A Survey of Blockchain-Based Identity Anonymity Research

Fa Fu[1], Gaoshang Lu[1], Jianqiang Huang[2] and Thomas Dreibholz[3]

[1]Hainan University, Haikou Hainan 570228, China

[2]China Telecom Corporation Limited Hainan Branch, Haikou Hainan 570125, China

[3]Simula Metropolitan Centre for Digital Engineering A/S Pilestredet 52, 0167 Oslo, Norway

fufa@hainanu.edu.cn

Abstract. With the booming development of blockchain technology, blockchain-based data transactions have been applied in many fields such as finance, healthcare and logistics. It can help users to realize data transactions and management more conveniently, securely, transparently and efficiently. However, there is a certain problem of identity privacy leakage when data transactions are conducted on blockchain. Therefore, the issue of user identity privacy protection has become the core issue of data transactions on the blockchain, which is crucial to the sustainable development and wide application of the blockchain. This paper discusses the privacy protection in the process of data transactions on blockchain in terms of user identity anonymity, introduces and analyzes in detail the current research status and implementation technologies for realizing identity anonymity on blockchain, explains the threats and challenges for realizing identity anonymity, analyzes the existing problems, and gives an outlook and summary of the future research directions for realizing identity anonymity on blockchain.

Key Words: Blockchain, Identity Anonymous, Data Transaction.

1 Introduction

A key feature of blockchain technology is decentralization, which allows participants to conduct transactions without a central control authority and also means that transactions and data records are open and transparent to all node chains, participants' identity data may be exposed to others. However, since some private information may be involved, such as transaction amounts, medical consultation records, and trade secrets, such users want to protect their identity information from disclosure. However, when using blockchain addresses to participate in

blockchain business, users need to frequently perform input and output operations. Analyzing this information can indirectly associate the true user identity of the account address, which poses a threat to privacy leakage for blockchain participants' accounts. Therefore, there is still a risk of leaking sensitive user identity information in blockchain transactions, such as the propagation trajectory of the transaction at the network layer, this information may be used to infer the true identity of the blockchain address. Therefore, how to protect user identity privacy data, prevent user identity information from being identified and leaked, and achieve identity privacy to protect users' real identity and private information is crucial for the sustainability and wide application of the blockchain.

2 Blockchain Technology

2.1 Blockchain

User identity privacy refers to mapping the real-world user's real identity to his or her address information on the blockchain [1], which contains personal information such as the user's identity and address that are recorded in detail and not publicly available. Among them, the user identity information refers to the basic personal information entered by the user when applying for access to the blockchain [2], while the user address information refers to the place where the individual belongs when participating in the blockchain data storage and transmission, and usually contains two accounts used for input and output in transactions. To protect identity anonymity, users usually use random addresses or pseudonyms for transactions in the blockchain [3]. A blockchain address is a pseudonym used by users in the blockchain system and is usually used as an account number for input and output during transactions. Compared to traditional account numbers, blockchain addresses are superior in concealing the user's identity [4].

2.2 Smart Contracts

Smart contracts are automated contracts that enable the signing and execution of contracts on the blockchain, a concept first introduced by Nick Szabo in 1996 in his paper "Smart Contracts: Building Blocks for Digital Markets". In Szabo's definition, a smart contract is an automatically executed contract based on a computer protocol that represents and enforces the terms of the contract in digital code. These codes allow for automated and decentralized execution of the contract and protect the security and privacy of the contract through encryption. However, in the late 1990s, computer technology was not mature enough to implement the concept of smart contracts. It was not until 2009 that the advent of Bitcoin made smart contracts possible. Born in

2013, Ether has revolutionized the face of smart contracts. Ether introduced a high-level programming language called Solidity, making it easier for developers to write more complex smart contracts and implement more features on the Ether blockchain. Since the birth of Ether, the applications of smart contracts have been expanding. Smart contracts have also become one of the most representative blockchain technologies.

3 Research and Analysis of Identity Anonymization Techniques

Currently, the main technologies applied in blockchain to achieve identity anonymity include blind signature, group signature, and aggregate signature technologies. In this section, we will comprehensively analyze the advantages and disadvantages of the main signature technologies in the blockchain.

