Secure Data Sharing and Prediction with Digital Twin and Blockchain in Healthcare

Yongyi Tang, Kunlun Wang, Dusit Niyato, Jie Li, Octavia A. Dobre, and Trung Q. Duong

***Abstract*—The rapid implementation of the fifth generation wireless networks has driven advances in digital twin (DT) technique, which has been widely used, especially in healthcare. However, the accessibility of data raises concerns about privacy, security, and accountability among participants, affecting overall security and performance of the healthcare DT system. In this paper, we investigate a blockchain-based secure healthcare digital twin data (HDTD) sharing framework to address data privacy concerns. In the blockchain-based secure HDTD sharing model, we propose the access control scheme through cloud storage and attribute encryption to realize the secure data interaction between different users. Based on this, in order to solve the problem of missing valid data due to data tampering or loss with limited resources, we design an HDTD missing value prediction algorithm to meet the real-time requirements of data interaction in DT. The experimental results show that compared with the existing schemes, the proposed blockchain-based secure HDTD sharing scheme has superior performance in improving data security and reducing data interaction delay. The paper outlines key technical challenges and future directions for blockchain-based HDTD research.**

***Index Terms*—Digital twin, healthcare, blockchain, data sharing, missing value, prediction.**

I. Introduction

THE global deployment of fifth generation wireless networks (5G) has stimulated research into sixth-generation wireless networks (6G) to identify their requirements and use cases. A prominent application of 6G is the digital twin (DT), which has significant utility in digital healthcare. A DT integrates a physical entity with its digital counterpart, enabling bi-directional communication that fosters their co-evolution. By harnessing advanced digital technologies, physical characteristics and relationships are translated into high-fidelity virtual models. In healthcare, DT models accurately reflect individuals' physiological and genetic traits, thereby facilitating diverse medical services such as disease prediction and treatment simulation.

While DT has been implemented in some healthcare scenarios using 5G, it has yet to leverage real-time synchronization fully [1]. To enhance data accuracy and reliability, healthcare digital twin data (HDTD) must be aggregated from various patients and institutions. The limited computational resources in typical medical facilities, combined with the extensive nature of DT data, necessitate the distribution of computational tasks across multiple edge nodes. This calls for robust data sharing among medical devices, their respective services, and healthcare data owners to support comprehensive lifecycle management in healthcare. Consequently, substantial data interactions throughout a healthcare DT's lifecycle demand meticulous attention to data integrity, security, and real-time transmission capabilities [2]. Traditional 5G models often rely on centralized data sharing, which may introduce security vulnerabilities in DT applications. Thus, there is increasing interest in employing 6G-enabled blockchain solutions to create a decentralized, reliable framework for medical DTs, addressing the need for high throughput.

Simultaneously, 6G's enhanced connectivity expands the Internet of Things (IoT) ecosystem, enriching healthcare DT data sources via widespread sensing devices. Resource-constrained healthcare devices, like wearable monitors, can use energy-efficient protocols such as Bluetooth for data exchange. However, the complexity of wireless sensor networks, along with the limitations of edge devices, heightens the risk of data loss and tampering, adversely affecting the efficacy of DT models. Therefore, implementing effective recovery strategies for missing data is crucial to ensure the precision of DT model.

A. Motivation

To more effectively demonstrate the significance of this study, we present a shared case of DT data in healthcare scenarios in Fig. 1. To achieve optimal precision and reliability for DT-based healthcare services, data is collected from multiple sources, including disparate devices, patients, and even medical institutions. Furthermore, the necessity for extensive training in DT models results in a significant increase in the amount of DT required data [3].

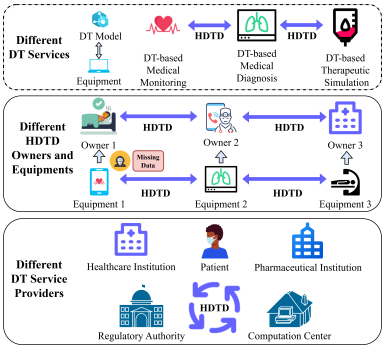


Fig. 1. DT data sharing in healthcare scenarios.

(1) The DT model generated by medical devices offers the unique advantage of providing accurate dynamic data references, facilitating services such as medical monitoring, diagnosis, and treatment simulation [2]. Specifically, shared detection data from HDTD enables physicians to quickly identify changes in patient status and trace physiological

conditions. Multiple specialized healthcare professionals can collaboratively analyze patient data to enhance diagnostic accuracy and simulate various treatment options and their potential effects on the DT, thereby reducing the side effects of medical interventions.

