Blockchain-Enabled Privacy-Preserving Access Control for Data Publishing and Sharing in the Internet of Medical Things

Guangjun Wu, *Member, IEEE*, Shupeng Wang, *Member, IEEE*, Zhaolong Ning, *Senior Member, IEEE*, and Jun Li

***Abstract*—Recently, the rapid developments in the Internet of Medical Things (IoMT) enable smart devices to generate and transmit massive personal electronic medical records (EMRs). However, there are many sensitive attributes in an EMR, which could be accessed by external or internal unauthorized users for malicious purposes. In this article, we present a triple subject purpose-based access control (TS-PBAC) model, which is compatible with a blockchain-enabled reliable transaction network, and design an individual-centric security and privacy-preserving mechanism for access control with different purposes and roles in IoMT scenarios. Specifically, we design hierarchical purpose tree (HPT) and related policies to guarantee the legality of an external user with different purposes. To improve the privacy for sensitive attributes against an internal attacker, we design a local differential privacy (LDP)-based policy and role-based access control scheme in an edge computing paradigm to grant fine-granularity rights for authorized users. In addition, we introduce mutual evaluation metrics to evaluate data quality from a patient-and-medical-service level in an open anonymous network, only using logs kept in the blockchain. We test our approach by real-world EMRs with 100 000 patients. The experimental results show that the proposed privacy-preserving scheme can better protect patient's privacy than traditional access control policies in IoMT environments, and can make reliable and stable access control decisions between data publishers and data requesters with different purposes.**

***Index Terms*—Blockchain, electronic medical records (EMRs), Internet of Medical Things (IoMT), privacy-preserving information.**

I. Introduction

The Internet of Medical Things (IoMT) specifically designed for healthcare needs and medical applications has led to exciting achievements in healthcare monitoring, telemedicine consultation, and individual clinic [1], [2]. Smart devices along with wearables generate and deliver large volume and valuable electronic medical records (EMRs). Consequently, patients, doctors, and even academic researchers can cooperatively collect, exchange, and even share EMRs with common interests for convenient transmission, personalized diagnosis, and academic research applications. The collected EMRs from a patient include personal ID, address, age, gender, physiological, and diagnosis information. If these sensitive information is accessed by unauthorized users for malicious purposes, it may lead to personal privacy violation or even e-fraud events.

The smart IoMT systems consist of complex medical data application scenarios. We present three intuitive example cases.

1) *Case 1:* In a hospital-clinical system, physicians need to read the exact EMRs and update their diagnosis and drug treatment information. In this case, doctors need the read and write authorization on EMRs.
2) *Case 2:* In a pharmaceutic system, pharmacists need to read the drug treatment information signatured from a doctor, and they do not need to read the sensitive attributes of a patient.
3) *Case 3:* In an electronic survey system, researchers need to conduct a population survey statistic, such as biostatistical programs, infectious diseases, and drug development. In this case, they do not require the exact privacy attributes of a patient, and they can tolerate some errors of the statistical medical analysis.

Therefore, it is crucial to ensure the security, integrity, and privacy of EMRs in different IoMT systems with different authorizations. In this article, we aim to integrate the authorization management and privacy-preserving mechanism together to protect individual privacy from external and internal attackers in the process of cooperatively data collecting, publishing, and sharing in IoMT environments.

A. Motivation

To achieve security and privacy-preserving data publishing and sharing for different intensions in IoMT environments, two basic problems should be addressed, i.e., as follows.

1) An efficient authorization management (i.e., authorities grant and revocation) and privacy-preserving mechanism should be proposed for patients with massive smart

Manuscript received August 12, 2021; revised September 26, 2021 and November 11, 2021; accepted December 20, 2021. Date of publication December 24, 2021; date of current version May 23, 2022. This work was supported in part by the National Key Research and Development Program of China under Grant 2021YFB3101305; in part by the National Natural Science Foundation of China under Grant 61931019 and Grant 61971084; and in part by the Dalian Young Science and Technology Star under Grant 2020RQ002. *(Corresponding author: Guangjun Wu.)*

Guangjun Wu, Shupeng Wang, and Jun Li are with the Institute Information of Engineering, Chinese Academy of Sciences, Beijing 100093, China (e-mail: wangshupeng@iie.ac.cn; wuguangjun@iie.ac.cn; lijun@iie.ac.cn).

Zhaolong Ning is with the School of Software, Dalian University of Technology, Dalian 116620, China (e-mail: z.ning@ieee.org).

Digital Object Identifier 10.1109/JIOT.2021.3138104

devices, as well as users with different purposes [3]. These different purposes require different authorities, i.e., read authority (for healthcare purposes), read and write authorities (for diagnosis purpose), and download authority (for academic research purpose). These different usages would bring challenges for a unified framework with security and privacy-preserving policies for fine-granularity sensitive attribute protection for a patient.

2) A reliable evaluation metric for participants is in an open and anonymous environment. A patient wants to select a professional doctor for remote medical consultation. Meanwhile, an academic researcher wants to collect high-quality EMRs from patients for accurate medical statistics. A selfish data publisher might publish low-quality EMRs, and an incompetent consultant would present error consultation for individual benefits, which might aggravate stability of an open transaction network.

Medical data publishing and sharing between patients and doctors associated with security and privacy concerns have attracted considerable attentions [4], [5]. Methods of cryptography are typical solutions for sensitive information protection. However, cryptographic methods are not suitable for large-scale data publishing and sharing in the IoMT environment, owning to the disadvantage of its complicated nature, high level of computational overhead, and complex key protection and management problems. The access control mechanism is one of the promising security solutions to grant authorities for external legal users [6]. Traditional access control models [such as DAC, MAC, and role-based access control (RBAC) models] are identity based and the authorization is performed directly or via the roles assigned to the subjects. These models are suitable for a limited set of users with a small size of access lists. In an open network, such as the IoMT environment, where different parties of communications (patients, healthtakers, researchers, as well as smart devices and servers) are connected and unknown to each other, it would be better to determine access control rules based on the purposes of users [6], in addition to the roles of access devices [7]. Currently, the techniques of differential privacy (DP) and local DP (LDP) have been increasingly accepted as a de facto standard to protect user privacy in cloud computing [8] and edge computing [9] platforms.

These approaches can protect individual sensitive attributes from internal attackers in large-scale data collection and aggregation without relying on a trusted third party [10], [11]. Therefore, it is a promising work to introduce LDP into the start-of-the-art access control model to provide security and privacy-preserving protection simultaneously, and provide efficient and reliable framework for massive users with different application purposes.

To address the problem of mutual evaluation in an individual-centric transaction network, methods of matching decision are proposed recently [12], [13]. In an individual-centric transaction network, a seller publishes his goods, and a buyer broadcasts his interest of requirements. When the demand is matched, the requester will pay the publisher for the goods using money or coins. The fundamental elements of stable matching decisions are mutual evaluation metrics. However, the current metrics of mutual evaluation are conducted under a centralized manner, where buyers bid goods directly with each other without any privacy concerns. Thus, it is important to address the beneficial concerns in a matching decision with privacy elements and keep stability of a transaction network.

The emergence of blockchain has attracted growing attention and research work in the context of data sharing and exchanging, because of the characteristics of decentralization, anonymity, and tamper resistant. Blockchain can facilitate establishing a secure, reliable, and decentralized ecosystem [14], to address these problems. Some research studies are also conducted under the context of medical data management, such as data transmission, exchanging and sharing, clinical trial management, and medical record management [15]–[17]. Currently, blockchain applications are popular in the Internet of Things [18], [19] and intelligent wireless networks [20]–[22].

Motivated by the challenging problems of security and privacy preserving in IoMT scenarios, we design a blockchain-enabled framework using a novel access control model, and conduct purposes and roles-based access control, LDP privacy policies, and mutual evaluation of anonymous participants.

B. Contributions

The main contributions of this article are as follows.

1) We propose a triple subject purpose-based access control (TS-PBAC) model in IoMT environments. The model includes three submodels, which are deployed at data publishers' site (e.g., patients), blockchain site, and data requesters' site, respectively. The proposed TS-PBAC model can satisfy two requirements simultaneously: a) providing security and privacy-preserving access control decisions for different purposes and b) presenting mutual anonymous evaluation metrics for stable beneficial matching.

2) We design a hierarchical purposes tree (HPT) to depict the classification and relations of different purposes in TS-PBAC. Based on the proposed HPT, we design different purpose computation, matching, and purpose-based access control (PBAC) decisions. We introduce time-based authority revocation mechanism into the control decision, such that the published data can only be accessed in a time period as a publisher desired. We also design privacy policies merging and updating principles to make the minimum reserved policy modifications, when a patient updates his privacy preferences.

3) We present individual centric privacy-preserving policies and anonymous evaluation metrics. We design LDP-based privacy policies, such that a patient can define a local privacy budget to protect his sensitive attributes before data are published. Also, we design mutual evaluation metrics among data publishers and data requesters. Utilizing these descriptions and computations, we design a transaction matching decision algorithm to ensure the

convergence and stability of a network for trade-like purpose applications.

4) We conduct extensive theoretical and experimental analysis to examine the efficiency and effectivity of our approach. The theoretical analysis exposes that our approach can make secure and privacy-preserving data sharing among data publishers and requesters with different purposes. We also design a consortium blockchain-enabled prototype to keep and validate privacy preferences, transaction logs, authorization rules, etc. We evaluate system efficiency using real-world and synthetic data sets. The evaluated results show that our approach can make reliable and stable matchings between anonymous participants with applicable performance.

The remainder of this article is organized as follows. we first briefly go over system architecture of our approach in Section II, and present the detailed TS-PBAC model in Section III. We then present LDP-based privacy polices and PBAC decisions in Sections IV and V, respectively. The detailed evaluation metrics in the TS-PBAC model and related soundness analysis are presented in Sections VI and VII. Finally, we evaluate our approach with real-world and synthetic data sets in Section VIII.

