Review on Information Privacy Protection based on Blockchain Principle

Shihua Huang *

School of Management Science and Engineering, Anhui University of Finance and Economics, Bengbu 233030, China

* Corresponding author Email: 2532400875@qq.com

Abstract: As a distributed general ledger technology, a basic platform of smart contract, and a new distributed computing paradigm, blockchain is bound to have an impact on the development of all walks of life, and drive a new round of technology and application changes. To have an impact on the development of all walks of life and drive a new round of technology and application changes, blockchain technology improves efficiency, reduces costs, and enhances data security. However, it also faces serious information security problems. However, while blockchain improves efficiency, reduces costs and improving data security, block technology also faces serious information privacy leakage problems. promote the research, development and application of blockchain in information privacy protection. This paper uses the literature analysis method to analyze the advantages and disadvantages of the blockchain in the information privacy protection from the aspects of the blockchain privacy protection. This paper uses the literature analysis method to analyze the advantages and disadvantages of the blockchain in the information privacy protection from the aspects of the blockchain privacy protection, analyze the privacy leakage threat of the blockchain in the network layer, transaction layer and application layer, and summarizes the main technologies of the blockchain in the information privacy protection of data users.

Keywords: Blockchain; Data Privacy; Information Privacy; Information Protection.

1. Introduction

Recently, Bitcoin[1], Monero[1], Dash[2], and other cryptocurrencies have taken the world by storm, and the blockchain technology involved has attracted widespread global attention. Blockchain is considered a breakthrough technology because of its ability to transform the Internet of Information into the Internet of Value[3]. Originally designed to enable decentralized and unencumbered electronic payments, blockchain combines various technologies such as Merkle trees[4], PoW (Proof of Work) proofs, and others. Blockchain technology utilizes various technologies such as Merkle trees and PoW (Proof of Work) proofs. It also incorporates cryptographic techniques like zero-knowledge proofs, hash functions, digital encryption, and signatures. This combination forms a new computing paradigm[6] and distributed architecture[7]. The public indication of countries’ attitudes towards Bitcoin and the volatility of the world's economic development have played a significant role in the rapid development of Bitcoin and the increased visibility of blockchain technology. It can be said that Bitcoin is one of the most successful applications of blockchain technology[8].

Along with the emergence of decentralized applications like Ether and the open-source project Hyperledger Fabric, blockchain technology has started to be implemented in various industries, leading to a new wave of technological advancements and application transformations. Blockchain technology is the latest breakthrough in decentralized, secure computing in open networks. From the perspective of data management, Blockchain is a distributed database that stores each transaction record separately in blocks after performing a hash calculation. Nodes then connect the current block to the chain after verifying the transaction. From a security standpoint, blockchain creates and maintains a network through peer-to-peer connections, utilizing cryptography and consensus protocols to safeguard information and data.

Blockchain technology is a systematic integration of various innovative technologies, including distributed databases, consensus mechanisms, P2P networks, cryptographic algorithms, and smart contracts. Its core advantage lies in decentralization and the establishment of a credible process. As a result, it has gained significant attention and widespread concern in the global fields of science and technology, investment, and has been extensively applied in sectors such as healthcare, finance, e-government, and logistics supply chain.

However, while blockchain technology is widely used to address various challenges across industries, it also faces a significant issue of information privacy leakage. This concern has been extensively discussed in recent literature research and is currently a major focus of scholars. The decentralized nature of the blockchain mechanism allows it to avoid the single-point failure of centralized servers that rely on a centralized architecture. It also addresses the issue of data leakage. However, decentralized independent blockchain nodes are required to disclose all transaction records in order to achieve consensus. This clearly increases the risk of information privacy breaches. Starting from blockchain technology, this paper reviews the current status of blockchain applications in information privacy protection and analyzes the advantages and shortcomings of its implementation. The aim is to provide a reference for future research and development in the field of information privacy using blockchain technology.

