# An IPFS Privacy Storage Sharing Scheme Based on SM4 Algorithm

Jiazhen Song[1], Yutong Li[2], Haori Lu[2], Jingrong Wang[2], and Peng Nie[3](✉)

[1] College of Computer Science, Nankai University, Tianjin 300350, China
2013904@mail.nankai.edu.cn
[2] College of Cyber Science, Nankai University, Tianjin 300350, China
{1911551,1913106,1911569}@mail.nankai.edu.cn
[3] College of Cyber Science and Tianjin Key Laboratory of Network and Data Security Technology, Nankai University, Tianjin 300350, China
niepeng@nankai.edu.cn

**Abstract.** Nowadays, based on the characteristics of immutability and decentralization, IPFS as a distributed file system is applied in more and more scenarios. However, the storage mode of IPFS is designed for all Internet users and is open. Therefore, the protection of personal privacy data stored on IPFS and the sharing of privacy information within a specific range have become the problems that need to be solved in the current application of IPFS. Based on the above requirements, this paper proposes a scheme based on SM4 domestic commercial security algorithm, which realizes the storage of private information in the form of ciphertext on IPFS, ensures its privacy and realizes the sharing of plaintext content only within the specified range. After analysis, our proposed scheme has reached a high level in algorithm performance and process security. And the scheme has been applied to the self-realized rumour management platform, which shows that it has a wide application prospect.

**Keywords:** IPFS · SM4 domestic commercial security algorithm · Privacy protection

## 1 Introduction

In recent years, IPFS technology has developed rapidly. IPFS is the abbreviation of interplanetary file system (IPFS). It is a global, point-to-point distributed version file system that attempts to connect all computing devices with the same file system. IPFS has the advantages of immutability and decentralization [8]. It is widely used in many fields, especially in the scenario of the combination of IPFS and blockchain. Because IPFS allows all nodes to access and find information. This information may contain personal privacy, which is related to privacy protection [9]. An effective solution is to use the national secret algorithm to encrypt the private information, and then upload it to IPFS. Only the trusted information applicant can obtain the key, and then decrypt and view the plaintext information.


Z. Zhan et al. (Eds.): EIMSS 2022, AHCS 7, pp. 641–649, 2023.
https://doi.org/10.2991/978-94-6463-024-4_67

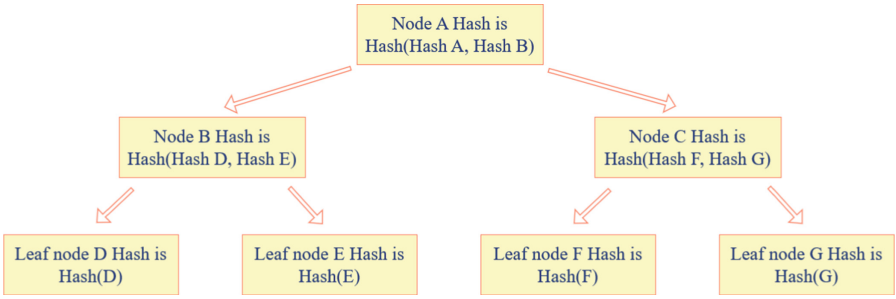


**Fig. 1.** Merkle tree structure.

Therefore, we propose an IPFS privacy data storage and sharing scheme based on domestic commercial SM4 security algorithm, which can simultaneously ensure the privacy of information on IPFS and the sharing within a certain range.

### 1.1 Inter Planetary File System, IPFS

IPFS, the interplanetary file system, is a network transmission protocol designed to create persistent and distributed storage and shared files. This technology is a content addressable peer-to-peer hypermedia distribution protocol. In IPFS network, all nodes will form a distributed file system. With the proposal of IPFS in 2014, many scholars have studied IPFS as a storage and communication platform in various distributed environments. Confais B et al. use IPFS as the storage system to provide object storage services, combined with NAS to provide first-class storage services for fog and edge computing architecture to assist in data sharing [1]. Alam S et al. Proposed an IPFS-based stargate service, which processes and uploads the contents of WARC files to IPFS, so that network files can be permanently stored [7].