II. System Architecture

Our prototype can has three parts: 1) data publishers; 2) blockchain-enabled middleware; and 3) data requesters. The data publishers, i.e., patients, have multiple IoT terminals, such as heart beat sensors, temperature sensors, etc., which generate large-volume data sets. We consider the edge computing paradigm, in which smart devices are first linked to an edge server, and the edge server is responsible for raw data collection and extraction. The blockchain-enabled middleware connects data publishers and data requesters, and provides security and privacy-preserving access control decisions, as well as reliable storage services between them. As the requester part, it constitutes data consumers, such as doctors, pharmacist, members of family, and even remote medical consulators. They interact with data publishers with the help of blockchains. The prototype is shown in Fig. 1. We mainly focus our work on the principle of security and privacy-preserving access control decision during the process of data publishing and exchanging. The work flow and components are shown in Fig. 2, and details are illustrated as follows.

1) *Edge Server:* A patient uses multiple smart devices and sensors to collect personal information, physical information, as well as diagnosis information. We use the edge computing paradigm, which provides the ability of local domain data processing and formal EMR extraction. The extracted records are pushed into blockchains, while the raw data sets are stored into security file servers. We also keep user's account, configure list of smart devices, and personal privacy preference information at the edge server.

2) *Blockchain-Enabled Manager:* We design a consortium blockchain to keep published EMRs, privacy policies, and model parameters of access control. The policies and accessing operations are conducted using smart contracts (SCs) by the consensus of the blockchain nodes. The manager serves as a middleware, i.e., blockchain together with security storage servers, which support large-file storage. The kept files are referenced by a pointer in EMRs, and the authorized users can assess the raw packed files. Some common files storage systems, such as interplanetary file system (IPFS), are proposed to provide versioning large files storage [23], [24].

3) *Anonymous Evaluation:* We design mutual evaluation metrics for anonymous participants. From the perspective of patient, he can evaluate a remote doctor repuation and medical experiences; from the perspective of data requester, he can quantitatively evaluate the effect of privacy disturbance from a patient using the utility function. Since evaluations are conducted using operations and logs kept in the blockchain, participants do not communicate directly.

4) *Hierarchical Purposes Tree Construction:* We divide application purposes into a series of different purposes with parent–child relationships. The HPT module accepts domine knowledge and medical subject value, and constructs different purposes in the format of a tree structure to support the PBAC decision.

5) *Policy and Decision Manager:* This module accepts privacy preference and policies from patients, classification

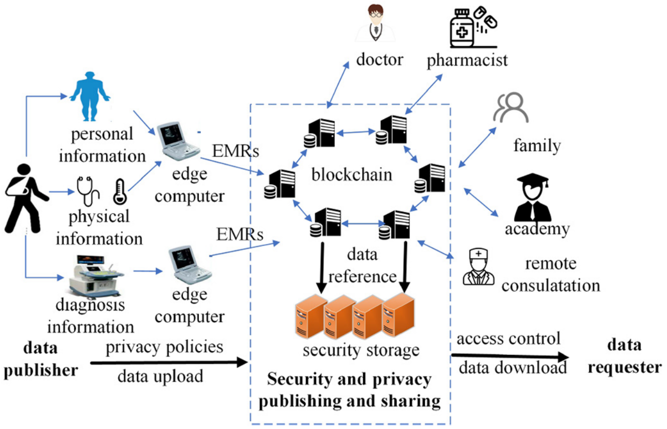


Fig. 1. Blockchain-enabled differential services for medical data publishing and sharing in IoMTs.

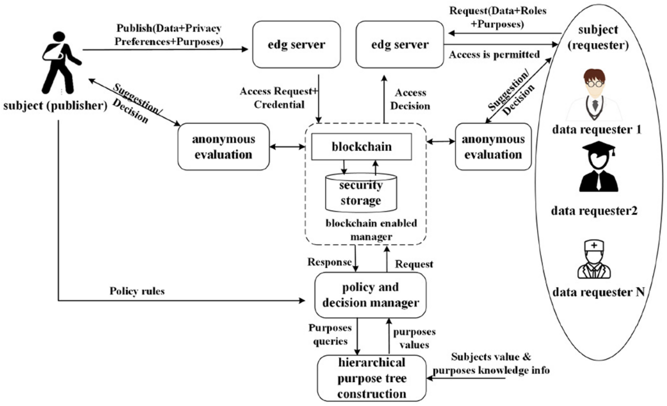


Fig. 2. Architecture and workflow of our prototype.

from HPT, and merges them to create access control policies. A patient configures LDP parameters for privacy attributes before data publishing, and adds his privacy rules into the manager. When a user launches a request, the manager makes a permit or forbidden decision according to privacy policies. Also, the manager keeps model parameters and privacy policies into blockchain ledgers. We intent to design SCs to interact with the blockchain-enabled middleware.

In this article, we first design the TS-PBAC model to filter out unauthorized illegal requesters. The purposes of TS-PBAC are predefined by medical knowledge. The purpose-based authorization can be considered as a type of coarse-grained classification. We further design the role-based privacy-preserving data access mechanism for permitted requesters after purpose-based authorization. The permitted requesters can read different versions of EMRs according to their role identifiers. For example, a permitted hospital doctor can read the exact version of EMRs, while a permitted hospital pharmacist can only read the LDP-based version of EMRs with the privacy-preserving mechanism.

III. TS-PBAC Model

The core components in TS-PBAC contain three parts: 1) working at data publisher site; 2) blockchain site; and 3) data requester site, and they are defined as follows.

Definition 1 [Purposed-Based Access Control at Publishers Site (PPBAC)]: A data publisher s_i publishes his data D with predefined purposes, roles, and the time when authorization revocation.

Formally, PPBAC $=$ $<D, \mathrm{IP}, R, T>$, where $D = \{r_1^i, r_2^i, \ldots, r_k^i, \ldots\}$, and r_k^i is the k_{th} data from publisher s_i; $\mathrm{IP} = \{\mathrm{EP}_{pt}, \mathrm{EP}_{fd}\}$, EP_{pt} is a set of the expected permit purposes and EP_{fd} is a set of the expected forbidden purposes; $R = \{\mathrm{role}_{pt}, \mathrm{role}_{fd}\}$ is a set that a publisher defines which roles can have specified authorizations; $T = \{t_s, t_r\}$, t_s is the time of published data, t_r is the length of authorized time, and after time t_r, the publisher revokes the authorization.

Definition 2 [Purposed-Based Access Control (RPBAC) at Requester Site]: A data requester p_j with role r_j requests data D with the intended access purpose AP at time t_j, i.e., RPBAC $= <D, \mathrm{AP}, r_j, t_j>$.

Using the definitions of PPBAC and RPBAC, we construct the blockchain site access control model to realize security and privacy-preserving access control between data publishers and data requesters.

Definition 3 [Purposed-Based Access Control (BPBAC) at Blockchain Site]: The submodel accepts the input privacy policies of PPAC from publisher s_i, with expected purpose EP, desired time period T, and target role R. Meanwhile, the submodel accepts access request of RPAC of a requester r_i with intended purpose AP for data D at time t_j. The objective of BPBAC is to make access decision using matching predicts, such as policy(IP, AP), policy(R, r_j), and policy(T, t_j), and provide anonymous evaluation metrics for data D and service level for r_j. Formally, BPBAC $=$ $<S, D, \mathrm{Policy}(<\mathrm{IP}, \mathrm{AP}>, <R, r_j>, <T, t_j>), \mathrm{Eval}(D/r_j)>$.

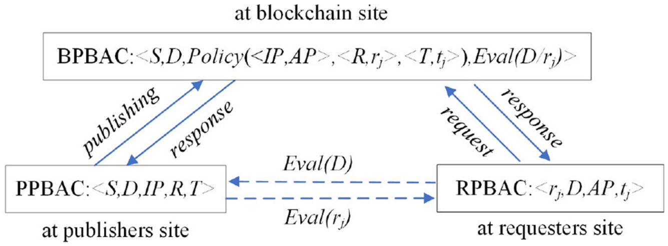


Fig. 3. Workflow of triple subjects in TS-PBAC model.

EMR data classification M{
I: ID information:{ID, account...}
D: Physical information{gender, age, time…}
C: Diagnosis information{temperature, disease treatment…}}

Fig. 4. Classification of EMR privacy attributes.

Considering these definitions and the framework proposed in the previous section, the formal definition of our TS-PBAC model is presented as follows.

Definition 4: The model of TS-PBAC includes three subject submodels, PPBAC, RPBAC, and BPBAC, which are deployed at data publisher site, data requester site, and blockchain site, respectively. TS-PBAC provides the interaction among three submodels, supports anonymous evaluation metrics Eval(D) and Eval(R), and makes reliable access and matching decision between participants with specified purposes.

The relationship of the three submodels in TS-PBAC is shown in Fig. 3. Notice that during the process of authorization, a data publisher and a data requester do not communicate with each other directly. The triple subject design in TS-PBAC can freely scale its framework to be compatible with the IoMT environment, where users increase the number of access terminals at the publisher site or at the requester sites freely. Next, we present details of the TS-PBAC model design, such as privacy preferences and policies, purposes-based access control, and the mutual evaluation metrics in Sections IV–VI, respectively.

IV. Priacy Policies in TS-PBAC

In this section, we specify LDP-based privacy preferences, which are predefined at the publisher site before data publishing. To establish a user-centric privacy protection strategy, we classify the personal privacy attributes in EMR M into a triple (I, D, C), where I indicates personal identifier information, which has a high-level privacy, such as name, social ID, account number, etc. D indicates an individual physical index information, which has a medium-level privacy, such as age, gender, etc. C indicates medical diagnosis and treatment information, which has a low-level privacy, such as treatment plan, medication and doctor diagnosis, etc. Each type of information can be further divided into several subattributes. Fig. 4 shows a schema of EMR privacy attribute classification. According to the identification, we design a specific LDP-based privacy preferences for high-level and medium-level sensitive attributes with different privacy budgets.

A. LDP-Based Privacy Preferences

At present, LDP [8] has been increasingly considered as a de facto standard to protect each user's privacy, when collecting aggregation information from a population, without relying on a trusted third party [10]. We apply local different privacy protection strategies for high-level and medium-level privacy attributes before data publishing to prevent the internal or even conspired attackers.