3.1 Blind Signature Technology

Blind signature is a digital signature technique that allows a signer to sign a message without knowing its content [5]. It has a wide range of applications in privacy protection and authentication authorization, especially in electronic cash, digital certificates, and anonymity networks. Rivest, R. L proposed the RSA blind signature scheme in 1982 [6], which is an implementation of blind signatures based on the RSA cryptographic algorithm with better security and efficiency. This is another important contribution of the blind signature technique. Chaum, D proposed the original blind signature scheme [7] in 1983, i.e., using a randomization technique so that the signer cannot know the content of the signature, thus enabling untraceable payments. This is one of the seminal works in blind signature techniques.

The basic principle of blind signature is that first, the signer Bob generates a pair of public and private keys [8] and publishes the public keys. Then, user Alice generates a random number as a "blind factor", and the message M to be signed is blinded using the blind factor, the message M is multiplied by the blind factor to obtain a blind message M'. The signer Bob signs the blind message M' with his private key to get a blind signature Sig(M'), and sends the blind signature Sig(M') to Alice. to obtain the signer's signature Sig(M) for the original message [9]. Since the blind signature process does not require the original message to be revealed to the signer, the privacy of the message can be guaranteed. The flowchart of the blind signature is shown in Figure 1.

First, blind signatures have strong privacy; during the blind signature process, the signer does not know the specific content of the message, and thus the privacy of the message can be guaranteed [10]. Second, blind signatures have strong anonymity, and users can obtain the

signer's signature without revealing their identity, thus achieving anonymous authentication authorization. Finally, blind signatures also have high security; blind signatures have the same security as ordinary digital signatures [11], i.e., they prevent forgery and tampering. However, in addition to this, blind signatures also have some disadvantages, such as slow processing speed, blind signature processing requires blind and anti-blind operations, and thus is slower compared to ordinary digital signatures. Secondly, the complexity of the blind signature operation is high, and blind signature technology is more complex than other digital signature technologies, requiring more calculations and communications. Finally, blind signatures are irrevocable, i.e., once the signer signs, the signature cannot be revoked. If the identity of the user is exposed, the reputation of the signer may be damaged.

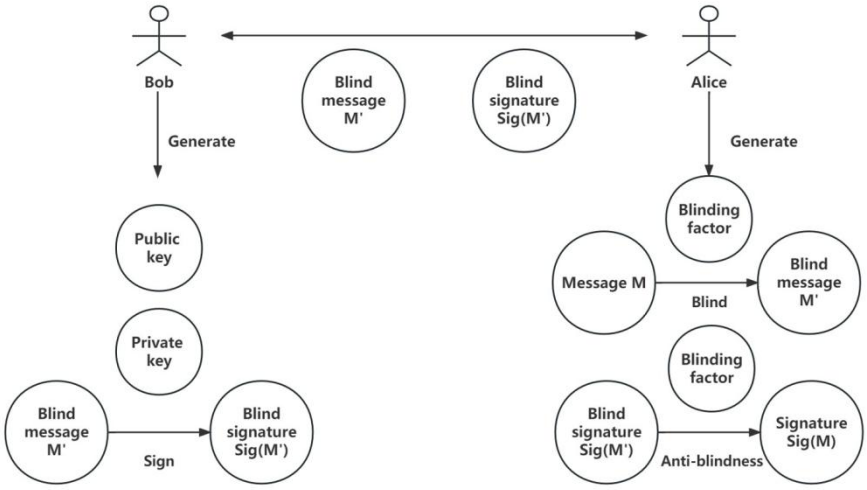

Fig. 1. Schematic diagram of the blind signature process.

3.2 Group Signature Technology

A group signature is a digital signature mechanism used to verify the integrity and origin of a message and to prove that a particular signer belongs to a specific group. Unlike ordinary digital signatures, group signatures allow any member of a group to sign a message [12] while maintaining individual privacy. In simple terms, a group signature is a digital signature scheme that decouples the signature of a group from the identity information of a single individual.

David Cham first introduced the concept of group signatures in 1991 [13] and introduced a cryptography-based group signature scheme that allows a group of members to publish a message using a group signature without revealing the identity of the individual. The scheme

also has a revocation function, i.e., the signer's signature can be revoked when necessary. Ronald Cramer proposed a multi-authority election scheme based on group signatures in 1994 [14], which employs multiple authorities to enhance the security and reliability of the scheme. Jan Camenisch proposed an efficient group signature scheme based on cryptography in 2004 [15], where the signature length of the scheme is independent of the swarm size and is only related to the security parameters. The scheme is efficient and secure and supports the revocation of the signature function.

