The following publication M. Li, A. R. Harish, C. Yu, Y. Yu, R. Y. Zhong and G. Q. Huang, "Blockchain-Based Medical Data Asset Sharing Framework for Healthcare 4.0," in IEEE Transactions on Industrial Informatics, vol. 21, no. 4, pp. 2779-2788, April 2025 is available at https://doi.org/10.1109/TII.2024.3488795.

Blockchain-Based Medical Data Asset Sharing Framework for Healthcare 4.0

Ming Li, *Senior Member, IEEE*, Arjun Rachana Harish, Chenglin Yu, Ying Yu, Ray Y. Zhong, *Senior Member, IEEE,* George Q. Huang, *Fellow, IEEE*

***Abstract*—To promote the sharing of medical data assets (MDAs) in a more secure and sustainable manner, this paper presents a blockchain-based MDA sharing framework. The contributions of this paper are threefold. First, we designed a layered-architecture to decouple the privacy-preserving responsibilities among technologies considering the incentive rewarding and parallelization of execution. Second, we introduce zero-knowledge proofs in smart contracts with a group signature to construct a supervisory privacy-preserving sharing mechanism, which can be executed in a decentralized environment to protect the privacy of MDAs. Third, we introduce an incentive mechanism that motivates MDA sharing by capturing the decentralized features of the participants to deliver fair rewards. The experiments show that our framework achieves a comprehensive privacy protection on sharing MDAs, comparing with single blockchain sharing schema, with only 2.2% sacrifice on TPS (throughput/second). Moreover, our framework has better potential for large-scale application due to the paralleled execution on ZKP (zero-knowledge proof)-based smart contracts.**

***Index Terms*— Blockchain, Medical data asset, Sharing mechanism, Incentive mechanism, Healthcare 4.0**

This work is partially supported by the Natural Science Foundation of Guangdong Province, China (No.2023A1515011203), the Zhejiang Provincial Natural Science Foundation of China (No. LQ23G020007), two grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. PolyU/15208824 and No. T32-707/22-N). (Corresponding author: Ying Yu)

Ming Li and George Q. Huang are with the Department of Department of Industrial and Systems Engineering, with the Research Institute for Advanced Manufacturing, and with the Research Centre for Digital Transformation of Tourism, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China. (ming.li@polyu.edu.hk; gq.huang@polyu.edu.hk)

Arjun Rachana Harish was with the Department of Data and Systems Engineering, The University of Hong Kong, China. He is now with the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China. (arjun.rachanaharish@polyu.edu.hk)

Chenglin Yu and Ray Y. Zhong are with the Department of Data and Systems Engineering, The University of Hong Kong, China (cl0415@connect.hku.hk; zhongzry@hku.hk)

Ying Yu is with College of Economics and Management, Zhejiang Normal University. (yuying429hk@126.com)

I. Introduction

The healthcare industry is currently heading into a new era, motivated by the revolution of Industry 4.0 and Logistics 4.0. This era has been defined as Healthcare 4.0 [1], which promotes the intelligent transformation of medical informatization. In such a context, the Internet of Things (IoT), radio frequency identification (RFID), medical wearables, and other smart devices have been integrated with traditional healthcare equipment and systems to construct a ubiquitous perceptual environment so that extensive data can be generated and collected to increase the accuracy of medical diagnosis and treatment. Additionally, cloud computing, big data analysis and artificial intelligence (AI) have been introduced to facilitate smart and connected healthcare delivery [2].

Medical data has been recognized as the cornerstone of Healthcare 4.0 [3]. The massive amounts of available medical data have promoted the acquisition and representation of knowledge for healthcare analysis and diagnosis. In recent years, machine learning, especially deep learning, has rapidly developed into a medical data analysis hotspot [4]. Machine learning is capable of automatically identifying the underlying diagnostic features of diseases or predicting treatment features from medical data. In turn, this capability has stimulated the commercialization and industrialization of machine learning applications in Healthcare 4.0. Medical data, which has significant value, possesses beneficial features. First, it has a definite provider, e.g., a patient or a medical institution. Second, medical data is valuable because of its potential academic and commercial uses. Third, medical data can be authorized or transferred among individuals and organizations. Hence, effective medical data may be recognized as a kind of medical data asset (MDA), which is the basic unit of useful information that can be labeled, indexed, stored, retrieved and manipulated based on the observed needs of medical data management practices [5].

However, MDA sharing does not come without challenges arising from medical privacy issues [6]. First, associating the identity information of an owner with the MDAs being shared is deemed sensitive, thereby inhibiting sharing. However, the identity of the owner of an MDA should be traceable by a specific manager on a limited basis in case a potential disease warning is encountered or an epidemiological investigation is conducted. Second, on the basis of identity privacy, the process privacy of MDAs is also a concern. The ownership of an MDA should be able to be verified in a more secure way to prevent MDA leakage. Furthermore, an ownership claim can be verified by an intended party without showing the MDA itself. Moreover, when an MDA is authorized to a third party, this process should be conducted with the minimal amount of data needed, and the authorized party should be able to provide

effective authorization to obtain access to this MDA.

Additionally, several real-life concerns make the above challenges more complicated. The major concern is the distribution of MDAs. A decentralized MDA organization paradigm has been strongly suggested in recent research. MDA stakeholders also follow a weakly centralized distribution. Since sharing behavior is highly associated with stakeholders and MDAs, it is necessary to consider both decentralized features when designing a sharing mechanism. Hence, technology is no longer the factor restricting MDA sharing; rather, it brings greater convenience to sharing. Currently, participants show interest in using the data made available through sharing platforms but refrain from sharing their own private MDAs. This behavior is attributed to the lack of incentives, as perceived by MDA contributors [7]. A large segment of the current literature uses the participation level or the amount of data uploaded by contributors to determine rewards [8], [9]; thus, these platforms limit themselves to the quid pro quo of compensating other contributors. Such situations produce platform overcrowding with irrelevant and poor-quality data and deter active participation in sharing platforms [10]. Hence, the need for a reward mechanism that takes the specific attributes that drive the benefits and costs of individual MDA contributors into account has become imperative. First, we should measure the factors that drive the MDA sharing behaviors of contributors and then reward/compensate participants with fair value.

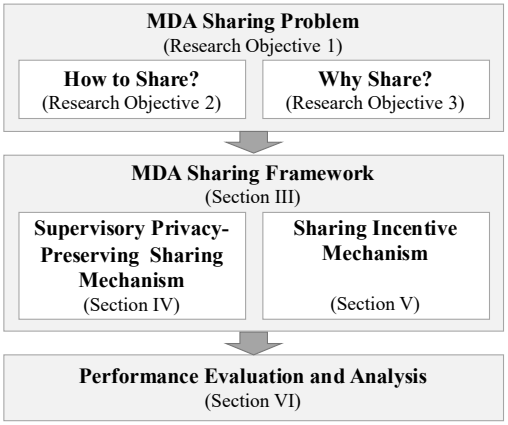


Fig. 1. Research roadmap

In summary, the challenging scenario is to shield the privacy concerns of MDA owners (i.e. patients or medical institutions) and promote their willingness when contributing their MDAs for third-party organizations (i.e. data analyzers). Thus, the underlying research questions are extracted as follows.

1) How can provable privacy guarantees be associated with MDA sharing to eliminate concerns over both identity and management privacy in a decentralized manner?

2) How can a reasonable incentive mechanism be designed to motivate MDA sharing under a decentralized environment while taking the specific drivers of sharing behavior into account?

To address these questions, this paper presents a blockchain-based MDA sharing framework as an integrated solution to facilitate the sharing of MDAs in a decentralized healthcare network, and the following research objectives are explored:

1) To propose an overall technical sharing framework for decentralized MDA sharing.
2) To design a privacy-preserving sharing mechanism to shield both identity and process information during MDA sharing.
3) To build an incentive mechanism and analyze its effectiveness in promoting and motivating MDA sharing behavior by the participants.

To achieve these objectives, the research pathways are depicted in Fig. 1. The blockchain-enabled MDA sharing framework is designed as an integrated solution. Blockchain is adopted as the underlying technology due to two reasons. First, the decentralized feature of blockchain can shield the concerns of MDA providers (MDAPs) that no one can solely dominate their shared MDAs. Otherwise, there must be a party, who will hold and manipulate all shared MDAs and it is difficult for this party to obtain sufficient trust from MDAPs because self-proof of innocence and self-regulation are usually weak. Second, blockchain is suitable for decoupling the relationship among multiple stakeholders [11]. This will facilitate the establishment of transparent and creditable cooperation among MDAPs, MDA storage provider, all kinds of data users. Our sharing framework provides a novel privacy-preserving solution to decouple the responsibilities of MDA storing, sharing and incentives. Based on this sharing framework, two enabling mechanisms are designed. The first one is a supervisory privacy-preserving sharing mechanism, which allows individuals to register and authorize their MDAs and enables authorized parties to access them in a privacy-preserving manner. This mechanism provides provable guarantees to eliminate concerns over identity and process privacy for total sharing management. Furthermore, it also maintains limited and reliable MDA identity traceability. The second mechanism is a sharing incentive mechanism, which enhances the sustainability of MDA sharing behaviors. We build a model for an individual-based reward mechanism that facilitates the codification of MDAs into a blockchain platform. Blockchain technology provides (1) convenient recording of information, with ownership being tied to the contributors, and (2) ease of access for users while tracking their use of MDAs, as prescribed by the codification strategy [12], [13]. Finally, we design and conduct a series of performance tests to verify and optimize the supervisory privacy-preserving sharing mechanism and analyze the impact of factors such as privacy, cost, and the interdependence of MDAs on the sharing behavior of MDA contributors.

The novelty of this paper is threefold. First, a layered architecture is adopted in the MDA sharing framework to decouple the relationship among multiple technologies and perform a fine-grained responsibility assignment for practicing privacy-preserving MDA management. Second, zero-knowledge proof (ZKP) has been employed for constructing MDA sharing with least privacy disclosure, with group signature approach for ensuring limited identity traceability. Third, with the above traceable identity, an individual-based reward mechanism through the blockchain is designed to encourage and stimulate MDA sharing behaviors.

The rest of the paper is organized as follows. Section II reviews related works about general privacy protection. Section III illustrates the overall framework of MDA sharing. Section IV specifies the supervisory privacy-preserving sharing mechanism. Section V describes and proves the sharing

incentive mechanism. Section VI evaluates both mechanisms, and Section VII draws the conclusions of this study and notes future work ideas.

II. Related Works

Studies on privacy protection have long been concerned with the wide applications of information systems and increasing privacy protection awareness. The research on privacy protection can be categorized into three directions: protection through policies, protection through statistical methods, and protection through technology.

