMDC. In the United States, the National Weather Service (NWS) may assume this role.

- Social network service (SNS) companies These provide lifelog data, including SNS data (SNS_{SNS}) generated from social media accounts (such as Twitter and Facebook) of individuals in the cohorts, to the C-PMDC.

- Data management and processing techniques

There are six core techniques for managing and processing data that are required to establish a precision medicine system; these include the IoT, cloud, big data, mobile, genetic analysis, and AI.

- Data security techniques

The data security techniques that must be implemented in a precision medicine system include standards, information security, data security policy principles and frameworks, and privacy policy principles and frameworks as defined by the Office of National Coordinator (ONC) for health information technology. Through privacy-related techniques, sensitive data types related to personal information, collected and processed in a precision medicine system, are strictly regulated.

Fig. 2 shows a type of sensitive data flow in a precision medicine system that is composed of defined entities.

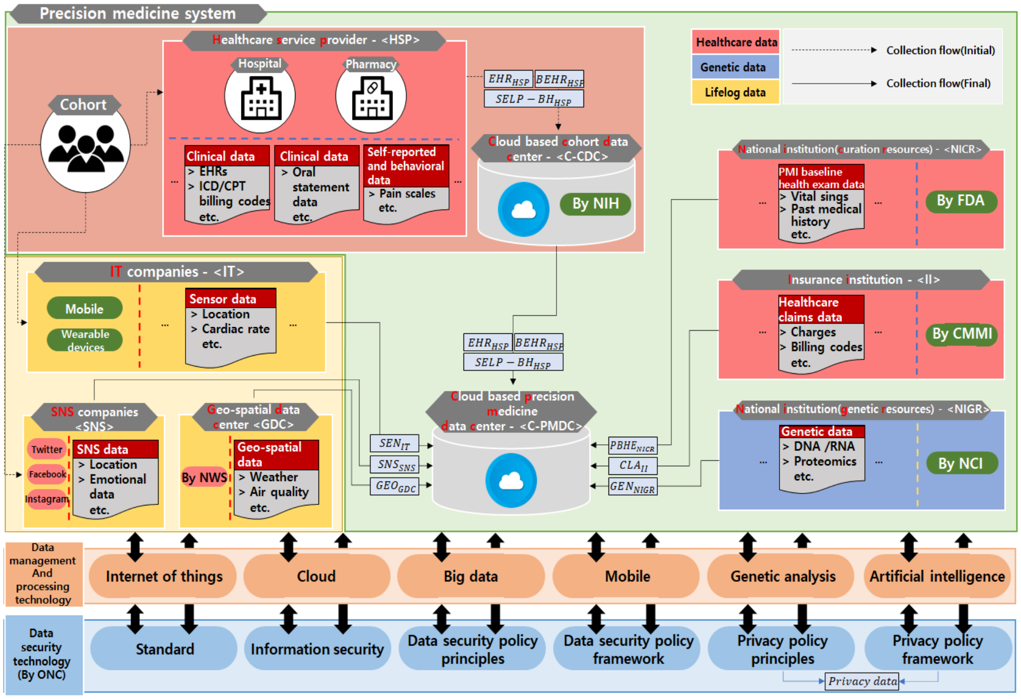

Figure 2: Proposed sensitive data flow in a precision medicine system

4.2 Reestablishment of sensitive data of a precision medicine system

This section presents the reestablishment of the data structure and content of the PMI project into four categories, namely healthcare, genetic, lifelog, and privacy data.

Fig. 3 illustrates the reestablished data for a precision medicine system.

- Healthcare data

Healthcare data are reestablished into clinical, self-reported and behavioral, PMI baseline health examination, and healthcare claims data.

- Clinical data

Clinical data consist of a dataset, referred to as EHR_{HSP}. They are reestablished into nine examples of EHRs, disease data, and medications. Another dataset, referred to as $BEHR_{HSP}$, is reestablished into two examples of oral statement and other images data. The data are provided by the HSPs and include the core data demanded by the C-CDC and subgroup data required for specific studies.

- Self-reported and behavioral data

These are reestablished into twelve examples, including pain scales, gender identity, alcohol, and smoking, and are referred to as $SELF-BH_{HSP}$. The data are provided by the HSPs and comprise core data demanded by the C-CDC and subgroup data required for specific studies.

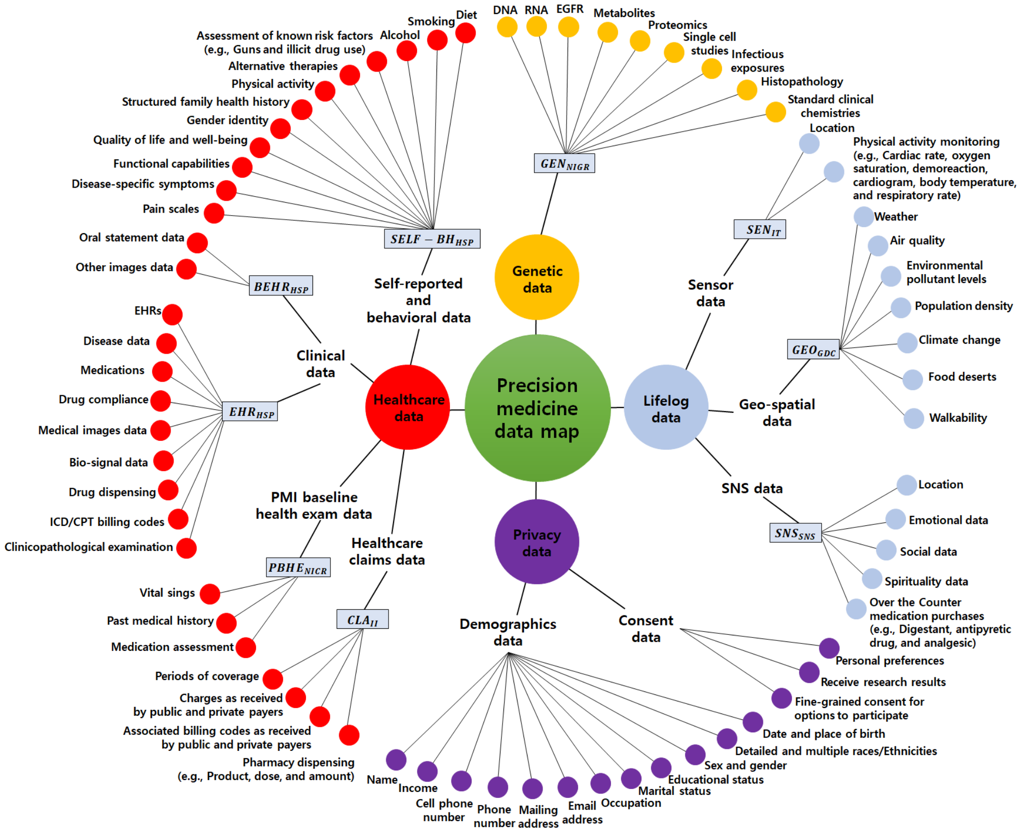

Figure 3: Precision medicine data map reestablished by the proposed architecture

- PMI baseline health exam data

Health examination data comprise a dataset reestablished into four examples including vital signs and past medical history. They are referred to as $PBHE_{NICR}$. The data are provided by the NICR providers and comprise core data demanded by the C-PMDC.

- Healthcare claims data

Claims data comprise a dataset that is reestablished into four examples, including periods of coverage, charges as received by public and private payers, and associated billing codes; they are referred to as CLA_{II}. The data are provided by the II and comprise core data demanded by the C-PMDC.

- Genetic data

Genetic data, referred to as $GEN_{NIGR,}$ comprise a dataset that is reestablished into nine examples including DNA, proteomics, and histopathology. These data are provided by the NIGR and comprise core data demanded by the C-PMDC.

- Lifelog data

Lifelog data comprise a dataset that is reestablished into sensor data, geo-spatial data, and SNS data.

- Sensor data

Sensor data, referred to as SEN_{IT}, comprise a dataset that is reestablished into two examples of location and physical activity monitoring. These data are provided by IT companies and comprise core data demanded by the C-PMDC and subgroup data necessary for specific research.

