



BDTS: Blockchain-Based Data Trading System

Erya Jiang¹, Bo Qin¹(✉), Qin Wang², Qianhong Wu³, Sanxi Li⁴,
Wenchang Shi¹, Yingxin Bi¹, and Wenyi Tang⁵

¹ School of Information, Renmin University of China, Beijing, China
{jiangey2018202191,bo.qin,wenchang,biyingxin11}@ruc.edu.cn

² CSIRO Data61, Eveleigh, Australia

³ Beihang University, Beijing, China

Qianhong.wu@buaa.edu.cn

⁴ School of Economics, Renmin University of China, Beijing, China
sanxi@ruc.edu.cn

⁵ University of Notre Dame, Notre Dame, USA
wtang3@nd.edu

Abstract. Trading data through blockchain platforms is hard to achieve *fair exchange*. Reasons come from two folds: Firstly, guaranteeing fairness between sellers and consumers is a challenging task as the deception of any participating parties is risk-free. This leads to the second issue where judging the behavior of data executors (such as cloud service providers) among distrustful parties is impractical in the context of traditional trading protocols. To fill the gaps, in this paper, we present a blockchain-based data trading system, named BDTS. BDTS implements a fair-exchange protocol in which benign behaviors can get rewarded while dishonest behaviors will be punished. Our scheme requires the seller to provide consumers with the correct encryption keys for proper execution and encourage a rational data executor to behave faithfully for maximum benefits from rewards. We analyze the strategies of consumers, sellers, and dealers in the trading game and point out that everyone should be honest about their interests so that the game will reach Nash equilibrium. Evaluations prove efficiency and practicability.

Keywords: Data Trading · Blockchain · Fair Exchange

1 Introduction

Data has risen to a new factor of production alongside traditional factors such as land, labor, capital and technology. Consumers, sellers, and data trading intermediaries together form a thriving data trading ecosystem, in which the consumer has to pay a fortune to the seller for acquisition, the seller could make some profits by providing the appropriate data and data trading intermediaries earn agent fees between sellers and consumers. However, such a high density of centralization is likely to be the weak spot to be attacked. On the one hand, any participating roles may act maliciously in the unsupervised system. The sellers

may provide fake data for profits as they may not own data as they claimed. The consumer may refuse to pay after receiving the appropriate data. The intermediaries such as cloud service providers may manipulate the stored data without permission from users [1, 2]. On the other hand, relying on centralized servers confront heavy communication and security pressure, greatly constraining the efficiency of the entire trading system. These challenges lead to the following questions,

Is it possible to propose a protocol in the data trading system with guaranteed fairness for all parties without significantly compromising efficiency?

Traditional solutions using cryptography and relying on trusted third parties (TTP) [3, 4] lack practical significance because finding such a TTP is reckon hard in practice. Instead of using *gradual release* method¹ [5, 6], many solutions [7–12] have been proposed by leveraging blockchain technologies [13] for better decentralization. Blockchain provides a public bulletin board [14] for every participating party with persistent evidence. A normal operating blockchain platform greatly reduces the risk of being attacked like a single-point failure or compromised by adversaries. Self-executing smart contracts always act benign and follow agreed principles, with transparency and accountability.

Based on such investigations, we adopt the blockchain technique as our baseline solution, with smart contracts acting as data executors. Specifically, we implement our scheme with strict logic of fair exchange on the Hyperledger blockchain. [15] Overall, implementing a data trading system with both exchange fairness and efficiency for real usage is the key task in this paper. To fill the gap, our *contributions* are as follows.

- We purpose BDTS, an innovative data trading system based on blockchain (Sect. 3 & 4). The proposed scheme realizes the exchange fairness for *all* participated parties, namely, *consumer*, *seller* and *service provider*. Each party has to obey the rules, and benign actors can fairly obtain incentivized rewards. Every data can only be sold once since each transaction is unique in the blockchain systems. Notably, we use the uniqueness index mechanism [16] and compare Merkle roots of different data to prevent someone from reselling data purchased from others.
- We prove the security of our scheme majorly from the economical side based on game theory (Sect. 5). Our proof simulates the behaviors of different parties, which is an effective hint to show actual reflection towards conflicts as well as real action affected by competitive participants. The proofs demonstrate that our game reaches *the subgame perform equilibrium (SPE)* [17, 18].
- We implement our scheme on the Hyperledger Fabric blockchain platform with comprehensive evaluations (Sect. 6). Experimental results prove the efficiency and practicability. Compared to existing solutions with complex crypto-algorithms (e.g. zero-knowledge proof), our scheme is sufficiently fast for lightweight deceives.

¹ The *gradual release* method means each party in the game takes turns to release a small part of the secret. Once a party is detected doing evil, other parties can stop immediately and recover the desired output with similar computational resources.

2 Related Work

In this section, we provide related primitives surrounding *fair exchange* protocols and *blockchains*. We compared the differences and main pros and cons between the references in Table 1.

Blockchain in Trading Systems.

Due to its non-repudiation, non-equivocation and non-frameability, blockchain has been widely used in trading systems [14]. Jung et al. [19] propose AccountTrade, an accountable trading system between customers who distrust each other. Any misbehaving consumer can be detected and punished by using book-keeping abilities. Chen et al. [16] design an offline digital content trading system. If a dispute occurs, the arbitration institution will conduct it. Dai et al. [20] propose SDTE, a trading system that protects data and prevents analysis code from leakage. They employ trusted execution environment(TEE) to protect data in an isolated area at the hardware level. Similarly, Li et al. [21] leverage the TEE-assisted smart contract to trace the evidence of investigators' actions. Automatic executions enable warrant execution accountability with the help of TEE. Zhou et al. [22] introduce a data trading system that prevents index data leakage where participants exchange data via smart contracts. These solutions rely on blockchain to create persistent evidence and act as a transparent authority to solve disputes. However, they merely perform effectively in trading the *text* data, rather than data cast in streaming channels such as TV shows and films, which are costly. The fairness issue has neither been seriously discussed.

Fair Exchanges using Blockchain. Traditional ways of promoting fair exchange across distrustful parties rely on trusted third parties because they can monitor the activities of participants, judging whether they have faithfully behaved. However, centralization is the major hurdle. Blockchain can perfectly replace the role of TTP. The behaviors of involved parties are transparently recorded on-chain, avoiding any types of cheating and compromising. Meanwhile, a predefined incentive model can be automatically operated by smart contracts, guaranteeing that each participant can be rewarded according to their contributions. Dziembowski et al. [7] propose Fairswap, utilizing the smart contract to guarantee fairness. The contract plays the role of an external judge to resolve the disagreement. He et al. [8] propose a fair content delivery scheme by using blockchain. They scrutinize the concepts between exchange fairness and delivery fairness during the trades. Eckey et al. [12] propose a smart contract-based fair exchange protocol with an optimistic mode. They maximally decrease the interaction between different parties. Janin et al. [23] present FileBounty, a fair protocol using the smart contract. The scheme ensures a buyer purchases data at an agreed price without compromising content integrity. Besides, blockchains are further applied to multi-party computations of trading systems [9, 10, 24].