Studies on privacy protection policy generally utilize laws and regulations to restrain the collection and management of private data, assuming that policies have significant effects on reasonable people [14]. In this context, technology can be introduced as a supplementary means of collecting, storing, retrieving, and analyzing private data. For the digitalization of privacy policies, two privacy policy description languages, the platform for privacy preferences (P3P) and the enterprise privacy authorization language (EPAL), have been proposed to make such policies more readable and enforceable. P3P was further standardized by the World Wide Web Consortium (W3C) so that it could be widely deployed between online websites and end users [15]. This approach has been employed to formulize privacy practices into a definite format so that both parties can identify and interpret their privacy policies to determine whether the privacy preferences of end users match the privacy requirements of websites. EPAL uses a logic program model to ensure abstract-level access control for privacy so that it can only be invoked when specific user roles or conditions indicated by the EPAL format are satisfied [16]. To enhance the compatibility of policy languages across different platforms, the eXtensible Access Control Markup Language (XACML) was designed as a platform-independent language using the eXtensible Markup Language (XML) to enhance its structurization. Moreover, role-based access control models have attracted more attention in the field of privacy policy in recent years [17]. For example, [18] proposed a knowledge-constrained role-based access control model to reduce unnecessary access to private medical information. [19] also contributed an attribute-based access control model using an access control markup language based on XML to achieve fine-grained access control for cloud-based electronic health records. Recently, blockchain technology has been adopted in access control. [20] designed a digital asset access control solution to migrate existing e-health systems to a unified blockchain-based model so that digital assets could be accessed seamlessly and securely. Moreover, [21] provided an effective method for conducting data interoperation and synchronization between a traditional relational database and a distributed ledger while considering secure access control. Upon blockchain, smart contract has been integrated to perform attribute-based access control due to definite execution feature after deployment for achieving security and privacy commitment as agreed for cloud-edge computing of IoT systems [22]. Afterwards, zero-knowledge proof was also explored with smart contract to realize to transparent access policies evaluation without disclosing the value of such sensible attributes under XACML [23]. In general, policy-based privacy protection studies have sufficiently explored policy implementations to protect user privacy from unconventional collection, unauthorized access, and accidental disclosure. The assumption is that the accessed party should hold the data. However, in MDA sharing, most of the MDA owners would be patients who do not have the responsibility to keep the MDAs after their healthcare diagnosis. Thus, policy-based privacy protection studies, especially using access control approaches, has limitation on disposing privacy protection among multiple parties.

To prevent excessive privacy disclosure, statistical methods have been introduced to obfuscate original private data. Such obfuscation has been carried out in two ways. The first approach involves the interpolation or obfuscation of original data to maintain its statistical characteristics so that the data receiver cannot infer sensitive data about individuals [24]. This method depends more on the data analysis requirements of data receivers in performing statistical computations. Moreover, the statistical features of MDAs, especially medical images, are difficult to extract, which becomes a limitation of this approach. The second approach focuses on data anonymization to shield identities or identifiable data through a trusted party [25]. In this regard, studies use statistical models to construct identity-independent disclosure mechanisms for specific attack models or data mining and learning models. However, these methods only provide effective protection for specific data analysis models, and the potential attacker can make use of marginal data and background knowledge to overflow the identity information.

Cryptography has attracted considerable attention for resolving privacy-preserving issues. Asymmetric and symmetric cryptographic algorithms are the most lightweight methods for securing private communications [26], but they are not capable of supporting complicated MDA operations because a single key is difficult to manage among multiple stakeholders. Thus, complex encryption mechanisms have been designed. Attribute-based encryption (ABE) maintains a series of keys with labeled descriptive attributes and takes one of the keys for the given plaintext; only the correct receiver with a matching key can decrypt the ciphertext [27]. Two enhancing mechanisms, the key policy and ciphertext policy, were proposed based on ABE to optimize encryption performance under different application scenarios. Recently, with the integration of blockchain technology, secure multiparty computing (SMC) has received increased attention [28]. SMC research can be categorized into three directions: oblivious transfer, oblivious polynomial evaluation, and homomorphic encryption. Homomorphic encryption is regarded as the most promising privacy preservation method, and its integration into blockchain has been recommended to address privacy issues [29]. For example, [30] presented a lightweight privacy-preserving protocol based on a labeled homomorphic encryption approach to protect IoT data between data owners, a third-party cloud service and data users. [31] proposed three data protection methods based on differential privacy and homomorphic encryption to protect existing data and conduct model aggregation in federated learning. Besides homomorphic encryption, blockchain was also taken for distributing cryptographic keys among multiple stakeholders. For instance, [32] proposed the using of smart contract to conduct group key

TABLE I
SUMMARY OF RELATED WORKS

	Problem Domain				Methodology Domain						
Paper	**Access Privacy**	**Content Privacy**	**Identity Privacy**	**Privacy Sharing Incentive**	**Language**	**Attribute Control**	**Blockchain**	**Smart Contract**	**Zero-knowledge Proof**	**Statistical Method**	**Cryptography**
[19]	√	×	×	×	√	×	×	×	×	×	×
[20]	√	×	×	×	√	√	×	×	×	×	×
[21]	√	×	×	×	√	√	×	×	×	×	×
[22]	√	×	×	×	×	√	×	×	×	×	×
[23]	√	×	×	×	√	√	×	×	×	×	×
[13]	√	×	×	×	×	√	√	×	×	×	×
[24]	√	×	×	×	×	√	√	×	×	×	×
[25]	√	×	×	×	×	√	√	√	×	×	×
[26]	√	√	×	×	√	√	√	√	√	×	×
[27]	×	√	×	×	×	×	×	×	×	√	×
[28]	×	×	√	×	×	×	×	×	×	√	×
[29]	×	√	×	×	×	×	×	×	×	×	√
[30]	×	√	×	×	×	√	×	×	×	×	√
[31]	×	√	×	×	×	×	√	×	×	×	√
[32]	√	√	×	×	×	√	√	×	×	×	√
[33]	×	√	√	×	×	√	×	×	×	×	√
[34]	×	√	×	×	×	×	×	×	×	×	√
[35]	√	×	√	×	×	×	√	√	×	×	√
[36]	×	√	×	×	×	√	√	×	×	×	√
[37]	×	√	√	×	×	×	√	×	√	×	×
This Paper	√	√	√	√	×	×	√	√	√	×	×

distribution with group signatures for authenticating user roles in vehicular ad hoc networks. Then, considering the more complex mesh-to-mesh network, blockchain is adopted as information sharing method for generating and verifying key pairs for cryptography. Concerning to data sharing, [33] integrated blockchain with ABE for addressing the policy hiding and keyword search when sharing medical data. [34] introduced ZKP as the blinding mismatching mechanism on blockchain to bridge data sharers and requesters with privacy and security protection.

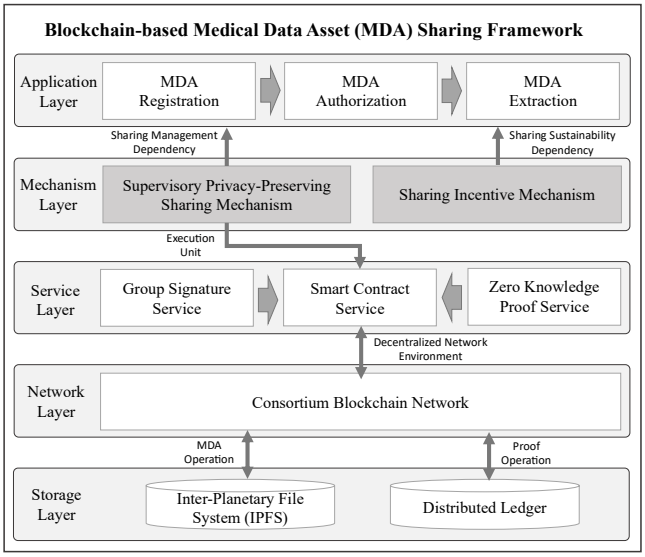


Fig. 2. Blockchain-based MDA sharing framework

However, there approaches are difficult to migrate to address our MDA sharing scenario as summarized in the problem domain of Table I, which considers the full lifecycle privacy protection of MDA circulation. Moreover, the scenario domain drives the innovation of methodology domain. Thus, this study aims to take the advantages of blockchain, smart contract and ZKP to be coupled to bridge the above gap. Preliminary studies have integrated blockchain as the enabling technology to address MDA management problem [20], [35], [36]. It allows MDAPs (i.e. patients) to generate tokenization for their MDAs and designate a third party (i.e. a caregiver) to access their MDAs through a permission-based mechanism. However, these frameworks still depend on single blockchain to bridge MDAPs and MDA users, which means blockchain needs to undertake both data storage and sharing logic disposal. Even though blockchain can provider an ingenious infrastructure for practicing privacy protection approaches, it would be worth exploring further separation on blockchain's responsibilities, which will be better to control MDA sharing, according with the Single Responsibility Principle of software engineering [37]. Also, as a script-based automata, smart contracts can provide transparent and regulatory of processes. They are limited to offer encapsulation for data protection. Thus, ZKP is introduced to realize the tokenization of MDAs. [34] has made a preliminary exploration on using the verification feature of ZKP for privacy-preserving demand matching. However, in our scenario, the management of MDAs is a series of jobs, not just a single matching step. This takes us to figure out whether the constrain system of ZKP can be further developed and how it should be to suit for our scenario.

III. THE OVERALL FRAMEWORK

To achieve secure and sustainable MDA sharing in a decentralized environment, a blockchain-based MDA sharing framework is proposed as an integrated solution to eliminate privacy concerns and encourage sharing behaviors. Table II gives the notations of all symbols used in this paper. This framework enables MDA owners to conduct privacy-preserving MDA sharing third-party organizations with incentives. This framework adopts a five-layer architecture to decouple the relationship for practicing privacy protection from infrastructure perspective, as shown in Fig. 2. The responsibility of each layer is specified as follows.

The application layer encapsulates the business logic of MDA sharing and provides user interactions. There are three key operations for the sharer and user. The first is MDA

registration, which enables the sharer to declare the ownership of a definite MDA. The second operation is MDA authorization, which allows a sharer to authorize registered MDAs to a third-party user (e.g., a medical institution or data analysis department). The third operation is for a third-party user to access authorized MDAs by showing the corresponding authorization proof.