- Geo-spatial data

Geo-spatial data, referred to as GEO_{GDC}, comprise a dataset that is reestablished into seven examples including weather, air quality, and food desserts. These data are provided by the GDC and comprise core data demanded by the C-PMDC and subgroup data necessary for specific research.

- SNS data

SNS data, referred to as SNS_{SNS}, are reestablished into five examples including location, emotional, and spirituality data. These data are provided by SNS and comprise subgroup data necessary only for specific research demanded by the C-PMDC.

- Privacy data

Privacy data comprise a dataset that is reestablished into demographics and consent data. The demographics data are reestablished into twelve examples including the name, contact details, occupation, and race of the patient. The consent data are reestablished into three examples including fine-grained consent for options to participate and receive the research results. In the case of privacy data, data providers and demanders are not specified because they are collected and processed by all entities constituting the precision medicine system. Privacy data protection follows the data protection technique specified by the standard authority for medical information techniques.

4.3 Proposal of scheme for secure sharing of sensitive data in a precision medicine system

In this section, a scheme is proposed for securely sharing sensitive data in a precision medicine system. The entities that constitute the scheme are listed in Tab. 1. The scheme consists of three phases, namely the registration phase, authentication phase, and data transfer phase.

Table 1: Notations used in the scheme

Notation	Description
ID_X	Identity of an entity X
pw	Password of provider
N	Random nonce
x	Secret value of demander
y	Secret value of KSI server
$\rho, \sigma, \tau, \varphi, \omega$	Value of authentication for each entity
$H(\cdot)$	One-way hash function

$PRNG(\cdot)$	Pseudo random number generator
PU_X / PR_X	Encryption/decryption function of public key cryptosystem of an entity X
T	Time stamp
$\oplus$	Exclusive-OR operation
$\vert\vert$	Concatenation operation
$Data_X$	Sensitive data provided by entity X
$Signature(\cdot)$	Signature value
$Sign_{sk}$	Secret key of provider used to generate signature value
$KSI\_Sign(\cdot)$	KSI signature on signature(·) using KSI
$KSI\_Token(\cdot)$	Publication of token for including a path on the Merkle root by KSI server
C_{number}/D_{number}	Encrypted/decrypted message

4.3.1 Registration phase

Fig. 4 presents the proposed registration phase for performing authentication between the entities constituting the precision medicine system.

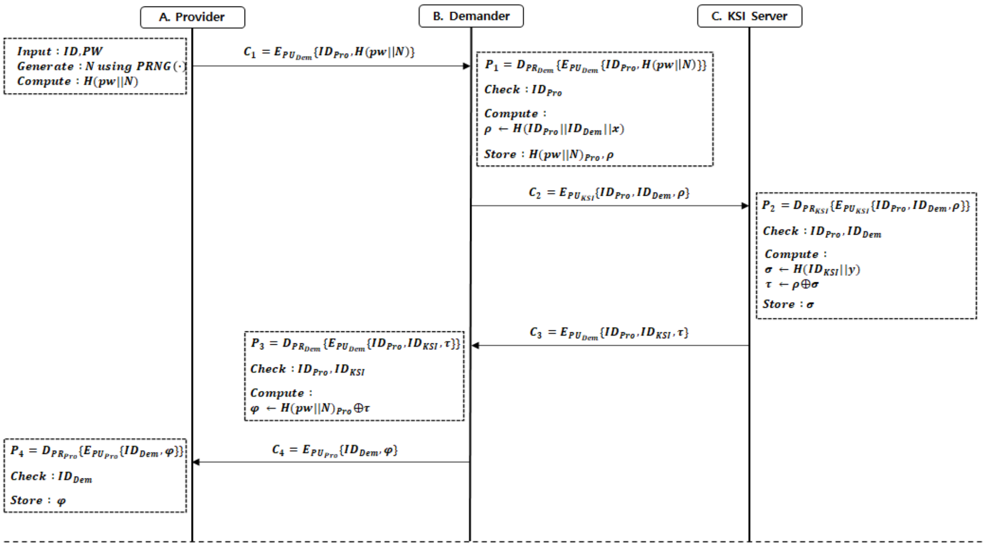
A. Provider
B. Demander
C. KSI Server
Input : ID,PW
Generate : N using PRNG(·)
Compute : H(pw||N)
$C_1 = E_{PU_{Dem}}\{ID_{Pro}, H(pw||N)\}$
$P_1 = D_{PR_{Dem}}\{E_{PU_{Dem}}\{ID_{Pro}, H(pw||N)\}\}$
Check : ID_{Pro}
Compute :
$\rho \leftarrow H(ID_{Pro}||ID_{Dem}||x)$
Store : $H(pw||N)_{Pro}, \rho$
$C_2 = E_{PU_{KSI}}\{ID_{Pro}, ID_{Dem}, \rho\}$
$P_2 = D_{PR_{KSI}}\{E_{PU_{KSI}}\{ID_{Pro}, ID_{Dem}, \rho\}\}$
Check : ID_{Pro}, ID_{Dem}
Compute :
$\sigma \leftarrow H(ID_{KSI}||y)$
$\tau \leftarrow \rho \oplus \sigma$
Store : σ
$C_3 = E_{PU_{Dem}}\{ID_{Pro}, ID_{KSI}, \tau\}$
$P_3 = D_{PR_{Dem}}\{E_{PU_{Dem}}\{ID_{Pro}, ID_{KSI}, \tau\}\}$
Check : ID_{Pro}, ID_{KSI}
Compute :
$\varphi \leftarrow H(pw||N)_{Pro} \oplus \tau$
$C_4 = E_{PU_{Pro}}\{ID_{Dem}, \varphi\}$
$P_4 = D_{PR_{Pro}}\{E_{PU_{Pro}}\{ID_{Dem}, \varphi\}\}$
Check : ID_{Dem}
Store : φ

Figure 4: Registration phase in the scheme

Step 1: <*Provider* sends C_1 to *Demander*>

The *Provider* enters ID, pw and generates a random nonce N with $PRNG(\cdot)$. Accordingly, $H(pw||N)$ is calculated, and $C_1 = E_{PU_{Dem}}\{ID_{Pro}, H(pw||N)\}$, encrypted with the Demander's public key, is sent to the *Demander*, along with the Provider's identity ID_{Pro}.

$$\{Input : Provider\ input\ to\ \boldsymbol{ID/pw}$$

$$\left\{\begin{array}{l} Generate : Random\ nonce\ \boldsymbol{N}\ using\ \boldsymbol{PRNG}(\cdot) \\ Compute : \boldsymbol{H(pw||N)}\ using\ Provider's\ \mathbf{pw}\ and\ random\ nonce\ \boldsymbol{N} \\ Encrypt : \boldsymbol{E_{PU_{Dem}}\{ID_{Pro}, H(pw||N)\}}\ with\ \boldsymbol{ID_{Pro}, H(pw||N)}\ using\ \boldsymbol{PU_{Dem}} \end{array}\right.$$

$$\boldsymbol{C_1} = E_{PU_{Dem}}\{ID_{Pro}, H(pw||N)\} \tag{1}$$

$$\{Send : \boldsymbol{C_1}$$

Step 2: <*Demander* checks P_1 and sends C_2 to *KSI server*>

The *Demander* obtains plain text P_1 by running $D_{PR_{Dem}}\{E_{PU_{Dem}}\{ID_{Pro}, H(pw||N)\}\}$ and decrypting cipher text C_1 received from the *Provider* with its private key. The identity ID_{Pro} of the *Provider* is then checked, and ρ, which is used as the authentication value, is calculated, as shown below, using its own identity ID_{Dem} and secret value x.

$$\rho = H(ID_{Pro}||ID_{Dem}||x) \tag{2}$$

Then, $H(pw||N)_{Pro}$, which is obtained by decrypting the cipher text of the *Provider*, and the authentication value ρ are stored. Next, the *Demander* sends $C_2 = E_{PU_{KSI}}\{ID_{Pro}, ID_{Dem}, \rho\}$, which denotes the encryption of the Provider's identity ID_{Pro}, Demander's own identity ID_{Dem}, and the authentication value ρ with the public key of *KSI server*, to the *KSI server*.