Table 1. Reference Summary

Reference	[19]	[16]	[20–22]	[7]	[8, 12]	This work
Decentralization	✓	✓	✓	✓	✓	✓
Fairness	×	×	×	✓	✓	✓
Prevent Resale	×	✓	×	×	×	✓
Multimedia Data	✓	×	×	×	✓	✓

3 Architecture and Security Model

Entities. First of all, we clarify the participating roles in our scheme. A total of three types of entities are involved: *consumer (CM)*, *seller (SL)*, and *service provider (SP)*². Consumers pay for data and downloading service with cryptocurrencies such as Ether. Sellers provide encrypted data as well as expose the segment of divided data when necessary to guarantee correctness. Service providers take tasks of storage and download services, and any participant who stores encrypted data can be regarded as a service provider. Miners and other entities participating in systems are omitted as they are out of the scope of this paper.

Architecture. We design a novel data trading ecosystem that builds on the top of the blockchain platform. A typical workflow in BDTS is that: *Sellers upload their encrypted data and description to service providers. Service providers store received data and establish download services. Consumers decide which pieces of data to purchase based on related descriptions. At last, the consumer downloads from service providers and pays the negotiated prices to sellers and service providers.* Our fair exchange scheme is used to ensure every participant can exchange data with payments without cheating and sudden denial.

Data Upload. The seller first sends his description of data to the blockchain. Description and other information such as the Merkle root of data would be recorded by blockchain. Here, the seller must broadcast the Merkle root and they are demanded to expose a certain number of plaintext data pieces. Service providers can decide whether they are going to store it by the stated information. At the time, the seller waits for the decision from the service providers. If a service provider decides to store encrypted data for earning future downloading fees, he first sends his information to the blockchain. The seller will post encrypted data to the service provider and the service provider starts to store the data. Notably, the seller can also become a self-maintained service provider if he can build up similar basic services.

Data Download. The consumer decides to download or not according to the description and exposed parts provided by the seller. Before downloading, the consumer should first store enough tokens on the smart contract. Then, the consumer sends a request for data from service providers. Service providers will send it to the consumer after encrypting the data with the private key. For security and efficiency, these processes will be executed via smart contracts, except for data encryption and downloading.

Decryption and Appealing. The consumer should pay for data and get the corresponding decryption key. The service provider and seller will provide their decryption key separately. The decryption key is broadcast through the blockchain so that it cannot be tampered with. The consumer can appeal whether

² Service providers generally act as the role of centralized authorities such as dealers and agencies in a traditional fair exchange protocol.

it is due to the receipt of a false decryption key or the verification finds that the data has been falsified or fabricated. The smart contract will arbitrate based on evidence provided by the consumer.

Security Assumption. We have three basic security assumptions. (i) *The blockchain itself is assumed to be safe.* Our scheme operates on a safe blockchain model with well-promised *liveness* and *safety* [25]. Meanwhile, miners are considered to be honest but *curious*: they will execute smart contracts correctly but may be curious about the plaintext recorded on-chain. (ii) *The basic crypto-related algorithm is safe.* This assumption indicates that the encryption and decryption algorithm will not suffer major attacks that may destroy the system. Specifically, AES and the elliptic curve, used for asymmetric encryption algorithms, are sufficiently supposed to be safe in cryptography. (iii) *Participants in this scheme are rational.* As the assumption of game theory, all players (consumer, seller, and service provider) are assumed to be rational: these three types of players will act honestly but still pursue profits within the legal scope.

Security Model. We dive into the strategies of each party.

Seller intend to obtain more payment by selling their data. In our scheme, a seller needs to provide mainly three sectors: *data*, *description*, and *decryption-key*. To earn profits, a seller would claim the data is popular and deserved to be downloaded, but he may provide fake data. The exchange is deemed as *fair* if consumers obtain authentic data that is matched with claimed descriptions. Then, the seller can receive rewards. Encryption is another component provided by the seller. Only the correct decryption key can decrypt the encrypted data, whereas the false one cannot. In summary, there are four potential strategies for sellers: a) *matched data (towards corresponding description) and matched key (towards data)*, b) *matched data and non-matched key*, c) *non-matched data and matched key*, and d) *non-matched data and non-matched key*.

Consumer intend to exchange their money for data and downloading services. Downloading ciphertext and decrypting it to gain plaintext is in their favor. Consumers provide related fees in our scheme and then download encrypted data from service providers who store the uploaded data. To earn profits, they intend to obtain data without paying for it or paying less than expected. Paying the full price as expected for data is a sub-optimal choice. The payment of consumers can be divided into two parts: paying the seller for the decryption key and paying service providers for the downloading service. Based on that, there are four strategies for consumers: a) *pay enough for sellers*, b) *pay less for sellers*, c) *pay enough for service providers*, and d) *pay less for service providers*.

Service Providers intend to provide the downloading service and earn profits. Service providers are like platforms, by storing as much data as possible and offering download service, they can ultimately attract clout and make a profit from the download fees. For uploading, service providers can choose whether to store data or not. Here, a seller can act as a service provider if he provides similar services of storage and download. For downloading, service providers will

provide encrypted data and the corresponding decryption key. The strategies for service providers are listed as follows: a) *authentic correct data and matched key*, b) *authentic data and non-matched key*, c) *fake data and matched key*, and d) *fake data and non-matched key*. The first two need the premise of storing the seller’s data.

Strategy Assumption. For security, an ideal strategy for the system is to reach a Nash equilibrium for all participants: sellers adopt the *correct data and matched key* strategy, consumers adopt the *pay enough for sellers, paying enough for service providers* and service providers who provide storing services adopt the *authentic correct data and matched key* strategy (discussed in Sect. 5).

4 The BDTS Scheme

In this section, we provide the concrete construction. To achieve security goals as discussed, we propose our blockchain-based trading system, called BDTS. It includes four stages: *contract deployment*, *encrypted data uploading*, *data downloading*, and *decryption and appealing*. Our scheme involves three types of contracts. Here, we omit the procedures such as signature verification and block mining because they are known as common sense.

Module Design. The system contains three types of smart contracts: seller-service provider matching contract (SSMC), service provider-consumer matching contract (SCMC), and consumer payment contract (CPC). Table 2 outlines the notation used in the module description.

Table 2. Notation

Notation	Description
SL, SP, CM	seller/service provider/consumer
K_{role}	the key of symmetric encryption algorithm of role
$Data_i$	the i -th unit of data plaintext in binary form
D_i	$Enc_{K_i}^{AES}(Data_i)$, the i -th unit of data encrypted by seller
DD_i	$Enc_{K_{sp}}^{AES}(D_i)$, encrypted D_i by service provider
Pub_{role}, Pri_{role}	the public/private key of asymmetric encryption algorithm of role
A_{role}	Ethereum address of role
IP_{role}	IP address of role
ID_{data}	the data ID, index of data in SSMC
$Desc$	the seller’s description of data
$Mtree(\Delta), Mproof(\Delta), Mvrfy(\Delta)$	the Merkle tree algorithms
Tkn_{role}	the token sent by role
$Price$	the price of entire data
$Unit Price$	downloading price for each unit

Seller-Service Provider Matching Contract. SSMC records the description and the Merkle root of data. The seller is required to broadcast certain

parts of data and the index of these parts should be randomly generated by the blockchain. Notably, these indexes cannot be changed once they have been identified. Last, SSMC matches service providers for every seller.

