Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2023.047660

ARTICLE

A Fair and Trusted Trading Scheme for Medical Data Based on Smart Contracts

Xiaohui Yang and Kun Zhang*

School of Cyber Security and Computer, Hebei University, Baoding, 071000, China

*Corresponding Author: Kun Zhang. Email: zhangkunzk@stumail.hbu.edu.cn

Received: 13 November 2023 Accepted: 15 December 2023 Published: 27 February 2024

ABSTRACT

Data is regarded as a valuable asset, and sharing data is a prerequisite for fully exploiting the value of data. However, the current medical data sharing scheme lacks a fair incentive mechanism, and the authenticity of data cannot be guaranteed, resulting in low enthusiasm of participants. A fair and trusted medical data trading scheme based on smart contracts is proposed, which aims to encourage participants to be honest and improve their enthusiasm for participation. The scheme uses zero-knowledge range proof for trusted verification, verifies the authenticity of the patient's data and the specific attributes of the data before the transaction, and realizes privacy protection. At the same time, the game pricing strategy selects the best revenue strategy for all parties involved and realizes the fairness and incentive of the transaction price. The smart contract is used to complete the verification and game bargaining process, and the blockchain is used as a distributed ledger to record the medical data transaction process to prevent data tampering and transaction denial. Finally, by deploying smart contracts on the Ethereum test network and conducting experiments and theoretical calculations, it is proved that the transaction scheme achieves trusted verification and fair bargaining while ensuring privacy protection in a decentralized environment. The experimental results show that the model improves the credibility and fairness of medical data transactions, maximizes social benefits, encourages more patients and medical institutions to participate in the circulation of medical data, and more fully taps the potential value of medical data.

KEYWORDS

Blockchain; data transactions; zero-knowledge proof; game pricing

1 Introduction

In today's information age, data has been regarded as an important asset [1]. However, in the medical field, the safe and efficient use of digital assets and transaction management have always faced challenges. According to the statistical results of 2,410 hospitals [2], less than 10% of patients with access rights try to access their electronic medical records. In stark contrast, hackers have relatively easy access to 399.5 million medical records [3]. The transaction flow of medical big data has become a key trend in the field of health care. Under the premise of ensuring patient privacy and data security, most patients are willing to share personal medical data [4].

Healthcare Medical Data Compliance Circulation Standards provide a regulatory framework for the circulation of medical data [5]. It is a groundbreaking milestone in the field of data circulation within the healthcare industry. According to the market research report [6], the global market value of electronic health records in 2021 is about 29,417.2 billion and is expected to grow to $ 42,203.5 billion by 2029. These data not only represent the individual's health status but also provide the basis for medical research, clinical decision-making, and public health policymaking. Especially for certain significant diseases, their connection to society is profound. For instance, utilizing blockchain technology to trace COVID-19 infectious disease information [7]. Tracing infectious disease information is necessary, but protecting patient privacy is also crucial. Therefore, when considering the research on sharing medical data, it is essential to comprehensively balance individual privacy and societal interests.

Medical data includes sensitive data such as disease history and personal health information. Once illegally obtained, it not only violates personal privacy but also may lead to serious consequences such as identity theft and credit crisis. At the same time, due to the lack of sufficient trust between data owners and consumers, the authenticity of medical data will be questioned, and effective transactions will be hindered. Therefore, there is a trade-off between privacy protection and data credibility, and it is difficult to verify the validity of the data before the transaction is completed. In addition, due to information asymmetry, data buyers may also have price fraud. Delgado-Segura et al. [8] proposed a fair data transaction protocol based on the Bitcoin system. The delivery of data and the execution of payments are atomic, because the seller cannot redeem the payment unless the buyer obtains the data, and the buyer cannot obtain the data without paying. Although it avoids the refusal of payment in data transactions, it is impossible to avoid price fraud due to information asymmetry in the pricing of data transactions. The lack of a fair pricing strategy may result in a lack of reasonable returns for data providers and a lack of sharing motivation, which limits the full sharing of data [9]. To promote the effective and secure sharing of medical data, new technologies, and strategies must be used to solve these problems.

Given the above problems, this paper proposes a credible and fair medical data transaction scheme based on the intelligent contract technology of decentralized automatic execution. This scheme uses zero-knowledge proof contracts before transactions to ensure the authenticity and credibility of data under the premise of privacy protection and realizes a fair incentive game pricing strategy to achieve a balance between data credibility and privacy protection and injects new vitality into data management and transactions in the health care field. Through experimental verification, the advantages and practicability of the scheme in data privacy protection, trusted verification, and fair incentives are proved. The main contributions of this paper are summarized as follows.

Zero-knowledge proof does not need to expose specific data content when verifying the authenticity of data, to achieve effective verification of data under the premise of protecting privacy. A solid privacy protection barrier is established between data owners and users before the transaction.

The medical data transaction pricing strategy based on the improved Stackelberg game enables the transaction participants to achieve the maximum benefits, thereby maximizing social welfare. It provides a new idea to encourage data sharing, encourage data providers and data consumers to actively participate in data transactions, and fully tap the potential value of data.

The structure of this paper is as follows. The second part introduces the related research work. The third part provides the basic knowledge of the theoretical model. The fourth part first introduces the framework of the scheme and introduces the key processes such as trusted verification and fair

pricing strategy in detail. The fifth part includes experimental verification and analysis. Finally, the sixth part summarizes and discusses.

2 Related Work

Recent years have seen a rapid development in data transactions. Traditional centralized third-party trading platforms have been proposed, but their opaque transaction processes may lead to dishonesty and harm participants' interests. With the advancement of blockchain decentralization technology, data transactions are shifting towards decentralization.