A group contains multiple members, who together form the group. A member of the group signs the message, and members outside the group can verify that the message has been signed by the group, but they do not know exactly which member has signed it. This approach conceals the true signature identity and achieves the unity of anonymity and super visibility. The signature process of the group signature is shown in Figure 2.

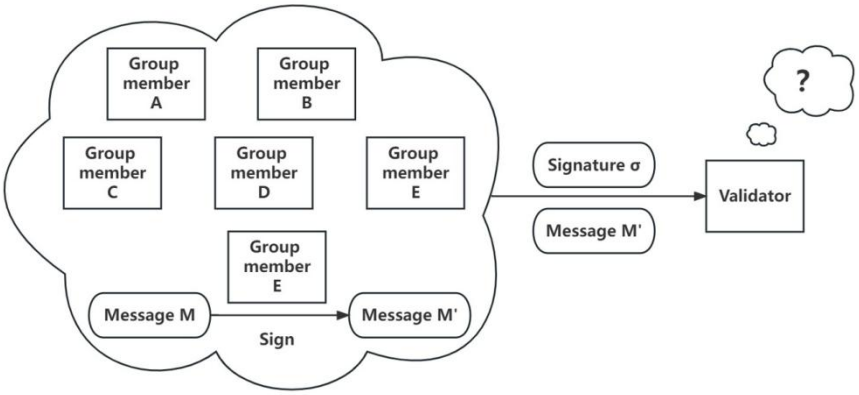

Fig. 2. Schematic diagram of the group signature process.

Group signature techniques have strong anonymity and do not require the identity of the signer to be revealed, so the signer can remain anonymous, which is important in cases where privacy needs to be protected. Group signature techniques have verifiability, i.e., the recipient can verify that the signature belongs to the group and verify the integrity and origin of the message, which ensures the authenticity and trustworthiness of the message. And finally, group signature techniques have non-repudiation i.e., signers cannot deny the messages they sign [16]. This is because the signature mechanism makes the signature unforgeable and the signers cannot claim that they did not sign the message. In addition to this, group signature techniques have some disadvantages, such as the possibility of abuse, as group signatures can be used for criminal activities or other unethical practices due to the anonymity of the signers. The signers need to be trusted, and the validity of group signatures depends on the trust of the signers. If one or more of

the signers behave maliciously, it may negatively affect the validity of the signature. It is difficult to revoke. Unlike ordinary digital signatures, group signature technology is difficult to be revoked because the identity of the signer is anonymous, and if malicious behavior occurs among the signers, it is difficult to find the responsible person and revoke the signature.

3.3 Aggregate Signature Technology

Aggregate signature is a digital signature mechanism that can significantly reduce transaction storage space and transmission costs, and improve verification efficiency. Careful consideration needs to be given to usage scenarios and signer trust when using this technology. It allows multiple signers to sign the same message and aggregates these signatures into a single signature. In simple terms, aggregated signatures are digital signature schemes that aggregate multiple signatures into a single signature [17] and are mainly used to achieve bulk verification of transactions. Dan Boneh et al. proposed a bilinear mapping-based aggregated signature scheme [18] in 2003, which is not only efficient but also verifies the signer identity and signature integrity.

Signer A first hashes the message M to get the message digest M', and then signs the message digest M' with his private key to get the signature σA, other signers also hash and sign their respective messages in this way to get the signature σB, signature σC, etc. All signers send their signatures to a centralized aggregator. The Aggregator combines all signatures into one signature and makes the signature public. The flowchart of aggregated signatures is shown in Figure 3.

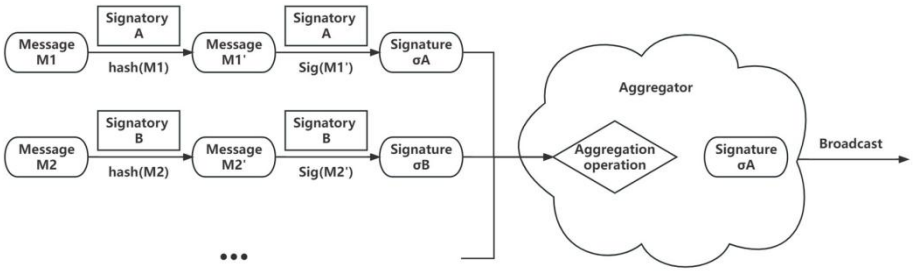

Fig. 3. Schematic diagram of the aggregated signature process.

