

A Survey of Blockchain-based Schemes for Data Sharing and Exchange

Rui SONG, Bin XIAO, Yubo SONG, Songtao GUO, and Yuanyuan YANG

Oct. 2023



Opening Minds • Shaping the Future • 啟迪思維 • 成就未來

Contents

- Introduction
- Blockchain-based Data Sharing
- Blockchain-based Data Exchange and Marketplace
- Challenges
- Conclusion

Introduction

Background

- **Data Sharing and Exchange**
 - Needs from *enterprises, governments, and academia*
 - *Difficult* for entities to meet data requirements relying on their own datasets
 - *Data sharing* and *exchange* have become huge industries
- **Definitions**
 - Data sharing: *sharing* digital assets between *different entities*
 - Data exchange: *trading* digital assets between *data owners* and *demanders*

Introduction

Traditional Schemes

- Centralized sharing and exchange platforms
 - Transactions depend on the *honesty* of *centralized service providers*
 - *Attacks* against platforms may cause *data breaches* and *property damage*
 - *High transaction fee* for data services

Introduction

Blockchain-based Schemes

- Blockchain-based schemes
 - No *reliance* on *authorities* or *trusted parties*
 - *Auditable* data transactions
 - *Protection* against *data tampering*
- Challenges
 - Low *transaction frequency*
 - High data *storage* and *communication costs*
 - Data *privacy* and *security concerns*

Introduction

Previous Work

- **Blockchain and Data Industries**
 - Deepa et al.¹ investigated the *applications of blockchain* in the *collection, storage, analysis*, and *privacy protection* of big data in different domains.
 - Xie et al.² investigated *blockchain-based schemes* of *cloud data storage* and *exchange marketplaces*.
 - Berdik et al.³ Investigated solutions for *interoperability, system efficiency, fault tolerance*, and *data integrity* in *blockchain-based data-sharing schemes*.

1. N. Deepa et al., "A survey on blockchain for big data: Approaches, opportunities, and future directions," arXiv preprint arXiv:2009.00858, 2020.

2. S. Xie et al, "Blockchain for cloud exchange: A survey," Computers & Electrical Engineering, vol. 81, p. 106526, 2020.

3. D. Berdik et al, "A survey on blockchain for information systems management and security," Information Processing & Management, vol. 58, no. 1, p. 102397, 2021.

Introduction

Contributions

- A survey of applications and schemes of blockchain in data sharing and exchange
 - **Compile and summarize highly cited papers** on blockchain-based data sharing and exchange schemes since 2015.
 - **Summarize system models** of blockchain-based systems, including **entities, objectives, and security properties**.
 - **Analyze** blockchain-based data-sharing schemes, focusing on their **architectures, design principles, and issues on data privacy and security**.
 - **Analyze** blockchain-based data marketplaces and exchange platforms, focusing on the **exchange process, monetization, and copyright protection**.

