Original Paper

Data Validation and Verification Using Blockchain in a Clinical Trial for Breast Cancer: Regulatory Sandbox

Tomonobu Hirano[1], BSc; Tomomitsu Motohashi[1], MSc; Kosuke Okumura[1], BSc; Kentaro Takajo[1]; Taiyo Kuroki[1], MSc; Daisuke Ichikawa[1], MD, PhD; Yutaka Matsuoka[2], MD, PhD; Eisuke Ochi[2,3], PhD; Taro Ueno[1], MD, PhD

[1]SUSMED Inc, Tokyo, Japan

[2]Division of Health Care Research, Center for Public Health Sciences, National Cancer Center Japan, Tokyo, Japan

[3]Faculty of Bioscience and Applied Chemistry, Hosei University, Tokyo, Japan

Corresponding Author:
Taro Ueno, MD, PhD
SUSMED Inc
3-8-5 Honcho, Nihonbashi
Chuo-ku
Tokyo,
Japan
Phone: 81 3 3527 3593
Email: t-ueno@umin.ac.jp

Abstract

Background: The integrity of data in a clinical trial is essential, but the current data management process is too complex and highly labor-intensive. As a result, clinical trials are prone to consuming a lot of budget and time, and there is a risk for human-induced error and data falsification. Blockchain technology has the potential to address some of these challenges.

Objective: The aim of the study was to validate a system that enables the security of medical data in a clinical trial using blockchain technology.

Methods: We have developed a blockchain-based data management system for clinical trials and tested the system through a clinical trial for breast cancer. The project was conducted to demonstrate clinical data management using blockchain technology under the regulatory sandbox enabled by the Japanese Cabinet Office.

Results: We verified and validated the data in the clinical trial using the validation protocol and tested its resilience to data tampering. The robustness of the system was also proven by survival with zero downtime for clinical data registration during a Amazon Web Services disruption event in the Tokyo region on August 23, 2019.

Conclusions: We show that our system can improve clinical trial data management, enhance trust in the clinical research process, and ease regulator burden. The system will contribute to the sustainability of health care services through the optimization of cost for clinical trials.

Introduction

Clinical trials involve a large flow of medical information, and it is necessary to secure the transparency and traceability of clinical data. The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use–Good Clinical Practice (ICH-GCP) provides guidance for monitoring the conduct of a clinical trial to verify that reported clinical data are complete, accurate, and accounted for by source records. The long-standing practice in the pharmaceutical and medical device industries has been frequent site visits and 100% source data verification (SDV) on site to ensure that data captured in source records are transferred correctly to the case report forms. It has been estimated that clinical trial monitoring by SDV can consume one-quarter of the sponsor's entire budget for a clinical trial [1]. As the complexity and size of clinical trials grow, it has become increasingly expensive to apply the 100% SDV approach [2,3]. For regulators of trials, such as the US Food and Drug Administration (FDA), European Medicines Agency (EMA) in the European Union, and Pharmaceuticals

and Medical Devices Agency (PMDA) in Japan, data auditing is challenging since there is no easy and secure way to access or view complex data. As such, process improvement is an active area of research [4]. With the involvement of more parties and more exchanges, the risk for human-induced error, whether unintentional or malicious, will increase. Previous studies have shown that 17% of authors of clinical trials reported they were personally aware of intentional fabrication in research [5,6]. In the case of misconduct in a clinical trial by an employee of Novartis Pharma, clinical data were manipulated to attribute qualities to the hypertension drug valsartan that it did not possess, such as preventing stroke [7]. The fabricated clinical data were used in an advertising campaign for the drug, and patients were prescribed the drug based on incorrect information. Following misconduct in clinical trials, the Japanese government established and enforced the Clinical Trials Act in 2018 [8]. It obligates clinical researchers to monitor and assure quality, observe practice criteria, and manage conflicts of interest.

Blockchain is emerging as a groundbreaking technology for secure data control in different areas. Bitcoin was the first application of blockchain as a digital currency in extensive use [9]. Recently, researchers have started to focus on using blockchain methodology for building a cryptographic proof of medical systems [10]. They have applied blockchain technology in various health care systems, with potential applications in supply chain management of health care products [11-13], insurance claims processing [14], management of electronic health records and electronic medical records [15-18], maintenance of protocols in clinical trials [19-21], and data management in clinical trials [22-25]. Although there can be many applications of blockchain technology, it is necessary that blockchain be an appropriate technical solution to a particular problem. If there are no incentives for data tampering and all writers can be trusted, blockchain technology is not necessary. On the other hand, if there are incentives for data tampering and it costs a lot to use a trusted third party, such as the Clinical Research Organization, using blockchain makes sense [26]. Due to the frequent occurrences of misconduct in clinical trials and the large consumptions of clinical trial budgets and time by current SDV practices, blockchain is an appropriate technical application in the data management of clinical trials to solve these problems.

We propose a solution to challenges in the current clinical trial system by using blockchain technology coupled with technologies such as client hashchain, encryption protocol, and health check function in servers for hazard management. The system was used in a clinical trial at the National Cancer Center of Japan that investigated the effect of home-based high-intensity interval training for breast cancer survivors. The project was conducted to demonstrate clinical data management using blockchain technology under the regulatory sandbox of the Japanese Cabinet Office [27,28]. This sandbox allows organizations to apply for demonstration and evaluation of new technology, such as blockchain and Internet of Things, without being subject to existing regulations. It simultaneously opens up the possibility for future deregulation measures. During the trial, we experienced the disruption of Amazon Web Services (AWS) cloud servers in the Tokyo region on August 23, 2019

[29]. We report the effect of the blockchain network and health checkup function, which resulted in system survival with zero downtime and secure clinical data registration during the hazardous server shutdown event.

Methods

Breast Cancer Clinical Trial

A clinical trial that investigates the effect of intervention by home-based high-intensity interval training for breast cancer patients was conducted. The study was a parallel-group, single-blind randomized controlled trial. Patients were randomly assigned to the active group with the habit-B program (high-intensity interval training, exercise counseling and guidance, home-based exercise support using information and communication technology, and a wearable device) or treatment as usual with a wearable device.

The eligibility criteria for participants were (1) female, aged between 20 and 59 years at diagnosis; (2) diagnosed with stages I to IIa breast cancer and currently 2 to 13 months postsurgical treatment; (3) not requiring cancer chemotherapy aside from hormone therapy; (4) ability to read, write, and understand Japanese; (5) ability to complete an electronic Patient Reported Outcome Questionnaire via smartphone; (6) consent to trial participation obtained in writing from the patient themselves; and (7) currently engaging in not more than moderate intensity exercise for 30 minutes on two separate days per week (total of 60 minutes).

The researcher provided computer-generated random allocation. Using the generated account, the contents of the app that participants used during the clinical trial were assigned automatically to either the habit-B program or the control. When a participant first visits the hospital, she is randomly assigned to an intervention, either with the habit-B program or treatment as usual with a wearable device, after signing informed consent. An independent data center provides computer-generated random allocation as a log-in account for the app. Based on the allocation sequences, the contents of the app participants use during the trial are assigned automatically to either the habit-B program or the control.

The study received ethical approval from the institutional review board of the National Cancer Center Japan and was registered at University Hospital Medical Information Network–Clinical Trial Registry [UMIN000036400]. Participants provided written informed consent to take part in the study. The study began in May 2019 and is ongoing [30].

Architecture of the Data Management System

The data management system that collects the clinical trial data consists of client smartphones, relay servers, and a blockchain network (Figure 1). The relay servers and blockchain network are built on AWS in the Tokyo region. We use Hyperledger Fabric v1.0 (Linux Foundation), an open-source blockchain platform that has been widely used, to operate the blockchain network. Hyperledger Fabric enables throughput of more than 3500 transactions per second in popular deployment configurations [31].