Service Provider-Consumer Matching Contract. SSMC helps consumers and service providers reach an agreement. It receives and stores the consumers' data, including required data and related information. The contract requires consumers to send payment. Then, the payment is sent to CPC.

Consumer Payment Contract. CPC works to command consumers to pay for data and command sellers to provide the decryption key. It achieves a fair exchange between decryption key (data downloading) and payment.

Encrypted Data Upload. In this module, a seller registers on SSMC and the service provider stores encrypted data (cf. Fig. 1.a).

Step1. When a seller expects to sell data for profits, he should first divide data into several pieces and encrypt them separately with different keys (denoted as K_i , where $i = 1, 2, \dots, n$), which is generated based on K . Such pieces of data should be valuable so that others can judge the quality of full data with the received segments. Here, $D_i = Enc_{K_i}^{AES}(Data_i)$ is the encrypted data.

Step2. The seller sends a registration demand in the form of a transaction. The registration demand includes the seller's information and data description. The seller information consists of A_{seller} and IP_{seller} . Data description includes four main parts: *content description*, *data size*, *the root r_d* and *the root r_{ed}* . Here, r_d is the root of M_d and r_{ed} is the root of M_{ed} , where $M_d = Mtree(Data_1, Data_2, \dots, Data_n)$ and $M_{ed} = Mtree(D_1, D_2, \dots, D_n)$. They will be recorded in SSMC. Tokens will also be sent as deposit in this step and may be lost later if the data is found resold. SSMC will reject the request if the corresponding r_d is the same as that of data recorded before. This mechanism prevents reselling on the blockchain platform.

Step3, 4. After approving the seller's registration demand, SSMC stores useful information. Blockchain generates the hash of the next block and uses it as a public random *seed*.

Step5, 6. The seller runs $Rand(seed)$ to get a sequence of random numbers I_{rand} . The number of random numbers generated is the number of data units that need to be exposed. We assume that this number can support semantic comparison with the data description and data plagiarism detection without disclosing too much plaintext data. The seller provides $(Data_{I_{rand}}, P_d, P_{ed})$ to SSMC, where $P_d = Mproof(M_d, I_{rand})$ and $P_{ed} = Mproof(M_{ed}, I_{rand})$. The contract SSMC checks $Mvrfy(i, r_d, Data_i, P_{d_i}) == 1$ and $Mvrfy(i, r_{ed}, D_i, P_{ed}) == 1$. If not, SSMC stops execution and returns error. Then, the exposed pieces of data will be compared to other pieces by utilizing the uniqueness index. Data plagiarism will result in deposit loss, preventing the reselling behavior. The authenticated data will be assigned an ID.

Step7, 8. The SP registration demand can be divide into IP_{sp} , A_{sp} , ID_{data} and *unit price*.

Step9, 10, 11. The seller sends encrypted data and Merkle proof to the service provider according to IP_{sp} and confirms the registration demand so that the corresponding service provider can participate in the next stage.

Matching and Data Downloading. In this module, a consumer registers on SCMC and selects the service provider to download data (see Fig. 1.b).

Step1, 2. The consumer queries for data description and the exposed pieces of data. The consumer compares the description with exposed data content and selects data once receiving feedback from SSMC.

Step3, 4, 5. The consumer stores the tuple $(IP_{sp}, A_{cm}, ID_{data})$ on SCMC and sends enough tokens to pay for the download service. These tokens will be sent to CPC and, if unfortunately the service provider or seller cheats on this transaction, will be returned to the consumer. When receiving the demand, SCMC queries SSMC with ID_{data} to obtain *price*, *datasize* and *unit price*. Then, SCMC will verify $Tkn_{cm} \geq price + size * unit\ price$. Failed transactions will be discarded while the rest being broadcast. The seller can determine the piece of data and the service providers by giving index i and the corresponding address.

Step6, 7, 8. The consumer contacts the service provider based on IP_{sp} , received in *Step2*. In *Step7*, a service provider encrypts data D with the random key K_{sp} . The service provider will calculate M_{eed} , where $M_{eed} = Mtree(DD_1, DD_2, \dots, DD_n)$, with the Merkle root r_{eed} and upload P_{eed_i} to SCMC, where $P_{eed_i} = Mproof(M_{eed}, i)$ and i is the index.

Step9,10. The selected service provider information is provided. It is composed of A_{sp} and the index of downloading pieces from service providers. The consumer can download data from multiple providers for efficiency. The service provider sends $DD = Enc_{K_{sp}}^{AES} D$ to consumers.

Step11, 12. The consumers need to verify whether or not $Mvrfy(i, r_{eed}, DD, P_{eed_i}) == 1$. If not, the (double-)encrypted data will be considered as an error if it cannot pass the verification and the consumer, as a result, will not execute *step14*.

Decryption and Appealing. In this module, the consumer pays both the service provider and the seller.

Payment to the service provider involves the following steps. (see Fig. 1.c)

Step1, 2, 3. SSMC transfers tokens and $(A_{cm}, A_{sp}, ID_{data})$ to CPC. The consumer generates a key pair (Pub_{cm}, Pri_{cm}) and broadcasts Pub_{cm} to CPC. CPC waits for the service provider to get $Enc_{Pub_{cm}}(K_{sp})$.

Step4, 5, 6. The consumer obtains $Enc_{Pub_{cm}}(K_{sp})$ from CPC. Then, he decrypts data with K_{sp} to get D'_i . If $Mvrfy(i, r_{ed}, D'_i, P_{ed}) \neq 1$, the consumer executes the appealing phase. Appeal contains (Pri_{cm}, i, DD_i) . Here, Pri_{cm} is generated in every download process. Otherwise, it indicates the decryption key and encrypted data received by the consumer are true, and CPC will send tokens to the service provider directly.

Step7, 8, 9. CPC calculates K_{sp} and D'_i , where $D'_i = Dec_{K_{sp}}^{AES}(DD)$ while the decryption key $K_{sp} = Dec_{Pri_{cm}}(Enc_{Pub_{cm}}(K_{sp}))$. Then, CPC verifies whether $Mvrfy(i, r_{ed}, D'_i, P_{ed}) \neq 1$. If it passes the verification, CPC withdraws the tokens to SSMC. Otherwise, CPC will pay the service providers.

Paying the seller is similar to paying the service providers, the differences between mainly concentrate on Step2, Step3, Step4, Step7, and Step8 (see Fig. 1.d).

Step2, 3, 4. The consumer generates a new public-private key pair (Pub_{cm}, Pri_{cm}) and broadcasts Pub_{cm} to CPC. After listening to CPC to get Pub_{cm} , the seller calculates $Enc_{Pub_{cm}}(K_{seller})$ and send it to CPC.