Data Sharing Schemes

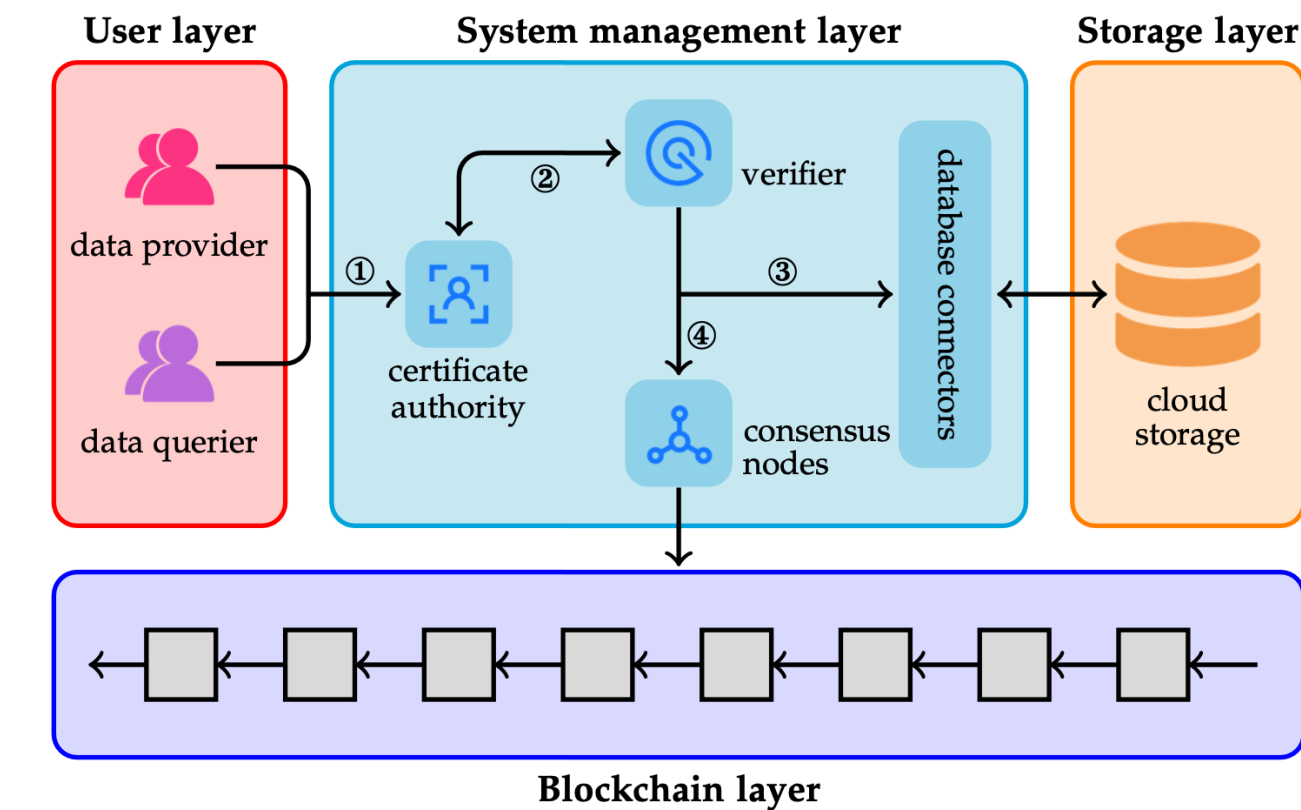
Comparison of blockchain-based data sharing schemes

Author(s)	Ref.	Year	Application Scenario	Architecture		Storage		
				Private/Consortium	Public/Contract	Private	Cloud	Local
Zyskind et al.	[35]	2015	Personal Data	✓			✓	
Peterson et al.	[52]	2016	Healthcare	✓		✓		
Azaria et al.	[53]	2016	Healthcare		✓	✓		✓
Xia et al.	[46], [54]	2017	EMR Sharing	✓			✓	
Shafagh et al.	[44]	2017	IoT	✓			✓	
Dorri et al.	[45]	2017	Smart Vehicle	✓				✓
Gordon et al.	[49]	2018	Healthcare	✓				✓
Zhang et al.	[42]	2018	Healthcare	✓		✓		
Chowdhury et al.	[55]	2018	Personal Data	✓		✓		
Cui et al.	[43]	2018	File Sharing	✓			✓	
Ali et al.	[47]	2018	PingER	✓		✓		
Zhang et al.	[56]	2018	Healthcare		✓			✓
Kang et al.	[57]	2018	Smart Vehicle	✓	✓	✓		
Theodouli et al.	[58]	2018	Healthcare		✓	✓		✓
Sultana et al.	[59]	2020	IoT		✓			✓
Xiao et al.	[60]	2020	Personal Data		✓		✓	✓
Feng et al.	[61]	2021	5G Drones	✓			✓	
Yu et al.	[62]	2021	IoT	✓				✓
Tan et al.	[63]	2021	COVID-19 EMR	✓				✓
Li et al.	[64]	2022	IoT	✓				✓