Figure 1. Data management system including client smartphones, relay servers, and blockchain network. The collected data from patient smartphones were sent to the blockchain network via relay servers. We used three relay servers, and the app randomly selected two relay servers to send the data after the authentication of the client device. We have configured the blockchain network to span multiple availability zones in Amazon Web Services.
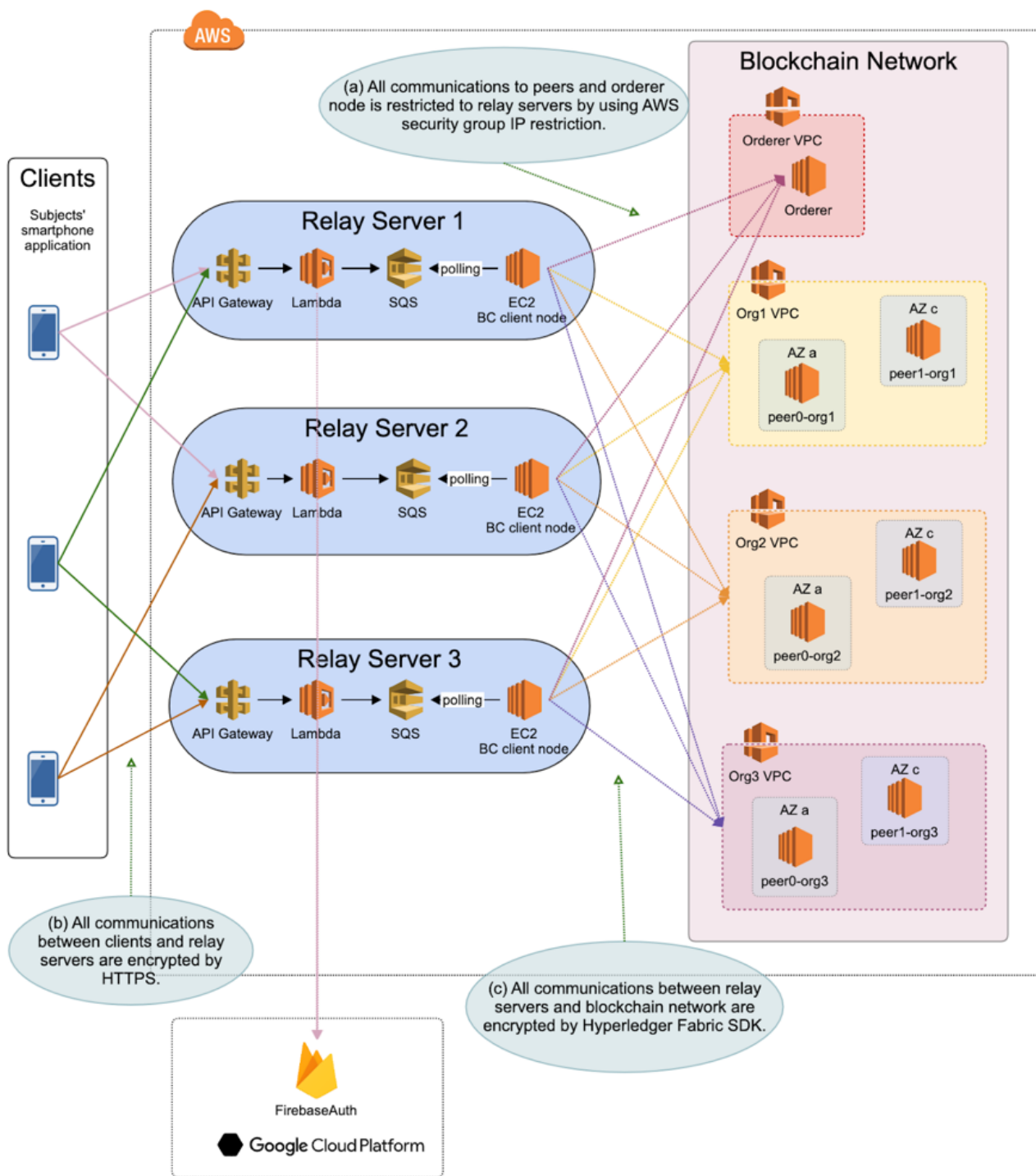
Data collected from patient smartphones are sent to the blockchain network via relay servers. We use three relay servers, and the app randomly selects two relay servers to send the data after the authentication of the client device. The client app contains an authentication key, and the relay servers verify the key so that the relay server only accepts access from the app for the clinical trial. The relay servers also verify the account for the app using Firebase authentication. These communications between clients and relay servers are encrypted using https. By deploying the relay server and setting the blockchain software development kits to append-only mode, the relay servers send the received data to the blockchain network. The data sent from the relay servers to the blockchain network are encrypted using the authentication protocol by the Hyperledger Fabric. By configuring the internet protocol address restriction to the listed relay proxy, the blockchain network, which contains the clinical trial data, is protected against malicious attack from the external network. The blockchain network is made up of three organizations, each of which contains two validating peers. Each identification and password for relay servers, blockchain

network organizations, and nodes of the blockchain network is managed by two independent departments in SUSMED Inc and the National Cancer Center of Japan [22].

We have configured the blockchain network to span multiple availability zones (AZs) in AWS. Each AZ is made up of one or more data centers equipped with independent power, cooling, and networking to ensure fault tolerance. We have also implemented a periodic health check function using the Amazon EC2 Auto Scaling Service. The function continues to maintain a fixed number of instances even if an instance becomes unhealthy. If an instance becomes unhealthy, the group terminates the unhealthy instance and launches another instance to replace it.

In order to increase system durability and prevent data loss, we have configured two kinds of queues in our system. The mobile app puts clinical data into the queue before sending, dequeues and sends the message, and then removes the data from the queue. If relay servers are unavailable, the sending process fails and the data remain available in the queue until the relay servers process it. The other queue is in a relay server system, and the queue is configured using AWS Simple Queue Service (SQS). When a relay server system receives the clinical data, the data are enqueued. In the relay server system, an EC2 instance is responsible for dequeuing and sending the message to the blockchain network. The data are removed from the queue after being successfully processed by the blockchain network. Otherwise, the data remain available in the queue until the sending process succeeds or the SQS retention period expires. In this system, we set the SQS retention period to 7 days. In doing so, we have seven days' time to recover the blockchain network if there is a breakdown.

Data Collection Using Smartphone

The Global Physical Activity Questionnaire is an internationally standardized questionnaire for surveying physical activity level [32-34]. Fear of cancer recurrence is assessed by the overall fear index score on the Concerns About Recurrence scale [35,36]. Depression is assessed using the Patient Health Questionnaire–9 [37]. Fatigue is assessed by the Cancer Fatigue Scale. Sleep is assessed by the Athens Insomnia Scale [38]. Quality of life is assessed using the EuroQol 5-Dimensions questionnaire [39,40]. All data listed above are stored in JavaScript Object Notation format in the database.

Data Registration to the Blockchain Network

When a participant first logs in to the system, the app generates a secret hash key and preserves it in the client device until the end of the study. The client device calculates a hash value based on the data generated by the patient and the secret hash key as well as the previous hash value using the SHA-256 hash algorithm [22,41]. Thus, the hash value in the client device comprises the chain structure. The hash value is also registered in the blockchain network along with the clinical data in order to guarantee tamper resistance of the value, although the secret hash key was preserved in the client device until the end of study. The collected data and client hash value are sent to the blockchain network via relay servers. We use three relay servers, and the app randomly selects two relay servers to send the data

after the authentication of the client device. Client nodes of the blockchain network are settled in the relay servers and the data are sent from the relay server to the blockchain network. The transactions are validated and accepted by the blockchain network in the following processes:

- Proposal: transaction is sent from client app to the endorsers in each organization
- Endorse: each endorser verifies that (1) the transaction proposal is well formed, (2) it has not already been submitted, (3) the signature is valid, and (4) the client is properly authorized to perform the proposed operation, which is described in the chaincode. If the transaction is validated, the chaincode is executed and the result with the signature is returned to the client
- Submit: client verifies that the number of signatures from the organizations satisfies the endorsement policy. If it is satisfied, the transaction is sent to the ordering service, which orders the series of transactions in chronological order and creates the block of transactions
- Broadcast: the block is delivered to all nodes
- Commit: if each block is well formed and validated to fulfill the endorsement policy, the block is appended to the chain in each node [22]

Data Validation and Verification

When participants complete the data input for the clinical trial, it is necessary to verify all data records sent to the blockchain network to satisfy the following conditions:

- Blockchain network has received all required data types for each participant
- All data records for each participant are correct as a client hashchain
- Blockchain network has received clinical records from two relay servers for each record, and the records are the same