TABLE II
NOTATION

Symbol	Description
A	An MDAP
$h()$	A hash function (i.e., SHA256)
T	A trusted medical authority
Pk_A^T, Sk_A^T	The group-usable public key and private key generated by T for A
$zkp_g(inp_1, inp_2, inp_3)$	A ZKP function for generating the proof. inp_1 denotes the public inputs, inp_2 represents the private inputs, and inp_3 signifies the constraints to be verified.
zkp_c	A ZKP function for verifying the proof
Pk_A , Sk_A	The public key and private key for A, where: $Pk_A = h(Sk_A)$
$st(msg, a, b)$	A way to safely transmit a message from a to b
σ	A random salt
α	The unique nonsubstitutable ID of the MDA from A for registration
S_*	Source data of an MDA
Re_*	A record on a distributed ledger
P_*	Proof generated by the ZKP service
$e(msg, key)$	Group signature encryption function
A_{info}	Data after group signature encryption
$Addr_*$	The block address of a record
$Addr_{*.brother}$	The brother nodes of the corresponding address
M_*	The root of a Merkle tree for a record
$t(msg)$	Function of the group administrator for decrypting the group message
H	A third party for MDA acquisition
$c(Re_*, Addr_{*.brother})$	A calculation function for producing the Merkle root
R	A Boolean value for the algorithmic result
g_i	The total number of unique MDAs with MDAP i
$h_i(\leq g_i)$	The number of MDAs out of the total g_i that are held by them
$F(\boldsymbol{h}; \boldsymbol{g})$	The function of FSP for producing an additional payoff from MDA sharing
$B(g_i)$	The inherent utility function of privacy for MDAP i obtains by keeping its unique MDAs private
$C(h_i)$	The cost function for sharing h_i
$V_i(h_i)$	The gained compensation function for contributing h_i

* A symbol may have the following subscripts:
(i) α means it is related to the registration of MDA α
(ii) α. H means it is related to the authorization of MDA α to H
(iii) α. H. E means it is related to the extraction of the authorized MDA α for H

The mechanism layer is the core of this framework. It implements two enabling mechanisms as backend dependencies to support the operations in the application layer. The supervisory privacy-preserving sharing mechanism is a sharing management dependency that protects identity and process privacy for MDA registration, authorization and extraction. Furthermore, it guarantees limited identity traceability for shared MDAs to prevent malicious sharing or to enable epidemiological investigations. The sharing incentive mechanism is also dependent on the application layer because it encourages owners to provide MDAs so that this sharing behavior is sustainable.

The service layer abstracts the near-minimum amount of services that can provide the critical functionalities needed to deliver mechanisms for execution on the sharing network. It consists of three services. The smart contract service generates a smart contract, which is an executable instance for completing the mechanism logic on the network layer. The group signature service is responsible for maintaining identity privacy and traceability. The ZKP service is integrated to guarantee the process privacy of MDA sharing. Both of these last two services are invoked by a smart contract so that they can handle the related contract logic that is encapsulated and required by the mechanism layer.The network layer consists of a consortium blockchain network. The consortium blockchain is selected because this framework aims to share MDAs with multiple stakeholders, and this scenario falls between public and private chains. In general, the network layer constructs a decentralized network to organize sharing stakeholders and provides a secure and transparent environment to execute smart contracts. More specifically, three functions held by this consortium blockchain are important for the upper layers. First, a peer-to-peer (P2P) protocol should be integrated to construct the P2P network so that the storage layer can be deployed for qualified peers. Second, a distributed identity protocol is essential for providing the underlying interfaces for users and user role management in terms of key generation, certificate issues and access control. The distributed message protocol enables asynchronous message distribution and routing in the distributed network.

The storage layer is responsible for archiving and extracting the two types of data in a distributed manner. The interplanetary file system (IPFS) is used to slice the original MDAs, and the generated slices can then be distributed to geographically dispersed peers for storage. The adoption of slicing makes MDA storage more secure and private because each peer can only obtain parts of the MDAs, and no peer can obtain a full MDA [38]. The distributed ledger acts as the basis of the blockchain, which only keeps various kinds of proofs. Since a proof is lighter than the original MDA, it decreases storage waste and improves the scalability of the blockchain network.

IV. SUPERVISORY PRIVACY-PRESERVING SHARING MECHANISM

The supervisory privacy-preserving sharing mechanism consists of five operations. The responsibility of each operation is described as follows.

1) Group enrollment

A trusted MDA authority, such as a hospital, a clinic, or another medical institution, is introduced as a kind of group to manage its internal members. An MDAP should first enroll in a group with a unique ID. Then, the group generates a pair (Pk_A^T, Sk_A^T) based on the ID and invokes $st((Pk_A^T, Sk_A^T), T, A)$.

2) MDA registration

MDA registration depends on the following assumptions. (i) An α must be generated. (ii) A must set a private key Sk_A for him/herself; Sk_A does not need to be stored, but the owner must remember the value of the key. (iii) Only someone who knows

Sk_A can register A's MDA. (iv) An α can only be registered once. The specific registration process is shown below:

Algorithm 1: User-side MDA registration process

Input: $S_\alpha, Pk_A, Sk_A, \sigma$
Output: P_α
set $\alpha \leftarrow h(S_\alpha)$;
set $Re_\alpha \leftarrow h(\alpha|Pk_A|\sigma)$;
set $inp_1 = \{Pk_A, Re_\alpha, \alpha\}$;
set $inp_2 = \{Sk_A, \sigma\}$;
set *constraints* $\leftarrow \{inp_1[1] \equiv h(inp_2[1]), inp_1[2] \equiv h(inp_1[1]|inp_1[3]|inp_2[2])\}$;
set $P_\alpha \leftarrow zkp_g(inp_1, inp_2, constraints)$;
return P_α

A generates a proof P_α through Algorithm 1. Then, A sends a request with P_α to the smart contract and sends S_α to the smart contract for the IPFS.

Algorithm 2: Node-side MDA registration process

Input: $P_\alpha, h(S_\alpha)$
Output: R
set α from P_α;
set $R \leftarrow False$;
if $\alpha \equiv h(S_\alpha)$ & $zkp_c(P_\alpha) \equiv True$ & $a\ is\ not\ in\ the\ registration\ record$, **then**
  set $R \leftarrow True$;
end
return R

The nodes run Algorithm 2 to ensure that A's request is legitimate. If the result of Algorithm 2 is true, Re_α is put into the request pool, and user A's request is broadcast to all other nodes and goes through the same process. Subsequently, Re_α is recorded in the distributed ledger, and S_α persists to the IPFS. Then, $Addr_\alpha$ is sent back to A.

3) MDA authorization

The MDA owner can anonymously authorize his or her MDAs to any organization or individual. This process stores an authorization record on the distributed ledger without additional information about the authorization process. The authorization procedure has two assumptions. (i) Only a registered α can be authorized. (ii) A knows the private asset key Sk_A and the σ corresponding to α. The authorization process for A to H is specified as follows:

Algorithm 3: User-side MDA authorization process

Input: $\alpha, Pk_A, Sk_A, \sigma, \sigma', Re_\alpha, Addr_{a.brother}, Sk_A^T, Pk_H, S_\alpha, M_\alpha$
Output: $P_{\alpha.H}$
set $msg \leftarrow h(S_a)$;
set $A_{info} \leftarrow e(msg, Sk_A^T)$;
Do $st(\{\sigma', A_{info}, \alpha, Pk_A\}, A, H)$;
set $Re_{\alpha.H} \leftarrow h(\alpha|Pk_A|Pk_H|\sigma')$;
set $inp_1 = \{Pk_A, Pk_H, Re_\alpha, M_\alpha, Addr_{a.brother}, \alpha, Re_{\alpha.H}\}$;
set $inp_2 = \{Sk_A, \sigma', \sigma\}$;
set *constraints* $\leftarrow inp_1[1] \equiv h(inp_2[1]), inp_1[3] \equiv h(inp_1[1]|inp_1[6]|inp_2[3]), inp_1[4] \equiv c(inp_1[3], inp_1[5]), inp_1[7] \equiv h(inp_1[1]|inp_1[2]|inp_1[6]|inp_2[2])\}$;
set $P_{\alpha.H} \leftarrow zkp_g(inp_1, inp_2, constraints)$;
return $P_{\alpha.H}$

First, A queries the Merkle root M_α of the block via $Addr_\alpha$. Then, A runs Algorithm 3. In Algorithm 3, A creates a message through the group signature encryption function and sends it to H. Then, similar to MDA registration, A sends a request with $P_{\alpha.H}$ to the smart contract.

Algorithm 4: Node-side MDA authorization process

Input: $P_{\alpha.H}, Addr_a$
Output: R
set α, M_α from $P_{\alpha.H}$;
set $R \leftarrow False$;
if $a\ in\ the\ record$ & $zkp_C(P_{\alpha.H}) \equiv True$ & $M_\alpha\ is\ the\ root\ of\ Addr_a$, **then**
  set $R \leftarrow True$;
end
return R

The node uses Algorithm 4 to check the request and then uses the same process as in *operation (2)* to process the result of Algorithm 3. If the request is successfully stored, A will be able to view the block address $Addr_{\alpha.H}$ of its request record.

When H is the administrator of A's group, he or she can identify the owner of the authorized MDA through *operation (5)*. When H is an organization/person outside A's group or is not the administrator of A's group, he or she can still use A_{info} in *operation (5)* and verify whether the MDA is from the expected group, but he or she cannot identify the specific owner.

4) MDA extraction

If H wishes to obtain A's authorized MDA, H can access S_α by presenting the proof of authorization. The whole process is designed as follows:

Algorithm 5: Node-side MDA extraction process

Input: $\alpha, Pk_A, Pk_H, Sk_H, \sigma', Re_{\alpha.H}, Addr_{a.H.brother}, M_{\alpha.H}$
Output: $P_{\alpha.H.E}$
set $inp_1 = \{Re_{\alpha.H}, Pk_A, Pk_H, M_{\alpha.H}, Addr_{a.H.brother}, \alpha\}$;
set $inp_2 = \{Sk_H, \sigma'\}$;
set *constraints* $\leftarrow inp_1[3] \equiv h(inp_2[1]), inp_1[4] \equiv c(inp_1[1], inp_1[5]), inp_1[1] \equiv h(inp_1[2]|inp_1[3]|inp_1[6]inp_2[2])\}$;
set $P_{\alpha.H.E} \leftarrow zkp_g(inp_1, inp_2, constraints)$;
return $P_{\alpha.H.E}$

H generates a proof $P_{\alpha.H.E}$ through Algorithm 5 and sends $P_{\alpha.H.E}$ to the smart contract for the IPFS.

The IPFS nodes verify the validity of the inputs in $P_{\alpha.H.E}$ and $P_{\alpha.H.E}$ with two conditions:

> $M_{\alpha.H}$ is the root of the corresponding $Addr_{\alpha.H}$ block.
> $P_{\alpha.H.E}$ can pass the proof verification.

If successfully passed, the smart contract for the IPFS returns S_α to H.

5) MDA ownership traceability

During the authorization process, A sends a message encrypted by the group signature to the authorized party. This message can be used to trace the MDA ownership of the group administrator or the group ownership of outside members to perform supervision using the following process:

(a) Calculate α=$h(S_\alpha)$.

(b) Find the A_{info} corresponding to α.

(c) Administrator: Calculate $t(A_{info})$ to obtain the identity of A.

Others: Calculate $t'(A_{info})$ to obtain only the group identity of A.