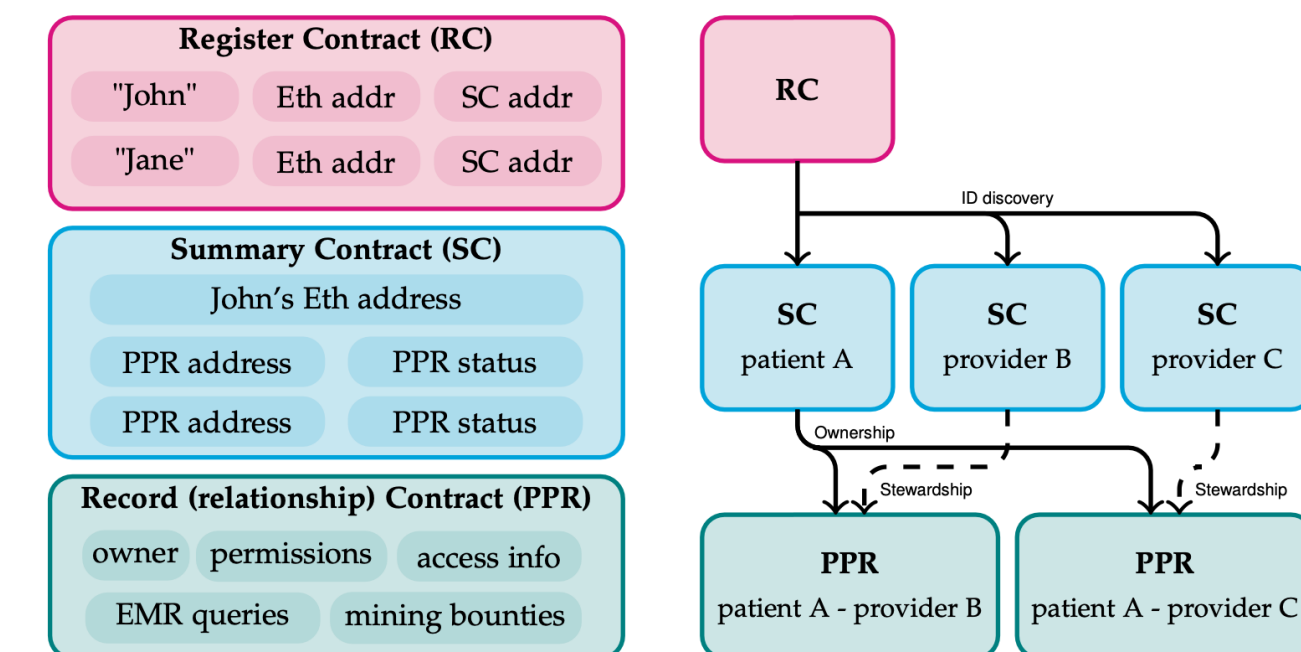
Data Sharing Schemes

System Architectures

- Underlying blockchain architectures
 - Permissioned blockchains
 - *Hyperledger fabric* or *self-built blockchains*
 - **Metadata** on blockchain, **data contents** on cloud/
distributed storage
 - Authorities for **access control** and **authentication**
 - Permissionless blockchains
 - Ethereum **smart contracts**
 - Contracts for **access control** and **operation recording**



Typical framework of data sharing systems based on permissioned blockchains.



The architecture of a smart contract-based data sharing system.^{1,2}

1. A. Azaria et al, "Medrec: Using blockchain for medical data access and permission management," in IEEE OBD. IEEE, 2016, pp. 25–30.

2. A. Ekblaw et al, "A case study for blockchain in healthcare: "MedRec" prototype for electronic health records and medical research data," in Proceedings of IEEE Open & Big Data Conference, vol. 13, 2016, p. 13.