(2) The healthcare DT model utilizes data from various sources to predict health outcomes, offering a comprehensive view that aids healthcare teams in making informed decisions. Due to privacy and security concerns, data silos exist among different equipment and data owners. To expedite the aggregation of DT data and model generation, it is essential to share HDTD from various equipment and data owners in order to meet the data requirements for DT-based services [2].

(3) There exists a collaborative relationship among different DT service providers. Through HDTD sharing, healthcare institutions can promptly obtain real-time health information about patients, thereby forming accurate DTs that aid physicians in developing personalized treatment plans. Pharmaceutical institutions can leverage data from healthcare providers and patients to conduct more precise clinical trial designs and drug development. Computation centers can undertake complex medical computational tasks. Regulatory authorities require access to relevant data to ensure that medical products and services comply with legal and regulatory requirements.

In this scenario, HDTD primarily derives from physiological data provided by wearable devices, such as glucose monitors. However, due to the inherent limitations of the device's operational capabilities, it may be susceptible to data loss resulting from availability and integrity attacks or malfunctions [1]. For instance, the Dexcom G6 glucose sensor experiences a data loss of 5% to 15% under availability and integrity attacks and signal interference. The DT model may inadequately capture real-world conditions, leading to biased decision-making. It is therefore necessary to develop a secure method for sharing HDTD and a recovery mechanism for missing HDTD in order to enhance the quality of DT services. The requirements for DT data in healthcare are summarized in Table I.

TABLE I
REQUIREMENTS AND ANALYSIS OF DT DATA IN HEALTHCARE

Requirements	Analysis
Multi-Party Participation Sharing	To create greater value for stakeholders in healthcare, HDTD must be shared among different parties.
Low-latency Real-time Information Update Sharing	In low-latency healthcare scenarios (e.g., real-time monitoring), sub-second multiple sharing is required from data sources (e.g., sensors) to target recipients (e.g., applications).
Support for Multiple Data Types	HDTD encompasses various data types, including structured, semi-structured, and unstructured data.
Scalability for Exponential Growth	HDTD's large volume necessitates solutions that support its exponential growth.
Integrity Verification	Participants in HDTD sharing cannot tolerate incomplete or forged data, as it may lead to serious medical incidents.
Privacy and Security Measures	HDTD is shared only with authorized parties that have owner permission, and resource-constrained users cannot afford complex computations in privacy protection schemes.
Missing Data Prediction	As a key component of HDTD, sensor data from physical devices is often susceptible to tampering or loss, affecting DT mapping. Therefore, solutions should support predicting missing HDTD.

B. Existing HDTD Security Issues and Solutions

Since HDTD arises from various systems, it necessitates central management, IoT device-based collection, and cloud-based storage in accordance with conventional data-sharing methods. The cloud subsequently incorporates the data with external knowledge to deliver services to DT users. For instance, a cloud healthcare system framework based on DT healthcare was introduced in [4], enabling data sharing among users through platform operators that manage medical cloud services. However, centralized designs may introduce security and trust risks in healthcare DT applications, such as node failures and unauthorized data alterations. While distributed database technology can improve latency and access speed by geographically distributing data, it still adheres to a master-slave architecture, creating a single point of failure that can compromise the entire cluster. Additionally, increased system complexity and communication overhead among nodes can elevate failure rates, failing to meet the robustness and security needs of healthcare scenarios [5].

Blockchain, as a distributed ledger technology, can establish trust among anonymous users and ensure the security and traceability of healthcare data by creating reliable transaction records. Patients maintain control over their data, while medical facilities can grant research institutions access to specific data via smart contracts. In [6], a blockchain-assisted eHealth framework was proposed to facilitate the sharing of medical records among healthcare participants. However, blockchain solutions fail to address the issue of data movement, particularly when patients transfer between different medical institutions, necessitating uninterrupted data sharing across various organizations. This can lead to unnecessary communication overhead. Hence, exploring integrated cloud-based solutions to eliminate redundant data-sharing requests is crucial. The use of cryptography in blockchain to enhance the privacy and security of DT data has been widely validated, including attribute-based encryption (ABE) [7] and proxy re-encryption (PRE) [8]. In [8], the PRE scheme was used to manage manufacturing DT data within cloud and blockchain environments. However, the authority structures in medical contexts are more complex, involving factors like patient identity and doctor authority, and ABE can effectively address these issues. Traditional ABE requires a central authority for key management, limiting its interoperability across heterogeneous IoT systems. Instead, blockchain can replace conventional central servers, allowing information management and data sharing among diverse systems. However, the high computational complexity of ABE can place a burden on edge healthcare DT participants, resulting in the need to outsourced computing resources.