The design of IPFS integrates the advantages of distributed hash table, BitTorrent, git and self-certified file system technology. It is considered to be the new generation Internet protocol that is most likely to replace HTTP, and provides a permanent decentralized method of storing files. The core data structure of IFS is Merkle directed acyclic graph.

As the core data structure of IFPS, Merkle directed acyclic graph coordinates the strengths of Merkle tree and directed acyclic graph. Figure 1 shows that Merkle tree consists one root node, a set of internal nodes and a set of leaf nodes. Leaf nodes placed to the bottom contain either data or its hash value, while other nodes contain the hash values of the its children's content.

As shown in Fig. 1, a Merkle tree is usually a binary tree, but multiway tree is also acceptable. Each branch of a Merkle tree is also a Merkle tree. The root nodes of two Merkle trees being the same means that the data represented must be the same. Using root-node verification can greatly reduce the scale of data transmission and the complexity of calculation. Merkle directed acyclic graph integrates the hash calculation method of Merkle tree and the structure of directed acyclic graph. Its structure is shown in Fig. 2. Each node can store data and hash values pointing to its child nodes. Merkle directed acyclic graph has three characteristics, namely 1) content addressing, using

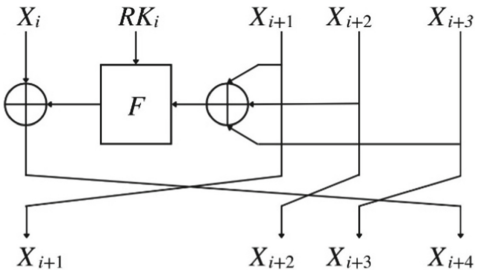


**Fig. 2.** SM4 algorithm process.

multiple hashing to uniquely identify the content of a data block; 2) data authenticity verification by checking the hash value, which empowers confirming whether the data has been tampered with; 3) duplication deletion, since the hash value of data blocks with the same content is the same, it is easy to remove duplicate data and thereafter save storage space.

### 1.2 SM4 Encryption Algorithm

The SM4 algorithm was drafted by Data Assurance & Communication Security Center, CAS, and Commercial Cryptography Testing Center, National Cryptography Administration. This is a symmetric encryption algorithm of the national standard commercial cryptography algorithm series [6]. The SM4 cipher has a key size and a block size of 128 bits each. In SM4 algorithm standard and this paper, byte data vector representing 1 bit is defined. The data vector composed of 32 bit is called "word", and a word is composed of 4 bytes. According to the standard description of SM4 algorithm, both encryption algorithm and key expansion algorithm adopt 32 rounds of nonlinear iterative structure. The process of decryption algorithm is the same as that of encryption algorithm, except that the order of using round key is opposite. The specific process is shown in the Fig. 2.

(1) Key extension

Before encrypting and decrypting the data, first expand the 16 byte encryption key to 32 round keys with a length of 4 bytes $rk_0, rk_1, \ldots, rk_{31}$.

Set encryption key:

$$MK = (MK_0, MK_1, MK_2, MK_3)$$
$$MK_i \in Z_2^{32}, \quad i = 0, 1, 2, 3 \tag{1}$$

The generation method of round key is as follows:

First, according to the primary key MK and the system parameter FK, the calculation is as follows:

$$(K_0, K_1, K_2, K_3) = (MK_0 \oplus FK_0, MK_1 \oplus FK_1, MK_2 \oplus FK_2, MK_3 \oplus FK_3) \tag{2}$$

Then, calculate the round key $rk_i$ for $i = 0, 1, \ldots, 31$ in turn:

$$rk_i = K_{i+4} = K_i \oplus T'(K_{i+1} \oplus K_{i+2} \oplus K_{i+3} \oplus CK_i) \tag{3}$$

The transformation $T'$ is a coincidence transformation, which is composed of a linear transformation L and a nonlinear transformation S. $CK_i$ is the fixed parameter of the algorithm.