Data Sharing Schemes

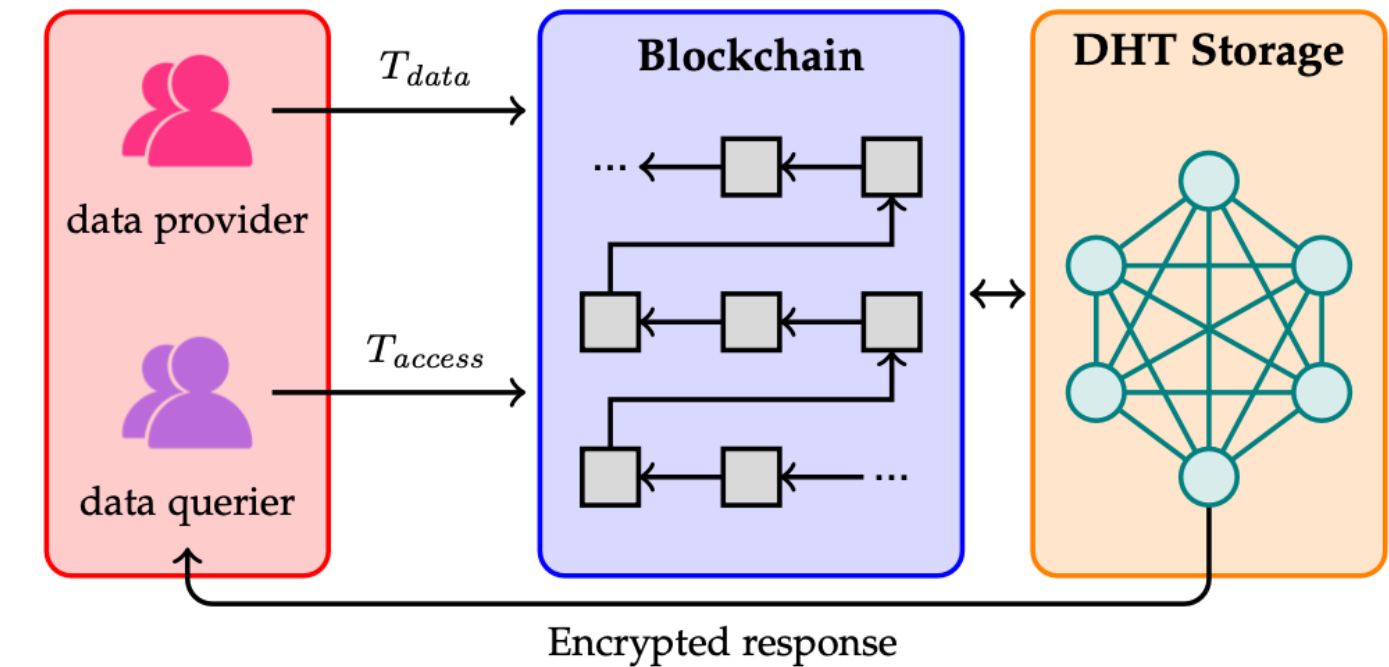
Access Control

- **Blockchain-based access control**

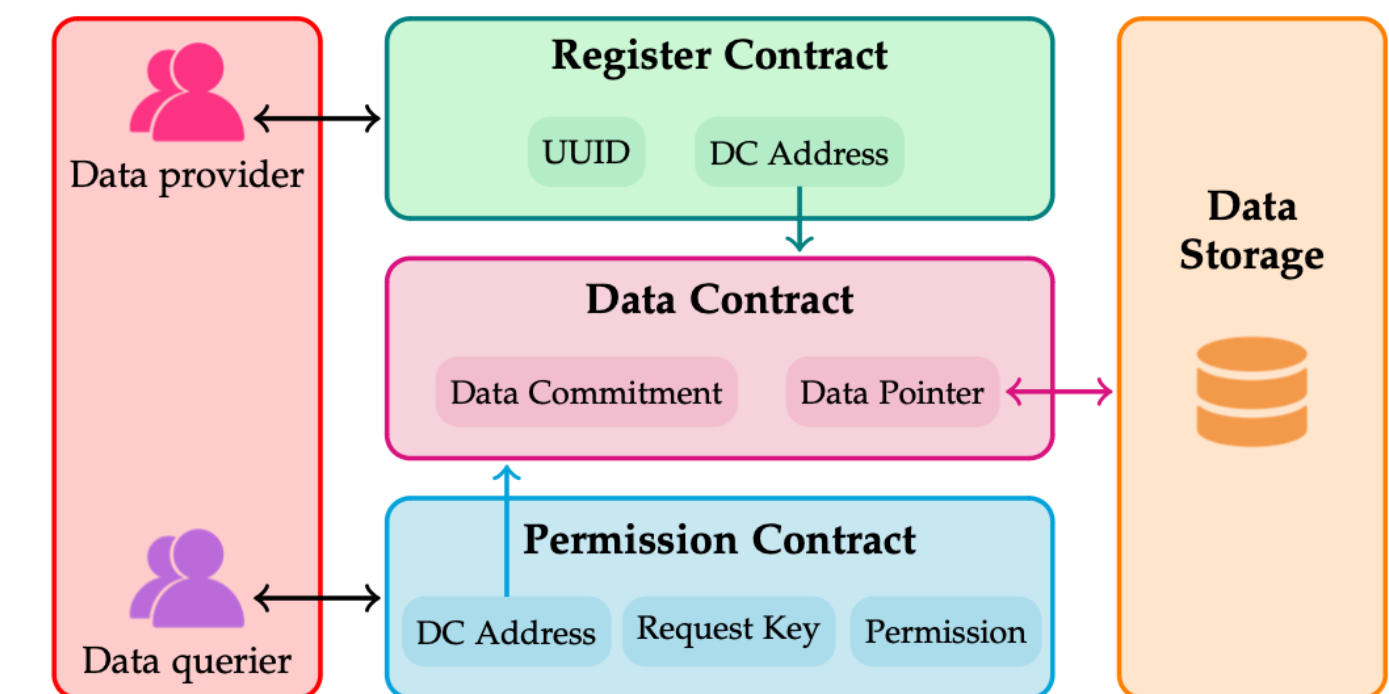
1. **Blockchain transactions** to define access permissions
2. **Smart contracts** to grant access permissions
3. **Token-based** access control