Juang et al. [10] proposed a secure data transaction scheme for cloud storage, where encrypted data verification is challenging. Additionally, the involvement of a bank as a third party introduces susceptibility to collusion attacks, resulting in an opaque transaction process. Chen et al. [11] proposed an offline digital content trading system, entailing direct customer-store transactions with bank-managed funds. Hwang et al. [12] proposed a provable fair file exchange protocol, enabling the file owner to iteratively exchange notarized documents with various parties while preserving source and participant anonymity. Jung et al. [13] proposed a data transaction liability agreement where brokers monitor and penalize dishonest consumers during transactions by identifying misconduct. However, the inclusion of third-party brokers escalates transaction costs and introduces potential trust issues. Su et al. [14] believed that in the traditional data market, data buyers and data sellers must use a centralized trading platform, which may be dishonest, affecting the fairness of data transactions and damaging the interests of both parties. Li et al. [15] proposed a blockchain-based distributed IoT data trading system, where users benefit from selling their data, and service providers purchase data for access. Bajoudah et al. [16] proposed a market model to solve the problem that participants rely on untrusted third parties for transactions while reducing transaction costs. Guan et al. [17] proposed a secure, fair, and efficient data transaction scheme that does not rely on any third party, which solves the problems of complex transaction processes, high transaction costs, and possible unfair exchange in data transactions. Additionally, Nguyen et al. [18] introduced a solution for IoT data transactions using distributed ledger technology, addressing inefficiencies and security concerns arising from the reliance on centralized third-party entities in IoT systems. Nawaz et al. [19] proposed a data transaction scheme for edge IoT, in which edge devices directly conduct data transactions with third parties without intermediaries.

Data trading is different from general commodity trading [20], and data quality and value cannot be specifically perceived and defined. Buyers want to buy the real data they need, while sellers want to share the data and get reasonable rewards.

Kiyomot et al. [21] proposed a fair-trade protocol for anonymous data sets. Without exposing data privacy, buyers need to verify the data multiple times to ensure that they have purchased the desired data. Zhao et al. [22] proposed a fair data transaction protocol based on blockchain. It integrates ring signatures, preventing double authentication signatures and similarity learning. It ensures the privacy of data providers and the availability of transaction data. Sheng et al. [23] proposed a blockchain-based copyright protection crowdsourcing data transaction framework and designed an auction algorithm based on semantic similarity to ensure the authenticity and individual rationality of the auction. Galteland et al. [24] constructed a fair-trading protocol based on smart contracts, which provides participants with better privacy and solves the problem of privacy leakage of transaction-sensitive data. Gao et al. [25] proposed a secure, fair, and instant data transaction scheme, which solves the problems of unfair transaction, transaction delay, and illegal transaction data faced by bitcoin-based data transaction schemes. Xue et al. [26] proposed a blockchain-based data trading framework that

aims to ensure security, fairness, and privacy. While it supports fine-grained sales using medical data for verification, there is still a privacy concern with partial attribute verification potentially revealing sensitive information. Chen et al. [27] introduced a comprehensive data transaction framework for vehicle networking based on blockchain, ensuring both the security and authenticity of data transactions. To further improve transaction efficiency, the authors proposed an iterative double auction mechanism, designed to incentivize increased participation in data trading. Zheng et al. [28] proposed a decentralized blockchain-based data trading platform with smart contracts for efficient distributed transactions and fair data rewards distribution. Li et al. [29] proposed a secure decentralized data transaction model based on blockchain to solve the trust problem between buyer, seller, and agent nodes and increase the incentive of user transaction data. Luong et al. [30] designed an unmanned aerial vehicle data auction scheme based on deep learning, which maximizes the income of energy service providers while ensuring incentives.

In summary, while decentralized blockchain architectures have addressed fairness and trust issues in data trading, research on medical data sharing lacks sufficient focus on data credibility and price fairness. This gap impacts data consumer confidence, making it challenging to balance privacy protection, data authenticity verification, and meeting specific needs. The imbalance in transaction price incentives reduces participant willingness, affecting enthusiasm and harming the interests of both parties. Further development in blockchain-based medical applications is crucial to ensure secure transactions, implement fair circulation schemes, and maximize the overall value of medical data.

3 Preliminary

3.1 Stackelberg Game

The Stackelberg game is a two-stage game with the concepts of leader and follower [31]. Leaders and followers try to maximize their profits. Therefore, the game provides dual benefits for both players. The Stackelberg game has been widely used to solve resource management problems in network systems in a distributed manner. Cardellini et al. [32] formulated the single cloud multi-service resource supply and pricing problem as a Stackelberg game to minimize service costs while maximizing provider revenue. Guo et al. [33] proposed a hierarchical architecture of smart home based on mobile edge computing and used the Stackelberg game to solve the problem of resource purchase and pricing of access points and user devices. Li et al. [34] proposed a game theory model for coordinating the exchange of EMR data among healthcare institutions to maximize social welfare. Therefore, this paper proposes a pricing strategy based on the Stackelberg game to solve the problem of fair pricing of medical data transactions, so that both data providers and data consumers have the best choice strategy.

3.2 Blockchain

Nakamoto first proposed the development of a cryptocurrency called Bitcoin in 2008 [35]. Transactions can be conducted through this system, and transaction records are stored as immutable records on the blockchain ledger. Therefore, it has been applied in areas such as copyright protection [36]. Smart contracts [37] are stored and executed on the blockchain without the need for a trusted third party, ensuring honesty. In this transaction model, we use smart contracts for verification and negotiation to ensure trust and the unmodifiable on-chain records can serve as evidence in transaction disputes.

3.3 ZoKrates

Based on the zk-SNARK algorithm Groth16 [38], ZoKrates [39] enables off-chain computing and on-chain verification in Ethereum. It provides a toolkit with domain-specific languages (DSLs), compilers, and witness and proof generators. ZoKrates seamlessly integrates with Ethereum's contract deployment and invocation process. For trusted verification using zk-SNARK, ZoKrates adopts a black-box approach. Specific calculations are written in a high-level language, compiled into flat code, and converted to a rank-1 constraint system (R1CS). ZoKrates follows the standard setup phase to generate common reference strings (CRS) and public keys—proof key and verification key. The verification key is deployed as a contract on the blockchain, while the proof key is distributed to the prover. The prover inputs information to calculate the zk-SNARK witness, generating proof based on the witness and CRS. The proof, combined with public input, undergoes verification using the stored verification key in the contract. Participants with the proof file can invoke the contract for verification.

4 Proposed Scheme

4.1 Scheme Framework

This paper proposes a smart contract-based application for trusted verification and fair trading of medical data. Here, patients are treated in various medical institutions. For patients with a willingness to share high-quality medical data, hospitals can generate a zero-knowledge proof contract for their authenticity and specific attributes. Medical institutions, as data generators, provide guarantees for their authenticity and data quality. In a decentralized blockchain, any participant who obtains proof can verify medical data and trade. To facilitate the description of the scheme, some symbols are defined in Table 1.

Table 1: Symbol definition

Symbol	Definition
Cid	The file identifier returned by IPFS.
Sign_Cid	The Cid of the patient and the doctor's private key are signed in turn.
Proof	Proof for verification.
prk_d, puk_d	Doctor's private key and public key.
prk_o, puk_o	The private key and public key of the data owner.
Attr	Specific attributes of data.
[*min, max*]	The scope of specific attribute declarations.