The first condition means that the blockchain network has received all records that are expected to be sent from clients. There are three types of records in the clinical trial. To begin with, a client sends a record that indicates that a secret hash key has been generated. When clinical data are created, the data are sent to the blockchain network via two relay servers. Finally, when the participants complete the clinical trial, the client sends the secret hash key to the blockchain network.
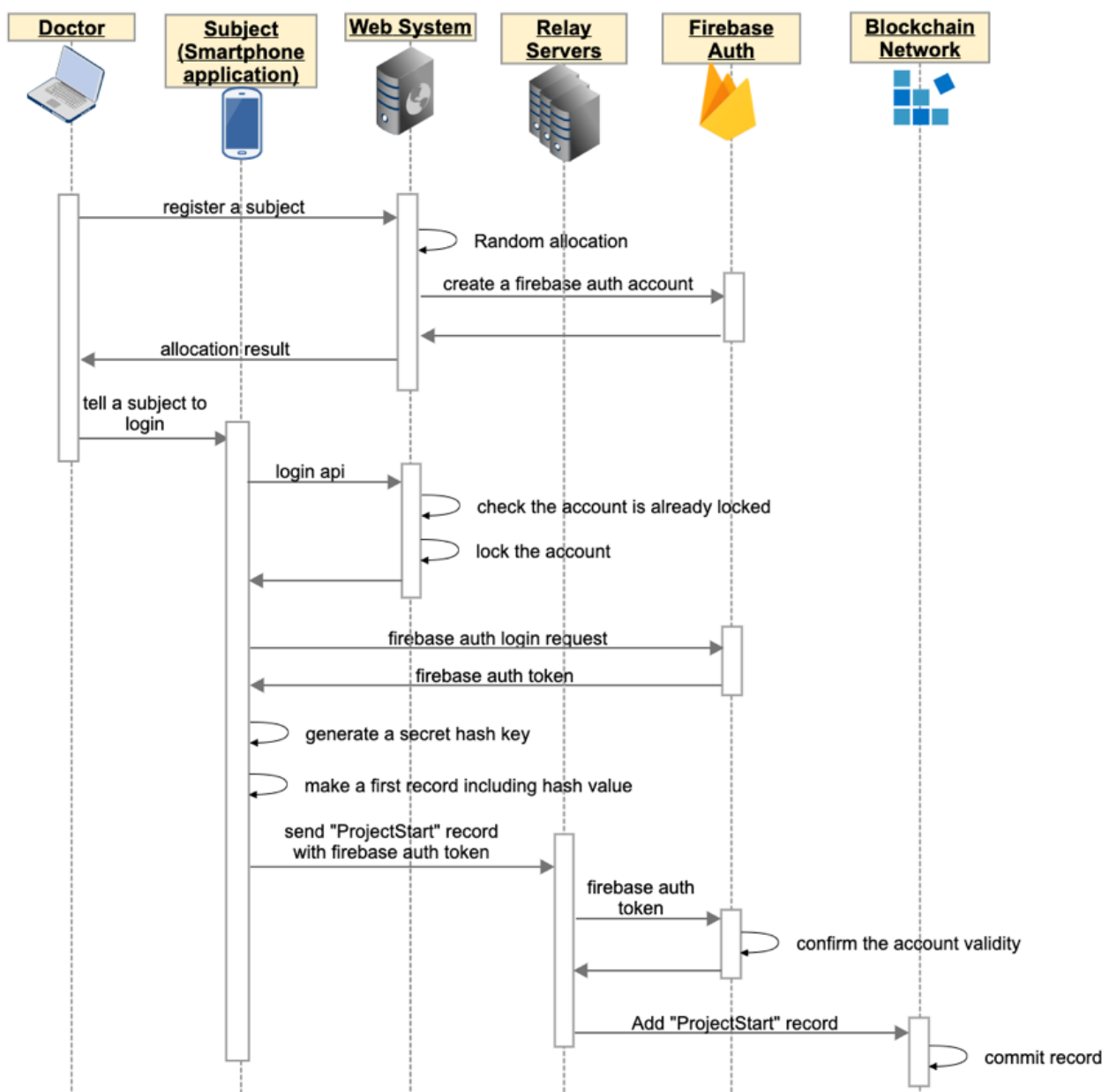
The second condition verifies that all data records registered to the blockchain network are correctly linked as a client hashchain. We sort all data records by their generated time and make a hash value from the clinical data, the secret hash key preserved in the client device until the end of the study, and the previous hash value. Then, we compare the results with the hash value that has been registered to the blockchain network. Due to the sensitivity of a hash function output in relation to its input, changing the data will result in a completely different hash string. Therefore, we can confirm these hash values are generated by the client device that preserves the secret hash key. By verifying this condition, we can detect if a malicious attacker has falsified the clinical data using hash values and clinical data that were stolen.

The purpose of the last condition is to confirm that the relay servers have not been hacked. Clients send their clinical data to two relay servers that are randomly selected from the three relay servers. If malicious users hack a relay server and they send some data from the relay server, the blockchain network will store the data. In order to detect the fraud access and verify the data, we need to confirm that all clinical data is received from two relay servers simultaneously and the data records are identical.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Results

Registration and Assignment

The participant downloads the app to their own smartphone through the internet and logs in to the system using the account provided by the researcher. When the participant first logs in to the system, the app generates a secret hash key and preserves it in the client's device until the end of the study. The account is locked when the participant first logs in to the system in order to prevent impersonation by stolen account or brute force attack (Figure 2).

Figure 2. When a participant first visits the hospital, a medical doctor decides if they fit the eligibility criteria. After obtaining informed consent, the doctor registers the patient to the clinical trial using a web-based allocation system. The system automatically creates an account for the participant, and the doctor gives the account to the participant. The participant downloads the app to their smartphone, and they can log in to the app using the account created by the doctor. When the participant first logs in to the app, the account is locked to prevent impersonation, and the secret hash key is then generated. The secret hash key will be preserved in the client app until the participant completes the trial. Using the secret hash key in the client device, the app calculates a hash value. The app sends the first record to the blockchain network via relay servers, and the data is registered in the tamper-resistant blockchain system.

Data Collection

During the trial, the participant inputs the data using the app, and the data are registered to the blockchain network via relay servers (Figure 3). The secret hash key is preserved only in the client's device during the clinical trial. The client app calculates a hash value based on the medical data, the secret hash key, and the previous hash value using the SHA-256 hash algorithm. The client hash value is also sent to the blockchain network along with the clinical data.

Figure 3. During the clinical trial, participant inputs clinical data using the mobile app. The app calculates the hash value based on the clinical data, the secret hash key preserved in the smartphone until the end of the study, and the previous hash value. The app sends the clinical data and a hash value to the blockchain network via relay servers.
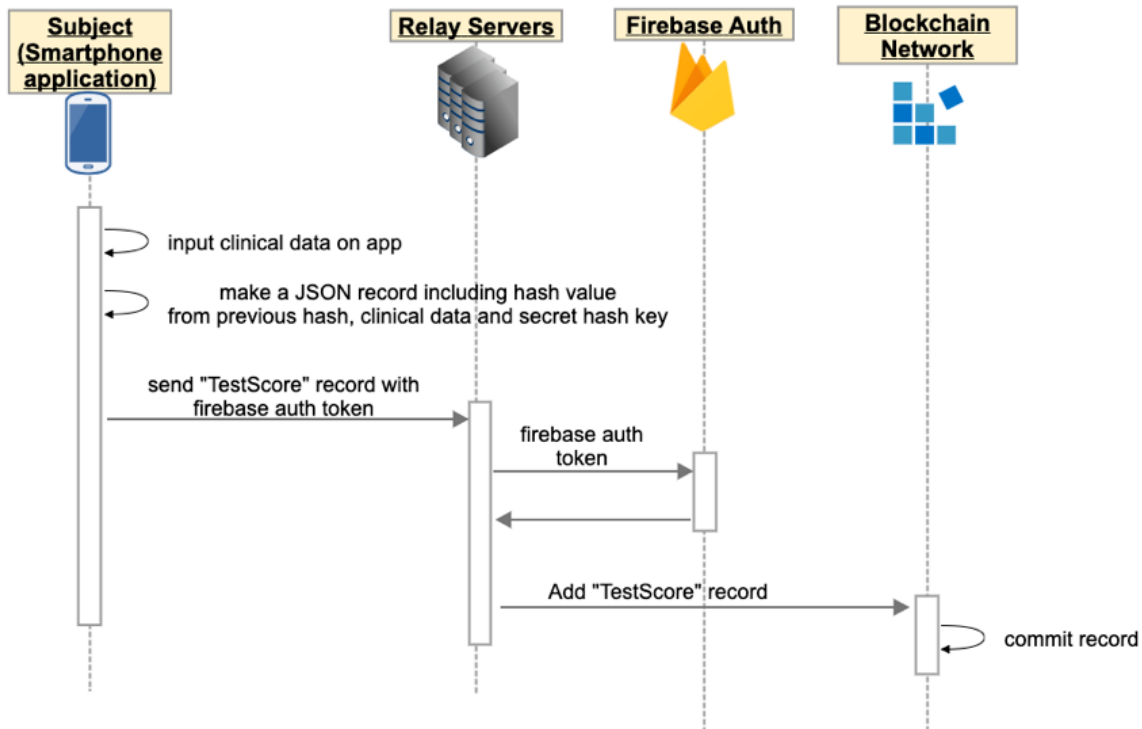
Completion of the Clinical Trial

When the participant completes the clinical trial, the app sends the secret hash key to the blockchain network, and all data of the participant is registered to the blockchain network (Figure 4).