Step7, 8. During the appealing phase, the consumer relies on his private key to prove his ownership. CPC verifies the encryption of the corresponding data, which is similar to the step of paying for service providers. The verification will determine the token flow.

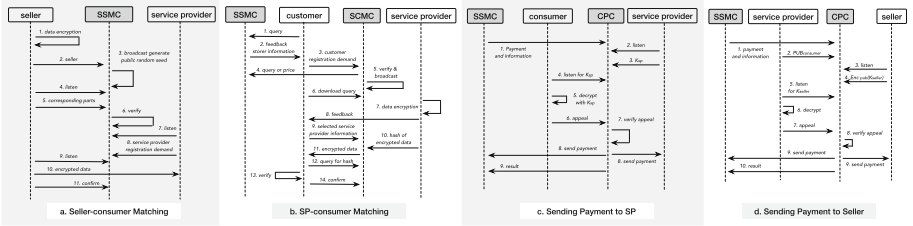


Fig. 1. Component Workflow

5 Security Analysis

In this section, we provide the analysis of BDTS based on game theory. The basic model of our solution is a *dynamically sequential game* with *multiple players*. The analyses are based on *backward induction*. We prove that our model can achieve a subgame perfect Nash equilibrium (SPNE) if all participants honestly behave.

Specifically, our proposed scheme consists of three types of parties, including seller (SL), service provider (SP), and consumer (CM) as shown in Fig. 2. These parties will act one by one, forming a sequential game. The following party can learn the actions from the previous. Specifically, A SL will first upload the data with the corresponding encryption key to the SP (workflow in *black* line). Once receiving data, the SP encrypts data by his private key and stores the raw data locally while related information is on-chain. CM searches online to find products and pay for the favorite ones both to SP and SL via smart contracts (in *blue* line). Last, the SP sends the raw data and related keys to CM (in *brown* line). Based on that, we define our analysis model as follows.

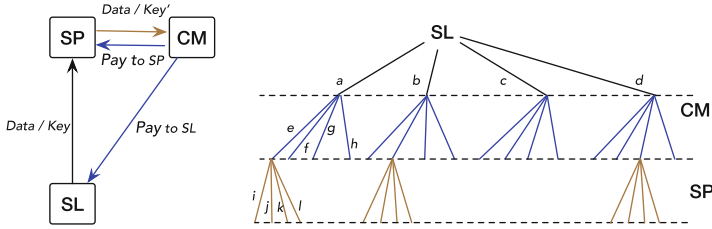


Fig. 2. Game and Game Tree

Definition 1. SM-SP-CM involved system forms an extensive game denoted by

$$\mathcal{G} = \{\mathcal{N}, \mathcal{H}, \mathcal{R}, P, u_i\}.$$

Here, \mathcal{N} represents the participated players where $\mathcal{N} = \{SL, SP, CM\}$; \mathcal{R} is the strategy set; \mathcal{H} is the history, P is the player function where $P : \mathcal{N} \times \mathcal{R} \rightarrow \mathcal{H}$; and u_i is the payoff function.

Each of participating parties, they have four strategies as defined in Sect.3 (security model). SL has actions on both updated data and related decryption keys (AES for raw data), forming his strategies \mathcal{R}_{SL} , where $\mathcal{R}_{SL} = \{a, b, c, d\}$. Similarly, CM has strategies $\mathcal{R}_{CM} = \{e, f, g, h\}$ to show his actions on payments to SL or SP. SP has strategies $\mathcal{R}_{SP} = \{i, j, k, l\}$ for actions on downloading data and related keys. We list them at Table 3.

However, it is not enough for quantitative analysis of the costs of these actions to be unknown. According to the market prices and operation cost, we suppose that a piece of raw data worth 10 units, while generating keys compensates 1 unit. The service fee during the transactions is 1 units for each party. Thus, we provide the cost of each strategy in the Table 3. The parameters of x and y are actual payments from CM, where $0 \leq x < 20, 0 \leq y < 4, x + y < 24$.

Then, we dive into the history set \mathcal{H} that reflects the conducted strategies from all parties before. For instance, the history aei represents all parties performing honestly. There are a total of 64 possible combinations (calculated by $64 = 4 * 4 * 4$) based on sequential steps of SL, SM, and SP. We provide their game tree in Table 3. We omit their detailed representation due to their intuitive induction. Our analysis is based on these fundamental definitions and knowledge. We separately show the optimal strategy (with maximum rewards) for each party, and then show how to reach a subgame perfect Nash equilibrium, which is also the Nash equilibrium of the entire game. Before diving into the details of calculating each subgame, we first drive a series of lemmas as follows.

Table 3. Strategies and Costs (i. The cost of -11 units are short for -11, applicable to all; ii. Data is sold at 20 (to SL) while the service fee is 4 (to SP))

<i>SL Strategy</i>	Matched data	Non-matched data
Matched key	a, -11	b, -1
Non-matched key	c, -10	d, 0
<i>CM Strategy</i>	Sufficiently	Insufficiently (to SL)
Sufficiently (to SP)	e, -24	f, -(x+4)
Insufficiently Paid	g, -(y+20)	h, -(x+y)
<i>SP Strategy</i>	Authentic data	Non-authentic data
Matched key	i, -2	j, -1
Non-matched key	k, -1	l, 0

Lemma 1. *If one seller provides data not corresponding to the description, the seller cannot obtain payments.*

Proof. The description and Merkle root of data are first broadcast before the generation of random indexes. Once completing the registration of the seller, the blockchain generates a random index. Exposed pieces are required to match the Merkle roots so that the seller cannot provide fake ones. Meanwhile, these pieces ensure that data can conform to the description. Otherwise, consumers will not pay for the content and service providers will not store it, either. \square

Lemma 2. *If one seller provides a decryption key not conforming to the description, the seller cannot obtain payments.*

Proof. The seller encrypts data (segmented data included) with his private keys. The results of both encryption and related evidence will be recorded by the smart contract, which covers the Merkle root of encrypted data and the Merkle root of data. If a seller provides a mismatched key, the consumer cannot decrypt the data and he has to start the appealing process. As D_i and receipt are owned by the consumer, if the consumer cannot obtain correct data, the consumer can appeal with evidence. The smart contract can automatically judge this appeal. If the submitted evidence is correct and decryption results cannot match the Merkle root of data, the contract will return deposited tokens to the consumer. \square

Lemma 3. *A consumer without sufficient payments cannot normally use data.*

Proof. The consumer will first send enough tokens to SCMC and this code of the smart contract is safe. The smart contract will verify whether the received tokens are enough for the purchase. After the seller and consumer provide their decryption key through the smart contract, the consumer can appeal at a certain time, or it's considered that the key is correct and payments will be distributed to the seller and service providers. \square

Lemma 4. *If one service provider provides data not conforming to that of the seller, he cannot obtain payments.*

Proof. This proof is similar to Lemma 1. \square

Lemma 5. *If one service provider provides a decryption key not conforming to data, he cannot obtain payments.*

Here, Lemma 1 to Lemma 5 prove the payoff function of each behavior. Based on such analyses, we can precisely calculate the payoff function of combined strategies in our sequential game. As discussed before, a total of 64 possible combinations exist, and we accordingly calculate the corresponding profits as presented in Table 4. We demonstrate that the system can reach the subgame perfect Nash equilibrium under the following theorem.