The aggregated signatures have some advantages. Firstly, it can reduce transaction storage space and transmission costs. Aggregated signatures allow multiple signers to aggregate their signatures into a single signature, which greatly reduces the storage space and transmission costs of the transaction. Secondly, aggregated signature technology also improves verification

efficiency. A single signature of an aggregated signature can reduce the verification workload because the verifier only needs to verify one signature instead of verifying multiple. Finally, the aggregated signature technique increases privacy protection, as aggregated signatures can add anonymity and privacy protection to the signer since they can aggregate multiple signatures into a single signature. In addition, aggregated signatures rely on the trust of all signers, and their validity depends on the integrity and security of all signers. If one of the signers acts maliciously, it may negatively affect the validity of the entire signature. Unlike traditional digital signatures, aggregated signatures are difficult to be revoked, because signers cannot revoke their signatures individually, if one of the signers behaves maliciously, revoking the entire signature may be very difficult. Finally, aggregated signatures may require more complex implementations and higher computational costs, and therefore may not be suitable for use in certain scenarios. According to the above analysis, the advantages and disadvantages of the three signature technologies are summarized in Table 1.

Table 1. Comparison of advantages and disadvantages of three technologies.

	Advantages	Deficiencies
Blind Signature	High privacy Strong anonymity High security	Slow speed High complexity Irrevocable
Group Signature	Strong anonymity Verifiability Non-deniability	Slow speed Unprecedented overheads Large signature length
Aggregate Signature	Reduce transaction storage space and transmission costs Improve validation efficiency Increase privacy protection	Potential for abuse Signers need to be trusted Difficult to revoke

4 Future Research Directions

Through the comparative analysis of different blockchain identity anonymity technologies, we can see that many researchers have proposed various identity anonymity technologies on the blockchain to guarantee the privacy and security of users, but there are still several aspects that need further research.

A) Performance problem: Since all data on the blockchain is public, achieving anonymity requires broadcasting encrypted transactions in the network and waiting for some time for each identity verification, which can increase the burden of network transmission and computation and lead to performance degradation. As proposed in the literature [19] based on homomorphic encryption, each participant in this scheme needs to perform a large number of encryption and decryption operations, which also affects the performance of the system due to the slow encryption and decryption speed of homomorphic encryption. Then there is a scheme based on the obfuscation technique proposed in the literature [20], in this scheme, all participants need to perform obfuscation operations, and the obfuscation operations consume a large amount of computational resources, which also affects the performance of the system. Therefore future research work needs to seek more efficient anonymity guarantee schemes and explore more efficient consensus algorithms to improve transaction processing speed. For example, using zero-knowledge proofs to protect privacy [21] without using homomorphic encryption or obfuscation techniques can achieve efficient privacy protection with high performance, and using cryptographic multi-party computation to protect privacy [22] can compute the corresponding results without exposing the original data, and the performance can be improved by parallel computation. More research is still needed on performance optimization and evaluation methods.

B) Implementing identity anonymity may involve legal compliance issues and anonymous identities may be used for illegal activities. Therefore, to achieve sustainable development of blockchain, future research efforts need to target technical means and solutions to achieve privacy protection while achieving technical controllability, such as using identity to authenticate and authorize participants, using traceability to track participants' behavior, and helping regulators identify illegal activities through ways and means such as government certification and blockchain identity certification agencies. Thus, how to develop regulatory standards to further ensure the legitimacy and transparency of data usage to avoid data misuse and mishandling is an important research issue.

C) Compatibility issues: Implementing anonymity protection in current blockchain technologies may encounter compatibility issues. Public data on the blockchain can improve transparency and trust, but some sensitive data, just like the privacy of users need to be protected. This requires appropriate encryption measures to ensure data security and privacy protection while keeping the data open. Therefore, future research work needs to develop more compatible blockchain technologies to solve this problem, such as promoting trusted blockchain technologies, such as federated chains and side chains, to meet the demand for identity anonymity in different scenarios.

5 Summary

This paper compares and contrasts different technologies of protecting identity anonymity for data transactions on the blockchain, analyzes the advantages and disadvantages of each technology and the applicable environment, and provides an outlook on the future direction of implementing identity anonymity for data transactions on the blockchain, to help researchers quickly and comprehensively understand the basic content and development trend of blockchain identity anonymity technology and future research directions. With the maturity of blockchain technology and its wide application in various industries, the realization of identity anonymity is of great research significance for the sustainable development of blockchain, and we still need to continue to study this area and create a more perfect and practical identity anonymity solution.