If the patient disturbs the sensitive attributes with predefined privacy budget before data publishing, a requester can compute aggregates (i.e., count, sum, avg, frequency, etc.), with error-guaranteed estimates from the collected data, while a requester cannot infer the patient sensitive attributes. The frequency estimation is a basic task for aggregation, and other complex problems relying on it, such as heavy hitter identification and frequent itemset mining, can also be improved. We first present a formal LDP-based sensitive attribute protection definition.

Definition 5: Assuming that a sensitive attribute of an EMR published from a publisher is encoded into V, algorithm A adds noise to V, $A(V) \rightarrow Y$, Y is sent to a receiver, for $\forall y \in Y$, $\forall v1, v2 \in V, v1 \neq v2$, when algorithm A satisfies $([P(A(v1) = y)]/[P(A(v2) = y)]) \leq e^{\epsilon}$, then it satisfies the condition of ϵ-LDP.

The definition above constrains the disturbance impact. Since the noise added by the ϵ-LDP algorithm is random, the distribution of the original data cannot be inferred even when the complicity of multiple requesters occurs. The error can be guaranteed when using output Y for statistical analysis. Unary encoding is leveraged to implement LDP privacy protection.

We define the input value from the publisher as v, and the LDP disturbance probability as p and q, respectively. The privacy budget ϵ can be expressed as follows:

$$\epsilon = \ln\left(\frac{p(1-q)}{(1-p)q}\right). \tag{1}$$

The input value is encoded as $\mathrm{Encode}(v) \rightarrow x$. Then, we digitize the publisher's data value to generate vector x with dimension d, which is shown as $x[0, 0, \ldots, 1, \ldots, 0]$. The vth bit is 1 and the rest is 0. To protect publisher's privacy, we add disturbance $x \rightarrow x'$, x' satisfies

$$\mathrm{pr}[x' = 1] = \begin{cases} p, & \text{if } \ x(i) = 1 \\ q, & \text{if } \ x(i) = 0. \end{cases} \tag{2}$$

A compression algorithm, such as snappy [25], can be used to compress x' before publishing it. Correspondingly, the requester decompresses the received data before analysis. The recipient uses the set of vectors to perform statistics, e.g., counting frequency of a publisher appears with value i

$$\widetilde{c}(i) = \frac{\sum_j Y(i) - nq}{p - q}. \tag{3}$$

In (3), the first term represents the statistics of all vectors whose ith bit is 1, and n is the number of collected vectors collected (i.e., the number of publishers). The second term in the denominator represents the noise reduction introduced by nq.

B. Privacy Policy Management

After a patient disturbs his sensitive information, he will set privacy preference along with his published data. Our model TS-PBAC generates privacy policy PPSet_D for data D, and keeps them in the blockchain (storing full policies) and edge server (cached for partial users in a domain), respectively. TS-PBAC creates one set of privacy policies for data D. When a patient updates his privacy preference of D, the privacy policies maintained at edge server and blockchain should be modified accordingly, i.e., the modification is done on PPSet_D, including removing, adding, and merging policies. A privacy policy set PPSet_D for D can be updated by a new policy PP_1 in the following principles.

1) *Removing a Privacy Policy:* A privacy policy rule PP_1 can be removed from privacy policy set PPSet_D, if $\mathrm{PP1} \in \mathrm{PPSet}_D$. In this case, we have

$$\mathrm{PPSet}_D \leftarrow \mathrm{PPSet}_D - \mathrm{PP}_1. \tag{4}$$

2) *Adding a New But No Conflicting Privacy Policy:* When a privacy policy rule is required to be added to PPSet_D, it is necessary to check that whether it is duplicated or could be interfered from the previously defined privacy policy rules in PPSet_D. If there exists no similar rule in the set, the new privacy policy rule is added to set PPSet_D. The formal definition is as follows:

$$\begin{aligned} &\mathrm{PPSet}_D \leftarrow \mathrm{PPSet}_D \cup \mathrm{PP}_1 \\ &\text{For } \ \forall \mathrm{PP}_i, \mathrm{PP}_i \in \mathrm{PPSet}_D, \mathrm{PP}_i \neq \mathrm{PP}_1. \end{aligned} \tag{5}$$

3) *Adding New and Conflicting Privacy Policy:* When the new rule PP_1 is conflicting with an existing policy rule PP_2, we use a similar operation for administrative rules in cloud environments as in [6]. Also, we add role-based and privacy budget parameters into the merging principles. We can merge them together to construct a new policy rule. Define a permit purpose as EP_{pt}, forbidden purpose as EP_{fd}, and the formal definition is as follows:

$$\begin{aligned} &\mathrm{PPSet}_D \leftarrow (\mathrm{PPSet}_D - \mathrm{PP}_2) \cup \\ &\mathrm{PP}_1 = \begin{cases} <\mathrm{EP}^1_{pt} \cap \mathrm{EP}^2_{pt}>, & \text{if } \mathrm{EP}^1_{fd} = \varnothing \wedge \mathrm{EP}^2_{fd} = \varnothing \\ <\mathrm{EP}^1_{fd} \cup \mathrm{EP}^2_{fd}>, & \text{if } \mathrm{EP}^1_{pt} = \varnothing \wedge \mathrm{EP}^2_{pt} = \varnothing \\ <\mathrm{EP}^1_{pt} - \mathrm{EP}^2_{pt}>, & \text{if } \mathrm{EP}^2_{pt} = \varnothing \wedge \mathrm{EP}^1_{fd} = \varnothing \\ <\mathrm{EP}^2_{pt} - \mathrm{EP}^1_{pt}>, & \text{if } \mathrm{EP}^1_{pt} = \varnothing \wedge \mathrm{EP}^2_{fd} = \varnothing \end{cases} \\ &R_{\mathrm{new}} \leftarrow R_1 \cap R_2 \\ &\epsilon_{\mathrm{new}} \leftarrow \min\{\epsilon_1, \epsilon_2\} \\ &T_{\mathrm{new}} \leftarrow < \max\{t_{s1}, t_{s2}\}, \min\{t_{r1}, t_{r2}\}>. \end{aligned} \tag{6}$$

V. Access Control Decisions in TS-PBAC

The access control decision is made by the predict of expected purpose *IP* from patients and access purposes *AP* from requesters. In this section, we present the different purpose classification and decision principles in TS-PBAC. The main notations used in trade-like purpose decision are shown in Table I.

Let a general purpose set be *GP*, and *GP* is a set of objects with parent–child relations. An intuitive way of expression is to organize elements and their relationships of *GP* by a

TABLE I
MAIN NOTATIONS

Notations	Meanings
GP	The whole set of general purposes
IP	The set of expected purposes defined by a data publisher
AP	The set of access purposes defined by a data requester
DMP	The set of expected purposes depicted in HPT
S	The set of data publishers, S_i represents the i-th publisher
P	The set of data requersters, P_i represents the i-th requester
Φ	The set of matching functions
U	Utility function
γ_i^k	Published data and $\|\gamma_i^k\|$ is its size
$\beta_{i,j}^k$	The connection between data publisher i and requester j
v	Normalization coefficient
$\Re$	Computation function of publisher reputation
Q	Satisfaction level evaluation function
ω, λ	Discretization coefficient
ζ	The set of matching vectors
Δ	Step price
V	Sensitive attributes of publisher information
Y	Privacy attribute collected at the requester
p	Disturbance probability
ϵ	Privacy budget
H	The set of wallets

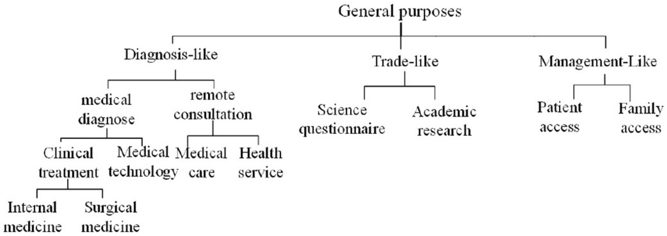


Fig. 5. Example of medical HPT.

hierarchical tree structure. As shown in Fig. 5, each node in the purpose tree represents a purpose in the *GP*, and each edge in the tree represents the parent–child relationship between two purposes. Also, in the perspective of medical application, the set *GP* can be divided according to medical knowledge and different applications; therefore, a node of the tree can be subdivided according to medical scenarios.

In this article, we mainly design three major classifications for different purposes and related usages; the first type is diagnosis-like purposes, which are related to medical treatment and healthcare between patients and doctors; the second type is trade-like purposes, which are related to financial transactions between patients and academic researchers or a remote consolation; and the third type is management-like purposes, which are related to manage patient himself data or accessed by the members of patient's family.

A purpose node can be further divided into child purposes. For example, we can design "late-stage cancer" and "early-stage cancer" as two child nodes succeeding from the parent node "cancer" in the HPT, such that the matching and evaluation between patients and doctors are more accurate. TS-PBAC can apply different access control policies on a purpose node and they are discussed in the following sections.

A. Diagnosis-Like Purpose Decision

The goal of diagnosis-like purpose is to make a fast access control decision for a doctor. Recall that individual purposes are defined in the privacy preference policies from a patient, together with the published data sets. Meanwhile, the access purpose (AP) is proposed when a data requester (i.e., a doctor) requires desired data. Thus, the diagnosis-like purpose decision is first to judge whether the AP from a requester is matched with the IP predefined by a data publisher, and make a permit or reject decision. Based on the previous definition, we can define the purpose-based decision. Define different medical purpose (DMP) as a subcollection of GP, and DMP is inferred from the purpose IP defined by a data publisher. Using the predefined medical HPT structure shown in Fig. 5, the DMP can be expressed as follows.

1) Let DMP be a set of different purposes, and each *P* in DMP represents a purpose.
2) Let child(*P*) be a set of purposes including *P* and all *P*'s descendant nodes are in the purpose tree HPT.
3) Let parent(*P*) be a set of purposes consisting of *P* and all *P*'s ancestors in the purpose tree HPT.
4) Let related(*P*) be a set of purposes consisting of *P*, *P*'s ancestors, and *P*'s descendants in the purpose tree HPT, i.e., Related(*P*) = child(*P*) ∪ parent(*P*).