2. Statistical Analysis of Domestic and International Literature

2.1. Quantitative Statistics of the Literature

After reading and organizing the literature, we found that the attention to blockchain at home and abroad is increasing, and at the same time, the topic of "blockchain privacy" is also
gradually attracting people's attention. Therefore, on December 06, 2021, we conducted a title path search using Web of Science with "blockchain privacy" as the search term, and the time span was 2016-2021, and a total of 310 articles were retrieved, excluding newspapers, magazines, and reports, and screened out 310 articles about "blockchain privacy". Excluding newspapers, magazines and reports, a total of 300 foreign language literature on "blockchain privacy" was screened, including 3 books and 288 journal articles. Using China Knowledge Network (CNKI) with "blockchain privacy" as the search term, a title path search was carried out for the period of 2016-2021, also excluding newspapers, magazines and reports, and a total of 182 articles were selected from the Chinese literature on "blockchain privacy", including 0 books and 288 journal articles. The total number of Chinese literatures on "blockchain privacy" is 182, including 0 books, 104 journals, and 88 theses. Therefore, this study focuses on analyzing 300 foreign and 182 Chinese articles from 2016 to 2021. The statistical results of specific literature publication year are shown in Figure 1.

![Figure 1. Annual distribution of domestic and international blockchain privacy literature](image)

As can be seen from Figure 1, first, domestic and international research on blockchain privacy began in 2016, and because blockchain is an emerging technology, international attention is paid to its development. And as people's self-awareness grows stronger, their concern for the privacy of their personal information gradually increases. This has led to a growing trend of applying new technologies, such as blockchain, to various industries. Consequently, research related to blockchain privacy has been increasing year by year. Secondly, while the overall number of domestic publications on blockchain privacy has been increasing in 2019, it is still significantly lower than the number of publications abroad. This indicates that the majority of blockchain research is still being conducted in foreign countries. Therefore, there is a need for increased efforts in domestic research on emerging technologies. Third, the number of articles on blockchain privacy at home and abroad shows an overall increasing trend. This indicates that blockchain privacy has gradually become a popular research area in academic circles both domestically and internationally. This is closely related to the fact that both the government and enterprises currently place great importance on the application of blockchain.

### 2.2. Journal Distribution of Literature

Through the statistics, it can be seen that in 2016-2021, there are 300 foreign language journals on blockchain research, and 121 of them are published in computer science, accounting for 40.33% of the total literature, ranking the first, and mainly focusing on "IEEE ACCESS", "IEEE INTERNET OF THINGS JOURNAL" and "FUTURE GENERATION COMPUTER SYSTEMS". "IEEE ACCESS", "IEEE INTERNET OF THINGS JOURNAL", and "FUTURE GENERATION COMPUTER SYSTEMS" are three journals where the research was published. Out of the total 95 publications in the field of telecommunications, these journals accounted for 31.67% of the total. These publications were listed in the second category and were primarily focused on the "INFORMATION PROCESSING MANAGEMENT" domain, "INFORMATION PROCESSING MANAGEMENT" and "SECURITY AND COMMUNICATION NETWORKS"; 70 articles were published in engineering electrical and electronic journals, accounting for 23.33%. 70 articles were published in engineering, electrical and electronic journals, accounting for 23.33%, which is the third, and mainly concentrated in "IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS", "IEEE INDUSTRIAL ELECTRONICS MAGAZINE". See Table 1 for the types and distribution of key journals on blockchain privacy.

<table>
<thead>
<tr>
<th>Distribution of foreign language journals</th>
<th>Journals with more publications</th>
<th>Total number of publications (articles)</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Science</td>
<td>IEEE ACCESS</td>
<td>121</td>
<td>40.33</td>
</tr>
<tr>
<td></td>
<td>IEEE INTERNET OF THINGS JOURNAL</td>
<td>70</td>
<td>23.33</td>
</tr>
<tr>
<td></td>
<td>future generation container systems</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>information processing management</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>security and communication networks</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>......</td>
<td>95</td>
<td>31.67</td>
</tr>
<tr>
<td></td>
<td>(sth. or sb) else</td>
<td>14</td>
<td>4.67</td>
</tr>
<tr>
<td></td>
<td>(grand) total</td>
<td>300</td>
<td>100</td>
</tr>
</tbody>
</table>

