A composite blockchain associated event traceability method for financial activities

Junlu Wang (wangjunlu@lnu.edu.cn)
Liaoning University
Su Li
Liaoning University
Wanting Ji
Liaoning University
Dong Li
Liaoning University
Baoyan Song
Liaoning University

Research Article

Keywords: Composite chain structure, Private chain, Alliance chain, Traceability of associated events, Reinforcement learning

Posted Date: July 19th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1846793/v1

A composite blockchain associated event traceability method for financial activities

Junlu Wang[1] • Su Li[1] • Wanting Ji[1] • Dong Li[1] • Baoyan Song[1]*

Abstract
The existing blockchain system mostly adopts the equal mining mode. All bookkeepers (entities) record the ledger books on a single master chain. The data storage is random. Moreover, in complex or classified financial scenarios, master chain data is difficult to be correlated or stored regularly, resulting in low efficiency of storage and query; At the same time, in the existing blockchain system, event traceability is mostly only found in the source block, and the implicit association between entities cannot be identified. To solve these problems, this paper proposes a composite blockchain associated event traceability method. This method firstly constructs the blockchain composite chain storage structure model, proposes the concept of private chain and alliance chain, and realizes the adaptive data association storage in complex or classified scenarios. Secondly, on the basis of obtaining the entity block of event source, auxiliary storage space is established to transfer storage relevant data. A query method of associated entity block based on Apriori algorithm is proposed, and the obtained traceability entity block is constructed to construct the source event association diagram, so as to describe the association relationship between event entities. Finally, a risk assessment system based on reinforcement learning is proposed to realize the risk assessment of traceability entity. Experiments show that the composite blockchain associated event traceability method proposed in this paper can reduce 60% of the storage overhead, improve 90% of the query accuracy and 50% of the security.

Keywords Composite chain structure • Private chain • Alliance chain • Traceability of associated events • Reinforcement learning

1 Introduction

Blockchain [1][2] is a new computing paradigm and collaboration model for establishing trust at low cost in an untrusted competitive environment. Due to its features such as high storage density [3], tamper-proof [4] and traceability, blockchain technology has been more and more widely applied. Blockchain carries out data storage by adding blocks [5], and generally, data is stored on a single chain. However, with the expansion of time and transaction data, data expansion [6] may lead to the reduction of storage and query efficiency. At the same time, single chain storage mode cannot realize associated storage or regular storage in complex or classified scenarios [7]. When tracing a source query, the existing blockchain system mostly takes the time stamp and hash pointer as clues [8][9], and realizes tracing to a certain transaction data on a certain block by looking for transactions in the block body, which makes it difficult to trace the implicit relationship between entities in the blockchain and provides insufficient support for subsequent event analysis.

For example, in the financial activities blockchain system, every financial enterprises(entities) if adopt the way of equal [10] with a single mode for data storage, can lead to financial enterprises (entities) transaction data chaotic and random [11], and not stored by law or association storage lead to low query efficiency [12], will cause inconvenience to maintenance. At the same time, according to the single mode stored transaction data of each financial enterprises (entities), when there is an association between financial enterprises (entities), for example, the headquarters and the branches of a financial enterprise, when the total entity (the headquarters of a financial

[1] Baoyan Song bysong@lnu.edu.cn
Extended author information available on the last page of the article

enterprise) legally transfers bad debts to a sub-entity (the branch of a financial enterprise), in the process of tracing query, according to the existing block chain technology, only the sub-entity (namely the branch of the financial enterprise) can be found, and the total entity (namely the headquarter of the financial enterprise) cannot be found out by associative tracing query, resulting in low query accuracy. Therefore, how to establish an efficient blockchain storage mode and associated traceability query method has always been a difficult point in the blockchain field research.

To solve the above problems, this paper proposes a composite blockchain associated event traceability method for financial activities, and its main contributions are as follows:

- In order to solve the problem of low storage and query efficiency caused by the single blockchain storage structure, a blockchain composite chain storage structure model is proposed. In combination with the different application patterns of blockchain and the characteristics of financial entities that need to be stored, the alliance chain [13] is constructed among different entities, and the private chain is constructed within the entity, and the communication consensus mechanism of the compound chain structure is given.
- In order to solve the problem of the current single chain tracing pattern of blockchain, a block query method of associated entity based on Apriori algorithm is proposed. Combining with the traceable property of blockchain, the event source entity block data can be traced out. Auxiliary storage space is established to transfer storage the event-related entity transaction data on the alliance chain, and the Apriori association rule algorithm is introduced to trace and query other entity block data associated with the event source entity block data.
- On this basis, the source event association diagram is constructed for all the traceable entity block data, and the association relationship between event entities is represented by the source event association diagram. The block data of all entities in the association diagram are represented as feature vectors in numerical form, and the risk assessment system based on reinforcement learning is constructed for the entities to realize the risk assessment of traceability entities.

2 Related Work

At present, many scholars have conducted in-depth studies on blockchain storage, related traceability methods and the application of blockchain in finance, and have achieved certain research results.

In terms of blockchain storage, literature [14] proposes a data storage method based on blockchain technology. This method proves that blockchain storage data has the characteristics of high storage density, traceability and tamper-proof, which provides ideas for subsequent research in the storage field. However, this method does not explain the specific storage mechanism. Literature [15] gives a detailed introduction of the data storage mechanism used in the current popular blockchain system, and points out that due to the limitation of data storage mode, the existing blockchain system has the problems of simple query function and low query performance. Literature [16] proposed a method for data storage with blockchain structure of single-chain mode. This method is simple and has high storage efficiency. However, under complex application scenarios, it cannot accurately and completely reflect the association and implicit relationship between entities. Literature [17] proposed a method of building multi-fork chain blockchain structure to store data. This method can store complex and huge data, but using multi-fork chain structure to store data reduces the efficiency of storage and query.

In blockchain traceability queries, literature [18] put forward a kind of high efficient retrieval method of block chain, including range query and Top - k queries, this method has good flexibility, but its solution is synchronizing the data from the blockchain and key-value database to another database, the use of the database for related query operation, not fundamentally solve the problem of query block chain; Literature [19] proposes a single chain pattern block chain traceability method, which can realize basic query application, but cannot realize the traceability query of related entities in complex scenarios related to concatenation and implied relations； Literature [20] proposed a

multi-fork chain pattern blockchain traceability method, which can realize complex traceability query, but the query efficiency is low due to the complex storage structure of multi-fork chain; Literature [21] proposes a data association query method based on data mining technology, which can realize association query under complex scenarios. However, the application of blockchain technology has the disadvantage of inconsistent data model; Literature [22] proposed that data with unique identification should be added to the blockchain to solve the problem of data traceability, but no specific traceability scheme was proposed; Literature [23] proposed a traceability system based on double blockchain technology, TSPPB, which uses the private chain to store the hash value of traceability information, and the detailed traceability information is stored in the IPFS[24] system.

Although blockchain technology has a broad application prospect, it still faces many key challenges such as security, regulation, capacity and timeliness. For example, large-scale application block chain technology inside the bank, especially our country commercial bank building electronic over 30 years, has achieved all focused concentration of processing and data sharing of the business, if you use the blockchain technology "decentralized" model for internal redeployment, their manpower cost to be reckoned with. At the same time, blockchain technology has high requirements on computing resources and storage resources of all distributed participating nodes, and transaction broadcasting has great pressure on the network. Therefore, the large-scale application of blockchain technology in the financial field still needs a process. However, the pace of application of the technology is likely to accelerate further with the extensive involvement of financial institutions.

To sum up, existing methods in blockchain storage and traceability query have problems such as low storage efficiency, complex storage structure and inaccurate traceability query. Therefore, considering the efficiency and accuracy of storage and traceability queries, this paper proposes a composite blockchain associated event traceability method for financial activities.

3 Composite Blockchain Model Construction

For complex entities or classification financial mode data source widely, the huge amount of storage characteristics, based on Merkle tree private chain block structure and based on Merkle Patricia tree alliance chain block structure. The data layer encryption mechanism, network layer and consensus layer block communication and consensus mechanism of the model are designed to realize the construction of composite blockchain model.

3.1 Composite chain structure

The composite chain structure model is composed of private chain and alliance chain. The private chain is built inside the entity to represent the transaction information of the entity. On the basis of private chain, the alliance chain between entities is constructed to form the composite chain structure model. Entity users interact with the private chain in which they are located, while other private chains are transparent to users. Entity users' requirements are realized by corresponding operations of the private chain. The schematic diagram of composite chain structure model is shown in Figure 1.