Recall that an expected purpose set *IP* from a publisher can be expressed as $<\mathrm{EP}_{pt}, \mathrm{EP}_{fd}>$, where EP_{pt} is a set of expected permit purposes in *IP*, and EP_{fd} is a set of forbidden purposes in IP. According to previous assumptions, the purposes in a HPT can be divided into three parts according to a given IP: 1) EP^{α}; 2) EP^{β}; and 3) unspecified purpose EP^{γ}. EP^{α} and EP^{β} are sets of permitted purposes and forbidden purposes, respectively. Formally, $\mathrm{EP}^{\alpha} = \mathrm{child}(\mathrm{EP}_{pt}) - \mathrm{related}(\mathrm{EP}_{fd})$, i.e., EP^{α} includes all child permitted purposes while excluding forbidded purposes; $\mathrm{EP}^{\beta} = \mathrm{related}(\mathrm{EP}_{fd})$, i.e., if purpose EP_{fd} is forbidden, all EP_{fd} related purposes are also forbidded purposes.

Therefore, the PBAC decision can be made using the parent–child relationship. A data requester can access the data if and only if $\mathrm{AP} \in \mathrm{EP}^{\alpha}$. Combining with the specific needs of secure access control, TS-PBAC defines unspecified purpose EP^{γ} as forbidded purposes, i.e., if $\mathrm{AP} \in \mathrm{EP}^{\gamma}$, a data requester cannot access the data.

B. Management-Like Purpose Decision

A patient can use different smart devices to connect to the blockchain to publish individual medical data. Also, a patient can delete or modify the published data set. Therefore, different users or devices need different authorizations for data management purposes, such as authority of download, update, read, and write. We introduce role-based authority assignment for massive IoMT smart devices, together with different authorized users, to compact the length of access control list for management-like purpose decision.

In terms of RBAC policy in our model, it first makes a purpose-based decision. If the decision is permitted, then we use the RBAC policy to determine the specified authority for the requester. The core idea of RBAC is to assign a role with a specified authority, and the same role identification can be assigned to a family of different devices. As shown in Tables II and III, the RBAC model can compact the access control list to

TABLE II
SMART DEVICES ACCESS CONTROL LIST

Devices or users	Location	Role ID
smart device 1	access point 1	$role_1$
smart device 2	access point 2	$role_1$
doctor 1	hospital 1	$role_2$
doctor 2	hospital 2	$role_2$
remote doctor 1	network access 1	$role_3$
remote doctor 2	network access 2	$role_3$
family member 1	access point 3	$role_4$
family member 2	access point 3	$role_4$

TABLE III
ROLES-BASED ACCESS CONTROL LIST

Role ID	Authorities
$role_1$	write
$role_2$	read + write
$role_3$	read
$role_4$	read + download

a compact RBAC list, and the same role identification, such as role_1 and role_2, can be assigned to a family of devices. In our model, the smart device configure information, i.e., Table III, is stored in an edge server, only the RBAC list is kept at the blockchain. When a new smart device connects to the blockchain, the edge server can assign a role identification to the new device, and there is no need to change the predefined privacy policies at blockchain.

C. Trade-Like Purpose Decision

The goal of a trade-like purpose decision is to make beneficial matching between data publishers and data requesters to obtain the maximum benefit.

We use the same symbols as in the matching theory literature [12], [26], and make necessary supplementaries. We design an algorithm to realize an automatching decision (shown in Algorithm 1). The main steps are listed as follows.

1) Step 1 is an initial stage of the matching algorithm. We empty $\zeta_{i,j}^{k,t}$, and set $\beta_{i,j}^{k,t}$, MATLIST_c, and SLIST_s be initial values.
2) Step 2 is to achieve mutual benefit via bargaining price. A publisher gives an initial price of data via a data utility function. Requester s_i checks his wallet to find whether he has enough money to meet the evaluation requirement. Once a requester s_i has enough money, he can calculate his candidate via goods demand function $\Omega_i(\beta_i^t, k)$

$$\Omega_i\left(\beta_i^t, k\right) = \left\{\begin{matrix} u^{\text{REQ}}, & \text{if } u^{\text{REQ}} > 0 \\ s_0, & \text{otherwise} \end{matrix}\right\}. \tag{7}$$

By this computation, we can filter out illegal requesters, who do not have enough money to bargain price.
3) Step 3 is a typical bidding routine. After requesters express their wishes, a publisher can make its matching choice. In each iteration, we only make one matching decision between a publisher and a requester until the candidate set of publisher or requester is empty, or the money of the requester is exhausted.

Algorithm 1 ITM Algorithm for the Trade-Like Purposes

```
Input: $S$, $P$, $\Gamma$, $H$.
Output: $\Phi$, $\zeta_{m,n}$.
Step 1: Initialization
for $\gamma_k \in \Gamma$ do
    Set $\beta_{i,j}^{k,t} = \beta_{min,j}^{k,t}, \zeta_{i,j}^{k,t} = 0, \forall i \in \emptyset^s, j \in \emptyset^P$;
    Set $\text{MATLIST}_c = \left\{p_j\right\}_{j=1}^{|\text{P}|}$, $\text{SLIST}_s = \{s_i\}_{i=1}^{|S|}$.
end for
Step 2: Bargaining price
for $\gamma_k \in \Gamma$ do
    $p_j \in \text{MATLIST}_c$ announces $\beta_{i,j}^{k,t}$ to each $s_i$;
    for unmatched requester $s_i, \forall i \in \emptyset^s$ do
        Calculate its demand $\Omega_i\left(\beta_{i,k}^t\right)$;
        if $\Omega_i\left(\beta_{i,k}^t\right) = p_0$ then
            Set $\zeta_{i,j}^{k,t} = 0, j \in \emptyset^\theta$, $\Phi^k(s_i) = \{p_0\}$;
            Remove $s_i$ from SLIST $_s$;
        else
            if wallet($s_{i,k}^t$) has enough money then
                $s_{i,k}^t$ bids for $p_j$, set $\zeta_{i,i}^{k,t} = 1, j \in \Omega_i\left(\beta_{k,t}^t\right)$.
            end if
        end if
    end for
end for
Step 3: Bidding step
for $\gamma_k \in \Gamma$ do
    for $p_i \in$ MATLIST $_c$ do
        if $\sum_{i=1}^m \zeta_{i,j}^{k,t} = 0$ and $\beta_{i,j}^{k,t} > \beta_{\text{min},j}^{k,t}$ then
            $\Phi^k\left(p_j\right) = \left(s_{i*}, \beta_{i*j}^{k,t-1}\right)$, $i* = \max\left\{\Phi_{j\to i}\right\}$;
            Remove $s_i$ from SLIST $_s$;
            if $|\Phi^k\left(p_j\right)| = q_j^k$ then
                Remove $s_i$ from SLIST $_s$;
                Set $\zeta_{i*,j^+}^{k,t} = 0$, where $j^+ = \Omega_{i*}\left(\beta_{i*}^{k,t}\right)$;
            end if
        else if $\sum_{i=1}^m \zeta_{i,j}^{k,t} > 1$ then
            Set $\zeta_{i,j}^{k,t+1} = \zeta_{i,j}^{k,t} + \Delta$;
        else if $\sum_{i=1}^m \zeta_{i,j}^{k,t} = 1$ then
            $\Phi^k\left(p_j\right) = \left(s_i, \beta_{i,j}^{k,t}\right)$, where $i = arg\zeta_{i^-}\zeta_{i^-,j}^{k,t} = 1$;
            Remove $s_i$ from $\text{SLIST}_s$;
            if $|\Phi^k\left(p_j\right)| = q_j^k$ then
                Remove $p_j$ from $\text{MATLIST}_c$;
            end if
        end if
    end for
end for
for $\gamma_k \in \Gamma$ do
    if $\text{MATLIST}_c = \emptyset$ or $\text{SLIST}_s = \emptyset$ or H(SLIST)=$\emptyset$ then
        Terminate the algorithm.
    else
        Set $t = t + 1$ and go back to Step 2;
    end if
end for
```

We briefly analyze the complexity of the algorithm. The complexity of algorithm initialization in step 1 is $|\gamma_k| \times \max\{|P|, |S|\}$. The mutual bargaining times of step 2 can be expressed as $\gamma_k \times |P| \times |S|$. In step 3, a requester can bid all data and select the desired data as requirement; thus. the bidding times can be depicted as $|\gamma_k| \times |P|$. Therefore, by combining the complexity in each step, the complexity of our proposed matching algorithm is $t \times |\gamma_k| \times |P| \times |S|$, where t is the number of iterations.

We next briefly analyze the value of t. Because the algorithm iteration candidate set is not empty, we express limiting condition as $U^{\text{REQ}} \geq 0$, i.e., iteration terminates when $U^{\text{REQ}} < 0$. The initial value is $\beta_{\min}$, we increase Δ each time, and the maximum number of iterations can be expressed as: $t_{\max} = (1/\Delta)(u^{\text{REQ}} - \beta_{\min})$, i.e.,

$$t_{\max} = \frac{1}{\Delta}\left(\mu Q_{s_i^k}\left(p_j^k, \gamma_k\right)/S_{\gamma_k} - \beta_{c,j}^k - \beta_{\min}\right). \tag{8}$$

VI. Evaluation Metrics in TS-PBAC

We design two major evaluation metrics Eval(D) and Eval(R), to evaluate the quality of privacy disturbed data and the level of medical service of a remote doctor, respectively. Note that the evaluation is conducted under anonymous networks, and a publisher and a requester cannot communicate directly. We design an evaluation measurement, which only uses statistics of historical transactions.

A. Data Quality Evaluation From Requesters

We quantitatively express elements of data quality from a patient, such as utility function, privacy budgets, and reputation of a data publisher in the view of data requesters.