- **Advantages**

- **Tracking** and **auditing** of data access requests
- **Fine-grained control** of specific data assets and users
- **Revoking** or **modifying** granted access permissions



Transaction-based data access control and permission management.¹



Smart contract-based data access control and permission management.²

1. G. Zyskind et al, "Decentralizing privacy: Using blockchain to protect personal data," in 2015 IEEE Security and Privacy Workshops. IEEE, 2015, pp. 180–184.

2. A. Azaria et al, "Medrec: Using blockchain for medical data access and permission management," in IEEE OBD. IEEE, 2016, pp. 25–30.

Data Sharing Schemes

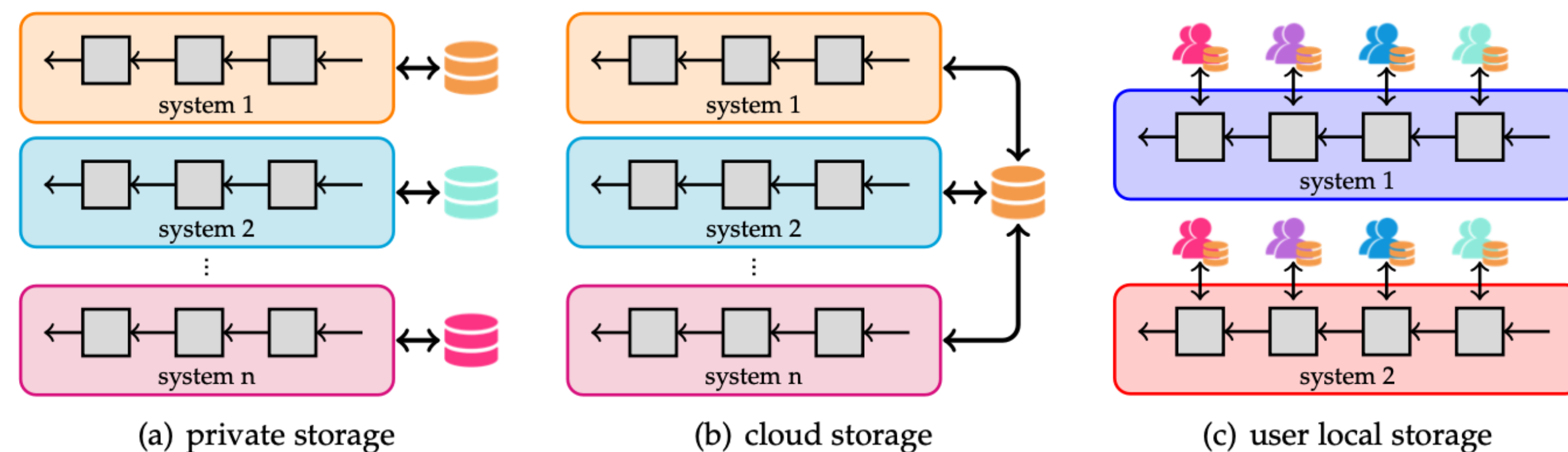
Interoperability

- Collaboration between data entities
 - **Syntactic interoperability**: data formats, protocols, programming interfaces
 - **Semantic interoperability**: data availability and validity
- Scenarios
 - Healthcare
 - IoT
 - Smart vehicle
 - ...

Data Sharing Schemes

Data Storage

- **Storage infrastructure for data-sharing systems**
 - **Private storage**: databases maintained by sharing platforms
 - **Local storage**: users store data assets locally and expose interfaces to sharing systems
 - **Cloud storage**: cloud services for data storage (or **decentralized storage** e.g., IPFS or Filecoin)