Figure 4. At the end of the clinical trial, the app extracts the secret hash key from the smartphone and sends it to the blockchain network via relay servers.
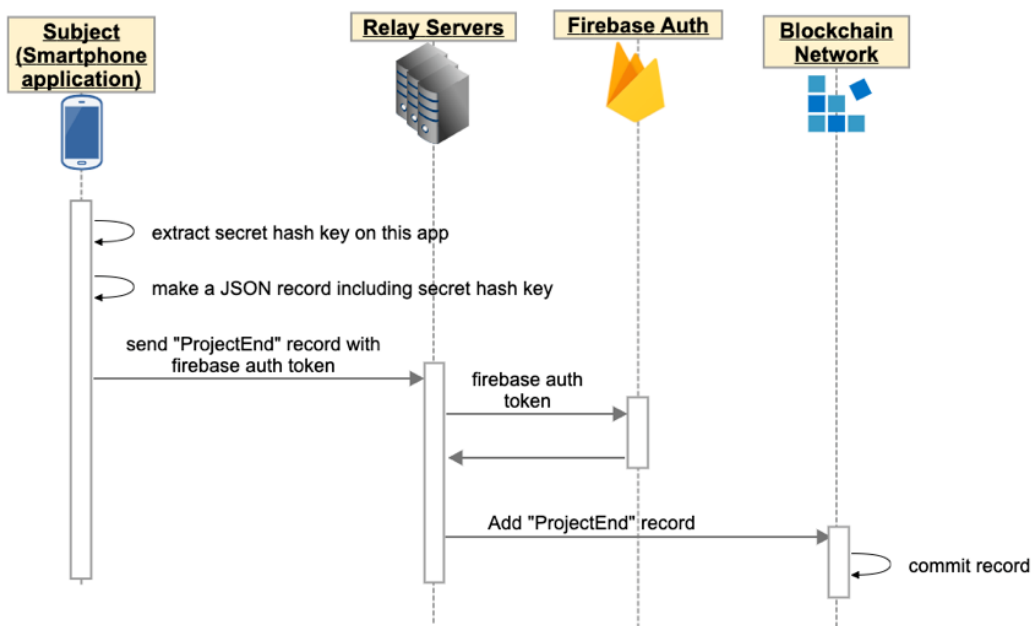
Data Validation and Verification

We verify and validate the patient's data collected in the clinical trial using a predetermined protocol. For regulators such as FDA, EMA and PMDA, the full transaction history since the beginning of the study is readily available with precise log data, and the auditing process can be done quickly and with the reliance that all clinical data are guaranteed or version controlled [23].

Blockchain technology ensures a secure and tamper-proof transaction history so that the blockchain network provides an integrity-protected data storage and process transparency [24]. However, there can be vulnerability before registration to the database in the blockchain network. The data can be tampered with and impersonation could occur, resulting in impairment of reliability of the data. To prevent impersonation of the client, we have configured the mechanism so that the account is locked when the participant first logs in to the system. The client app contains the authentication key, and the relay server checks the key so that the relay server only accepts access from the app for the clinical trial. By configuring the internet protocol address restriction to the listed relay server, the blockchain network, which contains the medical data, is protected against external attack. This configuration makes our system highly secure. Even if an attacker breaks the above security system under impersonation or unauthorized access to the relay server, we can validate and distinguish the correct data using the client hashchain function. In order to complement the tamper resistance of the blockchain network and ensure reliability of the entire system, we verified and validated the data integrity using client hashchains and multiple relay servers. User data along with the client hashchain are registered to the blockchain network. The app sends the user data and client hashchain to the blockchain network during the clinical trial. At the end of the trial, the app sends the secret hash key preserved in the client device to the blockchain network. Using the secret hash key, we verify the integrity of the data by calculating the hashchain retrospectively (Figure 5). User data and the client hashchain are sent from multiple relay servers to the blockchain network. A client app sends its clinical data to two relay servers, which are randomly selected from the three relay servers. In order to detect fraudulent access and verify the data sent from the relay servers, the data are compared with each other to confirm that the records are the same (Figure 6).

Figure 5. Data verification and validation using client hashchain. The first hash value is made from the first data and secret hash key. Subsequent hash values are made from clinical data, the previous hash key, and secret hash key.
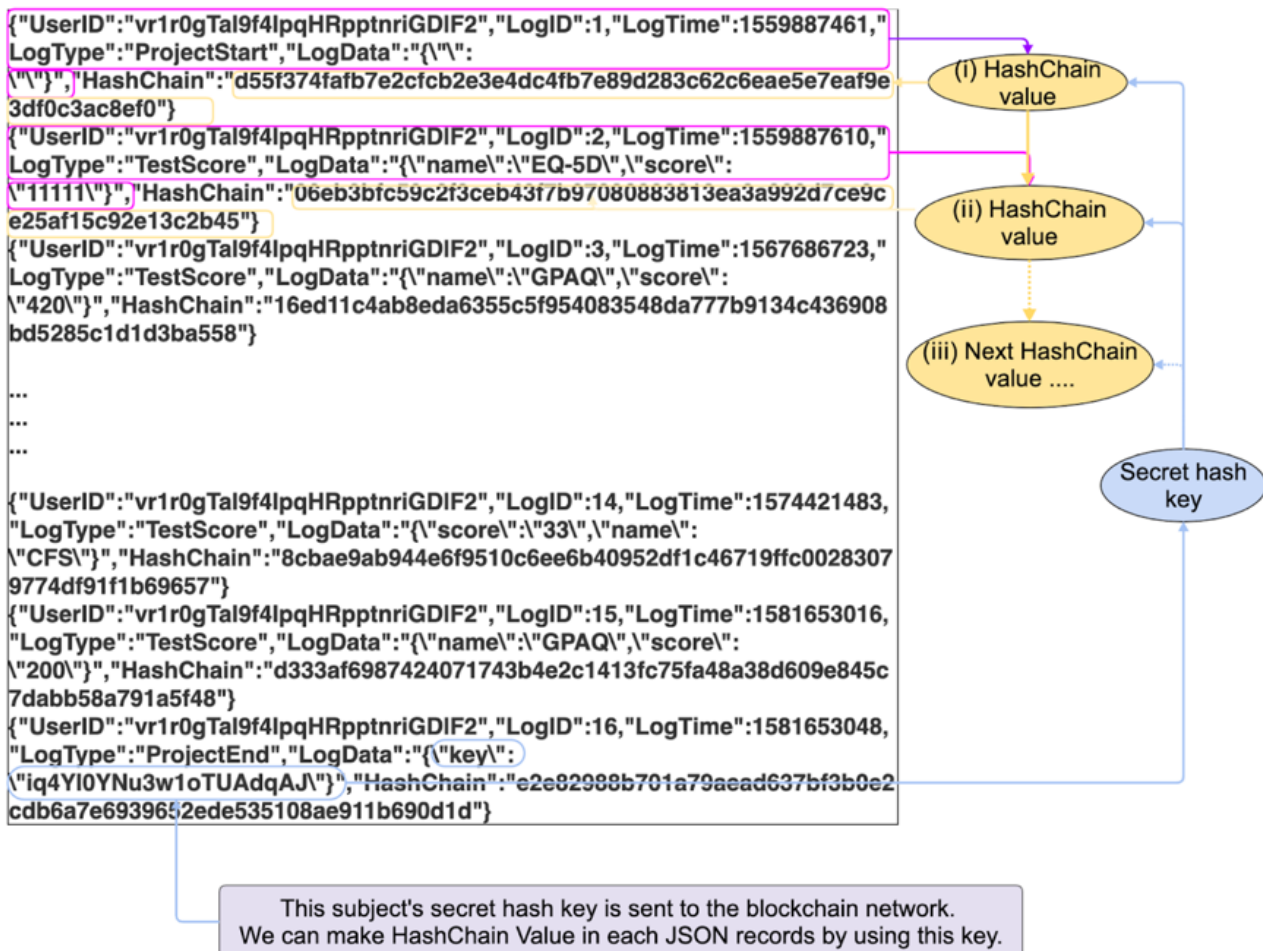
Figure 6. Data verification and validation using records from relay servers.