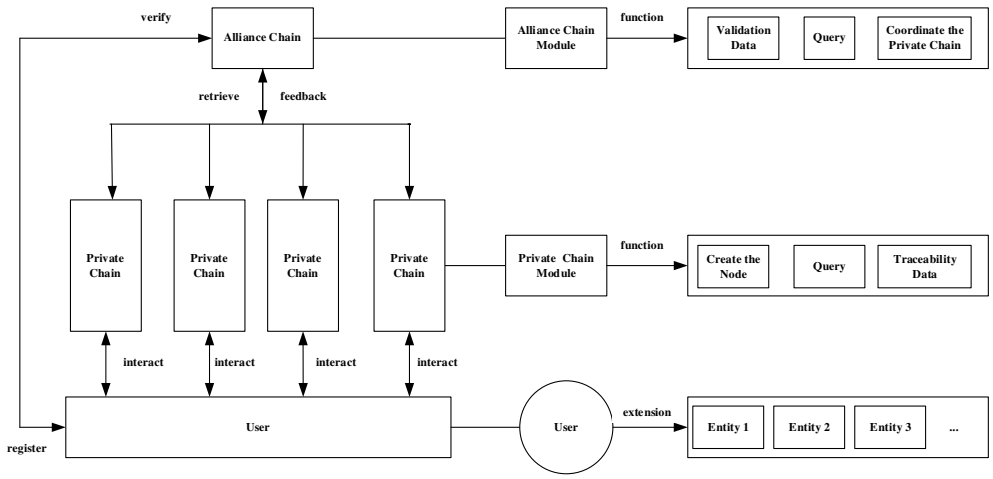

Figure 1 Composite chain structure model

First, build an internal private chain for each entity. In the chain of the private network, according to the transaction frequency entity set a reasonable period, each period the entity transactions as an independent node, Entity collection, certification and update transaction information, and then categorize them by date, all transaction data by the Hash algorithm to carry on the Hash algorithm to generate the corresponding only Hash value, through the cryptography algorithm to encrypt data to generate the key pair, and cover the timestamp, trade is stored in each node local books. Each node in the private chain network within a single entity records all transaction types and their corresponding transaction information through network webcasting and consensus mechanism validation. Then, an alliance chain is built between entities. All entities reach a unified blockchain technical standard and a unified industry standard through agreement to ensure the behavior and technical standardization of participating members. Based on the agreement reached, an alliance chain is built between different entities. In the alliance chain, each entity acts as an independent blockchain large node, and its internal private chain network is interconnected to form the alliance chain network. Collection entity authentication of trading information, after the certification by internal private blockchain, again through the alliance store chain distribution in each entity node local books, based on the all entities of the alliance chain node will be stored in each entity to collect transaction information, finally realizes the transaction data information between entities connected to be shared. In this composite chain structure, data are stored in the underlying block of the blockchain. The block is not only used to store data, but also to verify the validity of trading information and generate the next block. Combined with different storage and query requirements, this paper proposes a private + alliance composite structure for data storage.

3.1.1 Block structure of private chain based on Merkle tree

The private chain uses ECDSA Digital Signature Algorithm based on secP256K1 mathematics to generate two different keys (public key and private key). The private key encrypts the data and decrypts it with the public key when the transaction data needs to be verified. Each block consists of two parts of the block head and the block body, head by the hash value of the last block (Prev Hash), time stamps, random number (Nonce) and trade to the Root of the Hash (Root Hash), the "transaction type" index table, "suspicious transactions" index table, by putting on the hash of the last block, Root Hash and random number information through the hash algorithm to generate the hash value of the current blocks, each block of the block before hash pointer according to time order links constitute the entire private chain link relations, a private chain block header data information as shown in table 1.

Table 1 Private chain block header data

Attribute	Implication
The Version Number	The version number of the data block
PreHash	Hash value obtained by the hash algorithm such as Merkle Root and timestamp of the

	previous block
Timestamp	Approximate time of the block generation
Nonce	Random number of solutions for the current block consensus process
Merkle Tree Root	The root of the Merkle tree for all transactions in the block body through the hash algorithm
"Transaction Types" Index Table	Record the type of transaction to which the transaction for that block belongs
"Suspect Transactions" Index Table	Hash value of a suspicious transaction

The block body stores all the transaction information, and each transaction information is converted into a string of unique hash values through the hash algorithm and stored on the leaf node of the Merkle tree. The hash value of the upper node is generated layer by layer through the hash algorithm, and each data set corresponds to a unique hash root. If the underlying transaction record is tampered, the value of Merkle root will also change. The diagram of the private chain block structure is shown in Figure 2.

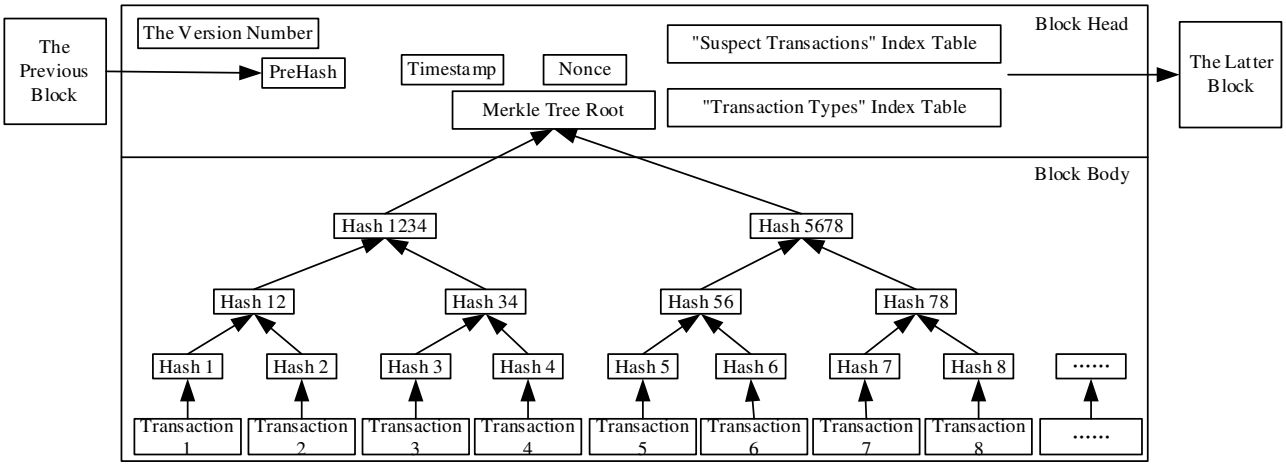

Figure 2 Block structure of private chain

Before the transaction data is stored in the block, first of all, all transaction types carried out by the entity should be counted and numbered uniformly. When the transaction data is stored in the block body, the transaction type information of the transaction should be added to the "Transaction Type" index table of the block head. In the process of data tracing query, the "transaction type" index table is used to query the transaction type of the transaction data to be traced; And then make rules for suspicious transactions, when the transaction data is to be stored in the block body, it shall judge whether the transaction is a suspicious transaction according to the suspicious transaction rules. If it is a suspicious transaction, after calculating the Hash value of the transaction data, the Hash value shall be stored in the Merkle tree and the Hash value shall also be stored in the "suspicious Transaction" index table of the block head. When a data traceability query is made, the suspicious Transactions index table is retrieved.

3.1.2 Block structure of alliance chain based on Merkle Patricia Tree

As described in section 3.1.1, each entity private chain will act as an account in the alliance chain, the entities shall establish the contact of transactions in the form of signing contracts, and establish the block structure of alliance chain based on Merkle Patricia tree for data storage. The characteristics of alliance chain are to emphasize the value between institutions or organizations in the same industry or across industries, the strong correlation of coordination, and the weak centralization within the alliance. The block structure diagram of the alliance chain is shown in Figure 3.

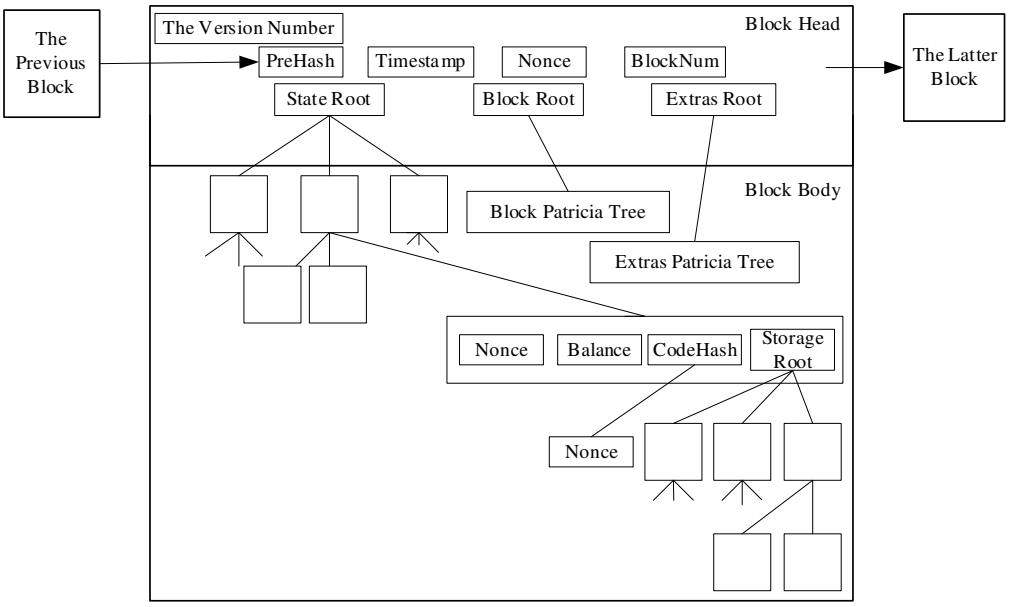

Figure 3 Block structure of alliance chain

The alliance chain also uses ECDSA algorithm to encrypt data. The block head consists of the Prev Hash of the previous block, timestamp, random number (Nonce), and the root Hash of three Merkle Patricia trees, which correspond to the state tree, the transaction tree, and the receipt tree respectively. The transaction information is stored in the block body. Three levelDB databases are established in the alliance chain, namely BlockDB, StateDB and ExtrasDB. BlockDB stores block heads and transaction records, StateDB stores entity status data, and ExtrasDB stores contract information signed between entities. Based on this, the underlying database of the alliance chain is built. The stored content and functions of each Merkle Patricia tree are shown in Table 2, and each block contains the root hash of the entire state tree, where the state tree is updated with period T.