V. Proposed Sharing Incentive Mechanism

With medical data contributors becoming increasingly aware of the value of their unique information, providing fair rewards or compensation has become imperative for MDA acquisition. For instance, new third-party platforms such as DNASimple reward donors with fixed rewards for their donations of biological samples [39]. Similar practices also exist in other industry sectors [40]. However, most of these programs merely focus on the quid pro quo of compensating contributors and fail to consider the drivers of MDA sharing, who could determine sufficiently valuable compensation to drive MDAP behavior. These include the disutility of MDAPs breaching the privacy of their unique MDAs, as well as the inherent time and energy costs linked to MDA sharing. Moreover, the influence of complementarity among the MDAPs in the reward mechanism is present as a blockchain platform integrates MDAs from different MDAPs. Finally, we introduce a model that endogenizes the decisions of MDAPs on MDA sharing. That is, an MDAP determines its optimal amount of MDA sharing to derive optimal individual-based rewards, while a blockchain platform leverages the MDAs to deliver knowledge and actionable insights to consumers such as healthcare providers for generating value.

Consider a framework service provider (FSP) with n registered MDAPs. The MDAPs are participants who interact with the FSP through their devices (registered smartphones, PCs, etc.). Each MDAP is assumed to be a self-utility maximizer. A model built on the self-utility maximization motivation not only produces valuable insights but can also easily be integrated into traditional management theories [41]. Similarly, the FSP is assumed to target the incremental payoff that it may generate, as effective medical image management will ultimately be reflected in its financial performance (by serving it to consumers interested in medical images). MDA consumers can use the valuable insights made available through the platform to either improve the value of their services or decrease their costs. Additionally, we assume that each MDA is unique to the person or patient to whom it belongs.

A. The FSP's Objective

Let $\boldsymbol{g} = (g_1, g_2 \ldots\ldots g_n)$, where g_i represents the total number of unique MDAs with MDAP i. For example, in the medical imaging context, each MDAP acquires MDAs from the radiographic scanning procedures in which it participates. Therefore, the number of images that the MDAP has acquired can be used as a measure of the number of MDA units possessed by the MDAP. The MDAPs share $\boldsymbol{h} = (h_1, h_2 \ldots\ldots h_n)$, where $h_i (\leq g_i)$ is the number of MDAs out of the total g_i that are held by them. For analytical simplicity, we treat both g_i and h_i as continuous variables. Subsequently, the FSP produces an additional payoff from MDA sharing, given by $F(\boldsymbol{h}; \boldsymbol{g})$, which is increasing and concave in h_i. This assumption is widely adopted in the literature [42], [43] and is suitable for scenarios that focus on incremental performance improvement.

Now, the MDAs obtained from each MDAP are heterogeneous in terms of their potential contributions to the platform payoff. For instance, an MDAP who contributes MRI images may contribute more to the payoff than an MDAP that contributes X-ray images of the same body part because of MRI's capability to depict even the most minute details and intricacies of the body part [44]. Hence, the potential impact of an MDAP on FSP performance may be expressed as $\frac{\partial F(\boldsymbol{h};\boldsymbol{g})}{\partial h_i}$, which represents the marginal MDA contribution from MDAP i to the payoff of the FSP.

Additionally, the MDAs managed through an FSP often tend to be interdependent and may complement one another in terms of the functions they serve. For example, multiple images of the same body part obtained from different MDAPs are required to train AI models for effective medical diagnosis [45], [46]. Hence, we introduce an additional facet called MDA interdependence. Here, $\frac{\partial^2 F(\boldsymbol{h};\boldsymbol{g})}{\partial h_i \partial h_j} > 0$ identifies a positive MDA interdependence between MDAPs i and j. Hence, the higher the value of $\frac{\partial^2 F(\boldsymbol{h};\boldsymbol{g})}{\partial h_i \partial h_j}$ is, the higher the level of complementarity among the MDAs contributed by the two MDAPs. For analytical simplicity, we rule out the case of negative interdependencies even though the results remain the same either way.

B. Costs Incurred by MDAPs

This study takes two factors that impose costs on MDAPs into account. First, MDAPs enjoy the inherent utility $B(g_i)$ of privacy, which is the utility of privacy that MDAP i obtains by keeping its unique MDAs private. Consequently, MDAP i incurs a cost $B(g_i) - B(g_i - h_i)$ when sharing h_i units of MDAs. Here, $B(\cdot)$ is increasing and concave in h_i; i.e., $B'(g_i) > 0$, and $B''(g_i) < 0$. It is reasonable to assume that the utility of privacy increases with the increase in the number of MDAs, i.e., $B'(g_i) > 0$. Furthermore, we assume that $B''(g_i) < 0$ because an MDAP with many MDAs experiences a lower increase in utility from an additional unit of MDA than an MDAP with a small number of MDAs. Hence, the more an MDAP engages in MDA sharing, the more rapidly its utility decreases.

Second, disutility also arises because of the time and effort that needs to be put in by MDAPs to identify, segregate, prepare and publish MDAs to the FSP. Hence, $C(h_i)$ represents the cost or disutility experienced by MDAP i when sharing h_i units of MDAs, where $C(\cdot)$ is convex in h_i.

Now, the resulting incentive-compatible MDA sharing mechanism functions as follows. First, the FSP introduces a valorization system to promote MDA sharing. Subsequently, the MDAPs determine the number of MDAs to publish, $h_i > 0$, if they decide to participate. In the case of nonparticipation, their utility remains unchanged. Finally, the FSP valorizes the MDAPs for their contributions.

C. MDA Valorization Mechanism

Assuming that V_i is the value extended to MDAP i based on the valorization system, the MDAP determines the optimal level of MDA sharing that can maximize its total utility. When sharing h_i units of MDAs, MDAP i incurs a cost $C(h_i)$, its remaining utility of privacy becomes $B(g_i - h_i)$, and it gains compensation $V_i(h_i)$ for its contributions. Hence, MDAP i's maximization problem may be written as:

$$\max_{h_i} Z = B(g_i - h_i) - C(h_i) + V_i \qquad ---(1)$$

Furthermore, its corresponding first-order condition may be written as follows:

$$\frac{\partial B(g_i - h_i)}{\partial h_i} - C'(h_i) + \frac{\partial V_i}{\partial h_i} = 0 \quad ---(2)$$

From Equation (2), when the MDAP determines the optimal h_i to share, it balances its utility from the marginal benefits $\frac{\partial V_i(h_i)}{\partial h_i}$ and the costs of privacy, $-\frac{\partial B(g_i - h_i)}{\partial h_i}$, and effort, $C'(h_i)$.

Consequently, the FSP maximizes its net payoff while considering the MDAPs' incentive to maximize their individual utilities. Hence, the FSP's maximization problem may be written as:

$$\max_{\substack{h_1 \ldots h_n \\ V_1 \ldots V_n}} \pi = F(\boldsymbol{h}) - \sum_{i=1}^{n} V_i \quad ---(3)$$

s.t.

$$B(g_i - h_i) - C(h_i) + V_i \geq B(g_i);\ \forall i \quad ---(4)$$

$$\frac{\partial B(g_i - h_i)}{\partial h_i} - C'(h_i) + \frac{\partial V_i}{\partial h_i} = 0;\ \forall i \quad ---(5)$$

Equations (4) and (5) represent the individual rationality constraint (IRC) and incentive compatibility constraint (ICC) of the agent, respectively. Here, the IRC ensures that an MDAP does not face a negative utility from MDA sharing, while the ICC considers the self-utility maximization of the MDAPs. These constraints are commonly used in principal agent studies [47], [48] to monetize an agent's contribution to the principal. At the optimum, the first-order condition for the FSP's maximization problem may be derived by introducing the binding constraint in (4) to Equation (3) as follows:

$$\frac{\partial F(\boldsymbol{h})}{\partial h_i} + \frac{\partial B(g_i - h_i)}{\partial h_i} - C'(h_i) = 0;\ \forall i \quad ---(6)$$

In Equation (6), the benefits are represented by the marginal increase in payoff $\frac{\partial F(\boldsymbol{h})}{\partial h_i}$, while $-\frac{\partial B(g_i - h_i)}{\partial h_i} + C'(h_i)$ is the marginal cost increase incurred by the MDAPs, which should be compensated to promote MDA sharing.

Let us denote $\boldsymbol{h}^* = (h_1{}^*, h_2{}^* \ldots . h_n{}^*)$ as the optimal solutions to the first-order conditions of the FSP. When binding constraint (4) is seen alongside constraints (5) and (6), the resulting incentive-compatible valorization system is given by the proposition below.

Proposition: The incentive-compatible valorization system is given as follows:

$$V_i(h_i) = x_i h_i + y_i \quad ---(7)$$

where

$$x_i = \frac{\partial F(\boldsymbol{h}^*)}{\partial h_i}\, h_i$$

$$y_i = B(g_i) - B(g_i - h_i{}^*) + C(h_i{}^*) - \frac{\partial F(\boldsymbol{h}^*)}{\partial h_i}\, h_i{}^*$$

Hence, Equation (7) provides a simple individual-based incentive-compatible valorization system that is linear in the number of MDAs shared. Notably, this study does not explicitly define any functional form for V_i, but the representation $V_i(h_i)$ here is based on the observed dependence of V_i on h_i in Equation (7). In Equation (7), x_i and y_i represent the marginal and basic values, respectively, given to MDAP i for its MDA contribution. These values are dependent on the number of shared MDAs h_i and its corresponding productivity $\frac{\partial F(\boldsymbol{h})}{\partial h_i}$. Interestingly, the marginal value set for the MDAs is equivalent to their productivity at the optimum.

Furthermore, when all MDAPs share the same optimal number of MDAs $h_i{}^*$, the optimal value assigned to the MDAPs from Equation (7) and the net payoff of the FSP converge to the following:

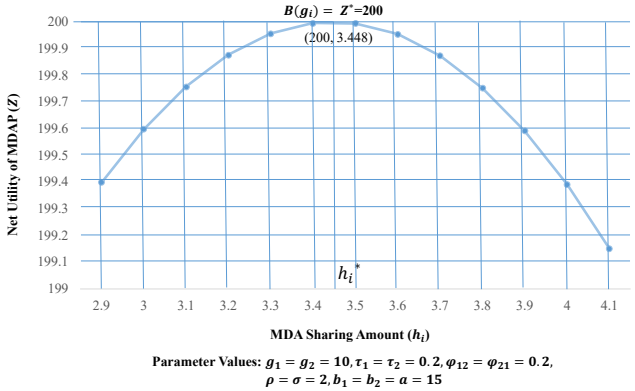


Fig. 3. Effect of the number of MDAs shared on the MDAP's net utility

$$V_i(h_i{}^*) = B(g_i) - B(g_i - h_i{}^*) + C(h_i{}^*) \quad ---(8)$$

$$\pi^* = F(\boldsymbol{h}^*) - \sum_{i=1}^{n} B(g_i) - B(g_i - h_i{}^*) + C(h_i{}^*) \quad (9)$$

From Equation (8), at the optimum, the compensation received by the MDAPs balances out the costs that they incur. Hence, the MDAPs enjoy a zero net surplus (i. e., $Z^* = B(g_i)$) from sharing $h_i{}^*$, whereas they are exposed to a net utility deficit otherwise, as shown in Fig. 3 (refer to Appendix A.2 for the functional forms used). Consequently, the FSP keeps the entire surplus, excluding the compensation provided to the MDAPs.