Admin@org1.eprobc.susmed.io{"UserID":"1MHPAbYtVxR6eVh9j98qOh3G3Y33","LogID":11,"LogTime":1578085488,"LogType":"TestScore","LogData":"{\"name\":\"GPAQ\",\"score\":\"420\"}","HashChain":"bb96882e1913b4e5ba24b67042e086be5e8f7dd5b607772100db52a01769b1e2"}

Admin@org1.eprobc.susmed.io{"UserID":"1MHPAbYtVxR6eVh9j98qOh3G3Y33","LogID":12,"LogTime":1578085549,"LogType":"TestScore","LogData":"{\"name\":\"AIS\",\"score\":\"2\"}","HashChain":"21fe640b64b9b6d9882e865560f5dc1e552f756cb7b985a49fb8d904fbedd511"}

...

Admin@org2.eprobc.susmed.io{"UserID":"1MHPAbYtVxR6eVh9j98qOh3G3Y33","LogID":1,"LogTime":1563516422,"LogType":"ProjectStart","LogData":"","HashChain":"19c8855619c870ec96794303485388a4421072c4740add386e9734f962b4fc0e"}

Admin@org2.eprobc.susmed.io{"UserID":"1MHPAbYtVxR6eVh9j98qOh3G3Y33","LogID":10,"LogTime":1578085307,"LogType":"TestScore","LogData":"{\"name\":\"CFS\",\"score\":\"17\"}","HashChain":"622963b3687e6860f1cc3b7a792b59eab5400109fd973f972d05922c23751234"}

...

Admin@org3.eprobc.susmed.io{"UserID":"1MHPAbYtVxR6eVh9j98qOh3G3Y33","LogID":1,"LogTime":1563516422,"LogType":"ProjectStart","LogData":"","HashChain":"19c8855619c870ec96794303485388a4421072c4740add386e9734f962b4fc0e"}

Admin@org3.eprobc.susmed.io{"UserID":"1MHPAbYtVxR6eVh9j98qOh3G3Y33","LogID":10,"LogTime":1578085307,"LogType":"TestScore","LogData":"{\"name\":\"CFS\",\"score\":\"17\"}","HashChain":"622963b3687e6860f1cc3b7a792b59eab5400109fd973f972d05922c23751234"}

...

Every record starts with Relay Server ID like "Admin@org1.eprobc.susmed.io". We can confirm that every record is sent from at least 2 relay servers and the data records are the same.

To evaluate the validation, we have simulated and tested client impersonation and data falsification by fraudulent access to the relay server. The integrity of the data can be maintained even if attacker tries to manipulate the data by impersonating the client device (Figure 7) or fraudulently accessing the relay servers (Figure 8). The result suggests that our proposed function of client hashchain and multiple relay servers can be a complement to the tamper resistance of the blockchain network to ensure the reliability of the entire system.

Figure 7. Simulation of data falsification by client impersonation. If an attacker successfully logs into the app using a stolen account or brute force attack, the attacker can create clinical data from another client device. The data from the impersonated client will be denied since only the true participant's app preserves the secret hash key that is required to calculate the hashchain.
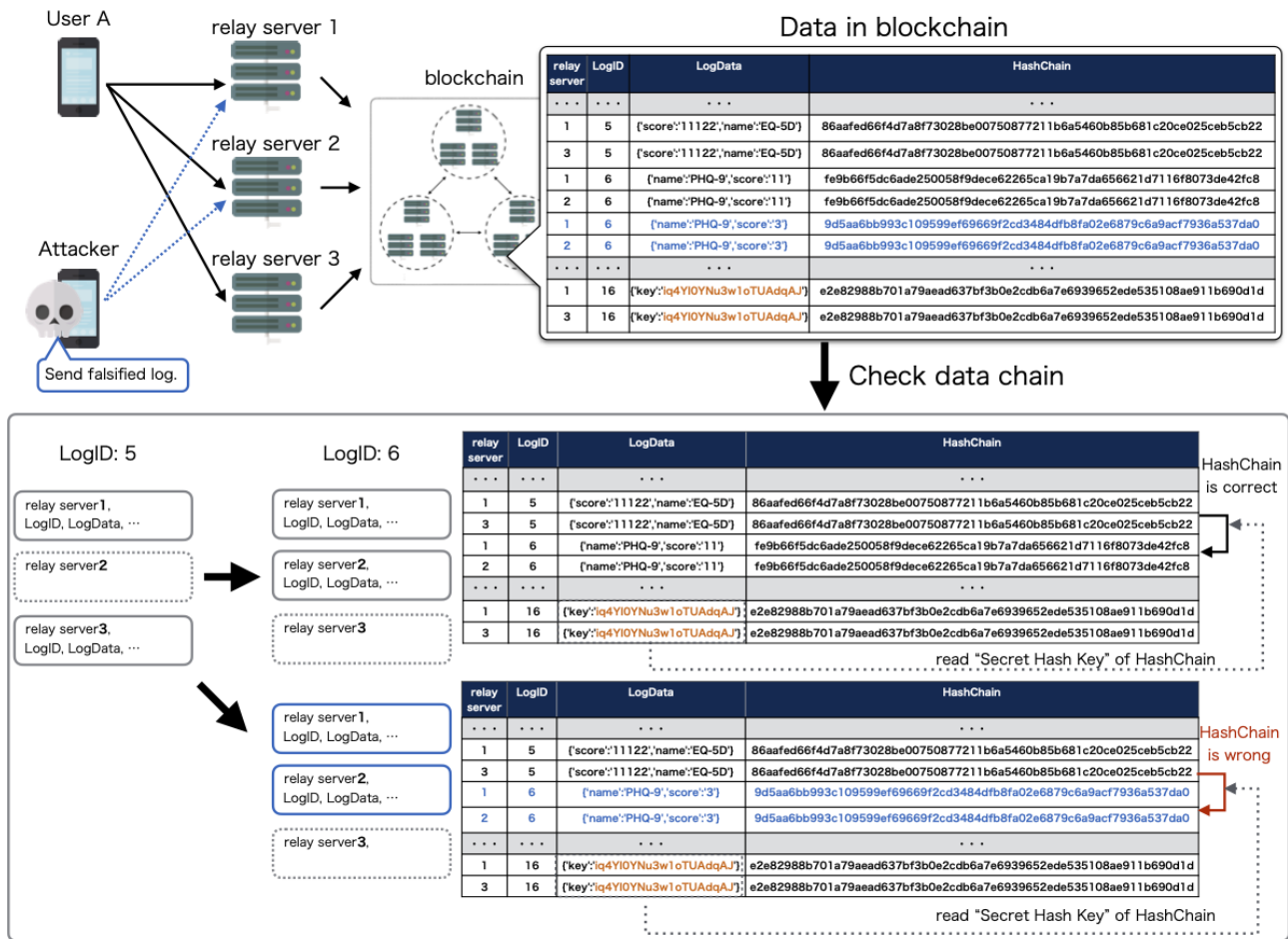
Figure 8. Simulation of data falsification by fraudulent access to the relay server. If an attacker successfully accesses the relay server fraudulently, the attacker can send falsified data to the blockchain. The falsified data from the hacked server will be detected by the integrity of the hashchain.
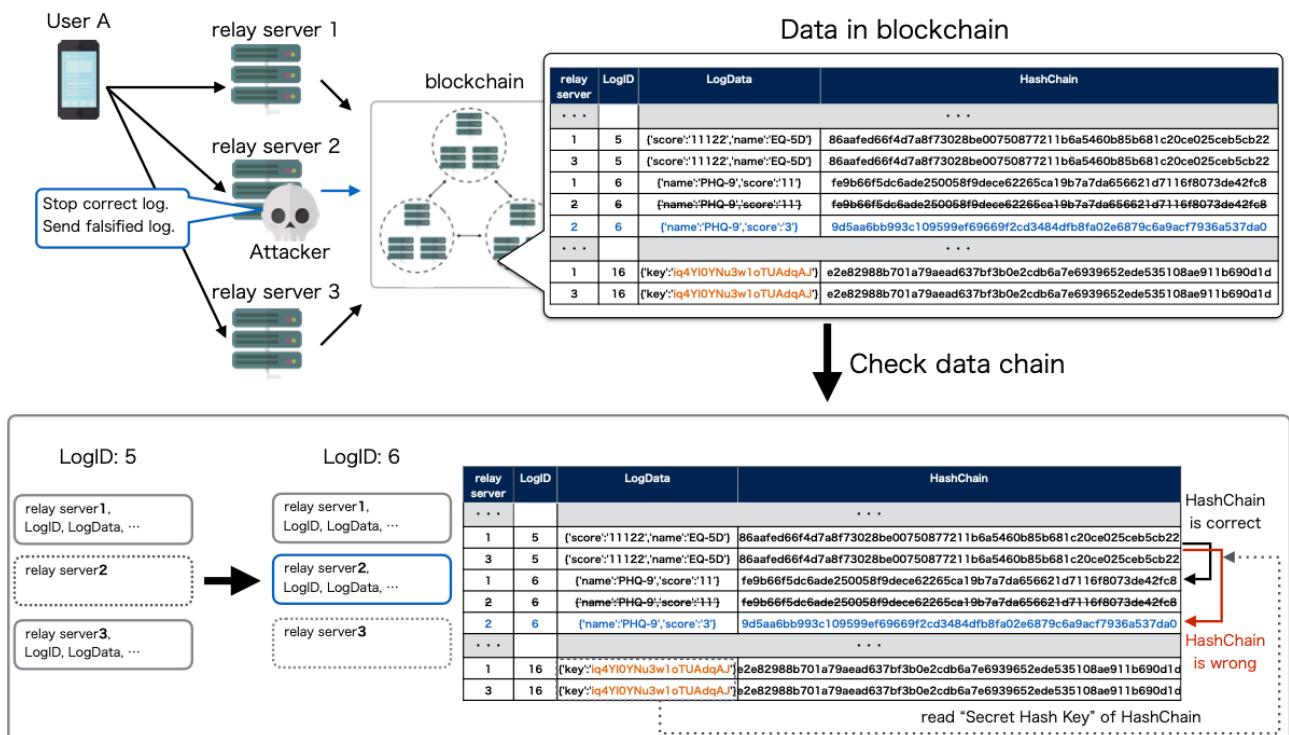
Data Integrity During Amazon Web Services Disruption Event