$$\{Decrypt : \boldsymbol{D_{PR_{Dem}}\{E_{PU_{Dem}}\{ID_{Pro}, H(pw||N)\}}\ using\ \boldsymbol{PR_{Dem}}$$

$$\boldsymbol{P_1} = D_{PR_{Dem}}\{C_1\} \tag{3}$$

$$\left\{\begin{array}{l} Check : Provider's\ identity\ \boldsymbol{ID_{Pro}} \\ Compute : \boldsymbol{\rho \leftarrow H(ID_{Pro}||ID_{Dem}||x)}\ using\ \boldsymbol{ID_{Pro}, ID_{Dem}}\ and\ Demander's\ \boldsymbol{x} \\ Store : \boldsymbol{H(pw||N)_{Pro}, \rho} \\ Encrypt : \boldsymbol{E_{PU_{KSI}}\{ID_{Pro}, ID_{Dem}, \rho\}}with\ \boldsymbol{ID_{pro}, ID_{Dem}, \rho}\ using\ \boldsymbol{PU_{KSI}} \end{array}\right.$$

$$\boldsymbol{C_2} = E_{PU_{KSI}}\{ID_{Pro}, ID_{Dem}, \rho\} \tag{4}$$

$$\{Send : \boldsymbol{C_2}$$

Step 3: <*KSI server* checks P_2, and sends C_3 to *Demander*>

The *KSI server* obtains plain text P_2 by executing $\mathrm{D_{PR_{KSI}}\{E_{PU_{KSI}}\{ID_{Pro}, ID_{Dem}, \rho\}\}}$ and decrypting cipher text C_2 sent from the *Demander* with its private key. The identity of *Demander* ID_{Dem} is then checked, and σ, which is used as the authentication value, is calculated using its own identity ID_{KSI} and secret value y. τ, which is used as the authentication value, is calculated as shown below by using ρ, which is obtained by decrypting the Demander's cipher text.

$$\sigma = H(ID_{KSI}||y) \tag{5}$$

$$\tau = \rho \oplus \sigma \tag{6}$$

The calculated σ is then stored, and $C_3 = E_{PU_{Dem}}\{ID_{Pro}, ID_{KSI}, \tau\}$, which is the encryption of the Provider's identity ID_{Pro}, KSI's own identity ID_{KSI}, and the authentication value τ with the public key of the *Demander*, is sent to the *Demander*.

$$\{Decrypt : \boldsymbol{D_{PR_{KSI}}\{E_{PU_{KSI}}\{ID_{Pro}, ID_{Dem}, \rho\}\}}\ using\ \boldsymbol{PR_{KSI}}$$

$$\boldsymbol{P_2} = D_{PR_{KSI}}\{C_2\} \tag{7}$$

$$\begin{cases} Check: Provider's\ identity\ \boldsymbol{ID_{Pro}}\ and\ Demander's\ identity\ \boldsymbol{ID_{Dem}} \\ \begin{cases} Compute: \boldsymbol{\sigma} \leftarrow \boldsymbol{H(ID_{KSI}||y)}\ using\ \boldsymbol{ID_{KSI}}\ and\ KSI\ server's\ \boldsymbol{y} \\ \qquad\qquad\ \boldsymbol{\tau} \leftarrow \boldsymbol{\rho \oplus \sigma}\ using\ authentication\ value\ \boldsymbol{\rho, \sigma} \end{cases} \\ Store: \boldsymbol{\sigma} \\ Encrypt: \boldsymbol{E_{PU_{Dem}}\{ID_{Pro}, ID_{KSI}, \tau\}}\ with\ \boldsymbol{ID_{Pro}, ID_{KSI}, \tau}\ using\ \boldsymbol{PU_{Dem}} \end{cases}$$

$$\boldsymbol{C_3} = E_{PU_{Dem}}\{ID_{Pro}, ID_{KSI}, \tau\} \tag{8}$$

$\{Send : \boldsymbol{C_3}$

Step 4: <*Demander* checks P_3 and sends C_4 to *Provider*>

The *Demander* obtains plain text P_3 by executing $D_{PR_{Dem}}\{E_{PU_{Dem}}\{ID_{Pro}, ID_{KSI}, \tau\}\}$ and decrypting cipher text C_3 sent from the *KSI server* with its private key. The identity of *Provider,* ID_{Pro}, and that of the *KSI server,* $\mathrm{ID_{KSI}}$, are then checked. Next, φ, which is used as the authentication value, is calculated as shown below using τ that is obtained by decrypting $H(pw||N)_{Pro}$ stored in Step 2 and the KSI server's cipher text.

$$\varphi = H(pw||N)_{Pro} \oplus \tau \tag{9}$$

Then, $C_4 = E_{PU_{Pro}}\{ID_{Dem}, \varphi\}$, which denotes the encryption of the Demander's own identity $\mathrm{ID_{Dem}}$ and authentication value φ with the public key of the *Provider*, is sent to the *Provider*.

$\{Decrypt : \boldsymbol{D_{PR_{Dem}}\{E_{PU_{Dem}}\{ID_{Pro}, ID_{KSI}, \tau\}\}}\ using\ \boldsymbol{PR_{Dem}}$

$$\boldsymbol{P_3} = D_{PR_{Dem}}\{C_3\} \tag{10}$$

$$\begin{cases} Check: Provider's\ identity\ \boldsymbol{ID_{Pro}}\ and\ KSI\ server's\ identity\ \boldsymbol{ID_{KSI}} \\ Compute: \ \boldsymbol{\varphi} \leftarrow \boldsymbol{H(pw||N)_{Pro} \oplus \tau} \\ \qquad\qquad\qquad using\ \boldsymbol{H(pw||N)_{Pro}}, authentication\ value\ \boldsymbol{\tau} \\ Encrypt: \boldsymbol{E_{PU_{Pro}}\{ID_{Dem}, \varphi\}}\ with\ \boldsymbol{ID_{Dem}, \varphi}\ using\ \boldsymbol{PU_{Pro}} \end{cases}$$

$$\boldsymbol{C_4} = E_{PU_{Pro}}\{ID_{Dem}, \varphi\} \tag{11}$$

$\{Send : \boldsymbol{C_4}$

Step 5: <*Provider* checks $\mathrm{P_4}$ and the end of the registration phase>

The *Provider* obtains plain text P_4 by executing $D_{PR_{Pro}}\{E_{PU_{Pro}}\{ID_{Dem}, \varphi\}\}$ and decrypting cipher text $\mathrm{C_4}$ sent by the *Demander* with its private key. The Demander's identity $\mathrm{ID_{Dem}}$ is then checked, and the authentication value φ is stored to end the registration phase.

$\{Decrypt : \boldsymbol{D_{PR_{Pro}}\{E_{PU_{Pro}}\{ID_{Dem}, \varphi\}\}}\ using\ \boldsymbol{PR_{Pro}}$

$$\boldsymbol{P_4} = D_{PR_{Pro}}\{C_4\} \tag{12}$$

$$\begin{cases} Check: Demander's\ identity\ \boldsymbol{ID_{Dem}} \\ Store: \ \boldsymbol{\varphi} \end{cases}$$

4.3.2 Authentication phase

Fig. 5 presents the proposed authentication phase to perform authentication between the entities constituting the precision medicine system.