Table 4. Payoff Function and Profits (blue texts reach Nash Equilibrium)

\mathcal{H}	Payoff in the form of (SL, CM, SP)						
<i>aei</i>	(9, -4, 2)	<i>bei</i>	(19, -24, 2)	<i>cei</i>	(10, -24, 2)	<i>dei</i>	(20, -24, 2)
<i>aej</i>	(9, -24, 3)	<i>bej</i>	(19, -24, 3)	<i>cej</i>	(10, -24, 3)	<i>dej</i>	(20, -24, 3)
<i>aek</i>	(9, -24, 3)	<i>bek</i>	(19, -24, 3)	<i>cek</i>	(10, -24, 3)	<i>dek</i>	(20, -24, 3)
<i>ael</i>	(9, -24, 4)	<i>bel</i>	(19, -24, 4)	<i>cel</i>	(10, -24, 4)	<i>del</i>	(20, -24, 4)
<i>afi</i>	(x-11, 16-x, 2)	<i>bfi</i>	(x-1, -24, 2)	<i>cfi</i>	(x-10, -24, 2)	<i>dfi</i>	(x, -24, 2)
<i>afj</i>	(x-11, -24, 3)	<i>bfj</i>	(x-1, -24, 3)	<i>cfj</i>	(x-10, -24, 3)	<i>dfj</i>	(x, -24, 3)
<i>afk</i>	(x-11, -24, 3)	<i>bfk</i>	(x-1, -24, 3)	<i>cfk</i>	(x-10, -24, 3)	<i>dfk</i>	(x, -24, 3)
<i>afl</i>	(x-11, -24, 4)	<i>bfl</i>	(x-1, -24, 4)	<i>cfl</i>	(x-10, -24, 4)	<i>dfl</i>	(x, -24, 4)
<i>agi</i>	(9, -y, y-2)	<i>bgi</i>	(19, -24, y-2)	<i>cgi</i>	(10, -24, y-2)	<i>dgi</i>	(20, -24, y-2)
<i>agj</i>	(9, -24, y-1)	<i>bgj</i>	(19, -24, y-1)	<i>cgj</i>	(10, -24, y-1)	<i>dgj</i>	(20, -24, y-1)
<i>agk</i>	(9, -24, y-1)	<i>bgk</i>	(19, -24, y-1)	<i>cgk</i>	(10, -24, y-1)	<i>dgk</i>	(20, -24, y-1)
<i>agl</i>	(9, -24, y)	<i>bgl</i>	(19, -24, y)	<i>cgl</i>	(10, -24, y)	<i>dgl</i>	(20, -24, y)
<i>ahi</i>	(x-11, -x-y, y-2)	<i>bhi</i>	(x-1, -24, y-2)	<i>chi</i>	(x-10, -24, y-2)	<i>dhi</i>	(x, -24, y-2)
<i>ahj</i>	(x-11, -24, y-1)	<i>bhj</i>	(x-1, -24, y-1)	<i>chj</i>	(x-10, -24, y-1)	<i>dhj</i>	(x, -24, y-1)
<i>ahk</i>	(x-11, -24, y-1)	<i>bhk</i>	(x-1, -24, y-1)	<i>chk</i>	(x-10, -24, y-1)	<i>dhk</i>	(x, -24, y-1)
<i>ahl</i>	(x-11, -24, y)	<i>bhl</i>	(x-1, -24, y)	<i>chl</i>	(x-10, -24, y)	<i>dhl</i>	(x, -24, y)

Theorem 1. *The game will achieve the only subgame perfect Nash equilibrium (SPNE) if all three parties act honestly: sellers upload the matched data and matched key, service providers adopt the authentic data, and matched decryption key, and consumers purchase with sufficient payments. Meanwhile, the SPE is also the optimal strategy for the entire system as a Nash Equilibrium.* □

Proof. First, we dive into the rewards of each role, investigating their payoffs under different strategies. For the seller, we observe that the system is not stable (cannot reach Nash equilibrium) under his optimal strategies. As shown in Table 4, the optimal strategies for sellers (*dei, dej, dek, del, dgi, dgj, dgk, dgl*) is to provide mismatched keys and data, while at the same time obtain payments from consumers. However, based on Lemma 1 and Lemma 2, the seller in such cases cannot obtain payments due to the punishment from smart contracts. These are impractical strategies when launching the backward induction for the subgame tree in Fig. 2. Similarly, for both consumers and service providers, the system is not stable and cannot reach Nash equilibrium under their optimal strategies. Based on that, we find that the optimal strategy for each party is not the optimal strategy for the system.

Then, we focus on strategies with the highest payoffs (equiv. utilities). As illustrated in Table 4 (red background), the strategies of *aei, afi* and *agi* hold the maximal payoffs where $u_{aei} = u_{afi} = u_{agi} = 7$. Their payoffs are greater than all competitive strategies in the history set \mathcal{H} . This means the system reaches Nash equilibrium under these three strategies. However, multiple Nash equilibriums cannot drive the most optimal strategy because some of them are impractical.

We conduct the backward induction for each game with Nash equilibriums. We find that only one of them is the subgame perfect Nash equilibrium with feasibility in the real world. Based on Lemma 3, a consumer without sufficient payments, either to the seller or service provider, cannot successfully decrypt the raw data. He will lose all the paid money $(x+y)$. This means both *afi* and *agi* are impractical. With the previous analyses in the arm, we finally conclude that only the strategy *aei*, in which all parties act honestly, can reach the subgame perfect Nash equilibrium. This strategy is also the Nash equilibrium for the entire BDTS game. \square

6 Implementation and Evaluation

Implementation and Configurations. We provide the detailed implementation of three major functions, including *sharding encryption* that splits a full message into several pieces, *product matching* to show the progress of finding a targeted product, and *payment* that present the ways to pay for each participant. Our full practical implementation is based on Go language with 5 files, realizing the major functions of each contact that can be operated on Hyperledger platform³. We provide implementation details in Appendix A.

Our evaluation operates on Hyperledger Fabric blockchain [15], running on a desk server with Intel(R) Core(TM) i7-7500U CPU@2.70 GHz and 8.00 GB RAM. We simulate each role (*consumer*, *seller* and *service providers*) at three virtual nodes, respectively. These nodes are enclosed inside separated dockers under the Ubuntu 18.04 TLS operating system.

Computational Complexity. Firstly, we provide a theoretical analysis of computational complexity and make comparisons with competitive schemes. We set τ_E , τ_{EA} , τ_D , τ_{DA} , τ_M and τ_V to separately represent the asymmetric encryption time, the symmetric encryption(AES) time, the asymmetric decryption time and the symmetric encryption time, the Merkle tree merging operation time and the Merkle proof verification time. We give our theoretical analysis of each step in Table 5.