The architecture of trusted fair trading based on blockchain is shown in Fig. 1. The proposed architecture consists of six components: data owner, data collector, data consumer, doctor, blockchain network, and IPFS. Assuming that all participants have completed the registration and other steps, the program runs as follows four steps.

Step 1 As the owner of the data, Step 1 patients look for medical institutions for medical diagnosis and decide whether to share the generated medical data to obtain the corresponding rewards. The rewards are not actual monetary compensation but are solely used to offset medical expenses.

Step 2 The doctor generates an electronic medical record for the patient encrypts it uploads it to IPFS, which returns the corresponding *Cid*. Then the doctor generates *Proof* and guarantees information for patients who are willing to share personal medical data. By re-signing the *Cid* of the patient's private key *prk_o* signature to generate *Sign_Cid*, and generating zero-knowledge proof contracts and proof for it, the contract is deployed to the blockchain network to verify the data attribute range characteristics. *Proof* and *Sign_ Cid*, *puk_d* will be shown back to the patient.

Step 3 If the data owner has the willingness to share personal medical data and obtain certain benefits, *Proof* , *Sign_Cid*, and *puk_d* can be sent to the data collector. The data collector is a trusted medical institution, which can only verify the data based on the re-signature verification and the proof call verification contract. The verification data is owned by the patient and has specific attributes and integrity. After the verification is valid, it is saved as medical data to be traded. During the verification process, in addition to verifying whether the patient's statement information is correct or not and whether it is owned by the patient, no personal information about the patient will be obtained to ensure the patient's privacy protection before the transaction.

Step 4 Data consumers can be many medical research institutions, pharmaceutical research and development companies, or medical insurance companies, and there is a competitive relationship between data consumers. When they need a certain type of medical data with specific attributes, they will negotiate with data collectors, reach Nash equilibrium through game bargaining of smart contracts, and seek pricing strategies to maximize the interests of all parties.

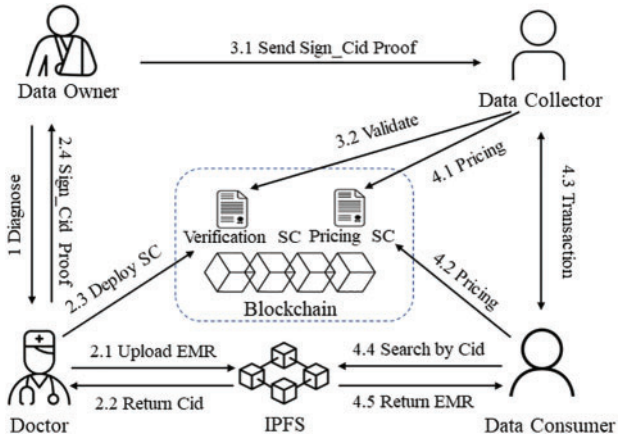

Figure 1: Scheme framework

4.2 Data Validation with Zero Knowledge Proof

In blockchain-based systems, despite transactions being securely recorded and tamper-proof, the intricate audit trail process remains a challenge for detecting illegal or malicious transactions. Malicious activities by anonymous participants often manifest as the exchange of false data. Particularly in the context of medical data exchange, data consumers face challenges in evaluating the authenticity, quality, and conformity of specific attributes (such as age, blood pressure, and blood glucose) with expectations. Additionally, due to the sensitivity of medical data, data owners are reluctant to disclose specific information before transaction finalization, resulting in diminished trust among transaction participants.

To address these challenges, employing zero-knowledge proof to verify specific attribute values in medical data aims to ensure data authenticity and alleviate privacy concerns. This trustworthy verification is conducted under the premise of safeguarding data privacy. The process of proof generation and verification using ZoKrates is illustrated in Fig. 2.

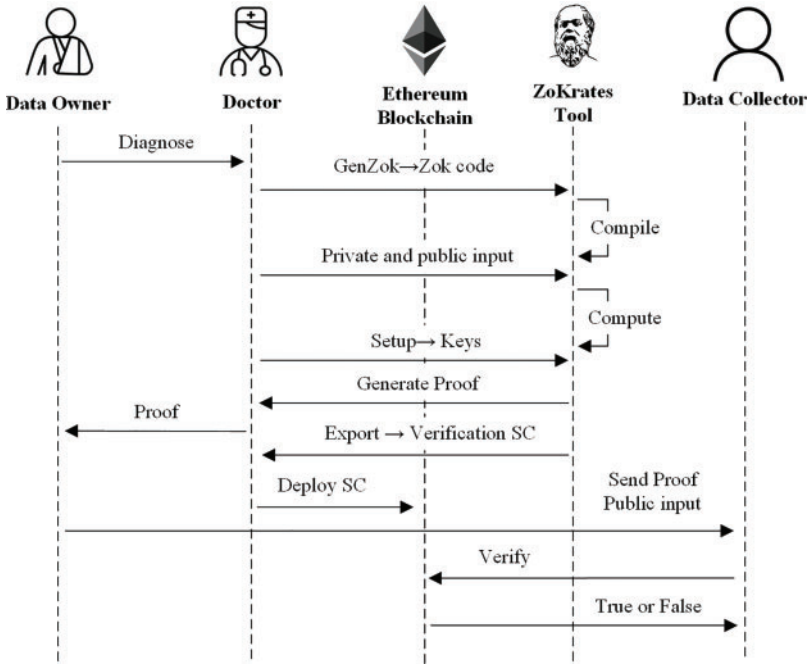

Figure 2: Zero-knowledge proof generation and verification

GenZok → Zok code: Doctors utilize the ZoKrates' DLS language to describe the zero-knowledge proof constraint verification program Zok. The arithmetic circuit enables users to prove the specific attribute value *Attr* in the patient's data within the declared range without disclosing the exact values of the attributes. Additionally, it includes the doctor's private key *prk_d* for signing the data identifier *Cid*, ensuring the authenticity and integrity of the medical data. The medical data is generated by the doctor, who bears responsibility for the data, guaranteeing its truthfulness and completeness.

Private and public input: According to the above Zok function description, the private input of the program Zok should include several attribute values Attr of the patient data, and the public input should include the maximum value max and the minimum value min file identifier Cid of the declared attribute value range to calculate and generate Witness.