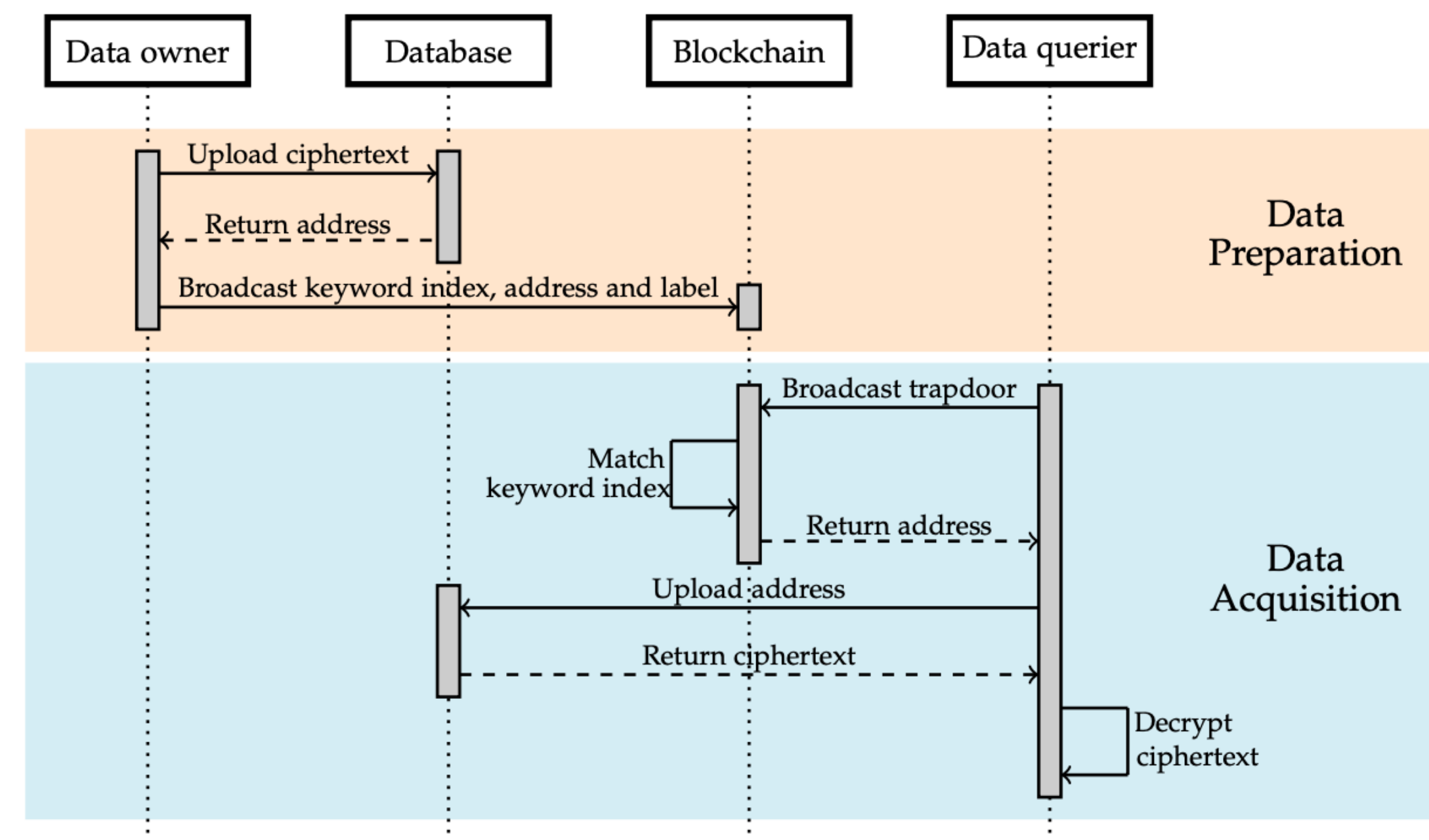


Data storage schemes commonly used in blockchain-based data-sharing systems.

Data Sharing Schemes

Data Privacy

- Privacy-preserving data sharing
 - *Identity-based encryption (IBE)*
 - *Attribute-based encryption (ABE)*
- Data querying and utilization
 - *Searchable encryption*
 - *Homomorphic encryption*
 - *Zero-knowledge proofs*



The sequential diagram of data querying in a trusted data sharing scheme.¹

1. X. Ma et al, "The data sharing scheme based on blockchain," in Proceedings of the 2nd ACM International Symposium on Blockchain and Secure Critical Infrastructure, 2020, pp. 96–105.