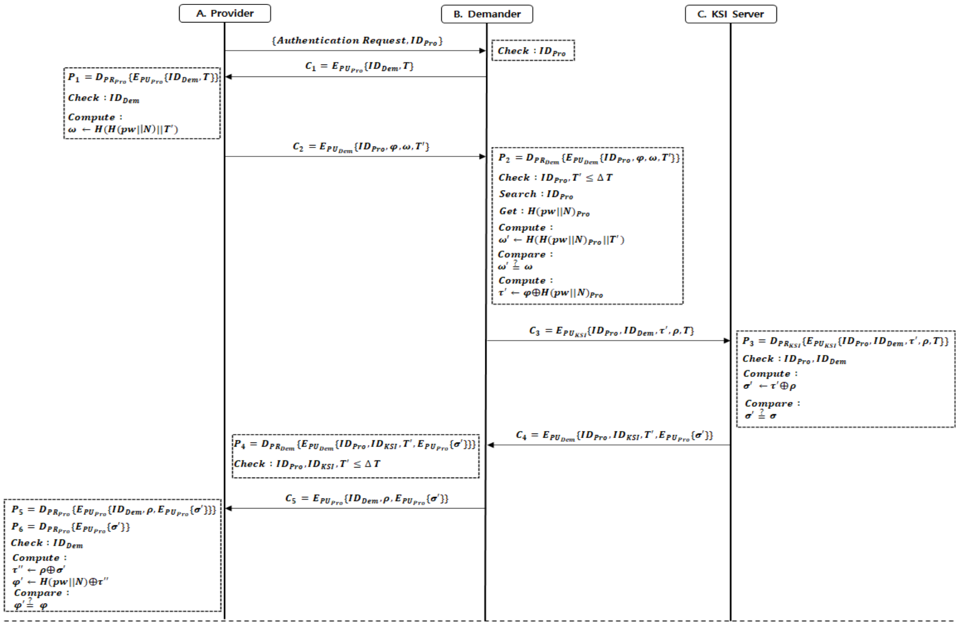
A. Provider
B. Demander
C. KSI Server
$\{Authentication\ Request, ID_{Pro}\}$
$Check : ID_{Pro}$
$C_1 = E_{PU_{Pro}}\{ID_{Dem}, T\}$
$P_1 = D_{PR_{Pro}}\{E_{PU_{Pro}}\{ID_{Dem}, T\}\}$
$Check : ID_{Dem}$
$Compute :$
$\omega \leftarrow H(H(pw||N)||T')$
$C_2 = E_{PU_{Dem}}\{ID_{Pro}, \varphi, \omega, T'\}$
$P_2 = D_{PR_{Dem}}\{E_{PU_{Dem}}\{ID_{Pro}, \varphi, \omega, T'\}\}$
$Check : ID_{Pro}, T' \leq \Delta T$
$Search : ID_{Pro}$
$Get : H(pw||N)_{Pro}$
$Compute :$
$\omega' \leftarrow H(H(pw||N)_{Pro}||T')$
$Compare :$
$\omega' \overset{?}{=} \omega$
$Compute :$
$\tau' \leftarrow \varphi \oplus H(pw||N)_{Pro}$
$C_3 = E_{PU_{KSI}}\{ID_{Pro}, ID_{Dem}, \tau', \rho, T\}$
$P_3 = D_{PR_{KSI}}\{E_{PU_{KSI}}\{ID_{Pro}, ID_{Dem}, \tau', \rho, T\}\}$
$Check : ID_{Pro}, ID_{Dem}$
$Compute :$
$\sigma' \leftarrow \tau' \oplus \rho$
$Compare :$
$\sigma' \overset{?}{=} \sigma$
$C_4 = E_{PU_{Dem}}\{ID_{Pro}, ID_{KSI}, T', E_{PU_{Pro}}\{\sigma'\}\}$
$P_4 = D_{PR_{Dem}}\{E_{PU_{Dem}}\{ID_{Pro}, ID_{KSI}, T', E_{PU_{Pro}}\{\sigma'\}\}\}$
$Check : ID_{Pro}, ID_{KSI}, T' \leq \Delta T$
$C_5 = E_{PU_{Pro}}\{ID_{Dem}, \rho, E_{PU_{Pro}}\{\sigma'\}\}$
$P_5 = D_{PR_{Pro}}\{E_{PU_{Pro}}\{ID_{Dem}, \rho, E_{PU_{Pro}}\{\sigma'\}\}\}$
$P_6 = D_{PR_{Pro}}\{E_{PU_{Pro}}\{\sigma'\}\}$
$Check : ID_{Dem}$
$Compute :$
$\tau'' \leftarrow \rho \oplus \sigma'$
$\varphi' \leftarrow H(pw||N) \oplus \tau''$
$Compare :$
$\varphi' \overset{?}{=} \varphi$

Figure 5: Authentication phase in the scheme

Step 1: <*Provider* sends $authentication\ request$ message to *Demander*>

The *Provider* sends its own identity $\mathrm{ID_{Pro}}$ and a request message $authentication\ Request$ to the *Demander* to request for authentication.

$$\begin{cases} Request : Provider\ \boldsymbol{authentication\ request}\ to\ Demander \\ Send : Provider's\ identity\ \boldsymbol{ID_{Pro}} \end{cases}$$

Step 2: <*Demander* checks the identity and sends $\mathrm{C_1}$ to *Provider*>

The *Demander* checks the received Provider's identity $\mathrm{ID_{Pro}}$ and sends $C_1 = E_{PU_{Pro}}\{ID_{Dem}, T\}$, which denotes the encryption of its identity ID_{Dem} and time stamp T with the Provider's public key, to the *Provider*.

$$\begin{cases} Check : Provider's\ identity\ \boldsymbol{ID_{Pro}} \\ Encrypt : \boldsymbol{E_{PU_{Pro}}}\{\boldsymbol{ID_{Dem}}, \boldsymbol{T}\} with\ \boldsymbol{ID_{Dem}}\ and\ Timestmap\ \boldsymbol{T}\ using\ \boldsymbol{PU_{Pro}} \end{cases}$$

$$\boldsymbol{C_1} = E_{PU_{Pro}}\{ID_{Dem}, T\} \tag{13}$$

$$\{Send : \boldsymbol{C_1}$$

Step 3: <*Provider* checks P_1 and sends C_2 to *Demander*>

The *Provider* obtains plain text $\mathrm{P_1}$ by executing $D_{PR_{Pro}}\{E_{PU_{Pro}}\{ID_{Dem}, T\}\}$ and decrypting cipher text $\mathrm{C_1}$ sent from the *Demander* with its private key. Then, the Demander's identity $\mathrm{ID_{Dem}}$ is checked, and using $H(pw||N)$, calculated in Step 1 of the registration phase, and time stamp value T', authentication value ω is calculated as follows:

$$\omega = H(H(pw||N)||T') \tag{14}$$

Then, $C_2 = E_{PU_{Dem}}\{ID_{Pro}, \varphi, \omega, T'\}$, which denotes the encryption of the Provider's identity $\mathrm{ID_{Pro}}$, authentication value φ stored in Step 5 of the registration phase, authentication value ω, and time stamp value T' calculated in this phase with the public key of the *Demander,* is sent to the *Demander*.

$$\{Decrypt : \boldsymbol{D_{PR_{Pro}}}\{\boldsymbol{E_{PU_{Pro}}}\{\boldsymbol{ID_{Dem}}, \boldsymbol{T}\}\ using\ \boldsymbol{PR_{Pro}}$$

$$\boldsymbol{P_1} = D_{PR_{Pro}}\{C_1\} \tag{15}$$

$$\begin{cases} Check : Demander's\ identity\ \boldsymbol{ID_{Dem}} \\ Compute : \boldsymbol{\omega} \leftarrow \boldsymbol{H}(\boldsymbol{H}(\boldsymbol{pw}||\boldsymbol{N})||\boldsymbol{T''})\ using\ \boldsymbol{H}(\boldsymbol{pw}||\boldsymbol{N}) \\ \qquad\qquad\qquad\qquad\qquad\qquad and\ updated\ timestamp\ \boldsymbol{T'} \\ Encrypt : \boldsymbol{E_{PU_{Dem}}}\{\boldsymbol{ID_{Pro}}, \boldsymbol{\varphi}, \boldsymbol{\omega}, \boldsymbol{T'}\}\ with\ \boldsymbol{ID_{pro}}, \boldsymbol{\varphi}, \boldsymbol{\omega}, \boldsymbol{T'}\ using\ \boldsymbol{PU_{Dem}} \end{cases}$$

$$\boldsymbol{C_2} = E_{PU_{Dem}}\{ID_{Pro}, \varphi, \omega, T'\} \tag{16}$$

$$\{Send : \boldsymbol{C_2}$$

Step 4: <*Demander* checks $\mathrm{P_2}$ and sends C_3 to *KSI server*>

The *Demander* obtains plain text $\mathrm{P_2}$ by executing $D_{PR_{Dem}}\{E_{PU_{Dem}}\{ID_{Pro}, \varphi, \omega, T'\}\}$ and decrypting cipher text $\mathrm{C_2}$ sent from the *Provider* with its private key. Then, the Provider's identity $\mathrm{ID_{Pro}}$ and validity of time stamp value T' are checked. A search on $\mathrm{ID_{Pro}}$ is performed to obtain $H(pw||N)_{Pro}$, which was stored in Step 2 of the registration phase. Using this, the verification value ω' is calculated as follows:

$$\omega' = H(H(pw||N)_{Pro}||T') \tag{17}$$

The calculated ω′ and ω obtained by decrypting the Provider's cipher text are compared for verification. The verification value τ' is then calculated, as shown below, using the Provider's authentication value φ and $H(pw||N)_{Pro}$.