In the healthcare DT lifecycle, data collection and uploading are prerequisites for predicting the DT model. However, limited resources of edge devices constrain the implementation of security measures, potentially exposing the system to integrity attacks that could result in loss and tampering of healthcare data used for DT [1]. Developing an analytical model through machine learning (ML) to directly address missing data can provide an effective solution [9], [10]. However, this approach does not analyze variable relationships and may not be suitable for DT requiring diverse data sources. An alternative method is K-nearest neighbor (KNN) interpolation, which utilizes correlations between examples. A prediction framework based on KNN was introduced in [11] to predict missing diabetes data. Nevertheless, KNN solutions often face inefficiencies due to the time required to determine optimal parameters.

Deep learning techniques, such as recurrent neural networks and gated recurrent units (GRU), offer advanced solutions for time series data and gradient issues, especially in missing medical data. In [12], the authors employed GRU to predict monitoring data loss in diabetic patients. However, these neural network-based solutions also have their own drawbacks, including difficulties in implementation on low-power devices and reliance on large datasets for efficacy.

C. Our Contributions

Based on the above challenges, we propose a blockchain-based healthcare DT architecture that allows for secure sharing of HDTD and meets the processing requirements for time-sensitive data. We investigate the security issues encountered by the above architectures and propose a strategy based on ABE and fine-grained access control to ensure the security of HDTD. In addition, we investigate the resource constraints and massive data interaction problems that the proposed encryption algorithm may encounter in HDTD, and propose a missing HDTD prediction algorithm to balance the real-time requirements and privacy protection of DT data, and further ensure the accuracy of DT model. Finally, simulations are performed to verify the validity of our proposal, and we discuss future research directions.

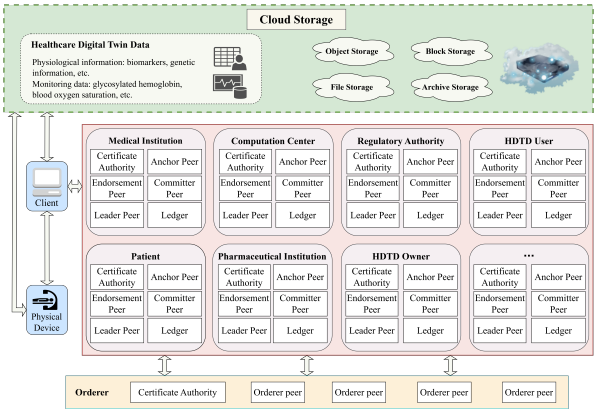


Fig. 2. Framework of blockchain enabled secure sharing of HDTD. Participants and physical devices can establish ownership of the HDTD by sending the hash value, timestamp, and owner's signature to the consortium blockchain network via a client. The network verifies this information and packages it into new blocks. Subsequently, the HDTD is uploaded and stored in the cloud. The HDTD owner can set pricing through tokenization, and sharing occurs based on ABE and smart contracts.

II. Overview of Blockchain Based Framework of Secure HDTD Sharing

The diagram in Fig. 2 illustrates our proposed framework for securely sharing HDTD using blockchain technology. The framework involves various participants, a client, physical device, and cloud storage.

A. Participants of Blockchain

Typically, participants in healthcare DT include healthcare institutions, computing centres, regulatory authorities, patients, pharmaceutical institutions, and others. They serve as crucial contributors to HDTD sharing as well as primary members of a consortium blockchain network. Only authorized users are allowed to join the network to ensure the security of data sharing. The orderer is responsible for ensuring that transactions on the blockchain network are packaged into blocks in a consistent order, and then broadcast to other peers. In addition to the orderer, which consists of a certificate authority and orderer peers, each participant includes a certificate authority, leader peer, endorsement peer, anchor peer, committing peer, and ledger [5], [8]. The peers within a participant are assigned various responsibilities. Anchor peers act as the participants' entry nodes into the network and provide a stable identity. The endorsement peer is accountable for verifying and endorsing transactions. In the blockchain system, transactions need to be approved by a designated number of peers. This shows the validity and legality of the transaction. The committer peer is accountable for the final confirmation and endorsement of transactions. In addition, the leader peer is responsible for the creation of new blocks and informing other peers on the network. The sequencing of transactions from clients is the responsibility of the orderer peer in Hyperledger Fabric [13]. To maintain consistency across all peers, transactions are ordered according to specific rules.

B. Client and Cloud Storage

Clients are responsible for creating transaction proposals, generating transactions, and sending them to orderer peers in our blockchain-based secure sharing framework. The HDTD sharing application runs on the client side, connecting to an organization's peer to communicate with the blockchain network. Clients also upload, preprocess, and encrypt data for physical devices and can manage and distribute the HDTD by connecting to the cloud. Cloud storage provides various services, including object, file, block, and archive storage. Unlike local storage, it offers benefits like scalability, high availability, and multi-device synchronization.