From 2016 to 2021, a total of 182 journals on blockchain privacy research were published in China. Out of these, 92 journals were published in the computer science category, accounting for 50.55% of the total number of journals and ranking first. These journals mainly focused on "computer research and development," "computer application research," and "small microcomputer systems." Additionally, there were 33 information journals, accounting for 18.13% and ranking second, and 30 telecommunications technology journals, accounting for 16.48% and ranking third. It mainly focuses on "Computer Research and Development", "Computer Application Research", and "Small and Microcomputer Systems"; 33 articles are in information journals, accounting for 18.13%, and listed in the second place; 30 articles are in telecommunication technology journals, accounting for 16.48%, and listed in the third place. See Table 2 for the types and distribution of key journals specifically related to blockchain privacy.
This can be seen by comparing Tables 1 and 2. (1) The literature on blockchain privacy by foreign scholars is primarily published in computer science and electronic information journals. Among these, more than 50% of the literature is ranked first in computer science journals. To a certain extent, this indicates that both domestic and foreign computer scholars are increasingly focusing on blockchain privacy. (2) It can also be observed from both domestic and international comprehensive journals that research on blockchain technology is experiencing a cross-disciplinary development trend. The exploration of this issue is not limited to computer science but also requires the involvement of engineering disciplines such as electronic information technology and the integration of knowledge and theoretical systems from other disciplines. This will contribute to the advancement of research on this topic.

2.3. Institutional Distribution of Authors of Literature

Table 3. Statistics on the authors' organizations in domestic and foreign literature

<table>
<thead>
<tr>
<th>Type of organization</th>
<th>Number of foreign publications (articles)</th>
<th>Share of foreign countries (%)</th>
<th>Number of domestic publications (articles)</th>
<th>Domestic share (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schools and colleges</td>
<td>242</td>
<td>80.67</td>
<td>161</td>
<td>88.46</td>
</tr>
<tr>
<td>research organization</td>
<td>48</td>
<td>16.00</td>
<td>15</td>
<td>8.24</td>
</tr>
<tr>
<td>corporations</td>
<td>10</td>
<td>3.33</td>
<td>6</td>
<td>3.30</td>
</tr>
<tr>
<td>add up the total</td>
<td>300</td>
<td>100</td>
<td>182</td>
<td>100</td>
</tr>
</tbody>
</table>

Through statistical analysis of the type of institution of the authors (first author only) in 300 foreign literature publications from 2016 to 2021, the following distribution was observed: 242 authors (80.67% of all authors) were affiliated with university faculties and departments, 48 authors (16.00% of all authors) were affiliated with scientific research institutions, and 10 authors (3.33% of all authors) were affiliated with enterprises. Statistical analysis of the institutional affiliation of the authors (first author only) of 182 domestic literature pieces from 2016 to 2021 reveals that 161 authors from university faculties comprised 88.46% of all authors. Additionally, 15 authors from scientific research institutions accounted for 8.24% of all authors, while 6 authors from enterprises represented 3.30% of all authors. See Table 3 for details.

Table 3 shows that the authors of domestic and foreign literature are primarily concentrated in university faculties, followed by scientific research institutions and enterprises. However, the number of authors in these institutions is significantly lower compared to that of university faculties. It is evident that university faculties play a crucial role in researching blockchain privacy. However, it also highlights the need for scholars from enterprise backgrounds to prioritize blockchain privacy. Additionally, addressing the issue of blockchain privacy requires not only systematic research but also practical collaboration between universities, research institutions, and enterprises. This collaboration is essential to enhance the practical value of the theory. Therefore, it is imperative to actively encourage cooperation between enterprises, research institutions, and universities to effectively promote the development and advancement of research in this discipline. To better promote the development and improvement of disciplinary research.

2.4. Thematic Distribution of Literature

Through the statistical analysis of 300 foreign literature sources, the author found that foreign scholars’ research on blockchain privacy primarily focuses on topics such as blockchain, privacy protection, privacy protection technology, and information protection, as shown in Table 4. Similarly, through the statistical analysis of 182 domestic literature sources, the author found that domestic scholars’ research on blockchain privacy also revolves around topics such as blockchain, privacy protection, privacy protection technology, and information protection, as indicated in Table 5.