(2) Encryption (decryption) algorithm

Suppose the plaintext input be $(X_0, X_1, X_2, X_3) \in (Z_2^{32})^4$, the ciphertext output be $(Y_0, Y_1, Y_2, Y_3) \in (Z_2^{32})^4$, and the round key obtained after key expansion be $rk_0, rk_1, \ldots, rk_{31}$. Then the encryption transformation of each round of the algorithm is:

$$\begin{aligned} X_{i+4} &= F(X_i, X_{i+1}, X_{i+2}, X_{i+3}, rk_i) \\ &= X_i \oplus T(X_{i+1} \oplus X_{i+2} \oplus X_{i+3} \oplus rk_i) \end{aligned} \tag{4}$$

After 32 rounds of encryption transformation, four 4-byte data outputs $(X_{32}, X_{33}, X_{34}, X_{35})$ are obtained, and the final encryption result is arranged in reverse order:

$$(Y_0, Y_1, Y_2, Y_3) = (X_{35}, X_{34}, X_{33}, X_{32}) \tag{5}$$

The decryption transformation of SM4 algorithm has the same structure and algorithm flow as the encryption transformation, except that the use order of round key is opposite to the encryption process.

When encrypting, the use order of round key is:

$$rk_0, rk_1, \ldots, rk_{31};$$

When decrypting, the use order of round key is:

$$rk_{31}, rk_{30}, \ldots, rk_0;$$

## 2 System Design

### 2.1 Overall Frame of the System

The scheme provided in this paper mainly involves three subjects:

Information Generator: Generates information containing private content that needs to be uploaded to IPFS.

Trusted Agent: After obtaining the authorization of the information generator, SM4 algorithm is used for encryption.

Information Requester: Request the required information from IPFS and convert it into corresponding plaintext information using SM4 decryption algorithm.

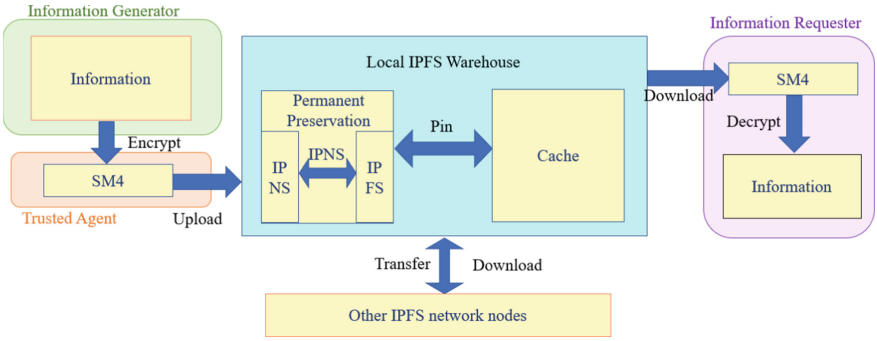


**Fig. 3.** System flow chart.

### 2.2 Scheme Description

The scheme proposed in this paper is shown in the Fig. 3.

In the information uploading stage, the information generator generates information and provides it to the trusted agent. The trusted agent encrypts the plaintext information with SM4 national secret algorithm, and then uploads it to the local IPFS warehouse.

The information stored in IPFS warehouse, whether permanently stored or cached, is encrypted. The communication between the local IPFS warehouse and other IPFS network nodes, that is, the information used during conversion or download, is also encrypted.

In the process of retrieving information from IPFS, first of all, we need to ensure that the requester of information is completely trusted. After the trust is established, the information requester can obtain the encrypted information it wants from IPFS and download it. Since the information requester is fully trusted, it is allowed to obtain the round key usage order of SM4 when the trusted agent uses encryption. The decryption transformation and encryption transformation of SM4 algorithm have the same structure and algorithm flow, and the difference is only the round key. The order of use is opposite to the encryption process. Therefore, the information requester can decrypt the information into the corresponding plaintext by calling the opposite key use sequence.