III. Implementation Method

The proposed framework consists of five steps for the interaction between its components: ownership determination of HDTD, HDTD storage, HDTD tokenization, smart contract design, and smart contract operation.

A. Ownership Determination of HDTD

The transmission of the hash value, timestamp, and signature of HDTD by its owner is mandatory for the consortium blockchain network to verify HDTD's ownership. When the blockchain network receives the hash value of HDTD, its immutability is guaranteed owing to the irreversibility of hash values and the incorruptibility of blockchains. Depending on the timing of health-related information, it may be classified as timely or non-timely data. Time-sensitive data comprises chiefly physiological data gathered by sundry sensors, such as glycated haemoglobin. Non-time sensitive data encompasses fixed data, such as genetic information. In the sphere of healthcare DT, the preponderance of the interchanging data

entails time-sensitive data. In light of this, we have formulated two regulations for block creation that enhance recognition and validation of HDTD ownership and meet transaction demands for time-sensitive information. The two regulations are as follows.

1) Level Classification: The HDTD comprises two distinct levels: low-level and high-level, referring to the non-time sensitive and time-sensitive data, respectively. The level of HDTD can be chosen by its owners according to their specific needs. This offers HDTD owners the flexibility to choose the category that suits them best.

2) Fee and Priority: The HDTD owner can pay a service fee to annotate high-level data. The orderer peer collects the service fee and subsequently prioritizes high-level data with higher service fees. Low-level data cannot be annotated using the service fee. Additionally, low-level data in new blockchain blocks must exceed a specific proportion.

The transaction initiated by the HDTD owner to establish ownership on the consortium blockchain network includes several key components. Specifically, it incorporates the hash of the HDTD provided by the data owner, the timestamp indicating when the transaction took place, high-level data that is relevant to the transaction, the service fee paid by the HDTD owner, and the signature of the HDTD owner for authentication purposes.

B. HDTD Storage

The cloud offers HDTD owners adjustable storage amenities and eradicates the obstacle of HDTD local storage quantity restraints. The HDTD owner pays the cloud for the service based on the required storage space and usage time. The blockchain procedure of transaction between the HDTD owner and the cloud is as follows.

1) The HDTD owner submits transactions related to payment for HDTD storage to the consortium blockchain. These transactions include several key elements: the account of the HDTD owner, the account associated with the cloud service, the size of the storage space utilized, the duration for which the storage is required, and the amount of payment involved in the transaction.

2) The transaction sent by the HDTD owner is audited, verified, and then packaged into a new block. When this new block is connected to the blockchain, the transaction is recorded in an unchangeable form.

3) The cloud must communicate with the consortium blockchain network upon receipt of a data storage request to confirm that the owner of the HDTD paid for the storage. If the response from the network confirms that the fees have been paid, the HDTD owner will be granted permission to upload their HDTD.

4) The HDTD owner encrypts HDTD using the ABE encryption algorithm and confuses the key [13]. Finally, it uploads the encrypted HDTD to the cloud.

C. HDTD Tokenization

In blockchain, data tokenization converts real-world data or assets into digital tokens, representing tangible items, rights, certificates, virtual currencies, and more, facilitating tracking, recording, and trading. The HDTD tokenization process involves the following steps:

1) An HDTD owner submits a transaction authorizing the HDTD to the consortium blockchain. This transaction includes the owner's account, the cloud storage address of the HDTD, the HDTD hash, the decryption policy, and the associated price.

2) The consortium blockchain network interacts with the cloud for transaction auditing. Upon signature verification and smart contract compliance, an orderer peer packages the transaction into a new block. Other peers collectively assess the block's validity, including signature checks and policy adherence. Once consensus is reached, the block is added to the blockchain, making the transaction immutable.

3) After the transaction is recorded, the HDTD is finalized as a data asset.

To protect the HDTD owner's interests, we establish three levels of access: "Fully Open," where HDTD is unrestricted; "Partially Open," where HDTD is shared with select authorized parties; and "Not Open," where HDTD is not shared at all. Access permissions are implemented through policies defined by the owner.

D. Smart Contract Design

Smart contracts are programs designed to automatically execute contract terms on blockchain, aiming to automate and enforce agreements without intermediaries [5]. Their creation involves negotiating transaction costs and actions between HDTD owners and users. Once agreed, these terms are converted into code with specific execution outcomes, verified through the blockchain framework to ensure legitimacy and security. Finally, both parties review and sign the agreement before its deployment on the blockchain.