Table 4. Distribution of topics in foreign blockchain privacy literature

<table>
<thead>
<tr>
<th>Foreign Language Subject Classification</th>
<th>Number of communications (articles)</th>
<th>Share (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>blockchain</td>
<td>116</td>
<td>38.67</td>
</tr>
<tr>
<td>Privacy</td>
<td>99</td>
<td>33.00</td>
</tr>
<tr>
<td>Privacy Protection Technology</td>
<td>73</td>
<td>24.33</td>
</tr>
<tr>
<td>Information protection</td>
<td>12</td>
<td>4.00</td>
</tr>
<tr>
<td>(grand) total</td>
<td>300</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 5. Topic distribution of domestic blockchain privacy literature

<table>
<thead>
<tr>
<th>Chinese Topic Classification</th>
<th>Number of communications (articles)</th>
<th>Share (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>blockchain</td>
<td>72</td>
<td>39.56</td>
</tr>
<tr>
<td>Privacy Protection Technology</td>
<td>59</td>
<td>32.41</td>
</tr>
<tr>
<td>Privacy</td>
<td>41</td>
<td>22.53</td>
</tr>
<tr>
<td>Information protection</td>
<td>10</td>
<td>5.49</td>
</tr>
<tr>
<td>(grand) total</td>
<td>182</td>
<td>100</td>
</tr>
</tbody>
</table>

By summarizing, consolidating, and conducting further analysis of the literature on blockchain privacy from 2016 to 2021, both domestically and internationally, the author discovered differences in research focus. Foreign countries tend to focus more on blockchain and its related technologies, while domestic countries demonstrate a greater interest in the application of blockchain privacy technology. This shows that
foreign research focuses more on the technical level, while domestic research is more inclined towards the practical application level.

3. Blockchain Information Privacy Protection

3.1. Advantages and Shortcomings of Blockchain Information Privacy Protection

The characteristics of blockchain technology enable it to solve the problem of information privacy leakage faced by some centralized servers. As a result, it has been widely applied to various scenarios that require information privacy protection, including healthcare information systems. However, it is also due to its decentralized and distributed storage mechanism that new information privacy protection problems arise.

Advantages of Blockchain Technology in Information Protection:

1) P2P networks prevent network eavesdropping. A blockchain network is a peer-to-peer (P2P) network structure. When a transaction is required between nodes, the initiator of the transaction first sends the transaction message to a node. The node verifies it and forwards the transaction message to neighboring nodes. The receiving node then continues this process until the message is broadcasted throughout the network. The transaction receiver eventually receives the transaction message from the network without directly communicating with the initiator. Therefore, it is difficult for an attacker to eavesdrop on the true source and destination of transaction messages in the blockchain network using the previous method of intercepting network traffic to uncover the communication relationship between users.

2) Support for anonymous transactions. Blockchain transactions are usually conducted using "addresses", which are similar to bank card accounts. Addresses are created and stored by the user without the involvement of a third party, and the way blockchain addresses are generated makes collisions between addresses very unlikely, greatly enhancing the anonymity of transactions. For example, the private key space corresponding to a Bitcoin address is 2256, which is 1077 in decimal, and the visible universe is estimated to contain only 1080 atoms, so there is enough address space in the Bitcoin system to support a one-time address strategy.[9] Therefore, there is enough address space in the Bitcoin system to support a one-time address strategy.

3) Decentralized architecture to cope with cyber-attacks. Programs that utilize blockchain technology typically employ decentralized architectures. In these architectures, users' information is stored in mutually independent nodes, eliminating the need for third-party servers to manage the storage of accounts, passwords, and other sensitive information. This approach significantly reduces the risk of traditional servers being attacked and information being leaked.

Flaws in blockchain technology for information security:

1) Nodes are vulnerable to attacks. Nodes in a blockchain are typically users' personal computers, which have lower performance but are more resistant to attacks compared to dedicated servers in traditional network architectures. In addition, nodes in a blockchain network are geographically dispersed and have equal status. Unlike in a traditional centralized architecture, it is not possible to target specific nodes for protection. This means that attackers can identify vulnerable nodes and launch attacks, potentially compromising the entire blockchain network.

2) Transaction correlations are easily exploited. All transactions in the blockchain are stored in the public global ledger. An attacker can easily obtain all transaction information and gradually erode anonymity in the blockchain by analyzing the correlations in the transactions, potentially even uncovering the user's true identity.