## 3 Evaluation

### 3.1 Algorithm Security and Performance

The encryption algorithm used in this system is the symmetric encryption algorithm SM4 in the national secret series algorithm. For symmetric key encryption algorithm, its encryption and decryption speed is higher than asymmetric key encryption algorithm in practical application. Therefore, based on the application of this system and the fast query speed of IPFS, the symmetric key encryption algorithm is selected.

The commonly used symmetric key encryption algorithms in the world include DES, [10] AES, [3] 3DES, [2] etc. DES algorithm adopts 64-bit packet length and 56-bit key length, [5] but it is finally replaced by AES algorithm because the key length is too short

| | DES | SM4 |
|---|---|---|
| Basis of calculation | Binary | Binary |
| Algorithm structure | Use standard arithmetic and logic operations, replace first and then replace, without nonlinear transformation | nonlinear transformation |
| Is the encryption and decryption algorithm the same | yes | Yes |
| Calculate the number of rounds | 16 rounds (3DES is 16 rounds × 3) | 32 rounds |
| Packet length | 64 bits | 128 bits |
| Key length | 64 bits (3DES is 128 bits) | 128 bits |
| The effective key length | 56 bits (3DES is 112 bits) | 128 bits |
| Implementation difficulty | Easy to implement | Easy to Implement |
| Achieve performance | Slow software and fast hardware | Fast software and hardware |
| Security | Lower (3DES higher) | Higher |

**Fig. 4.** Comparison of DES and SM4.

to resist cryptographic attacks. SM4 algorithm and AES algorithm have the same key length and packet length. Both packet length and key length are 128 bits. The principle of the algorithm is public and easy to implement, and the implementation process of the algorithm is more simplified than AES algorithm. Therefore, SM4 algorithm is superior to DES algorithm in security, equivalent to AES algorithm in anti-password attack level, and the algorithm is simple and easy to implement, which is suitable for being selected as the algorithm of data encryption. The main characteristics of DES algorithm and SM4 algorithm are compared as follows:

### 3.2 Process Security

When the information requester requests to query the information stored on IPFS, he needs to be trusted by the whole system. After being trusted, the trusted agent will give the information requestor the order of use of the round key used in the encryption process. The information requester decrypts with the round key of the opposite round. In the whole process, the plaintext information is not disclosed to third parties to ensure that the information is shared only within the specified scope of the system (Fig. 4).

## 4 Application Scenario

As shown in Fig. 5, the following will take the speech information storage of social platform as an example to illustrate the possible application scenarios of our proposed solution.

Online rumours refer to rumours fabricated and disseminated using the Internet as a carrier, that is, information fabricated in the Internet space, widely disseminated, unverified or refuted, and causing or likely to cause a certain legal interest to be infringed [4].

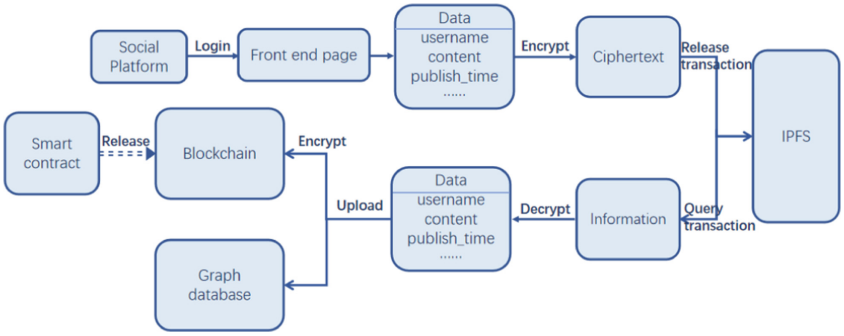


**Fig. 5.** Application scenario example diagram.

Nowadays, we are in the information age, and social media such as Weibo has given ordinary netizens a platform to publish their opinions, so that there are a large number of unproven information posted by netizens on the Internet. Rumours provide more developmental opportunities, faster spread, wider spread and greater real-world impact. Online rumours have the characteristics of confusion, incitement, anonymity, rapid spread, wide influence, dual identity of subject and object, etc. Therefore, managing online rumours has become an important issue to protect people's lives and property, and maintain national and social stability.