E. Smart Contract Operation

If the smart contract meets the specified conditions, the results will be executed automatically, triggering the application programming interfaces of cloud storage and blockchain to complete communication via network protocols. The input comprises transactions created by both HDTD owners and consumers. The output shall detail the execution status of the smart contract. The smart contract operation process can be summarized as an HDTD user sending transaction regarding data payment to the blockchain network. As the transaction is entered, the state of the smart contract changes and the smart contract begins execution. The HDTD owner then sends the data to the HDTD user via the cloud. To ensure the confidentiality of data, the transmission process uses ABE, as shown in Fig. 3. The HDTD transmission process is as follows.

1) An HDTD owner generates a random key k, and then use it to encrypt its data with a symmetric algorithm as $D = Enc_k(HDTD)$. When k is encrypted using a specific policy, the HDTD owner can ensure that only HDTD users who meet the policy can access the data. The HDTD owner can complete the encryption itself, but if the HDTD owner needs

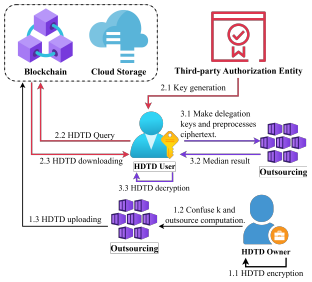


Fig. 3. HDTD sharing scheme based on ABE.

to outsource encryption calculations, the HDTD owner need to confuse k to prevent the outsourced calculation undertaker from illegally obtaining data [13]. We define the confused k as C and the access policy as $Policy$. We define C and the $Policy$ for outsourcing encryption as CT, which contains the hidden data key, data attributes, the third-party authorization agency that signed the attributes, and the decryption time of the authorized user. After receiving the data upload request, the outsourcing computing undertaker executes outsourcing encryption with CT. Finally, the HDTD encrypted and uploaded by the HDTD owner is expressed as $D = Enc_{CT}(HDTD)$.

2) HDTD users who want to obtain data must first obtain their user attributes from a third-party authorization agency (e.g, encryption service provider) and prove that they have the corresponding permissions. The third-party authorization agency will calculate the key to obtain $K_{uid,u}$ and $L_{uid,u}$ for the attribute u of each HDTD user uid. The HDTD user then performs an HDTD query to the blockchain network and obtains the ciphertext.

3) The HDTD user uses $K_{uid,u}$ and $L_{uid,u}$ to decrypt the ciphertext and obtain the data key k. Then, the HDTD user uses k to decrypt the data to obtain the original HDTD. Similarly, the HDTD user can also outsource decryption calculations. First, it needs to generate an authorization key to the outsourcing computing undertaker, and then preprocess the ciphertext sent to the outsourcing computing undertaker. The outsourced calculation undertaker uses the authorization key to complete the calculations and transmits the median result to the HDTD user. Finally, the HDTD user can easily decrypt the ciphertext.

During HDTD sharing, the cloud interacts only with ciphertext, avoiding access to private data. After receiving encrypted data, HDTD users relay related transactions to the blockchain network. Once verified and recorded, payment is automatically executed within the smart contract.

IV. Proposed Missing Value Design of HDTD

In Section III, we have achieved secure sharing of blockchain-based HDTD. In addition, to mitigate the impact of data loss on DT modeling and further improve the quality of healthcare service, we propose a prediction method for HDTD called prediction_LSH, which is derived from locality-sensitive hashing (LSH) technology [14] and can be summarized as follows:

Step 1: Modelling Healthcare Surveillance Data. In this step, we create a DT data model using monitored human physiological data, taking diabetes monitoring data as an example. We construct a matrix to represent this data over time, where each row corresponds to daily monitoring data from sensors and each column represents different time windows across various days. To account for significant variations in the HDTD over time, we normalize the matrix so that all values fall within the range of 0 to 1. Each normalized row is treated as an individual entry in the model.

Step 2: Entry Index Creation. Following the DT data model, we employ LSH to create an index for each entry. Each element of the normalized matrix is transformed into a vector before indexing. This process involves generating multiple random vectors with values between -1 and 1, where the number of dimensions corresponds to the number of columns in the matrix. We then compute the dot products of these random vectors with each entry's corresponding vector. If the result of the dot product is positive, the sub-index for that entry is set to 1. Otherwise, it is set to 0. This results in a series of sub-indices that together form a vector representing the entry's index. We repeat this for all entries to create a comprehensive index representation.

Step 3: Missing Value Prediction. We generate a hash table for all item indices obtained in Step 2 and replicate this process multiple times to create several hash tables. According to the principles of LSH, if two index values match in any of the hash tables, it indicates that the corresponding entries are similar. This similarity indicates that the associated HDTD is consistent over the observed days, allowing effective prediction of missing values within the specified time window and accurately forecasting the missing data points for the DT model.