Table 2 Three Merkle Patricia trees of the alliance chain

Identifier	State tree	Transaction tree	Receipt tree
Key	The account address	Transaction number	The index number
Value	The account content	Trading content	The contract content
Storage database	StateDB	BlockDB	ExtrasDB
Uniqueness	Block chain overall a tree	One tree per block	One tree per block
Supported queries	The transaction status of the entity and whether the entity exists	Whether the transaction exists in the block	An event instance of the address

3.2 Composite chain structure communication consensus mechanism

As described in section 3.1, all the transaction data will be stored in the block. On this basis, the encryption mechanism of the transaction data in the block will be optimized at the data layer. And the communication and consensus between blocks are designed at the network layer and the consensus layer, and all transaction information is time-stamped and broadcast in the network in real time and sent to each node in the network, and then all nodes jointly verify and form the "consensus", so as to realize the blockchain system with "no trust". As shown in Figure 4, the composite chain block structure is constructed and the data encryption mechanism is designed at the data layer, and the block structure and chain structure are encapsulated; By constructing P2P networking mechanism, data propagation and verification mechanism and node consensus algorithm of network layer and consensus layer, the communication consensus of composite chain structure is realized based on this.

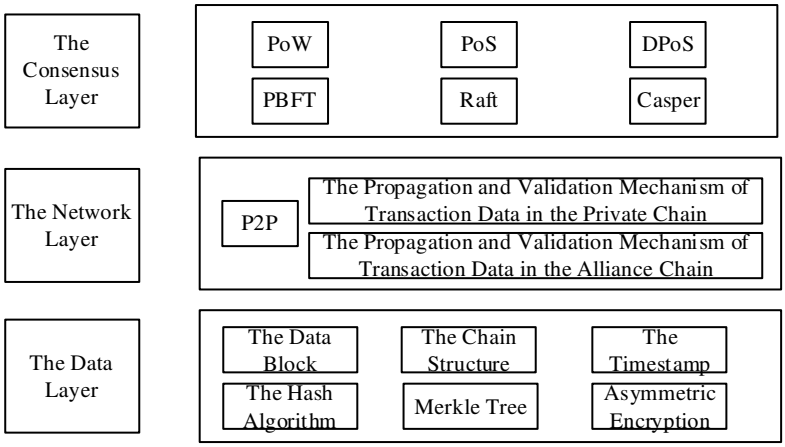

Figure 4 Blockchain system infrastructure model

3.2.1 Data encryption mechanism at the data layer

On the data layer, the traditional blockchain using pseudo random number encryption mechanism to ensure system security, but the pseudo random number encryption mechanism defense level is low, in a high level of security requirements of blockchain system, this kind of encryption mechanism can meet the demand of security, aiming at this problem, this paper proposes a true random number of encryption mechanism based on chaos principle, the process is as follows:

Step1: Deterministic chaotic discrete time dynamical system, f represents chaotic mapping state from $S—>S$, X_n represents the system state after n times of substitution. No new information is generated within the whole system, which is expressed by mathematical formula (3.1) :

$$X_{n+1} = f(X_n) \qquad X_{n+1}, X \in S \subseteq R^n \quad (1)$$

The results are determined by the initial value of the system:

$$H(X_n|X_0) = 0 \qquad (2)$$

Step2: Divide S into m disjoint states:

$$\beta = \{\beta_1, \beta_2, \beta_3, \ldots, \beta_m\} \qquad (3)$$

Since X_n is not equal in each of the intervals divided by β, a random sequence in base m is generated.The interval that produces the maximum variation of the function f is defined as the generation partition.There are

$$\cup_{i=1}^{m} \beta_i = S \quad (4)$$

$$\beta_i \cap \beta_j = \emptyset \qquad \forall i \neq j \quad (5)$$

Step3: The piecewise linear chaotic mapping function is adopted. X_n represents the result obtained by repeatedly substituting the piecewise function n times. After substituting the function again, X_{n+1} is obtained. The value of the parameter B in the expression determines the random mass of the whole X, and A is a scalar value. The piecewise function is as follows:

$$X_{n+1} = \begin{cases} BX_n + A, A \leq X_n \leq 0 \\ BX_n - A, 0 \leq X_n \leq A \end{cases} \qquad \text{n=0,1,2, } \ldots \qquad (6)$$

In this piecewise function, the generation of a generated non-memory discrete information source in the chaotic state is determined to be divided into $\beta = \{[-A, 0], [0, A]\}$, from which we can see that the system can generate a random sequence for encryption.

3.2.2 Block communication mode at the network layer

In the network layer, it is necessary to construct the networking mode of the network and the communication between nodes. The composite chain structure is based on the peer-to-peer network. There is no central node in the

network, and the P2P protocol which can tolerate single point of failure is used as the network transmission protocol. And because the private chain dynamically adds new blocks at any time, the network scalability, reliability and maintainability requirements are high, the alliance chain needs to support complex queries, so the private chain and the alliance chain are proposed to introduce different network communication structures to achieve communication.

Each node in the private chain is assigned a hash block of its own, and each node manages its own hash block. All hash blocks are combined into a hash table, and all nodes in the private chain jointly maintain the hash table, forming a network communication structure based on the fully distributed structured topology (DHT). Hash the IP address of each node on the private chain to get the node value of that node, the nodes are scaled from small to large to form a Chord loop, and the distance between each node and the next neighboring node is obtained based on the node value. Based on this, the value interval that each node is responsible for is obtained, and the keywords on each node are extracted. By hashing the keywords, the Hash value obtained is distributed according to the interval that each node is responsible for, so that the stored information of each resource is stored on one node. When searching a resource, it first hashes its keyword and compares the obtained value with the value interval table of the current node to get the most likely index information of the resource. Then, it queries the node to get the index of the resource. According to the index, the node where the resource is located can be found and communication can be established.

In the alliance chain, the master entity is regarded as the super node and the child entity as the ordinary node. A high-speed forwarding layer is formed between the super nodes, and then several layers are formed between these super nodes and the ordinary nodes they are responsible for. The information of other nodes in the system is stored on each super node, and the discovery algorithm forwards the query request among the super nodes, and the super nodes forward the query request to the appropriate leaf nodes, forming a network communication structure based on the semi-distributed topology. The communication structure diagram of the alliance chain is shown in Figure 5.

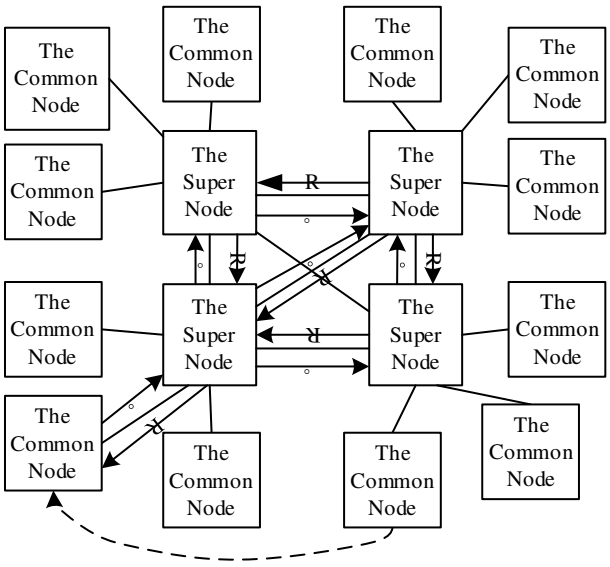

Figure 5 Schematic diagram of communication structure of alliance chain

3.2.3 Block consensus algorithm at the consensus layer

Decentralized composite chain structure run jointly by various maintenance, the network nodes can be provided by any party, part of the node may not be credible, in order to ensure the high availability of the data stored composite chain structure nodes and consistent trust and the security of the whole system, after establishing the communication between blocks, it is necessary to establish the consensus algorithm between blocks in the consensus layer. However, because of the different application requirements of private chain and alliance chain, different consensus algorithms are introduced.