Data Exchange Schemes

Comparison of blockchain-based data exchange schemes and marketplaces

Author(s)	Ref.	Year	Application Scenario	Technical Details				Marketplace Design			
				Access Control	Privacy	Fairness	Integrity	Data Quality	Copyright	Monetization	Data Service
Chen et al.	[100]	2017	Big Data		✓				✓		
Nasonov. et al.	[101]	2018	Big Data	✓	✓		✓	✓			✓
Ozyilmaz et al.	[102]	2018	IoT							✓	✓
Yang et al.	[103]	2018	Smart City Edge Computing		✓			✓			
Hynes et al.	[104]	2018	Healthcare		✓					✓	✓
Ramachandran et al.	[41]	2018	Smart City	✓	✓			✓		✓	
Park et al.	[105]	2018	IoT				✓	✓			
Travizano et al.	[106]	2018	Big Data		✓					✓	
Savelyev et al.	[40]	2018	Digital Assets						✓	✓	
Xiong et al.	[107]	2019	Machine Learning	✓						✓	
Bajoudah et al.	[108]	2019	IoT			✓					
Banerjee et al.	[109]	2019	Big Data		✓	✓					
Chen et al.	[110]	2019	Big Data		✓	✓					
Liu et al.	[111]	2019	IoT							✓	
Dai et al.	[112]	2020	Big Data	✓							✓
Hu et al.	[113]	2021	Big Data							✓	✓
Abdellatif et al.	[114]	2021	Edge Computing Healthcare		✓						✓
Liu et al.	[115]	2022	IoT	✓	✓			✓			

Data Exchange Schemes

Marketplaces

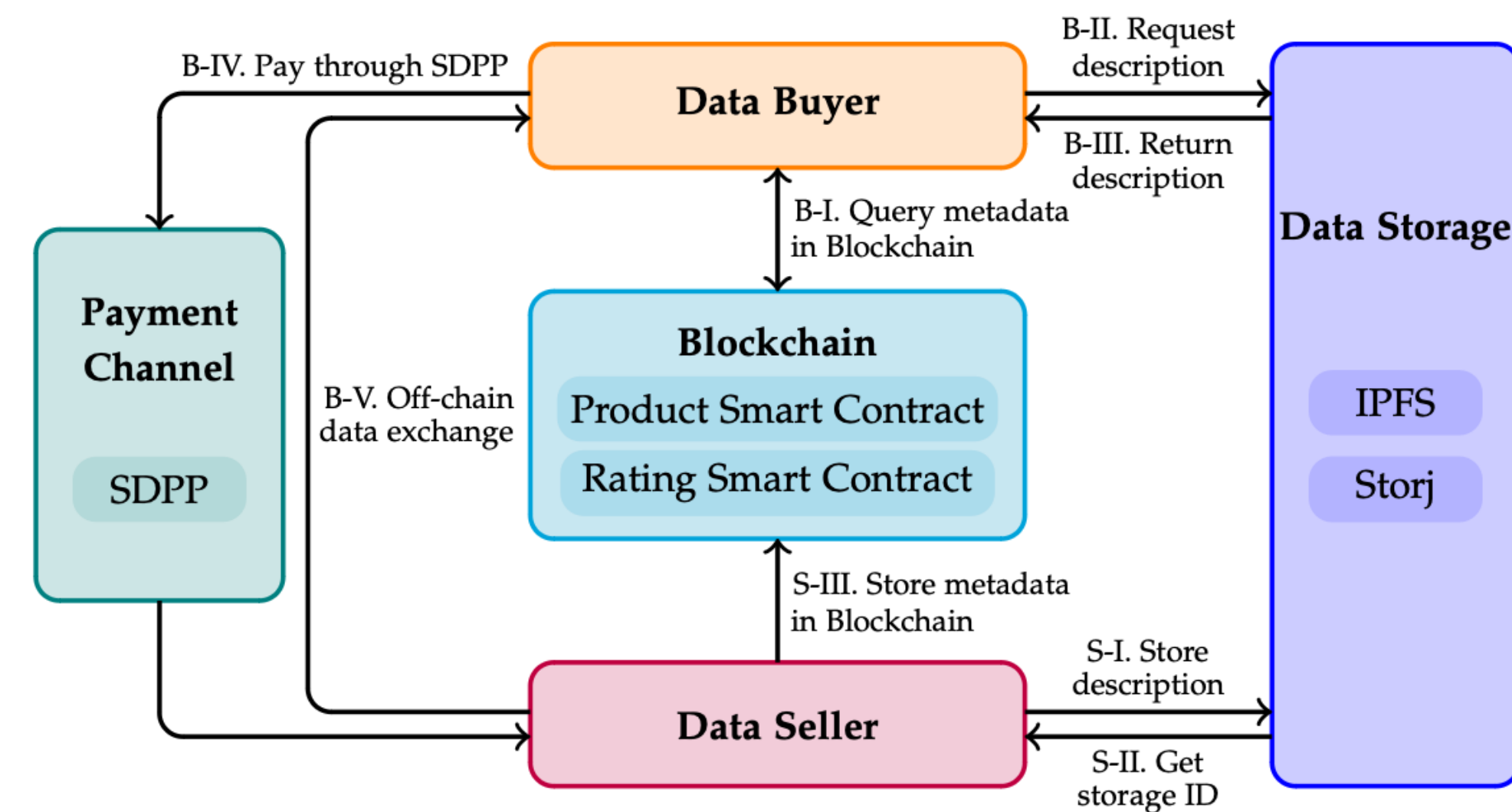
- Marketplace for data exchange and trading
 - *Intermediaries* for data transactions, but not possess data commodities, or
 - *Fully control* all data assets submitted by sellers
 - *Services* of statistics, cleaning, filtering, searching, monetization, etc.