V. Performance Evaluation

To validate the proposed framework and algorithm, we analyze the security of the HDTD shared framework in Subsection A. Additionally, we develop an evaluation system using Hyperledger Fabric and Alibaba Cloud [8] to assess time sensitivity. In Sub-section B, we evaluate the efficiency of the proposed HDTD algorithm for predicting missing values using medical monitoring data simulations from the BIG IDEAs Lab [15]. The study employs the Dexcom G6 continuous glucose monitor and the Empatica E4 wearable device, which measure interstitial glucose at five-minute intervals and electrodermal activity at 4 Hz. Data collected by these sensors is shared through a decentralized application developed in Golang.

A. Performance and Analysis of HDTD Sharing Framework

1) Security Analysis: The essential prerequisites for the secure sharing of DT data have also been described in [3], including data accessibility, integrity, confidentiality, and the capacity to safeguard privacy. Compared to conventional centralized methodologies that rely on authoritative third parties and distributed databases that rely on a master-slave architectural framework, the blockchain-based distributed methodology utilized in our framework has superior security [13].

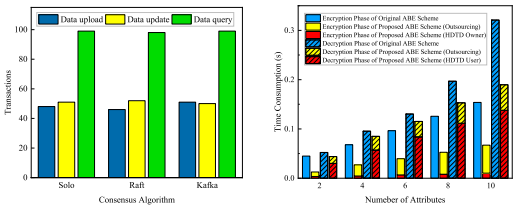
In terms of availability, our framework enhances security against denial-of-service attacks by eliminating single points

of failure, as the HDTD shared service only fails when all blockchain nodes are unavailable. To ensure data integrity and mitigate data tampering attacks, HDTD owners should hash the data before uploading it to the cloudand encapsulate ciphertext and hash when it is published to the blockchain. HDTD users can then verify the source and integrity using signatures and hashes. To protect data confidentiality and privacy from unauthorized access, our framework employs an ABE encryption scheme for encrypted data in the cloud, while outsourced computation utilizes a confused key and access policy. The security of the ABE encryption scheme has been fully demonstrated in [13]. The design of outsourcing encryption and decryption tasks has been incorporated into the framework, but this process does not reveal additional information to potential attackers. Consequently, the underlying security assumptions remain valid. Furthermore, in our framework, smart contracts facilitate the automatic execution of transactions, thereby reducing the risk of human error. Determining ownership can detect of any attempt to change owner information, effectively preventing fraudulent behavior [13].

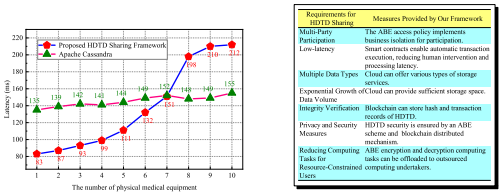
Fig. 4(a) illustrates the secure throughput achievable by our proposed framework under the three default consensus mechanisms of Hyperledger Fabric: Solo, Raft, and Kafka [13]. In this scenario, ten physical medical devices act as clients, while three malicious devices exist to perform blockchain operations of data upload, update, and query. Notably, our framework demonstrates similar secure throughput across different consensus mechanisms, reflecting its robustness. Because uploading and updating blood glucose and skin electrical activity data requires changes across all blockchain nodes, and data queries only retrieve results and are not committed, performance differences occur.

Fig. 4(b) shows the computational costs of encryption and decryption for the outsourced ABE scheme simulated with the Charm-Crypto 0.50 framework [7], we can observe that a linear increase in cost with the number of attributes due to the additional computational steps. The higher cost of decryption stems from the need to verify user attributes for access compliance, which requires additional computations to associate the user's private key with the ciphertext. Furthermore, our proposed ABE scheme achieves an average reduction of 57.6% in computational time compared to original ABE scheme.

2) Time Sensitivity Evaluation: Our primary objective is to facilitate practical real-time data sharing. Thus, our proposed solution provides a streamlined approach for evaluating data without requiring HDTD encryption and decryption processes, as compared to the methodologies in Section III. Fig. 4(c) shows the latency of our proposed blockchain-based sharing framework when sharing blood glucose and electrodermal activity data between one to ten physical medical devices. Obviously, as the number of physical medical devices increases, so does latency of HDTD sharing in our framework, mainly due to the fact that the consensus mechanism typically requiring more time to validate transactions. In contrast, the latency of the distributed database solution based on Apache Cassandra [5] remains relatively stable as the system can introduce additional nodes to handle requests, thus distributing the load. While the distributed database solution offers advantages in terms of performance and data processing speed, it may not meet the security, transparency, and trust requirements that are essential in healthcare scenarios. The sub-second latency in the simulation indicates that our proposed blockchain-based sharing framework allows for multiple HDTD sharing per second between different devices, thus meeting the exponential time sensitivity requirements of HDTD [2]. Fig. 4(d) demonstrates our proposed framework meets the secure HDTD sharing requirements listed above.