Definition 6 (Degree of Data Satisfaction): We establish a 0-1 distribution sigmoid function to represent the satisfaction level of data requester p_i^k to data γ_i^k released by data publisher s_i^k as

$$Q_{s_i^k}\left(p_j^k, \gamma_i^k\right) = \alpha \frac{1}{1 + e^{-\omega\left[\mathrm{R}_{j\to i}\left(p_j^k \cdot v_k\right) - \mathrm{R}_{j\to i}^{\min}\left(\gamma_i^k\right)\right]}} + (1-\alpha)\frac{1}{1 + e^{-\lambda\left[\epsilon_{\mathrm{i}\to j}\left(p_j^k \cdot v_k\right) - \epsilon_{i\to j}^{\min}\left(\gamma_i^k\right)\right]}}. \tag{9}$$

In (9), Q is a format of sigmoid function. ω and λ are discretization coefficients, and α is a normalized coefficient to make value Q arranged in [0, 1]. s_i^k indicates the ith data provider, sharing data γ_k with data requester p_j^k; $\mathrm{R}_{j\to i}^{\min}(\gamma_k)$ is the value of minimum expectation, and $\varepsilon_{i\to j}^{\min}(\gamma_k)$ is the value of expected privacy budget.

In (9), R is the reputation of a data publisher computed from the statistics of historical transactions. $R_{j\to i} = |CL_{P_{j\to i}}| + |CL_{N_{j\to i}}|$, and $CL_{P_{j\to i}}$ represents the ratio of requester j' successful transactions with publisher i to all requesters' successful transaction with publisher i; $CL_{N_{j\to i}}$ represents the ratio of requester j' failed or negative transaction with publisher i to the ratio of all requesters failed or negative transaction for publisher i. Using the previous definition, we present the data evaluation metrics from a requester.

Definition 7 (Data Quality Evaluation): A data requester p_i^k accepts data r_i^k from a publisher s_i^k and the evaluation function is defined as

$$\text{Eval}^{\text{REQ}}\left(r_i^k\right) = \mu_{s_i^k}^{\text{REQ}}\left(p_i^k, \gamma_i^k\right) = \mu Q_{s_i^k}\left(p_j^k, \gamma_i^k\right) - \beta_{i,j}^k\left|\gamma_i^k\right|. \tag{10}$$

In (10), the first item indicates the degree of satisfaction of requester p_i^k about data γ_i^k; the second item indicates the consumption of accepting data with size $|\gamma_i^k|$, and $\beta_{c,j}^k$ is a connection consumption between the requester and system. A patient can increase the value of privacy budget to improve data quality for a data requester, while it will decrease the level of privacy protection for sensitive attributes.

B. Medical Service Evaluation From Patients

We consider the differential service levels of a remote doctor from two aspects of evaluations. The first is the purpose and rule-based classification among profession groups; the second is the detailed service-level evaluation within the same profession group.

1) Purpose and Role-Based Classification Among Profession Groups: We use the purpose-based classification to decide whether the requester can obtain authorization for an EMR. However, the purpose-based classification is coarse grained, such that different medical professions might have different evaluation metrics. We also propose the fine-grained role-based selectivity to distinguish professions with the same purpose. A patient can select a remote doctor profession via the registered role identifier, and filter out the unwanted doctors, such that a headache patient does not select a doctor with cancer identifier.

2) Service-Level Evaluation in the Same Profession Group: A patient needs a measure to evaluate a remote unknown doctor whether the unknown doctor has the ability to provide the professional healthcare or consultation. We design an anonymous and offline metric to evaluate a remote unknown doctor via historical transactions. More details are as follows:

$$\text{Eval}^{\text{PUB}}(p_i) = \left(a\left(\frac{\text{TN}_{p_i}^{\text{suc}}}{\text{TN}_{p_i}^{\text{tot}}}\right) + b\left(\frac{\text{TH}_{p_i}^{\text{suc}}}{\text{TH}_{p_i}^{\text{tot}}}\right) + cEV_{p_i}\right) \tag{11}$$

where:

1) $(\text{TN}_{p_i}^{\text{suc}}/\text{TN}_{p_i}^{\text{tot}})$ is the ratio of the successful number of transactions of doctor p_i to his total number of transactions (successful and unsuccessful transactions). This index reflects that if a remote terminal is honest, such that an attacker of Denial of Service (DoS), which has multiple historical unsuccessful requests or transactions, can be filtered out;
2) $(\text{TH}_{p_i}^{\text{suc}}/\text{TH}_{p_i}^{\text{tot}})$ is the ratio of the successful number of transactions of doctor p_i to the total number of successful transactions in his department or hospital. This value indicates the influence of a doctor in his department;
3) EV_{p_i} is the medical experience value of a remote doctor. We use an anonymous method to evaluate the professional experience for an unknown doctor, such as ranks, titles, and the time when a doctor starts medical profession. This index is designed to depict a doctor medical experience in his profession;
4) a, b, and c are coefficient weight parameters, such that $a + b + c = 1$.

A patient can predefine a metric value for a role of a remote unknown doctor, along with his privacy preferences. If a remote doctor's AP matches the expected purpose EP and the evaluated metric value is larger than the predefined value from a patient, the remote doctor can obtain the authority.

TABLE IV
PRIVACY PRESERVED ACCOUNT ADDRESS DESIGN

security parameters	contents
security seed	xpjJ2s6FxKybu7M6shKXhDTK5E9Tr
account address	zQ9UuJcwCsg2CsBWjoqCaSvDtuLihxnuPS
public key	n9J8eWbJ1VimZ6fa7Piz4oFHbD6rmiXQz9cpzjigJmBNksnvMbur
private key	pcTtUibGougNgjMohyAcHcdfnZpvie6kWqR3DnwoWK7XW4NxJCU

Note that the metric value of computation and evaluation is from the logs and preference kept in the blockchain; therefore, we can conduct offline authorization management between a patient and a remote doctor.

VII. SECURITY ANALYSIS

In this section, we present a logical analysis of the proposed approach from three aspects: 1) soundness of access control model; 2) LDP-based privacy-preserving policies; and 3) stability of transaction network.

A. Soundness of Access Control Model

A common architecture of data collecting and publishing in IoMT might consider two fundamental requirements: 1) security and 2) availability. The feature of security guarantees that the published data are shared between authorized users only. The TS-PABC model can be considered as a combination of PBAC and RBAC model [6]. In our model, the patient data are only shared to the allowed requesters, who are matched with the purpose set IP as a patient predefined. Thus, our model can protect the typical external threats, such that the published data can only be accessed by an authorized user with permitted purposes and authorized roles. The feature of availability guarantees that a user can access the network and upload/download data at any time. Our model is suitable for edge computing paradigm to boost the performance of data collecting and exchanging with massive smart devices. The smart device list is kept in an edge server, and only the model parameters and privacy policies are kept in the blockchain. A user can connect any node of the blockchain to validate the legality and search the on-chain EMRs. Compared with traditional systems, the data are stored in a central location and may affect the system availability. Additionally, the users of TS-PBAC are anonymous and logs of operation are kept in the blockchain. Such design can provide system-level security and event traceability for different anonymous users in IoMTs.

B. Malicious Evaluation Prevention

The "malicious evaluation" exists in real life and it may be created by external and internal users. We prevent the malicious evaluation from external and internal users simultaneously via two designs.

1) *Secure Account Address Design:* In our blockchain prototype, the system creates a root account and multiple user accounts. The root account is used in the initialization work for a user account when a user first joins into the system. If a new user intends to join into the system, the agent of the user fetches the predefined cryptology algorithm, such as secp256k1, to create its local account address. After the local address is created successfully, the root account will transfer some predefined money/coins to the new account address. Then, the user agent can conduct transactions with other account address freely. The detailed information of a user is not required in the initialization work and the first-round money transfers. This design keeps the user privacy in an anonymous transaction network. An example of a generated account address from security seed using secp256k1 is shown in Table IV. The agent of a user account records the user account address, public key, private key, role information, etc. The agent of the root account can search the full ledger. We design SCs to compute the statistical information using the root account. When a user wants to connect to the blockchain system, the agent of the user posts parameters, i.e., the account address and public key to the blockchain. If the validation passes, the agent can connect to the blockchain and conduct the following operations. Thus, only legal internal users can connect to the blockchain and obtain the statistics of transactions.
2) *Reliable Validation and Evaluation:* The ledgers are kept in all peers of a consortium blockchain system. A peer might join the blockchain, and accept and validate the integrity of ledgers. The elements of evaluation metric, such as coefficient weights and function bodies [depicted in (11)], are implemented with SCs in solidity language, which can also be encoded into bytecodes and deployed over the blockchain. To evaluate a remote doctor, the agent of a user account can call the address of predefined SCs to request the statistical information from blockchain. The core elements of evaluation metrics, such as transaction logs, coefficient parameters, and function bodies, are all kept in the blockchain, which cannot be tampered by malicious users. Thus, the evaluation result is reliable for participants.

C. LDP-Based Privacy Preserving

We introduce LDP into traditional PBAC and RBAC models, such that a patient disturbs his sensitive attributes before publishing data. This improvement can enable a patient to rapidly publish his data using IoT devices without the complicated nature of computation overhead and key management using cryptology methods. Recall that the sensitive attributes of EMRs are divided into three parts, i.e., $M(I, D, C)$. For attributes I, the edge server uses hash-based pseudonym instead of practical ID to connect with the blockchain, such that a requester cannot obtain the real ID of a patient in the following transactions. For personal sensitive information D, we design privacy policies using LDP, which disturbs sensitive information using randomized parameters (p, q) locally. This design can ensure that a requester cannot infer the individual sensitive information from the collected data, for the reason that each bit in the collected vector is randomized with probability p or q. Finally, we design SCs to keep the privacy policies, function body access controls, as well as the key parameters for evaluations. The SCs are deployed and implemented over blockchain. Only authorized users can access the published data and all the operations are also kept in blockchain to ensure traceability.

Next, we briefly discuss evaluations about the data quality of collected data in the view of a requester. We adopt L_2 error metrics, i.e., variance Var[.], to present the accuracy of estimation at a requester. According to Definition 1, let input values be v_1 and v_2, and the output be Y. Since each bit is independently and identically distributed, according to the probability distribution [10], the variance of frequency $\tilde{c}(i)$ can be depicted as

$$\mathrm{Var}\left[\tilde{c}(i)\right] = \frac{nq(1-q)}{(p-q)^2} + \frac{nf_i(1-p-q)}{p-q} \tag{12}$$

where p and q are LDP disturbance probabilities, n is the number of data collected at the requester, and f_i is the fraction that true value i occurs. Since f_i is generally small, (12) can be expressed as

$$\mathrm{Var}\left[\tilde{c}(i)\right] = \frac{nq(1-q)}{(p-q)^2}. \tag{13}$$

We now briefly discuss the value assignment of p and q. In the Rappor algorithm [11], $p + q = 1$, p is larger than q, and $(p/q) <= e^{\epsilon}$, where ϵ is depicted as privacy budget in LDP techniques. A typical assignment is $p = 1 - (f/2)$, $q = (f/2)$; thus, $\epsilon = \ln(([1 - (f/2)]/(f/2))^2)$.