$$\tau' = \varphi \oplus H(pw||N)_{Pro} \tag{18}$$

Then, $C_3 = E_{PU_{KSI}}\{ID_{Pro}, ID_{Dem}, \tau', \rho, T\}$, which denotes the encryption of the Provider's identity $\mathrm{ID_{Pro}}$, Demander's identity $\mathrm{ID_{Dem}}$, calculated verification value τ', authentication value ρ stored in Step 2 of the registration phase, and time stamp T with the public key of the *KSI Server,* is sent to the *KSI Server.*

$$\{Decrypt : \boldsymbol{D_{PR_{Dem}}}\{\boldsymbol{E_{PU_{Dem}}}\{\boldsymbol{ID_{Pro}}, \boldsymbol{\varphi}, \boldsymbol{\omega}, \boldsymbol{T'}\}\}\ using\ \boldsymbol{PR_{Dem}}$$

$$\boldsymbol{P_2} = D_{PR_{Dem}}\{C_2\} \tag{19}$$

$$\begin{cases} Check : Provider's\ identity\ \boldsymbol{ID_{Pro}}\ and\ Timestamp\ validation\ \boldsymbol{T'} \leq \Delta\boldsymbol{T} \\ Search : \boldsymbol{ID_{Pro}}\ to\ get\ H(pw||N)_{Pro} \\ Get : \boldsymbol{H}(\boldsymbol{pw}||\boldsymbol{N})_{\boldsymbol{Pro}} \\ Compute : \boldsymbol{\omega'} \leftarrow \boldsymbol{H}(\boldsymbol{H}(\boldsymbol{pw}||\boldsymbol{N})_{\boldsymbol{Pro}}||\boldsymbol{T'})\ using\ \boldsymbol{H}(\boldsymbol{pw}||\boldsymbol{N})_{\boldsymbol{Pro}}, Timestamp\ \boldsymbol{T'} \\ Compare : \boldsymbol{\omega'}\ ?= \boldsymbol{\omega} \\ Compute : \boldsymbol{\tau'} \leftarrow \boldsymbol{\varphi} \oplus \boldsymbol{H}(\boldsymbol{pw}||\boldsymbol{N})_{\boldsymbol{Pro}}\ using\ authentication\ value\ \boldsymbol{\varphi}, \\ \qquad\qquad\qquad\qquad\qquad\qquad and\ \boldsymbol{H}(\boldsymbol{pw}||\boldsymbol{N})_{\boldsymbol{Pro}} \end{cases}$$

$$\begin{cases} Encrypt : \boldsymbol{E_{PU_{KSI}}}\{\boldsymbol{ID_{Pro}}, \boldsymbol{ID_{Dem}}, \boldsymbol{\tau'}, \boldsymbol{\rho}, \boldsymbol{T}\} \\ \qquad\qquad with\ \boldsymbol{ID_{Pro}}, \boldsymbol{ID_{Dem}}, \boldsymbol{\tau'}, \boldsymbol{\rho}, \boldsymbol{T}\ using\ \boldsymbol{PU_{KSI}} \end{cases}$$

$$\boldsymbol{C_3} = E_{PU_{KSI}}\{ID_{Pro}, ID_{Dem}, \tau', \rho, T\} \tag{20}$$

$$\{Send : \boldsymbol{C_3}$$

Step 5: <*KSI server* checks P_3 and sends C_4 to *Demander*>

The *KSI server* obtains plain text P_3 by executing $D_{PR_{KSI}}\{E_{PU_{KSI}}\{ID_{Pro}, ID_{Dem}, \tau', \rho, T\}\}$ and decrypting cipher text C_3 sent from the *Demander* with its private key. The Provider's identity ID_{Pro} and Demander's identity ID_{Dem} are then checked. Using the verification value τ' and authentication value ρ obtained through decryption, the verification value σ' is calculated as follows:

$$\sigma' = \tau' \oplus \rho \tag{21}$$

The calculated σ' and authentication value σ stored in Step 3 of the registration phase are compared for verification. Then, $C_4 = E_{PU_{Dem}}\{ID_{Pro}, ID_{KSI}, T', E_{PU_{Pro}}\{\sigma'\}\}$, which denotes the encryption of the Provider's identity ID_{Pro}, KSI's identity ID_{KSI}, time stamp value T', and $E_{PU_{Pro}}\{\sigma'\}$, the encryption of the verification value σ' with Provider's public key, with the Demander's public key, is sent to the *Demander*.

$$\{Decrypt : \boldsymbol{D_{PR_{KSI}}\{E_{PU_{KSI}}\{ID_{Pro}, ID_{Dem}, \tau', \rho, T\}\}}\ using\ \boldsymbol{PR_{KSI}}$$

$$\boldsymbol{P_3} = D_{PR_{KSI}}\{C_3\} \tag{22}$$

$$\left\{\begin{array}{l} Check : Provider's\ identity\ \boldsymbol{ID_{Pro}}\ and\ Demander's\ identity\ \boldsymbol{ID_{Dem}} \\ Compute :\ \boldsymbol{\sigma' \leftarrow \tau' \oplus \rho}\ using\ authentication\ value\ \boldsymbol{\tau', \rho} \end{array}\right.$$

$$\left\{\begin{array}{l} Compare :\ \boldsymbol{\sigma'\ ?= \sigma} \\ Encrypt : \boldsymbol{E_{PU_{Dem}}\left\{ID_{Pro}, ID_{KSI}, T', E_{PU_{Pro}}\{\sigma'\}\right\}} \\ \qquad\qquad with\ \boldsymbol{ID_{Pro}, ID_{KSI}, T', E_{PU_{Pro}}\{\sigma'\}}\ using\ \boldsymbol{PU_{Dem}} \\ \qquad\qquad\qquad \boldsymbol{C_4} = E_{PU_{Dem}}\{ID_{Pro}, ID_{KSI}, T', E_{PU_{Pro}}\{\boldsymbol{\sigma'}\}\} \end{array}\right.$$

$$\boldsymbol{C_4} = E_{PU_{Dem}}\{ID_{Pro}, ID_{KSI}, T', E_{PU_{Pro}}\{\boldsymbol{\sigma'}\}\} \tag{23}$$

$$\{Send : \boldsymbol{C_4}$$

Step 6: <*Demander* checks P_4 and sends C_5 to *Provider*>

The *Demander* obtains plain text P_4 by executing $D_{PR_{Dem}}\{E_{PU_{Dem}}\{ID_{Pro}, ID_{KSI}, T', E_{PU_{Pro}}\{\sigma'\}\}\}$ and decrypting cipher text C_4 sent from the *KSI server* with its private key. The Provider's identity ID_{Pro}, KSI server's identity ID_{KSI} , and validity of the time stamp value T' are then checked, and $C_5 = E_{PU_{Pro}}\{ID_{Dem}, \rho, E_{PU_{Pro}}\{\sigma'\}\}$, which denotes the encryption of the Demander' s identity ID_{Dem}, authentication value ρ, and cipher text $E_{PU_{Pro}}\{\sigma'\}$ with the Provider's public key, is sent to the *Provider*.