3) Facing multiple security threats. Blockchain technology is still in the development stage, and there are multiple security flaws that could potentially pose threats to its security. For example, DAO, an Ethereum-based crowdfunding application, had over 3 million Ether coins stolen as a result of hacking attacks[10].

3.2. Threats to Privacy on the Blockchain

Blockchain technology has privacy leakage problems in practical application scenarios. In summary, these problems can be categorized into three aspects: data, identity, and transaction. This paper utilizes the blockchain architecture to classify privacy threats in blockchain into three categories: privacy threats at the network layer, privacy threats at the transaction layer, and privacy threats at the application layer[11].

3.2.1. Privacy Threats at the Network Layer

The network layer encompasses the entire process of underlying communication, including the setup mode of blockchain nodes, the mechanism for node communication, and the mechanism for data transmission. Taking Bitcoin as an example, there are several privacy threats that primarily occur at the blockchain network layer, as outlined below.

1) IP address leakage. An attacker can deploy probe nodes to easily detect the IP address of a node.

2) Node Topology Relationship Exposure. Attackers use probe nodes to actively acquire and passively listen to obtain topological relationships between nodes.

3) Leakage of propagated transaction information. Because the transaction information propagated in the network is not encrypted at the network layer, an attacker can easily read the transaction information transmitted through the network.

3.2.2. Privacy Threats at the Transaction Layer

The entire transaction history in the current global blockchain ledger is public. This means that anyone can see how funds are transferred from one pseudonym to another. Additionally, an attacker can link different pseudonyms of the same user together. As a result, the user's data information and knowledge behind the data are highly susceptible to privacy threats.

1) Transaction information is analyzed. An attacker can analyze transaction records and filter valuable information, posing a threat to transaction privacy.

2) Identity information is presumed. Given that the current transaction data volume of the Bitcoin system is approximately 300 GB, encompassing all transaction records since 2008, an attacker can only analyze the transaction data but also potentially deduce the identity information of the trader based on it, provided they possess some background knowledge. On top of that, an attacker can analyze the data in the ledger and steal all the transactions from any account. Even for transaction records involving multiple accounts, if a customer uses multiple accounts to make transactions, an
attacker can still use address clustering techniques to identify transactions belonging to the same user[12]. Even if a customer uses multiple accounts to conduct transactions, an attacker can utilize address clustering to identify multiple accounts belonging to the same user, thereby creating a risk to their identity and privacy.

3.2.3. Privacy Threats at the Application Layer

The application layer involves the external utilization of blockchain technology, which may potentially compromise the privacy of user-related information. This is specified below.

(1) The threat of privacy leakage due to improper use of applications. When users utilize blockchain technology, they may engage in actions that could potentially expose private information due to a lack of understanding of the security mechanisms of blockchain. Consequently, malicious users may employ various methods to obtain this private information and carry out harmful activities, thereby jeopardizing the property security of individual users.

(2) The threat of malicious attacks on application usage. When users utilize websites that offer blockchain services, they are susceptible to malicious attacks and the theft of personal privacy information. This is because these websites may not have implemented adequate and efficient privacy technology protection, thereby posing a potential risk of privacy breach.

3.3. Blockchain Information Privacy Protection Technology

3.3.1. Data-oriented Information Privacy Protection Techniques

(1) Data encryption. Homomorphic encryption[13][14] allows for computation on ciphertext data without decrypting the data, providing a high level of security and enabling computational operations on private data. However, it is known for its inefficiency, high computational overhead, and the challenge of verifying computation results. Secure Multi-Party Computing[15] allows multiple users to collaborate on a computational task in an untrusted network environment without revealing their private inputs.

(2) Data distortion. Differential Privacy[16],[17] involves adding noise to data to provide privacy protection. It can effectively defend against background knowledge attacks and is built on strict mathematical foundations. However, it faces challenges in publishing complex and continuous data, as well as difficulties in distributed differential computation.