An optimal assignment of p and q is presented in [10]. The minimum variance value is achieved via the derivation of variance in (13), at this time $p = (1/2)$ and $q = [1/(e^{\varepsilon} + 1)]$.

D. Stability of Transaction Network

We now present detailed analysis for a transaction network with trade-like purposes. We introduce mutual beneficial trade-like purpose decision and the related evaluation metrics, which can accommodate our model to individual-centric applications in IoMT. For the decision of trade-like purpose, both data publisher and data requester intend to maximize their benefits. A publisher and requester might leave existing matching objects, when there are better choices, and it causes the entire network not to be converged [12], [27]. A transaction network can be depicted as $G(S,\ P,\ \phi,\ U)$, where S represents a set of data publishers, and $S = \{s_1, \ldots, s_i, \ldots\}$, s_i represents the ith data publisher; P represents a set of data requesters, $P = \{p_1, \ldots, p_j, \ldots\}$, p_j represents the jth data requester; and ϕ represents a matching relationship between a requester and a publisher depicted by the number of matches successfully. The traditional matching network has a one-to-one or one-to-many matching relationship. Let s_0 be a special requester and p_0 be a special publisher without any transactions; thus, ϕ can be expressed as $S \times P \rightarrow S \times P$; U represents the utility measurement in a transaction network. In an EMR data transaction network, $U = \{\mu^{\mathrm{PUB}}, \mu^{\mathrm{REQ}}\}$, where μ^{PUB} and μ^{REQ} represent benefits brought into a network by the publisher and requester, respectively. As with μ^{PUB}, it is the benefit for a patient to publish his data. It can be zero for diagnose-like and management-like purposes usage, while for trade-like purposes, the benefit can be depicted as follows.

Definition 8 (Utility Function From a Publisher): When the data publisher s_i^k and data requester p_j^k transmit data γ_i^k, the utility function is $\mu_{p_i^k}^{\mathrm{Pub}}(s_i^k, \gamma_i^k)$, which is expressed as

$$\mu_{p_j^k}^{\mathrm{PUB}}\left(s_i^k, \gamma_i^k\right) = \beta_{i,j}^k\left|\gamma_i^k\right| - \nu\beta_i^c\left|\gamma_i^k\right| \tag{14}$$

where the first item represents the benefit for sharing data γ_i^k between publisher i and requester j; the second item represents the cost that publisher i publishes the data, such as the fee or gas in a public blockchain, and β_i^c represents the connection between data publisher and EMR blockchain. At present, a client connects blockchain with multiple synchronization mechanisms, such as synchronization, asynchronous, and blocking manner. Each mode has a different cost. $|\gamma_i^k|$ represents the size of data γ_i^k, and ν is a normalized coefficient. Therefore, we can depict the maximum benefit of a transaction network as

$$\begin{aligned}
&\max\ \ \mathrm{G}(S, P, U)\\
&\quad= \max\sum_{k=1}^{T}\sum_{i=1}^{|S|}\sum_{j=1}^{|P|}\zeta_{ij}^k\left[\mu_{p_j^k}^{\mathrm{PUB}}\left(s_i^k, \gamma_i^k\right) + \mu_{s_i^k}^{\mathrm{REQ}}\left(\gamma_i^k\right)\right]\\
&\quad= \max\sum_{k=1}^{T}\sum_{i=1}^{|S|}\sum_{j=1}^{|P|}\zeta_{ij}^k\left[\beta_{i,c}^k\left|\gamma_i^k\right| - \nu\beta_i^c\left|\gamma_i^k\right| - \beta_{c,j}^k\left|\gamma_i^k\right|\right.\\
&\qquad\qquad\left. + \mu Q_{s_i^k}\left(P_j^k, \left|\gamma_i^k\right|\right)\right].
\end{aligned} \tag{15}$$

In a large-scale data publishing and sharing network, the data transaction task is a many-to-many matching decision. We further introduce the following constraints.

C_1: $\sum_{s_i^k \in S} \zeta_{ij}^k \leq n$.

C_2: $\sum_{\theta_j^k \in P} \zeta_{ij}^k \leq m$.

C_3: $\zeta_{ij}^k = \{0, 1\}$.

C_4: $\sum_{\theta_j^k \in \Theta} \zeta_{ij}^k\ \eta_{ij} \leq H(\theta_j)$

where constraint C_1 indicates that when exchanging data γ_i^k, there are at most q data subscription query data; constraint C_2 indicates that for a piece of data γ_i^k, there are m providers that can match with a provider (i.e., the requester can get multiple data from one publisher); condition C_3 is a matching matrix, the value of the vector is 0 or 1, 0 means no match, and 1 means match; constraint C_4 means that for requester θ_j, the maximum number of requests is restricted by its total token wallet $H(\theta_j)$ and η_{ij} is data price from publisher i. Compared with the traditional one-to-many matching network [26], we add constraints C_1 and C_4 to boost the performance of decision computation.

Generally, there are two kinds of selfish behaviors in a transaction network, leading to network unstable, called blocked, i.e., as follows.

1) The utility function of a publisher or a requester is less than zero, that is

$$\mu^{\text{Pub}}_{p^k_j}\left(s^k_i, \gamma_k, \beta^k_{ij}\right) < 0 \text{ or } \mu^{\text{Pub}}_{p^k_i}\left(s^k_j, \gamma_k, \beta^k_{ij}\right) < 0.$$

2) There are other matchings, such that the current publisher and requester can obtain larger benefits. Thus, they will leave the existing matching, which results in network unstable. More formally, there are s^k_i and p^k_j, such that

$$\exists\beta^k_{ij}\mu^{\text{REQ}}_{s^k_i}\left(p^k_j, \gamma_{k'}\beta^k_{ij}\right) > \mu^{\text{REQ}}_{s^k_i}\left(\Phi^k\left(s^k_i\right), \gamma_k\right)$$
$$\text{and } \mu^{\text{Pub}}_{p^k_j}\left(s^k_i, \gamma_k, \beta^k_{ij}\right) > \mu^{\text{Pub}}_{p^k_j}\left(\Phi^k\left(p^k_i\right), \gamma_k\right)$$
$$\text{where } \Phi^k\left(s^k_i\right) = \left(s^k_i, \beta^k_{ij}\right), \Phi^k\left(p^k_j\right) = \left(s^k_i, \beta^k_{ij}\right) \text{ and}$$
$$\beta^k_{ij} \neq \beta_{ij}.$$

We design the matching algorithm, which can solve the two blocked cases. For case 1, we utilize (7) to filter the block. In step 1, we set $\beta^{k,t}_{i,j} = \beta^{k,t}_{\min j}$, where $\beta^{k,t}_{\min j}$ is the smallest cost of the entire network, i.e., $\beta^{k,t}_{i,j} = vP^c_i|\gamma_k|$, satisfying $\mu^{\text{Pub}}_{p^k_j}(s^k_i, \gamma_k, \beta^k_{ij}) \geq 0$. Similarly, in step 2, we calculate the demand function of each requester (10). We add requester to candidate set SLISTs only when demand is a positive value, so we can guarantee $\mu^{\text{REQ}}_{s^k_i} \geq 0$. For the paired block case 2, the intelligent transaction matching (ITM) algorithm classifies the decisions of publishers into three types: 1) if there is no matching, a publisher selects the best utilization of a requester to establish the current matching; 2) if there are multiple requesters, a requester automatically increases price Δ, and continues to find requester q^k_j with complex conditions before terminating matching; and 3) if there is exactly one matching, the matching is added into a transaction. Meanwhile, in each biding iteration, a requester selects a publisher with the best reputation to match, so that we can establish an optimal mutual benefit matching between a publisher and a requester.

VIII. Performance Evaluations

A. Testbed

The prototype is deployed under linux platform. Eight servers run as blockchain middleware and consensus nodes, and two servers run as edge servers connected by data requesters or data publishers.

We select Chainsql [28] as a blockchain platform. Chainsql is developed from ripple networks, which use XRP ledger structure, encoded by C++ language. Meanwhile, Chainsql uses RECP as the consensus protocol, which is an optimization of the practical Byzantine fault-tolerant (PBFT) protocol. We also examine the efficiency of our EMR blockchain in our testbed. We create "privacy policies," "accesslist," and "parameters" tables in Chainsql to keep module parameters of our approach. As with the security storage system, we use the distributed file system to keep the archiving files in our evaluation. Also, other security storage service, such as IPFS, can be attached to our prototype for massive versioning file storage. The testbed architecture illustration in our experiments is shown in Fig. 6.

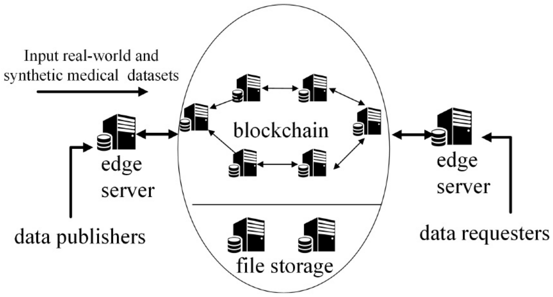


Fig. 6. Testbed architecture in the experiments.

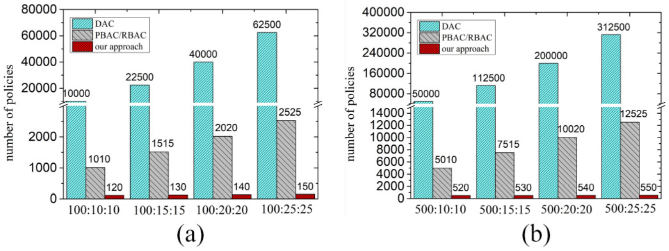


Fig. 7. Comparison with typical models of DAC, PBAC for massive simulated smart devices with different roles and purposes. (a) 100 simulated smart devices. (b) 500 simulated smart devices.

