Research Progress of Heterogeneous Blockchain Data Migration Technology

Yutong Xie[1,2,*][a], Shuai Chen[1,2][b] and Xinnan Wang[2,3][c]
[1]*State Grid Digital Technology Holding Co., Ltd., Beijing China*
[2]*State Grid Blockchain Technology (Beijing) Co., Ltd, Beijing, China*
[3]*Blockchain Technology Laboratory of State Grid Corporation of China, Beijing, 100053, China*

Keywords: Blockchain, Blockchain Data Migration, Heterogeneous Blockchain.

Abstract: In recent years, with the rapid development of blockchain technology, blockchain underlying architecture, algorithm and other technologies are constantly upgraded. The upgrade of technology drives the innovation of business. The old blockchain cannot meet the demand of new business, so it is necessary to carry out application update and blockchain data migration. However, due to the immutability and other characteristics of blockchain technology, it is very difficult to migrate data on the blockchain. At present, due to the lack of effective migration mechanism, there are a series of problems in the process of blockchain data migration, such as inconsistent structure of heterogeneous blockchain blocks and transaction data, modified proof of existence of data, and difficult migration of smart contracts. Therefore, the research of heterogeneous blockchain data migration technology is very important. This paper analyzes the current domestic and foreign researches on the unified abstraction of heterogeneous blockchain data structure, transformation of smart contract, proof of the existence of data and other technologies, and focuses on the basic problems such as cryptography algorithm, smart contract and blockchain data structure, and puts forward the technology of transferring heterogeneous blockchain block and transaction data, smart contract code and status data, so as to provides ideas for solving the problem of heterogeneous blockchain data migration and promote the further development of blockchain technology.

1 INTRODUCTION

In the development process of blockchain, due to the continuous emergence of new technologies, and there is no standard way to implement blockchain, both the public chain and the alliance chain have appeared the phenomenon of contention. With the evolution and replacement of technology, it is inevitable to upgrade, replace and migrate the original blockchain system. In this process, the data of the original blockchain application needs to be secure and available.

However, due to the characteristics of blockchain technology, data migration on the blockchain is very difficult. Heterogeneous blockchain data migration has seriously threatened the sustainable and healthy development of blockchain technology. At present, due to the lack of effective migration mechanism, The following problems exist during data migration.

- Heterogeneous blockchain block, transaction and other data structure is inconsistent, the data structure is difficult to convert.
- The existence proof of the original data is modified and the authority of the data is destroyed.
- Heterogeneous blockchain smart contract migration is difficult and it is difficult to verify the function consistency of smart contract.
- Heterogeneous blockchains provide incompatible interfaces.

With the practice and development of blockchain technology, State Grid Corporation of China has put forward higher requirements for the promotion and application of blockchain prospective technology, further deepening the research on blockchain

[a] https://orcid.org/0000-0002-8003-4176
[b] https://orcid.org/0000-0003-2377-4000
[c] https://orcid.org/0000-0002-9246-4819

Xie, Y., Chen, S. and Wang, X.
Research Progress of Heterogeneous Blockchain Data Migration Technology.
DOI: 10.5220/0012036600003620
In *Proceedings of the 4th International Conference on Economic Management and Model Engineering (ICEMME 2022)*, pages 548-554
ISBN: 978-989-758-636-1

technology, improving the independent mastery of core technology, and strengthening the application in more businesses of State Grid Corporation of China, such as energy trading (Zhang, Hou 2021), energy metering, etc.

On the basis of investigation and research, this paper puts forward the independent controllable heterogeneous blockchain data migration technology, which combined with proof mechanism of cryptography and data migration technology, solving the data migration problems that may occur in the process of heterogeneous blockchain upgrade and replacement, realizing the high availability of data in the process of blockchain upgrade, replacement and data migration, and improving the compatibility and collaboration between blockchain systems.

2 BLOCKCHAIN OVERVIEW

Blockchain originates from Bitcoin (Satoshi Nakamoto 2008) and is the underlying technology of Bitcoin. Bitcoin is the earliest application of blockchain technology. Blockchain is a distributed ledger that uses cryptography to append blocks confirmed by consensus in sequence. As the name implies, in terms of data structure, blockchain is a blockchain structure that uses hash Pointers instead of traditional Pointers. Each block contains a block header and a block body, the block header contains the version number, the hash value of the previous block, the root hash of the Merkle tree of the block, the timestamp of the generated block, the difficulty value, and the random number. The block body contains the transactions recorded in the block. Blockchain relies heavily on encryption algorithms, peer-to-peer communication technologies, and innovates the smart contracts (Vitalik Buterin 2013) implemention. The organic combination of various technologies gives blockchain features such as decentralization, immutability and traceability. Blockchain is divided into public chain, alliance chain and private chain, with the degree of openness decreasing successively. At present, alliance chain has a wide range of landing scenarios in China.