In order to timely store online rumours, trace the source, and build a dissemination node map, it is necessary to encrypt and store relevant information on social platforms, such as the publisher's name, and privacy content into IPFS.

In this application scenario, IPFS uses hash-based content addressing. Social platforms are information producers, rumour management platforms uploaded to them are trusted agents, and regulatory authorities such as the government are information requestors.

The social platform uploads the relevant information of the speech to the rumour governance platform, which uses SM4 algorithm to encrypt the content involving privacy, and then uploads it to IPFS. When the government and other information requesters who have obtained the trust of the rumour governance platform need to retrieve the information of the certificate, the rumour governance platform will give them the round key sequence used in SM4 encryption of the information they want to call. By using round keys in reverse order, plaintext information can be successfully obtained.

## 5 Conclusion

IPFS can solve many problems in specific life practice because of its good characteristics such as decentralization, permanent storage of data, and non-tampering of data. Therefore, it has been gradually applied to a wide range of scenarios since its birth. However, because the storage mode of IPFS is designed for all Internet users, allowing all nodes to access at will, and the nodes that have been connected to the IPFS network can freely find content, it is open, so it is not suitable for some scenarios that need to

store privacy messages. Based on the protection of personal privacy information, we propose an IPFS privacy storage and sharing scheme based on SM4 algorithm to meet the needs of privacy protection and privacy information sharing within the specified range in the above scenario. This scheme ensures that the private information in IPFS is not disclosed in plaintext, but stored in IPFS in the form of encryption by SM4 algorithm. Only the trusted information requester can obtain the round key in the encryption process of the requested message, and then decrypt it. It ensures that plaintext information is only disclosed within a specific range, and better privacy protection. Based on the characteristics of IPFS technology and a wide range of application scenarios, the needs of privacy protection and the needs of information sharing within a certain range, this scheme has broad application prospects.

**Acknowledgements.** The project is funded by National Key Project of China (No. 2020YFB 1005700) and Undergraduate Education and Teaching Reform Project of Nankai University in 2020 (No. NKJG2020255). Besides, we would like to thank other bachelors for their help without reservation.

## References

1. Confais B, Lebre A, Parrein B (2017) An object store service for a fog/edge computing infrastructure based on IPFS and a scale-out NAS. In: 2017 IEEE 1st international conference on fog and edge computing (ICFEC), pp 41–50. https://doi.org/10.1109/ICFEC.2017.13
2. Chen J (2020) Research and application of dynamic encryption system of 3DES algorithm. Appl Single Chip Microcomput Embed Syst 20(08):4–6
3. Gao G, Li Z (2020) Research and design of authentication encryption algorithm based on AES round function. J Netw Inf Secur 6(02):106–115
4. Huang Y (2022) Research on Internet rumor spread and governance in public crisis events. Legal Syst Econ 31(01):9–14
5. Jia W, Zhu L (2020) Implementation of DES encryption algorithm in network communication. Netw Secur Technol Appl 03:34–36
6. Lu S, Su B, Wang P, Mao Y (2016) Overview of SM4 block cipher algorithm. Inf Secur Res 2(11):995–1007
7. Alam S, Kelly M, Nelson ML (2016) Interplanetary wayback: the permanent web archive. In: 2016 IEEE/ACM joint conference on digital libraries (JCDL), pp 273–274
8. Tan H et al (2019) Archival data protection and sharing method based on blockchain. J Softw 30(09):2620–2635. https://doi.org/10.13328/j.cnki.jos.005770
9. Wang L, Guo Y, Zhu Y, Duan Z (2022) Monitoring information privacy protection mechanism based on blockchain and IPFS. Electron Des Eng 30(11):178–182+188. https://doi.org/10.14022/j.issn1674-6236.2022.11.038
10. Xiong Y (2021) Secrecy technology of computer information system based on DES data encryption algorithm. Inf Comput (Theoret Ed) 33(17):60–62