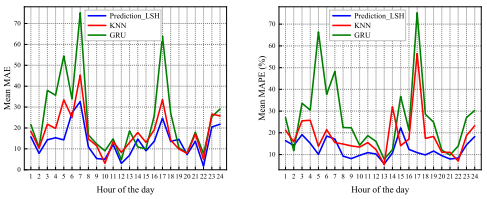


(a) Secure throughput. (b) Computation cost of ABE.

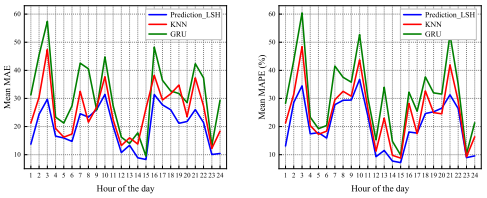


(c) HDTD shared latency. (d) Analysis of sharing performance.

Fig. 4. Performance of HDTD sharing framework.



(a) Blood glucose MAE. (b) Blood glucose MAPE.



(c) Electrodermal activity MAE. (d) Electrodermal activity MAPE.

Fig. 5. Prediction accuracy of vital sign value with respect to monitor hours.

B. Performance and Analysis of Prediction_LSH

To demonstrate the Prediction_LSH scheme's effectiveness, we conduct experiments using real blood glucose variability and wearable device data from the BIG IDEAs Lab [15]. The data used in the simulation consisted of interstitial glucose concentrations and electrodermal activity. Evaluation indicators include time cost, mean absolute error (MAE) and mean absolute percentage error (MAPE). Our goal is to pursue the

lowest values of these three indicators as much as possible. Comparative methods use GRU [12] and KNN [11].

1) Prediction Accuracy Evaluation: High prediction accuracy is crucial for healthcare DT to accurately assess physical entities. The missing value prediction algorithm utilizes a hash function value of 25 and a hash table size of 30. This study evaluates the prediction accuracy of interstitial glucose and electrodermal activity across different monitoring durations by partitioning daily data into hourly windows.

As shown in Fig. 5, the Prediction_LSH scheme outperforms GRU and KNN in accuracy across different time windows. Its superior performance stems from LSH's ability to identify the most similar days based on daily interstitial glucose and electrodermal activity data, demonstrating strong applicability for periodic HDTD.

2) Time Cost Analysis: Our Prediction_LSH scheme is time-efficient compared to complex deep learning methods, as it allows for offline HDTD index creation with a time complexity of $O(1)$, leading to an overall algorithmic complexity of $O(n)$ that meets the low time cost requirements of HDTD.

VI. Conclusions and Directions for Future Research

This article discussed the need for data sharing in the healthcare domain and proposed a secure sharing framework and implementation methodology for supporting multi-party secure data interactions. The framework was designed using blockchain, cloud storage, and outsourced encryption-decryption computation for ABE. To address the problems of data tampering and data loss caused by limited resources, we proposed a missing value prediction algorithm based on time-aware LSH. The study demonstrated the effectiveness of our proposed approach for DT in healthcare scenarios. Additionally, there are still some challenges and potential directions for future research:

1) Development of Hybrid Models in ML: The development of hybrid models that combine the advantages of different ML algorithms can improve prediction performance. However, since data on the blockchain can come from multiple sources of different qualities, there is a risk to the model's generalization capability [10]. To address this issue, techniques such as cross-validation or regularization can be considered to enforce model simplification and enhance generalization ability.

2) Processing Non-Periodic Time Series DT Data: ML anomaly detection can estimate true values from altered or incomplete non-periodic data, but this presents challenges for users with limited computing resources. In the future, task offloading strategies can be considered in edge computing scenarios [1].

3) Data Consistency: The data uploaded by the user is not synchronized, resulting in inconsistency between DT simulations and physical entities. Although controlling upload rate can reduce large-scale asynchronous upload and mitigate the impact of traffic attacks, it cannot meet the real-time requirements of DT. It may be beneficial in the future to consider utilising edge artificial intelligence for preliminary data processing and analysis at edge nodes in close proximity to the data source. This could potentially reduce the volume of data that needs to be uploaded to the central server, thereby enhancing response speed [2].

Acknowledgment

This work was supported by the Open Project Program of Guangxi Key Laboratory of Digital Infrastructure under Grant GXDIOP2023002, as well as by the National Research Foundation, Singapore, and the Infocomm Media Development Authority through its Future Communications Research & Development Programme. Additional support was provided by Defence Science Organisation (DSO) National Laboratories under the AI Singapore Programme (FCP-NTU-RG-2022-010 and FCP-ASTAR-TG-2022-003), the Singapore Ministry of Education (MOE) Tier 1 (RG87/22), and the NTU Centre for Computational Technologies in Finance (NTU-CCTF). Furthermore, the work of T. Q. Duong was partially supported by the Canada Excellence Research Chair (CERC) Program (CERC-2022-00109).