As a brand-new information storage, dissemination and management mechanism, blockchain technology has attracted great attention in various fields. In recent years, with its potential value and favorable policies, the blockchain industry has ushered in the best opportunity for industrial development, and the blockchain technology has also ushered in continuous development and innovation. By May 2019, more than 30 provinces and regions in China have issued policy guidance documents and carried out the layout of blockchain industrial chain, combining blockchain technology with local characteristics, which playing a positive role in serving economic and social development. Blockchain, as an emerging technology, has come into public view and become the focus of social attention.

With the development and in-depth application of smart contract technology, blockchain, as a ledger, has a strong programmable ability, which has broadened the original simple transaction function, and started to realize more complex functions such as complex conditional payment, business logic, automatic execution of scripts, multi-party agreements that conform to legal relations.

In China, enterprise application is the main battlefield of blockchain, and alliance chain are widely used. In the coming period of time, blockchain applications will be used to reduce costs, improve collaboration efficiency, and stimulate real economic growth. Different from the public chain, in enterprise applications, people pay more attention to the control, regulatory compliance, performance, security and other factors of blockchain.

3 RESEARCH STATUS OF HETEROGENEOUS BLOCKCHAIN DATA MIGRATION

At present, there are three kinds of technologies in the field of heterogeneous link data migration. First, how to unify and abstract the data structure of blockchain. It is necessary to study the differences in data structures of different blockchains, and conduct unified abstraction of data structures to build a unified data interface layer for the convenience of management of the business layer, so as to solve the compatibility problem of business layer caused by different data structures of heterogeneous blockchains in the process of data migration. Second, how to complete the migration of heterogeneous blockchain smart contract. It is necessary to study the current smart contract technology of heterogeneous blockchain, and explore how to achieve smart contract migration. The third is the difference of existence proof of different heterogeneous blockchain data.

3.1 Unified Abstraction Approach for Heterogeneous Blockchain Data Structures

In terms of the unified abstraction of heterogeneous blockchain data structure, there are many cross-chain protocols. Cross-chain protocol (Li, Qiu, Xu, Song, Liu 2021) is usually a trusted source oriented interoperation protocol, aiming to build a set of flexible, unified, reliable interoperation protocol, to achieve convenient access and reliable operation to different trusted sources. Trusted sources refer to software, hardware, or other types of entities that can provide trusted data. Trusted sources can be distributed or centralized. Common trusted sources include blockchain, oracle machine and so on.

The unified abstraction protocol in cross-chain protocol includes four protocols that define the unified abstraction among trusted sources, account services, cross-chain routes, and applications. As shown in figure 1.

- Unified Account Protocol: A unified abstraction of various trusted sources accounts, enabling the operation of different trusted sources with a unified account.
- Unified Addressing protocol: A unified abstraction of various trusted sources smart contracts (chain codes) and other operable objects to achieve unified addressing with the concept of "resources".
- Unified Invocation Protocol: A unified abstraction of the various trusted sources invocation protocols to invoke resources with a unified interface and parameters.
- Unified Access Protocol: An abstraction of various trusted source access protocols to implement unified access adaptation of different trusted sources. Different trusted sources develop plug-ins based on this protocol to achieve adaptation and access.

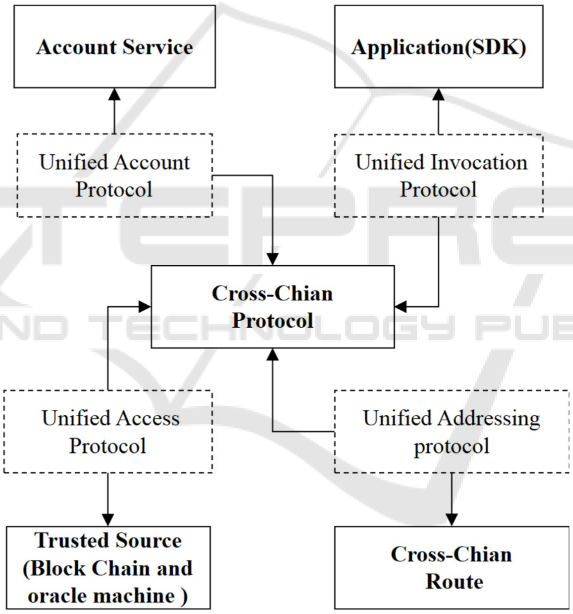

Figure 1: The unified abstraction protocol.