Firstly, at the *encrypted data uploading* module, the seller will divide the entire data into several pieces of data and upload their proofs on-chain. We assume the data has been split into n pieces, and every piece of data $Data_i$ needs to be encrypted into D_i . Then, these encrypted data have been stored at the Merkle leaves, merging both $Data_i$ and D_i to obtain M_d and r_{ed} . Secondly, at the *matching and data downloading* module, the consumer can select service providers to download different data segments from them. Before providing the service, the service provider needs to encrypt the received D_i with their private keys, accompanied by corresponding Merkle proofs as in the previous step. Here, the encryption is based on a symmetric encryption algorithm. Once completed, multiple downloads occur at the same time. More service providers will improve the efficiency of downloading because the P2P connection can make full use

³ https://github.com/YXJpYQ/BDTS_Blockchain_based_Data_Trading_System.git.

of network speed. Last, at the *decryption and appealing* module, the consumer obtains each encrypted piece of data and starts to decryption them. They need to verify whether the received data and its proof are matched. If all pass, they can use the valid keys (after payment) for the decryption. Here, the appeal time is related to the number of appeal parts instead of the appeal size.

We further make a comparison, in terms of on-chain costs, with existing blockchain-based fair exchange protocols. Gringotts [26] spends $O(n)$ as they store all the chunks of delivering data on-chain. CacheCas [27] takes the cost at a range of $[O(1), O(n)]$ due to its *lottery tickets* mechanism. FairDwload [8], as they claimed, spends $O(1)$. But they separate the functions of delivering streaming content and download chunks. Our protocol retains these functions without compromising efficiency, which only takes $O(1)$.

Table 5. Computational Complexity and Comparison (i is the number of segmented data; n represents a full chunk of data)

Algorithm		Complexity	Schemes	On-chain Cost
Encrypted data uploading		$i\tau_E + 2\tau_M + 2\tau_V$	Gringotts [26]	$O(n)$
Matching and Data downloading		$i\tau_{E_A} + 2\tau_M + 2\tau_V$	CacheCash [27]	$[O(1), O(n)]$
Encryption and appealing		$i\tau_D + \tau_{D_A} + 2\tau_M + 2\tau_V$	FairDwload [8]	$O(1)$
			<i>BDTS(Ours)</i>	$O(1)$

Efficiency. Then, we launch experimental tests to evaluate efficiency in multi-dimensions. We focus on the *download* functionalities, the most essential function (due to high frequency & large bandwidth) invoked by users.

Data Type. We evaluate three mainstream data types, covering text, image, and video. The text-based file is the most prevailing data format in personal computers. As a standard, a text file contains plain text that can be edited in any word-processing program. The image format encompasses a variety of different subtypes such as TIFF, png, jpeg, and BMP, which are used for multiple scenarios like printing or web graphics (e.g., NFT [28]). We omit subtle differences between each sub-format because they perform equivalently in terms of download services. Similarly, video has a lot of sub-types including MP4, MOV, MP4, WMV, AVI, FLV, etc. We only focus on its general type. From the results in Fig. 3, we can observe that all three types of data have approximately the same performance, under different configurations of data size and storage capacity. The results indicate that *the performance of the download service has no significant relationship with the data type*. This is an intuitive outcome that can be proved by our common sense. The upload/download service merely opens a channel for inside data, regardless of its content and types. This also shows that our BDTS system can support multiple types of data without compromising efficiency.

Data Size. We adjust data sizes at three levels, including 10M, 100M, and 1G, to represent a wide range of applications at each level. As shown in Fig. 3,

10M data (Text, 1 storage) costs at most no more than 2 s, 100M data in the same format spends around 18s, and 1G data costs 170s. The results indicate that *the download time is positively proportional to its data size*. The larger the data, the slower it downloads. This can also apply to different types of data and different storage capacities. A useful inspiration from evaluations of data size is to ensure a small size. This is also a major consideration to explain the reasons for splitting data into pieces in our BDTS. The splitting procedure can significantly improve service quality either for uploading or downloading. Sharded data can be reassembled into its full version once receiving all pieces of segments.

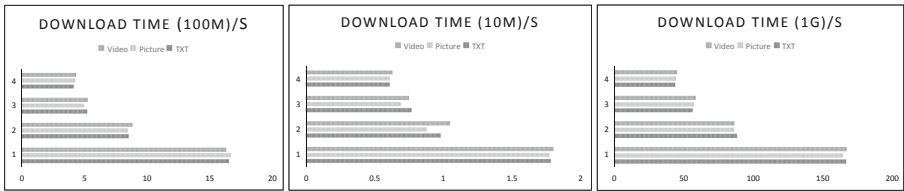


Fig. 3. Download Times of Different Data Type, Data Size and Storage Capacity: We evaluate three types of data formats including video (*grey*), image (*orange*), and text (*blue*). For each type, we test download times in distinguished data size with 10M (*left*), 100M (*middle*) and 1G (*right*). Meanwhile, we also investigate the performance along with increased number of storage devices (*from 1 to 4*), or equiv. the number of service providers.

Storage Capacity. The storage capacity refers to the number of storage devices that can provide download services. The device is a general term that can be a single laptop or a cluster of cloud servers. If each service provider maintains one device, the number of devices is equal to the number of participating service providers. We adjust the storage capacity from 1 device to 4 devices in each data type and data size. All the subfigures (the columns in *left*, *middle* and *right*) in Fig. 3 show the same trend: *increasing the storage capacity over the distributed network will shorten the download time*. The result can apply to all the data types and data sizes. The most obvious change in this series of experiments is adding devices from 1 to 2, which is almost short half of the download time. A reasonable explanation might be that a single-point service is easily affected by other factors such as network connection, bandwidth usage, or propagating latency. Any changes in these factors may greatly influence the download service from users. But once adding another device, the risk of single-point diminishes as the download service becomes decentralized and robust. More connections can drive better availability, as also proved by setting devices to 2, 3 and 4. This is why BDTS allows consumers to download data from multiple providers.

Average Time. We dive into one of the data types to evaluate its i) average download times that are measured in MB/sec by repeating multiple times of experiments under different data sizes; and ii) the trend along with the increased

Table 6. Average Download Time

Storage	Data Size (Text)						Average Time (s)
	1M	10M	50M	100M	500M	1G	
1	0.16	1.78	7.96	16.55	80.52	166.45	0.167
2	0.10	0.98	4.89	8.60	43.48	88.04	0.102
3	0.07	0.77	2.54	5.29	27.44	56.15	0.068
4	0.05	0.61	2.03	4.21	22.22	43.51	0.051
5	0.04	0.38	1.79	3.33	18.88	34.52	0.039
6	0.03	0.32	1.56	2.88	14.69	29.48	0.031

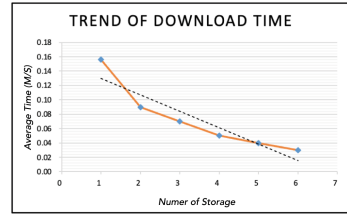


Fig. 4. Download Time

number of storage devices. Compared to previous evaluations, this series of experiments scrutinize the subtle variations under different configurations, figuring a suite of curves. As stated in Table 6, the average downloading times under the storage capacity (from 1 to 6) are respectively 0.167s, 0.102s, 0.068s, 0.051s, 0.039s, and 0.031s. Their changes start to deteriorate, approaching a convex (downward) function as illustrated in Fig. 4. This indicates that the trend of download time is not strictly proportional to the changes in storage capacity. They merely have a positive relation, following a diminishing marginal effect.