VI. Mechanism Evaluations

A. Evaluation of the Supervisory Privacy-Preserving Sharing Mechanism

TABLE III
SPECIFICATIONS

Hardware Environment	Dell Precision T7920 (Intel Xeon Silver 4214 CPU*2, 64 GB RAM, 1 TB SSD)*2
Development Tools	Python 3.8, Flask 1.1.2, Go-IPFS 0.10.0, ZoKrates 0.7.7, RabbitMQ-3.10.7, LevelDB-0.201, Gevent-20.0.0
Deployment Environment	VMware ESXi 6.7 Virtual machine for each node (10v CPU, 8 GB RAM, 80 GB SSD)
Network Configuration	TP-Link TL-R479G+
Consensus Algorithm	PBFT
Testbed Settings	7 Nodes (1 Primary and 6 Replica)

To further evaluate the computing performance of this mechanism, a prototype is implemented as the testbed to verify the feasibility of real-life applications. The specifications of this prototype are given in Table III.

The development tools for this prototype are based on Python 3.8, Flask 1.1.2, Go-IPFS 0.10.0, ZoKrates 0.7.7, RabbitMQ-3.10.7, LevelDB-0.201 and Gevent-20.0.0. FISCO BCOS v.2.9.0 is used to build the distributed ledger and consortium blockchain network in our framework. The computing

performance of this mechanism is evaluated based on four aspects. First, the overall system performance regarding the authorization verification operation is tested in terms of its transaction throughput and latency. Two parameters are considered key factors for designing experiments. One is the block size, which is widely discussed and analyzed in blockchain performance evaluations. It is measured by the number of transactions in a block in our experiments. The other parameter is the number of workers in each node that undertake proof operations. This is because proof operations are computationally intensive, and the parallelization of proof processes may be beneficial for achieving performance improvements. Thus, two groups of experiments are designed to verify the influences of the block size and number of workers.

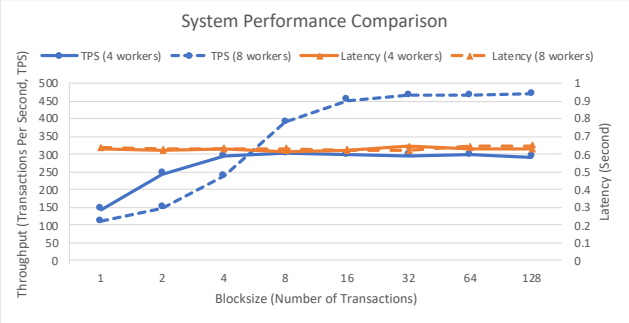


Fig. 4. System performance comparison

The first group focuses on the block size. The number of workers is set to 4 and 8 separately with different transaction sending rate strategies. For throughput testing, a total of 8192 requests are released at one time with 20 repetitions to obtain the throughput ceiling. Then, latency testing is conducted with an adaptive sending rate under the ceiling to obtain the general latency level. The results in Fig. 4 show that an increase in the block size can raise the throughput ceiling to some extent, and the latency remains stable at approximately 0.61 seconds when the sending rate does not exceed the throughput ceiling. Additionally, an increase in the number of workers can significantly enlarge the throughput ceiling. However, the increase in the number of workers has a negative effect on throughput when the block size is relatively small. This is because with 8 workers, proof operations occupy more CPU resources, so other threads for broadcasting and consensus are blocked. Furthermore, a small block size with a constant number of requests generates too many blocks concurrently, which aggravates the burden of the CPU with respect to handling communications for block broadcasting and consensus. Thus, the bottleneck of the CPU affects the throughput of the 8 workers. To avoid this, it is better to assign more CPUs to blockchain nodes and not set a block size that is too small, especially when the timeout value of the transaction of packing into a block can be controlled.

TABLE IV
BLOCKCHAIN PERFORMANCE ACHIEVED UNDER DIFFERENT PEER CONFIGURATIONS
(ZKP= GROTH16, BLOCK SIZE = 16)

Number of Workers	Throughput (TPS)	Latency (Seconds)	Average CPU Utilization
1	67.1573	0.6296	19.10%
2	161.8195	0.6212	33.60%
4	298.9398	0.6194	50.70%
6	386.6623	0.6322	78.40%
8	450.8449	0.6247	94.20%
10	447.9366	0.6096	94.50%
12	447.3179	0.6469	94.90%

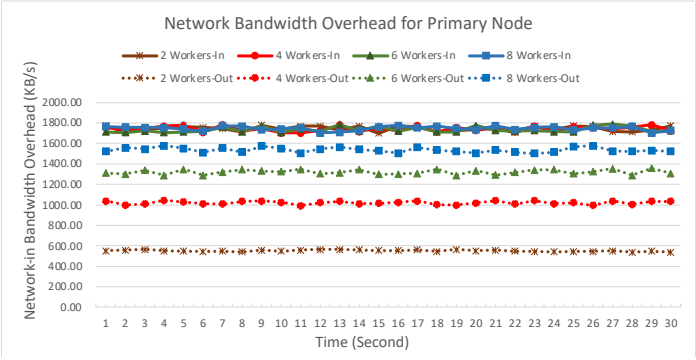
The second group of experiments aims to study the specific influence of parallelization on system performance. As shown in Table IV, an increase in the number of workers can indeed improve throughput before the CPU reaches saturation. However, even though the CPU is already fully loaded when 8 workers are enabled, the throughput remains stable. This is because the block size of 16 decreases the computing dependency of communication. Combined with the first group, it is preferable to first enlarge the block size to 16 or above and then set the number of workers according to the CPU performance to implement our framework. Second, we compare the specific performances of two ZKP schemes, Groth16 [49] and GM17 [50]. Their computing performance for the three key proof operations is summarized in Table V. One hundred sets of proof generation and verification processes are tested for registration, authorization and extraction, and the average time consumption levels are calculated. Large proof generation performance differences are observed. However, the schemes have similar proof verification performance. In addition, the proof sizes are the same for both schemes. For the three operations, the proof sizes are 3 kB, 6 kB and 6 kB for both Groth16 and GM17 [51]. Since all generations are performed on local devices, Groth16 is preferred considering the future adoption of smartphones, which are limited by their computational capacities and batteries, for MDA operations

TABLE V
PERFORMANCE COMPARISON BETWEEN DIFFERENT ZKP SCHEMES

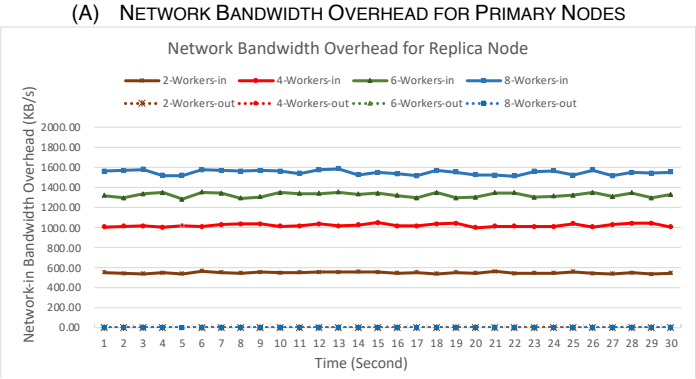
Average Processing Time	Groth16 (seconds)	GM17 (seconds)
Registration Generation	9.9936	102.6935
Registration Verification	0.0192	0.0195
Authorization Generation	58.8134	519.9867
Authorization Verification	0.0206	0.0211
Extraction Generation	54.6818	482.6898
Extraction Verification	0.0197	0.0199

Third, the network bandwidth overhead is measured under different numbers of workers when the block size is set to 16 transactions, as shown in Fig. 5. Fig. 5(A) is the network bandwidth overhead for primary nodes with different worker numbers. The inflow overhead is almost the same and stable because the inflows mainly consist of proof-related requests from clients, and the sending rate of clients has been set with a ceiling value (e.g., 550 requests/second, larger than the ceiling TPS). In addition, the inflows involve acknowledgments for consensus, which may cause slight fluctuation. The outflow overhead will increase with the worker number, but it will not have a linear ratio. This is because the outflow overhead is composed of verified proofs sent to others and acknowledgments. Ideally, the outflow overhead may increase with the worker number. However, the increase in the worker number will increase the CPU load to the bottleneck value, so we can see that the increased amplitude between outflow overheads will narrow. Fig. 5(B) is the network bandwidth overhead for replica nodes with different worker numbers. Generally, its inflow depends on the broadcasting rate of primary nodes. By the same analysis as for Fig. 5(A), it

approximates the outflows of primary nodes. Since replica nodes only generate acknowledgments (i.e., consensus, proof verification or confirmation), the outflow is significantly small, which is consistent with the above analysis. In general, the bandwidth overhead has a positive correlation with the number of workers because worker concurrency can accelerate the processing of transactions, which increases the communication overhead required for broadcasting and consensus. Since we use proofs instead of the MDAs themselves for blockchain network transmission, the small sizes of proofs greatly reduce the network bandwidth consumption. In addition, the inbound and outbound network bandwidth follow a fixed proportion because the message content and distribution scheme in practical Byzantine fault tolerance (PBFT) are stable. In conclusion, network bandwidth is unlikely to become the bottleneck of the system compared with the CPU. Additionally, a short group signature approach is adopted [52]. The size of the group signature is fixed to 92 bytes in our experiments because the input scales for signature parameters are consistent. Since signatures will be attached to proofs at the kilobyte level, the influence of signature size can be neglected. Furthermore, some limitations are observed due to the hardware constraints. The major limitation is the throughput ceiling. Since the physical workstations are limited, our experiments already make sufficient use of the performance of current hardware, especially the CPU. Thus, it is difficult to improve this performance. However, in real-life scenarios, the computing infrastructure, which usually involves a computing cluster or at least a server instead of one virtual machine, will have better CPU performance and will be able to avoid computation bottlenecks with an uncapped number of workers. Another limitation is the small network scale. The testing network contains 7 nodes, which can only simulate an early consortium network. However, since the network bandwidth overhead is very small, the expansion of the network scale will not have an appreciable impact on the overall system performance.