3.2 Transformation of Heterogeneous Blockchain Smart Contract

Smart contract can be regarded as a computer program running on the blockchain with preset rules, with status and conditional response, which can encapsulate, verify and execute complex behaviors of distributed nodes, and realize information exchange, value transformation and asset management. The core component of smart contract is smart contract programming language and its execution engine. The smart contract of blockchain needs to run in a completely isolated environment to ensure effective resource isolation between contracts, and between the contract and the host system, so as to ensure its controllable security. Blockchain virtual machines are one of the ways to execute blockchain smart contracts. Different smart contracts have different levels of security and richness of expression.

At present, the research of smart contract code transformation at home and abroad is still in an early stage. There are two main reasons: new technologies keep emerging, and the original contract language is also changing. For example, Solidity itself, the Ethereum smart contracts language, comes out with new versions and changes every once in a while. Now the main smart contract transformation is still in two categories. One is the transformation of Solidity language into WASM (Yang, Liu, Li, Zheng, Wang, Chen 2020) intermediate code, such as foreign Yul, which is an intermediate language only for Ethereum. Solidity compiler compiles Solidity source code into an intermediate representation of Yul and then converts it into EVM instructions. It is finally converted to bytecode. In addition, RuneEVM makes the EVM interpreter compatible with the interface of WASM, so that the EVM contract can be run on WASM. But both projects are currently inactive. The other is to compile EVM code into RISC-V instructions. Currently, some teams using RISC-V (Waterman, Lee, Patterson 2014) as virtual machines are doing related research, such as Nervos blockchain in China, but it is still in an early state.

3.3 Proof of Existence of Heterogeneous Blockchain Data

At present, the mainstream blockchain mainly proves the existence of Merkle Trees, but in a few blockchains, such as Fabric, there may be no relevant cryptography proof, which requires corresponding measures during the processing of data migration. In Fabric, every transaction needs to meet some predefined endorsement policy. When a transaction is executed, it will be signed by multiple endorsement nodes. When the signatures of all parties meet the endorsement policy, the transaction will be considered valid. Fabric stores the endorsement node signature information in the block as part of the transaction. Multiple transactions make up a list of transactions within a block. The transaction list computes a hash value in binary form, which is recorded in the block header. WeCross cross-chain protocol of WeBank in China has been studied in this area. As shown in figure 2.

- Block continuity verification: FISCO BCOS verifies that this block is the block of the other bolckchain by comparing the parent block hash in the block header with the real parent block hash.
- Block consensus verification: By checking the signature list of the current block, we can judge whether the number of legitimate signatures meets the PBFT consensus condition, and confirm that the current block represents the overall will of the other blockchain.
- Transaction existence verification: By verifying that the Merkle Path of the transaction hash to the transaction root is correct, we can know whether the transaction already exists on the blockchain.
- Transaction correctness verification: By verifying the correspondence among business expectation, transaction binary, and transaction hash, we can know whether transaction is the operation expected by the business.

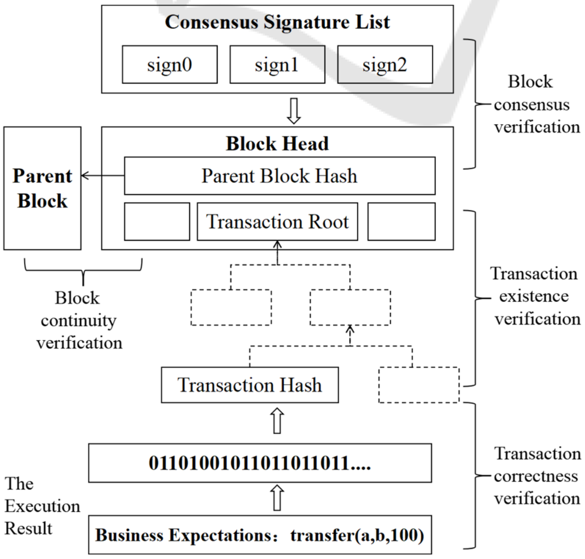
Consensus Signature List
sign0
sign1
sign2
Block consensus verification
Block Head
Parent Block
Parent Block Hash
Transaction Root
Block continuity verification
Transaction existence verification
Transaction Hash
01101001011011011011....
Transaction correctness verification
The Execution Result
Business Expectations：transfer(a,b,100)

Figure 2: The verifications of existence.