Each node of the private chain has three states: Follower, Candidate, and Leader. Through Leader Election and Log Replication, a consensus mechanism among nodes based on RAFT consensus is formed on the basis of ensuring security. At the beginning, all nodes are set to the follower state. If they do not receive messages from the leader, one of the nodes will become a candidate. The candidate node sends a request vote message to the other nodes. The other follower nodes will return a message of approval to the candidate node. If more than half of the followers agree, the node will become the leader. The client sends a message to the leader node, and the leader node will add an entity to the log. The leader will broadcast the entity to the other follower nodes. When the leader node receives a reply from most of the nodes, the leader needs to commit the entity and then broadcast it to the other follower nodes. The follower node receives the commit message, commits locally, and the process ends.

Consensus process of the alliance chain is carried out in three stages: pre-prepare, prepare and commit. The pre-prepare and prepare phases order the requests sent in the same view, so that each replicas node recognizes the sequence and executes it accordingly. The prepare and commit phases ensure that requests that have reached the commit state remain in the same sequence in the new view even after a view change. Forming a consensus mechanism between nodes based on the Practical Byzantine Fault Tolerance (PBFT) consensus, this mechanism reduces the complexity of the original byzantine fault tolerant algorithm from exponential to polynomial. Take one replica as the primary node and the others as the backup, the client sends a request to the master node to use the service operation, the master node broadcasts the request to other replicas, all replicas execute the request and send the result back to the client, and the client waits for F+1 different replica nodes to send the same result back as the final result of the whole operation.

3.3 Instance

In the establishment of blockchain for financial sector activities, assume that there are two example financial institutions, bank and securities, the transaction types of banks include deposit business, loan business, loan business, securities investment business, bank card business, guarantee business, electronic banking business, financial management business, etc. The transaction types of securities include stock trading, bond trading, fund trading, and other financial derivatives. Bank A is the head office, Bank B and Bank C are branches, Securities A is the head office, and Securities B and C are branches. In the construction of the composite chain structure blockchain storage structure, the transaction types of banks and securities will be counted and numbered respectively, and two transaction type index tables will be established. According to the identity and behavior of the clients of the bank and securities, the source, amount, frequency, flow direction and nature of transaction funds, etc. To judge whether it complies with the anti-money laundering and anti-terrorist financing regulations and guidelines, risk tips, money-laundering type analysis report and risk assessment report issued by the People's Bank of China and its branches; Or to judge whether it conforms to the criminal situation analysis, risk tips, crime type reports and work reports issued by public security organs and judicial organs, etc. If they meet the requirements, they will be listed as suspicious transactions, based on which suspicious transaction rules will be formulated.

Firstly, build the internal private chain of each organization, the transaction information collected, authenticated and updated by bank A, B, C and securities A, B, C are statistically sorted into their respective transaction information tables. A reasonable period is set according to the trading frequency of each institution, and the transactions of this institution within each cycle are an independent node. When the transaction data is stored in the block, the hash algorithm is used to hash each transaction data in each cycle of each institution to generate the corresponding and unique hash value. Cryptography algorithm is used to encrypt the transaction data and generate key pairs (public and private keys). Determine the type of the transaction and mark the transaction type index table in the block header. And judge whether it is a suspicious transaction according to the suspicious transaction rules. If so, the hash value of the transaction is stored in the suspicious transaction index table of the block header and the block body. If not, the transaction is stored directly on the block body. According to the previous block of each

block, the hash pointer is chronologically linked into a private chain, through network broadcast and consensus mechanism verification, each node of the internal private chain network of each institution will record all transaction types and their corresponding transaction information.

Then, the alliance chain between institutions is constructed. Banks and securities institutions reach a unified blockchain technical standard and a unified industry standard through agreement to ensure the normative behavior and technology of participating members, and the alliance chain is constructed based on the agreement reached by them. Statistic and collate the information of trade contracts signed between institutions, and set up a table of trade contracts. Each institution acts as an independent blockchain node and an account in the alliance chain, and its internal private chain network is interconnected to form the alliance chain network. Account information is stored in StateDB in the alliance chain, block header and transaction information is stored in BlockDB in the alliance chain, and contract information in the transaction contract table is stored in ExtrasDB in the alliance chain. Bank A and Security A are super nodes in the communication network, bank B and C and securities B and C as ordinary nodes, each bank and securities institution collects authenticated transaction information after the internal private blockchain is authenticated, verified by webcasts and consensus mechanisms. Then distributed and stored in the local ledger of each entity node through the Alliance Blockchain, based on this, all nodes of the entire blockchain alliance will store all transaction information of all institutions. The composite chain structure of financial institutions is shown in Figure 6.

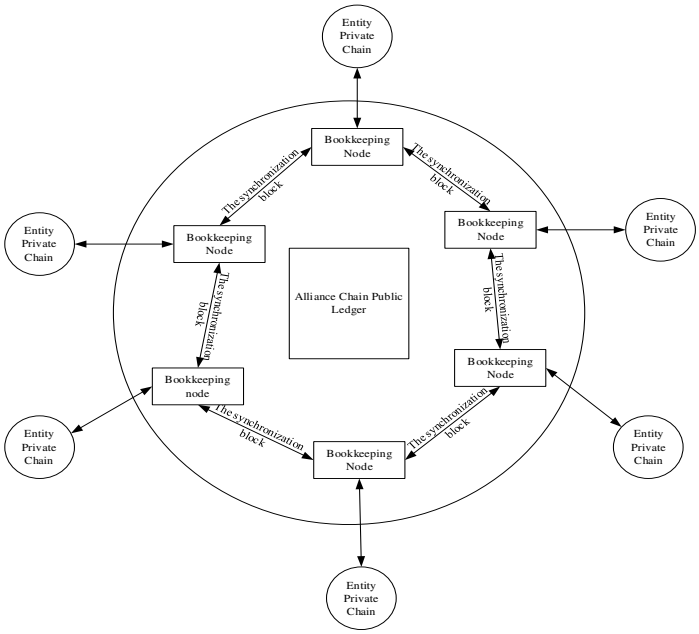

Figure 6 Composite chain structure of financial institutions

4 Associated Event Traceability Query Mechanism

On the basis of the establishment of the composite chain structure, first query the transaction data of the block entity of the event source. Then obtain other entity blocks that have an implicit relationship with the transaction data of the source entity block, the source event correlation diagram is drawn for all the entity block data from the traceability query, and then the risk assessment system based on reinforcement learning is constructed.

4.1 Event source entity block traceability

After the composite chain structure is established, all transaction information for each entity is stored in the private chain within the entity, the source entity of the problem exposed in the event is that after the transaction data is stored on the chain, it secretly and illegally modifies its books, resulting in inconsistent data information on the

chain and its own books under the chain, or abnormal transactions of the entity within a certain period of time. Since the composite chain structure data traceability not only includes data traceability to a single node, but also includes data validation, query and synchronization to each distributed node, so when tracing to the event source entity, with the time stamp and the previous block hash pointer as clues, the transaction in the block body is decrypted by the public key, and converted to the data format before storing in the private chain by the inverse hash algorithm. Transaction data that is inconsistent with the entity's own transaction ledger is the problem transaction data for the block of the entity at the source of the event. Specific situations are divided into the following three:

(1) Problem transactions may appear in the "Suspicious Transactions" index table of the entity's private chain block header, the transaction data of each "suspicious transaction" index table is decrypted and the inverse hash algorithm is compared with the entity's own ledger. Inconsistent transaction data is the problem transaction data of the event source entity.

(2) If no problems are found in the index table of "Suspicious Transactions", problem transactions maybe occur in transactions within the entity's private chain block, after decryption and inverse hash algorithm, the transaction data of each entity's private chain block body is compared with the entity's own ledger. Inconsistent transaction data is the problem transaction data of the event source entity.

(3) If no problems are found in the transaction of the entity private chain block body, the transaction in question would occur in transactions over a certain period of time in an entity's private chain, horizontal observation of transaction data on the entity private chain, global observation after decryption and inverse hashing algorithm of all transaction data of the entity private chain, if the transaction data in a certain period of time is relatively frequent, abnormal, or the transaction amount is too large, then the transaction in that period is the problem transaction data of the event source entity.

In the composite chain blockchain structure, the levelDB database is used to store data at the bottom of the system, and on this basis, the primary and secondary index method is introduced to further improve the search efficiency. The primary and secondary index method uses LevelDB as the primary key index, and sets the fields related to the query in the chain as the secondary index. The query layer consists of this index structure. The structure diagram of the primary and secondary index method is shown in Figure 7. First, the query command is sent to the query module, and the Key set of the query result is determined by the auxiliary index. Then use the Key Value to find the Value on the primary Key index, levelDB, and return the result.