Data Exchange Schemes

Exchange Process

- **System flow of data marketplaces**

- Data submission
- Commodity releasing
- Supply and demand matching
- Commodity Pricing
- Data Services
- Purchase and payment



System flow of a typical blockchain-based data marketplace.¹

1. G. S. Ramachandran et al, "Towards a decentralized data marketplace for smart cities," in 2018 IEEE International Smart Cities Conference (ISC2). IEEE 2018, pp. 1-8.

Data Exchange Schemes

Monetization

- Payment methods
 - *Cryptocurrencies* or *tokens*
 - *Payment channel* for high-frequency micro-transactions
 - Streaming data payment protocol (SDPP)
- Commodity pricing
 - *Seller pricing* (buyers rate sellers based on historical behaviors)
 - *Auction*
 - *AI-assisted pricing*
 - *Game-based pricing*

Data Exchange Schemes

Copyright

- **Copy control and data traceability**
 - *Binding* data assets to *metadata* (commitments) on blockchains
 - Generating a *unique identifier* for all data assets
 - Tracing illegal copiers by data *identifiers*
 - Traceability of data assets using *zero-knowledge proofs*
- **Strict control of original data assets**
 - Providing data *computation services* instead of raw data assets
 - Providing *programming interfaces*

Data Exchange Schemes

Existing commercial blockchain-based data marketplaces

Platform	Release	Blockchain	Token	Data storage	Remarks
Datum	2017	Ethereum	DAT	BigChainDB/IPFS	incentives mechanism
GXChain	2017	GXChain	GXS & GXC	IPFS/BaaS	WebAssembly execution MPC and HE
Databroker	2018	Ethereum	DTX	local & private storage	data matching
Datapace	2017	Hyperledger Fabric	TAS	local & private storage	Maniflux-based platform for sensor connection
IOTA	2016	Tangle	IOTA	Tangle	DAG for parallelized transactions
Streamer	2019	Ethereum	DATA	local storage & DHT	streaming data exchange publish/subscribe mode

Challenges

Platform Designs

- Insufficient discussion on platform designs
 - Most domains **lack** *unified data templates, interfaces, and interoperability standards*.
 - Data **privacy requirement** **contradicts** data querying, indexing, and supply-demand matching.
 - Lack of clear **motivation** and **incentives** for sellers
 - Need to penalize **malicious behaviors**
- Copyright and data traceability
 - Modifications to data assets could **invalidate tracking** and **auditing** techniques
 - Difficult to **detect** and **prevent** data **duplication** and **redistribution**
 - Difficult for buyers to **evaluate the value** of data commodities based on the contents

Challenges

Exchange Fairness

- **Fair data exchange issues**

- Providing buyers with validation methods could *reveal the privacy* of data commodities
- Predicates can only prove the correctness or validity of data, but not that it *meets buyers' needs*.
- Difficult to guarantee the *correctness and validity* of the predicates.

- **Performance issues**

- ZK-SNARKs-based schemes are *less efficient in proof generation* and support only *small data chunks*.
- Merkle proof-based schemes have a *big proof size* and a *long verification time*.

Conclusion

- Blockchain provides effective solutions for data sharing and exchange processes.
- This paper reviews **designs and schemes** of blockchain-based sharing and exchange systems:
 - System architecture and flow
 - Access control, authentication, and authorization
 - Data interoperability
 - Payment and monetization
- Comprehensive **knowledge** of applications for enterprises and researchers
- **Challenges** and **future research directions**

Thank you!



Opening Minds • Shaping the Future • 啟迪思維 • 成就未來



Department of Computing
電子計算學系