4 APPORACH OF HETEROGENEOUS BLOCKCHAIN DATA MIGRATION

For blockchain, there are mainly two types of data, block and transaction data, and contract code and contract status data. The block contains the signature of the consensus node and the transaction contains the signature of the user, both of which are the basis of the blockchain's history. The contract and contract status are calculated by the node based on the transaction records.

This section proposes a heterogeneous blockchain data migration method. The blocks and transactions of the original blockchain are extracted and loaded into the traditional database. Each consensus node stores the full block and transaction data respectively, and queries and verifies the block and transaction data through the consensus node of the original blockchain. After ensuring data consistency, the hash of the last block of the source blockchain is stored in the new blockchain genesis block, and the height of the block is added on the basis of the height of the last block of the source chain, so as to ensure the continuity of the block and reach a consensus on the signature of the Genesis block file. When the height is less than the height of the Genesis block, search the target block from the block database.

4.1 Migration of Blockchain and Transaction Data

For block and transaction data, the transaction contains the signature of the transaction sender, and the block contains the signature and time stamp of the consensus node, which determines that the transaction and block data are the unchangeable basis of the blockchain. If the data of the original transaction and block are reconstructed and migrated to the new blockchain, the invariance characteristics of the blockchain will be inevitably modified. Therefore, this block and transaction can be synchronized to a separate database for storage, which includs structured database, KV, file database.

At the same time, in the generation process of the new blockchain genesis block, the hash of the previous block of genesis block is the hash of the last consensus block of the original blockchain, and the height of Genesis block is increased by one on the basis of the original blockchain height. In this way, the new blockchain is connected with the old blockchain, and the old blockchain data is not falsified.

The historical height of block chain can be read directly from the database, and the new height can be obtained from the new block chain. Users can obtain signatures in blocks, signatures in transactions, and generate MPT proofs using block and transaction data to prove that data has not been falsified. At the same time, the replay of historical transactions can be used to generate the historical state, and the authenticity of the historical state can be verified by the state root and other related fields in the block. As shown in Figure 3.

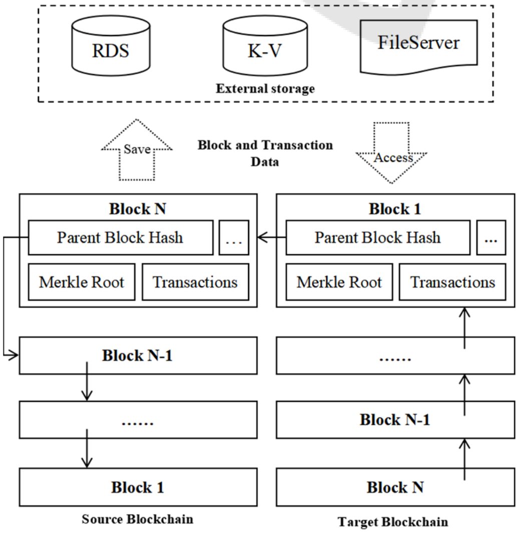

Figure 3: The technical principles of migration of blockchain and transaction data.

4.2 Migration of Smart Contract Code

For smart contract code, if the original blockchain and target blockchain use the same virtual machine, the smart contract code can be directly written into the initial file of the genesis block of the new blockchain and given the contract its original address, so that the original contract exists in the new blockchain when the new blockchain is started.

For target blockchain using different virtual machines, it is necessary to realize the translation of contract languages, which can be divided into manual translation and automatic translation. For manual translation, the implementation personnel need to manually write code according to the original contract logic. For automatic translation, it is necessary to use existing tools. For example, solidity2wasm already has the related tools. Or users can write automatic translation tools by themselves, through which the original contracts on the blockchain can be translated and initialized into the genesis block. As shown in Figure 4.