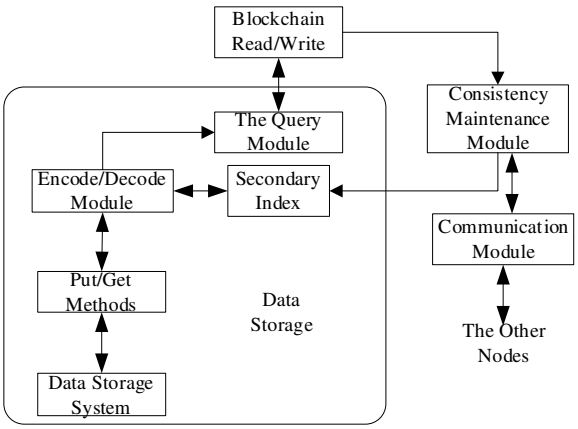

Figure 7 The primary and secondary index structure

4.2 Associated entity block query

On the basis of finding out the transaction data of the problem entity of the event source entity, it is further necessary to search out the transaction data of other entity blocks that have implicit relationship with the problem transaction data of the event source entity block. Firstly, auxiliary storage space is established to transfer store transaction data

of entities related to the event on the alliance chain. Then the transaction data of the implicit relational entity can be traced by using the query method of the associated entity block based on the Apriori algorithm. Finally, the source event correlation diagram is constructed for the traceable entity block transaction data, and then the correlation relationship between entities in the event is described. The query method for the associated entity block is as follows:

Step1: Establish an auxiliary storage space to store the entity transaction data related to the event in the alliance chain. After finding the problem transaction data for the event source entity block, query the transaction type of the problem transaction from the transaction type index table of the entity's private chain block header, find out all other entities in the alliance chain receipt database that have contractual relations with the event source entity. And then find the transaction data of other entities' private chains on these entities' private chains that have the same transaction type with the problem transaction data of the entity block in the source of the event, after decryption and inverse hashing algorithm, the above transaction data is stored in the auxiliary storage space.

Step2: Relevant definitions and rules of the query method of the associated entity block. Assume that $I = \{i_1, i_2, \dots, i_m\}$ is the set of transaction items for all entities in the auxiliary storage space, the transaction database $D = \{t_1, t_2, \dots, t_n\}$ is a transaction composed of the TID identity of the entity, each transaction $t_i (i = 1,2, \dots, n)$ is a set of items in I, that is all transactions in the entity, i.e. $t_i \in I$, the number of items in it is defined as the length of the transaction.

First, find out the frequent item set, and set D as the transaction database and I as the global item set. If the item set $L \in I, L \neq \emptyset$ and the support degree of L is not less than the minimum support, that is $support(L) \geq Minsupport$, L is called frequent item set. The following convention Lk refers to the k-item frequent set. If an item set L is frequent, then all its subsets (not including the empty set) must also be frequent; If the item set L is infrequent, then the superset containing L is infrequent, so L will not be included in the frequent item set and can be deleted when the frequent item set is generated in the future. If a transaction length is less than k, then the transaction cannot contain k-item frequent set Lk, and the transaction can be deleted when searching for Lk. Referencing the above features reduces the number of lookups when tracing the associated entity blocks.

Then determine the Apriori association rule, the implication expression of the association rule is $X => Y$, where$X, Y \subseteq I$ and $X \cap Y = \emptyset$. X is called an antecedent, Y is called a consequent, and the above association rule means that X can causes Y. $count(X)$ is the number of occurrences of item set X in D, and $||D||$ is the total number of transactional database records. If the association rule $X => Y$ is true, then: support degree $support(X => Y) = count(X \cup Y)/||D||$ and confidence degree $C\% = confidence(X => Y) = P(Y|X) = P(X \cup Y)/P(X)$. That is, if database D contains X records, at least $C\%$ also contains Y.

Step3: Carry out the query method of the associated entity block to find the associated entity private chain. According to the characteristics of the entity data, a threshold of support is set as $Minsupport$ and a threshold of confidence is set as $Minconfidence$. Through the method of layer by layer search iteration, the candidate set Ck is first generated, and then the support degree of all k item sets are calculated. If the support degree meets the requirement of the minimum support threshold, it becomes the frequent k itemset Lk. Then the candidate set $Ck + 1$ is generated on the basis of Lk, and the frequent $(k + 1)$ item set $Lk + 1$ is determined by judging the minimum support, and so on until the next frequent item set can not be found. The items in each frequent item set are composed of association rules, and the confidence of the rules is calculated respectively. If the confidence is greater than the minimum confidence threshold, the association rule is generated, the rule that meets the requirements of minimum support and minimum confidence at the same time is a strong association rule, indicating that the transaction data of the project has a high degree of association with the transaction data of the problem block occurring in the event source entity. In other words, it is determined that the entity of the transaction has an implicit association relationship with the event source entity.

On the basis of finding out all the transaction data of other entity blocks that have implicit relationship with the problem transaction data of the source entity block of the event, the corresponding relationship between the transaction data and the entity is constructed to build the correlation graph of the source event based on the

transaction time. By establishing the connection between the entities, the relevant facts can be formed. The source event correlation diagram consists of a set of facts, each of which can be represented as an SPO triplet (entity 1, relationship, entity 2) and (entity, attribute, and attribute value) to represent the relationship between the entity and the transaction in the event. Entities and their relationships are represented by (entity 1, relationship, entity 2), and entities and their transactions are represented by (entity, attribute, attribute value). A schematic diagram of the source event association graph is shown in Figure 8.

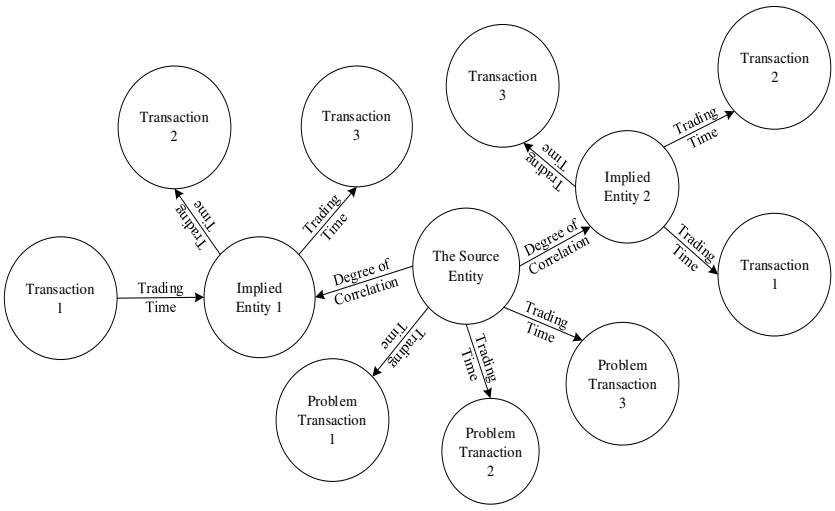

Figure 8 Source event association diagram

When creating the source event association diagram. First, create the source entity and its problem transaction data information, the line between the entity and the transaction indicates when the transaction took place, then create other entities that have an implicit relationship with the source entity, the line between the source entity and the implied entity represents the degree of correlation between the two, which is determined by the support and confidence calculated by the query method of the block of the associated entity. The higher the support and confidence, the higher the correlation. On this basis, establish a number of transactions related to the event in the implied entity, through the correlation diagram, the correlation relationship between entities in the event can be represented concretely.

4.3 Construction of risk assessment system based on reinforcement learning

As described in Section 4.2, on the basis of finding several transaction data related to this event, all entity block data in the source event correlation graph are represented as numerical eigenvectors, and the risk assessment system based on reinforcement learning is constructed for the entities in the source event correlation graph, so as to realize the risk assessment of traceable entities. The baseline model of the risk assessment system is shown in Figure 9.