During the trial, we encountered a disruption event of the AWS cloud server in the Tokyo region on August 23, 2019 [29]. This resulted in impaired EC2 instances for some resources in the affected area of the AZ. One of the blockchain network nodes shut down due to disruption of the data center. Since the blockchain network was composed of three organizations that contain two validating peers and we had configured the blockchain network to span multiple AZs in AWS, the redundancy maintained stable operation of the system during the event. Furthermore, our configuration of a periodic health check function using Amazon EC2 Auto Scaling Service automatically restored the affected node in a healthy data center (Figure 9). We have confirmed there were no error records during the disruption event.

Figure 9. System resilience during Amazon Web Services disruption event in the Tokyo region. To secure the redundancy of the system, we have configured the blockchain network to span multiple availability zones (AZs) in Amazon Web Services. Each AZ is made up of one or more data centers equipped with independent power, cooling, and networking to ensure fault tolerance. We have also implemented a periodic health check function using the Amazon EC2 Auto Scaling Service. When one of the nodes in the blockchain network broke down due to the severe disruption of the data center on August 23, 2019, our system maintained stable operation and lost no data by the redundancy of the blockchain network. Furthermore, the health check system using the Amazon EC2 Auto Scaling Group enabled the automatic recovery of the blockchain node in a healthy AZ after the data center disruption event.
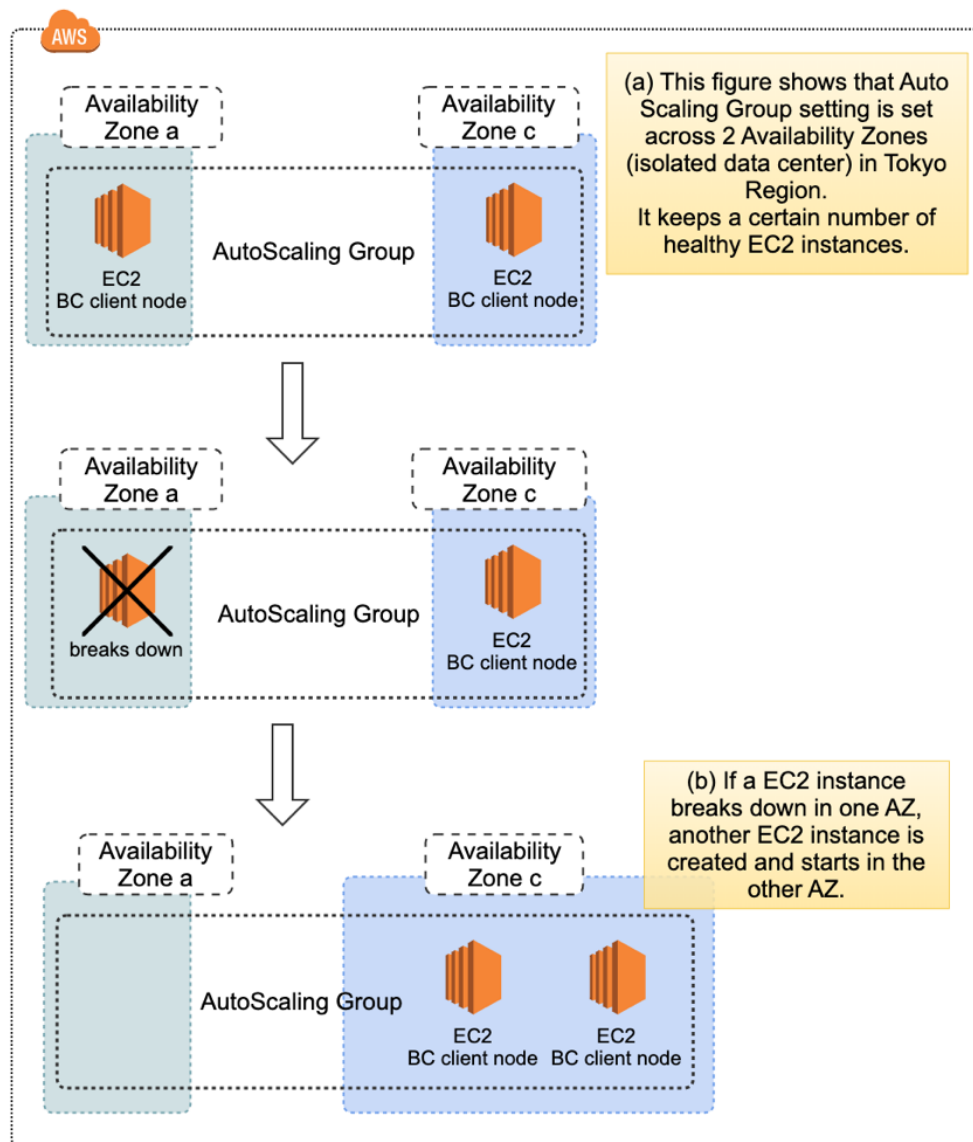
Even if further instances in the blockchain network or relay server components break down, we are able to prevent data loss. There are queues in the client smartphone app and relay servers, and the queues retain the clinical data until data are successfully sent. While the queues contain the data, we can recover the affected system.

Discussion

Principal Findings

In this study, we showed that a data management system using blockchain technology can reliably protect medical data and provide an immutable and fully traceable audit trail in a clinical trial. Due to the sensitivity of a hash function's output in relation

to its input, any change in clinical data will result in a completely different hash string. This string will be an input of the next block's hash function, and the resulting string will be completely different from what it was prior to the data modification. As such, data integrity in a blockchain-based data management system can be checked by simply comparing the hash values of a blockchain. Verifying integrity can be done quickly, as the regulator needs only check for hash string equivalence. Our system can thus offer an improvement in clinical trial data management, enhancing trust in the clinical research process and easing regulator burden [23]. Although we used Hyperledger Fabric for the blockchain platform, other open-source platforms such as Ethereum can also be applied [42,43]. Furthermore, our system can also be used in clinical trials that collect data via other devices like PC or Internet of Things.