(A) NETWORK BANDWIDTH OVERHEAD FOR PRIMARY NODES



(B) NETWORK BANDWIDTH OVERHEAD FOR REPLICA NODES

Fig. 5. Network bandwidth overhead for different worker numbers

TABLE VI
BENCHMARK EXPERIMENTS
(ZKP= GROTH16, BLOCK SIZE = 16)

Item	Baseline [12]	Our framework	Percentage Change	Numerical Change
TPS	460.7	450.8	↓2.2%	-9.9
Latency (s)	0.04	0.62	↑1450%	+0.58
Average CPU Utilization	34.03%	94.20%	↑176.8%	+60.17%
Network overhead-Primary node (in/out, Kb/s)	380/687	1746.67/1536.91	↑359.65% /↑123.71%	+1366.67 /+849.91
Network overhead-Peer node (in/out, Kb/s)	67/53	1550.46/3.54	↑2214.11% /↓93.32%	+1483.46 /-49.46

Fourth, a benchmark experiment has been conducted, comparing with single blockchain-based MDA sharing schema in [12]. The results are shown in Table VI. The benchmark approach employed blockchain for MDA sharing without privacy protection. The results show the exact overheads of TPS, latency, computing and networking for implementing privacy protection for MDA sharing. For TPS, our framework only sacrifices 2.2% on TPS, benefiting from the parallel execution. Latency will be increased significantly because of the generation and verification of proofs are computing-intensive, which takes more time for transaction disposal. This is also shown on the average CPU utilization, which means if more powerful CPU will be deployed, the performance of TPS and latency will be further optimized. For the networking overhead, proofs are needed to be attached in the transaction broadcasted so that both primary and peer nodes will have more overhead.

B. Impacts of the Number and Productivity of MDAs on MDA Sharing

Consider a scenario in which the productivity of MDAs obtained from MDAP i is increased while the remaining MDAPs display unaltered MDA productivity at the optimum. Here, at $\boldsymbol{h}^*$ (the optimum before an increase in productivity), Equation (6) takes a positive value on its left-hand side. For Equation (6) to hold, MDAP i needs to increase its number of shared MDAs, h_i, which in turn influences the numbers of MDAs shared by the remaining MDAPs because of MDA interdependence. Therefore, the corresponding h_j ($j \neq i$) increases from h_j^* ($j \neq i$) in the respective first-order conditions of the MDAPs. Furthermore, to compensate for the resulting decrease in $\frac{\partial B(g_i - h_i)}{\partial h_i} - C'(h_i)$, the value of $\frac{\partial F(\boldsymbol{h})}{\partial h_i}$ should increase at the optimum. As a result, the marginal benefits given to MDAP i as well as the remaining MDAPs see increases in value. Hence, it may be inferred that the optimal amount of MDA sharing increases with an increase in the productivity of MDAPs.

Now, the following lemma summarizes the impacts of the number of unique MDAs , g_i , held by MDAP i and its

productivity on the optimal number of shared MDAs h_i^* (refer to Appendix A.1 for the proof).

TABLE VII
THE OPTIMAL SHARING ANALYSIS

Number of MDAs	*Interdependence*	*MDA Productivity (MP)*		
		$MP_i > MP_j$	$MP_i = MP_j$	$MP_i < MP_j$
$g_i = g_j$	*Nonnegative*	$h_i^* > h_j^*$	$h_i^* = h_j^*$	$h_i^* < h_j^*$
$g_i > g_j$	*Zero*	$h_i^* > h_j^*$	$h_i^* > h_j^*$	*Mostly* $h_i^* > h_j^*$
	Positive	$h_i^* > h_j^*$ (mostly)	$h_i^* > h_j^*$ (mostly)	*Intermediate*

Lemma. The optimal number of shared MDAs h_i^* from an MDAP i is given in Table VII.

From the first row, for any level of interdependence, a more productive MDAP should engage in better sharing than its less productive counterparts when they all possess an equal number of MDAs ($g_i = g_j$ for all $i \neq j$). Now, to investigate how the valorization mechanism should be set up, consider a scenario in which MDAP i produces MDAs with higher productivity than those of MDAP j ($MP_i > MP_j$). Here, $h_i^* > h_j^*$ at the optimum. Observing (E.1) from the Appendix, the signs of the second and third parentheses are negative and positive, respectively, making the positivity of the first term apparent. Hence, x_i ($= \partial F(\boldsymbol{h}^*)/\partial h_i$) $>$ x_j ($= \partial F(\boldsymbol{h}^*)/\partial h_j$). Now, assume that for $y_i \geq y_j$, $V_i(h_i) = x_i h_i + y_i > V_j(h_i) = x_j h_i + y_j$ for all h_i. Furthermore, $V_i(h_j^*) = x_i h_j^* + y_i > V_j(h_j^*) = B(g) - B(g - h_j^*) + C \quad (h_j^*)$, producing a net positive utility for MDAP i, which shares h_j^*. This contradicts the result obtained from Section V(C), which identifies the net-zero utility experienced by the IPs at the optimum. Hence, $y_i < y_j$. As a result, the platform should set a higher marginal value and a lower base value for an MDAP with higher productivity given that the MDAPs display homogeneity over all other factors.

Now, the scenario in which MDAP i possesses more MDAs than MDAP j ($g_i > g_j$) is given in the second row of the table in the lemma. Here, MDAP i incurs lower costs from MDA sharing than MDAP j, and hence, it has lower associated marginal costs. For a scenario with nonexistent interdependence and $MP_i \geq MP_j$, the platform's payoff from MDAP i's MDAs is greater than or equal to that from MDAP j's MDAs. As a result, the platform should facilitate increased sharing of the MDAs from MDAP i in an attempt to balance the costs and benefits in Equation (6); i.e., $h_i^* > h_j^*$. For the remaining cases in the lemma, the relationship between h_i^* and h_j^* is not uniquely determined, even though $h_i^* > h_j^*$ for the majority of cases.

C. Discussion

The design, experiments, and analysis of the MDA sharing framework have produced insights and recommendations for FSPs and contributors aiming to maximize their benefits from MDA sharing. First, the novel supervisory privacy-preserving sharing mechanism, the technical backbone of the proposed framework, eliminates the privacy worries associated with contributors' MDA sharing. Second, the sharing incentive mechanism aligns the incentives to motivate the use of the platform by MDA contributors by capturing the attributes and factors that drive their sharing behavior. These preliminary observations and insights are the backbone for the implications we discuss below.

- FSPs can use two levers to maximize blockchain performance for MDA sharing: block size optimization and worker/processer parallelization on the blockchain node. Block sizes that are too small can inhibit the throughput of the underlying blockchain network, whereas large block sizes may result in empty blocks (from timeout) or blocks with very few transactions. Therefore, the FSPs should perform market testing to observe MDA contributor engagement, or the resulting transaction rate, to determine the ideal block size. Similarly, worker/processer parallelization for transaction execution is suitable for use alongside a zero-knowledge proof. The latter provides independent verifiability of transactions with minimal messaging of proofs to avoid network congestion.
- Notably, FSPs and participants may adjust the throughput ceiling of the mechanism by relaxing the hardware constraints. Unlike the physical workstation we used for the experiments in this study, an industrial computing infrastructure with a computing cluster (or at least a server) will produce high CPU performance. It can avoid computation bottlenecks with an uncapped number of workers. Therefore, the proposed mechanism can be easily scaled for industrial applications with many participants engaging with the sharing platform simultaneously.
- Finally, FSPs can set up a simple linear incentive mechanism that inherits the decentralized nature of the stakeholders by avoiding the centralization of the platform over time, i.e., a few participants taking all the rewards, as is typical in cryptocurrency networks such as Bitcoin. Although the incentive mechanism encourages contributors with more MDAs or more productive MDAs to contribute to FSPs, the more minor contributors still receive reasonable compensation or fair value for their contributions because of their interdependence with their larger counterparts. Therefore, more minor contributors or new contributors will have opportunities to join and benefit from FSPs.

VII. CONCLUSIONS

To facilitate the collection and sharing of big medical data, this paper presents a blockchain-based MDA sharing framework. The major contribution of this paper lies in three aspects. First, we designed a layered-architecture to decouple the privacy-preserving responsibilities among the employed techniques considering both the incentive rewarding and the parallelization of execution. Second, a supervisory privacy-preserving sharing mechanism is used to integrate the transparency and determinacy of smart contracts with the privacy-preserving ZKP and group signatures to create a provable and supervisory privacy protection guarantee for managing the MDA sharing process. Third, an MDA sharing incentive mechanism is presented to promote MDA sharing. Even though the analysis is based on a simplified model, it captures various important attributes that should be considered, such as the private utility derived by an MDAP from its MDAs, the time- and effort-intensive nature of MDA sharing, the impacts of MDAs on the financial performance of an FSP, and

the productivity of and interdependence among MDAs. The mechanisms introduced in this study can act as a stepping stone for further studies on MDA sharing and more sophisticated valorization mechanisms.

However, some future work ideas are worth further exploration. First, this paper does not consider an exit mechanism for current group members in terms of the group signature encryption function, which results from the shortcoming of the short group signatures. Thus, the current method can be further improved, or more kinds of group signature methods can be integrated to enhance the suitability of our framework for different application scenarios. Second, the security of sharing should be enhanced for real-life applications due to potential malicious sharers or repeated MDA sharing. Third, the shared MDAs may generate potential application value. The a posteriori incentive mechanism is promising for distributing reverse rewards; in particular, this paper contributes an identity traceability basis for obtaining these rewards.

APPENDIX A.1-PROOF OF LEMMA

Considering the first-order conditions in Equation (6) for MDAPs i and j together, we obtain the following:

$$\left\{\frac{\partial F(\boldsymbol{h}^*;\boldsymbol{g})}{\partial h_i}-\frac{\partial F(\boldsymbol{h}^*;\boldsymbol{g})}{\partial h_j}\right\}+\left\{\frac{\partial B(g_i-h_i^*)}{\partial h_i}-\frac{\partial B(g_j-h_j^*)}{\partial h_j}\right\} -\left\{C'(h_i^*)-C'(h_j^*)\right\}=0 \tag{E.1}$$

(1) **Case 1:** $g_i=g_j\,(=g)$. Assuming $MP_i>MP_j$, suppose that $h_i^*\le h_j^*$. In (E.1), the signs of the second and third expressionsin parentheses are negative and positive, respectively, making the positivity of the first term apparent. However,

$$\frac{\partial F(h_i^*,h_j^*;g,g)}{\partial h_i}\ge\frac{\partial F(h_j^*,h_j^*;g,g)}{\partial h_i}>\frac{\partial F(h_j^*,h_j^*;g,g)}{\partial h_j}\ge\frac{\partial F(h_i^*,h_j^*;g,g)}{\partial h_j} \tag{E.2}$$

which produces a contradiction. Thus, $h_i^*>h_j^*$.

Now, when $MP_i=MP_j$, $h_i^*=h_j^*$ by symmetry.