Setup → keys: The Setup step of ZoKrates generates CRS, that is, proving key proving key, and verifying key. According to the G16 [38] algorithm, the output form of a non-interactive linear proof for a quadratic arithmetic program is given by $R = (F, \ell, \{u_i(X), v_i(X), w_i(X)\}_{i=0}^{m}, t(X))$. The relationship is defined as $(a_1, \ldots, a_\ell) \in F^\ell$ and the witness $(a_{\ell+1}, \ldots, a_m) \in F^{m-l}$ such that $a_0 = 1$, and an $(n-2)$-degree polynomial $h(X)$ (where n is the degree of $t(X)$) satisfies the equation: $\sum_{i=0}^{m} a_i u_i(X) \cdot \sum_{i=0}^{m} a_i v_i(X) = \sum_{i=0}^{m} a_i w_i(X) + h(X)\, t(X)$. Here, F is the field, l is a parameter, and $u_i(X), v_i(X), w_i(X), t(X)$ are polynomials.

$(\sigma, \tau) \leftarrow \text{Setup}(R)$: Choose $\alpha, \beta, \gamma, \delta, x \leftarrow \mathbb{F}^*$, where $0 < l < m$, and set $\boldsymbol{\tau} = (\alpha, \beta, \gamma, \delta, x)$ and $\boldsymbol{\sigma} = \left(\alpha, \beta, \gamma, \delta, \{x^i\}_{i=0}^{n-1}, \left\{ \frac{\beta u_i(x) + \alpha v_i(x) + w_i(x)}{\gamma} \right\}_{i=0}^{l}, \left\{ \frac{\beta u_i(x) + \alpha v_i(x) + w_i(x)}{\delta} \right\}_{i=l+1}^{m}, \left\{ \frac{x^i t(x)}{\delta} \right\}_{i=0}^{n-2} \right)$.

The above expression includes random variables selected by the verifier from the finite field $\mathbb{F}^*$: $\alpha, \beta, \gamma, \delta, x$. Additionally, $u_i(x)$, $v_i(x)$, $w_i(x)$ and $t(x)$ are polynomials related to the zero-knowledge proof constraint verification program ZoK, generated through operations such as arithmetic circuit transformation, R1CS transformation, and quadratic arithmetic program (QAP) transformation.

Generate proof: The proof is generated by computing the witness in the second step and creating the proof key in the third step, then saved by the data owner.

$\pi \leftarrow \text{GenProof}(R, \sigma, a_1, \ldots, a_m)$: Choose $r, s \leftarrow \mathbb{F}$, and compute $\pi = \Pi\sigma = (A, B, C)$, where $A = \alpha + \sum_{i=0}^{m} a_i u_i(x) + r\delta$, $B = \beta + \sum_{i=0}^{m} a_i v_i(x) + s\delta$, $C = \frac{\sum_{i=l+1}^{m} a_i(\beta u_i(x) + \alpha v_i(x) + w_i(x)) + h(x)t(x)}{\delta} + As + rB - rs\delta$.

Export and deploy SC: The doctor uses ZoKrates to create a zero-knowledge proof verification smart contract (Verification SC). This contract utilizes the Ethereum blockchain's pre-compiled elliptic curve operation library for bilinear pairing and assessment. Its main purpose is to verify whether the input parameter, *Proof*, meets the pairing requirements. Once the Verification SC is successfully generated, doctors deploy it on the Ethereum blockchain.

Send proof and public input: After the data owner requests medical services, the doctor returns the generated proof to the user. When users are willing to share personal medical data but prefer not to disclose specific details before the transaction, they can send both the *Proof* and a public input to the data collector for verification. In this scenario, the public input represents a declared content, which is a certain range rather than specifying specific values.

Verify: The collector, serving as a pre-transaction verifier, utilizes the *Proof* provided by the data owner to invoke a smart contract for verifying the correctness of the public input declaration.

true/false $\leftarrow \text{Verify}(R, \sigma, a_1, \ldots, a_l)$: This involves computing a quadratic multivariate polynomial t such that $t(\sigma, \pi) = 0$. This equation corresponds to verifying whether $A \cdot B = \alpha \cdot \beta + \frac{\sum_{i=0}^{l} a_i (\beta u_i(x) + \alpha v_i(x) + w_i(x))}{\gamma} \cdot \gamma \cdot \delta$ is equal. If equality holds, the result is true, indicating that the data is trustworthy. Otherwise, it is false, signifying that the declared range is not trustworthy.

4.3 Stackelberg Game Pricing Strategy

In this section, the pricing strategy based on the Stackelberg game is introduced to help transaction participants determine the best data pricing and demand. By realizing the Nash equilibrium of the proposed game, a fair pricing strategy for data providers and data consumers can be achieved.

The entities involved in the pricing strategy game include medical data collectors and data consumers. The economic strategic game of Stackelberg game is used to model and describe how medical data maximizes the interests of all parties. Data collectors collect and verify medical data. As representatives of many medical data owners, they bargain with data consumers. In addition, in the proposed scheme, medical data is stored in IPFS, and the same medical data has the same index, which can effectively prevent the data owner from double sharing operations. In this pricing strategy model, the electronic medical record data collector plays the role of "leader" in making decisions, while the data consumer is the "follower" of the subsequent decision-making.

4.3.1 Definition of Game Pricing Strategy

First, the leader sets the unit data price strategy $\{p = p_i(i \in N): p_{min} \leq p_i \leq p_{max}\}$, where p_i is the price of the i-th ($i \in \{1, 2, \ldots, N\}$) data consumer. Here, the medical data owner determines the proportion $q(0 < q < 1)$ of the transaction price to the data collector as a fee and encourages the receipt collector to actively participate. The minimum price is the minimum cost requirement for all costs. Therefore, the unit additional cost of operation and maintenance costs $c < p_i$, by deducting the cost c, the expected return sum incentive function of the data provider (collectively, the collector and the data owner) can be expressed as:

$$R_P = p_i d_i - c d_i \tag{1}$$

Among them, d_i is the data demand of the i-th data consumer, and the demand is determined according to the action of the "leader". There are usually N data consumers competing for medical data resources. For specific consumers, it can be determined that as the amount of data of the same type increases, the new amount of information obtained by data consumers from the data will decrease. Initially, increasing the number of units of data may bring greater information gain than later, resulting in a larger increase in utility. However, as the amount of data increases, the newly added data may only bring a small amount of additional information, and the increase in utility gradually decreases. Therefore, the utility function of the i-bit data consumer is described by the natural logarithmic function as follows:

$$R_C = \alpha_i In\,(d_i + k_i) - p_i d_i\,(k_i > 1, d_i > 0) \tag{2}$$

where d_{min} is the minimum demand of the i-th consumer, α_i is a predefined non-zero positive factor, and k_i >1, ensuring that the satisfaction utility function is non-negative even when the demand is very small. $p_i d_i$ is the data cost paid by the i-th consumer for medical data.