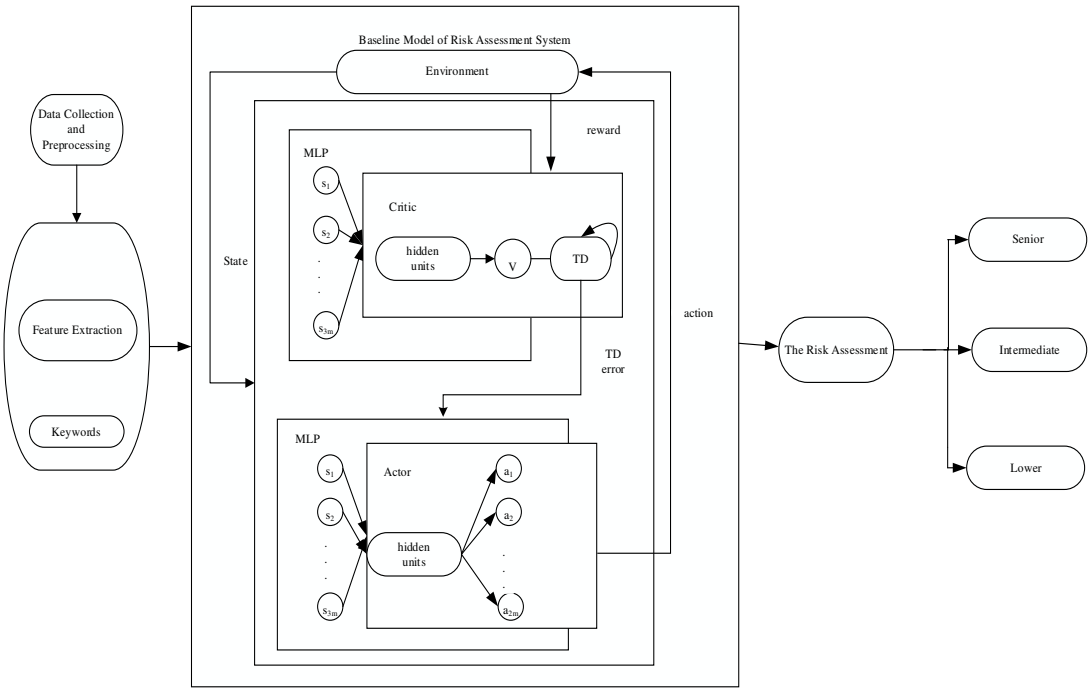

Figure 9 Baseline model of risk assessment system

As shown in Figure 9, data collection and preprocessing include obtaining the original data of entities in the correlation graph from the composite chain structure and related processing of text information; In feature extraction, this paper, aiming at text information, extracts keywords features, and converts text information into numerical feature vector form; Then the baseline model of the risk assessment system based on reinforcement learning is constructed to realize the risk assessment of traceability entities.

Entity transaction information includes text information such as the transaction object, the introduction of the transaction object, the transaction contract, and numerical information such as the transaction amount and the transaction time. The preliminary data processing is carried out to prepare for the subsequent feature extraction. After preprocessing, feature extraction is carried out on the text information, which is transformed into a numerical vector, features are extracted from the text information from the semantic level, and keywords are selected to represent the text features. In this paper, information gain method is used to extract key words.

After obtaining the above multiple text features, Max-Min ACLA(Actor-Critic Learning Automaton) algorithm is introduced to build the baseline model of risk assessment system. Let data set $\mathrm{D} = \left\{(x^1, y^1), (x^2, y^2), \ldots \left(x^i, y^i\right), \ldots (x^n, y^n)\right\}$ and n be the total number of samples, where x^i is the eigenvector of the ith sample, y^i is its target category, $y^i \epsilon \{0,1, \ldots, N-1\}$, and N is the total number of categories of 3. The algorithm constructs an agent for each target category. For each training sample, the intelligence with the same training sample category chooses actions to maximize its reward, while the intelligence with other categories chooses actions to minimize its reward. The Markov decision-making process of this algorithm is as follows:

Step1: Define the state set S. Usually it is continuous, for the length of the input vector x^i is m, state of $s^i \epsilon \mathrm{S}$ contains $3m$ elements. These elements are divided into three buckets. s_t represents the state vector at time t in a single iteration. For the input vector x^i, let the initial state $s_0 = \left(x^i, \vec{0}, x^i\right)$, and the size of the three buckets are all m.

Step2: Define action set A. There are a total of $2m$ actions, and each action sets the value of its corresponding bucket element. a_t represents the action selected at time t.

Step3: Define transfer function set T. At the next moment, the state $s_{t+1} = O(s_t, a_t)$, and the execution rule of operation O is: If the action is $0 \leq \mathrm{a}_t < \mathrm{m}$, set the $(\mathrm{m} + \mathrm{a}_t)$ th bucket element to the value of the a_t element

of the input vector; If the action satisfies $\mathrm{m} \leq \mathrm{a}_t < 2\mathrm{m}$, set the $(\mathrm{m} + \mathrm{a}_t)$ th bucket element to 0.

Step4: Define Instant Reward R. $r_t = 1 - \frac{z}{m}$, where z is the number of zeros in the state vector.

Step5: Define discount factor γ. In the training process, it is assumed that the number of actions executed in a single iteration is h. The agent interacts with the training sample, and each agent performs h actions and learns from the observed state transitions and the immediate rewards. Two different multilayer perceptrons (MLPs) are used to represent the state value function (critic) and the action selection function (actor). In this paper, the number of hidden layer nodes and the learning rate of the two MLPs are set to be the same in order to reduce the number of parameters. Assuming that $V_i(\cdot)$ represents the value function of agent AC_i of class i, and the sample error is expressed by TD error, AC_i will receive (s_t, a_t, r_t, s_{t+1}) after state s_t has performed its action, and use TD error δ_t to update the value function:

$$\delta_t = r_t + \gamma V_i(s_{t+1}) - V_i(s_t) \quad (7)$$

$$V_i(s_t) = V_i(s_t) + \alpha\delta_t \quad (8)$$

Where, α is the learning rate of critic. If the sample class $y = i$, then the MLP target value of AC_i selected action is:

$$\begin{cases} G = 1, \ if \ \ \delta_t \geq 0 \\ G = 0, \ if \ \ \delta_t < 0 \end{cases} \quad (9)$$

AC_i can maximize immediate rewards to learn higher state value functions. If the sample class $y \neq i$, then the MLP target value of AC_i selection action is:

$$\begin{cases} G = 0, \ if \ \ \delta_t \geq 0 \\ G = 1, \ if \ \ \delta_t < 0 \end{cases} \quad (10)$$

For samples of different categories, AC_i will receive a negative immediate reward, which will be passed to the value function of the initial state through TD learning. In the test phase, the agent does not need to choose the action. First, the value function $V_i(s_0)$ is calculated for all class i's AC_i, and the input sample will be predicted to be the class y_p represented by the agent with the maximum function:

$$y_p = arg \ \ i_{max} V_i(s_0) \quad (11)$$

A baseline model of reinforcement learning risk assessment system consisting of three agents was constructed according to the above process to conduct risk assessment on the entity.

4.4 Instance

In the above-mentioned composite chain structure of banks and securities, when a financial case or financial risk occurs, it is due to the problem of an institution object itself, unauthorized modification of a transaction data in the books, or deletion of relevant transaction records and the abnormal trading of the institution in a certain period of time, etc., Suppose the problem is that there are frequent and unusual roll-ins and roll-outs of securities B over a period of time.

In the retrospective inquiry of the financial case or financial risk, the source institution of the financial case or financial risk is firstly traced back, that is, Securities B. The transaction data in the index table of suspicious transactions in the private chain of each bank and securities institution is transformed into the data format before being stored in the block through decryption and inverse hashing algorithm, if no problems are found after comparing with the accounts of each institution; Then, the transaction data in the private chain block of each bank and securities institution is converted into the data format before the block by decryption and inverse hashing algorithm. If the transaction data is compared with the account of each institution, no problem is found; Further horizontal global observation of the transaction information of each private chain shows that Securities B frequently

enters the account and enter an item of expenditure in the accounts within a certain period of time and the transaction amount is too large. Therefore, Securities B is listed as the source problem organization of the financial case or financial risk, and the transactions during this period are problematic transactions.

Then trace back to other entities that have an implicit relationship with the problematic transaction in Security B. From the ExtrasDB database of the composite chain structure alliance chain, we find out the institutional securities A, Bank B and Bank C that have contract dealings with Securities B, and find out the transaction information of the frequent and abnormal transfer and transfer types and transaction time between Securities A and Bank B and Bank C and Securities B. After decryption and inverse hashing algorithm, the above transaction data is sorted and converted into the transaction format and stored in the established auxiliary storage space. The transaction information table after sorting and conversion is shown in Table 3. T1,T2,T3 and T4 are Securities B, Securities A, Bank B and Bank C respectively, and I1-I5 are transactions conducted by them. The minimum support is set as 50% and the minimum confidence is set as 70%. The Apriori algorithm is used to trace and query the implicit relational entities and their trading data.

Table 3 Transaction information table

TID	Attribute
T1	I1,I2,I3,I4
T2	I1,I2,I3
T3	I3,I4
T4	I1,I2,I5

Through layer by layer search iteration, according to Minsupport is 50%, frequent 3-item set $L3 = \{\{I1, I2, I3\}\}$ can be obtained, according to Minconfidence is 70%, $\{I1, I3\} -> \{I2\}$, $\{I2, I3\} -> \{I1\}$ is frequent association rule. In other words, there is a strong correlation between I1,I2, and I3, which proves that security A has an implicit correlation with security B. After finding out all the transaction data related to financial risks or cases, the source event correlation graph is constructed for the transaction data of securities A and B, and the correlation relationship between securities A and B is represented by the graph.