$$\{Decrypt : \boldsymbol{D_{PR_{Dem}}\{E_{PU_{Dem}}\{ID_{Pro}, ID_{KSI}, T', E_{PU_{Pro}}\{\sigma'\}\}\}}\ using\ \boldsymbol{PR_{Dem}}$$

$$\boldsymbol{P_4} = D_{PR_{Dem}}\{C_4\} \tag{24}$$

$$\begin{cases} \mathit{Check} : \mathit{Provider's\ identity}\ \boldsymbol{ID_{Pro}}, \mathit{KSI\ server'sidentity}\ \boldsymbol{ID_{KSI}}, \\ \qquad \mathit{and\ Timestamp\ validation}\ \boldsymbol{T'} \leq \Delta\boldsymbol{T} \\ \mathit{Encrypt} : \boldsymbol{E_{PU_{Pro}}}\left\{\boldsymbol{ID_{Dem}}, \boldsymbol{\rho}, \boldsymbol{E_{PU_{Pro}}}\{\boldsymbol{\sigma'}\}\right\} \\ \qquad \mathit{with}\ \boldsymbol{ID_{Dem}}, \boldsymbol{\rho}, \boldsymbol{E_{PU_{Pro}}}\{\boldsymbol{\sigma'}\}\ \mathit{using}\ \boldsymbol{PU_{Pro}} \end{cases}$$

$$\boldsymbol{C_5} = E_{PU_{Pro}}\{ID_{Dem}, \rho, E_{PU_{Pro}}\{\sigma'\}\} \tag{25}$$

$\{\mathit{Send} : \boldsymbol{C_5}$

Step 7: <*Provider* checks P_5 and P_6, and the end of the authentication phase>

The *Provider* obtains plain texts P_5 and P_6 by executing $D_{PR_{Pro}}\{E_{PU_{Pro}}\{ID_{Dem}, \rho, E_{PU_{Pro}}\{\sigma'\}\}\}$ and $D_{PR_{Pro}}\{E_{PU_{Pro}}\{\sigma'\}\}$ and decrypting cipher text C_5 sent from the *Demander* with its private key. The Demander's identity $\mathrm{ID_{Dem}}$ is then checked, and using the authentication value ρ and verification value σ', the verification value τ'' is calculated as follows:

$$\tau'' = \rho \oplus \sigma' \tag{26}$$

Then, using $H(pw||N)$ and the calculated verification value τ'', the verification value φ' is calculated as follows:

$$\varphi' = H(pw||N) \oplus \tau'' \tag{27}$$

The calculated φ' and authentication value φ stored in Step 5 of the registration phase are compared for verification, and the authentication phase is ended.

$\{\mathit{Decrypt} : \boldsymbol{D_{PR_{Pro}}}\{\boldsymbol{E_{PU_{Pro}}}\{\boldsymbol{ID_{Dem}}, \boldsymbol{\rho}, \boldsymbol{E_{PU_{Pro}}}\{\boldsymbol{\sigma'}\}\}\}\ \mathit{using}\ \boldsymbol{PR_{Pro}}$

$$\boldsymbol{P_5} = D_{PR_{Pro}}\{C_5\} \tag{28}$$

$\{\mathit{Decrypt} : \boldsymbol{D_{PR_{Pro}}}\{\boldsymbol{E_{PU_{Pro}}}\{\boldsymbol{\sigma'}\}\}\ \mathit{using}\ \boldsymbol{PR_{Pro}}$

$$\boldsymbol{P_6} = D_{PR_{Pro}}\{E_{PU_{Pro}}\{\sigma'\}\} \tag{29}$$

$$\begin{cases} \mathit{Check} : \mathit{Demander's\ identity}\ \boldsymbol{ID_{Dem}} \\ \begin{cases} \mathit{Compute} : \boldsymbol{\tau''} \leftarrow \boldsymbol{\rho} \oplus \boldsymbol{\sigma'}\ \mathit{using\ authentication\ value}\ \boldsymbol{\rho}, \boldsymbol{\sigma'} \\ \quad \boldsymbol{\varphi'} \leftarrow H(pw||N) \oplus \boldsymbol{\tau''}\ \mathit{using}\ \boldsymbol{H(pw||N)} \\ \qquad\qquad \mathit{and\ authentication\ value}\ \boldsymbol{\tau''} \end{cases} \\ \mathit{Compare} : \boldsymbol{\varphi'}\ ?= \boldsymbol{\varphi} \end{cases}$$

4.3.3 Data transfer phase

Fig. 6 presents the proposed data transfer phase required to authenticate the entities that form a precision medicine system. The involved procedure is detailed below.

Step 1:

The *KSI server* access is provided to the *Demander* and *Provider*. In addition, the *Demander* access is provided to the *Provider*.

Step 2: <*Provider* sends KSI_Sign to *KSI server*>

For the KSI-based signature required to transfer the data value $Data_{Pro}$, the *Provider* uses a secret key sk for the personal signature and KSI to compute the following:

$$Signature(Data_{Pro}) = Sign_{sk}(H(Data_{pro})) \tag{30}$$

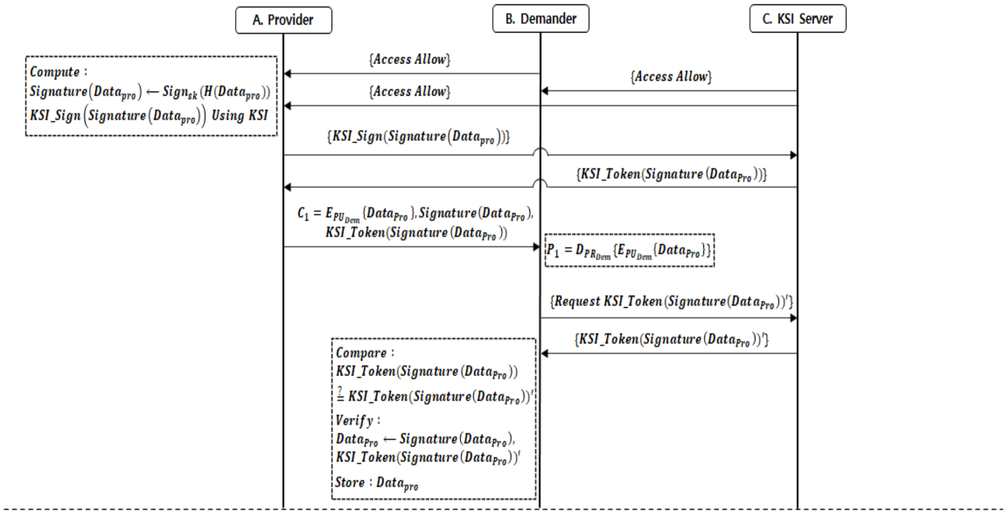
A. Provider
B. Demander
C. KSI Server
Compute :
$Signature(Data_{pro}) \leftarrow Sign_{sk}(H(Data_{pro}))$
$KSI\_Sign(Signature(Data_{pro}))$ Using KSI
{Access Allow}
{Access Allow}
{Access Allow}
$\{KSI\_Sign(Signature(Data_{pro}))\}$
$\{KSI\_Token(Signature(Data_{Pro}))\}$
$C_1 = E_{PU_{Dem}}\{Data_{Pro}\}, Signature(Data_{Pro}),$
$KSI\_Token(Signature(Data_{Pro}))$
$P_1 = D_{PR_{Dem}}\{E_{PU_{Dem}}\{Data_{Pro}\}\}$
$\{Request\ KSI\_Token(Signature(Data_{Pro}))'\}$
$\{KSI\_Token(Signature(Data_{Pro}))'\}$
Compare :
$KSI\_Token(Signature(Data_{Pro}))$
$\overset{?}{=} KSI\_Token(Signature(Data_{Pro}))'$
Verify :
$Data_{Pro} \leftarrow Signature(Data_{Pro}),$
$KSI\_Token(Signature(Data_{Pro}))'$
Store : $Data_{pro}$

Figure 6: Data transfer phase in the scheme

$\{KSI\_Sign\big(Signature(Data_{Pro})\big)\ Using\ KSI$

The computed KSI-based signature values $\{KSI\_Sign(Signature(Data_{Pro}))\}$ are then sent to the *KSI server*.

$$\left\{\left\{\begin{array}{l} Compute: \boldsymbol{Signature}(\boldsymbol{Data_{Pro}}) \leftarrow \boldsymbol{Sign_{sk}}(\boldsymbol{H}(\boldsymbol{Data_{Pro}}))\ using\ \boldsymbol{Data_{Pro}} \\ \qquad\qquad\qquad\qquad\qquad and\ Provider's\ \boldsymbol{sk} \\ \qquad\qquad \boldsymbol{KSI\_Sign}(\boldsymbol{Signature}(\boldsymbol{Data_{Pro}})) \leftarrow using\ \boldsymbol{KSI} \end{array}\right.\right.$$

$\{Send: \boldsymbol{KSI\_Sign}(\boldsymbol{Signature}(\boldsymbol{Data_{Pro}}))$

Step 3: <*KSI server* sends KSI_Token to *Provider*>

The *KSI server* publishes the token values $\{KSI\_Token(Signature(Data_{Pro}))\}$ including the Merkle root path for the KSI-based signature values $\{KSI\_Sign(Signature(Data_{Pro}))\}$ received from the *Provider* and sends them to the *Provider*.