(3) Restrictions on publishing. Off-chain storage[18] involves storing complete data off-chain and only keeping part of the information on the public ledger. This approach allows for private storage and can reduce the computing and storage overhead on the blockchain. However, it is important to note that all parties involved still need to maintain the original data. Segregation of Ledgers[19]: Put the ledgers with different privacy requirements on separate distributed ledgers; however, modifications to the blockchain structure may introduce new security concerns. Coalition, Private Chain[20]. The purpose of adding a node access mechanism to the blockchain is to fundamentally prevent unprivileged users from accessing it; however, this reduces the degree of decentralization.

3.3.2. User-oriented Information Privacy Protection Technology

(1) Access Control. Smart Contracts[21] aim to design the access control policy within the smart contract on the blockchain. This approach helps to avoid decision centralization and enables more flexible and automated access control. However, it is important to pay special attention to the security and privacy issues associated with the smart contract itself. Attribute Encryption[22] aims to match the access control policy with a set of attributes in order to achieve more flexible and fine-grained access control. However, bilinear encryption is known to be time-consuming.

(2) Transactions are anonymous. Coin obfuscation mechanism[23] obfuscates the sender and receiver of transactions to improve the privacy of blockchain transactions. However, it needs to be combined with other methods to enhance privacy protection. Zero-knowledge proof[24] validates messages without revealing any valid information, ensuring high privacy. However, it is relatively inefficient. Digital Signature[25],[26],[27][28] Participants are anonymous and can protect transaction information, but there are numerous regulatory challenges. Secure Channel Protocol[29] is an off-chain transaction and on-chain arbitration method that aims to protect privacy and improve transaction throughput. However, it has more restricted business scenarios and requires modifications to the underlying protocols. k-anonymization[30][31] is a method that groups data to achieve small computation overhead and accurate query results. However, it is vulnerable to attacks from background knowledge.

4. Conclusion

While the process of digital transformation in various industries is advancing rapidly, it also brings about numerous privacy and security challenges. The characteristics of blockchain, such as decentralization, immutability, anonymity, and traceability, offer a potential solution for ensuring information privacy and security. This paper begins with an analysis of the existing literature on blockchain privacy, focusing on the importance of protecting information privacy in blockchain and the technical aspects related to information privacy protection. The aim is to contribute to the application of information privacy protection in blockchain technology. Finally, the article presents an outlook on the future development of blockchain information privacy technology.

Blockchain technology has made significant advancements in information privacy research. However, it still faces several challenges in terms of security, privacy, transaction performance, and integration with other privacy protection technologies. These challenges are crucial to address when applying blockchain privacy technology. In the future development of blockchain privacy, we have the following outlook:

Accountability mechanisms for decentralized privacy protection. Privacy protection in blockchain is a "double-edged sword." On the one hand, a well-behaved user expects the blockchain to efficiently protect information, such as their identity and account. On the other hand, some malicious accounts may exploit the privacy-preserving mechanism to carry out illegal transactions. However, the analysis of the existing literature reveals that few studies have explored blockchain accountability mechanisms. This is problematic because accountability mechanisms are necessary in various application scenarios. Therefore, privacy protection in blockchain requires the establishment of accountability mechanisms. In the event that blockchain is used as a criminal platform, it is essential to disclose information, such as the
identity and accounts of malicious users, in order to alert other users. In addition, some scenarios require the assessment of user trust or the auditing of data, but blockchain lacks a strong and trustworthy centralized privacy assurance mechanism. Considering the need for accountability and the challenges it presents, decentralized privacy protection with accountability will be an important area of future research.

Smart Contract Privacy Protection. Smart contracts are a crucial technology for developing diverse blockchain-based applications. These applications necessitate miners to verify the accuracy of contract execution results, which poses challenges in safeguarding contract privacy. The contract itself and the data generated during contract execution should remain invisible to everyone except the contract creator. Currently, verifiable computing, secure multiparty computation (SMC), and trusted hardware have been adopted. From the analysis of the article, it is evident that they are not fully suitable for blockchain implementation due to their lack of efficiency and reliance on centralized parties. Therefore, there is a need to protect contract privacy in a decentralized and efficient manner, making it practical for deployment in the real world.

Acknowledgments
This study was funded by the Graduate Student Innovation Fund (ACYC2021430) of Anhui University of Finance and Economics.

References