ICH-GCP provides guidance for monitoring conduct of a clinical trial to verify that reported clinical data are complete, accurate, and accounted for by source records. Representatives from the pharmaceutical and medical device industries must visit hospitals frequently and verify data by SDV to confirm that collected medical data are transferred correctly to the case report forms. It has been reported that almost half of the total budget for a phase 3 clinical trial is GCP-related and half of the GCP-related cost is related to SDV [1,3]. The total cost for current SDV in a typical phase 3 trial is estimated at US $90 million, which will be added to the price of the drug [3]. As the size and complexity of clinical trials grow, the cost of monitoring by the traditional method has become increasingly expensive. EMA and FDA have published updated guidelines for the monitoring and oversight of clinical trials. Both agencies encourage study sponsors to implement remote and risk-based monitoring [44,45]. Although remote monitoring may be an attractive alternative to traditional SDV, some studies have shown that remote monitoring can be more time-consuming than traditional monitoring and may be insufficient as a complete alternative for SDV [46]. Another option, risk-based monitoring, also needs risk assessment before its implementation and will not eliminate the need for data verification by the human eye. Conversely, our blockchain-based system enables secure data management without labor-intensive monitoring operations in medical practice. It can drastically reduce the cost of human resources and the risk of human-induced error. Information technology systems for the management of clinical trials need to keep audit trails permanent, tamper-proof, and verifiable, as stipulated by FDA 21 CFR part 11. Blockchain for clinical research involves a promising set of technologies that may advance data integrity and efficiencies in clinical research.

This study was conducted to demonstrate clinical data management using blockchain technology under certification of the regulatory sandbox of the Japanese Cabinet Office. The Japanese government aims to actively explore innovative and disruptive technologies through this regulatory sandbox system under the Cabinet Secretariat to address challenges such as extreme aging of the population [28]. Japanese society is facing challenges like increased national public health expenditures, higher demand for health care services lacking appropriate cost controls, and shortage of health care workers. Some of these challenges are becoming more acute, with recent data indicating that expenditures in health care insurance are growing in Japan [47]. The data collected through the demonstrations in the regulatory sandbox will be used in deliberation for regulatory reform, which will facilitate the creation of innovative business activities with new technologies and new business models.

Severe disruption of cloud server services occurred in the AWS Tokyo region on August 23, 2019, during the clinical trial. Thousands of companies, ranging from internet retailers and smartphone payment platforms to game providers, were affected and their services were suspended. During the event, customers of these services could not access their own information and lost data that should have been generated during the event. Our system maintained stable operation and lost no data from the participants of the clinical trial during the disruption event due to the redundancy of the blockchain network. The configuration of the blockchain network to span multiple data centers in AWS, along with the health check system using Amazon EC2 Auto Scaling, enabled automatic recovery of the blockchain network after severe disruption of the data center. These features in our system, which provide zero downtime for the availability of patient data, are suitable to meet the requirements of clinical trials.

Limitations

Further studies are needed to verify the scalability of the system for conducting various kinds of clinical trials. Although the transaction throughput in the Hyperledger Fabric platform that we used is much higher than public blockchain, further studies like multiple clinical trials in a single system may be fruitful. Data collection via multiple devices should also be tested toward the dissemination of virtual clinical trials.

Conclusion

In our study, we ran a clinical trial for breast cancer using a blockchain-based clinical data management system. The data were validated by protocol, and traceability was secured. Even with the severe disruption event of cloud server services, our system preserved the integrity of the clinical trial data. The study was conducted as a regulatory sandbox project by the Cabinet Office of Japan and will open the possibility for future deregulation measures enabling efficient clinical development.

Authors' Contributions

TU designed the research; TH, TM, KO, KT, YM, and YO performed the research; TH and DI analyzed the data; and TH, TM, TK, DI, and TU wrote the paper.

Conflicts of Interest

TH, TM, KO, KT, TK, DI, and TU are members of SUSMED Inc. YM has received speaker fees from Suntory, Pfizer, Mochida, Eli Lilly, Morinaga Milk, and NTT Data and is conducting collaborative research with Morinaga Milk and SUSMED Inc.

References

1. Olsen R, Bihlet AR, Kalakou F, Andersen JR. The impact of clinical trial monitoring approaches on data integrity and cost: a review of current literature. Eur J Clin Pharmacol 2016 Apr;72(4):399-412. [doi: 10.1007/s00228-015-2004-y] [Medline: 26729259]

2. Scannell JW, Blanckley A, Boldon H, Warrington B. Diagnosing the decline in pharmaceutical R&D efficiency. Nat Rev Drug Discov 2012 Mar 1;11(3):191-200. [doi: 10.1038/nrd3681]

3. Funning S, Grahnén A, Eriksson K, Kettis-Linblad A. Quality assurance within the scope of Good Clinical Practice (GCP)—what is the cost of GCP-related activities? A survey within the Swedish Association of the Pharmaceutical Industry (LIF)'s members. Qual Assur J 2009 Jan;12(1):3-7 [FREE Full text] [doi: 10.1002/qaj.433]

4. Bhatt A. Quality of clinical trials: a moving target. Perspect Clin Res 2011 Oct;2(4):124-128 [FREE Full text] [doi: 10.4103/2229-3485.86880] [Medline: 22145122]

5. Gupta A. Fraud and misconduct in clinical research: a concern. Perspect Clin Res 2013 Apr;4(2):144-147 [FREE Full text] [doi: 10.4103/2229-3485.111800] [Medline: 23833741]

6. George SL, Buyse M. Data fraud in clinical trials. Clin Investig (Lond) 2015;5(2):161-173 [FREE Full text] [doi: 10.4155/cli.14.116] [Medline: 25729561]

7. McCurry J. Former Novartis employee arrested over valsartan data. Lancet 2014 Jun 21;383(9935):2111 [FREE Full text] [Medline: 24959638]

8. Yanagawa H, Nokihara H, Yokoi H, Houchi H, Nagai M, Yamashita R, Shikoku Collaborative Group for Promotion of Clinical Trials. Present status and perspectives on future roles of Japanese clinical research coordinators. J Clin Med Res 2018 Dec;10(12):877-882 [FREE Full text] [doi: 10.14740/jocmr3602w] [Medline: 30425759]

9. Nakamoto S. Bitcoin: a peer-to-peer electronic cash system. 2008. URL: https://bitcoin.org/bitcoin.pdf [accessed 2020-05-06]

10. Kuo T, Kim H, Ohno-Machado L. Blockchain distributed ledger technologies for biomedical and health care applications. J Am Med Inform Assoc 2017 Nov 01;24(6):1211-1220. [doi: 10.1093/jamia/ocx068] [Medline: 29016974]

11. Raghavendra M. Can blockchain technologies help tackle the opioid epidemic: a narrative review. Pain Med 2019 Oct 01;20(10):1884-1889. [doi: 10.1093/pm/pny315] [Medline: 30848821]

12. Sylim P, Liu F, Marcelo A, Fontelo P. Blockchain technology for detecting falsified and substandard drugs in distribution: pharmaceutical supply chain intervention. JMIR Res Protoc 2018 Sep 13;7(9):e10163 [FREE Full text] [doi: 10.2196/10163] [Medline: 30213780]

13. Tseng J, Liao Y, Chong B, Liao S. Governance on the drug supply chain via gcoin blockchain. Int J Environ Res Public Health 2018 Dec 23;15(6):1 [FREE Full text] [doi: 10.3390/ijerph15061055] [Medline: 29882861]

14. Zhou L, Wang L, Sun Y. MIStore: a blockchain-based medical insurance storage system. J Med Syst 2018 Jul 02;42(8):149 [FREE Full text] [doi: 10.1007/s10916-018-0996-4] [Medline: 29968202]

15. Wang H, Song Y. Secure cloud-based EHR system using attribute-based cryptosystem and blockchain. J Med Syst 2018 Jul 05;42(8):152. [doi: 10.1007/s10916-018-0994-6] [Medline: 29974270]

16. Mayer AH, da Costa CA, Righi RDR. Electronic health records in a blockchain: a systematic review. Health Informatics J 2019 Sep 30:1460458219866350. [doi: 10.1177/1460458219866350] [Medline: 31566472]

17. Beinke JH, Fitte C, Teuteberg F. Towards a stakeholder-oriented blockchain-based architecture for electronic health records: design science research study. J Med Internet Res 2019 Oct 07;21(10):e13585 [FREE Full text] [doi: 10.2196/13585] [Medline: 31593548]