B. Data Set

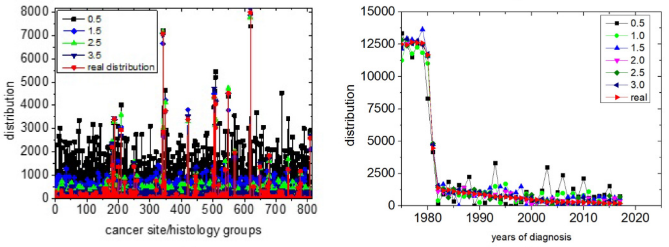
We use the real-world medical data set SEER [29], which is an open-source medical data set for academic research. SEER includes nearly 500 000 patients with 187 attributes totally. We select 100 000 patient information with 100 attributes for testing, including patient ID, address, age, survival time, race, gender, and other diagnosis or treatment attributes.

We classify them into three types, such as $M(I, D, C)$, and use different privacy protection strategies. We also test the efficiency of EMR publishing and sharing using the proposed intelligent matching decision. To examine the efficiency of our access control model, we use synthetic data sets to simulate 100–500 IoMT smart devices with different purposes and roles.

C. TS-PBAC Model Evaluation

Current models of access control in IoMT, such as PBAC [6] and RBAC [7], only provide access control decisions, i.e., permit or forbidden, to subject directly. Our TS-PBAC model designs three submodels, and each submodel is responsible for different elements of PBAC, such as patients' privacy preference, LDP privacy-preserving strategies, privacy policies maintaining, etc. The three submodels are deployed at the data publisher site, blockchain site, and data requester site, respectively.

We evaluate the length of access control list, which is the core elements measuring efficiency in access control models. In Fig. 7(a), we present the number of access list for models of DAC, RBAC/PBAC, and our TS-PBAC using different policies, roles, and purposes. We first simulate 100 smart devices connected to an edge server. Our model has the smallest length of access list, and achieves nearly 1/100 and 1/10 compression compared with traditional DAC and RBAC models, respectively.



(a) (b)

Fig. 8. LDP-based sensitive attributes privacy policies. (a) Cancer site/histology estimation. (b) Years of diagnosis estimation.

Also, when the number of smart device increases from 100 to 500, the length of privacy policies of TS-PBAC also increases from 120 to 520 accordingly as shown in Fig. 7(b). In this case, we can keep the device list in an edge server, and we only keep model parameters and role-based authority access list at the blockchain. Therefore, the TS-PBAC model facilitates deploying over blockchain and keeps relatively small size of model parameters into the blockchain ledger structure.

D. Privacy Policy Evaluation

We introduce LDP-based techniques to protect the high-level and medium-level sensitive attributes at the patient site before data publishing. We first evaluate the LDP-based privacy policy, where it is depicted as privacy budget ϵ. We perform statistical analysis over disturbed EMRs at the requester site. We use two kinds of attributes in D, i.e., cancer site, which is encoded for histology groups; year, which is the time a patient is diagnosed. In a randomized response LDP algorithm, ϵ is relative to the domain of input variable x. We use 100 000 EMRs and test cancer site/histology groups, years of diagnosis to expose the relationship.

Fig. 8(a) shows the statistics on histology groups of cancer diseases. We set ϵ as 0.5, 1.5, 2.5, and 3.5, and compute the statistical distribution at the requester site, respectively. Since the attribute of primary site of cancer site/histology groups is encoded into a value belonging to [0, 800], i.e., $|D| = 800$, when we set $\epsilon \in (2.0, 2, 5)$, we can obtain a relatively accurate statistical estimation. Fig. 8(b) demonstrates the estimated distribution on years of diagnosis, which belongs to [1980, 2020]. We encode the difference of maximum years and minimum years, i.e., $|D| = 40$. We compute the estimated aggregates at the requester site, when ϵ is with different values, i.e., 0.5, 1.0, 1.5, 2.0, 2.5, 3.0. In Fig. 8(b), if we set $\epsilon \in (1.0, 1.5)$, we can obtain a relatively accurate estimation.

We also examine the relationship between estimation variance and privacy budgets. In Fig. 9(a) and (b), we present the relationship between estimation variance and privacy budget in cancer site estimation and years of diagnosis estimation, respectively. Previous literature [10], [11] have exposed that privacy budget ϵ reflects privacy protection ability. In Fig. 9(a) and (b), when ϵ is small, the ability of privacy protection is strong; while the quality of data is weak, i.e., the variance of estimation at requester site is large. Otherwise, when ϵ is large, the ability of privacy protection becomes weak, while the quality of data is good, i.e., the variance of estimation at requester site becomes small.

Also, we present the evaluation on data quality computation using different privacy budget ϵ in Fig. 9(c). When other parameters are defined, the value of ϵ and the quality of data increases correspondingly. We notice that when the domain of input value $|D|$ is small, such as $|D| = 40$, a publisher can set a small ϵ, and the requester can obtain a relatively accurate estimation. When the domain of input value $|D|$ increases, we can increase the value of ϵ, such as $|D| = 800$ and $\epsilon = 2.0$, to achieve the same accuracy of estimation at the requester site. Thus, we can adjust the privacy budget parameter *epsilon* at the publisher site to achieve different estimation accuracies at the requester site, and the process of adjustment can be evaluated by the matching decision.

E. Trade-Like Purpose Decision Evaluation

The total benefit of a matching network is also known as welfare in the area of economics, which is used to evaluate the stability of a data transaction network. We evaluate our proposed transaction-like purpose decision on three aspects of welfare and the number of bargaining times in a matching decision.

Fig. 10(a) illustrates the welfare (total benefit of a transaction network) changes with the number of data publishers. When the maximum sell times (n) and bid times (m) are fixed, the welfare increases with the number of requesters. For the reason that if there are more requesters, there might be more successful matching pairs between requesters and publishers to increase the welfare. Also, when the matching decision reaches the limit of maximum sell times (n), the total benefit of a transaction can be stable. When we set a larger amount of money of a requester in each bid round, the welfare also increases accordingly. Therefore, the sell times (n) from publishers, and the money used in each round of bidding from requesters are two core elements in a matching decision model.

Fig. 10(b) illustrates the welfare changing with different ratios of the number of publishers to the number of requesters. We explore the impact ratio of the number of requesters to the number of publishers on network welfare. When we fix the maximum number of publishers and bidding time, at the beginning, the network revenue increases with the increase of the ratio, that is, when the number of publishers is smaller than the number of requesters, the network welfare increases with the value ratio. However, when the value of ratio is greater than or equal to 1, that is, the number of publishers is greater than or equal to the number of requesters, the total welfare of a transaction network tends to be stable.

Fig. 10(c) illustrates the number of bargaining times changing with step price in each iteration. We can find that the number of bargaining times gradually decreases when the step price size increases. In the beginning, the number of bargaining times decreases rapidly, and it gradually becomes stable, since there is no need to bargain price for the data to require a high price. At this time, the requesters bid their desired data from publishers directly.

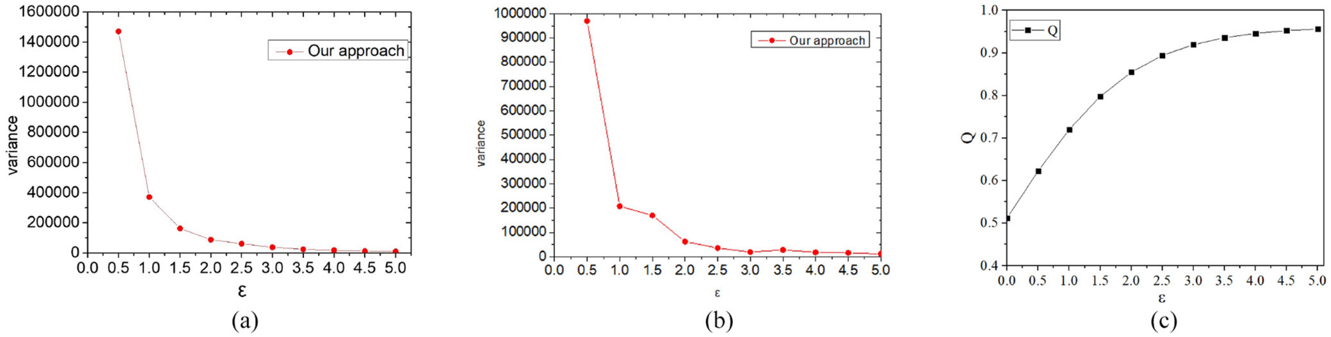


Fig. 9. Privacy budget versus corresponding variance distribution. (a) Privacy budgets in cancer site estimation. (b) Privacy budgets in years of diagnosis. (c) Privacy budgets versus theoretical data quality.

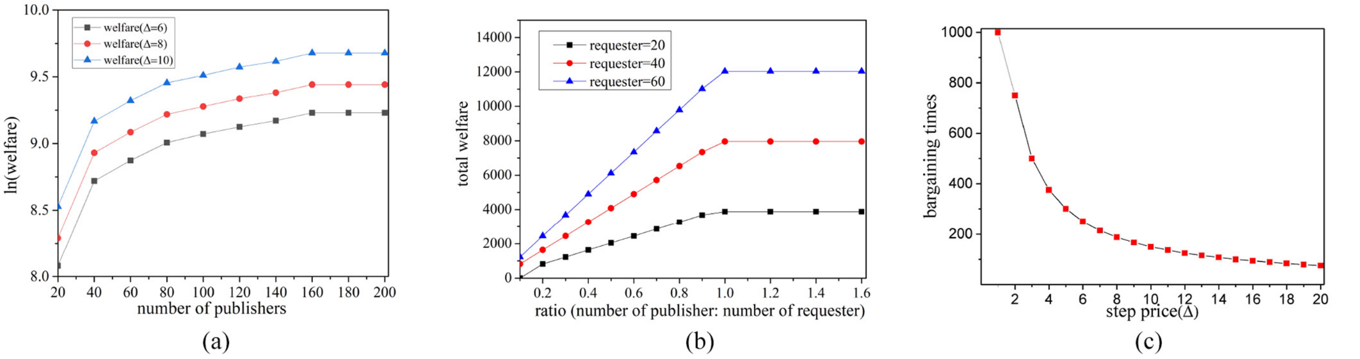


Fig. 10. Trade-like purposes access control decision evaluations. (a) Ratio of publishers and requesters. (b) Walfare for trade-like purposes. (c) Bargaining times.