Based on the above utility function, medical data providers and consumers will perform sequential decision-making processes until the optimal data price and quantity are achieved. The data provider first sets the price, and then the data consumer sets the demand according to the pricing strategy. Next, data collectors try to maximize their utility by maximizing the price and amount of data, while data consumers want to reduce the price and increase the amount of data to maximize their utility. When the proposed Stackelberg game reaches Nash equilibrium, the maximum return of "leader" and "follower" can be achieved. Then, medical data providers and consumers will be motivated by this strategy to achieve their best returns. The Nash equilibrium between data providers and consumers can be described as follows.

If any price p_i and demand d_i satisfy the following conditions, then point (p^*,d^*) is the Nash equilibrium point, where p^* represents the optimal unit price of resource data, and d^* represents the data demand of consumers.

$$R_{sum}(p^*,\ d^*) \geq R_P(p_i,\ d_i) + R_C(p_i,\ d_i) \tag{3}$$

$$\begin{cases} R_P\,(p^*,\ d^*) \geq R_P(p_i,\ d^*) \\ R_C\,(p^*,\ d^*) \geq R_C\,(p^*,\ d_i) \end{cases} \tag{4}$$

4.3.2 The First Stage of the Stackelberg Game

Calculate the optimal data demand of the i-th data consumer. The data consumer chooses the optimal demand d^* that can maximize the revenue according to the pricing. For the data demand d_i,

the first-order and second-order partial derivatives of the consumer utility function R_C are as follows:

$$\begin{aligned} \frac{\partial R_C}{\partial d_i} &= \frac{\alpha_i}{d_i + k_i} - p_i \\ \frac{\partial^2 R_C}{\partial d_i^2} &= -\frac{\alpha_i}{(d_i + k_i)^2} < 0 \end{aligned} \tag{5}$$

Therefore, it can be concluded that R_C is a convex function. At the same time, the demand p^* for price p_i can be expressed by solving $\frac{\partial R_C}{\partial d_i} = 0$.

$$d^* = \frac{\alpha_i}{p_i} - k_i \tag{6}$$

4.3.3 The Second Stage of the Stackelberg Game

The optimal data price of the itch data consumer is calculated. Since different pricing has an impact on the optimal demand of data consumers, we substitute the expression (6) of demand d^* about price p_i into (1), and the incentive function R_P of the data provider is converted into:

$$R_P = p_i \left(\frac{\alpha_i}{p_i} - k_i \right) - c \left(\frac{\alpha_i}{p_i} - k_i \right) \tag{7}$$

Then, we calculate the partial derivative of the activation function R_P concerning p_i. The first-order and second-order partial derivatives of Eq. (7) are as follows:

$$\begin{aligned} \frac{\partial R_P}{\partial p_i} &= \frac{c\alpha_i}{p_i^2} - k_i \\ \frac{\partial^2 R_P}{\partial p_i^2} &= -\frac{2c\alpha_i}{p_i^3} < 0 \end{aligned} \tag{8}$$

From Eq. (8), it can be concluded that R_P is a convex function curve that increases first and then decreases. Therefore, R_P can reach the maximum value at $\frac{\partial R_P}{\partial p_i} = 0$, and p^* can be obtained.

$$p^* = \sqrt{\frac{c\alpha_i}{k_i}} \tag{9}$$

In the proposed Stackelberg game-based pricing strategy, medical institutions negotiate with many data consumers as data collectors. Therefore, data collectors can find the optimal unit price p^* for data resources, and data consumers can also choose the optimal amount of data d^* according to the price. Therefore, when the Nash equilibrium point (p^*,d^*) is reached, data collectors and data consumers, as well as patients, can obtain the best revenue incentive. The pricing model takes into account the maximum benefits of all parties, has fairness, and encourages all parties to actively participate in the sharing and circulation of medical data.

5 Experiment and Analysis

The proposed scheme is implemented and tested on an AMD Ryzen 7 5800H with Radeon Graphics 3.20 GHz and 16 GB RAM computer. The operating system is Windows 10 Home Chinese version. The blockchain test network is a test network named “Ropsten”. The smart contract is compiled and

run using the online compiler Remix on the Chrome browser 116.0.5845.112 (official version 64-bit). zk-SNARK is initialized in Remix extension ZoKrates: 0.8.4. It is mainly divided into three aspects: zero-knowledge proof verification use case test, game pricing strategy numerical simulation, and zero-knowledge proof contract performance test to verify the feasibility and practicability of the proposed scheme.

5.1 Verification Case

At present, the algorithms that can realize the non-interactive zero-knowledge proof of zk-SNARKs mainly include PGHR13, GM17, G16, etc., and the scheme selection is G16. Compared with other algorithms, the G16 algorithm has the smallest verification time and concise zero-knowledge proof and has completeness and reliability [38].

In this test, only the specific attribute of age is verified, where the range [*min, max*] is publicly declared information. The specific age information is private input, revealed solely during the doctor's proof generation process, and remains concealed from the verifier. ZoKrates automatically computes the requisite witness based on the circuit. Subsequently, the global parameter configuration and proof generation are executed.

In the non-interactive zero-knowledge proof contract, the contract interface such as Fig. 3 is provided. During the verification process, only public inputs and *Proof* need to be sent to the verifier, as depicted in Fig. 4a. Any participant, including data collectors and data consumers, can use the proof information to call the contract for verification. If the verification is passed, True is returned as Fig. 4b, indicating that the declaration information is credible. Otherwise, False is returned as Fig. 4c, the declaration is incorrect, and the data is not credible.