(2) **Case 2:** $g_i>g_j$. When $h_i^*\le h_j^*$, as in case 1, the terms in the first set of parentheses of (E.1) must produce a negative value for the first-order condition to hold. Hence,

$$\frac{\partial F(h_i^*,h_j^*;g_i,g_j)}{\partial h_i}=\frac{\partial F(h_i^*,h_j^*;g_j,g_j)}{\partial h_i}\ge\frac{\partial F(h_j^*,h_j^*;g_j,g_j)}{\partial h_i} \tag{E.3}$$

- *Subcase 1*: *Zero interdependence*. For nonexistent interdependence,

$$\frac{\partial F(h_j^*,h_j^*;g_j,g_j)}{\partial h_j}=\frac{\partial F(h_i^*,h_j^*;g_i,g_j)}{\partial h_j} \tag{E.4}$$

If $MP_i\ge MP_j$, then $\partial F(h_j^*,h_j^*;g_j,g_j)/\partial h_i\ge\partial F(h_j^*,h_j^*;g_j,g_j)/\partial h_j$. Subsequently, from (E.3) and (E.4), $(h_i^*,h_j^*;g_i,g_j)/\partial h_i\ge\partial F(h_i^*,h_j^*;g_i,g_j)/\partial h_j$, which produces a contradiction. Therefore, $h_i^*>h_j^*$. If $MP_i<MP_j$, then $\partial F(h_j^*,h_j^*;g_j,g_j)/\partial h_i<\partial F(h_j^*,h_j^*;g_j,g_j)/\partial h_j$. From (E.3) and (E.4), $\partial F(h_i^*,h_j^*;g_i,g_j)/\partial h_i-\partial F(h_i^*,h_j^*;g_i,g_j)/\partial h_j$ becomes negative only when there is a substantial productivity difference. Although (E.1) does not hold for the majority of the cases resulting in $h_i^*>h_j^*$, it holds (that is, $h_i^*\le h_j^*$) for a substantial difference in productivity.

- *Subcase 2*: *Positive interdependence*. (E.4) can be modified to suit this subcase as follows:

$$\frac{\partial F(h_j^*,h_j^*;g_j,g_j)}{\partial h_j}>\frac{\partial F(h_i^*,h_j^*;g_j,g_j)}{\partial h_j}<\frac{\partial F(h_i^*,h_j^*;g_i,g_j)}{\partial h_j} \tag{E.5}$$

For $MP_i\ge MP_j$, taking (E.3) and (E.5) into account, $\partial F(h_i^*,h_j^*;g_i,g_j)/\partial h_i\ge\partial F(h_i^*,h_j^*;g_i,g_j)/\partial h_j$ is likely to hold for low-interdependence cases or when the difference between g_i and g_j is not very large, i.e., when $\partial F(h_i^*,h_j^*;g_i,g_j)/\partial h_j-\partial F(h_i^*,h_j^*;g_j,g_j)/\partial h_j$ is not very significant (the last inequality in (E.5)). As a result, $h_i^*>h_j^*$ in the majority of cases. For $MP_i<MP_j$, $\partial F(h_j^*,h_j^*;g_j,g_j)/\partial h_i<\partial F(h_j^*,h_j^*;g_j,g_j)/\partial h_j$. From (E.3) and (E.5), $\partial F(h_i^*,h_j^*;g_i,g_j)/\partial h_i-\partial F(h_i^*,h_j^*;g_i,g_j)/\partial h_j$ becomes nonpositive for a significant difference in productivity and nonsignificant interdependence. Although (E.1) is not satisfied in the majority of cases, producing $h_i^*>h_j^*$, it will hold (i.e., $h_i^*\le h_j^*$) given that there exists a very significant difference in productivity.

APPENDIX A.2-FUNCTIONAL FORMS

The functional forms used for the numerical study (or the numerical examples introduced in Fig. 2) are listed below.

$$F(\boldsymbol{h})=\sum_i(\tau_i/2)\left(b_i{}^2-(b_i-h_i)^2\right)\sum_{j\ne i}g_j+\sum_i\sum_{j>i}\varphi_{ij}h_ih_j$$
$$b_i\ge g_i,\tau_i>0\text{ and }\varphi_{ij}>0\text{ for all }i\ne j. \tag{E.6}$$

$$B(g_i)=(\rho/2)(a^2-(a-g_i)^2),\rho>0\text{ and }a\ge g_i\text{ for all }i. \tag{E.7}$$

$$C(h_i)=(\sigma/2)h_i{}^2,\sigma>0. \tag{E.8}$$

Since $\partial F(\boldsymbol{h})/\partial h_i=\tau_i(b_i-h_i)\sum_{j\ne i}g_j+\sum_{j\ne i}\varphi_{ij}h_j$, the productivity of h_i increases with increases in τ_i and φ_{ij}. Hence, these two parameters together represent the productivity of the given MDAs. In addition, φ_{ij} captures the interdependence between the MDAs of MDAP i and MDAP j (as $\partial^2F(\boldsymbol{h})/\partial h_i\partial h_j=\varphi_{ij}$). Furthermore, ρ and σ account for the importance of privacy and the significance of the MDAP's effort and time, respectively.

To derive the incentive-compatible valorization system introduced in the proposition, we consider a scenario with two MDAPs ($n=2$). Note that our derivation remains applicable for scenarios with more than two MDAPs, even though we skip the details here to conserve space.

Since $B(g_i-h_i)/\partial h_i=-\rho(a-g_i+h_i)$ and $C'(h_i)=\sigma h_i$, the first-order condition for MDAP i introduced in Equation (6) becomes the following:

$$\tau_ig_{3-i}(b_i-h_i)+\varphi_{12}h_{3-i}-\rho(a-g_i+h_i)-\sigma h_i=0,i=1,2. \tag{E.9}$$

Solving the equations for the two MDAPs in (E.9) simultaneously, we obtain the following:

$$h_i{}^*=\frac{\tau_{3-i}\rho g_i{}^2+(\tau_i\tau_{3-i}b_ig_{3-i}+\tau_{3-i}\varphi_{12}b_{3-i}-(\tau_{3-i}a-\rho-\sigma)\rho)g_i}{(\tau_{3-i}g_i+\rho+\sigma)(\tau_ig_{3-i}+\rho+\sigma)-\varphi_{12}{}^2} \tag{E.10}$$
$$+\frac{(\tau_ib_i(\rho+\sigma)+\varphi_{12}\rho)g_{3-i}-(\varphi_{12}+\rho+\sigma)\rho a}{(\tau_{3-i}g_i+\rho+\sigma)(\tau_ig_{3-i}+\rho+\sigma)-\varphi_{12}{}^2},i=1,2.$$

Plugging h_i^* from (E.10) into $\partial F(\boldsymbol{h})/\partial h_i$, $B(g_i-h_i)$, and $C(h_i)$, we obtain the marginal value $x_i=\partial F(\boldsymbol{h}^*)/\partial h_i$ and the base value $y_i=B(g_i)-B(g_i-h_i^*)+C(h_i^*)-(\partial F(\boldsymbol{h}^*)/\partial h_i)h_i^*$. Here, for brevity, we omit the detailed expressions.

REFERENCES

[1] Z. Pang, G. Yang, R. Khedri, and Y.-T. Zhang, "Introduction to the special section: convergence of automation technology, biomedical engineering, and health informatics toward the healthcare 4.0," *IEEE Reviews in Biomedical Engineering*, vol. 11, pp. 249–259, 2018.

[2] J. Li and P. Carayon, "Health Care 4.0: A vision for smart and connected health care," *IISE Transactions on Healthcare Systems Engineering*, pp. 1–10, 2021.

[3] H. Ghayvat *et al.*, "CP-BDHCA: Blockchain-based Confidentiality-Privacy preserving Big Data scheme for healthcare clouds and applications," *IEEE Journal of Biomedical and Health Informatics*, 2021.

[4] D. Sierra-Sosa *et al.*, "Scalable healthcare assessment for diabetic patients using deep learning on multiple GPUs," *IEEE transactions on industrial informatics*, vol. 15, no. 10, pp. 5682–5689, 2019.

[5] S. Sneha and A. Dulipovici, "Strategies for Working with Digital Medical Images Secure knowledge sharing: an organizational culture perspective View project Strategies for Working with Digital Medical Images," 2006, doi: 10.1109/HICSS.2006.439.

[6] J. Sun, Y. Yuan, M. Tang, X. Cheng, X. Nie, and M. U. Aftab, "Privacy-preserving Bilateral Fine-grained Access Control for Cloud-enabled Industrial IoT Healthcare," *IEEE Transactions on Industrial Informatics*, 2021.

[7] K. Zhang, N. Antonopoulos, and Z. Mahmood, "A review of incentive mechanism in peer-to-peer systems," presented at the 2009 First International Conference on Advances in P2P Systems, IEEE, 2009, pp. 45–50.

[8] Y. Zhang, M. Pan, L. Song, Z. Dawy, and Z. Han, "A survey of contract theory-based incentive mechanism design in wireless networks," *IEEE wireless communications*, vol. 24, no. 3, pp. 80–85, 2017.

[9] Y. Zhan, P. Li, Z. Qu, D. Zeng, and S. Guo, "A learning-based incentive mechanism for federated learning," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6360–6368, 2020.

[10] X. Liang, W. Chen, J. Li, Y. Mu, and Z. Tian, "Incentive Mechanism of Medical Data Sharing Based On Information Entropy in Blockchain Environment," presented at the Journal of Physics: Conference Series, IOP Publishing, 2019, p. 022056.

[11] L. Marques, D. Morais, and A. Terra, "More Than Meets the Eye: Misconduct and Decoupling Against Blockchain for Supply Chain Transparency," *Production and Operations Management*, p. 10591478231224928, 2024.

[12] K. M. Bartol and A. Srivastava, "Encouraging knowledge sharing: The role of organizational reward systems," *Journal of leadership & organizational studies*, vol. 9, no. 1, pp. 64–76, 2002.

[13] M. T. Hansen, "The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits," *Administrative science quarterly*, vol. 44, no. 1, pp. 82–111, 1999.

[14] C. J. Bennett and C. D. Raab, *The governance of privacy: Policy instruments in global perspective*. Routledge, 2017.

[15] L. F. Cranor, "P3P: Making privacy policies more useful," *IEEE Security & Privacy*, vol. 1, no. 6, pp. 50–55, 2003.

[16] P. Ashley, S. Hada, G. Karjoth, C. Powers, and M. Schunter, "Enterprise privacy authorization language (EPAL)," *IBM Research*, vol. 30, p. 31, 2003.

[17] A. H. Anderson, "A comparison of two privacy policy languages: EPAL and XACML," presented at the Proceedings of the 3rd ACM workshop on Secure web services, 2006, pp. 53–60.

[18] R. Zhang, D. Chen, X. Shang, X. Zhu, and K. Liu, "A knowledge-constrained access control model for protecting patient privacy in hospital information systems," *IEEE journal of biomedical and health informatics*, vol. 22, no. 3, pp. 904–911, 2017.

[19] K. Seol, Y.-G. Kim, E. Lee, Y.-D. Seo, and D.-K. Baik, "Privacy-preserving attribute-based access control model for XML-based electronic health record system," *IEEE Access*, vol. 6, pp. 9114–9128, 2018.

[20] S. Biswas, K. Sharif, F. Li, I. Alam, and S. Mohanty, "DAAC: Digital Asset Access Control in a Unified Blockchain Based E-Health System," *IEEE Transactions on Big Data*, 2020.

[21] S. Biswas, K. Sharif, F. Li, Z. Latif, S. S. Kanhere, and S. P. Mohanty, "Interoperability and synchronization management of blockchain-based decentralized e-health systems," *IEEE Transactions on Engineering Management*, vol. 67, no. 4, pp. 1363–1376, 2020.