5 EXPERIMENT AND ANALYSIS

The experimental data set is from the flush website data, and the experimental data crawled from the flush website is selected for calculation. The effectiveness of the composite blockchain event traceability method is verified through storage efficiency and storage overhead, association traceability query efficiency and accuracy, etc., and the single chain structure and multi-fork chain structure are compared for storage and traceability query.

Further, risk assessment and analysis experiments are carried out on entities. Data sets of entities in the source event correlation graph are divided into training set and test set in a ratio of 7:3, and each type of data generally maintains its proportion in the overall data. In the classification, 0 represents senior risk, 1 represents intermediate risk, and 2 represents low risk. For a variety of text features, the baseline model of risk assessment system is firstly used to classify related entities, and then for keyword features, in this paper, a variety of traditional supervised machine learning methods are used to do the same experiment for comparison, including support vector machine (SVM), logistic regression (LR), multi-layer perceptron (MLP), and Naive Bayesian. The parameter Settings of each model are shown in Table 4.

Table 4 Model parameter settings

The model name	parameter
Baseline model of reinforcement learning	Number of hidden layer nodes hu: 11;Learning rate α: 0.03;Number of actions executed h: 7;Discount factor γ: 0.9
SVM	Penalty factor C: 1.0;Kernel function(kernel): rbf
LR	Penalty: l2; Optimized fitting reference algorithm (Solver) : liblinear
MLP	Optimizer solver : lbfgs; The regularization parameter alpha:le-5; Number of nodes

	in hidden layer hu :5; Learning rate :0.001
Naive Bayesian	Method: Bernoulli; Smoothness factor :1.0

The effect measurement indicators used in this paper are defined as follows:

$$\text{accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \quad (12)$$

$$\text{precision} = \frac{TP}{TP+FP} \times 100\% \quad (13)$$

Where, TP represents the number of correctly classified positive examples, FP represents the number of wrongly classified negative examples, TN represents the number of correctly classified negative examples, and FN represents the number of wrongly classified positive examples. Accuracy represents the correct rate, that is, the overall performance of the model without distinguishing specific categories. The overall performance and convergence of the model can be obtained through accuracy; Precision represents accuracy rate, which can be obtained for each category. The classification ability of each category and the whole can be evaluated by precision.

5.1 Composite chain storage efficiency analysis

The experiment simulates the efficiency comparison between the composite chain structure and the blockchain storage structure with single chain and multi-fork chain. The abscissa represents the storage entity data set, and the ordinate represents the storage time. The experimental results are shown in Figure 10.

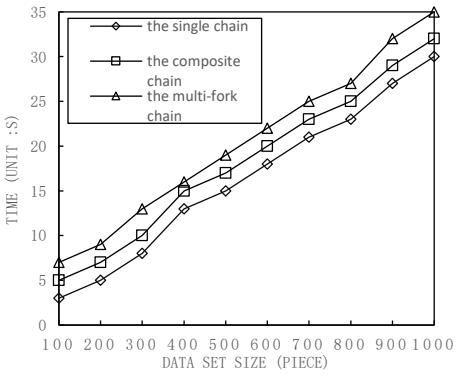

Figure 10 Composite chain storage efficiency comparison diagram

It can be seen from Figure 10 that the storage efficiency of the composite chain storage structure proposed in this paper is between the single chain structure and the multi-fork chain structure.

5.2 Composite chain storage overhead analysis

The experiment simulates the cost comparison between the composite chain structure and the blockchain storage structure using single chain and multi-fork chain. The abscissa represents the storage entity data set, and the ordinate represents the space required for storage. The experimental results are shown in Figure 11.

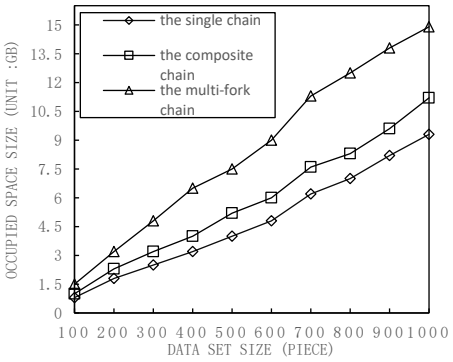

Figure 11 Composite chain storage overhead comparison diagram

It can be seen from Figure 11 that the storage cost of the composite chain storage structure proposed in this paper is between the single chain structure and the multi-fork chain structure. Multi-fork chain storage structures take up much more space than single and composite chain storage structures because of the redundant storage of physical data.

5.3 Associated traceability query efficiency analysis

Figure 12 shows the comparison of traceability query efficiency between the traceability method of composite blockchain associated events and single chain and multi-fork chain storage modes. Where, the abscissa represents the entity data set stored, and the ordinate represents the time required for query. The experimental results are shown in Figure 12.

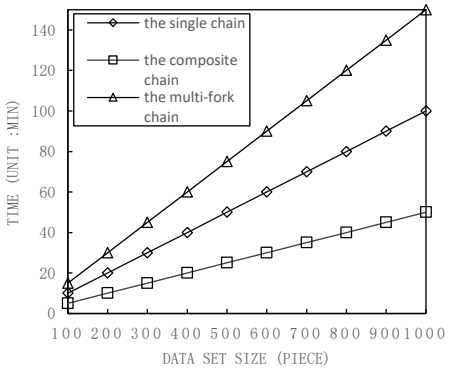

Figure 12 Associated traceability query comparison diagram

As can be seen from Figure 12, with the increase of data sets, the composite blockchain event traceability method proposed in this paper has more obvious advantages than the other two storage structure traceability query algorithms.

5.4 Associated traceability query accuracy analysis

Figure 13 is the comparison diagram of query accuracy. The abscissa represents the stored entity data set, and the ordinate represents the correlation traceability query accuracy. The experimental results are shown in Figure 13.

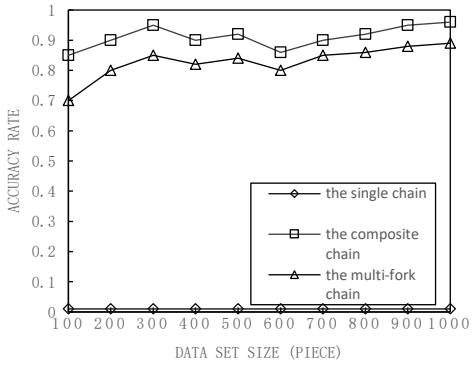

Figure 13 Associated traceability query accuracy rate comparison diagram

As can be seen from Figure 13, The composite blockchain event traceability method and the multi-fork chain model can be used to trace out the implied relationship between entities in the blockchain, but the former has a higher accuracy, and the single-chain model cannot be used to trace out the association relationship between entities.

5.5 Accuracy rate analysis of risk assessment

The training set accuracy rate results of experiments with different number of keywords are shown in Figure 14, where the abscissa represents the number of iterations, the ordinate represents correct rate, and curves with different marks represent model accuracy rate with different number of keywords. The experimental results are shown in Figure 14.

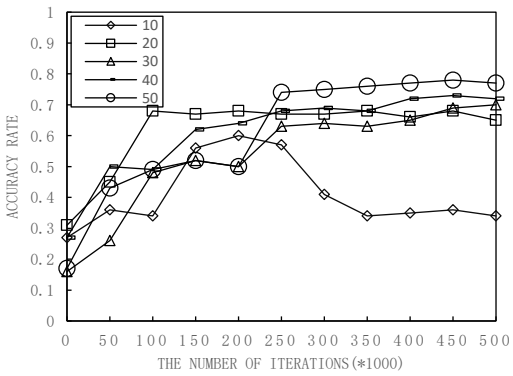

Figure 14 Risk assessment accuracy rate comparison diagram

As can see from Figure 14, when the model does not converge, accuracy shows an upward trend; when the model converges, accuracy tends to be stable.

5.6 Precision rate analysis of risk assessment

In this paper, four traditional machine learning algorithms are used to do the same experiment to verify the effectiveness of the baseline model of the risk assessment system based on reinforcement learning. Choosing Precision as the primary metric to identify institutions at risk, the abscissa represents the number of keywords, and the ordinate represents precision. Different marked curves represent the keyword characteristic precision of different algorithms, the experimental results are shown in Figure 15.

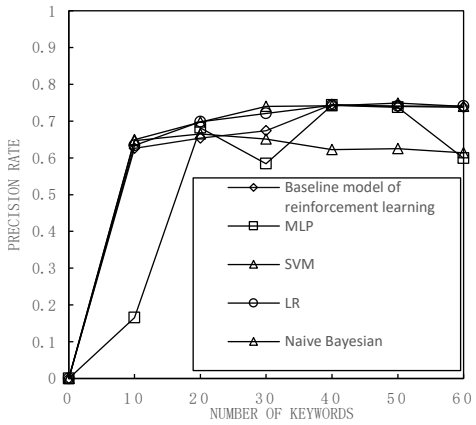

Figure 15 Risk assessment precision comparison diagram

The following figure shows the precision of each subclass when each algorithm has the optimal effect.