TABLE V
TPS EVALUATIONS AT REQUESTER SITE

Operations	Q_1	Q_2	Q_3	Q_4	Q_5
TPS	143	167	143	125	111

F. System Evaluation

We test our prototype system using 100 000 EMR records, each of which includes 100 key-value pairs. After running the decision model, a publisher can publish its EMRs into the blockchain. We conduct query transaction over the developed blockchain to fetch the published data. We examine TPS for different operations, such as searching in a ledger (Q_1), querying a historical transaction (Q_2), creating an account (Q_3), count the number of transaction (Q_4), and running a contract (Q_5). The TPS of the five operations is shown in Table V.

As a result, we can summarize that our prototype can publish 35 000 key-value pairs per second onto the proposed EMR blockchain. Meanwhile, we can conduct 125–167 TPS at the requester site for common data sharing and accessing transaction operations.

IX. CONCLUSION

In this article, we presented an individual-centric data publishing and sharing framework in the IoMT, aiming to provide security and privacy-preserving protection for differential purpose applications. Our proposed TS-PBAC model is compatible with blockchain. We presented the LDP-based privacy preferences, purposes, and roles-based access control decisions, and also designed an intelligent matching decision algorithm to maximize benefits for trade-like purposes of participants. Moreover, our approach showcases the leverage of access control model, blockchain, and edge computing paradigam into IoMT scenarios, to address the reliable data publishing and sharing network with traceability and tamperproof characteristics.

In the future, we intend to enhance the quality of healthcare services along with peer-evaluation metrics, such that doctors can conduct mutual evaluation without leaking privacy. Moreover, we intend to develop a full level prototype to collect and publish complex medical data, such as medical questionnaires and medical images in real-time scenarios.

REFERENCES

[1] G. E. Santagati, N. Dave, and T. Melodia, “Design and performance evaluation of an implantable ultrasonic networking platform for the Internet of medical things,” *IEEE/ACM Trans. Netw.*, vol. 28, no. 1, pp. 29–42, Feb. 2020.

[2] B. D. Deebak and F. Al-Turjman, “Smart mutual authentication protocol for cloud based medical healthcare systems using Internet of medical things,” *IEEE J. Sel. Areas Commun.*, vol. 39, no. 2, pp. 346–360, Feb. 2021.

[3] J. Srinivas, A. K. Das, M. Wazid, and N. Kumar, “Anonymous lightweight chaotic map-based authenticated key agreement protocol for industrial Internet of Things,” *IEEE Trans. Dependable Secure Comput.*, vol. 17, no. 6, pp. 1133–1146, Nov./Dec. 2020.

[4] A. Ali *et al.*, “Security, privacy, and reliability in digital healthcare systems using blockchain,” *Electronics*, vol. 10, no. 16, p. 2031, 2021.

[5] V. S. Naresh, S. Reddi, and V. Allavarpu, “Blockchain-based patient centric health care communication system,” *Int. J. Commun. Syst.*, vol. 34, no. 2, p. e4749, 2021.

[6] M. Amini and F. Osanloo, "Purpose-based privacy preserving access control for secure service provision and composition," *IEEE Trans. Services Comput.*, vol. 12, no. 4, pp. 604–620, Jul./Aug. 2019.

[7] B. S. Egala, A. K. Pradhan, V. Badarla, and S. P. Mohanty, "Fortified-chain: A blockchain-based framework for security and privacy-assured Internet of medical things with effective access control," *IEEE Internet Things J.*, vol. 8, no. 14, pp. 11717–11731, Jul. 2021.

[8] N. Li, M. Lyu, D. Su, and W. Yang, *Differential Privacy: From Theory to Practice* (Synthesis Lectures on Information Security, Privacy, & Trust). San Rafael, CA, USA: Morgan Claypool Publ., 2016.

[9] S. Wang, J. Li, G. Wu, H. Chen, and S. Sun, "Joint optimization of task offloading and resource allocation based on differential privacy in vehicular edge computing," *IEEE Trans. Comput. Soc. Syst.*, early access, May 13, 2021, doi: 10.1109/TCSS.2021.3074949.

[10] T. Wang, J. Blocki, N. Li, and S. Jha, "Locally differentially private protocols for frequency estimation," in *Proc. 26th USENIX Security Symp.*, 2017, pp. 729–745.

[11] Ú. Erlingsson, V. Pihur, and A. Korolova, "RAPPOR: Randomized aggregatable privacy-preserving ordinal response," in *Proc. ACM SIGSAC Conf. Comput. Commun. Security*, 2014, pp. 1054–1067.

[12] S. Bayat, Y. Li, L. Song, and Z. Han, "Matching theory: Applications in wireless communications," *IEEE Signal Process. Mag.*, vol. 33, no. 6, pp. 103–122, Nov. 2016.

[13] Z. Ning *et al.*, "Mobile edge computing enabled 5G health monitoring for Internet of medical things: A decentralized game theoretic approach," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 2, pp. 463–478, Feb. 2021.

[14] Z. Ning *et al.*, "Intelligent resource allocation in mobile blockchain for privacy and security transactions: a deep reinforcement learning based approach," *Sci. China Inf. Sci.*, vol. 64, no. 6, pp. 1–16, 2021.

[15] Y. Zhuang, L. R. Sheets, Y.-W. Chen, Z.-Y. Shae, J. J. P. Tsai, and C.-R. Shyu, "A patient-centric health information exchange framework using blockchain technology," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 8, pp. 2169–2176, Aug. 2020.

[16] X. Liu, P. Zhou, T. Qiu, and D. O. Wu, "Blockchain-enabled contextual online learning under local differential privacy for coronary heart disease diagnosis in mobile edge computing," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 8, pp. 2177–2188, Aug. 2020.

[17] S. Chaganti *et al.*, "Electronic medical record context signatures improve diagnostic classification using medical image computing," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 5, pp. 2052–2062, Sep. 2019.

[18] B. Cao *et al.*, "When Internet of Things meets blockchain: Challenges in distributed consensus," *IEEE Netw.*, vol. 33, no. 6, pp. 133–139, Nov./Dec. 2019.

[19] Y. Li *et al.*, "Direct acyclic graph-based ledger for Internet of Things: Performance and security analysis," *IEEE/ACM Trans. Netw.*, vol. 28, no. 4, pp. 1643–1656, Aug. 2020.

[20] Z. Ning *et al.*, "Blockchain-enabled intelligent transportation systems: A distributed crowdsensing framework," *IEEE Trans. Mobile Comput.*, early access, May 13, 2021, doi: 10.1109/TMC.2021.3079984.

[21] B. Cao, M. Li, L. Zhang, Y. Li, and M. Peng, "How does CSMA/CA affect the performance and security in wireless blockchain networks," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 4270–4280, Jun. 2020.

[22] J. Kang *et al.*, "Blockchain for secure and efficient data sharing in vehicular edge computing and networks," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4660–4670, Jun. 2019.

[23] O. A. Lajam and T. A. Helmy, "Performance evaluation of ipfs in private networks," in *4th International Conference on Data Storage and Data Engineering, DSDE '21*. New York, NY, USA: Assoc. Comput. Machinery, 2021, pp. 77–84, doi: 10.1145/3456146.3456159.

[24] R. Cao, Z. Tang, C. Liu, and B. Veeravalli, "A scalable multicloud storage architecture for cloud-supported medical Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 3, pp. 1641–1654, Mar. 2020.

[25] "Google/Snappy." [Online]. Available: https://github.com/google/snappy/

[26] S. Bayat, R. H. Y. Louie, Z. Han, B. Vucetic, and Y. Li, "Distributed user association and femtocell allocation in heterogeneous wireless networks," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 3027–3043, Aug. 2014.

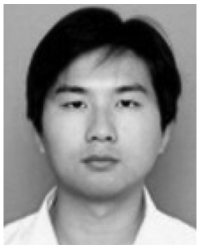
[27] C. Chen, C. Wang, T. Qiu, N. Lv, and Q. Pei, "A secure content sharing scheme based on blockchain in vehicular named data networks," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3278–3289, May 2020.

[28] M. Muzammal, Q. Qu, and B. Nasrulin, "Renovating blockchain with distributed databases: An open source system," *Future Gener. Comput. Syst.*, vol. 90, pp. 105–117, Jan. 2019.

[29] "SEER Incidence Data, 1975–2018." [Online]. Available: https://seer.cancer.gov/data/

Guangjun Wu (Member, IEEE) received the M.S. and Ph.D. degrees from Harbin Institute of Technology, Harbin, China, in 2006 and 2010, respectively.

He is currently a Senior Engineer with the Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China. His research interests include big data analysis, distributed storage, data security, and privacy computing.



Shupeng Wang (Member, IEEE) received the M.S. and Ph.D. degrees from Harbin Institute of Technology, Harbin, China, in 2004 and 2007, respectively.

He is currently a Senior Engineer with the Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China. His research interests include big data management and analytics, network security, and edge computing.

Zhaolong Ning (Senior Member, IEEE) received the Ph.D. degree from Northeastern University, Shenyang, China, in 2014.

He was a Research Fellow with Kyushu University, Fukuoka, Japan, from 2013 to 2014. He is an Associate Professor with the School of Software, Dalian University of Technology, Dalian, China. He has authored or coauthored more than 120 scientific papers in international journals and conferences. His research interests include Internet of Things, mobile-edge computing, deep learning, and resource management.

Jun Li received the B.S. degree in software engineering from Beijing Technology and Business University, Beijing, China, in 2018. He is currently pursuing the Ph.D. degree in cyberspace security with the Institute of Information Engineering, Chinese Academy of Sciences, Beijing.

His main research interests are edge computing and Internet of Vehicles.