Practicability. We further discuss the practicality of the system. We highlight several major features of BDTS by digging into its *usability*, *compatibility*, and *extensibility*.

Usability. Our proposed scheme improves usability in two folds. Firstly, we separately store the raw data and abstract data. The raw data provided by the sellers are stored at the local servers of service providers, while the corresponding abstract data (in the context of this paper, covering *data*, *description* and *proof*) is recorded on-chain. A successful download requires matching both decryption keys and data proofs under the supervision of smart contracts. Secondly, the data trade in our system includes all types of streaming data such as video, audio, and text. These types can cover the most range of existing online resources.

Compatibility. Our solution can be integrated with existing crypto schemes. To avoid repeated payment, simply relying on the index technique is insufficient. The watermarking [29] technique is a practical way to embed a specific piece of mark into data without significantly changing its functionality. It can also incorporate bio-information from users, greatly enhancing security. Beyond that, the storage (encrypted) data can leverage the hierarchical scheme [30] to manage its complicated data, as well as remain the efficiency of fast query.

Extensibility. BDTS can extend functionalities by incorporating off-chain payment techniques (also known as layer-two solutions [31]). Off-chain payment has the advantage of low transaction fees in multiple trades with the same person. Besides, existing off-chain payment solutions have many advanced properties such as privacy-preserving and concurrency [32,33]. Our implementation only set the backbone protocol for fair exchange, leaving many flexible slots to extend functionalities by equipping matured techniques.

7 Conclusion

This paper explores the fairness issue in current data transaction solutions where traditional centralized authorities are not subject to any oversight due to their superpowers. Our proposed scheme, BDTS, addresses such issues by leveraging blockchain technology with well-designed smart contracts. The scheme utilizes automatically operating smart contracts to act in the role of a data executor with transparency and accountability. Our analyses, based on strict game theory induction, prove that the game can achieve a subgame perfect Nash equilibrium with optimal payoffs under the benign actions of all players. Furthermore, we implement the scheme on the Hyperledger Fabric platform and evaluated that the system can provide users with fast and reliable service.

Acknowledgment. Thanks to Lixiaoyang Wang for his contributions to this article. This work was supported by National Key R&D Program of China (2020YFB1005600), National Natural Science Foundation of China (grant no.72192801), Natural Science Foundation of China (U21A20467, 61932011, 61972019), and Beijing Natural Science Foundation (Z220001, M21031) and CCF-Huawei Huyanglin Foundation (CCF-HuaweiBC2021009).

Appendix A. Implementation Details

We give more implementation details by focusing on three major components.

Sharding Encryption. Based on the real scenario, data transmitted in our system is large in scale. A promising way for transferring the data is delivering them in segments (also known as *data sharding*). Data sharding in BDTS does not affect the system consensus or consistency. Instead, data sharding is an off-chain operation conducted by sellers that will be processed before uploading. A full piece of data is split into several shards (or pieces, segments), being encrypted and stored in different memories. The blockchain only reserves its sequential orders and related evidence such as descriptions, addresses and proofs. When a consumer confirms the purchase, he needs to download the data on service providers according to the storage list and obtain the decryption key after successful payment. Then, he can decrypt the data in pieces and finally resemble them according to the sequences for the entire piece of data. We implement the data sharding with the logic in Alg.1. Given the size of a slot (indicating the expected size of a shard), we firstly calculate the number of data segments (*line 4*). Then, a full piece of data is split into n segments (*line 5*). The seller then create its encryption keys (K_1, K_2, \dots, K_n) based on his master private key K_{seller} (*line 6*). Once completed data splitting and key generation, the algorithm starts to encrypt each data segment under the seller's private keys (*line 8–11*). Encrypted data also generate its proofs for further verifications (*line 10*). Last, both raw data and encrypted data are stored on the leaves of Merkle tree to create on-chain roots $MT1$ and $MT2$ (*line 12–14*).

Product Matching. This function describes the process of searching for a targeted source from service providers. In the context of Alg.m.2, the terms *keyword*, *choice*, *ProductList*, *SPList*, *MD*, *Data* and *Desc* represent the searchable keywords, user’s preferences of products, product list, service provider list, data and product description. When a consumer inputs a keyword, the algorithm starts to search for matched ones (line 3) by ranging all descriptions in the product list (line 2–5). Matched products will be recommended to a channel called *showlist* for consumers. The algorithm then inputs *choice* requested from the consumer, and searches related sources (encrypted data, data, description) from service providers (line 8–13). The returned information is sent to the consumer.

Payment. This function mainly describes the method of making payments. The terms in Alg.m.3 *cmAddr*, *slrAddr*, *spAddr* and *Price* stand for the consumer’s address, the seller’s address, the service provider’s address and the price of commodities. The result (either True or False) represents the final result on whether the payment has been successfully executed. The algorithm inputs the addresses of all three entities and the commodity price (line 1). If the token amount of consumer is less than selling prices, the algorithm returns false and the transaction fails (line 1–3). Otherwise, the transaction proceeds. A major difference compared to traditional exchange protocols is that the consumer needs to pay both service providers for their on-chain services and the seller for his resources (line 4–9) (Fig. 5).

Algorithm 1 Sharding Encryption	Algorithm 2 Products Matching	Algorithm 3 Payment
Input: Private Key (seller), Data Output: MD, MT1, MT2 1: function ENCRYPT(K_{seller} , Data) 2: $n \leftarrow 1$ 3: if $slot < Data$ then 4: $n \leftarrow Data / slot$ 5: $Data_1, Data_2, \dots, Data_n \leftarrow Data$ 6: $K_1, K_2, \dots, K_n \leftarrow K_{seller}$ 7: end if 8: for $i = 1$ to n do 9: $D_i \leftarrow AES(K_i)(Data_i)$ 10: $MD \leftarrow D_i$ 11: end for 12: $MT1 \leftarrow Hash(Data_1, Data_2, \dots, Data_n)$ 13: $MT2 \leftarrow Hash(MD_1, MD_2, \dots, MD_n)$ 14: return MD, MT1, MT2 15: end function	Input: keyword, choice, ProductList, SPList Output: MD, Data, Desc 1: function MATCH(keyword, choice, ProductList, SPList) 2: for $i = 1$ to ProductList.length do 3: if ProductList[i].Desc == keyword then 4: ShowList \leftarrow ProductList[i] 5: end if 6: end for 7: ShowList show 8: if choice then 9: Map \leftarrow ProductList[choice].SPList 10: $Data_s \leftarrow$ ProductList[choice].Data 11: $Desc \leftarrow$ ProductList[choice].Desc 12: $D \leftarrow$ Map load 13: end if 14: return MD, Data, Desc 15: end function	Input: cmAddr, slrAddr, spAddr, Price Output: result (True/False) 1: function PAY(cmAddr, slrAddr, Price) 2: if cmAddr.amount < Price then 3: result \leftarrow false 4: else 5: slrAddr \leftarrow cmAddr.transfer(Price) 6: spAddr \leftarrow cmAddr.transfer(Price) 7: result \leftarrow true 8: end if 9: return result 10: end function