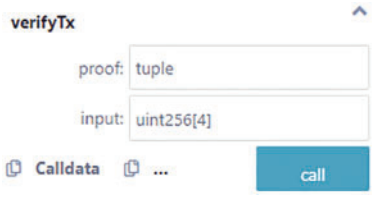

Figure 3: Verification contract interface

(a) Proof and public inputs (b) Log of verification success (c) Log of verification failure

Figure 4: Smart contract verification

5.2 Numerical Simulation of Pricing Strategy

According to the pricing strategy, data consumers first select the optimal amount of data according to the price set by the data collector. Simulate the impact of data demand on the satisfaction of the same data consumer at different prices set by the data collector. The non-zero positive factor $\alpha = 10000$ is set here, and the minimum demand d_{min} is set to 100. When the amount of data is lower than the minimum demand, the initial information gain k is greater than the minimum demand, so when the demand is lower than the minimum demand, the satisfaction is non-negative. To facilitate the simulation data, k is set to 1.5 times the minimum demand, $k = 150$, price $p = 10, 20, 30, 40, 50$. To facilitate the calculation and ensure that the cost c is lower than the changing price p, set the cost $c = 1$.

For the same data consumer, the impact of demand on satisfaction is shown in Fig. 5. In the case of different pricing by data collectors, data consumer satisfaction will increase first and then decrease with the increase of data demand, indicating that data consumer satisfaction will not continue to increase with the increase of data demand without considering the impact of data price. Therefore, data consumers have the optimal data demand to maximize their satisfaction. In addition, for the same data demand, data consumer satisfaction gradually decreases with the price increase, which means that data consumers need to pay more to obtain the same income. Under the same satisfaction, the lower the price, the higher the amount of data required.

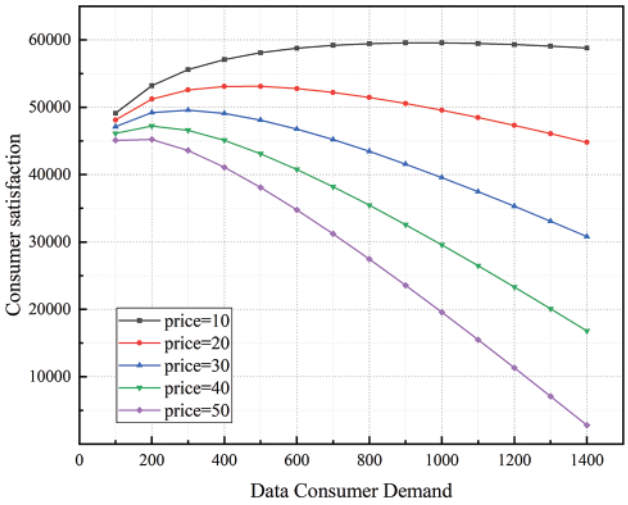

Figure 5: Impact of data demand on consumer satisfaction

As mentioned above, price has a great influence on data demand. To meet their maximum satisfaction, data consumers adjust the optimal demand strategy for different prices. Under the same parameter conditions, the impact of data price on the optimal demand of data consumers is shown in Fig. 6. When the data collector's pricing is low, the data consumer chooses to purchase a large amount of data to meet the maximum satisfaction. As the price of data increases, data consumers will reduce the amount of data purchased to balance the cost of expenditure to maximize their satisfaction. For data consumers, according to the pricing of data collectors, select the appropriate demand to achieve the maximum benefit.

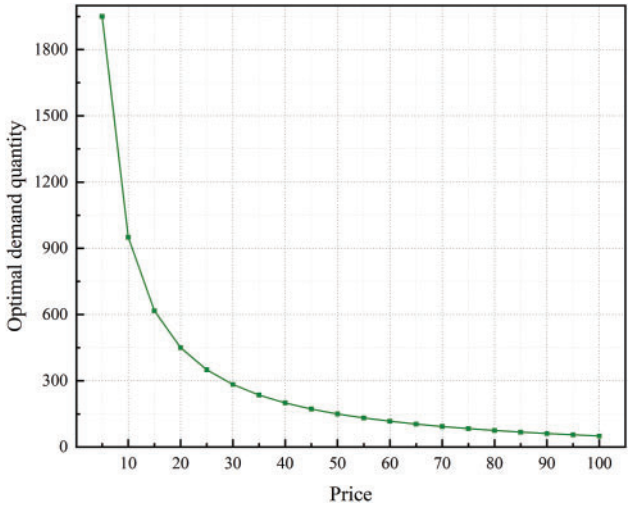

Figure 6: The impact of price on the optimal demand

Data demand is affected by data prices. For data providers, it is not that the higher the price, the greater the revenue. Set non-zero positive factors $\alpha = 10000, 15000, 20000, 25000$ to represent different data consumers. For any data consumer, the impact of price on the total revenue of the data provider is shown in Fig. 7. As the price increases, the total revenue increases first and then decreases due to the change in data consumer demand for data as shown in Fig. 6. Therefore, there is an optimal price to maximize the benefits of data providers.

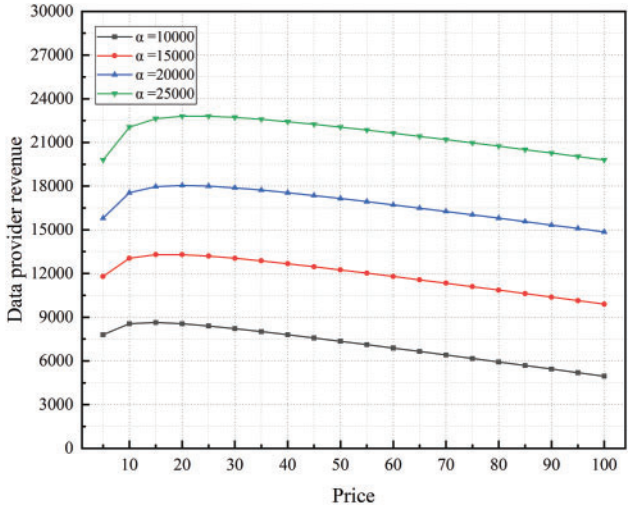

Figure 7: The impact of price on the total revenue of data provider

5.3 Smart Contracts Cost

We compiled the smart contract using the Solidity programming language in the Remix online compiler and deployed it. The expected gas costs for deploying the contract and executing it are shown in Table 2. The gas consumed here is only a test value and does not represent a real cryptocurrency. In the real system, the alliance chain and more efficient consensus algorithm can be used to further improve the efficiency of the blockchain network.

Table 2: Gas cost of the smart contract

Function	Deploy contract gas	Transaction gas
Verification SC	1485734	244927
Pricing SC	893573	23749

5.4 Schemes Comparison

To illustrate the advantages of the proposed scheme, it is compared with the data transaction-related schemes (different schemes are selected from the related work, as well as the traditional centralized architecture scheme). An analysis was conducted comparing five key aspects: Decentralized architecture, Tamper-proof records, Privacy protection for attribute information, Trustworthy validation of data on the blockchain, and Fair incentive pricing strategy.