Acknowledgment

This research was funded by Haikou Science and Technology Plan Project(2022-007)

References

1. Yu, X.: Blockchain privacy protection key technology research and application. Information Technology 3(5), 36-50 (2020).
2. Liu, H. O., Zhou, Y. Y., Zhou, X., et al.: A Review of Blockchain Privacy Threats and Protection Mechanisms Research. Computer Integrated Manufacturing System 3(9), 1-25 (2023).
3. Yu, P.: Blockchain-based Research on Privacy Data Protection and Sharing of Electronic Medical Records. Information Technology 1(8), 25-41 (2020).
4. Song, Y. C., Ning, X. Y.: Blockchain technology risk assessment and control. Finance & Accounting Monthly 14(9), 124-140 (2021).
5. Wu, Y. X.: Research on group blind signature and multi-bank electronic cash system. Information Technology 11(2), 31-45 (2014).
6. Rivest, R. L., Shamir, A., Tauman, Y.: How to Leak a Secret. In: Boyd, C. (eds.) ASIACRYPT 2001, LNCS, vol. 2248, pp. 552–565. Springer, Heidelberg (2001).
7. Chaum, D.: Blind Signatures for Untraceable Payments. Advances in Cryptology 15(2), 19-22 (1983).
8. Wang, Y.: Research on Automatic Trust Negotiation Protocol Based on Secure multi-party computation. Information Technology 6(1), 49-65 (2012).
9. Liu, D.: A Blind Signature Based on Combined Public Key Cryptography. Fujian Computer 31(2), 21-32 (2015).

10. Gong, Z. Y.: IoT privacy protection based on partial blind signature algorithm. Modern Trade Industry 41(25), 111-123 (2020).
11. He, B.: A study of forward-secure proxy blind signature scheme. Information Technology 3(2), 23-39 (2014).
12. Lei, Y. C.: Blockchain privacy protection scheme based on group signature. Information Technology 5(3), 42-58 (2020).
13. Chaum, D., Heyst, EV.: Group Signatures. Advances in Cryptology - EUROCRYPT '91, Lecture Notes in Computer Science 547, 257-265 (1991).
14. Cramer, R., Gennaro, R., Schoenmakers, B.: A Secure and Optimally Efficient Multi-Authority Election Scheme. Transactions on Emerging Telecommunications Technologies 8(5), 481-490 (2012).
15. Camenisch, J.: Efficient Group Signature Schemes for Large Group. Advances in Cryptology -CRYPTO'97, Lecture Notes in Computer Science 1294, 410-424 (1997).
16. Lu, D. J., Wang, Y.: An identity-based gated proxy signature scheme. Basic Science 26(01), 1-14 (2010).
17. An, T.: Research and Application of Data Security Privacy Protection Methods in Cloud Environment. Information Technology 5(9), 22-38 (2020).
18. Boneh, D., Gentry, C., Lynn, B., et al.: Aggregate and Verifiably Encrypted Signatures from Bilinear Maps. In: Biham, E. (eds.) EUROCRYPT 2003, LNCS, vol. 2656, pp. 416–432, Springer, Heidelberg (2003).
19. Miers, I., Garman, C., Green, M., et al.: Zerocoin: Anonymous Distributed E-Cash from Bitcoin. In: Symposium on Security and Privacy, pp. 397-411. IEEE, San Francisco (2013).
20. Bonneau, J., Narayanan, A., Miller, A., et al.: Mixcoin: Anonymity for Bitcoin with accountable mixes. In: Christin, N., Safavi-Naini, R,. (eds.) International Financial Cryptography Association 2014, LNCS, vol. 8437, pp. 486-504. Springer, Berlin, Heidelberg (2014).
21. Bowe, S., Chiesa, A., Green, M., et al.: ZEXE: Enabling Decentralized Private Computation. In: Symposium on Security and Privacy, pp. 947-964. IEEE, San Francisco (2020).
22. Damgård, I., Keller, M., Larraia, E., et al.: Practical Covertly Secure MPC for Dishonest Majority – or: Breaking the SPDZ Limits. Advances in Cryptology - EUROCRYPT 2013, Lecture Notes in Computer Science 7887, 463-480 (2013).