[22] C. Zhonghua, S. Goyal, and A. S. Rajawat, "Smart contracts attribute-based access control model for security & privacy of IoT system using blockchain and edge computing," *The Journal of Supercomputing*, vol. 80, no. 2, pp. 1396–1425, 2024.

[23] D. D. F. Maesa, A. Lisi, P. Mori, L. Ricci, and G. Boschi, "Self sovereign and blockchain based access control: Supporting attributes privacy with zero knowledge," *Journal of Network and Computer Applications*, vol. 212, p. 103577, 2023.

[24] F. Brunton and H. Nissenbaum, *Obfuscation: A user's guide for privacy and protest*. Mit Press, 2015.

[25] H. Xu and N. Zhang, "Implications of Data Anonymization on the Statistical Evidence of Disparity," *Management Science*, 2021.

[26] S. Singh, P. K. Sharma, S. Y. Moon, and J. H. Park, "Advanced lightweight encryption algorithms for IoT devices: survey, challenges and solutions," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–18, 2017.

[27] H. Li, K. Yu, B. Liu, C. Feng, Z. Qin, and G. Srivastava, "An efficient ciphertext-policy weighted attribute-based encryption for the Internet of health things," *IEEE Journal of Biomedical and Health Informatics*, 2021.

[28] Z. Guan, X. Zhou, P. Liu, L. Wu, and W. Yang, "A Blockchain based Dual side Privacy preserving Multi party Computation Scheme for Edge enabled Smart Grid," *IEEE Internet of Things Journal*, 2021.

[29] W. Liang, D. Zhang, X. Lei, M. Tang, K.-C. Li, and A. Zomaya, "Circuit copyright blockchain: Blockchain-based homomorphic encryption for IP circuit protection," *IEEE Transactions on Emerging Topics in Computing*, 2020.

[30] S. Li, S. Zhao, G. Min, L. Qi, and G. Liu, "Lightweight privacy-preserving scheme using homomorphic encryption in industrial Internet of Things," *IEEE Internet of Things Journal*, 2021.

[31] B. Jia, X. Zhang, J. Liu, Y. Zhang, K. Huang, and Y. Liang, "Blockchain-enabled federated learning data protection aggregation scheme with differential privacy and homomorphic encryption in IIoT," *IEEE Transactions on Industrial Informatics*, 2021.

[32] M. A. Shawky *et al.*, "Efficient blockchain-based group key distribution for secure authentication in VANETs," *IEEE Networking Letters*, vol. 5, no. 1, pp. 64–68, 2023.

[33] J. Liu, Y. Fan, R. Sun, L. Liu, C. Wu, and S. Mumtaz, "Blockchain-aided privacy-preserving medical data sharing scheme for e-healthcare system," *IEEE Internet of Things Journal*, 2023.

[34] X. Li, H. Zhao, and W. Deng, "BFOD: Blockchain-based privacy protection and security sharing scheme of flight operation data," *IEEE Internet of Things Journal*, 2023.

[35] S. Rouhani, L. Butterworth, A. D. Simmons, D. G. Humphery, and R. Deters, "MediChain TM: a secure decentralized medical data asset management system," presented at the 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), IEEE, 2018, pp. 1533–1538.

[36] A. Srivastava, K. Farooq, R. Kumari, A. Poornima, and M. Nirmala, "Blockchain based Medical Data Asset Management System.," *Grenze International Journal of Engineering & Technology (GIJET)*, vol. 8, no. 2, 2022.

[37] R. C. Martin, "The single responsibility principle," *The principles, patterns, and practices of Agile Software Development*, vol. 149, p. 154, 2002.

[38] P. Poornima Devi, S. A. Bragadeesh, and A. Umamakeswari, "Secure data management using IPFS and Ethereum," presented at the Proceedings of International Conference on Computational Intelligence, Data Science and Cloud Computing: IEM-ICDC 2020, Springer, 2021, pp. 565–578.

[39] DNA "DNAsimple - Speeding up science with you " DNA Simple, 2023. https://www.dnasimple.org/newhome. (accessed Apr. 20, 2024).

[40] Reach, "A Better Way to Reward Customers for Sharing Data | REACH," 2023. https://joinreach.com/blog/better-way-reward-customers-sharing-data/. (accessed Apr. 20, 2024).

[41] J. B. Barney, "The debate between traditional management theory and organizational economics: substantive differences or intergroup conflict?," *Academy of Management review*, vol. 15, no. 3, pp. 382–393, 1990.

[42] R. K. Chellappa and A. Mehra, "Cost drivers of versioning: Pricing and product line strategies for information goods," *Management Science*, vol. 64, no. 5, pp. 2164–2180, 2018.

[43] S. Samaddar and S. S. Kadiyala, "An analysis of interorganizational resource sharing decisions in collaborative knowledge creation," *European Journal of operational research*, vol. 170, no. 1, pp. 192–210, 2006.

[44] L. Fayad, "CT Scan Versus MRI Versus X-Ray: What Type of Imaging Do I Need?" Accessed: Jun. 23, 2021. [Online]. Available: https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/ct-vs-mri-vs-xray

[45] C. Guo, S. Su, K.-K. R. Choo, and X. Tang, "A fast nearest neighbor search scheme over outsourced encrypted medical images," *IEEE*

Transactions on Industrial Informatics, vol. 17, no. 1, pp. 514–523, 2018.

[46] K. Guo *et al.*, "MDMaaS: Medical-assisted diagnosis model as a service with artificial intelligence and trust," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 2102–2114, 2019.

[47] R. Zhang and Q. Zhu, "FlipIn: A Game-Theoretic Cyber Insurance Framework for Incentive-Compatible Cyber Risk Management of Internet of Things," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2026–2041, 2019.

[48] D. I. Singham, "Sample average approximation for the continuous type principal-agent problem," *European Journal of Operational Research*, vol. 275, no. 3, pp. 1050–1057, 2019.

[49] J. Groth, "On the size of pairing-based non-interactive arguments," presented at the Advances in Cryptology–EUROCRYPT 2016: 35th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Vienna, Austria, May 8-12, 2016, Proceedings, Part II 35, Springer, 2016, pp. 305–326.

[50] J. Groth and M. Maller, "Snarky signatures: Minimal signatures of knowledge from simulation-extractable SNARKs," presented at the Advances in Cryptology–CRYPTO 2017: 37th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 20–24, 2017, Proceedings, Part II, Springer, 2017, pp. 581–612.

[51] S. Atapoor and K. Baghery, "Simulation extractability in Groth's zk-SNARK," in *Data Privacy Management, Cryptocurrencies and Blockchain Technology*, Springer, 2019, pp. 336–354.

[52] D. Boneh, X. Boyen, and H. Shacham, "Short group signatures," presented at the Annual international cryptology conference, Springer, 2004, pp. 41–55.

Ming Li (Senior Member, IEEE) received his bachelor degree in Computer Science from South China University of Technology at 2012, and master's and Ph.D. degrees in Industrial and Manufacturing Systems Engineering from the Department of Data and Systems Engineering, The University of Hong Kong, in 2013 and 2018, respectively. He is currently a Research Assistant Professor with the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University. His research interests are blockchain and cyber-physical system in smart manufacturing and logistics. He has authored or coauthored more than 70 papers on international journals and top conferences, including IEEE T-II, IJPE, JMS, RCIM, etc. Dr. Li is the Core Member of the 2019 Guangdong Special Support Talent Program – Innovation and Entrepreneurship Leading Team. He is also a Senior Member of CCF and a Member of ASME and IISE, as well as Co-found of CommaTech.

Arjun Rachana Harish graduated with a Bachelor's degree in Mechanical Engineering from the Indian Institute of Information Technology, Jabalpur, in 2019 and completed his Ph.D. in Industrial and Manufacturing Systems Engineering from the Department of Data and Systems Engineering, The University of Hong Kong, in 2024. He is currently a Post-Doctoral Fellow with the Department of Industrial and Systems Engineering at the Hong Kong Polytechnic University, where he is actively pursuing his research interests in blockchain and industrial large models in sustainable logistics and operations management. Dr. Harish has authored or coauthored more than 10 articles in international journals and conferences, including IEEE T-II, CAIE, CII, and JIII, among others.

Chenglin Yu received his bachelor's degree in Computer Science from China University of Geosciences and his master's degree in Computer Science from The George Washington University. He is currently pursuing his Ph.D. in the Department of Data and Systems Engineering at The University of Hong Kong. His research focuses on blockchain design, intelligent logistics, and smart manufacturing. He has published nearly 20 SCI-indexed papers in international journals.

Ying Yu received her bachelor's degree in industrial engineering from Beijing University of Science and Technology in 2009, and her master's and Ph.D. degrees in Industrial and Manufacturing Systems Engineering from the Department of Data and Systems Engineering, The University of Hong Kong, in 2014 and 2020, respectively. She is currently a Lecturer at the College of Economics and Management, Zhejiang Normal University. Her research interests include intelligent data analytics, supply chain management and logistics, and robotics.

Ray Y. Zhong (Senior Member, IEEE) is an Assistant Professor at the Department of Data and Systems Engineering, The University of Hong Kong (HKU). He has an interdisciplinary education background in Computer Science and Technology (BSc), Digital Signal Processing (MSc), and Industrial Engineering (PhD). Before joining HKU, he was a lecturer in Department of Mechanical Engineering, University of Auckland, New Zealand. Since joining HKU in January 2019, Ray won over HK$ 20 million external peer-reviewed competitive grants. Ray is one of well-known leading researchers in systems analytics with the following research fields: 2nd in Construction Informatics, 7th in Manufacturing Systems, 20th in Big Data Analytics from Google Scholar. He authored over 270 peer-reviewed journal and conference papers with total citation from Google Scholar is over 18,500 (H-index: 64, i10-index: 180), and is one of the HKU researchers ranked by Clarivate Analytics in the top 1% worldwide by citations since 2019.

George Q. Huang (Fellow, IEEE) joined Department of Industrial and Systems Engineering at The Hong Kong Polytechnic University in December 2022 as Chair Professor of Smart Manufacturing. Prior to this appointment, he was Chair Professor of Industrial and Systems Engineering and Head of Department in Department of Industrial and Manufacturing Systems Engineering at The University of Hong Kong. He gained BEng and PhD in Mechanical Engineering from Southeast University (China) and Cardiff University (UK) respectively. He has conducted research projects in areas of Smart Manufacturing, Logistics, and Construction through IoT-enabled Cyber-Physical Internet and Systems Analytics. His research has been supported with substantial government and industrial grants. He has directed a strong research team and collaborated closely with leading academic and industrial organizations through joint projects and start-up companies. He has published extensively and his works have been highly cited by research communities. He serves as associate editors and editorial members for several international journals. He is Chartered Engineer (CEng), Fellow of IEEE, IISE, ASME, HKIE, IET and CILT.