The comparison results are shown in Table 3, from which we can see that the scheme has significant advantages. It should be particularly emphasized that compared with the comparison scheme, this paper adopts the decentralized architecture to verify the chain with zero knowledge proof and realizes the credible verification of data under the premise of attribute information privacy protection. At the same time, in terms of fairness and incentives for transaction prices, and from the perspective of maximizing social welfare, the improved Stackelberg game pricing strategy is used to enable all parties to achieve the best returns and encourage participants to actively participate and maintain honest behavior.

Table 3: Scheme comparison

Scheme	Decentralized architecture	Tamper-proof records	Privacy protection for attribute information	Trustworthy validation of data on the blockchain	Fair incentive pricing strategy
Traditional scheme	×	×	×	×	×
Nguyen et al. [19]	√	√	×	×	×
Zhao et al. [22]	√	√	×	√	×
Gao et al. [25]	√	√	×	×	×
Xue et al. [26]	√	√	×	√	×
Zheng et al. [28]	√	√	×	×	√
Proposed scheme	√	√	√	√	√

6 Conclusions

This paper proposes a trusted and fair medical data transaction scheme based on smart contracts, which solves the problem of untrustworthy data and unfair prices in the process of medical data transactions. In terms of data trustworthiness verification, zero-knowledge scope proof is used to verify data trustworthiness under the premise of protecting attribute privacy, which effectively protects patient privacy. In terms of data transaction pricing, the improved game model is used to select the best strategy for both parties to the transaction, which proves the existence of Nash equilibrium, so that all parties can obtain the maximum return and maximize social welfare. The research will provide useful guidance and innovative ideas for the future development of medical data transactions, to promote the field of health care towards more secure, credible, and fair data transactions.

Although blockchain technology provides a distributed and decentralized security guarantee for medical data transactions, its performance in dealing with large-scale transactions has been widely scrutinized and questioned. Due to limitations in the consensus mechanism, when transaction volume and load are high, the blockchain network may experience transaction delays, making it challenging to meet real-time demands. Additionally, there are difficulties in integrating with hospital systems in practical applications. Therefore, the next research direction will focus on optimizing consensus algorithms to enhance the efficiency of the blockchain network and exploring methods that facilitate easier integration with real-world hospital systems.

Acknowledgement: We would like to acknowledge the editors and anonymous reviewers.

Funding Statement: This research was funded by the Natural Science Foundation of Hebei Province (F2021201052). The author who received the grant is identified by X. Yang.

Author Contributions: Study conception and design: X. Yang, K. Zhang; formal analysis and validation: X. Yang, K. Zhang; methodology: X. Yang, K. Zhang; analysis and interpretation of results: X. Yang, K. Zhang; draft manuscript preparation: X. Yang, K. Zhang; writing-review & editing: X. Yang, K. Zhang. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The authors confirm that the data and materials supporting the findings of this study are not applicable as there were no specific datasets or materials used.

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

References

[1] J. M. Nolin, "Data as oil, infrastructure or asset? Three metaphors of data as economic value," *Journal of Information, Communication and Ethics in Society*, vol. 18, no. 1, pp. 28–43, 2020.

[2] S. C. Lin, C. R. Lyles, U. Sarkar and J. Adler-Milstein, "Are patients electronically accessing their medical records? Evidence from national hospital data," *Health Affairs*, vol. 38, no. 11, pp. 1850–1857, 2019.

[3] Researchers at greenbone networks vulnerability analysis and management company discovered 400 million medical radiological images exposed online via unsecured pacs servers, 2019. https://securityaffairs.com/91452/hacking/pacs-servers-unprotected-online.html (accessed on 19/05/2023).

[4] L. J. Kish and E. J. Topol, "Unpatients—why patients should own their medical data," *Nature Biotechnology*, vol. 33, no. 9, pp. 921–924, 2015.

[5] Healthcare Medical Data Compliance Circulation Standards, 2023. http://www.cinsa.org.cn/2023/0901/c33219a517345/page.htm (accessed on 04/12/2023).