Fig. 5. Major Functions

References

1. Wang, C., Wang, Q., et al.: Toward secure and dependable storage services in cloud computing. TSC 5(2), 220–232 (2011)
2. Zhu, Y., Ahn, G.-J., Hongxin, H., et al.: Dynamic audit services for outsourced storages in clouds. TSC 6(2), 227–238 (2011)
3. Kùpçü, A., Lysyanskaya, A.: Usable optimistic fair exchange. In: Pieprzyk, J. (ed.) CT-RSA 2010. LNCS, vol. 5985, pp. 252–267. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-11925-5_18
4. Micali, S.: Simple and fast optimistic protocols for fair electronic exchange. In: PODC, pp. 12–19 (2003)

5. Blum, M.: How to exchange (secret) keys. *TOCS* **1**(2), 175–193 (1983)
6. Pinkas, B.: Fair secure two-party computation. In: Biham, E. (ed.) *EUROCRYPT* 2003. LNCS, vol. 2656, pp. 87–105. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-39200-9_6
7. Dziembowski, S., Ekeley, L., Faust, S.: Fairswap: how to fairly exchange digital goods. In: *CCS*, pp. 967–984 (2018)
8. He, S., Lu, Y., Tang, Q., Wang, G., Wu, C.Q.: Fair peer-to-peer content delivery via blockchain. In: Bertino, E., Shulman, H., Waidner, M. (eds.) *ESORICS* 2021. LNCS, vol. 12972, pp. 348–369. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-88418-5_17
9. Shin, K., et al.: T-chain: a general incentive scheme for cooperative computing. *IEEE/ACM ToN* **25**(4), 2122–2137 (2017)
10. Choudhuri, A.R., Green, M., Jain, A., Kaptchuk, G., Miers, I.: Fairness in an unfair world: fair multiparty computation from public bulletin boards. In: *CCS*, pp. 719–728 (2017)
11. Kiayias, A., Zhou, H.-S., Zikas, V.: Fair and robust multi-party computation using a global transaction ledger. In: Fischlin, M., Coron, J.-S. (eds.) *EUROCRYPT* 2016. LNCS, vol. 9666, pp. 705–734. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-49896-5_25
12. Ekeley, L., Faust, S., Schlosser, B.: Optiswap: fast optimistic fair exchange. In: *AsiaCCS*, pp. 543–557 (2020)
13. Wood, G., et al.: Ethereum: a secure decentralised generalised transaction ledger (2022). <https://ethereum.github.io/yellowpaper/paper.pdf>
14. Li, R., et al.: How do smart contracts benefit security protocols? arXiv preprint [arXiv:2202.08699](https://arxiv.org/abs/2202.08699) (2022)
15. Androulaki, E., et al.: Hyperledger fabric: a distributed operating system for permissioned blockchains. In: *EuroSys*, pp. 1–15 (2018)
16. Chen, J., Xue, Y.: Bootstrapping a blockchain based ecosystem for big data exchange. In: *Bigdata Congress*, pp. 460–463. IEEE (2017)
17. Fang, F., Liu, S., et al.: Introduction to game theory. In: *Game Theory and Machine Learning for Cyber Security*, pp. 21–46 (2021)
18. Moore, J., Repullo, R.: Subgame perfect implementation. *Econometrica: J. Econometric Soc.*, 1191–1220 (1988)
19. Jung, T., et al.: Accounttrade: accountable protocols for big data trading against dishonest consumers. In: *INFOCOM*, pp. 1–9. IEEE (2017)
20. Dai, W., et al.: SDTE: a secure blockchain-based data trading ecosystem. *IEEE Trans. Inf. Forensics Secur. (TIFS)* **15**, 725–737 (2019)
21. Li, R., Wang, Q., Liu, F., Wang, Q., Galindo, D.: An accountable decryption system based on privacy-preserving smart contracts. In: Susilo, W., Deng, R.H., Guo, F., Li, Y., Intan, R. (eds.) *ISC 2020*. LNCS, vol. 12472, pp. 372–390. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-62974-8_21
22. Zhou, J., et al.: Distributed data vending on blockchain. In: *IEEE International Conference on Internet of Things (iThings)*, pp. 1100–1107. IEEE (2018)
23. Janin, S., Qin, K., Mamagishvili, A., Gervais, A.: Filebounty: fair data exchange. In: *EuroS&PW*, pp. 357–366. IEEE (2020)
24. Kiayias, A., Koutsoupias, E., Kyropoulou, M., Tselekounis, Y.: Blockchain mining games. In: *ACM EC*, pp. 365–382 (2016)
25. Garay, J., Kiayias, A., Leonardos, N.: The bitcoin backbone protocol: analysis and applications. In: Oswald, E., Fischlin, M. (eds.) *EUROCRYPT* 2015. LNCS, vol. 9057, pp. 281–310. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-46803-6_10

26. Goyal, P., et al.: Secure incentivization for decentralized content delivery. In: USENIX Workshop on Hot Topics in Edge Computing (HotEdge) (2019)
27. Almashaqbeh, G.: CacheCash: A Cryptocurrency-based Decentralized Content Delivery Network. Columbia University (2019)
28. Wang, Q., Li, R., Wang, Q., Chen, S.: Non-fungible token (NFT): Overview, evaluation, opportunities and challenges. arXiv preprint [arXiv:2105.07447](https://arxiv.org/abs/2105.07447) (2021)
29. Yang, R., Au, M.H., Yu, Z., Xu, Q.: Collusion resistant watermarkable PRFs from standard assumptions. In: Micciancio, D., Ristenpart, T. (eds.) CRYPTO 2020. LNCS, vol. 12170, pp. 590–620. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-56784-2_20
30. Gentry, C., Silverberg, A.: Hierarchical ID-based cryptography. In: Zheng, Y. (ed.) ASIACRYPT 2002. LNCS, vol. 2501, pp. 548–566. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-36178-2_34
31. Gudgeon, L., Moreno-Sanchez, P., Roos, S., McCorry, P., Gervais, A.: SoK: layer-two blockchain protocols. In: Boneau, J., Heninger, N. (eds.) FC 2020. LNCS, vol. 12059, pp. 201–226. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-51280-4_12
32. Malavolta, G., Moreno-Sanchez, P., Kate, A., et al.: Concurrency and privacy with payment-channel networks. In: CCS, pp. 455–471 (2017)
33. Green, M., Miers, I.: Bolt: anonymous payment channels for decentralized currencies. In: CCS, pp. 473–489 (2017)