References

[1] L. U. Khan *et al.*, "Digital-Twin-Enabled 6G: Vision, Architectural Trends, and Future Directions," *IEEE Commun. Mag.*, vol. 60, no. 1, pp. 74–80, Jan. 2022.

[2] J. Chen *et al.*, "Networking Architecture and Key Supporting Technologies for Human Digital Twin in Personalized Healthcare: A Comprehensive Survey," *IEEE Commun. Surv. Tutor.*, pp. 1–1, Sept. 2023.

[3] M. Zhang *et al.*, "Digital Twin Data: Methods and Key Technologies," *Digital Twin*, vol. 1, p. 2, Feb. 2022.

[4] Y. Liu *et al.*, "A Novel Cloud-Based Framework for the Elderly Healthcare Services Using Digital Twin," *IEEE Access*, vol. 7, pp. 49 088–49 101, Apr. 2019.

[5] S. Bergman *et al.*, "Permissioned Blockchains and Distributed Databases: A Performance Study," *Concurrency Comput. Pract. Exp.*, vol. 32, no. 12, p. e5227, Jun. 2020.

[6] S. Cao *et al.*, "Toward Secure Storage in Cloud-based eHealth Systems: A Blockchain-Assisted Approach," *IEEE Netw.*, vol. 34, no. 2, pp. 64–70, Apr. 2020.

[7] B. Gong *et al.*, "Towards Secure Data Storage in Web 3.0: Ciphertext-Policy Attribute-Based Encryption," *IEEE Netw.*, pp. 1–1, Oct. 2023.

[8] W. Shen *et al.*, "Secure Sharing of Big Digital Twin Data for Smart Manufacturing Based on Blockchain," *J. Manuf. Syst.*, vol. 61, pp. 338–350, Oct. 2021.

[9] G. Wang *et al.*, "Tackling Missing Data in Community Health Studies Using Additive LS-SVM Classifier," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 2, pp. 579–587, Dec. 2018.

[10] T. Emmanuel *et al.*, "A Survey on Missing Data in Machine Learning," *J. Big Data*, vol. 8, pp. 1–37, Oct. 2021.

[11] S. Suyanto *et al.*, "A New Nearest Neighbor-Based Framework for Diabetes Detection," *Expert Syst. Appl.*, vol. 199, p. 116857, Aug. 2022.

[12] T. Zhu *et al.*, "Blood Glucose Prediction for Type 1 Diabetes Using Generative Adversarial Networks," in *CEUR Workshop Proceedings*, vol. 2675, Santiago de Compostela, Spain, Aug. 2020, pp. 90–94.

[13] X. Wei *et al.*, "Secure Data Sharing: Blockchain-Enabled Data Access Control Framework for IoT," *IEEE Internet Things J.*, vol. 9, no. 11, pp. 8143–8153, Sept. 2022.

[14] L. Kong *et al.*, "LSH-Aware Multitype Health Data Prediction with Privacy Preservation in Edge Environment," *WWW*, pp. 1–16, Sept. 2021.

[15] C. Peter *et al.*, "BIG IDEAs Lab Glycemic Variability and Wearable Device Data (version 1.1.1)," PhysioNet, https://doi.org/10.13026/73s9-cw03, accessed Aug. 2024.

Author Biography

Yongyi Tang (51255904070@stu.ecnu.edu.cn) is currently pursuing a Master's degree at the School of Communication and Electronic Engineering, East China Normal University, Shanghai, China.

Kunlun Wang (klwang@cee.ecnu.edu.cn) is a Professor with the School of Communication and Electronic Engineering, East China Normal University, Shanghai, China.

Dusit Niyato (dniyato@ntu.edu.sg) is currently a professor in the School of Computer and Data Science, Nanyang Technological University, Singapore.

Jie Li (lijiecs@sjtu.edu.cn) is with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China, where he is a Chair Professor.

Octavia A Dobre (odobre@mun.ca) is a Professor and Canada Research Chair Tier-1 at Memorial University, Canada. She is an elected member of the European Academy of Sciences and Arts, a fellow of the Engineering Institute of Canada, a fellow of the Canadian Academy of Engineering, and a fellow of the Royal Society of Canada.

Trung Q. Duong (tduong@mun.ca) is a Canada Excellence Research Chair and professor at Memorial University, Canada. He is also an adjunct professor at Queen's University Belfast, UK and a Research Chair of Royal Academy of Engineering. He is also a visiting professor at Kyung Hee University, South Korea under the Eminent Scholar program.