[6] Electronic Health Records Market Size, Share, Growth Report, 2021. https://www.zionmarketresearch.com/report/electronic-health-records-market (accessed on 05/05/2023).
[7] P. Zhu, J. Hu, Y. Zhang and X. Li, "Enhancing traceability of infectious diseases: A blockchain-based approach," *Information Processing & Management*, vol. 58, no. 4, 2021. https://doi.org/10.1016/j.ipm.2021.102570
[8] S. Delgado-Segura, C. Pérez-Solà, G. Navarro-Arribas and J. Herrera-Joancomartí, "A fair protocol for data trading based on bitcoin transactions," *Future Generation Computer Systems*, vol. 107, pp. 832–840, 2020.
[9] W. G. Van Panhuis, P. Paul, C. Emerson, J. Grefenstette, R. Wilder *et al.,* "A systematic review of barriers to data sharing in public health," *BMC Public Health*, vol. 14, no. 1, pp. 1–9, 2014.
[10] W. S. Juang and Y. Y. Shue, "A secure and privacy protection digital goods trading scheme in cloud computing," in *2010 Int. Computer Symp. (ICS2010)*, Tainan, Taiwan, IEEE, pp. 288–293, 2010.
[11] C. Chen and J. Liao, "Fair offline digital content transaction system," *IET Information Security*, vol. 6, no. 3, pp. 123–130, 2012.
[12] R. J. Hwang and C. H. Lai, "Provable fair document exchange protocol with transaction privacy for e-commerce," *Symmetry*, vol. 7, no. 2, pp. 464–487, 2015.
[13] T. Jung, X. Y. Li, W. Huang, Z. Qiao, J. Qian *et al.,* "AccountTrade: Accountability against dishonest big data buyers and sellers," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 1, pp. 223–234, 2018.
[14] G. Su, W. Yang, Z. Luo, Y. Zhang, Z. Bai *et al.,* "BDTF: A blockchain-based data trading framework with trusted execution environment," in *2020 16th Int. Conf. on Mobility, Sensing and Networking (MSN)*, Tokyo, Japan, pp. 92–97, 2020.
[15] H. Li, L. Pei, D. Liao, X. Wang, D. Xu *et al.,* "BDDT: Use blockchain to facilitate IoT data transactions," *Cluster Computing*, vol. 24, pp. 459–473, 2021.
[16] S. Bajoudah, C. Dong and P. Missier, "Toward a decentralized, trust-less marketplace for brokered IoT data trading using blockchain," in *2019 IEEE Int. Conf. on Blockchain (Blockchain)*, Atlanta, GA, USA, IEEE, pp. 339–346, 2019.
[17] Z. Guan, X. Shao and Z. Wan, "Secure fair and efficient data trading without third party using blockchain," in *2018 IEEE Int. Conf. on Internet of Things and IEEE Green Computing and Communications and IEEE Cyber, Physical and Social Computing and IEEE Smart Data*, Halifax, NS, Canada, IEEE, pp. 1395–1401, 2018.
[18] L. D. Nguyen, I. Leyva-Mayorga, A. N. Lewis and P. Popovski, "Modeling and analysis of data trading on blockchain-based market in IoT networks," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6487–6497, 2021.
[19] A. Nawaz, J. Peña Queralta, J. Guan, M. Awais, T. N. Gia *et al.,* "Edge computing to secure IoT data ownership and trade with the Ethereum blockchain," *Sensors*, vol. 20, no. 14, 2020. https://doi.org/10.3390/s20143965
[20] P. Zhu, C. Miao, Z. Wang and X. Li, "Informational cascade, regulatory focus, and purchase intention in online flash shopping," *Electronic Commerce Research and Applications*, vol. 62, 2023. https://doi.org/10.1016/j.elerap.2023.101343
[21] S. Kiyomoto and K. Fukushima, "Fair-trading protocol for anonymised datasets requirements and solution," in *2018 4th Int. Conf. on Information Management (ICIM)*, Oxford, UK, IEEE, pp. 13–16, 2018.
[22] Y. Zhao, Y. Yu, Y. Li, G. Han and X. Du, "Machine learning based privacy-preserving fair data trading in big data market," *Information Sciences*, vol. 478, pp. 449–460, 2019.
[23] D. Sheng, M. Xiao, A. Liu, X. Zou, B. An *et al.,* "CPchain: A copyright-preserving crowdsourcing data trading framework based on blockchain," in *29th Int. Conf. on Computer Communications and Networks*, Honolulu, HI, USA, IEEE, pp. 1–9, 2020.
[24] Y. J. Galteland and S. Wu, "Blockchain-based privacy-preserving fair data trading protocol," *Cryptology ePrint Archive*, 2021. https://eprint.iacr.org/2021/1321

[25] J. Gao, T. Wu and X. Li, "Secure, fair and instant data trading scheme based on bitcoin," *Journal of Information Security and Applications*, vol. 53, 2020. https://doi.org/10.1016/j.jisa.2020.102511
[26] L. Xue, J. Ni, D. Liu, X. Lin and X. Shen, "Blockchain-based fair and fine-grained data trading with privacy preservation," *IEEE Transactions on Computers*, vol. 72, no. 9, pp. 2440–2453, 2023.
[27] C. Chen, J. Wu, H. Lin, W. Chen and Z. Zheng, "A secure and efficient blockchain-based data trading approach for Internet of Vehicles," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 9, pp. 9110–9121, 2019.
[28] S. Zheng, L. Pan, D. Hu, M. Li and Y. Fan, "A blockchain-based trading platform for big data," in *IEEE INFOCOM 2020-IEEE Conf. on Computer Communications Workshops*, Toronto, ON, Canada, IEEE, pp. 991–996, 2020.
[29] C. Li, S. Liang, J. Zhang, Q. E. Wang and Y. Luo, "Blockchain-based data trading in edge-cloud computing environment," *Information Processing & Management*, vol. 59, no. 1, 2022. https://doi.org/10.1016/j.ipm.2021.102786
[30] N. C. Luong, N. D. D. Anh, N. H. Sang, S. Feng, V. D. Nguyen *et al.,* "Optimal auction for effective energy management for UAV-assisted metaverse synchronization system," in *2023 IEEE 20th Consumer Communications & Networking Conf. (CCNC)*, Las Vegas, NV, USA, IEEE, pp. 392–397, 2023.
[31] H. Zhang, Y. Xiao, L. X. Cai, D. Niyato, L. Song *et al.,* "A multi-leader multi-follower stackelberg game for resource management in LTE unlicensed," *IEEE Transactions on Wireless Communications*, vol. 16, no. 1, pp. 348–361, 2017.
[32] V. Cardellini, V. di Valerio and F. L. Presti, "Game-theoretic resource pricing and provisioning strategies in cloud systems," *IEEE Transactions on Services Computing*, vol. 13, no. 1, pp. 86–98, 2016.
[33] S. Guo, X. Hu, G. Dong, W. Li and X. Qiu, "Mobile edge computing resource allocation: A joint Stackelberg game and matching strategy," *International Journal of Distributed Sensor Networks*, vol. 15, no. 7, pp. 1550147719861556, 2019.
[34] C. Li, M. Dong, J. Li, G. Xu, X. Chen *et al.,* "Healthchain: Secure EMRs management and trading in distributed healthcare service system," *IEEE Internet of Things Journal*, vol. 8, no. 9, pp. 7192–7202, 2020.
[35] S. Nakamoto and A. Bitcoin, "A peer-to-peer electronic cash system," *Bitcoin*, 2008. https://bitcoin.org/bitcoin.pdf (accessed on 20/06/2023).
[36] P. Zhu, J. Hu, X. Li and Q. Zhu, "Using blockchain technology to enhance the traceability of original achievements," *IEEE Transactions on Engineering Management*, vol. 70, no. 5, pp. 1693–1707, 2023.
[37] V. Buterin, "A next-generation smart contract and decentralized application platform," *White Paper*, 2014. https://finpedia.vn/wp-content/uploads/2022/02/Ethereum_white_paper-a_next_generation_smart_contract_and_decentralized_application_platform-vitalik-buterin.pdf (accessed on 20/06/2023).
[38] J. Groth, "On the size of pairing-based non-interactive arguments," in *Advances in Cryptology-EUROCRYPT 2016: 35th Annual Int. Conf. on the Theory and Applications of Cryptographic Techniques*, Vienna, Austria, pp. 305–326, 2016.
[39] J. Eberhardt and S. Tai, "ZoKrates-scalable privacy-preserving off-chain computations," in *2018 IEEE Int. Conf. on Internet of Things and IEEE Green Computing and Communications and IEEE Cyber, Physical and Social Computing and IEEE Smart Data*, Halifax, NS, Canada, pp. 1084–1091, 2018.