18. Vazirani AA, O'Donoghue O, Brindley D, Meinert E. Blockchain vehicles for efficient medical record management. NPJ Digit Med 2020 Jan 6;3(1):3. [doi: 10.1038/s41746-019-0211-0]

19. Nugent T, Upton D, Cimpoesu M. Improving data transparency in clinical trials using blockchain smart contracts. F1000Res 2016;5:2541 [FREE Full text] [doi: 10.12688/f1000research.9756.1] [Medline: 28357041]

20. Angeletti F, Chatzigiannakis I, Vitaletti A. The role of blockchain and IoT in recruiting participants for digital clinical trials. 2017 Presented at: 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM); 2017; Split. [doi: 10.23919/softcom.2017.8115590]

21. Benchoufi M, Porcher R, Ravaud P. Blockchain protocols in clinical trials: Transparency and traceability of consent. F1000Res 2017;6:66 [FREE Full text] [doi: 10.12688/f1000research.10531.5] [Medline: 29167732]

22.   Motohashi T, Hirano T, Okumura K, Kashiyama M, Ichikawa D, Ueno T. Secure and scalable mhealth data management using blockchain combined with client hashchain: system design and validation. J Med Internet Res 2019 May 16;21(5):e13385 [FREE Full text] [doi: 10.2196/13385] [Medline: 31099337]

23.   Wong DR, Bhattacharya S, Butte AJ. Prototype of running clinical trials in an untrustworthy environment using blockchain. Nat Commun 2019 Feb 22;10(1):917 [FREE Full text] [doi: 10.1038/s41467-019-08874-y] [Medline: 30796226]

24.   Ichikawa D, Kashiyama M, Ueno T. Tamper-resistant mobile health using blockchain technology. JMIR Mhealth Uhealth 2017 Jul 26;5(7):e111 [FREE Full text] [doi: 10.2196/mhealth.7938] [Medline: 28747296]

25.   Osipenko L. Blockchain's potential to improve clinical trials: an essay by Leeza Osipenko. BMJ 2019 Oct 03:l5561. [doi: 10.1136/bmj.l5561]

26.   Wust K, Gervais A. Do you need a blockchain? 2018 Presented at: 2018 Crypto Valley Conference on Blockchain Technology (CVCBT); 2018; Zug. [doi: 10.1109/cvcbt.2018.00011]

27.   Mackey T, Bekki H, Matsuzaki T, Mizushima H. Examining the potential of blockchain technology to meet the needs of 21st-century Japanese health care: viewpoint on use cases and policy. J Med Internet Res 2020 Jan 09;22(1):e13649 [FREE Full text] [doi: 10.2196/13649] [Medline: 31917371]

28.   Japan External Trade Organization. New regulatory sandbox framework in Japan. 2018 Jun 06. URL: https://www.jetro.go.jp/ ext_images/en/invest/incentive_programs/pdf/Detailed_overview.pdf [accessed 2020-02-23]

29.   Summary of the Amazon EC2 and Amazon EBS service event in the Tokyo (ap-northeast-1) region. 2019. URL: https:/ /aws.amazon.com/jp/message/56489/ [accessed 2020-02-23]

30.   Tsuji K, Ochi E, Okubo R, Shimizu Y, Kuchiba A, Ueno T, et al. Effect of home-based high-intensity interval training and behavioural modification using information and communication technology on cardiorespiratory fitness and exercise habits among sedentary breast cancer survivors: habit-B study protocol for a randomised controlled trial. BMJ Open 2019 Aug 22;9(8):e030911 [FREE Full text] [doi: 10.1136/bmjopen-2019-030911] [Medline: 31444192]

31.   Androulaki E, Manevich Y, Muralidharan S, Murthy C, Nguyen B, Sethi M. Hyperledger fabric. 2018 Presented at: Proc Thirteenth EuroSys Conf EuroSys; 2018; Porto Portugal p. 1-15. [doi: 10.1145/3190508.3190538]

32.   Cleland CL, Hunter RF, Kee F, Cupples ME, Sallis JF, Tully MA. Validity of the global physical activity questionnaire (GPAQ) in assessing levels and change in moderate-vigorous physical activity and sedentary behaviour. BMC Public Health 2014 Dec 10;14:1255 [FREE Full text] [doi: 10.1186/1471-2458-14-1255] [Medline: 25492375]

33.   Chu AHY, Ng SHX, Koh D, Müller-Riemenschneider F. Reliability and validity of the self- and interviewer-administered versions of the global physical activity questionnaire (GPAQ). PLoS One 2015;10(9):e0136944 [FREE Full text] [doi: 10.1371/journal.pone.0136944] [Medline: 26327457]

34.   Bull FC, Maslin TS, Armstrong T. Global physical activity questionnaire (GPAQ): nine country reliability and validity study. J Phys Act Health 2009 Nov;6(6):790-804. [Medline: 20101923]

35.   Vickberg SMJ. The Concerns About Recurrence Scale (CARS): a systematic measure of women's fears about the possibility of breast cancer recurrence. Ann Behav Med 2003;25(1):16-24. [Medline: 12581932]

36.   Momino K, Akechi T, Yamashita T, Fujita T, Hayahi H, Tsunoda N, et al. Psychometric properties of the Japanese version of the Concerns About Recurrence Scale (CARS-J). Jpn J Clin Oncol 2014 May;44(5):456-462. [doi: 10.1093/jjco/hyu032] [Medline: 24711631]

37.   Muramatsu K, Miyaoka H, Kamijima K, Muramatsu Y, Yoshida M, Otsubo T, et al. The patient health questionnaire, Japanese version: validity according to the mini-international neuropsychiatric interview-plus. Psychol Rep 2007 Dec;101(3 Pt 1):952-960. [doi: 10.2466/pr0.101.3.952-960] [Medline: 18232454]

38.   Soldatos CR, Dikeos DG, Paparrigopoulos TJ. Athens Insomnia Scale: validation of an instrument based on ICD-10 criteria. J Psychosom Res 2000 Jun;48(6):555-560. [doi: 10.1016/s0022-3999(00)00095-7] [Medline: 11033374]

39.   Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001 Jul;33(5):337-343. [Medline: 11491192]

40.   Tsuchiya A, Ikeda S, Ikegami N, Nishimura S, Sakai I, Fukuda T, et al. Estimating an EQ-5D population value set: the case of Japan. Health Econ 2002 Jun;11(4):341-353. [doi: 10.1002/hec.673] [Medline: 12007165]

41.   Gueron S, Johnson S, Walker J. SHA-512/256. 2011 Proc Eighth Int Conf Inf Technol New Gen 2011:354-358. [doi: 10.1109/itng.2011.69]

42.   Chowdhury MJM, Ferdous MS, Biswas K, Chowdhury N, Kayes ASM, Alazab M, et al. A comparative analysis of distributed ledger technology platforms. IEEE Access 2019;7:167930-167943 [FREE Full text] [doi: 10.1109/access.2019.2953729]

43.   Kuo T, Zavaleta Rojas H, Ohno-Machado L. Comparison of blockchain platforms: a systematic review and healthcare examples. J Am Med Inform Assoc 2019 May 01;26(5):462-478. [doi: 10.1093/jamia/ocy185] [Medline: 30907419]

44.   Food and Drug Administration. Oversight of clinical investigations? A risk-based approach to monitoring. 2013. URL: https://www.fda.gov/media/116754/download [accessed 2020-02-23]

45.   European Medicines Agencies. Reflection paper on risk based quality management in clinical trials. 2013. URL: https:/ /www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-risk-based-quality-management-clinical-trials_en. pdf [accessed 2020-02-23]

46. Mealer M, Kittelson J, Thompson BT, Wheeler AP, Magee JC, Sokol RJ, et al. Remote source document verification in two national clinical trials networks: a pilot study. PLoS One 2013;8(12):e81890 [FREE Full text] [doi: 10.1371/journal.pone.0081890] [Medline: 24349149]

47. Organisation for Economic Co-operation and Development. 2019. Health at a Glance 2019 OECD Indicators URL: https://www.oecd-ilibrary.org/deliver/4dd50c09-en.pdf?itemId=%2Fcontent%2Fpublication%2F4dd50c09-en&mimeType=pdf [accessed 2020-05-04]

Abbreviations

AWS: Amazon Web Services
AZ: availability zone
EMA: European Medicines Agency
FDA: US Food and Drug Administration
ICH-GCP: International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use–Good Clinical Practice
PMDA: Pharmaceuticals and Medical Devices Agency
SDV: source data verification
SQS: Simple Queue Service

XSL•FO

RenderX