Consistency verification is required for both manual and automatic translation to ensure that the contract logic is consistent. It is necessary for the implementer to write the test case dataset according to the original contract, and write the test code according to the original contract and the new blockchain contract respectively. When the code coverage of the test case dataset is high enough, the consistency of the contract can be guaranteed to a certain extent. At the same time, when the new blockchain contract does not meet the expectations of the original contract, the new blockchain provides the contract update function to ensure that the new blockchain contract can finally maintain the consistency with the original blockchain contract.

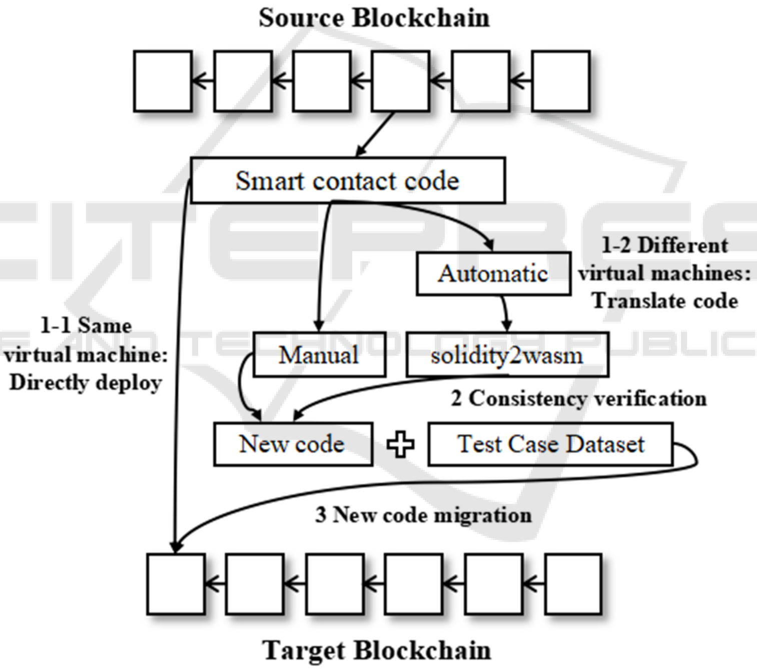

Figure 4: The technical principles of migration of smart contract code.

4.3 Migration of Smart Contract Status

For the contract status, it is necessary to traverse all the status data in the contract through the original blockchain contract interface and other methods, and write the data into the corresponding contract in the genesis block file of the new blockchain. When the new blockchain is started, the contract status can be written into the new blockchain.

5 CONCLUSIONS

A blockchain is a special database that stores transaction information on a block's data structure, while running smart contracts and keeping state within them. For data migration in the blockchain, there are usually several types of data: block and transaction data, smart contract code, and smart contract status data. Traditional databases can directly extract, translate, load data, and finally perform consistency verification. However, for heterogeneous blockchain, data migrantion inevitably involves data reconstruction and modification, which may make the blockchain lose the characteristics as an "invariant historical database".

Heterogeneous blockchain data migration technology ensures the integrity of historical data existence proof, the consistency of smart contract logic, and the compatibility of the interface. The research on migration technology of transaction and block content, smart contract logic and smart contract code data realizes the permanent availability of historical data. This research can break through the technical bottleneck that the current heterogeneous blockchain cannot guarantee the integrity of the original data and business consistency, the loss of the proof of the existence of data in the process of data migration. To form an overall plan for heterogeneous blockchain data migration, which can support the technological upgrade of blockchain platforms in different enterprises and industries, and promote the healthy development of blockchain technology.

REFERENCES

Li Ming,Qiu Honglin,Xu Quanqing,Song Wenpeng,Liu Baixiang. (2021). Research on cross-chain and interoperability for blockchain system. J. The Journal of China Universities of Posts and Telecommunications. 28(05), 1-17.

Satoshi Nakamoto. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. URL https://bitcoin.org/en/bitcoin-paper.

Vitalik Buterin. (2013). Ethereum: A Next-Generation Smart Contract and Decentralized Application Platform. URL https://github.com/ethereum/wiki/wiki/White-Paper.

Waterman A, Lee Y, Patterson D, et al. (2014). The RISC-V Instruction Set Manual. Volume 1: User-Level ISA, Version 2.0.

Yang Z., Liu H., Li Y., Zheng H., Wang L., Chen B.. (2020). Seraph: Enabling cross-platform security analysis for EVM and WASM smart contracts. J. *Proceedings - International Conference on Software Engineering*.

Zhang S, Hou C. (2021). Model of decentralized cross-chain energy trading for power systems. J. Global Energy Interconnection, 4(3), 11.