$\{Send: \boldsymbol{KSI\_Token}(\boldsymbol{Signature}(\boldsymbol{Data_{Pro}}))$

Step 4: <*Provider* sends C_1, $data\ signature$, and KSI_Token to *Demander*>

The *Provider* sends $C_1 = E_{PU_{Dem}}\{Data_{Pro}\}$, which denotes the data value $Data_{Pro}$ encrypted with the Demander's public key, the data signature value $Signature(Data_{Pro})$, and token value $KSI\_Token(Signature((Data_{Pro}))$ used for the signature verification to the *Demander*.

$\{Encrypt: \boldsymbol{E_{PU_{Dem}}}\{\boldsymbol{Data_{Pro}}\}\ with\ \boldsymbol{Data_{Pro}}\ using\ \boldsymbol{PU_{Dem}}$

$$\boldsymbol{C_1} = E_{PU_{Dem}}\{Data_{Pro}\} \tag{31}$$

$\{Send: \boldsymbol{C_1}, \boldsymbol{Signature}(\boldsymbol{Data_{Pro}}), \boldsymbol{KSI\_Token}(\boldsymbol{Signature}(\boldsymbol{data_{Pro}}))$

Step 5: <*Demander* checks P_1 and sends $KSI\_Token\ request$ message to *KSI server*>

The *Demander* obtains plain text P_1 by executing $D_{PR_{Dem}}\{E_{PU_{Dem}}\{Data_{Pro}\}\}$ and decrypting cipher text C_1 received from the *Provider* with its private key. Then, for integrity verification, it sends a message $\{Request\ KSI\_Token(Signature((Data\_Pro\ ))'\}$ requesting a token value from the *KSI server* before storing the data.

$$\{Decrypt : \boldsymbol{D_{PR_{Dem}}}\{\boldsymbol{E_{PU_{Dem}}}\{\boldsymbol{Data_{Pro}}\}\}\ using\ \boldsymbol{PR_{Dem}}$$

$$\boldsymbol{P_1} = D_{PU_{Dem}}\{C_1\} \tag{32}$$

$$\{Request : Demander\ \boldsymbol{KSI\_Token}(\boldsymbol{Signature}(\boldsymbol{Data_{Pro}}))''\ to\ KSI\ server$$

Step 6: <*KSI server* sends KSI_Token to *Demander*>

The *KSI server* sends the token value $\{KSI\_Token(Signature((Data_{Pro}))'\}$ to the *Demander* in reply to the message received from the *Demander*.

$$\{Send : \boldsymbol{KSI\_Token}(\boldsymbol{Signature}(\boldsymbol{Data_{Pro}}))''$$

Step 7: <*Demander* verifies $Data_{Pro}$ integrity, and the end of Data Transfer phase>

The *Demander* compares and verifies the token value $KSI\_Token(Signature((Data_{Pro}))$ received from the *Provider* in Step 4 and the token value $KSI\_Token(Signature((Data_{Pro}))'$ received from the *KSI server* in Step 6. It then verifies the data value $Data_{Pro}$ using the signature value $Signature(Data_{Pro})$ and token value $KSI\_Token(Signature((Data_{Pro}))'$. After a valid verification, the data value $Data_{Pro}$ is stored, and the data transfer phase is terminated.

$$\begin{cases} Compare : \boldsymbol{KSI\_Token}(\boldsymbol{Signature}(\boldsymbol{Data_{Pro}}))' \\ \qquad\qquad ?= \boldsymbol{KSI\_Token}(\boldsymbol{Signature}(\boldsymbol{Data_{Pro}})) \\ Verify : \boldsymbol{Data_{Pro}} \leftarrow \boldsymbol{Signature}(\boldsymbol{Data_{Pro}}), \\ \qquad\qquad\qquad\qquad \boldsymbol{KSI\_Token}(\boldsymbol{Signature}(\boldsymbol{Data_{Pro}}))' \\ Store : \boldsymbol{Data_{Pro}} \end{cases}$$

4.4 Security analysis

In this section, the analysis conducted in Section 2.2 to determine whether the proposed scheme meets the security requirements for each infringement threat to the sensitive data of the precision medicine system is verified.

- Data exposure (Threat 1)-Data confidentiality guarantee (Requirement 1)

The data sent or received between the entities during the authentication or data-transfer phase can be exposed through a sniffing attack by an attacker. To prevent such attacks, the proposed scheme applies public key cryptography that uses $\{PU_{Entity}, PR_{Entity}\}$ on each instance of data sent or received between entities during the authentication or data-transfer phase. Hence, the data sent or received represent a ciphertext $\{C_X = E_{PU_{Entity}}\{Data\}\}$. As only the valid entities encrypt and decrypt data for the authentication or data-transfer phase, data confidentiality is guaranteed.

- Data forgery and falsification (Threat 2)-Data integrity verification (Requirement 2)

Data can be forged or falsified by a forgery or falsification attack on the data sent and received between entities during the data-transfer phase or owing to errors that occur during the data communication process. This leads to reliability issues with respect to the data collected from the precision medicine system. As a preventive measure, the proposed scheme utilizes authentication values $\{\rho, \sigma, \tau, \varphi, \omega\}$ during the registration and authentication phases. An attacker cannot compute all the authentication values. If the data are forged or falsified, such values are not verified during the authentication phase, thereby guaranteeing data integrity. Furthermore, the demander leverages the value $\{Signature(Data_{Pro})\}$ signed by the provider and the signature token value $\{KSI\_Token(Signature(Data_{Pro}))\}$ for the corresponding signature value received from the KSI server on the data $\{E_{PU_{Dem}}\{Data_{Pro}\}\}$ sent by the provider during the data transfer phase. A hash calendar, which contains the signature value $\{Signature(Data_{Pro})\}$, is determined by the signature token value $\{KSI\_Token(Signature(Data_{Pro}))\}$. Given the corresponding signature value $\{s\}$ and a node value $\{u\}$ at the same level on the calendar, $\{s \rightarrow s_1 = h(u_1||\mathrm{s}) \rightarrow s_2 = h(s_1||u_{34}) \rightarrow \cdots \rightarrow s_{root}\}$ is performed according to the Merkle hash tree. When the s_{root} obtained in this process is verified to be valid, the integrity of signature value s is also verified. Therefore, data forgery or falsification threats can be prevented by verifying the integrity of the data $\{Data_{Pro}\}$ sent by the provider.

- Unauthorized entities (Threat 3) – Entity authentication (Requirement 3)

If data are shared by an unauthorized entity, reliability issues arise with respect to the data collected from the precision medicine system and infringement of the availability of precision medicine services. As a preventive measure, the proposed scheme performs mutual authentication in the registration and authentication phases.

During the registration phase, an entity is registered using the authentication values $\{\sigma\}$ generated with nonce values $\{N\}$ that are known only to the provider, authentication values $\{\rho\}$ generated by secret values $\{x\}$ that are known only to the demander, and secret values $\{y\}$ that are known only to the KSI server. Finally, the provider stores the authentication values $\{\varphi\}$ generated with the initial authentication values during the registration phase.