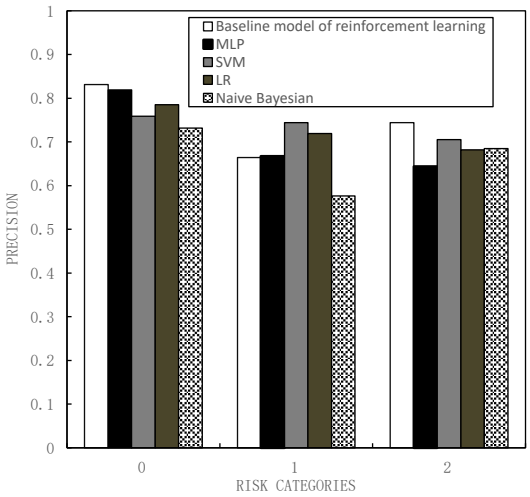

Figure 16 Precision rates for each risk category comparison diagram

As can be seen from Figure 15, in terms of features, the performance of most models reached the optimum when the keywords were 40. In terms of models, the baseline model of risk assessment system based on reinforcement learning is better than most traditional machine learning models in terms of overall precision. At the same time, it can be seen from Figure 16 that the baseline model of risk assessment system based on reinforcement learning is less affected by data imbalance, while most traditional machine learning models are affected by data imbalance, resulting in low precision in the category of less data.

6 Conclusion

Blockchain is an effective means for secure data storage and traceable query. Blockchain stores data by adding blocks. Data is stored in single-chain mode. As the expansion of time and transaction data leads to data bloat, storage and query efficiency may decrease. At the same time, the single chain storage mode cannot realize the adaptive data association storage under complex or classification scenarios. In traceability query, the existing blockchain takes timestamp and hash pointer as clues. By searching transactions in the block body, it can only be traced to a certain transaction data in a certain block, but cannot be traced to query the implicit relationship between entities in the blockchain, which provides insufficient support for subsequent event analysis. In view of these problems, this paper makes an in-depth study on the blockchain storage structure and query methods, and proposes a composite blockchain associated event traceability method. Firstly, the blockchain composite chain storage structure model is constructed to realize the adaptive data association storage under complex or classification scenarios. Secondly, in

the traceability query, an auxiliary storage space is established to transfer the relevant data after obtaining the entity blocks of the source of the event. A query method of the associated entity blocks based on the Apriori algorithm is proposed, and the traceable entity blocks obtained are constructed to describe the correlation between the event entities. Finally, the risk assessment system based on reinforcement learning is proposed to realize the traceability entity risk assessment. The experimental results show that the traceability method proposed in this paper has good effectiveness, accuracy and availability.

Acknowledgements This work was supported by the National Key R&D Program of China (Grant No. 2021YFF0901004), the National Natural Science Foundation of China (Grant Nos.62072220, 61502215), the Scientific Research Project of the Educational Department of Liaoning Province (Grant No. LJKZ0094), and the Central Government Guides Local Science and Technology Development Foundation Project of Liaoning Province (Grant No.2022JH6/100100032).

Author Contributions Author Junlu Wang and Dong Li are responsible for literature collection and writing. Author Su Li and Wanting Ji are responsible for writing ideas and revising the paper. Author Junlu Wang, Su Li, Wanting Ji, Dong Li and Baoyan Song are responsible for giving guidance and revising the paper. All authors reviewed the manuscript and agreed with the content.

Funding This work was supported by the National Key R&D Program of China (Grant No. 2021YFF0901004), the National Natural Science Foundation of China (Grant Nos.62072220, 61502215), the Scientific Research Project of the Educational Department of Liaoning Province (Grant No. LJKZ0094), and the Central Government Guides Local Science and Technology Development Foundation Project of Liaoning Province (Grant No.2022JH6/100100032).

Data availability Not applicable

Code Availability Not applicable

Declarations

Human and Animal Ethics Not applicable

Competing interests The authors declare that they have no competing interests.

Ethics approval and consent to participate Not applicable

Consent for Publication Not applicable

References

1. AHMAD R W, HASAN H, YAQOOB I, et al. Blockchain for Aerospace and Defense: Opportunities and Open Research Challenges[J]. Computers & Industrial Engineering, 2020.
2. TIAN G H, HU Y H, CHEN X F. Research progress of blockchain system attack and defense technology[J]. Journal of software,2021,32(5):0-0.
3. WAN P K, HUANG L, HOLTSKOG H. Blockchain-enabled Information Sharing within a Supply Chain: A Systematic Literature Review[J]. IEEE Access, 2020, PP(99):1-1.
4. CAI T, LIN H, CHEN W H, et al. Blockchain-enabled Efficient Data Sharing Scheme for Internet of Things [J]. Journal of Software, 2021,32(4):11-0.

5．GE Y, HUANG C L, CHEN M, et al. HACCP Quality Traceability Model and System Implementation Based on Blockchain [J]. Transactions of the Chinese Society for Agricultural Machinery, 2021(3).
6．BARTOLETTI M, BRACCIALI A, LANDE S, et al. A general framework for blockchain analytics[J]//Proceedings of the 1st Workshop on Scalable and Resilient Infrastructures for Distributed Ledgers, Las Vegas, Nevada, USA, 2017:11-15.
7．HE P, YU G, ZHANG Y F, et al. Review of Blockchain Technology and Application [J]. Computer Science,2017,44(4):1-7.
8．IEMIEUX V L. Trusting records：is blockchain technology the answer?[J].Records Management Journal,2016, 26(2):110-139.
9．WANG S, DINH T A, LIN Q, et al. ForkBase: An Efficient Storage Engine for Blockchain and Forkable Applications[J]. Proceedings of the VLDB Endowment, 2018, 11(10).
10．HALPIN H, PIEKARSKA M, Introduction to Security and Privacy on the Blockchain[C]// IEEE European Symposium on Security & Privacy Workshops. IEEE, 2017:1-3.
11．KRAFT D. Difficulty control for blockchain-based consensus systems[J]. Peer-to-Peer Networking and Applications, 2016,9(2): 397−413.
12．DANNEN C. Introducing Ethereum and Solidity: Foundations of Cryptocurrency and Blockchain Programming for Beginners[M]. Apress,2017.
13．DINH T T A, WANG J, CHEN G, et al. BLOCKBENCH: a framework for analyzing private blockchains[J]. //International Conference on Management of Data.2017:1085-1100.
14．CAO Y, MIAO Z G. An Improved Apriori Algorithm for Frequent Terms Optimization Based on Vector Matrix [J]. Journal of Jilin University (Natural Science Edition), 2016,54 (02):349-353.
15．BIAN G Q, WANG Y. An Improved Apriori Algorithm Based on Matrix and Weight [J]. Microelectronics & Computer, 2017,34 (1): 137-140.
16．SHAO Q F, JIN C Q, ZHANG Z, et al. Blockchain Technology: Architecture and Progress [J]. Chinese Journal of Computers, 2018, 041(005):969-988.
17．WANG Q G, HE P, NIE T Z, et al. Overview of Blockchain System Data Storage and Query Technology [J]. Computer Science, 2018,45 (012): 12-18.
18．WANG Z H, LIU P Z, SONG C B, et al. Research on Flexible Trustable Traceability System of Agricultural Products Based on Blockchain. Computer Engineering, 2020, v.46; No.521(12):319-326.
19．JIN H, DAI X, XIAO J. Towards a Novel Architecture for Enabling Interoperability amongst Multiple Blockchains[C] //International Conference on Distributed Computing Systems. IEEE Computer Society,2018:1203-1211.
20．LI Y, ZHENG K, YAN Y, et al. EtherQL: a query layer for blockchain system[C] // International Conference on Database Systems for Advanced Applications. 2017.
21．QI L, XIAO Z. Research on trust mechanism of cooperation innovation with big data processing based on blockchain[J]. Eurasip Journal on Wireless Communications & Networking, 2019, 2019(1).
22．LUU L, NARAYANAN V, ZHENG C, et al, A secure sharding protocol for open blockchains[C]// the 2016 ACM SIGSAC Conference. ACM, 2016.
23．LIU L J. Research and application of improved Apriori algorithm [J]. Computer Engineering and Design, 2017(12):142-146.
24．HAN H. Blockchain infiltration into data transaction to solve traceability and authorization "pain points" [J]. Communications World, 2017(19).

Authors and Affiliations

Junlu Wang[1] • Su Li[1] • Wanting Ji[1] • Dong Li[1] • Baoyan Song[1]

Baoyan Song

bysong@lnu.edu.cn

Junlu Wang

wangjunlu@lnu.edu.cn

Su Li

liisuu@163.com

Wanting Ji

wanting.ji@lnu.edu.cn

Dong Li

dongli@lnu.edu.cn
[1] School of Information, Liaoning University, Shenyang, 110036, Liaoning, China