During the authentication phase, each entity is authenticated. Comparing the final authentication values $\{\varphi\}$ stored during the registration phase and those values $\{\varphi'\}$ that were mathematically computed using other authentication values $\{\sigma\}$ generated during the authentication phase, a mutual authentication is carried out. Thus, threats from an unauthorized entity can be prevented.

- Replay attack (Threat 4)-Data-validity verification (Requirement 4)

By launching a replay attack on the data sent and received during the authentication or data-transfer phase, an attacker can threaten the authentication availability or make the precision medicine system collect data that are duplicated unnecessarily. Assuming that an attacker intercepts the ciphertexts sent or received between entities, the proposed scheme applies time stamps $\{T\}$ on the data shared between entities to carry out validation of data $\{T' \leq \Delta T\}$. Accordingly, the replay attack can be prevented if the time stamps are not valid.

In the data-transfer phase, the data can be protected from replay attacks using the time stamp values {Time: day of the week, month, day, HH:MM:SS, standard time, year}, aggregated or published during the data signature process via the KSI-based technique.

- Repudiation (Threat 5)-Nonrepudiation (Requirement 5)

In this situation, the provider, participating in the precision medicine system, or the demander may deny a history of sharing data. To prevent repudiation, the provider performs $\{Signature(Data_{Pro}) \leftarrow Sign_{sk}(H(Data_{Pro}))\}$ to generate signature values on the data using private keys $\{sk\}$ during the data-transfer phase. The signature values $\{KSI\_Sign(Signature(Data_{Pro})\}$ are then registered on the KSI server to use the KSI-based technique, and the corresponding signature token values $\{KSI\_Token(Signature(Data_{Pro}))\}$ are received from the KSI server. By sending the signature and token values with the data, non-repudiation is achieved against the provider.

Upon receiving the data from the provider, the demander receives the signature token values $\{KSI\_Token(Signature(Data_{Pro}))'\}$ that correspond to the data from the KSI server. Then, the demander performs integrity verification (Requirement 2). As the KSI server owns a log that keeps track of the data receipts at this point, non-repudiation is achieved against the demander as well.

5 Conclusions

In precision medicine, big data related to advanced science and medical services are incorporated into existing medical techniques to establish treatment objectives. This process is followed by precise targeted therapy. As the key data in this case include sensitive data, such as patient history, DNA, and personal information, data security must be ensured during the process of sharing. Therefore, in this study, possible infringement threats to sensitive data in a precise medicine system were outlined, and security requirements were established. Additionally, the sensitive data in the existing PMI were categorized and reestablished according to the proposed architecture of the precision medicine system and data flow. A scheme for securely sharing sensitive data was proposed, and security analyses were performed for the various infringement threats. The sensitive data used in precision medicine are considered to be big data; therefore, to reduce the workload of sharing sensitive data in such a system, a KSI-based technique was implemented to reduce the cryptographic computations while processing the data. The results of this study are expected to help in determining a secure and more efficient method of sharing sensitive data when establishing a precision medicine environment.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

Aboujaoude, E. (2019): Protecting privacy to protect mental health: the new ethical imperative. *Journal of Medical Ethics*, vol. 45, no. 9, pp. 604-607.

Beauvais, M.; Knoppers, B. M. (2020): When information is the treatment? precision medicine in healthcare. *Healthcare Management Forum*, vol. 33, no. 3, pp. 120-125.

Buchmannm, J.; Geihs, M.; Hamacher, K.; Katzenbeisser, S.; Stammler, S. (2019): Long-term integrity protection of genomic data. *EURASIP Journal on Information Security*, https://doi.org/10.1186/s13635-019-0099-x.

Buldas, A.; Kroonmaa, A.; Laanoja, R. (2013): *Keyless Signatures' Infrastructure: How to Build Global Distributed Hash-Trees*. Springer-Verlag Berlin Heidelberg, Germany.

Chen, F.; Jiang, X.; Wang, S.; Schilling, L. M.; Meeker, D. et al. (2018): Perfectly secure and efficient two-party electronic-health-record linkage. *IEEE Internet Computing*, vol. 22, no. 2, pp. 32-41.

Desmond-Hellmann, S.; Sawyers, C. (2011): *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. National Research Council, USA.

Ferryman, K.; Pitcan, M. (2018): *Precision Medicine National Actor Map*. Data & Society, USA.

Gu, K.; Yang, L. H.; Yin, B. (2018): Location data record privacy protection based on differential privacy mechanism. *Information Technology and Control*, vol. 47, no. 4, pp. 639-654.

He, S. M.; Zeng, W. N.; Xie, K.; Yang, H. M.; Lai, M. Y. et al. (2017): PPNC: privacy preserving scheme for random linear network coding in smart grid. *KSII Transactions on Internet and Information Systems*, vol. 11, no. 3, pp. 1510-1532. https://doi.org/10.1109/TDC.2018.8440380.

Hudson, K.; Lifton, R; Patrick-Lake, B. (2015): *The Precision Medicine Initiative Cohort Program*. National Institutes of Health, USA.

Klonoff, D. C.; Price, W. N. (2017): The need for a privacy standard for medical devices that transmit protected health information used in the precision medicine initiative for diabetes and other diseases. *Journal of Diabetes Science and Technology*, vol. 11, no. 2, pp. 220-223.

Liu, X.; Li, Y.; Qu, J.; Ding, Y. (2017): A lightweight pseudonym authentication and key agreement protocol for multi-medical server architecture in TMIS. *KSII Transactions on Internet and Information Systems*, vol. 11, no. 2, pp. 924-944.

Matrana, M. R.; Campbell, B. (2020): Precision medicine and the institutional review board: ethics and the genome. *Ochsner Journal*, vol. 20, no. 1, pp. 98-103.

Min, Z.; Yang, G.; Wang, J.; Kim, G. J. (2019): A privacy-preserving bgn-type parallel homomorphic encryption algorithm based on LWE. *Journal of Internet Technology*, vol. 20, no. 7, pp. 2189-2200.

Mylrea, M.; Gourisetti, S. N. G.; Bishop, R.; Johnson, M. (2018): Keyless signature blockchain infrastructure: facilitating NERC CIP compliance and responding to evolving cyber threats and vulnerabilities to energy infrastructure.

Noorbakhsh-Sabet, N.; Zand, R.; Zhang, Y.; Abedi, V. (2019): Artificial intelligence transforms the future of health care. *The American Journal of Medicine*, vol. 132, no. 7, pp. 795-801.

Qian, T.; Zhu, S.; Hoshida, Y. (2019): Use of big data in drug development for precision medicine: an update. *Expert Review of Precision Medicine and Drug Development*, vol. 4, no. 3, pp. 189-200.

Ra, G. J.; Lee, I. Y. (2018): A study on KSI-based authentication management and communication for secure smart home environments. *KSII Transactions on Internet and Information Systems*, vol. 12, no. 2, pp. 892-905.

Rubin, I. R.; Glusman, G. (2019): Opportunities and challenges in interpreting and sharing personal genomes. *Genes*, https://doi.org/10.3390/genes10090643.

Safavi, K.; Kalis, B. (2018): Digital health tech vision 2018. Accenture Consulting, Ireland.

Sankar, P. L.; Parker, L. S. (2017): The precision medicine initiative's all of us research program: an agenda for research on its ethical, legal, and social issues. *Genetics in Medicine*, vol. 19, no. 7, pp. 743-750.

Schaefer, G. O.; Tai, E. S.; Sun, S. (2019): Precision medicine and big data. *Asian Bioethics Review*, vol. 11, no. 3, pp. 275-288.

Whitsel, L. P.; Wilbanks, J.; Huffman, M. D.; Hall, J. L. (2019): The role of government in precision medicine, precision public health and the intersection with healthy living. *Progress in Cardiovascular Diseases*, vol. 62, no. 1, pp. 50-54.

Xafis, V.; Labude, M. K. (2019): Openness in big data and data repositories. *Asian Bioethics Review*, vol. 11, no. 3, pp. 255-273.

Yin, C. Y.; Shi, L. F.; Sun, R. X.; Wang, J. (2019): Improved collaborative filtering recommendation algorithm based on differential privacy protection. *Journal of Supercomputing*, https://doi.org/10.1007/s11227-019-02751-7.